Estimating Exponential Polynomial Signals

S. Golden and B. Friedlander

Electrical and Computer Engineering Dept.
University of California at Davis
Davis, CA 95616

Abstract

In this paper we approximate arbitrary complex signals by modeling both the logarithm of the amplitude and the phase of the complex signal as finite-order polynomials in time. We present a computationally efficient algorithm that estimates the unknown parameters by successively solving a series of optimization problems that are only a function of a single unknown complex parameter. At high signal-to-noise ratios, the mean-squared error of the estimates are shown to be close to the Cramer-Rao bound for a particular example by using a Monte Carlo simulation.

1 Introduction

It is well-known that an arbitrary complex signal can be represented by its magnitude and phase. In this paper we model complex signals by approximating the signal's phase as a finite-order Taylor expansion in time. Further, we model the logarithm of the time-varying amplitude of the signal as a finite-order Taylor expansion as well. For notational simplicity, we take the order of the two polynomials to be the same. The types of signals that we consider in this paper arise in various applications such as geophysical phenomena [1] and speech processing [2].

In this paper, we consider observing a complex signal, s_n, in complex additive white Gaussian noise, w_n. That is, suppose we observe

$$y_n = s_n + w_n, \tag{1}$$

where n ranges from $1, 2, ..., N$. We will assume that the logarithm of the complex signal is exactly representable by a finite-order Taylor expansion. Specifically, we express the signal as

$$s_n = \exp\left(a_0 + a_1 n + a_2 \frac{n^2}{2!} + ... + a_M \frac{n^M}{M!}\right), \tag{2}$$

where the coefficients of the Taylor expansion are unknown complex parameters. Note that the real parts of the Taylor expansion coefficients specify the envelope of the signal, while the imaginary parts of the Taylor expansion coefficients specify the phase of the signal. We refer to a signal that is in the form of Eq. (2) as an Exponential Polynomial Signal (EPS).

A popular method for estimating the parameters of non-linear signal models, e.g. the parameters in Eq. (2), is by using Maximum-Likelihood (ML) estimation. It is straight-forward to show that ML estimation for the unknown signal parameters corresponds to the following optimization problem

$$\min_{a_0,\dots,a_M} n \sum_{n=1}^{N} \left| y_n - \exp\left(a_0 + a_1 n + ... + a_M \frac{n^M}{M!}\right)\right|^2. \tag{3}$$

There are various iterative methods to solve general non-linear optimizations problems such as the one described by Eq. (3), e.g. the Gauss-Newton method. The difficulty with these algorithms is that a poor choice for the initial estimate may lead to convergence to a local, rather than global minimum. In this paper, we consider a method to obtain an initial estimate that is both computationally efficient and is shown to come close to this global minimum by using a Monte Carlo simulation. Specifically, we compare the Mean Square Error (MSE) of our initial estimate obtained from a Monte Carlo simulation to the Cramer-Rao bound using a particular example to evaluate the performance of the estimate.

[1]This work was supported by the University of California MICRO program and Applied Signal Technology, Inc., by the Office of Naval Research under contract No. N00014-91-J-1602, and by the National Science Foundation under grant No. MIP-90-17221.

2 Estimation Algorithm

Here, we concern ourselves with estimating the highest order coefficient, a_M. If this coefficient is known (or can be well-estimated) then the a_M term in Eq. (2) can be demodulated from the observed signal. Once the a_M term is demodulated from the observation then the a_{M-1} term can be estimated by considering the demodulated signal to be of order $M-1$.

An obvious method to determine a_M from s_n would be to take the Mth derivative of the logarithm of s_n. A more general method would be to note that a_M can be determined by defining the derivative without letting the step size approach zero. We denote the derivative defined in this fashion as a finite-difference. That is, the Mth finite-difference of the logarithm of the signal is

$$\nabla_{\tau_M}\nabla_{\tau_{M-1}}\cdots\nabla_{\tau_1}\ln s_n = a_M. \qquad (4)$$

The finite-difference is defined as

$$\nabla_\tau f_n = \frac{f_n - f_{n-\tau}}{\tau}.$$

The subscripts on the finite-difference in Eq. (4) are different since, in general, each finite-difference introduces a new parameter into the operation.

