Sparse Representations with Chirplets via Maximum Likelihood Estimation

Jeffrey C. O'Neill, Patrick Flandrin, and William C. Karl

Abstract— **We formulate the problem of approximating a signal with a sum of chirped Gaussians, the so-called *chirplets*, under the framework of maximum likelihood estimation. For a signal model of one chirplet in noise, we formulate the maximum likelihood estimator (MLE) and compute the Cramér-Rao lower bound. An approximate MLE is developed, based on time-frequency methods, and is applied sequentially to obtain a decomposition of multiple chirplets. The decomposition is refined after each iteration with the expectation-maximization algorithm. A version of the algorithm, which is $O(N)$ for each chirplet of the decomposition, is applied to a data set of whale whistles.**

I. Introduction

Chirplets are a class of signals that consists of Gaussians that are translated in time and frequency, scaled, and chirped. They are defined as

$$\boldsymbol{s}_{t,\omega,c,d} = s(n;t,\omega,c,d) = (\sqrt{2\pi}d)^{-\frac{1}{2}} \exp\left\{-\left(\tfrac{n-t}{2d}\right)^2 + j\tfrac{c}{2}(n-t)^2 + j\omega(n-t)\right\}.$$

where t, ω, and c are in $\mathbb{R}$ and d is in $\mathbb{R}^+$. The parameters t, ω, c, and d represent, respectively, the location in time, the location in frequency, the chirp rate, and the duration, and $s(\cdot)$ is defined such that

$$\|\boldsymbol{s}_{t,\omega,c,d}\|^2 = \sum_n |s(n;t,\omega,c,d)|^2 = 1.$$

In this paper we present a method for approximating a signal as a weighted sum of chirplets

$$x(n) \approx \sum_{i=1}^{q} a_i\, e^{j\phi_i}\, s(n;t_i,\omega_i,c_i,d_i).$$

using as few parameters as possible, i.e. a *sparse* approximation. The justification of sparse representations has been elegantly discussed by several authors [1], [2], [3], [4], to which we refer the reader for more detail. The result of this approximation can be applied to de-noising, data compression, and feature extraction [1], [3], [2], [4], [5].

Our method is to assume that the signal is actually a sum of q chirplets in complex, white, Gaussian noise[1]

$$X(n) = \sum_{i=1}^{q} a_i\, e^{j\phi_i}\, s(n;t_i,\omega_i,c_i,d_i) + W(n) \qquad (1)$$

where $\mathrm{Re}\{W(n)\}$ and $\mathrm{Im}\{W(n)\}$ are uncorrelated and $\sim \mathcal{N}(0,\sigma^2)$, and estimate the unknown parameters via maximum likelihood estimation. The unknown parameters are the chirp parameters a_i, ϕ_i, t_i, ω_i, c_i, and d_i for $i = 1,\ldots,q$, and the variance of the CWGN σ^2. We formulate the problem for a discrete signal to facilitate the implementation, and also for a complex signal as it simplifies the theory. We assume that any real signal of interest can be converted to a complex signal via a Hilbert transform [6] or other filtering. Several authors have proposed chirplet like decompositions [3], [7], [8], [9], [10], or more general methods for sparse representations [1], [2]. Our method is different, in that we formulate the problem using maximum likelihood estimation. Bayesian techniques for a similar decomposition have recently been proposed [11].

Chirplets have several appealing properties that motivate their selection for approximating signals.

- If the signal is translated in time or frequency, scaled, or chirped, then the representation will be covariant to these changes. This can be important in classification tasks [12].
- The chirplets are versatile in that we can approximate a variety of time-frequency structures [8].
- Subsets of the chirplets give frames of wavelets and Gabor functions [13].
- Chirplets are the only signals to satisfy the generalized uncertainty principle [14] with equality

$$\sigma_t^2\,\sigma_\omega^2 - \mathrm{cov}_{t\omega}^2 \geq \frac{1}{4}$$

where σ_t is the duration, σ_ω is the bandwidth, and $\mathrm{cov}_{t\omega}$ is the time-frequency covariance.[2]
- Chirplets are the only signals for which the Wigner distribution is non-negative [15].
- Calculations involving chirplets can often be expressed in closed form.

The paper proceeds as follows. In Section II we develop the maximum likelihood estimator (MLE) for a model of one chirplet in noise ((1) with $q = 1$), and in Section III we derive the Cramér-Rao lower bound (CRLB) for the same

J.C. O'Neill is with Boston University. Address: Electrical and Computer Engineering, 8 Saint Mary's Street, Boston, MA 02215-2421. Email: jeffo@bu.edu. Phone: (617) 353-0161. Fax: (617) 353-6440.

P. Flandrin is with the Ecole Normale Supérieure de Lyon.

W.C. Karl is with Boston University.

