Estimating and Interpreting the Instantaneous Frequency of a Signal—Part 2: Algorithms and Applications

BOUALEM BOASHASH, SENIOR MEMBER, IEEE

This paper, which addresses the important issue of estimating the instantaneous frequency (IF) of a signal, is a sequel to the paper which appears in this issue, and dealt with the concepts relating to the IF. In this paper the concept of IF is extended to be able to cope with discrete time signals. The specific problem explored is that of estimating the IF of frequency modulated (FM) discrete-time signals imbedded in Gaussian noise. There are many well established methods for estimating the IF—these methods include differentiation of the phase and smoothing thereof, adaptive frequency estimation techniques such as the phase locked loop (PLL), and extraction of the peak from time-varying spectral representations. More recently methods based on a modeling of the signal phase as a polynomial have been introduced. All of these methods are reviewed, and their performances are compared on both simulated and real data. Guidelines are given as to which estimation method should be used for a given signal class and signal-to-noise ratio (SNR).

I. INTRODUCTION

The estimation of the instantaneous frequency (IF) is a natural progression from steady-state sinusoidal frequency estimation which has been studied extensively for many years. The various methods which have been developed for obtaining good sinusoidal frequency estimates are briefly described in Section II as background information. For nonstationary frequency modulated (FM) signals, the FM law may be considered to be a continuum of different frequencies, with the frequency at a particular time being described well by the concept of "instantaneous frequency." These nonstationary signals are common both in nature and in many man-made processing environments. There is a great need, then, to develop effective techniques for IF estimation.

Section III includes a review of currently available IF estimation techniques. The coverage includes phase differencing of the analytic signal and smoothing thereof, counting the zero crossings, adaptive estimation methods based on the least mean square (LMS) algorithm, or the recursive least squares (RLS) algorithm, phase-locked loop techniques, estimation via time-frequency distribution (TFD) moments or peaks [20], and methods based on a modeling of the signal phase law as a polynomial. Only the single FM component case will be dealt with in detail, but a short description of how one may extend the algorithms to multicomponent signals will be provided.

Simulations are provided in Section IV to compare the various methods with one another and with the theoretical lower variance bounds. Finally, some discussion on the suitability of a particular method for given signal class and SNR range is presented.

The following model is used throughout the text: $s(n), n = 0, \ldots, N-1$ is a length, N, sequence. The analytic signal associated with $s(n)$ is a discrete-time sequence given by $z_s(n) = s(n) + jH[s(n)]$, where $H[\,]$ defines the discrete-time Hilbert transform operation [45]. $z(n), n = 0, \ldots, N-1$ is a length N sequence of complex data values corresponding to the undistorted analytic signal, $z_s(n)$, plus an additive complex Gaussian noise component, $\varepsilon(n)$. The undistorted complex signal has real part, $s(n)$ and imaginary part $q(n)$, while the distorted signal has real and imaginary parts given respectively by $x(n)$ and $y(n)$. The sequence $s(n)$ is modeled as having the form, $s(n) = A(n)\cos(\phi(n))$, where $A(n)$ and $\phi(n)$ are the amplitude and phase functions, respectively. It is assumed that $z_s(n)$, generated with a Hilbert Transformer, is of the form, $z_s(n) = A(n)e^{j\phi(n)}$. The conditions which are necessary for this last assumption to be valid are discussed in [12]. The sampling period (and the sampling frequency) is assumed to be unity throughout this paper.

II. FREQUENCY ESTIMATION TECHNIQUES FOR STATIONARY SIGNALS

Before considering the specific problem of frequency estimation and subsequently of IF estimation it is useful to recall some principles used in the process of estimation generally. These basic principles are outlined below. Several properties are usually sought for a "good" estimator. Typically, one seeks estimators that are *consistent,*

Manuscript received January 25, 1991; revised November 6, 1991.

The author is with Queensland University of Technology, Centre for Signal Processing Research, Brisbane Q.4001, Australia.

IEEE Log Number 9107703.

and statistically and computationally *efficient*. A consistent estimator is one which converges in probability to the true value asymptotically. For a consistent estimator, then,

$$\lim_{N\to\infty} \Pr\{|\hat{a} - a| > \epsilon\} = 0 \tag{1}$$

where $\hat{a}$ is the estimator of a, N is the number of samples in the observation sequence, Pr denotes probability, and ε is an arbitrarily small positive number [40, p. 45].

A statistically *efficient* estimator is one whose variance is the lowest theoretically possible (assuming the estimate is unbiased). This lower theoretical bound on the variance is called the Cramer–Rao (CR) bound [40, p. 46], [60, p. 72]. It is given by

$$\mathrm{var}(\hat{a}) \geq \frac{1}{E[(\,\partial \ln p(z;a)/\partial a)^2]} \tag{2}$$

or equivalently by,

$$\mathrm{var}(\hat{a}) \geq \frac{-1}{E[(\,\partial^2 \ln p(z;a)/\partial^2 a)]} \tag{3}$$

where $z = [z(1)z(2)\ldots z(n)]$ is the vector of observed samples, and $p(z;a)$ signifies the probability density function (pdf) of z given the parameter, a. E is the expectation operator. Since z is complex with real part x and imaginary part y, the pdf of z is the joint pdf of x and y. $p(z;a)$ is often referred to as the likelihood function.

It may be that no estimator will meet this bound, but if there is one, it can be produced by the use of maximum likelihood (ML) techniques, i.e., that estimate of the parameter is chosen which would have made the observed sequence most likely to occur [40, p. 47]. While the ML estimate is guaranteed to be statistically efficient for long data lengths, it may not be computationally efficient. It may in fact be computationally quite intensive. For this reason, ML estimators are sometimes discarded in favor of suboptimal but computationally simpler ones. The principles described above have been put to very wide use in estimating the frequency of a sinusoid in white Gaussian noise. The signal model which has often been used is

$$z(n) = Ae^{j2\pi fn} + \epsilon(n) \tag{4}$$

where A is the amplitude, f is the frequency, $z(n)$ is the discrete complex observation sequence and $\varepsilon(n)$ is the complex white Gaussian noise sequence. The ML estimate for the frequency of a single sinusoid in white Gaussian noise has been shown to be given by finding that frequency at which the periodogram, or "spectrum," attains its maximum [52]. This may be implemented with an initial coarse search on the bins of a Fast Fourier Transform (FFT) and a subsequent interpolation procedure [52]. As long as the "coarse" frequency estimate falls within the main lobe of the frequency response, this technique converges to the correct global maximum [53]. This estimate meets the CR bound above a SNR threshold, the bound being given by [52]:

$$\mathrm{var}[\hat{f}] \geq \frac{12}{(2\pi)^2(A^2/\sigma^2)N(N^2-1)} \tag{5}$$

where N is the number of independent samples in the data, A is the signal amplitude and $2\sigma^2$ is the complex noise variance.

The estimated variance departs quite dramatically from the CR bound once the SNR falls below a threshold value, a phenomenon which is common in nonlinear estimators [52]. (See also [2] for a detailed discussion on the CR bound for stationary signals).

As mentioned earlier, the ML method can be computationally too intensive in some applications. In an attempt to find frequency estimators which reduce computation and/or increase resolvability, many researchers have turned to parametric methods. These methods typically model the signal as having a rational transfer function. It is often computationally advantageous to assume that the numerator of the transfer function is a constant. Such models are said to be auto-regressive (AR), or alternatively, linear predictive. The frequency estimates are obtained by finding the roots of the polynomial denominator. For say a simple complex sinusoid, these methods are very computationally efficient, although they are generally not statistically efficient. They also allow closely spaced sinusoids to be well resolved. Several variants of this approach exist, and include maximum entropy methods, Prony's Method, etc. They are described in [40] and [43].

Other techniques which have found widespread use for spatial frequency estimation in the array processing field are the eigenbased methods, such as Pisarenko's Harmonic Decomposition and MUSIC [40, p. 431]. These methods assume that the observed signal can be decomposed into noise and signal components, and then use the fact that the signal vectors will be orthogonal to the noise vectors. Thus the MUSIC spectral estimator is formed as the inverse of the sum of inner products between signal vector and noise vector estimates. The frequencies of the signal components are taken to be the peaks of the spectral estimate.

Recently, Tretter introduced another frequency estimation technique [58]. He showed that for a complex sinusoid in white Gaussian noise at high SNR, the phase may be approximated well as a linear function of time, imbedded in an additive white Gaussian noise process. He then used a linear regression (i.e., least squares fitting) technique to estimate the frequency. Because the least squares technique is equivalent to the ML one for white Gaussian processes [40, p. 49], his estimator is also ML for the high SNR range under consideration. It thus approaches the CR bound for high SNR. One problem with Tretter's algorithm is that the first stage of the algorithm necessitates extracting the phase from the data. This is prone to significant numerical errors. Kay obtained a modified form of this estimator by fitting a model to adjacent phase difference estimates, rather than the phase values themselves, thus avoiding the phase unwrapping problem [39]. The resulting estimator is simply a smoothing of phase differences with a quadratic window.

The methods described in this section have dealt specifically with sinusoidal frequency estimation. They provide a good basis to understand the more complicated problem of estimating time-varying frequencies, which is considered in the next section.

III. Review of Instantaneous Frequency Estimation Techniques

A. Definition of the Discrete-Time IF

The definition for the IF of a real continuous-time signal, $s(t)$, was given by Ville as [62]:

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{6a}$$

where $\phi(t)$ is the phase of the analytic signal associated with $s(t)$. This was the definition used in [12].

To implement discrete-time IF estimators based on the definition in (6a), one must first address the question of how the differentiation operation may be realized in discrete time. One solution is to use a discrete finite impulse response (FIR) differentiator [45, p. 164]. The discrete-time IF may then be defined as

$$f_i(n) = \frac{1}{2\pi}\phi(n) * d(n) \tag{6b}$$

where $d(n)$ is the impulse response of an FIR differentiating filter, and $*$ denotes convolution in time. Such filters have practical problems, however, since they exaggerate the effects of high frequency noise [14]. Good approximations to the differentiation operation in discrete-time can be obtained by using a phase differencing operation. This approach is very computationally efficient and in general yields better noise performance than is obtainable with (6b). The forward and backward finite differences ((FFD) and (BFD), respectively) defined by (7) and (8), are two commonly used phase differencing operations:

$$\hat{f}_f(n) = \frac{1}{2\pi}(\phi(n+1) - \phi(n)) \tag{7}$$

$$\hat{f}_b(n) = \frac{1}{2\pi}(\phi(n) - \phi(n-1)). \tag{8}$$

One may also estimate the discrete-time IF using another single phase differencing operation, referred to as the central finite difference (CFD):

$$\hat{f}_c(n) = \frac{1}{4\pi}(\phi(n+1) - \phi(n-1)) \tag{9}$$

Of the three discrete IF estimators in (7), (8), and (9), the one defined by (9) has some distinct advantages. Firstly it is unbiased and has zero group delay for linear FM signals [12], and secondly it corresponds to the first moment in frequency of a number of TFD's [19], [13].

1) General Phase Difference Estimator: One can define a class of phase difference estimators which are unbiased for polynomial phases of arbitrary order. For phase given by

$$\phi(n) = \sum_{i=0}^{p} a_i n^i \tag{10}$$

the IF is obtained as

$$\hat{f}_i(n) = \frac{1}{2\pi}\sum_{i=1}^{p} i a_i n^{i-1}. \tag{11}$$

A qth order generalized phase difference estimator may be defined as [9]:

$$\hat{f}(n) = \frac{1}{2\pi}\sum_{k=-q/2}^{q/2} b_k \phi(n+k) \tag{12}$$

where q is an even integer. The b_k coefficients are to be found so that $\hat{f}(n) = \hat{f}_i(n)$, or:

$$\sum_{k=-q/2}^{q/2} b_k \phi(n+k) = \sum_{i=1}^{p} i a_i (n)^{i-1}. \tag{13}$$

For example, the second order phase difference estimator is just the CFD estimator. Note that the even ordered estimators are favored over the odd order ones because they introduce no group delay. A table of the coefficients for the first few even ordered phase difference estimators, each of which is unbiased for a polynomial phase function of the given order, is shown as follows:

Order	Coefficients						
2	$-\frac{1}{2}$	0	$\frac{1}{2}$				
4	$\frac{1}{12}$	$-\frac{2}{3}$	0	$\frac{2}{3}$	$-\frac{1}{12}$		
6	$-\frac{1}{60}$	$\frac{3}{20}$	$-\frac{3}{4}$	0	$\frac{3}{4}$	$-\frac{3}{20}$	$\frac{1}{60}$

The details are provided in [9].

One can also implement the phase differentiation using the classical continuous-time formula for FM discriminators [30], [1], [31]:

$$\phi'(t) = \frac{x(t)y'(t) - x'(t)y(t)}{x^2(t) + y^2(t)}. \tag{14}$$

This leads to the following discrete-time estimator [30, p. 72]

$$\hat{f}_i(n) = \frac{1}{2\pi}\frac{x(n)y'(n) - x'(n)y(n)}{x^2(n) + y^2(n)}. \tag{15}$$

FIR filters are also used to approximate the differentiation operations needed in order to evaluate the numerator [45].

B. Smoothed Versions of the Phase Difference Estimator

The CFD estimator given in (9) is unbiased for linear FM signals, but exhibits very high variance for noisy signals. A number of approaches may be used to reduce the variance. Firstly, if the signal's frequency is known to be constrained to a bandwidth, B, then the signal is filtered outside this bandwidth. Secondly, smoothing may be used to yield lower variance (but not necessarily unbiased) estimates.

A particular smoothed estimator has been proposed by Kay. He defined a "weighted phase difference" estimator, which meets the CR bounds for a stationary signal, as given in (5). He arrived at this estimator by trying to find the ML estimate of a sequence of local frequency estimates.

These local frequency estimates are obtained by forward finite difference operations. Because of the differencing operation, the sequence of estimates are correlated, or "colored." The solution for the optimal or ML value for the frequency, then, does not reduce to a linear averaging, but rather to a "weighted" averaging procedure. Kay's estimator is given by

$$\hat{f}_{(n)} = \frac{1}{2\pi} \sum_{n=0}^{N-2} h_n[\phi(n+1) - \phi(n)] \tag{16}$$

where h_n, the averaging or smoothing window, is given by [39]:

$$h_n = \frac{1.5N}{N^2-1}\left(1 - \left[\frac{n-((N/2)-1)}{N/2}\right]^2\right). \tag{17}$$

The variance reduction which is achieved by using the above window is [39]:

$$\frac{\mathrm{var}(\hat{f})|\text{no window}}{\mathrm{var}(\hat{f})|\text{window}} = \frac{N(N+1)}{6(N-1)} \simeq \frac{N}{6}. \tag{18}$$

The weighting function, h_n, can also be applied to AR based frequency estimates [39] in preference to the normal rectangular window, and simulations have shown that significantly enhanced performance results. As pointed out in [35], the estimator in (16) can easily be implemented as an IF estimator or tracker, by formulating it as a recursion in time. That is, a sliding window may be used to evaluate the frequency locally, in much the same way as the spectrogram is used to estimate the spectrum locally. This approach, however, will yield degraded estimates if there are significant frequency variations (particularly nonlinear ones) within the window. To overcome this limitation, one can reconsider the whole problem using a polynomial phase model for the signal rather than a linear phase one [9]. Both Tretter's linear regression fit estimator [58] and Kay's weighted phase difference estimator, in fact, may be extended to a polynomial phase signal model. Section IV describes both extensions in detail.

C. Zero-Crossing IF estimation

One means for estimating the local frequency of a narrowband process, which is common in seismic processing, is to measure the number of zero-crossings. For a sinusoidal signal, or a signal which can be considered to be locally stationary, the frequency is given by the inverse of the period, or alternatively by half the inverse of the interval between zero-crossings, i.e.,

$$f = \frac{1}{2T_z} \tag{19}$$

or

$$f = Z/2 \tag{20}$$

where T_z is the interval between zero crossings, $2T_z$ is the period, f is the frequency, and Z is the zero-crossing rate.

Since we are dealing with discrete-time signals with unit sampling rate, the value of T_z is actually given by the number of sample intervals, k, between zero-crossings, and therefore, (19) becomes $f = 1/2k$. (It is assumed for the present that the zero-crossings fall exactly on sample points). Then there will be exactly $k+1$ sample points in the interval between consecutive zero-crossings (including the two end points). This zero-crossing estimate is shown below to be a linear average of the FFD estimates within the interval [9].

Proof: Since $T_z = k$, the zero-crossing estimator in (19), assuming the first zero-crossing occurs at time index n, can be expressed as

$$\hat{f}(n) = \frac{1}{2k}. \tag{21}$$

The usual FFD estimator may be written as,

$$\hat{f}_f(n) = \frac{1}{2\pi}[\phi(n+1) - \phi(n)]. \tag{22}$$

The linear average of the k consecutive FFD estimators within the interval is

$$\begin{aligned}\hat{f}_{fk}(n) =& \frac{1}{2k\pi}[\phi(n+1) - \phi(n) + \phi(n+2) - \phi(n+1) \\ &+ \ldots + \phi(n+k) - \phi(n+k-1)]\end{aligned} \tag{23}$$

which may be rewritten as

$$\hat{f}_{fk}(n) = \frac{1}{2k\pi}[\phi(n+k) - \phi(n)]. \tag{24}$$

Now since $\phi(n+k) - \phi(n) = \pi$, (24) reduces to (21).Q.E.D.