To illustrate the algorithms that arise from finite-differencing, let us consider an example. We consider the case when $M = 2$. That is, the modeled signal is a linear frequency modulated signal (known as a chirp) with a Gaussian envelope. A natural method to determine a_2 from s_n would be to take the second derivative of the logarithm of the signal. Alternatively, the highest-order coefficient can be obtained by taking the second finite-difference of the signal. Specifically,

$$a_2 = \nabla_{\tau_2}\nabla_{\tau_1}\ln s_n = \frac{1}{\tau_1\tau_2}\ln\frac{s_n s_{n-\tau_1-\tau_2}}{s_{n-\tau_1}s_{n-\tau_2}}, \qquad (5)$$

or equivalently

$$s_n s_{n-\tau_1-\tau_2} - \exp(a_2\tau_1\tau_2)s_{n-\tau_1}s_{n-\tau_2} = 0. \qquad (6)$$

By substituting Eq. (1) into Eq. (6), we determine a transformed model of our data. We denote this transformation as the Exponential Polynomial Transform (EPT). Specifically, the transformed model is

$$y_n y_{n-\tau_1-\tau_2} - \exp(a_2\tau_1\tau_2)y_{n-\tau_1}y_{n-\tau_2} = \varepsilon_n \qquad (7)$$

where ε_n is a zero-mean noise process. To estimate the unknown parameter a_2, we use the method of least squares. That is, we minimize the sum of the squares of the zero-mean noise process. This optimization problem is stated as

$$\min_{a_2} n \sum_n \left|y_n y_{n-\tau_1-\tau_2} - \exp(a_2\tau_1\tau_2)y_{n-\tau_1}y_{n-\tau_2}\right|^2, \qquad (8)$$

where the limits on the summation are chosen to be over all non-zero values.

This optimization problem has a simple analytical solution. Specifically, to obtain the solution, differentiate the objective function with respect to the unknown parameter and set the result equal to zero. The estimate is taken to be the non-trivial solution to this equation. The closed form solution is a function of fourth-order autocorrelation functions and is expressed as

$$\hat{a}_2 = \frac{1}{\tau_1\tau_2}\ln\frac{\sum_n y_n y_{n-\tau_1-\tau_2} y^*_{n-\tau_1} y^*_{n-\tau_2}}{\sum_n \left|y_{n-\tau_1}y_{n-\tau_2}\right|^2}. \qquad (9)$$

The estimate in Eq. (9) will require on the order of N complex multiplies. Specifically, to compute the two fourth-order autocorrelation functions will require $4(N - \tau_1 - \tau_2)$ complex multiplies.

An alternative estimation algorithm would be based upon computing an estimate using weighted least squares. Ideally, the weights would be chosen such that the weighting matrix would be equal to the inverse of the covariance matrix of ε_n.

3 Relationship to Existing Algorithms

The algorithm that we have presented is closely related to the work of Peleg [3]. Peleg's work differed from our own in that Peleg concentrated on constant-amplitude polynomial phase signals. That is, only a_0 was allowed to be complex while the higher-order coefficients were constrained to be purely imaginary. Peleg made use of a different transformation that he denoted as the Discrete Polynomial Transform (DPT) to estimate the highest-order phase coefficient.

In order to make a comparison between the EPT and the DPT, we will derive the DPT algorithm from an optimization problem that is different but related to the optimization problem of Eq. (8). That is, consider an algorithm that determines the unknown parameter a_2 from solving the following optimization problem

$$\min_{a_2} n \sum_{n,\tau_2} |y_n y_{n-\tau_1-\tau_2} - \exp(a_2\tau_1\tau_2) y_{n-\tau_1} y_{n-\tau_2}|^2. \tag{10}$$

The difference between Eq. (8) and Eq. (10) is that in Eq. (10) the summation is performed over all possible values of τ_2 as well as n. Although the optimization problem of Eq. (10) does not have a closed-form solution, a solution to this optimization problem can be obtained by iterative methods.

If we make the assumption that a_2 is purely imaginary, e.g. the signal has constant-amplitude, the optimization problem of Eq. (10) becomes

$$\max_{\omega} x\, Re \left\{ \sum_{n,\tau_2} y_n y^*_{n-\tau_1} y^*_{n-\tau_2} y_{n-\tau_1-\tau_2} \exp(-j\omega\tau_1\tau_2) \right\} \tag{11}$$

where we have let ω represent the imaginary part of a_2. Now, by making a change of variables it can be shown that Eq. (11) is equivalent to

$$\max_{\omega} x \left| \sum_{n} y_n y^*_{n-\tau_1} \exp(-j\omega n\tau_1) \right|^2. \tag{12}$$

The optimization problem of Eq. (12) is, in fact, the DPT algorithm. The solution of Eq. (12) is obtained by using an iterative procedure. The Fourier transform operation is implemented by a Fast Fourier Transform (FFT) for computational efficiency considerations.

