

# Performance Bounds for Polynomial Phase Parameter Estimation with Nonuniform and Random Sampling Schemes

Jonathan A. Legg, *Member, IEEE*, and Douglas A. Gray, *Member, IEEE*

**Abstract**—Estimating the parameters of a cisoid with an unknown amplitude and polynomial phase using uniformly spaced samples can result in ambiguous estimates due to Nyquist sampling limitations. It has been shown previously that nonuniform sampling has the advantage of unambiguous estimates beyond the Nyquist frequency; however, the effect of sampling on the Cramér–Rao bounds is not well known.

This paper first derives the maximum likelihood estimators and Cramér–Rao bounds for the parameters with known, arbitrary sampling times. It then outlines two methods for incorporating random sampling times into the lower variance bounds, describing one in detail.

It is then shown that for a signal with additive white, Gaussian noise the bounds for the estimation with nonuniform sampling tend toward those of uniform sampling. Thus, nonuniform sampling overcomes the ambiguity problems of uniform sampling without incurring the penalty of an increased variance in parameter estimation.

**Index Terms**—Cramér–Rao bound, polynomial phase estimation.

## I. INTRODUCTION

THERE are situations when the parameters of a cisoid (rotating phasor) with unknown amplitude and polynomial phase need to be estimated when the signal's spectrum exceeds the maximum allowable sampling rate. An example is synthetic aperture radar [1], [2], where the radar's pulse repetition frequency (sampling rate) determines the unambiguous Doppler bandwidth. This is chosen to accommodate the spectrum of a stationary target, which has an approximately quadratic phase response. A moving target, on the other hand, has a Doppler shifted response that causes Nyquist ambiguities if the radar transmits at uniform intervals. Under these circumstances, a nonuniform sampling interval with specific properties permits unambiguous estimates; a reconstructed spectrum tends toward that of the underlying analog process [3].

It is useful to know the minimum possible variances of the amplitude and polynomial phase coefficient estimates;

Manuscript received May 21, 1998; revised June 24, 1999. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jean Jacques Fuchs.

J. A. Legg is with Surveillance Systems Division, Defence Science and Technology Organization, Salisbury, Australia. He is also with the Cooperative Research Centre for Sensor Signal and Information Processing, Adelaide, Australia.

D. A. Gray is with the Department of Electrical and Electronic Engineering, University of Adelaide, Australia. He is also with the Cooperative Research Centre for Sensor Signal and Information Processing, Adelaide, Australia.

Publisher Item Identifier S 1053-587X(00)00984-3.

these provide a benchmark with which estimator performance may be compared. They may also show whether or not the desired accuracy is theoretically possible. These variances are provided by the Cramér–Rao bounds, which give a quantitative measure of the ultimate accuracy with which a parameter may be estimated by an unbiased estimator.

Many papers have been written on the subject of polynomial phase parameter estimation, for example, Peleg and Porat [4] introduced the continuous and discrete polynomial phase transforms, and Peleg and Friedler [5] discussed the discrete polynomial phase transform in detail. Friedler and Francos [6] determined the Cramér–Rao bounds for the estimates of parameters in a multicomponent polynomial phase signal in Gaussian noise and presented an estimation algorithm, and Barbarossa *et al.* [7] applied the high-order multilag ambiguity function to synthetic aperture radar. This paper is concerned with theoretical estimation and estimation bounds rather than implementation methods. It differs from previous work in that it incorporates both arbitrary and random sampling times.

Section II describes the model of the signals of interest. Section III derives the maximum likelihood estimators for the unknown parameters in this model. Estimator bias is discussed in Section IV. Cramér–Rao bounds are discussed in Section V for deterministic sampling and Section VI for random sampling. These include a discussion of the optimal sampling scheme and examples of different sampling distributions, respectively. Section VII shows the output from simulations to illustrate the results, and Section VIII concludes the paper.

## II. SIGNAL MODEL

Suppose that the signal

$$y(t) = \rho e^{j2\pi(\phi + \theta_1 t + \theta_2 t^2 + \dots + \theta_K t^K)} + n(t)$$

is sampled at the times  $\{t_i\}$ ,  $i = 1, \dots, N$ , where  $\rho$  is an unknown, positive amplitude,  $j^2 = -1$ ,  $\phi$  and  $\{\theta_k\}$ ,  $k = 1, \dots, K$  are unknown polynomial phase coefficients, and  $n(t)$  is a known, zero mean, stationary, circular, Gaussian noise process. Without loss of generality, let  $\phi$  be restricted to  $0 \leq \phi < 1$ . Note that the unknown signal model parameters are nonrandom.

Throughout this paper, it is assumed that the sampling times are known to the estimator. Unless stated otherwise, they are arbitrary; in Section VI, they are governed by a probability distribution.

The sampled signal can be represented by

$$\mathbf{y}(\mathbf{t}) = \rho e^{j2\pi\phi} \mathbf{v}(\boldsymbol{\theta}, \mathbf{t}) + \mathbf{n}$$

where  $\mathbf{y}(\mathbf{t}) = [y(t_1), y(t_2), \dots, y(t_N)]^T$ . (Boldface indicates vectors or matrices, and the superscript  $T$  denotes the transpose operation.) The signal

$$(\mathbf{v}(\boldsymbol{\theta}, \mathbf{t}))_i = e^{j2\pi(\theta_1 t_i + \theta_2 t_i^2 + \dots + \theta_K t_i^K)}$$

with  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_K]^T$  and  $\mathbf{t} = [t_1, t_2, \dots, t_N]^T$ . The noise vector

$$\begin{aligned} \mathbf{n} &\sim \mathcal{CN}_N(0, \mathbf{R}) \\ &= \frac{1}{\pi^N} |\mathbf{R}^{-1}| \exp(-[\mathbf{x} - \mathbf{E}[\mathbf{x}]]^H \mathbf{R}^{-1} [\mathbf{x} - \mathbf{E}[\mathbf{x}]]) . \end{aligned}$$

$\mathbf{R} = \mathbf{E}[\mathbf{n}\mathbf{n}^H]$  is the noise covariance matrix, the superscript  $H$  denoting the Hermitian conjugate-transpose operator.  $\mathbf{E}[\cdot]$  denotes the expectation operator.