Thus the expression for the zero-crossing estimate is simply a linear average of k adjacent FFD estimates. The averaging does not incorporate the optimal quadratic weighting function, h_n described in the previous section, and hence is suboptimal. It is, however, extremely simple computationally. Note that if the interval between zero-crossings is not an integer number of samples, then in addition to the linear averaging produced by the estimator, quantization "noise" is introduced. To reduce the variance of the zero-crossing estimate, Rabiner and Schafer [46] have proposed taking the average number of zero-crossings within a window of length M. Their estimator is defined by

$$Z(n) = \sum_{m=-M}^{M} |\mathrm{sgn}[s(m)] - \mathrm{sgn}[s(m-1)]| h(n-m) \tag{25}$$

where

$$\begin{aligned}\mathrm{sgn}[s(n)] &= 1, \quad \text{for } \ s(n) \geq 0 \\ &= -1, \quad \text{for } \ s(n) < 0\end{aligned} \tag{26}$$

and where

$$\begin{aligned}h(n) &= \frac{1}{2M}, \quad \text{for } \ 0 \leq n \leq M-1 \\ &= 0 \quad \text{otherwise.}\end{aligned} \tag{27}$$

The previous discussion applies to locally sinusoidal processes. For nonstationary FM signals, the zero-crossing based instantaneous frequency estimator can be extended according to:

$$\hat{f}_i(n) = \frac{Z(n)}{2} \tag{28}$$

where $Z(n)$ is given by (25). By sliding the window, $h(n)$, with n, we obtain an estimate of the IF as a function of n. The finite length of the window, M, introduces a bias-variance trade-off in IF estimation. If M is large, and the IF law is not linear within the window, there will be a bias. If M is small the bias is likely to be reduced but at the cost of higher variance. Results are presented in Section V.

D. Adaptive IF Estimation

One approach to IF estimation is to formulate it as a problem of adaptively estimating the local frequency. This approach has given rise to the phase locked loop (PLL), which is used widely in communications systems [17]. The PLL adaptively demodulates the incoming signal to baseband, where it is filtered, and the output fed back into the demodulation stage. The standard PLL performs quite well in noise, but is unable to track very rapid changes in the IF. Special modifications are required to be able to deal with the latter.

Snyder has derived some useful adaptive estimators in [56]. He derived an estimator based on a nonlinear least squares criterion, and a linear approximation to this estimator (the *Extended Kalman Filter*). He also showed that the Extended Kalman filter reduces very nearly to the PLL in the stationary case. Other variants and extensions of the PLL have also been devised [33].

Another form of adaptive IF estimation is based on modeling the data as a linear predictive process. Two methods which may be used for this type of estimation are the LMS and the RLS algorithms [32]. Both are described as follows.

1) LMS Algorithm: Griffiths proposed an adaptive IF estimation algorithm in [29] based on a linear prediction filter which has its coefficients updated with each new data sample. Griffiths' method is based conceptually on extracting the peak of a short-time linear prediction based spectral estimate. Great computational savings are achieved in the process by recursively updating the spectral estimate as each new point is received rather than by recalculating it from the raw data each time. The resulting algorithm, which is based on gradient descent techniques, is quite simple. However, because the recursive algorithm is inherently an IF tracking process, it is unable to respond to very rapid (or noisy) IF changes. The estimate may therefore exhibit significant noise susceptibility. Details of the algorithm are given below.

The vector of data samples at time, n, is denoted by

$$\boldsymbol{z}_n = [z(n)\; z(n-1)\; ..\; z(n-L+1)]^T \tag{29}$$

where L is the linear prediction filter length, and ^{T} represents the transpose operation. The corresponding vector of linear prediction filter coefficients is

$$\boldsymbol{a}_n = [a_1(n)\; a_2(n)\; ..\; a_L(n)]^T. \tag{30}$$

As each new data sample is processed, the filter coefficients should ideally be updated so as to minimize the mean square prediction error. For stationary statistics the error is a unimodal function of the filter coefficient vector, and hence gradient descent techniques may be used to converge to the optimal values for the filter coefficients. The LMS algorithm of Widrow and Hoff [69] is used and the updated coefficients are given by the relation as [43, pp. 264–266]

$$\boldsymbol{a}_{n+1} = \boldsymbol{a}_n - 2\mu e_{n+1}\boldsymbol{z}_n^* \tag{31a}$$

$$e_{n+1} = z(n+1) + \boldsymbol{z}_n{}^T\boldsymbol{a}_n \tag{31b}$$

where e_{n+1} is the linear prediction error at time index, $n+1$, μ is the adaptation constant, and $*$ denotes the complex conjugate. A standard form of equations for the LMS algorithm is given in [32, pp. 302–304].

The IF estimate is determined from the peak of the linear prediction based spectrum, i.e.,

$$\hat{f}_{i(n)} = f_i \quad \text{which maximizes}$$
$$\left|1 + \sum_{k=1}^{L} a_k(n)\exp[-j2\pi f_i k]\right|^{-2}. \tag{32}$$

Where several frequencies are being estimated or tracked, the expression in (32) is modified to extract the different peaks corresponding to the individual frequency components. For tracking a single complex sinusoid in noise, the IF may be determined with great computational efficiency according to

$$\hat{f}_i(n) = \frac{1}{2\pi}\arg[a_1{}^*]. \tag{33}$$

The coefficient μ controls the rate of adaptation—if μ is close to its upper limit, adaptation is swift but steady state error may be large, while if μ is small, adaptation will be slow. The main advantage of this algorithm is its computational simplicity which is apparent from (31). In addition, better algorithms may be used for the adaptation [70].

2) RLS Adaptive Frequency Estimation: The RLS algorithm is a technique which again models the data as a linear prediction sequence, and which updates the linear prediction coefficients with each new data sample. The RLS algorithm differs from the LMS algorithm in that an exponentially weighted approximation to the inverse of the covariance matrix is used as the "adaptation coefficient," rather than a scalar coefficient. The advantage of the RLS algorithm over the LMS one is its improved speed of convergence and robustness to signal energy levels. The classical RLS algorithm requires order L^2 computations as opposed to order L for the LMS algorithm, but *fast RLS* algorithms have been developed which are only order L [18]. The updating of the algorithm parameters at time, n, is achieved by the following set of equations [43, pp. 267–269]:

$$\boldsymbol{a}_{n+1} = \boldsymbol{a}_n - e_{n+1}\boldsymbol{P}_n\boldsymbol{z}_n{}^* \tag{34a}$$

$$e_{n+1} = z(n+1) + \boldsymbol{z}_n{}^T\boldsymbol{a}_n \tag{34b}$$

$$P_n = \left[\alpha P_{n-1}^{-1} + z_n{}^* z_n{}^T\right]^{-1} \tag{34c}$$

where P_n is the exponentially weighted approximation to the covariance matrix inverse, and α is the *forgetting factor*. A standard form of equations for the RLS algorithm is given in [32, pp. 480–483]. By using the matrix inversion lemma, these equations reduce to:

$$a_{n+1} = a_n - e_{n+1} c_n \tag{35a}$$

$$c_n = \frac{P_{n-1} z_n{}^*}{\alpha + z^T{}_n P_{n-1} z_n{}^*} \tag{35b}$$

$$P_n = \frac{1}{\alpha}\left(I - c_n z_n{}^T\right) P_{n-1} \tag{35c}$$

where e_{n+1} is as defined in (29b), and I is the identity matrix.

For a single noisy complex sinusoid the above equations can be implemented very simply, and the resulting estimator is unbiased. The local IF may be obtained according to (33). For multiple components the IF estimates are extracted from the peaks of the linear prediction spectrum using (30). Results are presented in Section V.

E. IF Estimation Based on the Moments of TFD's

Cohen has formulated a class of two-dimensional functions (TFD's) which may be used to represent the distribution of signal energy in time and frequency [20]. The discrete time expressions for these functions were given in [13] as

$$\rho(n,k) = \sum_{m=-M}^{M} \sum_{m-M}^{M} G(p-n,m) z(p+m) \cdot z^*(p-m) e^{-j4\pi mk/N} \tag{36}$$

where $G(n,k)$ is a window function which selects a particular TFD, and $M = (N-1)/2$.

A number of TFD's (e.g., the Wigner–Ville Distribution (WVD)) yield the IF through their first moment [19], and many other TFD's (e.g., the Short-Time Fourier Transform [13]) yield approximations to the IF through their first moment. TFD first moments, then, provide another means of estimating the IF. In [64] White and Boashash considered the problem of estimating the IF of a Gaussian random process using WVD first moments. The particularly useful aspect of estimating the IF in this manner is that masking or other forms of preprocessing in the time-frequency plane can be performed so as to reduce noise effects or to estimate IF laws of the various components separately [13]. See for example, Figs. 1 and 2 which show respectively linear FM IF estimates obtained from the WVD first moment obtained with and without a preliminary masking operation. The SNR level was 3 dB. The masking or time-varying filtering operation has resulted in a significant variance reduction. This methodology is further described in Section VI-B and in [13]. The IF derived via the first moment of the discrete WVD is given by [6], [19]:

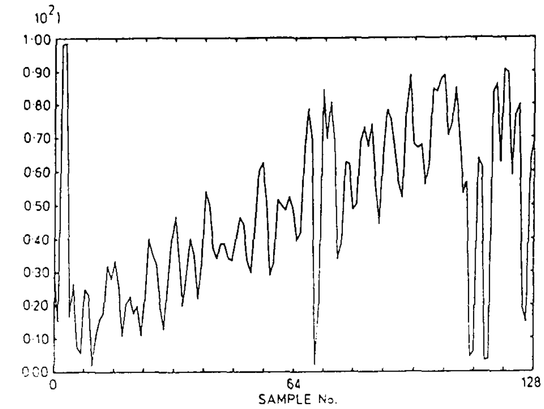


Fig. 1. IF law of a linear FM signal obtained by using the first moment of the WVD.

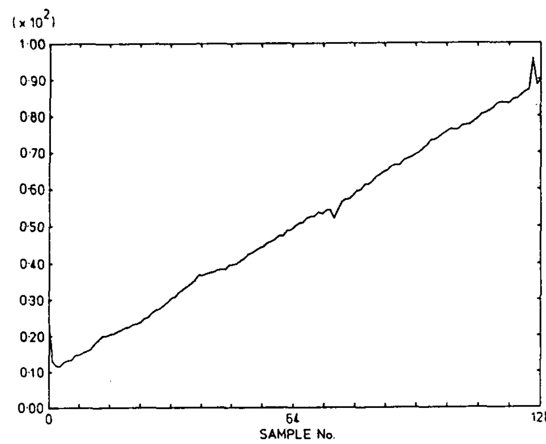


Fig. 2. IF law of a linear FM signal obtained by using the first moment of the WVD and time-frequency filtering.

$$\hat{f}_c(n) = \frac{M}{2\pi}\left(\sum_{k=0}^{M-1} e^{j2\pi k/M} W^g(n,k)\right) \bmod 2\pi. \tag{37}$$

where $W^g(n,k)$ is the discrete WVD, and is defined by

$$W^g(n,k) = \sum_{m=-M}^{M} z(n+m/2) z^*(n-m/2) e^{-j2\pi mk/M}. \tag{38}$$

Because this method is computationally demanding (requiring calculation of a TFD, masking and first moment calculations), and is not generally statistically optimal, other methods are preferred.

The IF, or an approximation thereof, may similarly be obtained via moments of the discrete time TFD's defined in (36) [6], [13]. Such moments can be shown to be approximately a smoothed CFD estimator [13]. That is,

$$m^1{}_\rho(n) \simeq \left[\hat{f}_c(n) * G(n,1)\right] \bmod \frac{f_s}{2}. \tag{39}$$

Here $\hat{f}_c(n)$ is as specified in (9) and $G(n,1)$ is the function $G(n,k)$ of (36) evaluated at $k = 1$. These results are detailed in [13]. In practice, this method is computationally demanding, since it requires calculation of a TFD, a masking operation and then an inversion procedure. Because of its equivalence with a smoothed CFD operation, it is used only in very specific applications.

F. IF Estimation Based on the Peak of TFD's

1) Peak of the Short-Time-Fourier-Transform: As indicated earlier, the ML estimate of the frequency of a complex sinusoid imbedded in a stationary white complex Gaussian noise sequence is given by the peak of the periodogram or "spectrum" [52]. It seems intuitively appealing to generalize this result to nonstationary signals by forming an IF estimate from the peak of a TFD, since TFD's were constructed especially to deal with nonstationary signals. Many authors have proposed such a technique. One of the most obvious TFD's to use is the STFT, which will perform very well if the signal being considered is quasistationary. However, because of the fact that the optimum window for the spectrogram is the reciprocal of the square root of the frequency rate, poor results are obtained for rapidly slowing FM signals [9]. Results are presented in Section V.

2) Peak of the WVD: The WVD peak was proposed as an IF estimation technique in [5], and it was applied to determining absorption and dispersion parameters in seismic processing. The WVD's usefulness arises from its ability to localize energy along the IF law. If the signal under consideration has a linear frequency law and constant amplitude, the WVD will reduce to a row of delta functions on the IF law—a situation which is very conducive to IF estimation. The existence of any nonlinearity in the IF law may cause the peak of the IF to be biased, so it is common to adjust the window so that the IF variation inside the window length is approximately linear [12].

Rao and Taylor showed further that WVD peak based IF estimation is optimal for linear FM signals with high to moderate SNR [51]. Wong and Jin also investigated the use of this estimator and compared its performance with the CR lower bounds for continuous time-varying frequency [71]. IF peak detection from an AR based WVD was also used in [68] and was seen to track nonlinearly varying FM laws better than the conventional WVD at high SNR. While the WVD peak extraction method has been shown to be an optimal IF estimator for linear FM signals at high SNR, it degrades significantly at low SNR. For this reason the use of the cross Wigner–Ville (XWVD) peak has also been proposed as an IF estimator [46], [47], [13], [10]. It is described below. Simulations may be found in Section V.

3) Peak of the XWVD: The XWVD between a reference signal, $s(n)$, and an observed signal, $r(n)$, is defined as

$$W_{z_s z_r}(n,k) = \sum_{m=-M}^{M} z_s(n+m/2) \cdot z_r{}^*(n-m/2) e^{-j2\pi mk/M} \tag{40}$$

where $z_s(n)$ and $z_r(n)$ are analytic signals corresponding to $s(n)$ and $r(n)$, respectively.

Normally the reference signal is not known, but it may be estimated from the observed signal. The estimation of the reference involves obtaining an initial IF estimate, $f_i(n)$, and reconstructing a unit amplitude signal according to

$$z_{\hat{s}} = \exp\left(j2\pi \sum_{k=0}^{n} \hat{f}_i(k)\right), \quad \text{for } n = 0, \ldots, N-1. \tag{41}$$

The proposed XWVD based IF estimation procedure is:

1) *Initialization:* Form a unit amplitude reference signal from the estimated IF;
2) *Estimation:* Form the XWVD between the reference signal and the observed signal; estimate the IF from the peak of the XWVD;
3) *Recursion:* Repeat Step 1 until the difference of IF estimates from successive iterations is less than a specified amount.

The rationale behind the method is that each time a new XWVD is estimated the signal energy concentration should increase, so that the probability of correctly estimating the IF in a noise background should also increase. Note that any estimator could be used as the starting value for the IF. If an STFT peak detection estimate is used, the algorithm converges typically within a couple of iterations. Details of this scheme, including an analysis of the performance, is provided in [10], [13], and [46]. If the signal is long and is analyzed via a sliding window, such that the FM law is linear within the window, then the method will attain the CR bound with an SNR threshold substantially lower than in the case of the WVD peak (see Table 1) [13]. Results are presented in Section V.

4) Peak of the generalized WVD: The generalized WVD (GWVD) provides a tool for the problem of time-frequency signal representations of polynomial nonlinear FM signals [15]. The GWVD is multilinear and it represents higher-order polynomial FM signals as delta functions in the time-frequency domain.

The IF estimate from the peak of the GWVD is unbiased and its variance meets the CR lower bounds at high SNR's [15].

5) Other TFD's: Further possibilities for TFD peak based IF estimates lie with some of the recently proposed TFD's [16]. The Zhao-Atlas-Marks Distribution, for example, seems to have good time-frequency localization and good noise performance [72], making it a useful prospect for IF estimation. It also tends to suppress cross-terms, so that it is suited to multicomponent signals. The signal-dependent TFD's of Baraniuk and Jones [3] and the adaptive techniques for high resolution time-varying spectral estimation of Fineberg and Mammone [26] also seem promising. The relationship between the IF and the wavelet transform [22] is an open question.

G. Time-Varying AR Model Based IF Estimation

Sharman and Friedlander [51] proposed an IF estimator based on time-varying AR modeling of the signal data. The authors developed an AR estimation procedure which specifically takes account of the signal nonstationarity, and arrived at a solution based on solving an adapted version of the Modified Yule-Walker equations. The IF was obtained from the roots of the time-varying AR polynomial in much the same way as frequency is derived from the stationary transfer function. The accuracy attainable is unfortunately lower than is attainable for the stationary case, due to the fact that it is difficult to obtain good "instantaneous" covariance estimates. That is, it is not possible in the nonstationary case to average over large data records.

H. Enhancement of IF Laws Through the Application of Tracking Algorithms

One can achieve significant improvement in the IF estimates if some *a priori* information about the rate of variation of the IF is available. This information usually involves some assumption about the IF being slowly varying, which can be used to "smooth" the sequence of IF estimates. One way to characterize this information is to assign probability distributions (known as *a posteriori* distributions) to various parameters in the IF law. The smoothing of the IF law is commonly referred to as "tracking" [4].

One technique for tracking is to model the IF estimate sequence as a Markov chain in which there is a fixed probability distribution associated with the IF changing from one point to the next. Normally small changes in consecutive IF estimates will be assigned high probabilities, while large changes will be given low probabilities. The optimal IF law or "track" is the one which makes the actual sequence of IF estimates most likely to have occurred. It can be obtained using a Hidden Markov Modeling algorithm. This kind of maximum *a posteriori* (MAP) IF estimation approach has been applied to spectrogram peak based IF estimates for single component signals in [4], and for multiple signals in [67]. MAP IF estimation algorithms have also been developed for use on the data directly. These latter algorithms, based on dynamic programming [66] and on Hidden Markov Modeling [66], [67], have the advantage that the IF law need only be slowly varying in general. Sharp transitions are possible, but are assigned a low probability.