The computational requirements of the DPT includes an FFT calculation that requires on the order of $N \log N$ complex multiplies to initialize the iterative algorithm. Further, each iteration of the algorithm will require on the order of N complex multiplies. Note that the DPT is more computationally efficient than an algorithm based upon solving Eq. (10) directly, since the DPT uses an FFT calculation. However, the computational requirements for the DPT are usually greater than the EPT estimate obtained in Eq. (9). Of course, a precise computational comparison would depend upon the length of the data segment being used.

Although the DPT was designed for constant-amplitude signals, Peleg suggests using the DPT for signals with time-varying amplitudes as well. Therefore, in our numerical simulation, we will compare the DPT's estimate for the time-varying phase to the phase estimate obtained from the EPT. Note that the DPT does not provide an estimate for the time-varying amplitude.

The signal model of Eq. (2) is not new. For example, Marques and Almeida proposed this model for harmonic coding of speech. However, in [2] the authors were concerned with the efficient generation of the signal in Eq. (2), rather than estimating the coefficients.

The work of Marques and Almeida was based on two earlier papers [4, 5] that only dealt with the $M = 2$ case. In [5], Kaiser proposed a method to estimate a_2. He replaced the signal with the observation in Eq. (5) and averaged a transformation of this equation over time with both of the delay parameters chosen to be equal to unity. Specifically, he obtained an estimate for a_2 by the following equation

$$\hat{a}_2 = \ln \left(\frac{1}{N-2} \sum_{n=3}^{N} \frac{y_n y_{n-2}}{y^2_{n-1}} \right). \tag{13}$$

Kaiser presented this algorithm for situations where the error was not due to additive measurement noise but only due to computer roundoff error. It therefore does not seem surprising that the EPT produced estimates with a smaller MSE then the estimate obtained from Kaiser's method when using the Gaussian noise model that we have presented. This comparison was made by numerical simulations.

4 General Highest-Order Coefficient

In this section, we extend the concepts of the previous sections to estimating the highest-order coefficient for a signal of arbitrary order M. That is, we determine an initial estimate for a_M where our model is given by Eqs. (1) and (2).

Theorem:

The Mth finite difference of the logarithm of s_n is

$$a_M = \nabla_{\tau_M} \cdots \nabla_{\tau_2} \nabla_{\tau_1} \ln s_n = \frac{1}{\gamma_M} \ln \frac{g^M_n(S)}{h^M_n(S)} \tag{14}$$

where $\gamma_M = \tau_1\tau_2 \cdots \tau_M$ and $g^M_n(S)$ and $h^M_n(S)$ are obtained recursively. Specifically, the recursions are

$$g^m_n(S) = g^{m-1}_n(S) h^{m-1}_{n-\tau_m}(S) \tag{15}$$

$$h^m_n(S) = h^{m-1}_n(S) g^{m-1}_{n-\tau_m}(S) \tag{16}$$

where these recursions are initialized by choosing $g^1_n(S) = s_n$, and $h^1_n(S) = s_{n-\tau_1}$. The recursive equations are applied $M - 1$ times, by setting m to $m = 2, 3, \cdots, M$. The functional argument S

is used to denote the signal for all time. That is, $S = \{s_1, s_2, \cdots, s_N\}$.

Proof:

This theorem is proved by induction in [6]. The derivation is based upon a related theorem in [3].

In order to show how to use the recursive equations, Eqs. (15) and (16), we again consider estimating a_2 when the signal is given by Eq. (2) with $M = 2$. The functions $g_n^2(S)$ and $h_n^2(S)$ are obtained from using the recursive equations with only one iteration when $m = 2$. In this manner, we obtain $g_n^2(S) = s_n s_{n-\tau_1-\tau_2}$ and $h_n^2(S) = s_{n-\tau_1} s_{n-\tau_2}$.

To derive estimation algorithms, we use Eq. (14) to write an expression that is a function of the noise-free signal and the unknown parameter that we would like to estimate. By substituting the measurement model of Eq. (1) into Eq. (14), we obtain a model that describes the general EPT model. That is,

$$g_n^M(Y) - \exp(a_M \gamma_M) h_n^M(Y) = \varepsilon_n \tag{17}$$

where ε_n is a zero-mean noise sequence. The functional argument Y is defined in a manner that is similar to the way we defined S, except Y represents the noisy observation rather than the noise-free signal.