EDICS 2-SREP Signal and Noise Representation and Modeling

[1] We use upper-case letters to denote random variables, lower-case letters to denote deterministic quantities or realizations of random variables, bold face to indicate vectors, and plain text to indicate scalars. Exceptions to this rule are the constants N, M, and P.

[2] The necessity of the equality condition is not stated in [14], but it follows directly from what is presented there.

model. Since the MLE developed in Section II is computationally prohibitive, in Section IV we develop an approximation to the MLE that is more efficient computationally, and in Section V we perform a simulation to compare the MLE, the approximate MLE, and the CRLB. In Section VI, we combine the approximate MLE of Section IV with the expectation-maximization algorithm to approximate a signal as a weighted sum of several chirplets. In Section VII we show examples of the method on synthetic data and whale whistles. In Section VIII we compare our method with related work.

II. Maximum Likelihood Estimation for One Chirplet

Consider one chirplet in noise

$$X(n) = a_o\, e^{j\phi_o}\, s(n; t_o, \omega_o, c_o, d_o) + W(n) \tag{2a}$$

$$\boldsymbol{X} = a_o\, e^{j\phi_o}\, \boldsymbol{s}_{t_o,\omega_o,c_o,d_o} + \boldsymbol{W} \tag{2b}$$

where $W(n)$ is defined above and $n \in \{1, \ldots, N\}$. We denote a realization of (2) as

$$x(n) = a_o\, e^{j\phi_o}\, s(n; t_o, \omega_o, c_o, d_o) + w(n)$$
$$\boldsymbol{x} = a_o\, e^{j\phi_o}\, \boldsymbol{s}_{t_o,\omega_o,c_o,d_o} + \boldsymbol{w}.$$

There are seven deterministic, unknown parameters

$$\boldsymbol{\theta}_o = \begin{bmatrix}\sigma_o^2 & a_o & d_o & c_o & t_o & \omega_o & \phi_o\end{bmatrix}$$

to be estimated. Denote $p_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ as the probability density function for $\boldsymbol{X}$, then given a realization of the data, the log likelihood function [16] is

$$\begin{aligned} l(\boldsymbol{x};\boldsymbol{\theta}) &= \log p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) \\ &= -N\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\left\|\boldsymbol{x} - a\, e^{j\phi}\, \boldsymbol{s}_{t,\omega,c,d}\right\|^2. \end{aligned} \tag{3}$$

The maximum likelihood estimator is the value of $\boldsymbol{\theta}$ that maximizes $l(\boldsymbol{x};\boldsymbol{\theta})$. Simplifying (3) results in

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{ml} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{x};\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \Big\{ &-N\log(2\pi\sigma^2) \\ &- \tfrac{1}{2\sigma^2}\left(a^2 + \|\boldsymbol{x}\|^2 - 2a\,\mathrm{Re}\left\langle \boldsymbol{x}, e^{j\phi}\, \boldsymbol{s}_{t,\omega,c,d}\right\rangle\right)\Big\}. \end{aligned} \tag{4}$$

This maximization is equivalent to the following sequence of operations

$$\begin{bmatrix}\hat{t}_{ml}\\ \hat{\omega}_{ml}\\ \hat{c}_{ml}\\ \hat{d}_{ml}\end{bmatrix} = \arg\max_{t,\omega,c,d} \left|\langle \boldsymbol{x}, \boldsymbol{s}_{t,\omega,c,d}\rangle\right|^2 \tag{5a}$$

$$z = \left\langle \boldsymbol{x}, \boldsymbol{s}_{\hat{t}_{ml},\hat{\omega}_{ml},\hat{c}_{ml},\hat{d}_{ml}}\right\rangle \tag{5b}$$

$$\hat{\phi}_{ml} = -\angle z \tag{5c}$$

$$\hat{a}_{ml} = |z| \tag{5d}$$

$$\hat{\sigma^2}_{ml} = \frac{\|\boldsymbol{x}\|^2 - |z|^2}{2N}. \tag{5e}$$

Thus, if we can solve (5a), then we can then straightforwardly solve the MLE in (4).

Consider the function

$$\begin{aligned} f(t,\omega,c,d) &= \left|\langle \boldsymbol{x}, \boldsymbol{s}_{t,\omega,c,d}\rangle\right|^2 \\ &= \left|\langle \boldsymbol{s}_{t_o,\omega_o,c_o,d_o}, \boldsymbol{s}_{t,\omega,c,d}\rangle\right|^2 + \\ &\quad 2\mathrm{Re}\{\langle \boldsymbol{s}_{t_o,\omega_o,c_o,d_o}, \boldsymbol{s}_{t,\omega,c,d}\rangle \langle \boldsymbol{w}, \boldsymbol{s}_{t,\omega,c,d}\rangle\} + \\ &\quad \left|\langle \boldsymbol{w}, \boldsymbol{s}_{t,\omega,c,d}\rangle\right|^2. \end{aligned}$$

In Appendix A we show that $g(t,\omega,c,d) = \left|\langle \boldsymbol{s}_{t_o,\omega_o,c_o,d_o}, \boldsymbol{s}_{t,\omega,c,d}\rangle\right|^2$ is unimodal. Thus for $\sigma = 0$, the likelihood function is also unimodal and any initialization can be used to find the maximum. For $\sigma > 0$ the noise induces local maxima in the likelihood function and a better initialization is needed.