Another type of tracking algorithm which may be used is based on Kalman filtering techniques. This type of approach is described in [36].

The application of tracking algorithms is a particularly attractive proposition for WVD peak based estimates. Often the artifacts inherent in the WVD create interpretation difficulties, and limit the usefulness of the WVD. However, these artifacts should be ignored by the tracker, because of their irregular behavior. This property was used to advantage in [20], where IF tracks were extracted from the WVD plane, based on the assumption that the IF laws were polynomial functions of time.

IV. IF Estimation Methods Based on Polynomial Phase Modeling

The estimators based on the direct definition of the discrete IF in Section III-A make no implicit assumptions about the form of the IF law, and for this reason, they exhibit high variance. Significant reductions in variance can be achieved by incorporating some form of *a priori* knowledge into the estimation procedure. One means of doing this is to assume that the IF law may be expressed as a finite order polynomial, which in turn implies a polynomial phase law. The selection of order of the polynomial is what allows the incorporation of the *a priori* information: if the signal's IF is known to be slowly changing, low orders can be chosen, whereas high orders can be chosen if the IF law is known to be changing rapidly. The signal model under the polynomial phase approximation is generalized from (4) and reads:

$$z(n) = A(n)\ \exp(j\phi(n)) + \epsilon(n) \tag{42}$$

$$= s(n) + jq(n) + \epsilon(n) \tag{43}$$

where

$$\phi(n) = a_0 + a_1 n + a_2 n^2 + a_3 n^3 + \ldots a_p n^p = \sum_{k=0}^{p} a_k n^k,$$

and where $A(n)$ is the amplitude, $\phi(n)$ is the phase, $\varepsilon(n)$ is a complex noise process of variance $2\sigma^2$, and $n = 0, N-1$. $s(n)$ and $q(n)$ are given by

$$s(n) = A(n)\cos(a_0 + a_1 n + a_2 n^2 + \ldots a_p n^p) \tag{44a}$$

$$q(n) = A(n)\sin(a_0 + a_1 n + a_2 n^2 + \ldots a_p n^p). \tag{44b}$$

To estimate the coefficients a_k, $(k = 0, \ldots, p)$, of the phase polynomial in (44), one possibility is to find the set of coefficients which minimize the squared error between the estimated signal and the observed one. This will result in a nonlinear least squares problem and would have to be solved numerically. An alternative is to "linearize" the problem by unwrapping the phase. Tretter [58] has shown that this approximation is valid for high SNR. Simple linear least squares estimation techniques may then be used [9].

A further possibility is to extend the ML parameter estimation approach of Rife and Boorstyn for a stationary tone [52] to the polynomial case. Since the observed signal is assumed Gaussian this will be equivalent to the nonlinear least squares solution [60, p. 61], but is in many respects more straightforward to implement. Once the polynomial phase law estimate $\hat{\phi}(n)$ has been calculated, the IF, $\hat{f}_i(n)$ is directly obtained from:

$$\hat{f}_i(n) = \frac{1}{2\pi}\frac{d\hat{\phi}(n)}{dn} = \frac{1}{2\pi}\sum_{k=1}^{p} k\ \hat{a}_k\ n^{k-1} \tag{45}$$

where $\hat{a}_k$ is the estimate of the coefficient a_k.

All of the above mentioned techniques are described in more detail in the following sections.

A. Least Squares Based Polynomial Coefficient Estimation Algorithms

1) Least Squares Estimation Methodology: The $p+1$ unknown parameters $(a_0, \ldots, a_p)$ in (44) may be obtained by minimizing the sum of squared error, E, according to:

$$E = \sum_{n=0}^{N-1} |\hat{z}(n) - z(n)|^2. \tag{46}$$

The solution for the parameters, $\hat{a}_k$, is obtained numerically by solving the system of equations, (47) and (48) below, derived in Appendix B [9]:

$$\sum_{n=0}^{N-1} z(n)\ \hat{A}(n)\ e^{j\hat{\phi}(n)}\ n^k = \sum_{n=0}^{N-1} \hat{A}^2(n) e^{2j\hat{\phi}(n)}\ n^k \tag{47}$$

$$\sum_{n=0}^{N-1} z(n)\ e^{j\hat{\phi}(n)}\ n^k = \sum_{n=0}^{N-1} \hat{A}(n)\ e^{2j\hat{\phi}(n)}\ n^k. \tag{48}$$

$\hat{A}(n)$ is the amplitude law estimate obtained as an order q polynomial in a similar manner to that used for obtaining the phase:

$$\hat{A}(n) = \sum_{k=0}^{q} \hat{b}_k\ n^k \tag{49}$$

(47) and (48) then become a system of $p+q+2$ highly nonlinear equations which are not easily solved. Furthermore, the statistical characteristics of the resulting estimators would not be analytically obtainable. A numerical solution may be obtained for the full vector of parameters, $\boldsymbol{\beta} = [b_0, b_1, b_2, \ldots b_q, a_0, a_1, a_2, \ldots a^p]^T$, from:

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - \boldsymbol{J}^{-1} \frac{dp(z;\boldsymbol{\beta})}{d\boldsymbol{\beta}} \tag{50}$$

where $p(z;\boldsymbol{\beta})$ is the pdf of the signal, $\boldsymbol{\beta}^k$ the kth iterative evaluation of $\boldsymbol{\beta}$ and $\boldsymbol{J}$ is the Fisher Information matrix [40, p. 47], with elements given by

$$J_{i,j} = E\left[\frac{\partial \log(p(z;\boldsymbol{\beta}))}{\partial \beta_i} \cdot \frac{\partial \log(p(z;\boldsymbol{\beta}))}{\partial \beta_j}\right] \tag{51}$$

where $E[\ ]$ denotes the expectation operator, and β_i is the ith element of the parameter vector, $\boldsymbol{\beta}$. Use of this method for determining the optimal parameter set, is problematical. There may be many local maxima to which the algorithm will converge, and one must then incorporate some means of trying to escape from these local equilibria. A good initial estimate is therefore crucial.

An alternative and simpler method is to unwrap the phase and model the phase function using regression techniques [40, p. 49]. The instantaneous phase is then obtained from the data as

$$\phi(n) = \arctan \frac{\mathrm{Im}[z(n)]}{\mathrm{Re}[z(n)]} \tag{52}$$

Solving (52) for $\phi(n)$ is not a straightforward procedure, since arctan is defined on $(0,\ 2\pi]$ [45, p. 216]. Thus there are numerical problems associated with "unwrapping" the phase, particularly at low SNR. Assuming constant amplitude, the polynomial phase model of (45) may be expressed in matrix form as

$$\boldsymbol{\phi} = \boldsymbol{X}\boldsymbol{a} \tag{53}$$

where $\boldsymbol{\phi}$ is the unwrapped phase observation vector, $\boldsymbol{a}$ is the parameter vector and $\boldsymbol{X}$ a matrix of constants. The three matrices are defined as follows:

$$\boldsymbol{\phi} = [\phi(0), \phi(1), \ldots \phi(N-1)]^T \tag{54}$$

$$\boldsymbol{a} = [a_0, a_1, \ldots a_p]^T \tag{55}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 2^p \\ \cdot & \cdot & \ldots & \cdot \\ 1 & N-1 & \ldots & (N-1)^p \end{bmatrix} \ldots \tag{56}$$

The least squares error solution for the vector of parameters, $\boldsymbol{a}$, assuming that any noise is additive zero-mean, is [40, p. 49]:

$$\hat{\boldsymbol{a}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\phi}. \tag{57}$$

1) Linear Least Squares Algorithm Description: A good description of the phase unwrapping problem along with a useful algorithm is described in [59]. Note that phase unwrapping is fraught with errors at low SNR, due to the very great susceptibility of the phase to noise at low SNR values.

The signal is segmented if the slope of the IF law changes too rapidly. For example, sinusoidal FM signals would be segmented at the frequency law peaks, the inter-peak segments being far easier to model than the entire signal. Auto-segmentation may be performed on the basis of "phase-acceleration."

Initialization: if the form of the IF law is known (e.g., linear, quadratic, etc.) then the model order, p, can be set manually. If the form is unknown, or does not conform to the model directly (e.g., sinusoidal, hyperbolic), then models of increasing order are fitted recursively, starting with $p = 0$. The algorithm is terminated when some criterion of fit is met. The algorithm is initialized by setting $p = 0$. The segments determined in step two are processed one at a time. Initially the $\boldsymbol{X}$ matrix in (53) is a row vector of 1's and hence $(\boldsymbol{X}^T\boldsymbol{X})^{-1} = 1/N$.

Recursion: the recursion involves repeated calculation of the parameter estimation formula in (53) for successively higher values of p. A normalized error measure is used to ascertain goodness of fit and to thereby determine the termination point for the algorithm [9].

In the iterative calculation of (53) it is the inverse matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ that is the most labor intensive task. It is performed using the matrix update lemma, from

[43]. Kitchen [41] has shown that if the phase function is modeled as a weighted sum of orthogonal polynomials the matrix inversion may be dispensed with and a simple algorithm derived. Furthermore the polynomial coefficients may be readily obtained from those of the orthogonal polynomials.

The algorithm works well for high SNR, where the parameter estimate variances meet the CR bounds. The algorithm fails when used for short signals at low SNR due to phase unwrapping errors [58]. The polynomial signal model is conducive to very long observation windows, and for long signals, consequently, the performance can be quite good across a broad SNR range. Simulations are presented in Section V.

B. IF Estimation by Fitting a Polynomial to Local IF Estimates

Kay's weighted phase difference IF estimator, as presented in Section II-B, may be extended to a polynomial phase law, by replacing the forward finite difference estimator by a qth order phase difference estimator. The signal model under the polynomial phase assumption was given in (44).

The local IF estimator used will be one of the qth order phase difference estimators defined in (12). For N discrete data samples, then, $N-q$ consecutive IF estimates will be available. They will be denoted by $\hat{f}_q(n)$, for $n = 0, \ldots, N-q$. Note that these estimators will be unbiased, but will contain a colored zero mean Gaussian noise component, i.e.,

$$\hat{f}_q(n) = \frac{1}{2\pi}\left[\left(a_1 + 2a_2 n + 3a_3 n^2 + \ldots + p a_p n^{p-1}\right)\right] + u(n) \tag{58}$$

where $u(n)$ is the zero mean colored Gaussian noise process, $\hat{f}(n)$ is obtained through a linear combination of Gaussian distributed phases.

The problem of estimating the parameters, $a_1, a_2, \ldots a_p$, becomes one of parameter estimation in colored Gaussian noise. The ML solution involves minimizing [40, p. 50]

$$\left(\hat{\boldsymbol{f}}_q - \boldsymbol{X a}\right)^t \boldsymbol{C}^{-1}\left(\hat{\boldsymbol{f}}_q - \boldsymbol{X a}\right) \tag{59}$$

where $\hat{\boldsymbol{f}}_q = [\hat{f}_q(1), \hat{f}_q(2), \ldots \hat{f}_q(N-1)]^T, \boldsymbol{C}$ is the $(N-1) \times (N-1)$ covariance matrix of $\hat{\boldsymbol{f}}_q, \boldsymbol{a} = [a_0 a_1 \ldots a_p]^T$, and the matrix (60) shown at the bottom of the page. The optimal set of parameter estimates for $\boldsymbol{a}$ is given by [40, p. 50]

$$\hat{\boldsymbol{a}} = \left(\boldsymbol{X}^T \boldsymbol{C}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^T \boldsymbol{C}^{-1}\hat{\boldsymbol{f}}_q\right) \tag{61}$$

and the variance of the estimator is given by [40, p. 50]

$$\operatorname{var}(\hat{\boldsymbol{a}}) = \left(\boldsymbol{X}^T C^{-1}\boldsymbol{X}\right)^{-1}. \tag{62}$$

The covariance matrix, $\boldsymbol{C}$, is a $q+1$ diagonal matrix given by [9]:

$$C_{ij} = \frac{\sigma^2}{A^2(2\pi)^2}\sum_{k=0}^{q-|j-i|} b_k b_{k+|j-i|}, \qquad \text{for } i,j \leq N-q. \tag{63}$$

Equation (61) may be solved in a straightforward fashion to yield the polynomial phase coefficient estimates, and hence the ML estimate of the IF law. These estimates will meet the CR bounds for high SNR. (See Appendix A for a derivation of lower variance bounds for the polynomial coefficient estimates).

At the time of finalizing this manuscript, another paper on frequency rate estimation appeared in the literature [24] which contains a number of useful insights into extending the Kay estimator to higher order phases. The discussion on obtaining robust phase difference estimates is especially helpful.

C. ML Based Polynomial Coefficient Estimation Procedure

Another IF estimator may be defined by deriving the ML solution for the polynomial phase coefficient estimates [9]. The method is a simple generalization of the ML solution for estimating the frequency of a stationary tone [52]. It will be seen that the ML estimates are those coefficients which maximize the magnitude of a *polynomial Fourier transform*, the latter being a correlation of the observed signal with a continuum of polynomial phase signals rather than with the usual sinusoidal tones. In other words, the ML solution finds the best correlation match of a sinusoidal tone to the observed signal in the stationary case, while in the nonstationary case it finds the best correlation match of a polynomial phase function with the observed signal. The details are provided below.

1) ML Estimation Algorithm: It is assumed that the signal to be analyzed is corrupted by white complex Gaussian noise of variance, $2\sigma^2$. The analytic version of the observed signal has a real part, $x(n)$, and an imaginary part, $y(n)$. The real and imaginary parts of the uncorrupted signal are $s(n)$ and $q(n)$, respectively. The parameter vector, assuming constant amplitude, A, is given by $\boldsymbol{\alpha} =$

$$\boldsymbol{X} = \frac{1}{2\pi}\cdot\begin{bmatrix} 1 & 0 & 0 & \cdot 0 \\ 1 & 2 & 3 & \cdot(p-1) \\ 1 & 4 & 12 & \cdot(p-1)2^{p-1} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot\cdot \\ 1 & 2(N-q-1) & 3(N-q-1)^2 & \cdot(p-1)(N-q-1)^{p-1} \end{bmatrix}. \tag{60}$$

$[A, a_0, a_1, a_2, \dots ap]^T$. The pdf of $z(n)$, given the parameter vector, $\boldsymbol{\alpha}$, is [40, p. 44]

$$p(\boldsymbol{z}; \boldsymbol{\alpha}) = \frac{1}{(\sigma^2 2\pi)^N} \exp\left[\frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} (x(n) - s(n))^2 + (y(n) - q(n))^2\right]. \quad (64)$$

The ML estimate of $\boldsymbol{a}$ requires that p, (or $\log(p)$), be maximized over $\boldsymbol{\alpha}$ [40, p. 47]. The maximum of p will occur at the maximum of

$$L_0 = -\frac{1}{N} \sum_{n=0}^{N-1} (x(n) - s(n))^2 + (y(n) - q(n))^2. \quad (65)$$

Since $\Sigma x(n)^2$ and $\Sigma y(n)^2$ are constants corresponding to the signal energy, maximization of L_0 is equivalent to maximizing the function

$$L = \frac{2}{N} \sum_{n=0}^{N-1} (x(n)s(n) + y(n)q(n)) - \frac{1}{N} \sum_{n=0}^{N-1} (s^2(n) + q^2(n)) \quad (66)$$

over all the parameters. Now substituting with the expressions for $s(n)$ and $q(n)$ from (44) into (66) yields

$$L = 2A \ \mathrm{Re}[\exp(-ja_0) D(a_1, a_2, \dots a_p)] - A^2 \quad (67)$$

where

$$D(a_1, a_2, \dots, a_p) = \frac{1}{N} \sum_{n=0}^{N-1} z(n) \cdot \exp\left(-j\left(a_1 n + a_2 n^2 + \dots a_p n^p\right)\right). \quad (68)$$

It can be easily seen that L is maximized over a_0 when $a_0 = \arg[\mathrm{D}(a_1, a_2, \dots a_p)]$. Also L is maximized over $[a_1, a_2, \dots a_p]$ when $D(a_1, a_2, \dots a_p) = D(\hat{a}_1, \hat{a}_2, \dots \hat{a}_p)$, where the " ^ " denote the parameters which maximize $D(a_1, a_2, \dots a_p)$. Thus the maximum of L over all a_i parameters is given by

$$\max_{[a_0, a_1, a_2, \dots a_p]} L = 2A|D(\hat{a}_1, \hat{a}_2, \dots \hat{a}_p,)| - A^2. \quad (69)$$

This expression is maximized over A when $A = |D(\hat{a}_1, \hat{a}_2, \dots \hat{a}_p)|$. Hence the maximum of L over all parameters is

$$\max \ L = \max |D(\hat{a}_1, \hat{a}_2, \dots \hat{a}_p)|^2. \quad (70)$$

This result is a simple generalization of the result derived in [52] for a stationary tone. Solving for the ML estimates necessitates maximizing $|D(\boldsymbol{\alpha})|$ over a p dimensional space, which is difficult for a large value of p. For this reason it was proposed in [9] that p be made small (e.g., $p = 2$ or 3), and a window length chosen so that the signal's IF law could be approximated by the polynomial within the window. A new estimation of the parameters is performed each time the window slides forward, and one searches over a fine grid centered on the existing parameters. This type of estimation is analogous to one of the adaptive IF estimation techniques. The size of the grid over which the search must be performed for each new data sample plays a role similar in some senses to the adaptation constant in the LMS algorithm. If the grid is chosen too small, accurate tracking will not occur; while if the grid is chosen to be too large, tracking will occur but there will be a large amount of computation required.

The maximization of $|D(\boldsymbol{\alpha})|$ in (69) can be implemented practically with the use of FFT's. If $p = 1$, one has only to find the a_1 which maximizes $|D(\boldsymbol{\alpha})|$, i.e., the frequency corresponding to the maximum value of the signal's spectrum. If $p = 3$, say, one must first dechirp the signal by the dechirping function, $\exp(ja_2 n^2 + ja_3 n^3)$, for all a_2 and a_3, and search for the global maximum spectrum value. Computation can be reduced very considerably by making a search over a coarse grid, followed by a fine search. Alternatively, one may use a multidimensional FFT implementation [44]. Details of the estimation procedure for the parameters vector, $\boldsymbol{\alpha}$, are given in Appendix C.