For convenience, the unknowns are stored in the vector  $\mathbf{u} = [\rho, \phi, \boldsymbol{\theta}^T]^T$ . It is assumed that the number of measurements  $N$  exceeds the number of unknowns.

### III. MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood (ML) estimator for these parameters is found by applying classical theory, for example, [8] and [9]. This estimator is of interest since it returns the most likely value of the parameters, given a uniform prior probability, with variances achieving the Cramér–Rao bound. The ML estimate corresponds to the parameters that maximize the probability density function (PDF) of the observations  $p(\mathbf{y}|\mathbf{t}; \mathbf{u})$  (which is termed the likelihood function). This PDF is conditioned on the random sampling times (whose PDF will be specified later) and is parameterized by the vector  $\mathbf{u}$ . Then

$$\begin{aligned} \hat{\mathbf{u}} &= \arg \max_{\mathbf{u}} p(\mathbf{y}|\mathbf{t}; \mathbf{u}) \\ &= \arg \max_{\mathbf{u}} \frac{1}{\pi^N} |\mathbf{R}^{-1}| \exp(-[\mathbf{y} - \rho e^{j2\pi\phi} \mathbf{v}]^H \\ &\quad \cdot \mathbf{R}^{-1} [\mathbf{y} - \rho e^{j2\pi\phi} \mathbf{v}]) \\ &= \arg \min_{\mathbf{u}} (\mathbf{y} - \rho e^{j2\pi\phi} \mathbf{v})^H \mathbf{R}^{-1} (\mathbf{y} - \rho e^{j2\pi\phi} \mathbf{v}) \end{aligned}$$

where  $\mathbf{v} = \mathbf{v}(\boldsymbol{\theta}, \mathbf{t})$ .

For any  $\boldsymbol{\theta}$ , the ML estimator of the amplitude and initial phase

$$\{\hat{\rho}, \hat{\phi}\} = \arg \min_{\rho, \phi} (\mathbf{y} - \rho e^{j2\pi\phi} \mathbf{v})^H \mathbf{R}^{-1} (\mathbf{y} - \rho e^{j2\pi\phi} \mathbf{v}). \quad (1)$$

We can consider the amplitude and initial phase to comprise a complex amplitude with real and imaginary components  $\rho_r$  and  $\rho_i$ . Due to the restrictions placed on  $\rho$  and  $\phi$ , it follows that

$$\hat{\rho} e^{j2\pi\hat{\phi}} = \widehat{\rho e^{j2\pi\phi}} = \hat{\rho}_r + j\hat{\rho}_i$$

and differentiating the right-hand side of (1) with respect to these components results in

$$\hat{\rho}_r = \frac{\Re\{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{y}\}}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}}$$

and

$$\hat{\rho}_i = \frac{\Im\{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{y}\}}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}}$$

where  $\Re\{\cdot\}$  and  $\Im\{\cdot\}$  return the real and imaginary parts, respectively. Since  $\mathbf{R}$  is positive definite,  $\mathbf{R}^{-1}$  is positive definite, which means that  $\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}$  is real. Therefore

$$\hat{\rho} e^{j2\pi\hat{\phi}} = \frac{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{y}}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}} \quad (2)$$

which is the least squares amplitude estimator. If the noise were white with variance  $\sigma^2$  (so that  $\mathbf{R} = \sigma^2 \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix), this would reduce to a matched filter:  $\hat{\rho} e^{j2\pi\hat{\phi}} = \mathbf{v}^H \mathbf{y}$ . If  $\boldsymbol{\theta} = [\theta_1]$ , this would be equivalent to evaluating the discrete time Fourier transform (DTFT) of the sampled signal at the frequency of interest.

The estimator for the polynomial phase coefficients

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}} (\mathbf{y} - \hat{\rho} e^{j2\pi\hat{\phi}} \mathbf{v})^H \mathbf{R}^{-1} (\mathbf{y} - \hat{\rho} e^{j2\pi\hat{\phi}} \mathbf{v}) \\ &= \arg \max_{\boldsymbol{\theta}} \frac{|\mathbf{y}^H \mathbf{R}^{-1} \mathbf{v}|^2}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}} \end{aligned} \quad (3)$$

using (2). Thus,  $\boldsymbol{\theta}$  can be estimated without knowledge of the amplitude or initial phase. Again, in the case of white noise and  $\boldsymbol{\theta} = [\theta_1]$ , this is the frequency that maximizes the squared magnitude of the DTFT. (In practice, a coarse frequency estimation may be performed by finding the maximum squared value of the discrete Fourier transform evaluated at a finite number of frequency bins. Unfortunately, as the signal-to-noise ratio is reduced, it suddenly becomes more likely that a noise spike will be larger than the value in the frequency bin nearest the signal, leading to a gross error. This is known as the ‘‘threshold effect’’ [8].)

The performance of this estimator will be considered in the following sections.

### IV. BIAS

The expected value of the complex amplitude estimate

$$\mathbf{E}[\hat{\rho} e^{j2\pi\hat{\phi}}] = \frac{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{E}[\mathbf{y}]}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}} = \rho e^{j2\pi\phi}$$

and therefore, the estimator (2) is unbiased, irrespective of the sampling times.