While the MLE is often consistent as $N \to \infty$, it is not for the signal model in (2). From the likelihood function in (3) it is seen that the MLE is a least squares estimator. A necessary condition for a least squares estimator to be consistent is that for $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, we must have

$$D_N(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \left\| a_1\, e^{j\phi_1}\, \boldsymbol{s}_{t_1,\omega_1,c_1,d_1} - a_2\, e^{j\phi_2}\, \boldsymbol{s}_{t_2,\omega_2,c_2,d_2}\right\|^2 \to \infty$$

as $N \to \infty$ [17]. This condition is clearly not satisfied for the model in (2).

III. The Cramér-Rao Lower Bounds for One Chirplet

In this section we calculate the Cramér-Rao lower bounds (CRLB's) corresponding to the signal model in (2). The bounds are more enlightening if one has an intuitive notion of what to expect from the bounds. To this end, we first present several examples that provide this intuition.

In Figure 1 we present the Wigner distributions [6] of six chirps. The first three examples have a chirp rate of zero, and the only difference between them is the value of the duration parameter. The last three examples have a chirp rate of 1/64, and again the only difference between them is the value of the duration parameter.

- When the duration is small, the signal is concentrated in time and thus it should be easy to estimate the location in time. As the duration increases, it should become harder to estimate the location in time.
- The estimation of the duration should be analogous to the estimation of the location in time.
- When the duration is small, the chirp rate should be very difficult to estimate (compare Figures 1(a) and (d)). As the the duration increases, it should become easier to estimate the chirp rate.
- When the chirp rate is zero, estimating the location in frequency should be dual to estimating the location in time. When the chirp rate is not zero, the signal is not concentrated in frequency for large values of the duration (see Figure 1(f)). In fact, the signal will be the most concentrated in frequency for some intermediate duration (see Figure 1(e)). Thus, for non-zero chirp rates, the estimation of the location in frequency is difficult for both small and large durations and easiest somewhere in between.

The Fisher information matrix is defined as [16]

$$I(\boldsymbol{\theta})_{i,j} = -E\left[\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\log p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta})\right],$$

and the CRLB is the inverse of the Fisher information

$$C(\boldsymbol{\theta}) = I(\boldsymbol{\theta})^{-1}.$$

This can be computed in closed form for the signal model in (2)

$$C(\boldsymbol{\theta}) = \frac{\sigma^2}{a^2}\begin{bmatrix} \frac{a^2\sigma^2}{N} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2d^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{d^4} & 0 & 0 & \frac{-1}{d^2} \\ 0 & 0 & 0 & 0 & 4d^2 & 4cd^2 & 4\omega d^2 \\ 0 & 0 & 0 & 0 & 4cd^2 & \frac{1+4c^2d^4}{d^2} & 4c\omega d^2 \\ 0 & 0 & 0 & \frac{-1}{d^2} & 4\omega d^2 & 4c\omega d^2 & \frac{3+8\omega^2d^2}{2} \end{bmatrix}.$$

The computations are based on the following approximations

$$\sum_{n=1}^{N}\frac{(n-t)^p}{\sqrt{2\pi}d}\exp\left\{-2\left(\frac{n-t}{2d}\right)^2\right\} \approx \begin{cases} 1 & p=0, \\ 0 & p=1,3, \\ d^2 & p=2, \\ 3d^4 & p=4 \end{cases}$$

that are valid as long as the signal is approximately time-limited with respect to the sampling interval and approximately band-limited with respect to the sampling rate. The diagonal elements of the CRLB matrix are bounds on the variance of unbiased estimators of the unknown parameters. Several of these are plotted in Fig. 2, and they are in agreement with the intuitive arguments presented above.

The CRLB's provide valuable information for using the estimated chirp parameters in a classification task. For instance, if the duration of a chirplet is very small, then we should not use the chirp rate for classification since our confidence in this parameter would be very low.

IV. Approximate MLE for One Chirplet

Solving (5a) exactly is computationally expensive since it would require an exhaustive grid search over the parameter space. Here, we propose a method for solving (5a) based on the idea of *zooming* to a time-frequency region. This zooming will allow high quality estimates with modest computations. This zooming will also provide robustness in the case where the signal is actually multiple chirplets in noise[3] Other authors [7], [8], [10], [19] have proposed methods for solving (5a). Of these, only [8] provides good estimates when there are multiple chirplets present and we will further discuss this method in Section VIII.