The next step is to minimize the least squares optimization problem based upon the EPT model that is described in Eq. (17). The analytical solution to this optimization problem is stated as

$$\hat{a}_M = \frac{1}{\gamma_M} \ln \frac{\sum_n (h_n^M(Y))^* g_n^M(Y)}{\sum_n |h_n^M(Y)|^2}. \tag{18}$$

The two autocorrelations in Eq. (18) can be computed using

$$2M\left(N - \sum_{i=1}^{M} \tau_i\right)$$

complex multiplies. This particular implementation is obtained by explicitly determining $g_n^M(Y)$ and $h_n^M(Y)$ from the recursive equations of Eqs. (15) and (16).

5 Cramer-Rao Bound

In this section we use the Cramer-Rao lower Bound (CRB) to provide a bound on the estimation accuracy of estimating the unknown parameters from the parametric model described by Eq. (1).

The Cramer-Rao bound for the variances of the unknown parameters can be obtained by taking the diagonal elements of the Fisher information matrix. A well-known form for the Fisher information matrix of a complex Gaussian signal with real unknown parameters can be expressed as

$$I = \frac{2}{\sigma^2} \sum_n Re\left\{ \left(\frac{\partial s_n(\theta)}{\partial \theta}\right)^H \frac{\partial s_n(\theta)}{\partial \theta} \right\}. \tag{19}$$

The argument of the signal, θ, is used to denote a real column vector of the unknown parameters. For our model, this column vector would be represented as $\theta = [Re(a) \quad Im(a)]^T$ where $a = [a_0 \quad a_1 \quad \cdots \quad a_M]^T$.

To represent the signal, using this notation, we will define the time-varying row vector

$$\omega_n = \begin{bmatrix} 1 & n & \frac{n^2}{2!} & \cdots & \frac{n^M}{M!} \end{bmatrix}^T.$$

We can now express the signal as

$$s_n(\theta) = \exp\left(\omega_n^T Re(a) + j\omega_n^T Im(a)\right).$$

The derivative of the signal with respect to the unknown parameters is

$$\frac{\partial s_n(\theta)}{\partial \theta} = [\omega_n^T \quad j\omega_n^T] s_n(\theta)$$

Thus, Eq. (19) becomes

$$I = \frac{2}{\sigma^2} \sum_n \begin{bmatrix} \omega_n \omega_n^T & 0 \\ 0 & \omega_n \omega_n^T \end{bmatrix} \exp\left(2\omega_n^T Re(a)\right). \tag{20}$$

From Eq. (20), we observe that the Fisher information matrix is block diagonal. Further, both of the diagonal blocks are identical. Implying that the variances for the real parts of the unknown parameters will be exactly equal to the variances of the imaginary parts of the unknown parameters. Note, that the Fisher information matrix is only a function of the real parts of the unknown parameters and not the imaginary parts. That is, the estimation accuracy for estimating both the time-varying amplitude and the time-varying phase is only a function of the time-varying amplitude and the order of the polynomial that approximates the phase.

6 Comparison by Simulation

Here, we compare the MSE of the estimates obtained from the EPT and DPT to the Cramer-Rao bound. To compare the performance of these algorithms we consider a numerical example where the signal is given by Eq. (2) with $M = 2$. The value of the parameters are $a_0 = -6 + 2j$, $a_1 = (4.8 + 8j) * 10^{-2}$, and $a_2 = (-2 + 4j) * 10^{-4}$. The number of samples was chosen to be $N = 500$. The noise free signal is shown in Fig. 1. A Monte Carlo simulation using 300 runs was performed for both methods.

For the DPT, we used $\tau_1 = 90$ samples which is considerably smaller than Peleg's suggestion of choosing $\tau_1 = N/2 = 250$ when the signal has a constant amplitude. We chose to use a smaller delay parameter since the energy of the signal is not uniformly distributed throughout the data segment, as in the constant-amplitude case, but is concentrated in a compact region of the data segment. The delay parameters for the EPT should be chosen in a similar manner. We chose the delay parameters to be $\tau_1 = 77$ and $\tau_2 = 93$. These values were obtained by using a first order perturbation analysis which is shown in [6]. Specifically, the values were obtained by minimizing the variance of the estimate given in Eq. (9) over a range of different delay parameters.