Ambiguities result when more than one set of model parameters can give rise to the observations. This is well known with uniform sampling since a cisoid with a constant frequency can be fitted to the data at the signal’s frequency or at multiples of the sampling frequency away from this. These aliases may be interpreted as estimation *biases* since, with no prior knowledge, a practical estimation strategy that searches for cisoids from a pre-defined starting point will consistently take one of them as the true value. (Another possibility, that the estimator has a random starting frequency, is not considered since the resulting error, which is a variance, depends on the strategy adopted. For the evaluation of the variance, it is assumed that the correct alias is chosen.) These biases will now be considered in detail.

The expected value of the polynomial phase coefficient estimator (3) is

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\theta}}] &= \mathbb{E} \left[ \arg \max_{\boldsymbol{\theta}} \frac{\mathbf{y}^H \mathbf{R}^{-1} \mathbf{v}}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}} \right] \\ &= \mathbb{E} \left[ \arg \max_{\boldsymbol{\theta}} \frac{1}{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}} \right. \\ &\quad \cdot \sum_{k,l} \left[ \rho e^{-j2\pi(\phi + \theta_1 t_k + \theta_2 t_k^2)} + \mathbf{n}_k \right] \\ &\quad \left. \cdot \mathbf{R}_{k,l}^{-1} \hat{\rho} e^{j2\pi(\hat{\phi} + \hat{\theta}_1 t_l + \hat{\theta}_2 t_l^2)} \right]^2 \end{aligned}$$

where only a second-order polynomial phase is being considered for brevity.

For white noise, this becomes

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\theta}}] &= \mathbb{E} \left[ \arg \max_{\boldsymbol{\theta}} \left| \sum_k \rho e^{j2\pi[\hat{\phi} - \phi + (\hat{\theta}_1 - \theta_1)t_k + (\hat{\theta}_2 - \theta_2)t_k^2]} \right. \right. \\ &\quad \left. \left. + \mathbf{n}_k e^{j2\pi(\hat{\phi} + \hat{\theta}_1 t_k + \hat{\theta}_2 t_k^2)} \right|^2 \right] \\ &= \mathbb{E} \left[ \arg \max_{\boldsymbol{\theta}} \left| \sum_k \rho e^{j2\pi[(\hat{\theta}_1 - \theta_1)t_k + (\hat{\theta}_2 - \theta_2)t_k^2]} + \mathbf{n}'_k \right|^2 \right] \end{aligned}$$

where the noise term  $\mathbf{n}'_k = \mathbf{n}_k e^{j2\pi(\hat{\phi} + \hat{\theta}_1 t_k + \hat{\theta}_2 t_k^2)}$  has the same statistics as  $\mathbf{n}_k$ . (This is because the amplitude is unchanged, and  $\mathbf{n}_k$  has a uniformly distributed random phase.)

For uniform sampling, where  $t_k = kT$ , the sum of the exponentials will reach a maximum when the exponential's arguments are multiples of  $j2\pi$ . This will happen when both

$$\left. \begin{aligned} \hat{\theta}_1 &= \theta_1 + x/T \\ \text{and } \hat{\theta}_2 &= \theta_2 + y/T^2 \end{aligned} \right\} \quad (4)$$

for arbitrary integers  $x$  and  $y$ . The estimated spectrum will reach the same value in each of these cases. Thus, *the estimator is biased*.

Any nonuniform or random sampling scheme will suppress these effects since the phases will then not be aligned. However, it may be important to suppress aliases as much as possible. This is especially true when there are several signals, which is a case not considered here. Bilinskis and Mikelsons [3, pp. 80-88] discuss the degree of aliasing, which is defined as the squared sum of the Fourier coefficients of the signal at aliasing frequencies, for a variety of different random sampling schemes. They show that random intervals with a positive correlation perform better than with no correlation or periodic sampling with jitter. (In general, sampling schemes that approximate a uniform probability of a sample occurring as a function of time have the best properties.) These schemes will be discussed later.

## V. CRB'S FOR DETERMINISTIC SAMPLING

There is a requirement for a figure of merit for the performance of the estimator (3) that takes into account the bias issues discussed above (4), such as the mean square error. However, as mentioned previously, the bias achieved depends on the specific implementation of (3); therefore, it will be assumed here that the correct alias is chosen. The mean square error is then equivalent to the variance.

The lower variance bound for an unbiased estimator, the Cramér–Rao bound<sup>1</sup> is given by

$$\text{CRB}[(\hat{\mathbf{u}})_k] = (\mathbf{J}^{-1})_{k,k} \quad (5)$$

(e.g., [10]), where the Fisher information matrix  $\mathbf{J}$  has the elements

$$(\mathbf{J})_{k,l} \triangleq \mathbb{E} \left[ \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_k} \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_l} \right]. \quad (6)$$

It can be shown that for  $\mathbf{y} \sim \mathcal{CN}(\mathbf{m}(\mathbf{u}), \mathbf{R})$ , (6) simplifies to

$$\begin{aligned} (\mathbf{J})_{k,l} &= 2\Re \left\{ \frac{\partial \mathbf{m}(\mathbf{u})^H}{\partial (\mathbf{u})_k} \mathbf{R}^{-1} \frac{\partial \mathbf{m}(\mathbf{u})}{\partial (\mathbf{u})_l} \right\} \\ &= 2\Re \left\{ \frac{\partial (\rho e^{j2\pi\phi} \mathbf{v})^H}{\partial (\mathbf{u})_k} \mathbf{R}^{-1} \frac{\partial (\rho e^{j2\pi\phi} \mathbf{v})}{\partial (\mathbf{u})_l} \right\} \end{aligned}$$

for this problem [2, App. D], where  $(\mathbf{v})_i = e^{j2\pi(\theta_1 t_i + \theta_2 t_i^2 + \dots + \theta_K t_i^K)}$ . All parameters are assumed to be unknown, and therefore,  $\mathbf{u} = [\rho, \phi, \boldsymbol{\theta}^T]^T$ .