D. IF Estimation for Multicomponent Signals

Many practical applications require that the IF laws of multicomponent signals be estimated. It is important, then, to consider which of the methods which have been presented can be readily extended to the multicomponent case. Those methods which are based on peak detection in the tf plane should be capable of processing multiple signals. One will encounter similar problems to those found in the stationary case, however. That is, useful estimates will only be able to be found if the frequency laws are clearly separable (or resolvable) [63]. Thus for two widely separated FM laws, one may use the ML method described in Section IV-C, and extract the two IF laws which maximize the *polynomial Fourier transform.* For closely spaced laws, parametric techniques such as the RLS or LMS algorithms will be necessary. Similarly, the method based on tf filtering and first moment calculation should be a viable way of processing multiple components [13]. The methods which could be used for the multiple component case, then, would be the adaptive RLS and LMS algorithms, the spectrogram and XWVD peak detection methods, the ML based polynomial phase coefficient estimation methods, and the TFD moment based technique [23], [37], [38].

An interesting approach to generalizing the IF for multicomponent signals has been provided in [28]. There the author reasoned that since the output of the FFT is simply a sequence of filtered subcomponents of the signal, one may differentiate the phase of these outputs with time to yield an "Instantaneous Frequency Distribution." The concept was applied to the tracking of formats in speech signals. McMahon and Barrett [44] used a very similar approach in deriving their "Phase Interpolation Estimators," and putting them to use in tracking underwater acoustic frequency lines.

The problem of tracking the individual IF's of a multicomponent signal in the presence of noise was also

Table 1 Comparison of IF Estimation Techniques (Obtained from simulations using a linear FM chirp signal in additive noise).

ESTIMATOR	SNR THRESHOLD (FOR N=128)	COMPUTATIONAL REQUIREMENTS
Phase Differencing	–	N complex multiplications, N phase conversion operations N real multiplications
Zero-Crossings	–	$3N$ real additions, N real multiplications
LMS (Order 1)	–	$3N$ complex multiplications, $2N$ complex additions N real multiplications, N phase conversion operations
RLS (Order 1)	–	$4N$ complex multiplications, $2N$ complex additions, $3N$ real multiplications, N real additions, N phase conversion operations
Smoothed Phase Differencing	18 dB	N complex multiplications, N phase conversion operations, NW real multiplications, NW real additions
Linear Least Squares Estimation of Phase Polynomial	8 dB	$N^2p^2 + p^2 + 2Np + Np^2$real multiplications, $N^2p^2 + p^2 + 2Np + Np^2$ real additions, N phase conversion operations
STFT Peak	–	$NW/2 \log W + FNW + NW/2$ complex multiplications, $NW \log W + FNW$ complex additions, N real multiplications
WVD Peak	1 dB	$NW/4 \log W + N(W+1)/4 + FNW/2$ complex multiplications, $NW/2 \log W + N(W+1)/4 + FNW/2$ complex additions, N real multiplications
ML Estimation of Phase Polynomial	-3 dB	$(G(N/2 \log N + FN + 3N/2))^{p-1}$ complex multiplications, $(G(N \log N + FN))^{p-1} + Np$ real additions, $N(p-1)G^{p-1} + Np$ real multiplications, $N(p-1)G^{p-1} + Np$ real additions NG^{p-1} phase conversion operations
Adaptive Estimation of Phase Polynomial	–	$2(N/2 \log N + FN + 3N/2)$ complex multiplications, $2(N \log N + FN)$ complex additions, $6N + 2$ real multiplications, $6N + 2$ real additions, $2N$ phase conversion operations
XWVD Peak	–4 dB	$INW/2 \log W + IN(W+1) + FINW + INW/2$ complex multiplications, $INW \log W + IN(W+1) + FINW$ complex additions, IN real multiplications, IN real additions, IN phase conversion operations

addressed by DiMonte and Arun [23]. The assumptions made were that the individual IF's are slowly varying and that the noise is white, zero-mean and Gaussian. The authors performed a singular value decomposition (SVD) of the data matrix and then approximated it by its principal singular vectors and singular values, in order to achieve noise suppression. The number of components and the IF tracks are estimated from the principal singular vectors of a Hankel matrix [23]. Detailed discussion of IF estimation for multicomponent signals is beyond the scope of this paper.

V. Statistical/Computational Comparison of IF Estimation Algorithms

In this section the various IF estimation algorithms are compared with one another. The two criteria of comparison used are the statistical performance and the computational complexity of the estimator. The computational evaluation is made through a table which details the approximate number and type of calculations required for each algorithm, while the statistical comparison is made by graphing the average variance of each method against the CR bound [14].

A. Comparison in Terms of Computation and Variance

Table 1 lists the various estimation methods described and gives for each an indication as to the SNR threshold, and computational requirements of the method [14]. All entries in the table are meant only as a guide, since each will depend critically on a number of things such as the type of implementation which is used, etc. The computational requirements listed are for the basic algorithm, without special modifications. These requirements have been calculated, based on the determination of the IF for N points within a window in the data sequence,

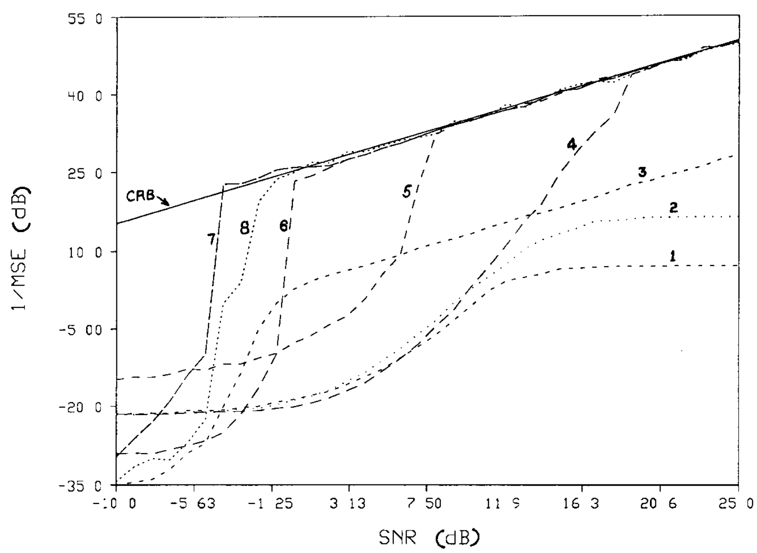


Fig. 3. Comparison with the CR bounds for IF estimates obtained using: 1) zero crossings; 2) RLS and LMS; 3) spectrogram peak; 4) CFD with Kay's Window; 5) linear least squares (estimating 2 parameters); 6) peak of WVD; 7) peak of XWVD; 8) ML techniques.

assuming that a complex analytic signal is available as input data. If a particular method does not require an analytic signal as input (e.g., the Zero-Crossing Estimator), then a small saving may be made. If the analytic signal is generated with an FIR Hilbert transformer then this saving would be approximately $(L_f + 1)N/2$ real multiplications and $(L_f + 1)N/2$ real additions, where L_f is the filter length. For comparison purposes, divisions are considered to be equivalent to multiplications, and subtractions are considered to be equivalent to additions. In order to keep the comparison simple, a number of factors such as the storage requirements, the number of comparison operations, etc., are not considered.

Symbols used in the table are: N is the length of the data sequence, W is the window length, I is the number of iterations in the XWVD based scheme, p is the order of the polynomial model for the phase, F is the number of "fine" (inter-bin) searches used to maximize an FFT, and G is the number of points in each dimension of the search grid for the ML polynomial phase method. The term phase conversion operation refers to a conversion between the rectangular and polar coordinate systems.

The statistical comparison has been made by plotting the variance obtained through simulation with theoretically derived lower variance bounds. The signal chosen for comparison was a linear FM signal, which chirped from 10 to 90 Hz, and had a sampling rate of 200 Hz. The IF at each time index was calculated using a 128 point running window. The reciprocal of the MSE for each estimate is plotted as a function of SNR in Fig. 3. Figure 3 shows that the XWVD based method gives best performance, in that it has the lowest threshold of those methods which meet the CR bound. It must be stressed, however, that this will only be the case if the signal is long compared with the window length used, so that end effects may be neglected [10]. The WVD, which has a threshold about 6 dB higher, would be preferred over the XWVD method for the SNR range in which it attains the CR bound, since it is computationally simpler. Kay's estimator, without specific extension to the polynomial phase model, meets the CR bound, but only at very high SNR. The ML and linear least squares based method also closely approach the CR bounds, with the ML method having a very low threshold. The spectrogram never meets the CR bound corresponding to the 128 point window, because the optimal window length for the spectrogram is shorter than for the other methods. The smoothed RLS, LMS, and zero-crossing estimators all perform quite poorly because of their suboptimal smoothing procedures.

The comparison which has been made, and the conclusions which have been drawn, are only strictly valid for signals with linear frequency laws. However, the table provides a good rule of thumb of what to expect for signals that may be locally approximated by a linear FM signal. Section V-B presents simulation results for a signal with nonlinear frequency law: a sinusoidal FM law.

B. Performance of Algorithms on Nonlinearly Modulated FM Signals

Computer simulations have been run to compare the performance of the IF estimation techniques on nonlinear FM signals—in particular, a sinusoidally modulated FM signal. The signal considered was a sinusoidal FM signal of 1024 points, the IF law being shown in Fig. 4(a). The signal was imbedded in 0 dB white Gaussian noise, and estimates based on b) the direct definition in (6), c) a

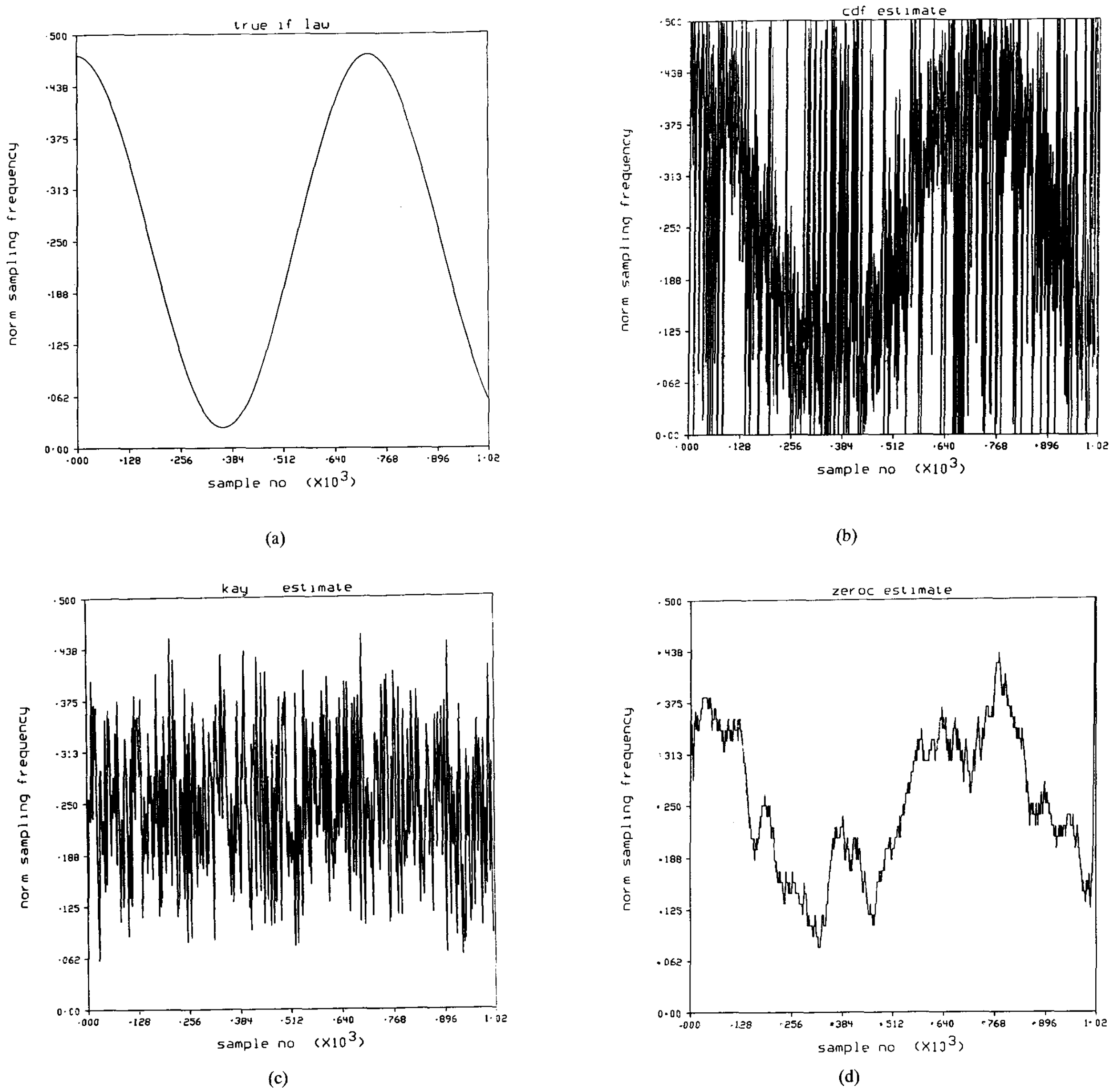
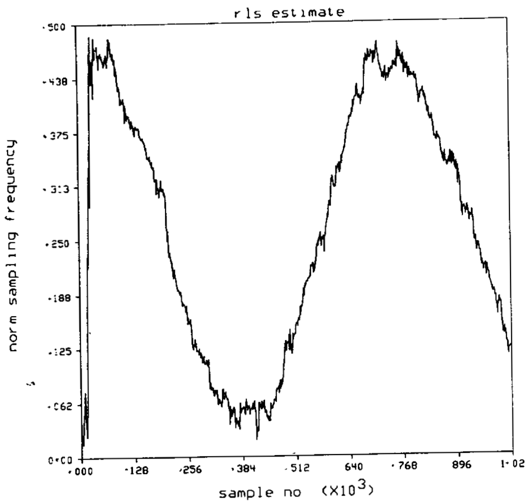


Fig. 4. (a) True instantaneous frequency law of an FM signal, $s_1(t)$. (b) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using the CFD estimator. (c) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using a smoothed phase difference estimator. (d) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using the Recursive Least Squares algorithm.

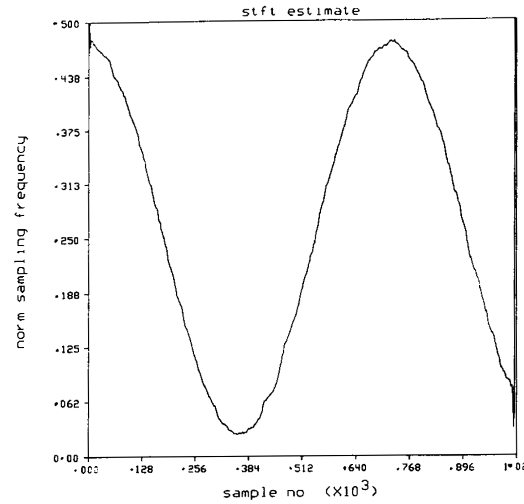
smoothed version of the direct definition in (16), d) the RLS algorithm, e) an averaged zero crossing measure f) the Spectrogram peak, g) the WVD peak, h) the XWVD peak, i) ML based phase polynomial coefficient estimation j) polynomial coefficient estimation using gradient descent algorithm, k) linear least squares polynomial coefficient estimation are shown respectively in Fig. 4 (b)–(k). It is seen that the only estimates which are low in variance are the XWVD based method and the ML polynomial phase model based methods.

The second set of simulations was performed on the same signal, but this time imbedded in −4 dB noise. The IF Estimates using the same methods as for signal 1 are shown in Fig. 5(a)-(j). For these simulations the spectrogram peak, the XWVD method and the ML method are seen to yield low variance estimates. The WVD peak and the RLS estimates are moderate in variance, while all others are high.

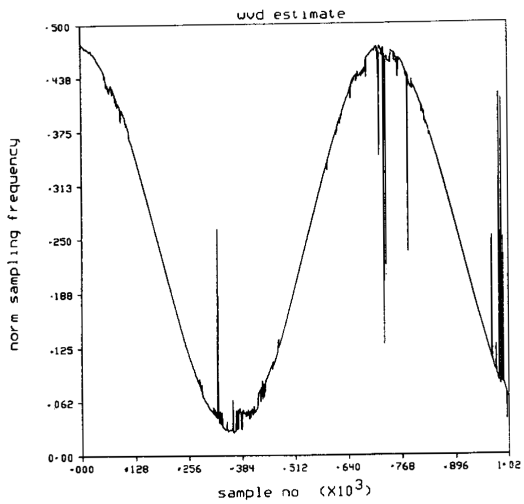
The linear least squares based estimates in Figs. 4(k) and 5(j) are seen to be a fairly poor estimates. This is as predicted for low SNR environments. However the algorithm performs well at high SNR (Table 1). Figures



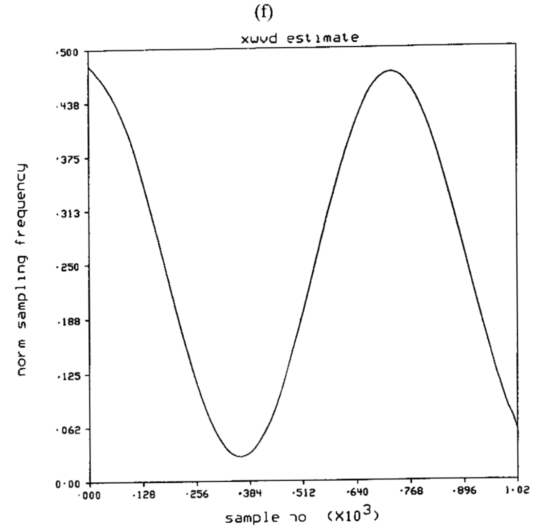
(e)



(f)



(g)



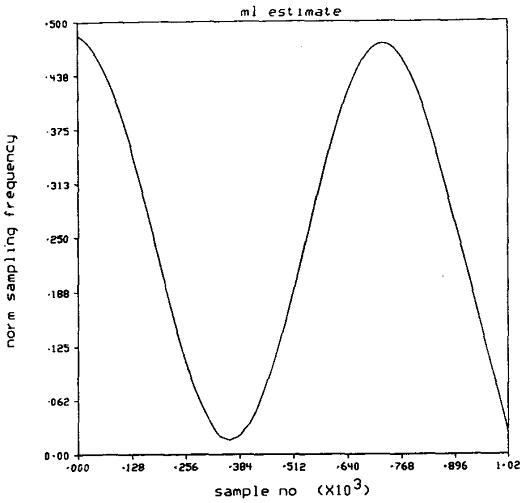
(h)

Fig. 4. *(continued)* (e) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using a zero-crossing estimator.(f) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using the STFT peak. (g) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using the WVD peak. (h) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using the XWVD peak.