The MSE of the estimates are shown in Fig. 2, along with the Cramer-Rao lower bound for the variance of the estimates. The results show that the MSE obtained from both the EPT algorithm and the DPT algorithm came close to achieving the Cramer-Rao lower bound at high signal-to-noise ratios.

We have shown that the estimates obtained from the EPT are computationally efficient. The estimates require on the order of only N complex multiplies where N is the number of data points. We have also shown for a particular example that the MSE of the estimate comes close to the Cramer-Rao bound at high signal-to-noise ratios. Further, unlike the DPT, the EPT also provides an estimate for the parameters that model the time-varying amplitude of the signal.

References

[1] B. Boashash, "Estimating and Interpreting The Instantaneous Frequency of a Signal," *Proceedings of the I.E.E.E.*, vol. 80, April 1992.

[2] J. S. Marques, L. B. Almeida, "A Fast Algorithm for Generating Sinusoids with Polynomial Phase," *ICASSP*, Toronto, Ont., Canada, 1991.

[3] S. Peleg, *Estimation and Detection with the Discrete Polynomial Transform*. Ph.D. Thesis, U.C. Davis, 1993.

[4] D. Jones, T. Parks, "On Computing Equally Spaced Samples of a Complex Gaussian Function," *IEEE Transactions on ASSP*, vol. 35, No. 10, pp. 1479-81, 1987.

[5] J. Kaiser, "On the Fast Generation of Equally Spaced Values of the Gaussian Function Aexp(-at*t)," *IEEE Transactions on ASSP*, vol. 35, No. 10, pp. 1479-81, 1987.

[6] S. Golden, *Exponential Polynomial Signals*. Ph.D. Thesis, U.C. Davis, (In Preparation).

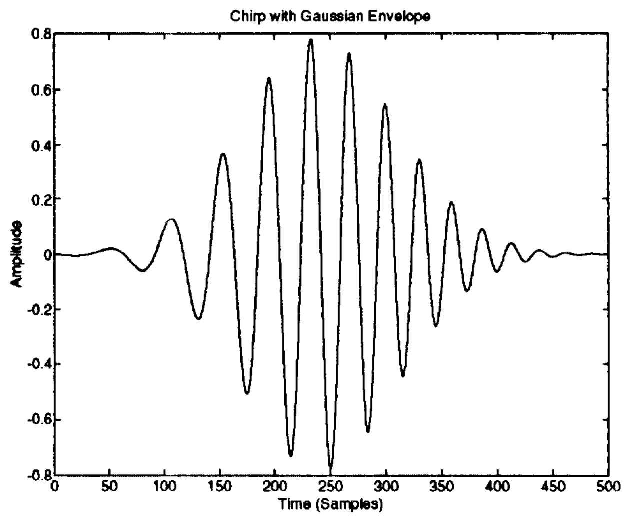

Fig. 1. The real part of a linear frequency modulated signal (chirp) with a Gaussian envelope. This particular signal was used as the error-free signal in numerical simulations.

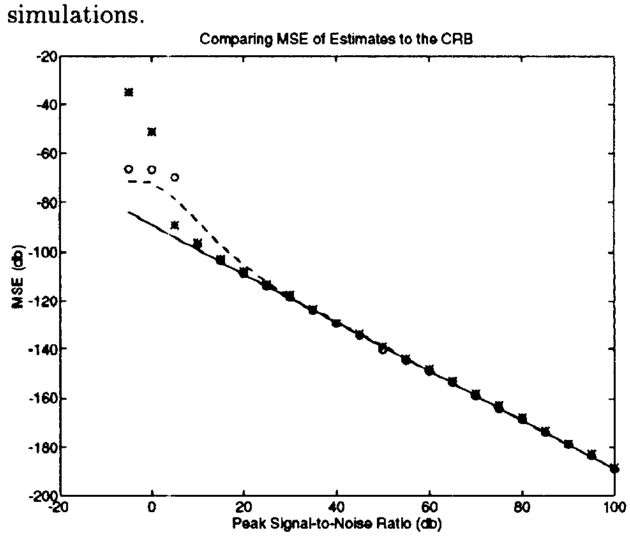

Fig. 2. The MSE for several different estimates are compared to the Cramer-Rao lower bound (solid line). The MSE for the $Re(a_2)$ is shown by: (dashes) EPT. The MSE for the $Im(a_2)$ is shown by: (circles) EPT; and (asterisks) Peleg.