Here, we have, using vector notation for brevity

$$\begin{aligned} \frac{\partial \rho e^{j2\pi\phi} \mathbf{v}}{\partial \rho} &= e^{j2\pi\phi} \mathbf{v} \\ \frac{\partial \rho e^{j2\pi\phi} \mathbf{v}}{\partial \phi} &= j2\pi \rho e^{j2\pi\phi} \mathbf{v} \end{aligned}$$

and

$$\frac{\partial \rho e^{j2\pi\phi} \mathbf{v}}{\partial \theta_n} = j2\pi \rho e^{j2\pi\phi} \mathbf{T}^n \mathbf{v}$$

where  $\mathbf{T}$  is a matrix containing the sampling times

$$\mathbf{T} = \begin{bmatrix} t_1 & & & 0 \\ & t_2 & & \\ & & \ddots & \\ 0 & & & t_N \end{bmatrix}.$$

It can then be shown that

$$\left. \begin{aligned} (\mathbf{J})_{1,1} &= 2\Re\{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{v}\} \\ (\mathbf{J})_{k+2,1} &= 4\pi \rho \Im\{\mathbf{v}^H \mathbf{T}^k \mathbf{R}^{-1} \mathbf{v}\}, & k \geq 0 \\ (\mathbf{J})_{1,l+2} &= -4\pi \rho \Im\{\mathbf{v}^H \mathbf{R}^{-1} \mathbf{T}^l \mathbf{v}\}, & l \geq 0 \\ (\mathbf{J})_{k+2,l+2} &= 8\pi^2 \rho^2 \Re\{\mathbf{v}^H \mathbf{T}^k \mathbf{R}^{-1} \mathbf{T}^l \mathbf{v}\}, & k, l \geq 0. \end{aligned} \right\} \quad (7)$$

<sup>1</sup>It is well known that the Cramér–Rao bounds represent the minimum variance with which a parameter may be estimated by an unbiased estimator. In this case, there is the potential for additive bias, as illustrated by (4). However, an additive offset will not affect the variance of a random variable so the lower variance bound is still the Cramér–Rao bound, despite the bias.

For white noise and  $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ , this is<sup>2</sup>

$$\mathbf{J} = \frac{1}{\sigma^2} \begin{bmatrix} 2N & 0 & 0 & 0 \\ 0 & 8\pi^2 \rho^2 N & 8\pi^2 \rho^2 \sum_k t_k & 8\pi^2 \rho^2 \sum_k t_k^2 \\ 0 & 8\pi^2 \rho^2 \sum_k t_k & 8\pi^2 \rho^2 \sum_k t_k^2 & 8\pi^2 \rho^2 \sum_k t_k^3 \\ 0 & 8\pi^2 \rho^2 \sum_k t_k^2 & 8\pi^2 \rho^2 \sum_k t_k^3 & 8\pi^2 \rho^2 \sum_k t_k^4 \end{bmatrix}. \quad (8)$$

Since absolute times  $\{t_k\}$  are referred to, it is necessary to specify a reference relative to the origin  $t_0$ . This will be considered next for the cases of uniform and nonuniform sampling.

### A. Uniform Sampling

If we let the sampling times be centered on  $t_0$ , we have

$$t_k = \left[ (k-1) - \frac{N-1}{2} \right] T + t_0, \quad k = 1, \dots, N \quad (9)$$

where  $T$  is the sampling interval.

Some of the Cramér–Rao bounds depend on  $t_0$ , which is not intuitive. The effect was observed for the simpler problem of estimating both  $\phi$  and  $\theta_1$  by Rife *et al.* [8, p. 592], who stated that the minimum Cramér–Rao bound on  $\hat{\phi}$  was obtained by choosing  $t_0 = -((N-1)/2)T$ . When (9) is used in (5) and (8), the lower variance bound for phase reaches a minimum when the sampling times are centered on  $t_0 = \pm \sqrt{T^2(N^2+1)/20}$ , which is a significant departure from 0. Although this is an interesting phenomenon, it will not be investigated further here.

The most appropriate  $t_0$  to use may be when as many cross terms in the Fisher matrix (8) as possible are forced to 0. This happens when  $t_0 = 0$ . Elements with sums of  $t$  to odd powers become zero, and therefore, the bounds become

$$\left. \begin{aligned} \text{CRB}[\hat{\rho}] &= \sigma^2 / (2N) \\ \text{CRB}[\hat{\phi}] &= \frac{3}{32\pi^2} \frac{\sigma^2}{\rho^2} \frac{3N^2 - 7}{N(N^2 - 4)} \\ \text{CRB}[\hat{\theta}_1] &= \frac{3}{2\pi^2} \frac{\sigma^2}{\rho^2} \frac{1}{T^2 N(N^2 - 1)}, \text{ and} \\ \text{CRB}[\hat{\theta}_2] &= \frac{45}{2\pi^2} \frac{\sigma^2}{\rho^2} \frac{1}{T^4 N(N^4 - 5N^2 + 4)}. \end{aligned} \right\} \quad (10)$$

### B. Nonuniform Sampling

For nonuniform sampling, we have

$$t_k = \sum_{i=0}^{k-1} \tau_i, \quad k = 1, \dots, N$$

where  $\tau_k$  is the interval between sampling times  $t_k$  and  $t_{k+1}$ ,  $\tau_0$  being the offset of  $t_1$  from the origin. Simultaneously zeroing

<sup>2</sup>The second-order case is of interest due to its relevance to synthetic aperture radar, where the distance to a target and, hence, the phase of its response may be approximated by a quadratic.