[3]In this section we are interested in finding an approximate MLE for a signal model of one chirplet in noise. However, since the ultimate goal is a decomposition of multiple chirplets we allow for this possibility.

In recent years there has been an explosion of methods for estimating parameters of chirps and more generally polynomial phase signals (e.g. [20], [21], [22], [23], [24], [25], [19], [26], [27], [28], [29], [30], [31], [32]). We chose and adapted methods from the literature based upon the following criteria:

- existence of a relationship with the MLE (i.e. not completely *ad-hoc*),
- extensibility to multi-component signals (to be discussed in a subsequent section),
- applicability to the specific problem of interest here (particularly the Gaussian amplitude modulation),
- robustness in low SNR (e.g. avoiding derivatives), and
- computational simplicity.

Our method for estimating the chirplet parameters is as follows.

1. Estimate, from a global measure, the chirp rate and duration (to increase the accuracy of the zoom).
2. Using the estimates of the chirp rate and duration, estimate the location in time and frequency (the actual zooming).
3. Re-estimate, from a local measure, the chirp rate and duration.
4. Use a quasi-Newton procedure to find the closest local maximum of the likelihood function.

Based on the SNR, one could choose to eliminate step 1 or repeat steps 2 and 3 several times. The computations of these steps depend on the number of data points N and a *resolution parameter* M that will generally be much smaller than N. The computations for each of these steps are listed in Table I, and we now investigate each of these steps in detail.

A. Global Estimation of Chirp Rate and Duration

We would like to estimate the chirp rate and duration parameters without any knowledge of the other parameters. Wang *et al.* [30] proposed the following method for estimating chirp rate[4]

$$\hat{c} = \arg\max_{c}\int |A_x(cr,r)|^2\,dr \tag{6}$$

where $A_x(\theta,\tau)$ is the ambiguity function. While there does not appear to be a straightforward connection with standard estimators (e.g. MLE), (6) performs well in low SNR and for multi-component signals.

In Appendix B we show that Wang's estimator is equivalent to

$$\hat{c} = \arg\max_{c}\int |X_c(\omega)|^4\,d\omega \tag{7}$$

where $X_c(\omega)$ is the Fourier transform of the chirped signal

$$X_c(\omega) = \int x(t)\,\exp\{-jct^2/2\}\,\exp\{-j\omega t\}\,dt.$$

[4]We present the time-frequency representations with continuous variables with the understanding that the implementation for discrete signals is straightforward [33].

Chirping is a unitary (energy preserving) operator, thus $\|X_c\|^2$ is the same for all c. However, the fourth power in (7) emphasizes peaky spectra and can thus be considered a measure of the concentration of X_c.

Computing the ambiguity function requires $O(N^2 \log N)$ computations, however solving (7) over M values of c requires only $O(MN \log N)$ computations. Since $M \ll N$ this significantly reduces the necessary computations.

Once the chirp rate has been estimated, we could estimate the duration with

$$y(t) = x(t) \exp\{-j\hat{c}t^2/2\}$$
$$\hat{d} = \arg\max_d \left\langle |r_y(\cdot)|^2, \boldsymbol{s}_{0,0,0,d} \right\rangle$$

where $r_y(\tau)$ is the auto-correlation function of $y(t)$. This *ad-hoc* processing removes information regarding the chirp rate and the location in time and frequency and gives an exact estimate in the absence of noise.

The statistics of $|r_y(\tau)|^2$ are complicated and we do not have closed form results. Empirical observations show that the CWGN has the most influence at $\tau = 0$ and that the weighted inner product

$$\hat{d} = \arg\max_d \left\langle |r_y(\cdot)|^2, \boldsymbol{s}_{0,0,0,d} \right\rangle_w \tag{8}$$

is more reliable, where $w(\tau)$ is of the form

$$w(\tau) = \begin{cases} 0 & |\tau| < \tau_0, \\ 1 & \text{otherwise.} \end{cases}$$

and τ_0 is small. If we solve (8) over M values of d, then the computations are $O(MN \log N)$.