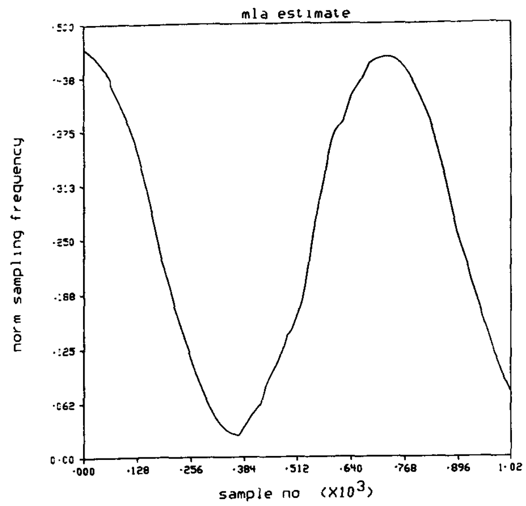
6 and 7 show IF estimates which have been estimated respectively from a linear FM and a sinusoidal FM signal. The estimates are very close reproductions of the true laws at high SNR.

C. Performance of Algorithms on Real Data

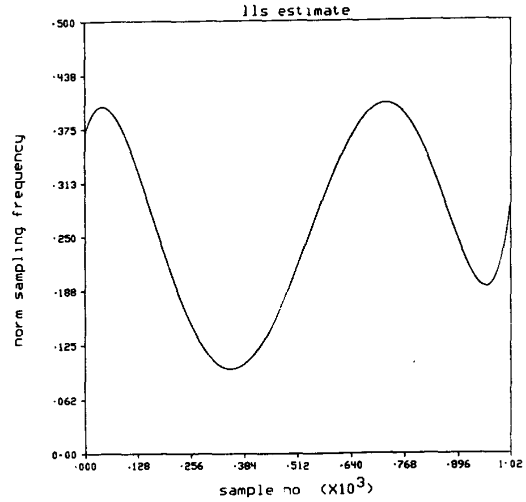
The real data set used for comparing the algorithms was an unknown chirp which corresponded to time-varying acoustic Doppler information, generated by a helicopter passing over an observer. It corresponded to a long (13 440 samples) comparatively slowly chirping signal, with an adjusted sampling rate of 444 Hz. Figures 8, 9, 10, 11, 12, 13, and 14 show respectively the IF estimates obtained using a smoothed phase differencing, smoothed RLS estimates, STFT peak estimates, WVD peak based estimates, XWVD peak based estimates, linear least squares estimation of polynomial phase model coefficients, and adaptive ML estimation of the polynomial phase model coefficients. The information which can be extracted from the IF measurements is 1) the speed of the aircraft which is related to the frequency corresponding to the flat region at the beginning of the plot, and 2) the altitude which is related to the frequency slope in the rapidly chirping section of the signal [25]. The XWVD method and the adaptive ML technique appear to give the best estimates, based on the expected IF law model.



(i)



(j)



(k)

Fig. 4. *(continued)* (i) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using ML estimation. (j) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise using ML estimation with adaptive searching. (k) Instantaneous frequency estimate of $s_1(t)$ in 0 dB white Gaussian noise by using Linear Least Squares Estimation.

VI. Applications

A. Use of the IF to Detect Harmonically Related Signals

The occurrence of harmonically related signal components is very common in many natural situations. When the fundamental is stationary, Fourier Transform based techniques give an effective measure of the harmonic content, as well as of the system noise. When, however, the signal spectrum is time-varying, these techniques can become inadequate due to poor resolution. Even the use of the STFT, which was introduced to take account of nonstationarities, has limited use. If the IF of the fundamental can first be estimated, though, it can be incorporated into the formation of the time-varying spectral representation, such that good resolution in time and frequency can be achieved. This may be done practically by using a smoothing procedure in the WVD plane, which tends to preserve energy in the region where the harmonics would be expected, and suppresses energy elsewhere [7]. The general MAP estimator for time-varying IF estimation of the fundamental of a harmonic series is also addressed in [67].

B. Automatic Time-Varying Filtering

In many applications such as speech, seismic or underwater acoustics, the ability to perform time-varying filtering can be very helpful. Signal enhancement and signal separation are two areas of application. To perform this time-frequency filtering, a time-varying system transfer

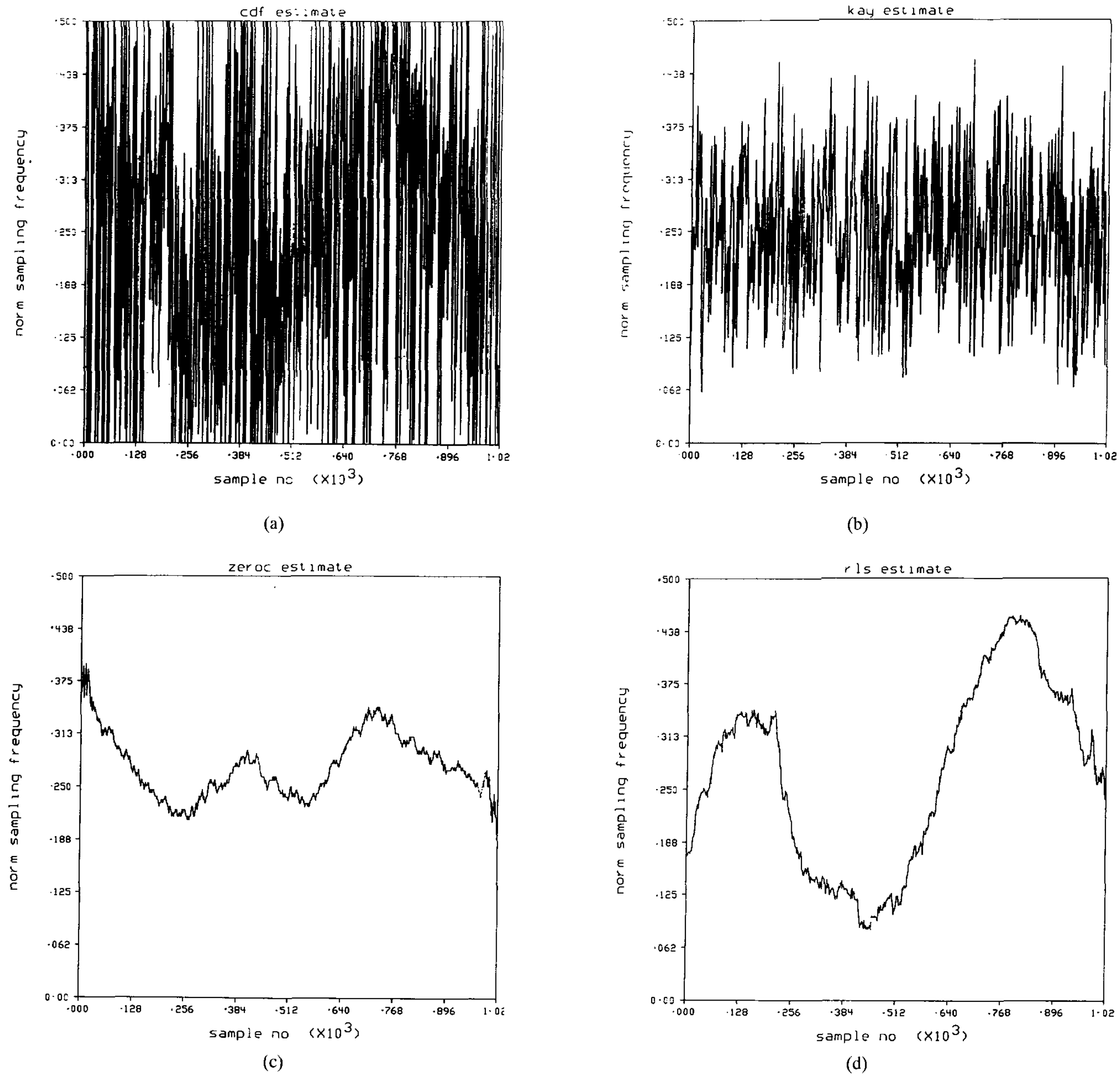


Fig. 5. (a) IF estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using the CFD estimator. (b) IF estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using a smoothed phase difference estimator. (c) IF estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained by using a zero-crossing estimator; (d) IF estimate of $s_1(t)$ in 0 dB white Gaussian noise obtained using the RLS's algorithm.

function, $H(t, f)$, must be specified in accordance with the application. The WVD, being an entirely real transform, is well suited to time-varying filtering, particularly for monocomponent signals. The time-varying filtering operation is given by

$$y(t) = W^{-1}[W(t, f) \cdot H(t, f)] \tag{71}$$

where $W(t, f)$ is the WVD of the input signal to be processed, and $W^{-1}[\;]$ represents the synthesis operator, i.e., the operator which finds the time domain analytic signal whose WVD is closest to the operand in a LMS's sense. If the operand is a valid WVD, the synthesis operator is equivalent to WVD inversion.

The time-varying filter operation, in the discrete domain, consists of the following steps:

1) Design of a time-varying transfer function, $H(n, m)$.
2) Calculation of the Discrete WVD (DWVD) or windowed DWVD of the input signal, $s(n)$.
3) Synthesis of $y(n)$ from the $W(n, m).H(n, m)$ product.

Ideally the time-varying filter would capture all the signal energy but reject all the noise. Clearly the ideal is not generally realizable, and one must settle for a trade-off in which the significant signal energy is captured, while most of the noise is eliminated. A useful approach is to window about the IF, since it is known that, for monocomponent

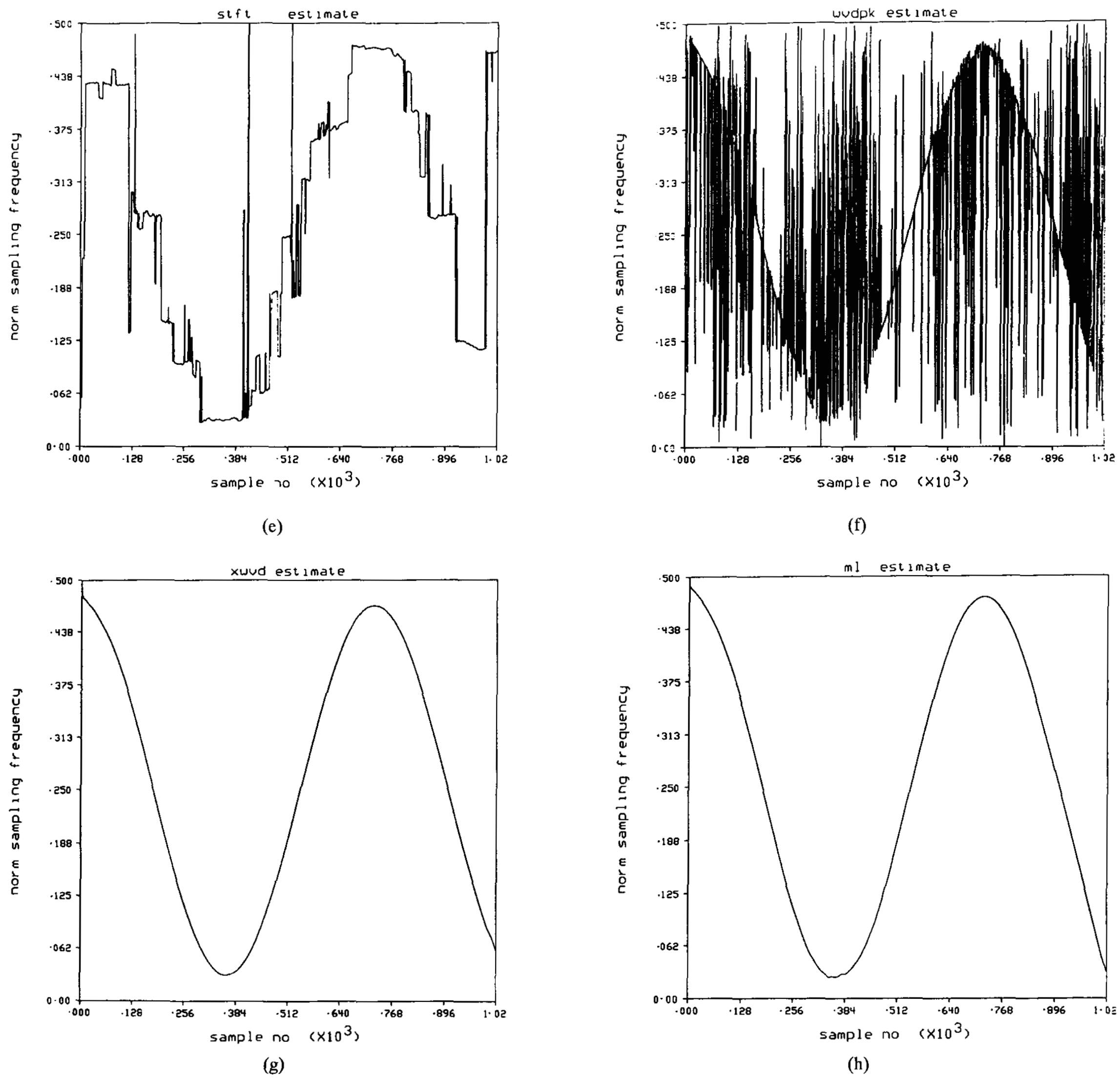


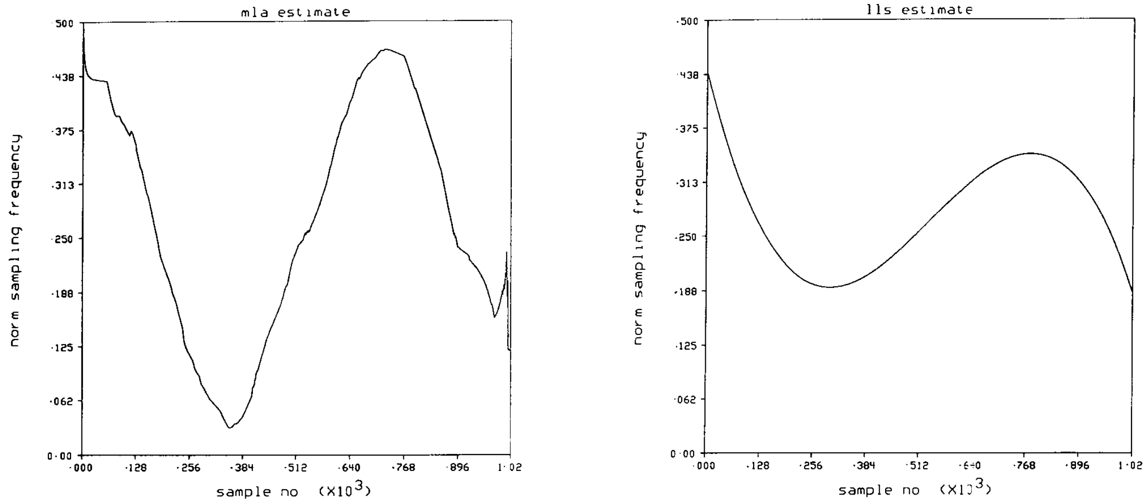
Fig. 5. *(continued)* (e) IF estimate of $s_1(t)$ in -4 dB white Gaussian noise obtained using the STFT peak. (f) IF estimate of $s_1(t)$ in -4 dB white Gaussian noise obtained using the WVD peak. (g) IF estimate of $s_1(t)$ in -4 dB white Gaussian noise obtained using the XWVD peak. (h) IF estimate of $s_1(t)$ in -4 dB white Gaussian noise obtained using ML estimation.

signals, the signal energy tends to be concentrated there. The various techniques proposed earlier, then, may be used for IF estimation as a first step in the design of the automatic filter. The bandwidth of the time-varying filter may be selected according to the application, as can the window type (e.g., Rectangular, Bartlett, Hanning, Hamming, Blackman, etc. [45]). An automatic time-varying filter was designed, using the above ideas, and applied to a monocomponent signal [11].

C. Seismic Processing

In a typical seismic situation, a vibrioses signal is transmitted into the earth and the various subterranean layers provide some degree of reflection of this source signal. The different arrival times of these reflected waveforms are often used to reveal information as to the geological structure of the region. The reflected returns can also be processed, however, to yield information as to the amount of dispersion present in the formation. This can be determined simply by comparing the IF of the received signal with that of the source.

IF and instantaneous phase attributes are generally used to highlight seismic events which are not apparent using normal amplitude based data processing. They have been proved useful in the detection of some hydrocarbons, and some geological structures. More recent work has shown that IF can be used to determine frequency dispersion in a vibrioses signal, which helps determine the absorbing



(i)

(j)

Fig. 5. *(continued)* (i) IF estimate of $s_1(t)$ in 0 dB white Gaussian noise using ML estimation with adaptive searching; (j) IF estimate of $s_1(t)$ in 0 dB white Gaussian noise by using Linear Least Squares estimation.