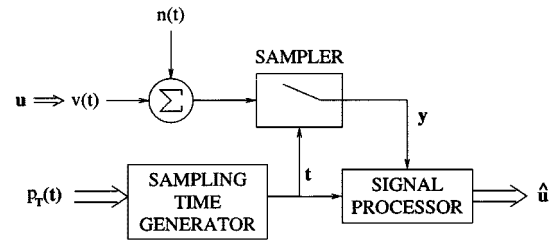


Fig. 1. Representation of the random sampling parameter estimation scenario. The sampling times are randomly distributed and assumed to be known to the signal processor.

the off-diagonal terms in (8) is a problem, for example, for the sum of the times to be zero, we have

$$\tau_0 = \frac{1}{N} \sum_{i=1}^{N-1} (i - N) \tau_i \quad (11)$$

so that the origin is positioned at the “center of gravity” of the times. Other criteria, such as setting the sum of cubed times to be zero [which corresponds to another off-diagonal term in (8)], gives different results to (11) in general. The value of  $\tau_0$  to be used is therefore not clear; the global minimum of the Cramér–Rao bounds may need to be found.

### C. On the Optimal Sampling Scheme

Apart from [11], little work has been reported on the problem of optimum sampling of polynomial phase signals. In [11], it is conjectured that the optimal sampling strategy is to cluster samples into  $K + 1$  groups evenly spaced across the sampling interval where  $K$  is the order of the polynomial phase. However, this strategy also imposes the physically unrealizable condition that samples within a cluster must be infinitely close together.

Further research, particularly considering the effect of noise, is required. One approach, for example, would be to maximize the determinant of the Fisher matrix (8) with respect to the sampling times.

## VI. CRB'S FOR RANDOM SAMPLING

Provided the sampling times are fixed, the Cramér–Rao bounds in the colored Gaussian noise case can be found using (7) and (5). However, if the times are specified in terms of distributions  $p_{\mathbf{T}}(\mathbf{t})$  rather than specific instants, it is desirable to know the *expected* Cramér–Rao bounds. *It is assumed that the actual times are known for estimating the parameters  $\mathbf{u}$ , but differ from one realization to the next.* This situation is shown in Fig. 1. The Cramér–Rao bounds for a particular realization can be found from the results of the previous section. However, what may also be required is a bound for the variance over the ensemble of realizations. Two possible approaches will be discussed here.

- 1) Find the expectation of the bounds themselves with respect to the timing distributions

$$E_{\mathbf{T}}[\text{CRB}[(\mathbf{u})_i]] = \int_{\mathbf{T}} \text{CRB}[(\mathbf{u})_i; \mathbf{t}] p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t}. \quad (12)$$

This is an intuitive method; however, its true meaning is not clear, and we do not further pursue this.

- 2) Since the likelihood function now contains random times,  $p_{\mathbf{T}}(\mathbf{t})$  should be incorporated in to the Fisher information matrix (6) giving

$$(\mathbf{J}')_{k,l} \triangleq \mathbb{E}_{\mathbf{Y},\mathbf{T}} \left[ \frac{\partial \ln p(\mathbf{y}, \mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_k} \frac{\partial \ln p(\mathbf{y}, \mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_l} \right]. \quad (13)$$

$p(\mathbf{y}, \mathbf{t}; \mathbf{u})$  is the joint probability density function of the sampled signal and the sampling times, parameterized by the unknowns. Note that the expectation is over the sample spaces of both the random measurements  $\mathbf{y}$  and the random measurement times  $\mathbf{t}$ .

We are now considering approach 2. Equation (13) can be simplified. Using a conditional probability expression

$$p(\mathbf{y}, \mathbf{t}; \mathbf{u}) = p(\mathbf{y}|\mathbf{t}; \mathbf{u}) p_{\mathbf{T}}(\mathbf{t}; \mathbf{u})$$

and  $p_{\mathbf{T}}(\mathbf{t}; \mathbf{u}) = p_{\mathbf{T}}(\mathbf{t})$  since the sampling times do not depend upon the parameters; then, from (13)

$$\begin{aligned} (\mathbf{J}')_{k,l} &= \mathbb{E}_{\mathbf{Y},\mathbf{T}} \left[ \frac{\partial \ln(p(\mathbf{y}|\mathbf{t}; \mathbf{u}) p_{\mathbf{T}}(\mathbf{t}))}{\partial (\mathbf{u})_k} \frac{\partial \ln(p(\mathbf{y}|\mathbf{t}; \mathbf{u}) p_{\mathbf{T}}(\mathbf{t}))}{\partial (\mathbf{u})_l} \right] \\ &= \mathbb{E}_{\mathbf{Y},\mathbf{T}} \left[ \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u}) + \ln p_{\mathbf{T}}(\mathbf{t})}{\partial (\mathbf{u})_k} \right. \\ &\quad \left. \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u}) + \ln p_{\mathbf{T}}(\mathbf{t})}{\partial (\mathbf{u})_l} \right] \\ &= \mathbb{E}_{\mathbf{Y},\mathbf{T}} \left[ \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_k} \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_l} \right] \\ &= \int_{\mathbf{T}} \int_{\mathbf{Y}} \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_k} \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_l} \\ &\quad \cdot p(\mathbf{y}, \mathbf{t}; \mathbf{u}) d\mathbf{y} d\mathbf{t} \\ &= \int_{\mathbf{T}} \left[ \int_{\mathbf{Y}} \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_k} \frac{\partial \ln p(\mathbf{y}|\mathbf{t}; \mathbf{u})}{\partial (\mathbf{u})_l} \right. \\ &\quad \left. \cdot p(\mathbf{y}|\mathbf{t}; \mathbf{u}) d\mathbf{y} \right] p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \\ &= \int_{\mathbf{T}} (\mathbf{J})_{k,l} p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \\ &= \mathbb{E}_{\mathbf{T}}[(\mathbf{J})_{k,l}] \end{aligned}$$

where  $\mathbf{J}$  was defined in (6). The final result is similar in concept to (12), except that the expectation is over the Fisher matrix elements rather than the Cramér–Rao bounds.