B. Estimation of Location in Time and Frequency

Once we have an estimate of the chirp rate and duration, we would like to use this information to estimate the location in time and frequency of the chirplet. Denote the spectrogram [6] of signal $x(t)$ with window $h(t)$ as $S_x(t, \omega; h)$. If $\hat{c} = c_o$ and $\hat{d} = d_o$, then

$$\begin{bmatrix} \hat{t} \\ \hat{\omega} \end{bmatrix} = \arg\max_{t,\omega} S_x(t, \omega; \boldsymbol{s}_{0,0,\hat{c},\hat{d}}) \tag{9}$$

is the MLE of the location in time and frequency [34]. If $\hat{c} \neq c_o$ or $\hat{d} \neq d_o$, then (9) performs worse than the MLE. However, we show in Lemma 1 of Appendix A that in the absence of noise, (9) has a unique maximum at t_o and ω_o for *any* values of $\hat{c}$ and $\hat{d}$. If we compute M frequency samples of the spectrogram, then this estimator requires $O(NM \log M)$ computations.

C. Local Estimation of Chirp Rate and Duration

Once we have an estimate of the location in time and frequency of the chirplet, we would like to re-estimate the chirp rate and duration. Assuming that we have good estimates of the location in time and frequency, we expect the local estimates of the chirp rate and duration to be better than the global estimates. Several authors have proposed using Radon transformations of time-frequency distributions as a means for estimating chirp rate [20], [27], [35], [36], [31].

One method[5] for estimating chirp rate is based on the Wigner distribution [27], denoted $W_x(t, \omega)$. Given estimates of the location in time and frequency, $\hat{t}$ and $\hat{\omega}$, then the chirp rate estimator is

$$\hat{c} = \arg\max_c \int W_x(\hat{t} + r, \hat{\omega} + cr)\, dr. \tag{10}$$

Once the chirp rate has been estimated, we use

$$\hat{d} = \arg\max_d \left| \langle \boldsymbol{x}, \boldsymbol{s}_{\hat{t},\hat{\omega},\hat{c},d} \rangle \right|^2 \tag{11}$$

to estimate the duration. Both (10) and (11) have relationships with MLE's that we now describe.

If $\hat{t} = t_o$ and $\hat{\omega} = \omega_o$, then (10) approaches the MLE as $d_o \to \infty$ [27]. For finite d_o, (10) performs worse than the MLE, but in the absence of noise, has a unique maximum at c_o

$$\int W_x(\hat{t} + r, \hat{\omega} + cr)\, dr = \frac{\pi}{1/d^2 + 2d^2(c - c_o)^2}.$$

Considering the examples in Fig. 1d and e, this is an unexpected, but desirable result. Even though the Wigner distributions of the chirps are not concentrated along the actual chirp rate, we can correctly estimate the chirp rate for any duration in the absence of noise.

Errors in $\hat{t}$ and $\hat{\omega}$ are generally such that $\hat{t}$ and $\hat{\omega}$ lie along the instantaneous frequency of the chirp. If this is so, then (10) still performs well. If the SNR is low enough so that $\hat{t}$ and $\hat{\omega}$ do not lie close to the instantaneous frequency of the chirplet, then (10) does not perform well.

Computing the Wigner distribution requires $O(N^2 \log N)$ computations, but by windowing the signal to M samples around $\hat{t}$, the computations can be reduced to $O(M^2 \log M)$. In practice, the estimator can be calculated without computing the Wigner distribution [40].

If $\hat{t} = t_o$, $\hat{\omega} = \omega_o$, and $\hat{c} = c_o$, then (11) is the MLE for the duration. By maximizing over M values of d and windowing the signal to M samples, the computations can be reduced to $O(M^2)$.

D. Quasi-Newton Maximization

The previous steps provide estimates of the location in time and frequency, the chirp rate, and the duration. However, these estimates will generally not correspond to the global or a local maximum of the likelihood function, and better estimates can obtained by applying a quasi-Newton maximization procedure. It is hoped that the previous steps provide good enough estimates so that the quasi-Newton maximization provides the MLE. If we perform the quasi-Newton maximization on M samples centered around $\hat{t}$, then this step requires $O(M)$ computations.

[5] Another method for estimating chirp rate is based on the recently introduced *local ambiguity function* [37]. This method can be more reliable for multi-component signals and non-linear chirps [38], [39], although it is not discussed here due to space limitations.

V. Simulation Results

To evaluate the performance of the approximate MLE developed in the previous section, we ran a simulation for the discrete signal defined with parameters $N = 127$, $t_o = 64$, $\omega_o = 0$, $c_o = 2\pi/127$, $d_o = 8$, $a_o = 1$, and $\phi_o = 0$. The discrete signal is in CWGN with standard deviations of 0.2, 0.4, 0.6, 0.8, 0.10, and 0.12, which correspond to SNR's ranging from 9.9 dB to -5.6 dB.[6] This was done for 5000 trials at each noise level.

The true MLE is too computationally expensive to simulate. We settled for finding the local maximum of the likelihood function closest to the true value of the parameters, which we call the *simulated MLE*. For high SNR this will be a good approximation to the MLE. For low SNR this will perform *better* than the MLE, since the local maximum of the likelihood function closest to the true value will not necessarily be the global maximum. For the lowest SNR, the approximate MLE from the previous section had a higher likelihood than the simulated MLE for 21 of the 5000 trials, thus the simulated MLE was *too good* at least this many times.