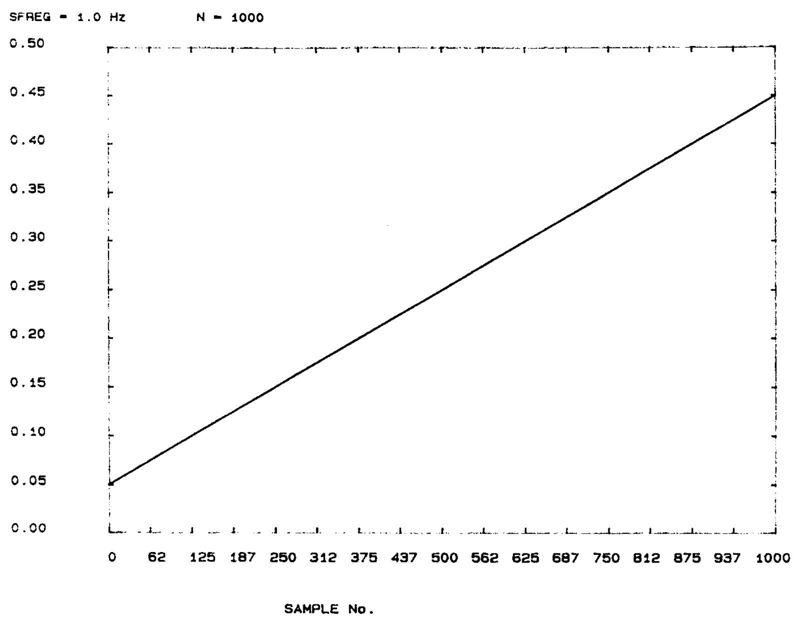


Fig. 6. Linear least squares IF estimate of a linear FM signal.

nature of a traversed rock.

1) Interpretation of Seismic IF and Phase: The instantaneous phase of a seismic signal emphasizes the continuity of geological events. Phase is independent of reflection strength and can make weak coherent events clearer [57]. Consequently, when weak reflections are received, phase is directly useful for delineating the normal geological structures such as discontinuities, faults, pinchouts, etc. that would otherwise interfere with each other. Traditionally, phase displays are colored using the colors of a color wheel [57].

IF is used to delineate more complicated events. Reflections are often composed of several individual reflections from several closely spaced reflectors. The individual reflections would normally be associated with reflectors which are nearly constant in acoustic impedance. This superposition of reflections can result in the production of a frequency pattern (i.e., IF function) which would

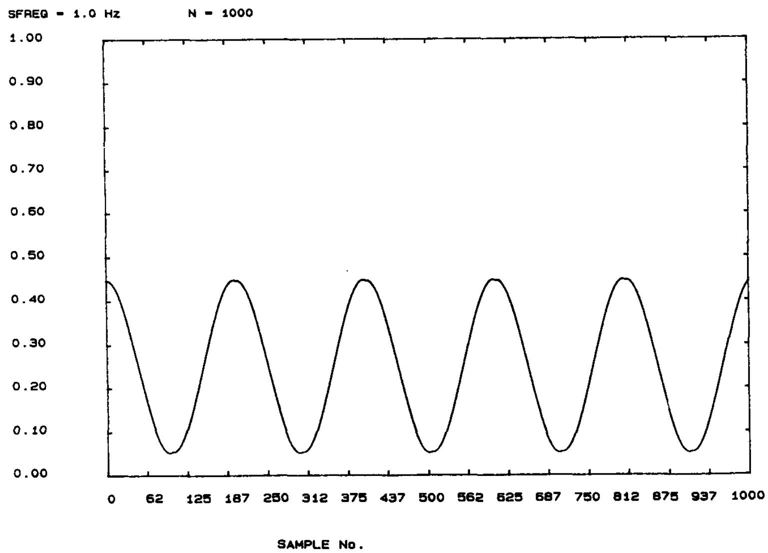


Fig. 7. Linear least squares estimate of a sinusoidal FM signal.

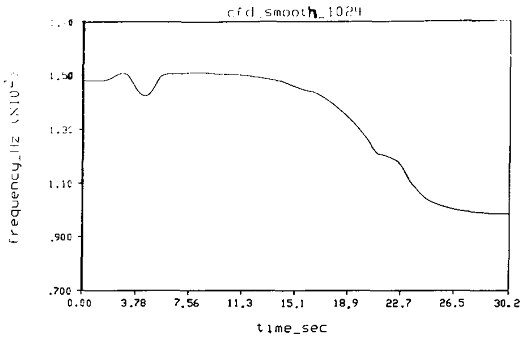


Fig. 8. Smoothed CFD estimate of the IF law for a time-varying Doppler shifted signal produced by a helicopter flying over an observer.

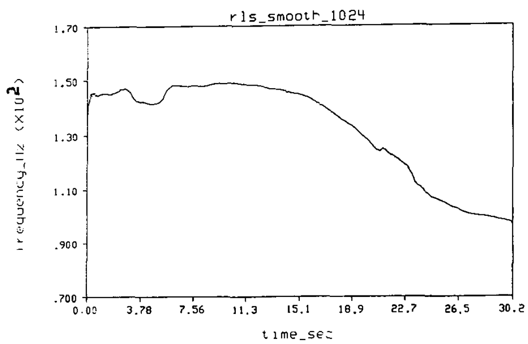


Fig. 9. Smoothed RLS estimate for the IF law of the time-varying Doppler shifted sound signal.

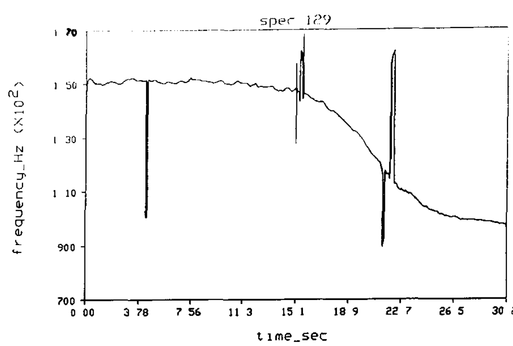


Fig. 10. Spectrogram peak estimate for the IF law of the time-varying Doppler shifted signal.

characterize a particular composite reflection. Thus $f_i(t)$ is often used to correlate from one set of seismic data to another. The frequency character will change as the sequence of layers forming a composite reflection changes in lithology or thickness. It has been observed [57] that pinchouts and edges of hydrocarbon-water interfaces (oil field extremity) change $f_i(t)$ more rapidly.

An empirical observation given by Taner [57] involves the shifting of frequency for certain geological features. Quite often, when a reflection is made below a gas sand, condensate or oil reservoir, a shift toward lower frequencies is observed. It has been further observed by many people that the shift occurs only for the reflection immediately below the hydrocarbon zone while a normal frequency character is observed for deeper reflections. No full explanation has been given for this. It is simply an empirical observation. Similar shifts toward lower frequencies are also sometimes associated with fracture zones and brittle rocks.

IF plots are made in color to observe the above mentioned features. Frequency is color coded in 2 Hz steps with the red end of the spectrum being used for low frequencies and the violet end being used for higher frequencies. Frequencies

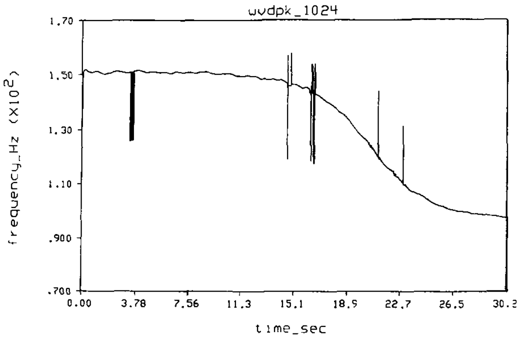


Fig. 11. WVD peak estimate for the IF law of the time-varying Doppler shifted signal.

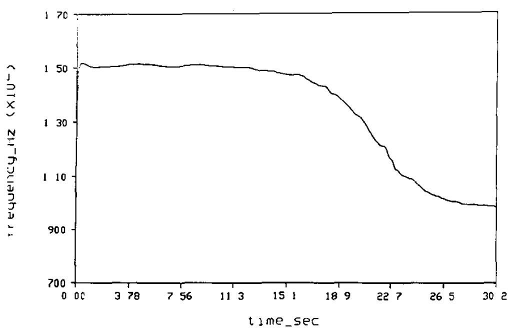


Fig. 12. XWVD peak estimate for the IF law of the time-varying Doppler shifted signal.

below 5 Hz are normally not colored.

The IF was used in [5] to study borehole recorded Vibroseis signals. A typical Vibroseis chirp has a bandwidth, B, and a duration, T, the product of which is large. The method was applied to a single component downgoing chirp (i.e., reflected components removed) to derive an empirical relationship between IF and dispersion effects. Frequency dispersion is simply described as an occurrence where different frequency components of a signal travel at different velocities [73].

D. IF in Radar Processing

In this section, two applications of the use of IF in radar signal processing are examined. We consider a target which is to be detected, tracked and imaged.

In the first application, the target will be considered to be a single point scatterer with its radial velocity being time-varying. It will be shown that in this case, the variations in radial velocity can be descripted by the IF. This provides information about the direction of the moving target [42].

In the second application, the radial velocity of the target is assumed constant and the target is considered to consist of a collection of point scatterers. It will be shown that in this case, the Doppler induced frequency from each of these point scatterers can be measured by their IF [42].

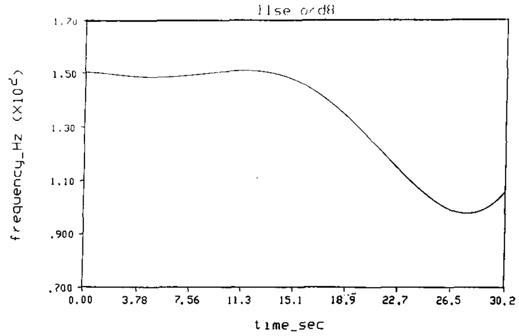


Fig. 13. Linear Least Squares estimate for the IF law of the time-varying Doppler shifted signal.

1) IF Applied to Frequency Tracking of the Target: In this application the target is described by a single point reflector which undergoes variations in its radial velocity. The range of the target becomes a function of time as indicated by (72).

$$R(t) = R_0 - v(t) \cdot t. \tag{72}$$

The variable v here is considered to be a vector. For example, if the target is traveling towards the receiver, then the Doppler induced frequency is positive, while when the target is traveling away from the receiver, the Doppler induced frequency is negative. The time-varying delay is given by (73) [61].

$$\tau(t) = \frac{2R}{c}\left(t - \frac{\tau(t)}{2}\right). \tag{73}$$

The $\tau(t)/2$ appears in (72) because of the variations in the range which occur at time $t - \tau(t)/2$. Substituting (72) into (73) yields:

$$\tau(t) = \tau - \frac{2v(t)}{c}t \tag{74}$$

c is the propagation velocity of the illuminating wave. The backscattered signal at the receiver is given by (75) [61].

$$s_r(t) = \mathrm{Re}\left\{\sqrt{E_t}z(t-\tau)\exp\left[j2\pi f_c\left(t + \frac{2v(t)}{c}t\right)\right]\right\} \tag{75}$$

where $\sqrt{E_t}$ is the energy of the transmitted signal, $z(t)$ is the analytic form of the transmitted signal and Re denotes the real part of the expression.

Since the radial velocity is time-varying, then, the Doppler induced frequency will also be time-varying producing a nonstationary spectrum. Consequently, IF can be used to track this variation in the frequency and give an indication on the direction of the target [42].

2) IF of Individual Point Scatterers: Consider a body of a target, whose surface is rough with respect to the wavelength of the illuminating wave, and which is under some form of rotation in space. The body illuminated by the incident wave is shown in Fig. 15.

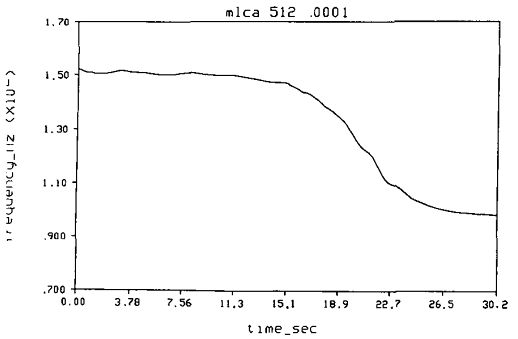


Fig. 14. ML (adaptive) estimate for the IF law of the time-varying Doppler shifted signal.

The distance, d, between the transmitter and the receiver corresponds to a delay time, τ, which is a function of the propagation velocity in the direction of the line of sight (the direct line between the transmitter and receiver). It is assumed that the surface variations on the body are such that when the incident wave undergoes reflection, the phase of the backscattered signal is uniformly distributed. Since a number of point scatterers may occupy the same time delay, then the resulting random variable representing the reflection process can be considered to be Gaussian distributed. Therefore, the backscattered signal due to point scatterers is nonstationary since the characteristics which describe the reflection process are nonstationary. For a nonrotating object the backscattered signal for each delay τ is given by (76):

$$r(\tau, t) = \mathrm{Re}\{b(\tau)z(t-\tau)\} \tag{76}$$

where $b(\tau)$ is the stochastic process at time delay τ from the receiver, $z(t)$ is the analytic form of the transmitted signal, $s(t)$. Hence the total backscattered signal from a nonrotating body is given by (75).

$$r(t) = \mathrm{Re}\left\{\int_{-\infty}^{+\infty} b(\tau)\, z(t-\tau)\, d\tau\right\}. \tag{77}$$

This equation can be considered to be a filtering operation which does not depend on time [42].

Because the body does undergo a rotation, the backscattered signal from each delay is effectively modulated by different numbers of point scatterers passing through the delay time. Equation (77) can be modified to include this modulation effect and is rewritten as (78).

$$r(t) = \mathrm{Re}\left\{\int_{-\infty}^{+\infty} b(t,\tau)\, z(t-\tau)\, d\tau\right\}. \tag{78}$$

This can be considered to be a time-varying filtering operation. If this random process is assumed to be a stationary process then (78) can be rewritten in the form given by (79).

$$r(t) = \mathrm{Re}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} D(\tau,y)\, z(t-\tau)\, e^{j2\pi yt}\, dy\, d\tau\right\}. \tag{79}$$

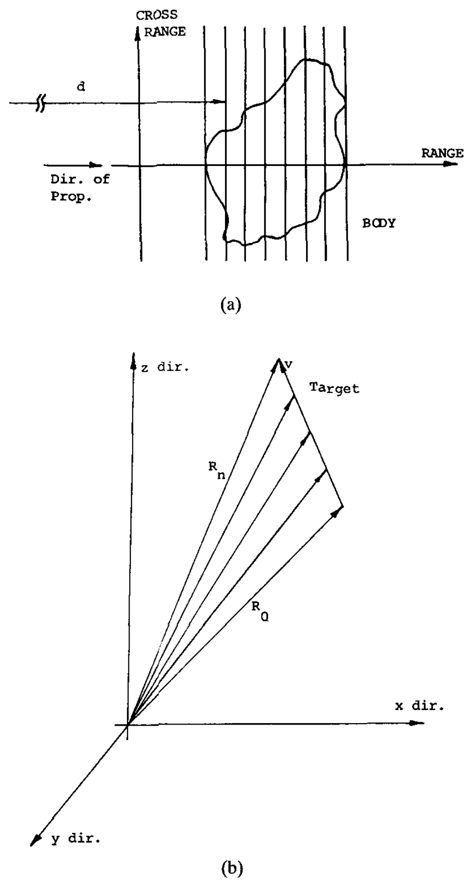


Fig. 15. (a) Rotating body in space. (b) Variations in radial velocity.

Here $D(\tau, y)$ is the reflectivity function of the target. Forming the expected value of the Wigner–Ville distribution of the backscattered signal, and assuming certain coherence conditions gives the relationship shown in (80) [8]:

$$E[W_r(t,f)] = \sigma(t,f) {*}{*} W_S(t,f) \tag{80}$$

where $**$ denotes convolution in both time and frequency. The function $\sigma(t,f)$ denotes the scattering function of the target. Equation (80) can be interpreted as the WVD of the transmitted signal moved to points in the time frequency plane representing points in the scattering function. The Doppler shift in frequency can be considered to be a *localized IF* which represents a point scatterer traveling at a certain velocity, v with the Doppler induced frequency given by

$$f_d = f_c \frac{2v}{c}. \tag{81}$$

Hence a point on the target traveling at velocity v and a given distance d from the receiver is mapped onto a location in the time-frequency plane described by the IF.

E. Other Applications

1) Sonar: Frequency line tracking is done very commonly in sonar environments [4]. The IF estimation techniques used in this paper may be used to provide initial estimates for the tracking process.

2) Biomedical Applications: Sclabassi and coworkers [54] used the IF to analyze electroencephalogram (EEG) signals, continuous time-varying voltages reflecting ongoing activity in the brain. The data set was recorded during sleep in infants. The IF provides useful information, because frequently, the neurophysiologist is interested in knowing the major frequency of the EEG at a given time regardless of the distribution of the spectrum. It was also reported in [54] that the IF is an effective tool for the analysis of the blood flow signals. The tracking of the IF is used to study blood flow in an area where the flow pattern is unpredictable due to the complex geometry.

3) Speech: In order to track the important time-varying information in speech, Friedman developed an "IF distribution." This is a technique which applies the normal definition of the IF (i.e., the time derivative of the phase) to the DFT. That is, rather than differentiating the phase of the signal itself, Friedman differentiated the phase of the Fourier components of the signal [28]. This enables robust IF estimation of the various subcomponents of the speech signal.

4) Underwater Acoustics: McMahon and Barrett [44] used a similar approach to Friedman in estimating the IF of narrowband frequency components in underwater acoustic signals. They were able to achieve very low variance estimates for both single and multicomponent signals. Barrett and Streit later applied hidden Markov model techniques to improve further the quality of their frequency law estimates [4].