$\mathbf{J}$ 's elements (7) can be written in the form

$$(\mathbf{J})_{k,l} = \alpha(k, l) f\{\mathbf{w}^H(k) \mathbf{R}^{-1} \mathbf{x}(l)\}$$

where

$$\alpha(k, l) = \begin{cases} 2, & k = l = 1 \\ 4\pi\rho, & k > 1, l = 1 \\ -4\pi\rho, & k = 1, l > 1 \\ 8\pi^2\rho^2, & k, l > 1. \end{cases}$$

$f\{\cdot\}$  extracts the real or imaginary part, and  $\mathbf{w}(k)$  and  $\mathbf{x}(k)$  are of the form  $\mathbf{T}^{k-2} \mathbf{v}$  with  $\mathbf{w}(1) = \mathbf{x}(1) = \mathbf{v}$ ; therefore

$$(\mathbf{J}')_{k,l} = \alpha(k, l) f \left\{ \int_{\mathbf{T}} \mathbf{w}^H \mathbf{R}^{-1} \mathbf{x} p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \right\}. \quad (14)$$

This cannot readily be simplified since  $(\mathbf{R})_{m,n} = r(t_m - t_n)$ ; it is a function of the variables being integrated over.

In the case of white noise,  $\mathbf{R} = \sigma^2 \mathbf{I}$ , and (14) becomes

$$\begin{aligned} (\mathbf{J}')_{k,l} &= \frac{\alpha(k, l)}{\sigma^2} f \left\{ \int_{\mathbf{T}} \mathbf{v}^H \mathbf{T}^{g(k,l)} \mathbf{v} p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \right\} \\ &= \frac{\alpha(k, l)}{\sigma^2} f \left\{ \sum_n \int_{\mathbf{T}} (\mathbf{v})_n^* t_n^{g(k,l)} (\mathbf{v})_n p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \right\} \\ &= \frac{\alpha(k, l)}{\sigma^2} f \left\{ \sum_n \int_{\mathbf{T}} t_n^{g(k,l)} p_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \right\} \end{aligned}$$

since the phase terms cancel. The superscript  $*$  denotes the conjugation operation, and  $g(k, l)$  expresses the powers of  $t$

$$g(k, l) = \begin{cases} 0, & k = l = 1 \\ k + l - 4, & k, l \geq 2 \\ \text{undefined,} & \text{otherwise.} \end{cases}$$

If the PDF's are independent, we have

$$\begin{aligned} (\mathbf{J}')_{k,l} &= \frac{\alpha(k, l)}{\sigma^2} f \left\{ \sum_n \int_{T_N} \cdots \int_{T_2} \int_{T_1} t_n^{g(k,l)} \right. \\ &\quad \left. \cdot \prod_m p_{T_m}(t_m) dt_1 dt_2 \cdots dt_N \right\} \\ &= \frac{\alpha(k, l)}{\sigma^2} f \left\{ \sum_n \int_{T_n} t_n^{g(k,l)} p_{T_n}(t) dt \right\} \end{aligned}$$

and

$$(\mathbf{J}')_{k,l} = \frac{\alpha(k, l)}{\sigma^2} N f \left\{ \int_{\mathbf{T}} t^{g(k,l)} p_{\mathbf{T}}(t) dt \right\}$$

if the PDF's are the same. (Note that the argument of  $f$  is the  $g(k, l)$ th moment of the PDF of  $t$ .)

#### A. Example Sampling Distributions

1) *Random Offsets:* If there are random offsets from uniformly spaced samples centered on  $t = 0$  (i.e., jitter), we have

$$t_n = \left( n - \frac{N-1}{2} \right) T + u_n, \quad n = 0, \dots, N-1$$

where  $u_n \sim \mathcal{U}_t(0, bT)$  and

$$\mathcal{U}_x(\mu, s) \triangleq \begin{cases} 1/(2s), & \mu - s \leq x \leq \mu + s \\ 0, & \text{otherwise} \end{cases}$$

for uniformly distributed offsets. Then

$$p_{T_n}(t) = \mathcal{U}_t([n - (N-1)/2]T, bT).$$

It can be shown that

$$\left. \begin{aligned} (\mathbf{J}')_{1,1} &= 2N/\sigma^2 \\ (\mathbf{J}')_{k+2,1} &= 0 \\ (\mathbf{J}')_{1,k+2} &= 0, \text{ and} \\ (\mathbf{J}')_{k+2,l+2} &= \frac{8\pi^2\rho^2}{\sigma^2} \frac{1}{2b(k+l+1)} \\ &\quad \cdot \sum_{n=0}^{N-1} \{ [n - (N-1)/2]T + b \}^{k+l+1} \\ &\quad - \{ [n - (N-1)/2]T - b \}^{k+l+1}. \end{aligned} \right\} \quad (15)$$

2) *Uniformly Distributed Over Whole Interval:* In this case, for all  $n$

$$p_{T_n}(t) = \begin{cases} 1/(2a), & -a \leq t \leq a \\ 0, & \text{otherwise.} \end{cases}$$

This distribution is impractical since it requires that some samples occur with intervals that are unrealistically small. However, it may be approximated in practice by the use of random intervals between samples; the distribution of the intervals is of no real importance when  $N$  is large and may incorporate a minimum value. (Bilinskis *et al.* [3] discuss this, methods of generating the sequences, and the relative merits of different sampling schemes in detail.)