In Fig. 3 we show the CRLB's, the bias and variance of the simulated MLE, and the bias and variance of the approximate MLE developed in the previous section. Only the results for the time, frequency, chirp rate, and duration parameters are shown, since the other three depend directly on these four. One can see that the approximate MLE performs nearly as well as the simulated MLE for SNR's down to about -4 dB. Note that for this problem, the MLE is *not* asymptotically efficient and thus does not attain the CRLB. There are several details that need to be clarified about these results.

The SNR's are not a good measure of the difficulty of the problem. For instance if we changed the simulation by using $N = 255$, then the SNR would decrease by about 3 dB, but the decrease in the performance of the estimator would be minimal (since the extra noise does not overlap the signal in time). The standard deviation of the CWGN is a better indicator of the difficulty of the problem, since it is independent of N.

In this simulation, the estimators of time and frequency have small bias. However this is due to the fact that the true value of these parameters lie in the center of their parameter spaces ($[1, 127]$ and $[-\pi, \pi]$, respectively). In another simulation where we set $t_o = 32$, the estimator was slightly biased towards 64 since there was more noise in this direction. Thus the estimators appear to be biased towards the center of the parameter space (though for the duration parameter, it is not clear where the center is).

VI. Multiple Chirplets

Now that we have a solution to the simpler problem of one chirplet in CWGN, we would like to solve the more complicated problem of q chirplets in CWGN as expressed in (1). Direct maximization over $6q + 1$ parameters (6 parameters for each chirplet and σ^2) is not feasible, so we propose a sequential algorithm that uses the expectation-maximization (EM) algorithm [41], [42], [43] as a refining step.

Suppose that we have $q-1$ chirplets fit to the signal $x(n)$ and that we desire to fit a qth chirplet. The procedure is
1. Compute the residual

$$e(n) = x(n) - \sum_{i=1}^{q-1} \hat{a}_i \, e^{j\hat{\phi}_i} \, s(n; \hat{t}_i, \hat{\omega}_i, \hat{c}_i, \hat{d}_i)$$

2. Apply the approximate MLE of Section IV to fit a chirplet to the residual.
3. Use the q component model and the EM algorithm to refine the estimates of the q chirplets.

Matching pursuit [3], [8] uses steps similar to 1 and 2 to obtain a decomposition of multiple chirplets. However, the addition of step 3 and the convergence properties of the EM algorithm ensures that we shall always have at least a local maximum of the likelihood function for the q chirplet model.

For this problem, there is a natural choice for the complete data [42] in the EM algorithm. The EM algorithm consists of two steps: an expectation step (E step) and a maximization step (M step). The E step calculates the complete data $x_i(n)$

$$e(n) = x(n) - \sum_{i=1}^{q} \hat{a}_i \, e^{j\hat{\phi}_i} \, s(n; \hat{t}_i, \hat{\omega}_i, \hat{c}_i, \hat{d}_i)$$

$$x_i(n) = \hat{a}_i \, e^{j\hat{\phi}_i} \, s(n; \hat{t}_i, \hat{\omega}_i, \hat{c}_i, \hat{d}_i) + \beta_i \, e(n) \quad i = 1, \ldots, q.$$

where $\beta_i \geq 0$ and $\sum_{i=1}^{q} \beta_i = 1$. The M step applies the approximate MLE of Section IV to each of the $x_i(n)$ to refine the estimate of this chirplet. It remains to specify the values of the β_i. A natural choice is $\beta_i = \frac{1}{q}$. However, since the M step is much slower than the E step, we propose the following for β_i

$$\beta_i = \begin{cases} 1 & \text{if } i \bmod k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

where k is the iteration of the EM algorithm. Thus, we update only one chirplet at each iteration (the approximate MLE is applied only once for each M step) and results in a q fold improvement in speed for each iteration. More iterations of the EM algorithm may be required, but our experiences show that the overall algorithm converges more quickly.

We shall not attempt to explicitly estimate q, but rather continue the algorithm until a specified stopping criteria has been reached. We shall not investigate specific stopping criteria here, but one could propose several: a test for the whiteness of the residual, a test comparing the estimated amplitude of the qth chirplet with the estimated noise variance, a test comparing the change in the norm of the residual with the estimated noise variance, or a test involving the Akaike information criterion [44].

[6] Calculated according to $10 \log_{10} \frac{a^2}{2N\sigma^2}$.

VII. Examples

To illustrate the algorithm of the previous section, we use an example consisting of a synthetic signal and two recordings of whale whistles.