5) Oceanography: Imberger and Boashash [34] used the IF to measure the kinetic energy dissipation of turbulent water using measurements obtained from temperature gradient microstructure recorders. The relationship between the kinetic energy and the IF parameter is detailed in [34].

VII. Conclusions

A comparison of available algorithms for IF estimation of a monocomponent nonstationary signal has been made and it has been seen that the choice of estimator depends on the SNR and the signal class.

The smoothed phase difference estimator has the advantage of being computationally very simple and meets the CR bound for high SNR signals. Thus for high SNR signals which are nearly stationary, this estimator may be used.

For nonstationary signals at high SNR one may also use polynomial phase modeling with the polynomial parameters being found by nonlinear regression of the phase, or by a nonlinear regression on local IF estimates.

The peak of the WVD has been shown to be optimal for linear FM signals at high to moderate SNR. Thus for high to moderate SNR signals which can be approximated well (at least locally) by linear FM signals, this estimator should be used.

For low SNR nonstationary signals, which are comparatively long, the iterative XWVD scheme may be used with a sliding window. For short very low SNR signals, it seems that only a combination of polynomial phase modeling and ML techniques is able to achieve low variance estimates.

It should also be noted that further improvement in the overall estimation of the IF law may be achieved by use of a tracking algorithm [4]. Because of the robustness of certain tracking algorithms, it may be acceptable in a given application, to use computationally efficient (but suboptimal) estimation procedures, and rely on the tracking phase to remove excessive noise.

Many applications use some sort of spectral estimation techniques and often, they have time-varying characteristics. The algorithms presented in this paper should then prove very useful for all these practical applications.

Appendix 1

Derivation of Lower Variance Bounds for Polynomial Phase Modeled Estimates

The derivation of lower variance bounds for estimating the parameters of an arbitrary polynomial phase FM signal presented here is a generalization of the derivation proposed by Rife and Boorstyn for a single stationary tone [52]. It first appeared in [9] and was derived by O'Shea [47].

Consider an arbitrary discrete time signal given by

$$s(n) = A\cos(a_0 + a_1 n + a_2 n^2 + a_3 n^3 + \dots a_p n^p). \quad \text{(A1.1)}$$

Its Hilbert transform is denoted by $q(n)$. The above signal model assumes the observed signal can be represented as a constant amplitude FM signal with frequency modulation law given by a finite dimension polynomial. If this signal is imbedded in white Gaussian noise of variance σ^2, the corresponding complex N point discrete sampled signal will be given by

$$z(n) = x(n) + jy(n). \quad \text{(A1.2)}$$

To evaluate the lower variance bounds the joint pdf of the sampled observations must be obtained. If the unknown parameter vector is

$$\boldsymbol{\alpha} = [A, a_0, a_1, a_2, a_3, \dots, a_p] \quad \text{(A1.3)}$$

then the joint pdf is given by [52]

$$p(\boldsymbol{z}; \boldsymbol{\alpha}) = \left(\frac{1}{\sigma^2 2\pi}\right)^2 \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n) - s(n))^2 + (y(n) - q(n))^2\right]. \quad \text{(A1.4)}$$

To find the variance bounds it is first necessary to determine the elements of the Fisher Information matrix, J. The elements are given by the following equation [40, p. 46]:

$$J_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left[\frac{\partial s(n)}{\partial \alpha_i} \frac{\partial s(n)}{\partial \alpha_j} + \frac{\partial q(n)}{\partial \alpha_i} \frac{\partial q(n)}{\partial \alpha_j} \right]. \quad \text{(A1.5)}$$

Inversion of this matrix and extraction of the diagonal elements will then yield lower bounds on estimate variances. (See also [2] for a discussion on how "tight" this type of variance bound is for the stationary case).

The lower bounds on the parameter estimates are derived in [9]. The *approximate* bounds are:

$$\mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{N} \quad \text{(A1.6)}$$

$$\mathrm{var}(\hat{a}_0) \geq \frac{\sigma^2}{A^2} \frac{9}{N} \quad \text{(A1.7)}$$

$$\mathrm{var}(\hat{a}_1) \geq \frac{\sigma^2}{A^2} \frac{192}{N^3} \quad \text{(A1.8)}$$

$$\mathrm{var}(\hat{a}_2) \geq \frac{\sigma^2}{A^2} \frac{180}{N^5} \quad \text{(A1.9)}$$

where the ˆ superscript denotes the estimator of the given parameter.

All of these results are verifiable by simulation, since the ML likelihood estimator (derived in Section IV-C) is known to meet these lower variance bounds asymptotically, i.e., for moderate to high SNR. A detailed derivation of lower variance bounds for a polynomial phase signal is also provided in [48].

In practice, it is often useful to segment the signal and try and fit lower order polynomials to the phase segments. This is done because the lower variance bounds increase significantly as the polynomial order increases. In fact, there will be a bias/variance trade-off. If the polynomial order is made high, there will be low bias, but high variance. The opposite will obviously be true if a low order is chosen.

APPENDIX 2

Derivation of the Nonlinear Least-Squares Estimator Equations for Polynomial Phase Modeling

Take a signal estimate of the form:

$$\hat{z}(n) = \hat{A}(n)\ \exp(j\hat{\phi}(n)) \quad \text{(A2.1)}$$

where $\hat{\phi}(n)$ and $\hat{A}(n)$ are both estimated as polynomials:

$$\hat{\phi}(n) = \sum_{h=0}^{p} \hat{a}_h n^h \quad \text{and} \quad \hat{A}(n) = \sum_{h=0}^{q} \hat{b}_h n^h \quad \text{(A2.2)}$$

Then the sum of the squared errors, E, is given by

$$E = \sum_{n=0}^{N-1} |\ \hat{z}(n) - z(n)\ |^2 \quad \text{(A2.3)}$$

$$E = \sum_{n=0}^{N-1} |\ \hat{A}(n)\ \exp(j\hat{\phi}(n)) - z(n)\ |^2 \quad \text{(A2.3)}$$

Minimizing with respect to a given phase parameter estimate, $\hat{a}_h$ (for $h = 0, \ldots, p$) yields:

$$\frac{dE}{d\hat{a}_h} = 2 \sum_{n=0}^{N-1} [\hat{z}(n) - z(n)] j n^h \hat{z}(n) \quad \text{(A2.4)}$$

$$= 2 \sum_{n=0}^{N-1} [\hat{A}^2 \exp(2j\hat{\phi}(n)) - n\hat{A}(n) z(n) \exp(j\hat{\phi}(n))] j n^h \quad \text{(A2.5)}$$

Equating this to zero gives, in vector form, for all h:

$$(\hat{\boldsymbol{z}} - \boldsymbol{z})^T \boldsymbol{y}_h = 0 \quad \text{(A2.6)}$$

where $\boldsymbol{z}$, $\hat{\boldsymbol{z}}$, and $\boldsymbol{y}_h$ are vectors containing $z(n)$, $\hat{z}(n)$, and $y_h(n) = z(n)n^h$, respectively. T represents transposition.

Expanding into matrix representation then:

$$\hat{\boldsymbol{z}}^T \boldsymbol{Y} = \boldsymbol{z}^T \boldsymbol{Y} \quad \text{(A2.7)}$$

where Y is the N by $(p+1)$ matrix:

$$\boldsymbol{Y} = \begin{bmatrix} \hat{z}(0) & 0 & \cdots & 0 \\ \hat{z}(1) & \hat{z}(1) & \cdots & \hat{z}(1) \\ \hat{z}(2) & 2\hat{z}(2) & \cdots & 2^p \hat{z}(2) \\ \cdot & \cdot & \cdots & \cdot \\ \hat{z}(N-1) & (N-1)\hat{z}(N-1) & \cdots & (N-1)^p \hat{z}(N-1) \end{bmatrix}. \quad \text{(A2.8)}$$

Equation (A2.4) must now be minimized with respect to the $q+1$ amplitude parameters $b_k (k = 0, \ldots, q)$:

$$(\hat{\boldsymbol{z}} - \boldsymbol{z})^T \boldsymbol{w}_k \quad \text{(A2.9)}$$

where $w_k(n) = n^k \exp(j\hat{\phi}(n))$, so that:

$$\hat{\boldsymbol{z}}^T \boldsymbol{W} = \boldsymbol{z}^T \boldsymbol{W} \quad \text{(A2.10)}$$

where $\boldsymbol{W}$ is the $N(q+1)$ matrix shown below. The matrices $\boldsymbol{W}$ and $\boldsymbol{Y}$ will only be invertible if there are as many parameters as data points, $p+1 \geq N$ or $q+1 \geq N$. In this case the trivial solution $\hat{\boldsymbol{z}} = \boldsymbol{z}$ is obtained.

$$\boldsymbol{W} = \begin{bmatrix} \exp\left(j\hat{\phi}(0)\right) & 0 & \cdots & 0 \\ \exp\left(j\hat{\phi}(1)\right) & \exp\left(j\hat{\phi}(1)\right) & & \exp\left(j\hat{\phi}(1)\right) \\ \exp\left(j\hat{\phi}(2)\right) & 2\exp\left(j\hat{\phi}(2)\right) & \cdots & 2^q \exp\left(j\hat{\phi}(2)\right) \\ \cdot & \cdot & \cdots & \cdot \\ \exp\left(j\hat{\phi}(N-1)\right) & (N-1)\exp\left(j\hat{\phi}(N-1)\right) & \cdots & (N-1)^q \exp\left(j\hat{\phi}(N-1)\right) \end{bmatrix}. \quad \text{(A2.11)}$$

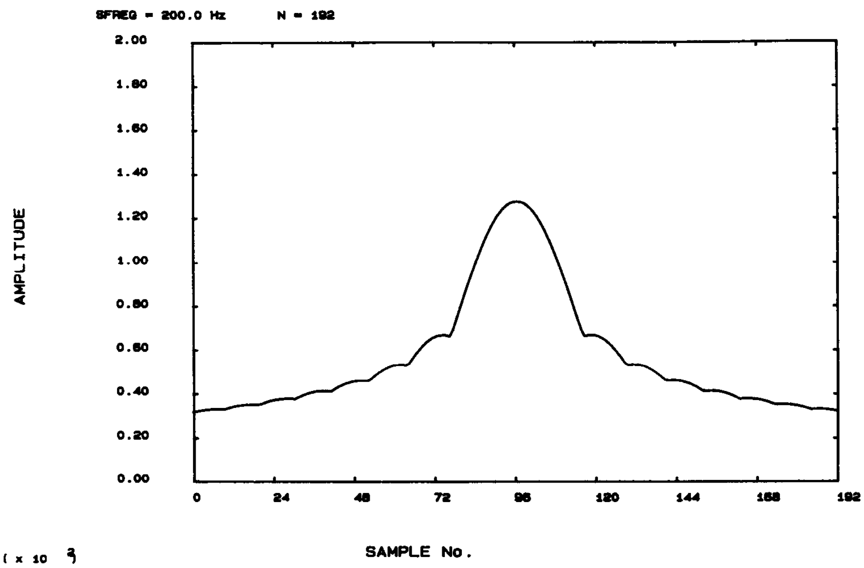


Fig. 16. $\hat{a}_2$ error function versus $\hat{a}_2$ (rectangular window).

It can also be noted that if the amplitude law is known, or can be assumed to be a constant $\hat{A}(n) = \hat{A}$ for all n, then $\boldsymbol{Y} = \hat{A}\boldsymbol{W}$ so that (A2.7) and (A2.10) become identical.

APPENDIX 3

Adaptive Estimation of Polynomial Coefficients

For the algorithm in Section IV-C to be practical, it is necessary to reduce the amount of computation associated with maximization of $|D(\boldsymbol{\alpha})|$ over the parameter vector, $\boldsymbol{\alpha}$, for ML based estimation. A number of approaches exist. An obvious technique is to limit p to 2. Many frequency laws can be approximated locally by a linear FM law, and so this is not an unreasonable constraint in many practical applications. One may then implement the $p = 2$ dimensional search by extracting the peak of a 2D Fourier transform, with frequency and frequency slope being the two transformed variables of the 2D FFT [44]. This approach is convenient in that there are efficient algorithms for implementing two-dimensional FFT's.

Alternatively, for $p = 2$, one may use gradient descent search techniques to determine a close approximation to the maximum of $|D(\boldsymbol{\alpha})|$. This reduces the computational complexity greatly. The procedure is described as follows [9], [47].

For quadratic phase functions the ML solution involves simply finding the maximum of $|D(a_1, a_2)|$ over all possible a_1 and a_2. The maximum of $|D(\boldsymbol{\alpha})|$ for a given value of a_2 is found by first dechirping (i.e., multiplying) the original signal by $\exp(-ja_2n^2)$, then Fourier transforming and extracting the peak. The maximum over all a_2 is found by performing this routine for all possible values of a_2 and extracting the global maximum. Conveniently, when the maximum of $|D(\boldsymbol{\alpha})|$ given a_2 is plotted as a function of a_2, it becomes very close to exhibiting only one maximum (see Fig. 16). There are in fact, very small local maxima due to the sidelobe phenomenon which arise in obtaining spectral estimates for finite duration signals. These local maxima may be eliminated by applying a "smooth" window, say a Hanning window, before forming the spectral estimate. Gradient based search techniques can then be applied to find the maximum of $|D(\boldsymbol{\alpha})|$ [27].

The use of gradient based search techniques can reduce the computational complexity of polynomial phase based IF estimation to only about twice the computational cost of the spectrogram for the same signal. This may be done by recursively updating the $\hat{a}_2$ estimate with each new data sample, based on a gradient measure. The unimodal nature of the function shown in Fig. 17 indicates that for a signal with unchanging a_1 and a_2 the procedure should eventually converge to the correct parameter estimates (providing the iteration size is made small enough). Figure 18(a) shows the true IF law estimate for a chirp signal of length 256 and unnormalized sampling frequency=200 Hz. The plots in (b) and (c) illustrate the adaptation to the correct IF estimate from start-up, for various values of "adaptation constant," u.

Several practical points should be made in using this approach. Firstly, the use of say, a Hanning window transforms the spectral peak versus the $\hat{a}_2$ error into an almost unimodal function. There will in fact be very small minor maxima for very large errors. Further, the curve will have a low gradient at these large error values. To steepen the gradient at these extremities, eliminate the nonmonotonicities and hence guarantee comparatively rapid convergence of the algorithm one can use very smooth windows, such as a squared Hanning window. Secondly, spurious effects due to outliers may be significantly reduced by limiting the admissable change in $\hat{a}_2$ for each iteration if the signal parameters are slowly changing.

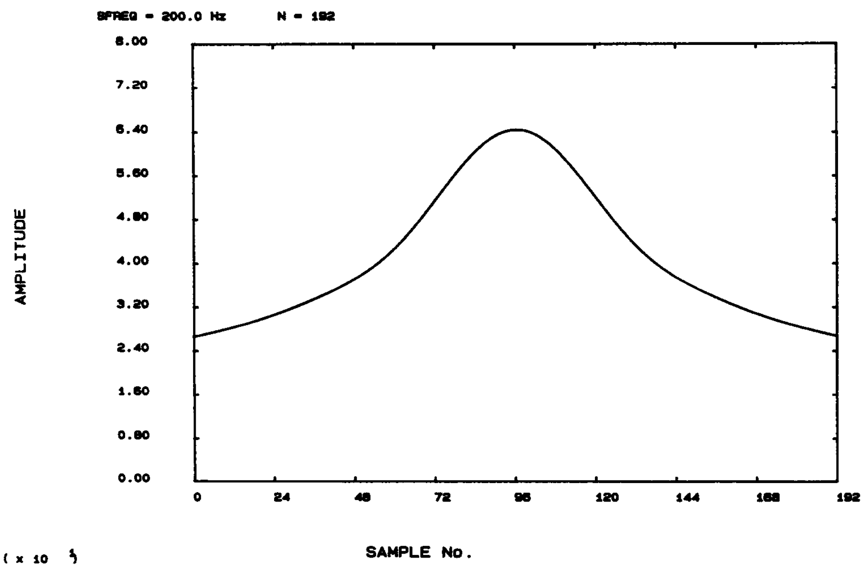
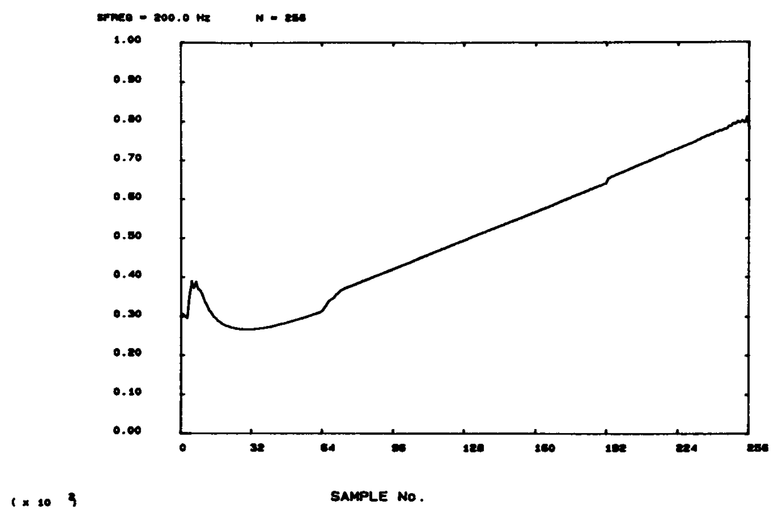


Fig. 17. $\hat{a}_2$ error function versus $\hat{a}_2$ (Hanning window).



(a)

Fig. 18. (a) True IF law of a linear FM signal.

Lower Variance Bounds for Polynomial Phase Modeled Estimates

The derivation of the lower variance bounds for estimating the parameters of an arbitrary polynomial chirp signal is presented in [9], [47] and independently in [48]. The results obtained therein are verifiable by simulation, since the ML estimator (derived in Section IV-B) is known to meet these lower variance bounds asymptotically.