It can be shown that

$$\begin{aligned} (\mathbf{J}')_{1,1} &= 2N/\sigma^2 \\ (\mathbf{J}')_{k+2,1} &= 0 \\ (\mathbf{J}')_{1,k+2} &= 0 \end{aligned}$$

and

$$(\mathbf{J}')_{k+2,l+2} = \begin{cases} 8\pi^2 \frac{\rho^2}{\sigma^2} \frac{Na^{k+l}}{(k+l+1)}, & k+l \text{ even} \\ 0, & \text{otherwise.} \end{cases}$$

This is the same as the result achieved with random offsets (15), where the offsets are  $\pm 0.5T$ . However, it is preferable to use random intervals in practice. The intervals may be constrained to have a minimum sampling time separation, unlike with random offsets, and there need only be a small degree of randomness; a large amount, such as  $\pm 0.5T$ , leads to a degree of unpredictability in performance [3].

If  $a = (N-1)T/2$ , making the time span the same as for a uniform sequence, and  $\theta = [\theta_1, \theta_2]^T$ , the corresponding Cramér–Rao bounds are

$$\left. \begin{aligned} \text{CRB}[\hat{\rho}] &= \sigma^2/(2N) \\ \text{CRB}[\hat{\phi}] &= \frac{9}{32\pi^2} \frac{\sigma^2}{\rho^2} \frac{1}{N} \\ \text{CRB}[\hat{\theta}_1] &= \frac{3}{2\pi^2} \frac{\sigma^2}{\rho^2} \frac{1}{T^2 N(N-1)^2} \\ \text{and} \\ \text{CRB}[\hat{\theta}_2] &= \frac{45}{2\pi^2} \frac{\sigma^2}{\rho^2} \frac{1}{T^4 N(N-1)^4}. \end{aligned} \right\} \quad (16)$$

This is equivalent to the bounds for uniform sampling (10) for large  $N$  and slightly larger for small  $N$ . Increasing the interval  $a$  will not make them the same; the effects are nonlinear.

Fig. 2 shows a comparison between the Cramér–Rao bounds corresponding to uniform sampling (10), the nonuniform sampling strategy discussed in Section V-C, and uniformly distributed random sampling (16) as a function of the number of samples  $N$ .  $\sigma^2 = 1$ ,  $\rho^2 = 1$ , and  $T = 1$ . There is very little difference between the uniform and uniformly distributed random schemes when  $N > 15$ . For  $\theta_1$  and  $\theta_2$ , the scheme of Section V-C consistently out performs the other two.

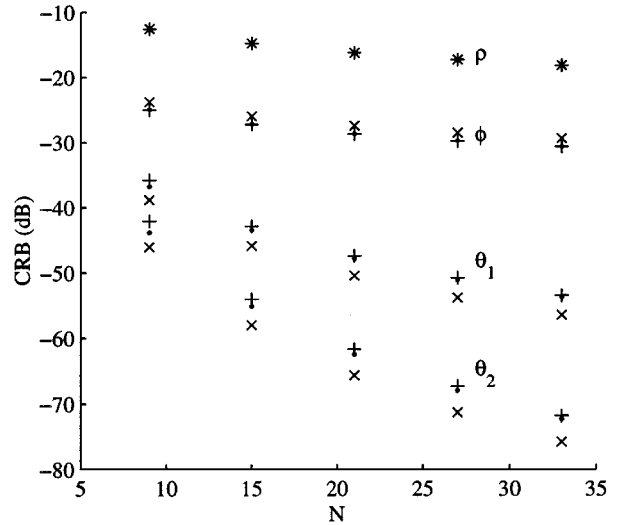


Fig. 2. Cramér–Rao bounds for uniform sampling ( $\cdot$ ), “optimum” nonuniform deterministic sampling ( $\times$ ), and uniformly distributed random sampling ( $+$ ) with normalized parameters.

3) *Known Times:* As a final example, nonrandom sampling times can be treated as a limiting case of random sampling with

$$p_{T_n}(t) = \delta(t - t_n) \forall n$$

where  $\delta(t)$  is the Dirac delta function. It can easily be shown that  $\mathbf{J}' = \mathbf{J}$ , which is the Fisher matrix discussed in Section V.

## VII. SIMULATIONS

Fig. 3 shows the estimation errors obtained as a function of the noise variance when the four parameters  $\rho$ ,  $\phi$ ,  $\theta_1$ , and  $\theta_2$  were estimated with simulated data and uniform sampling. The signal’s magnitude  $\rho$  was 1, the number of samples  $N$  was 20, the sampling interval  $T$  was 2, the phase parameter  $\phi$  was 0.123, the frequency  $\theta_1$  was 0.106, and the frequency rate  $\theta_2$  0.01. 15 estimations were performed for each noise variance; the vertical bars in the figure show the corresponding 99% confidence intervals. The noise values were the same for each of the three simulations.

Fig. 4 shows the errors obtained during a similar simulation using uniformly distributed sampling. The sampling times were the same for each of the three noise variances. The sample variances are very similar to the Cramér–Rao bounds.

## VIII. CONCLUSION

It was shown that under the right conditions, the parameters of a polynomial phase signal can be estimated unambiguously with nonuniform sampling to the same accuracy as with uniform sampling but without the bias errors due to aliasing effects. This permits the accurate, unambiguous measurement of signal parameters beyond the Nyquist limits encountered with uniform sampling.