The synthetic signal consists of four chirplets and the sum of the Wigner distributions of the four chirplets is shown in Fig. 4. The chirplets are close enough to each other so that it is very difficult for time-frequency distributions (TFD's) to resolve them. To illustrate this, we chose the time-varying AOK TFD [45], a method that often provides better resolution than other TFD's. In Fig. 4 we show the time-varying AOK TFD of the sum of the four chirplets and the time-frequency structure of the chirplets is not clear. Next we applied the algorithm of the previous section without the EM refinement (step 3) and the sum of the Wigner distributions of the four estimated chirplets is shown in Fig. 4. Without the EM refinement the estimated parameters are quite different from the true parameters. Finally we applied the algorithm of the previous section with the EM refinement and the parameters of the four chirplets were correctly estimated. This example illustrates the robustness of our approximate MLE when multiple chirplets are present.

The next two examples are two whale whistles. In the approximate MLE of Section IV, the most computationally expensive step is the global estimation of chirp rate, which is $O(MN \log N)$ (see Table I). However, we have some prior knowledge about the whale whistles. We know that the whale whistles will generally have long durations and small chirp rates. With this knowledge, we decide to forgo the global estimate of chirp rate and duration in Section IV and initialize with $\hat{c} = 0$ and $\hat{d} = 100$. The most expensive step is now the estimation of location in time and frequency and the computations are reduced to $O(NM \log M)$. Since $N \gg M$, this is a significant savings in computations.

In Fig. 5a we show the envelope of time series, the spectrum in dBs, and a spectrogram (50 dB dynamic range) of a whale whistle. With $M = 128$ we fit seven chirplets to the whale whistle (the stopping criteria was subjective), and display the results in Fig. 5b. For comparison we ran the algorithm of Section VI without the EM refinement, and eight chirplets were required for a similar decomposition (measured by the norm of the residual). With $N = 4000$ and $M = 128$, the computation time was approximately 5 seconds per chirplet on a Sparc Ultra 60 running Matlab.

In Fig. 6 we show an example with another whale whistle. With $M = 128$, we fit eight chirplets to the whale whistle. Without the EM refinement, ten chirplets were required to obtain a similar approximation.

It is not clear how to evaluate the de-noising capabilities of the algorithm on the whale whistles. The SNR is low enough that is not meaningful to compare the time-series or the spectrum of the data with the approximation. The spectrograms of the signal and the approximation are very similar in the regions where the signal is, and while this is encouraging, it is not a definitive answer. The original whistles and their approximations were played to an expert who stated that the approximations sounded like the original with the noise removed [46].

VIII. Discussion

In this last section we would like to compare the method proposed here with previously proposed methods with respect to performance (sparsity and high resolution) and computational complexity [1]. We would like to separate existing methods into two classes. The first are methods that use a global optimization strategy to provide high performance at the cost of high computational complexity [1], [2]. These methods will clearly perform better than the method proposed here but with much greater computational complexity. The second are those that use a sequential procedure (as done here in Section VI) to provide less sparsity and less resolution but with significantly less computation [3], [8], [9], [10], [7]. Matching Pursuit (MP) [3], [8] is representative of these methods and we shall present a detailed comparison between MP and the method presented here.

Overall, the method presented here and MP have a similar strategy. Although not explicitly stated, each iteration of MP is seeking to solve the MLE for one chirplet, discussed in Section II. There are two significant differences between our method and MP. The first difference is that MP essentially uses a different approximate MLE than the one that we developed in Section IV. In an attempt to find the maximum of the likelihood function over time, frequency, chirp rate, and duration, MP uses a coarse, four-dimensional grid search over the entire parameter space. The second difference is the addition of the EM refinement in Section VI. Without the EM refinement there is no connection with the likelihood function for the q chirplet model. With the EM refinement, we shall always have at least a local maximum of the likelihood function for the q chirplet model. We shall now compare our method with MP for the three criteria of sparsity, high resolution, and computational complexity.

Sparsity – Our method without the EM refinement will have similar sparsity to MP. However, the EM refinement can increase the sparsity as evidenced in the two whale whistle examples, where the number of chirplets was reduced from eight to seven, and from ten to eight, respectively.

High Resolution – If we assume that the signal is really a sum of chirplets in noise, then the question is "How close can the chirplets be and still be resolved by the method?" MP is well known to suffer from poor resolution properties. Our algorithm will have the potential to resolve chirplets that are very close, since this corresponds to the global maximum of the likelihood function. However, when chirplets are close, the likelihood function will have a complicated structure and many local maxima, so it will be likely that we find one of these local maxima. As demonstrated by the synthetic example in Fig. 4 the EM refinement step increases the resolution of our method. We are currently investigating estimators that are robust to misspecification [47] as a means for further improving the res-

olution.[7]

Speed – The computations for MP are a $O(N^2 \log N)$ initialization followed by $O(N \log N)$ per chirplet. Our algorithm does not have an initialization step, the computations are $O(N \log N)$ per chirplet, and by incorporating prior knowledge, as was done for the whale whistle examples, it can be reduced to $O(N)$ per chirplet.