The lower bounds for the parameter estimate, $\hat{a}_2$, in a quadratic phase polynomial, with unnormalized sampling frequency of 200 Hz, and duration 0.64 s are shown plotted in Fig. 19 as a function of SNR. The SNR is defined by

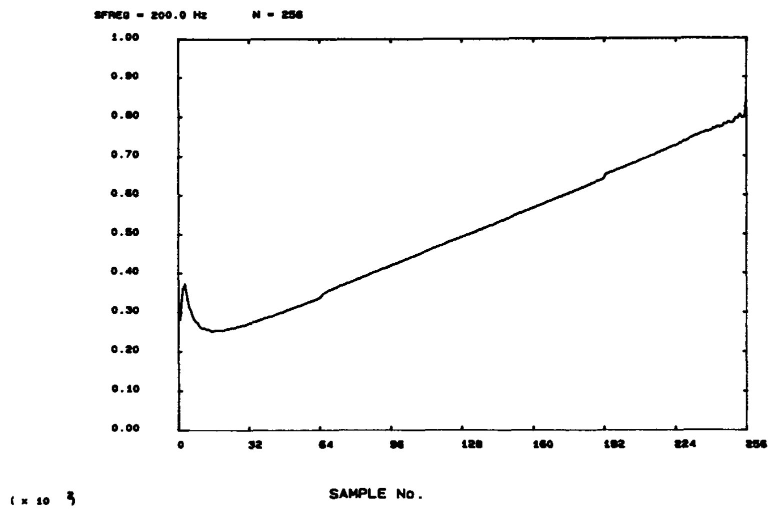
$$\mathrm{SNR} = 10 \log_{10} \frac{A^2}{E} \tag{A3.1}$$

where

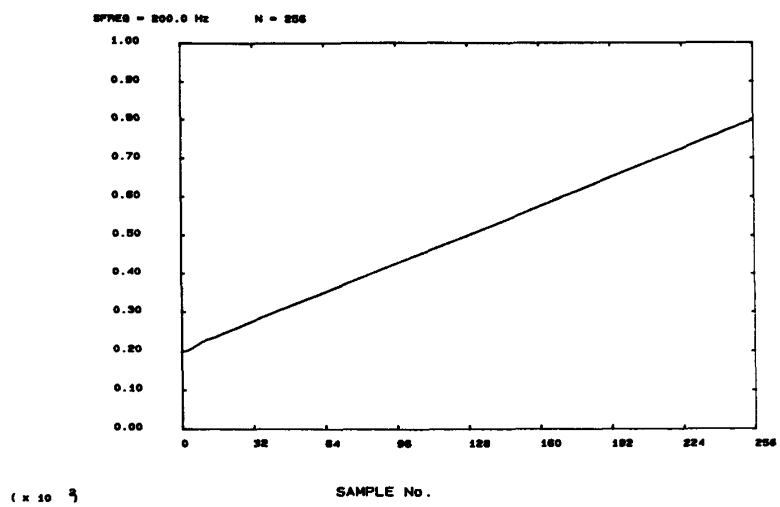
$$E = \frac{1}{K} \sum_{i=1}^{K} (a_2 - \hat{a}_2)^2 \tag{A3.2}$$

and where A is the amplitude of the real signal.

Each value of variance plotted for the ML estimate was the result of averaging over 200 simulations. The ML estimates are seen to be extremely close to the lower variance



(b)



(c)

Fig. 18. *(continued)* (b) and (c): IF law estimates of a linear FM signal using adaptive ML techniques with adaptive searching.

bounds for values of SNR= -1 dB and higher. At lower SNR the ML estimate variance increases dramatically. This deviation from the theoretical bounds below a certain SNR (the threshold) is typical of nonlinear estimators. A similar plot is shown in Fig. 20 for the simulated variance of $\hat{a}_1$. The threshold for this estimate is seen to be -3 dB, which is 2 dB lower than that for $\hat{a}_2$. It should be noted that the variances of both $\hat{a}_1$ and $\hat{a}_2$ exhibit thresholds at higher SNR than is observed for the ML estimate of the frequency of a stationary tone [52]. The threshold effect for the latter occurs at -5 dB for a signal of length 128. More generally, as the polynomial order increases, the variances of the parameters $\hat{a}_k$ also increase.

It should be noted that when true signal frequencies are close to either zero or half the sampling rate, small

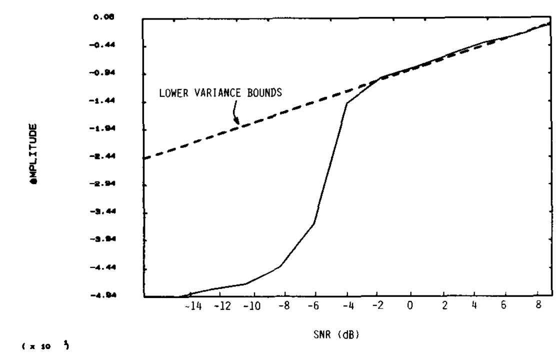


Fig. 19. $\hat{a}_2$ variance compared with lower variance bound.

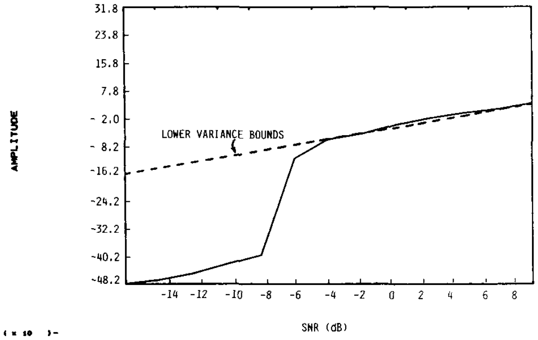


Fig. 20. $\hat{a}_1$ variance compared with lower variance bound.

estimation errors may cause frequency wraparound such that the discrete frequency estimates appear to be in error by almost 0.5. This can give a misleading idea of how great the estimation errors really are. When evaluating frequency errors, then, one should consider the frequency domain to be part of a circle, and frequency differences are measured as arc-lengths on the circle [35]. There is then no difference between the frequency values 0 and 0.5.

ACKNOWLEDGMENT

The author would like to acknowledge the Australian Defence Science and Technology Organization (DSTO) and the Australian Research Council (ARC) who have supported this work. The author also acknowledges P. O'Shea and B. Ristic for their contribution and help in the preparation of the manuscript as well as L. B. White and D. Gray for reviewing the manuscript.

In addition, the author thanks the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] H. Abu Salem, "A comparison of some FM discriminators," Master's thesis, San Diego State University, CA, 1988.

[2] M. Arnold and B. Boashash, "Variance bounds for single-tone parameter estimates," in *Proc. Int. Symp. Signal Processing and Its Applications*, Aug. 1990, pp. 194–196.

[3] R. Baraniuk and D. Jones, "Optimal Kernel design for time-frequency analysis," in *Proc. SPIE Conf. Advanced Algorithms and Architectures for Signal Processing V*, vol. 1348, San Diego, CA, July 1990.

[4] R. F. Barrett and R. Streit, "Frequency tracking using hidden Markov models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 586–598, 1990.

[5] B. Boashash, "Representation temps-frequence," Thes. Doct. Ing., Inst. Nat. Polytechnique, Univ. Grenoble, France, May, 1982.

[6] B. Boashash, B. Lovell, and P. Kootsookos, "Time-frequency signal analysis and instantaneous frequency: their inter-relationship and applications," presented at the *Proc. ISCAS*, OR, Apr., 1989.

[7] B. Boashash, G. Jones, and P. O'Shea, "Instantaneous frequency of signals: Concepts, estimation techniques and applications," Proc. of SPIE, "Advanced algorithms and architectures for signal processing IV," vol. 1152, pp 382–400, San Diego, CA, Aug. 1989.

[8] B. Boashash, O. P. Kenny, and H. J. Whitehouse, "Radar imaging using the Wigner-Ville Distribution," *Proc. SPIE* vol. 1154, Real-Time Signal Processing XII, pp. 282–294, 1989.

[9] B. Boashash, P. O'Shea, and M. Arnold, "Algorithms for instantaneous frequency estimation: A comparative study," in *Proc. SPIE Conf. Advanced Algorithms and Architectures for Signal Processing V*, vol. 1348, San Diego, CA, July 1990.

[10] B. Boashash and P. O'Shea, "Use of the cross Wigner-Ville distribution for estimating the instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, to be published.

[11] B. Boashash and L. B. White, "Instantaneous frequency estimation and automatic time-varying filtering," in *Proc. ICASSP-90*, pp. 1221–1224, Albuquerque, NM, 1990.

[12] B. Boashash, "Instantaneous frequency: Part 1: Fundamentals," *Proc. IEEE*, vol. 80, 1992, this issue.

[13] B. Boashash, "Time-frequency signal analysis," in *Advances in Spectrum Analysis and Array Processing*, vol. 1 of 2, S. Haykin, Ed. New York: Prentice Hall, 1990.

[14] B. Boashash, P. O'Shea, and B. Ristic, "Statistical/computational comparison of some estimators for instantaneous frequency," in *Proc. Int. Conf. Acoustic, Speech, Signal Processing*, Toronto, Canada, 1991, vol. 5, pp. 3193–3196.

[15] B. Boashash and P. O'Shea, "Time-varying higher order spectra," *Proc. SPIE, Conf. Advanced Algorithms and Architectures for Signal Processing VI*, San Diego, CA, July 1991.

[16] B. Boashash, Ed., *Time-Frequency Signal Analysis—Methods and Applications.* New York: Longman Cheshire, 1991.

[17] A. Carlson, *Communications Systems.* New York: McGraw-Hill, 1986.

[18] J. Cioffi and T. Kailath, "Fast recursive-least-squares, transversal filters for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 304–337, 1984.

[19] T. A. C. M. Classen and W. F. G. Mecklenbrauker, "The Wigner Distribution—Part II," *Philips Journal of Research*, vol. 35, pp. 276–300, 1980.

[20] F. Cohen, G. Boudreau-Bartels, and S. Kadambe, "Tracking of unknown non-stationary chirp signals using unsupervised clustering in the Wigner distribution space," in *Proc. ICASP '88*, pp. 2180–2183, 1988.

[21] L. Cohen, "Time-frequency distributions," *Proc. IEEE*, vol. 77, pp. 941–981, 1989.

[22] I. Daubechies, "The wavelet transform: A method for time-frequency localization," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice Hall, 1991.

[23] C. DiMonte and K. Arun, "Tracking the frequencies of superimposed time-varying harmonics," in *Proc. ICASSP '90*, Albuquerque, NM, 1990, pp. 2539–2542.

[24] P. Djuric and S. Kay, "Chirp parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, Dec. 1990.

[25] B. Ferguson, "A ground based acoustic technique for estimating the speed and altitude of an aircraft," *J. Acoust. Soc. Amer.*, to be published.

[26] A. Fineberg and R. Mammone, "An adaptive technique for high resolution time varying spectral estimation," in *Proc. Int. Conf. Acout. Speech, Signal Processing*, pp. 3185–3188, Toronto, 1991.

[27] R. Fletcher, *Practical Methods of Otpimisation.* New York: Wiley, 1981.

[28] D. Friedman, "Detection and frequency estimation of narrowband signals by means of the instantaneous frequency distribution (IFD)," presented at the IEEE ASSP Workshop on Spectral Estimation and Modeling, Minneapolis, MN, 1988.

[29] L. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 202–221, 1975.

[30] F. J. Harris, "Exact FM detection of complex time series," in *Proc. ISSPA '87*, pp. 70–73, Brisbane, Australia, 1987.

[31] F. J. Harris and H. Abu Salem, "Performance comparison of Wigner-Ville based techniques to standard FM-discriminators for estimating instantaneous frequency of a rapidly slewing FM sinusoid in the presence of noise," *SPIE*, Advanced Alg. and Archit. for Signal Processing III, vol. 975, 1988.

[32] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice Hall, 1991, 2nd ed.

[33] D. Huang and E. J. Hannan, "Estimation of Time-Varying Frequency," submitted to IEEE *Trans. Acoust., Speech, Signal Processing.*

[34] J. Imberger and B. Boashash, "Application of the Wigner-Ville distribution to temperature gradient microstructure: a new technique to study small scale variations," *J. Physical Oceanography*, vol. 16, no. 12, pp. 1997–2012, 1986.

[35] L. Jackson and D. Tufts, "Author's Reply," *IEEE Trans. Acoust.,*

Speech, Signal Processing, vol. 27, pp. 200, Apr. 1979.
[36] B. James and B. D. O. Anderson, "The amplitude, phase and frequency estimation of multiharmonic signals in noise—An investigation of the general phase-frequency estimator," in *Proc. ISSPA '90*, Brisbane, Australia, 1990, pp. 141–146.
[37] G. Jones, "Time-frequency analysis of multicomponent signals," in *Proc. EUSIPCO '90*, Sept, 1990, pp. 141–144,
[38] G. Jones, "Instantaneous frequency, time-frequency distributions, and the analysis of multicomponent signals," Ph.D. dissertation, Queensland University of Technology, Brisbane, Australia, 1992.
[39] S. Kay, "Statistically/computationally efficient frequency estimation," in *Proc. ICASSP '88*, New York, 1988, pp. 2292–2295.
[40] S. Kay, *Modern Spectral Estimation.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[41] J. Kitchen, private communication, DSTO, EWD, Adelaide, SA, 1990.
[42] O. P. Kenny, Ph.D. dissertation, Queensland University of Technology, Brisbane, Australia, 1992.
[43] S. L. Marple, *Digital Spectral Analysis.* Englewood Cliffs, NJ: Prentice Hall, 1987.
[44] D. McMahon and R. Barrett, "Generalization of the method for the estimation of the frequencies of tones in noise from the phases of discrete Fourier transforms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 12, pp .371–383, 1987.
[45] A. Oppenheim and R. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1975.
[46] P. O'Shea and B. Boashash, "Some robust instantaneous frequency estimation techniques with application to non-stationary transient detection," in *Proc. EUSIPCO '90*, Sept, 1990, pp. 165–168.
[47] P. O'Shea, "Detection and estimation methods for non-stationary signals," Ph.D. dissertation, University of Queensland, Brisbane, Australia.
[48] S. Peleg and B. Porat, "The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase," *IEEE Trans. Acoust., Speech, Signal Processing*, Apr. 1991, to be published.
[49] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1975, pp. 762.
[50] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals.* Englewood Cliffs, NJ: Prentice Hall, 1978.
[51] Rao and F. Taylor, "Estimation of the Instantaneous Frequency Using the Discrete Wigner Distribution," *Electron. Lett.*, vol. 26, no.4, pp. 246–248, 1990.
[52] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. 20, pp. 591–598, 1974.
[53] B. Quinn, "Convergence of the periodogram estimator," preprint.
[54] R. Sclabassi, M. Sun, D. N. Krieger, and M. S. Scher, "Time-frequency domain problems in the neurosciences," in *Time-Frequency Signal Analysis—Methods and Applications*, B. Boashash, Ed. New York: Longman Cheshire, 1991.
[55] K. Sharman and B. Friedlander, "Time-varying auto-regressive modeling of a class of non-stationary signals," in *Proc. ICASSP '84*, San Francisco, CA, vol. 22.2.1–22.2.4, 1984.
[56] D. L. Snyder, *A State Space Approach to Analog Communications Systems.*Cambridge, MA: MIT Press.
[57] M. T. Taner, F. Koehler, and R. E. Sheriff, "Complex Seismic Trace Analysis," *J. Geophys.*, vol. 44, no. 6, pp. 1041–1063, June 1979.
[58] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 832–835, Nov. 1985.
[59] J. M. Tribolet, "A new phase unwrapping algorithm," IEEE Trans. on Acoust., Speech, Signal Processing, vol. 26, pp. 170–177, 1977.
[60] H. Van Trees, *Detection, Estimation and Modulation Theory, Part I.* New York: John Wiley, 1968.
[61] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part III.* New York: Wiley, 1971.
[62] J. Ville, "Theorie et Application de la Notion de Signal Analytic," *Cables et Transmissions*, vol. 2A, pp. 61–74, 1948.
[63] A. Walker, "On the Estimation of a Harmonic component in a Time Series with Stationary Independent Residuals," *Biometrika*, vol. 58, pp. 21–36, 1971.
[64] L. B. White and B. Boashash, "Estimating the instantaneous frequency of a Gaussian random process," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1518–1521, Sept. 1988.
[65] L. B. White, "Some aspects of time-frequency analysis of non-stationary signals," Ph.D. dissertation, University of Queensland, Brisbane, Australia, 1988.
[66] L. B. White, "Estimation of the instantaneous frequency of a noisy signal," in *Time-Frequency Signal Analysis—Methods and Applications*, B. Boashash, Ed. New York: Longman Cheshire, 1991.
[67] L. B. White, "Cartesian hidden Markov models with applications," *IEEE Trans. Acoust., Speech, Signal Processing*, submitted for publication.
[68] H. Whitehouse, B. Boashash, and J. Speiser, "High resolution techniques for temporal and spatial signals," presented at the Workshop on High Resolution Methods for Underwater Acoustics, organized by GRETSI, Juan les Pins, France, June 16, 1989; also in the Springer-Verlag *Lecture Notes in Computer Science*, M. Bouvet and G. Bienvenu, Eds., 1990.
[69] B. Widrow and M. Hoff, "Adaptive switching circuits," in *IRE 1960, Wescon. Conv. Rec., Part 4*, 1960, pp. 96–104.
[70] B. Widrow and S. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.
[71] K. M. Wong and Q. Jin, "Estimation of the time-varying frequency of a signal: The Cramer–Rao bound and the application of the Wigner Distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 519–536, 1990.
[72] Y. Zhao, L. E. Atlas, and R. J. Marks II, "The use of cone-shape kernels for generalized time-frequency representation of non-stationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, June, 1990.
[73] P. Boles, private communication.

Boualem Boashash (Senior Member, IEEE), for a photograph and biography please see page 538 of this issue.