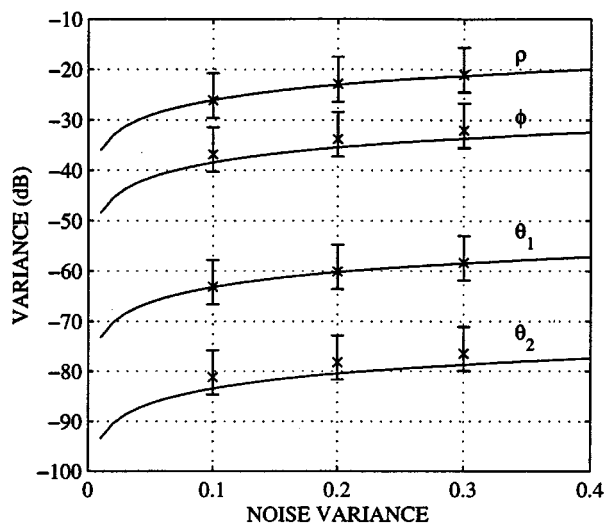


Fig. 3. Estimation errors found by simulation with uniform sampling. The solid curves show the Cramér-Rao lower variance bounds.

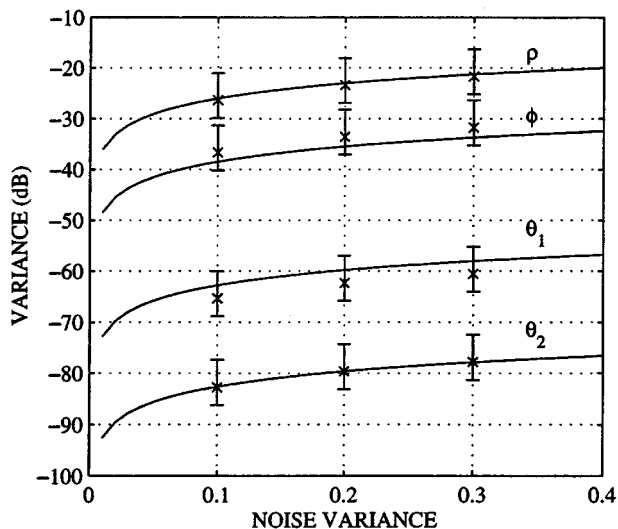


Fig. 4. Estimation errors found using uniformly distributed random sampling.

ACKNOWLEDGMENT

The authors thank the reviewers of this paper, who made valuable suggestions regarding its contents.

REFERENCES

[1] J. A. Legg, A. G. Bolton, and D. A. Gray, "SAR moving target detection using a nonuniform PRI," in *Proc. EUSAR*, March 1996, VDE-Verlag GmBH, pp. 423–426.

[2] J. A. Legg, "Synthetic aperture radar using nonuniform sampling," Ph.D. dissertation, Univ. Adelaide, Adelaide, Australia, Aug. 1997.  
 [3] I. Bilinskis and A. Mikelsons, *Randomized Signal Processing*. London, U.K.: Prentice-Hall, 1992.  
 [4] S. Peleg and B. Porat, "Estimation and classification of polynomial-phase signals," *IEEE Trans. Inform. Theory*, vol. 37, pp. 422–430, Mar. 1991.  
 [5] S. Peleg and B. Friedler, "The discrete polynomial-phase transform," *IEEE Trans. Signal Processing*, vol. 43, pp. 1901–1914, Aug. 1995.  
 [6] B. Friedler and J. M. Francos, "Estimation of amplitude and phase parameters of multicomponent signals," *IEEE Trans. Signal Processing*, vol. 43, pp. 917–926, Apr. 1995.  
 [7] S. Barbarossa, A. Porchia, and A. Scaglione, "Multiplicative multi-lag high order ambiguity function," in *Proc. ICASSP*, May 1996, pp. 3022–3025.  
 [8] D. C. Rife and R. R. Boorsteyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591–598, Sept. 1974.  
 [9] S. Peleg and B. Porat, "The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase," *IEEE Trans. Signal Processing*, vol. 39, pp. 749–752, Mar. 1991.  
 [10] L. L. Scharf, *Statistical Signal Processing; Detection, Estimation and Time Series Analysis*. Sydney, Australia: Addison-Wesley, 1991.  
 [11] J. Ångeby, "On identifiability, aliasing and the optimal sampling strategy for polynomial phase signals," in *Proc. SAMPTA Int. Workshop Sampling Theory Appl.*, June 1997, pp. 217–222.



maritime sensor fusion.

**Jonathan A. Legg** (M'88) received the B.Eng. degree in electronic engineering with distinction from the South Australian Institute of Technology, Adelaide, in 1988 and the Ph.D degree in radar signal processing from the University of Adelaide in 1997.

He has been with the Surveillance Systems Division of the Australian Defence Science and Technology Organization, Salisbury, since 1989 developing hardware and software for measurement radars and contributing to synthetic aperture radar signal processing. His more recent interests include



**Douglas A. Gray** (M'82) received the B.Sc. and Ph.D. degrees in mathematical physics from the University of Adelaide, Adelaide, Australia, in 1969 and 1974, respectively.

From 1973 to 1993, he was with the Australian Defence Science and Technology Organization, Salisbury, applying signal processing to sonar and electronic warfare. From 1977 to 1979, he was a Visiting Scientist at the Royal Aircraft Establishment, London, U.K., and in 1985, he was a Visiting Fellow at the Australian National University, Canberra. He is currently Professor of sensor signal processing and Deputy Director of the Cooperative Research Centre for Sensor Signal and Information Processing at the University of Adelaide. His research interests are in the application of signal processing to sonar, radar, and electronic warfare, particularly in adaptive processes, beam forming, and signal sorting and classification techniques.