Thus, the method described here will generally perform better than MP with less computation. A limitation of this method is that it assumes that the signal is well modeled as a sum of a small number of chirplets. It is straightforward to apply MP to other signal models, such as a sum of wavepackets [3] or a sum of splines [5]. In principle, our method can be applied to any model, but much of the work here is specific to the chirplet model and would not necessarily apply to other models.

Much of this paper can be characterized as a straightforward application of estimation theory and adaption of previous work for estimating chirp parameters. The significant innovations are the technical details discussed in the appendices and the formulation of the approximate MLE in Section IV. The primary contribution of this paper is the conglomeration of ideas from sparse representation theory, time-frequency analysis, and estimation theory to produce a powerful method for obtaining sparse representations.

Matlab software for implementing the chirplet decomposition is available at `http://mdsp.bu.edu/jeffo`.

Acknowledgements

The first author would like to thank Eric Chassande-Mottin, Phillip Ainsleigh, and Tod Luginbuhl for interesting conversations regarding this work, and also the Naval Undersea Warfare Center for providing the whale data.

Appendices

I. Likelihood Function

The relations

$$a_1(n-t_1)^2 + a_2(n-t_2)^2 = (a_1+a_2)\left(n - \tfrac{a_1t_1+a_2t_2}{a_1+a_2}\right)^2 + \tfrac{a_1a_2}{a_1+a_2}(t_1-t_2)^2 \quad (12)$$

and

$$\left| \int e^{-(t/2d)^2} e^{j(ct^2/2+\omega t)}\, dt \right|^2 = \tfrac{\pi}{|1/4d^2 - jc/2|} \exp\left\{ \tfrac{-2\omega^2}{d^{-2}+4c^2} \right\} \quad (13)$$

will be used in the proofs below. The first is simple algebra and the second comes from the Fourier transform of a chirp [6].

Lemma 1. *The function*

$$f(t,\omega) = \left| \langle s(n;t_o,\omega_o,c_o,d_o), s(n;t,\omega,c,d) \rangle \right|^2$$

has a unique maximum at $t = t_o$ and $\omega = \omega_o$ for any values of t_o, ω_o, c_o, d_o, c, and d.

Proof. Without loss of generality it may be assumed that $t_o = \omega_o = 0$. Let $\alpha = td_o^2/(d^2+d_o^2)$ then by applying (12)

$$f(t,\omega) = \tfrac{1}{2\pi d d_o} \exp\left\{ \tfrac{-t^2}{2(d^2+d_o^2)} \right\} \left| \sum_n \exp\Big\{ -\tfrac{d^2+d_o^2}{4d^2d_o^2}(n-\alpha)^2 + j\tfrac{c_o-c}{2}\left(n - \tfrac{ct}{c-c_o}\right)^2 + jn(\omega_o - \omega) \Big\} \right|^2$$

Further simplification provides

$$f(t,\omega) = \tfrac{1}{2\pi d d_o} \exp\left\{ \tfrac{-t^2}{2(d^2+d_o^2)} \right\} \left| \sum_n \exp\Big\{ -\tfrac{d^2+d_o^2}{4d^2d_o^2}n^2 + jn^2\tfrac{c_o-c}{2} + jn(\alpha(c_o-c) - (\omega - ct)) \Big\} \right|^2$$

The discrete approximation to (13) is accurate if we assume the signal is essentially band limited. Thus

$$f(t,\omega) = \tfrac{1}{|d^2+d_o^2 - j2d^2d_o^2(c_o-c)|} \exp\left\{ -\tfrac{t^2}{2(d^2+d_o^2)} - \tfrac{4d^2d_o^2}{d^2+d_o^2+8d^2d_o^2(c_o-c)^2}\left(t\tfrac{d^2c_o+d_o^2c}{d^2+d_o^2} - \omega \right)^2 \right\}$$

and the function is maximized at $t = 0$ and $\omega = 0$. □

Lemma 2. *The function*

$$g(c,d) = \left| \langle s(n;t_o,\omega_o,c_o,d_o), s(n;t_o,\omega_o,c,d) \rangle \right|^2$$

has a unique maximum at $c = c_o$ and $d = d_o$ for any values of t_o, ω_o, c_o, and d_o.

Proof. Without loss of generality it may be assumed that $t_o = \omega_o = 0$.

$$g(c,d) = \tfrac{1}{2\pi d d_o} \left| \sum_n \exp\left\{ -\tfrac{d_o^2+d^2}{4d^2d_o^2}n^2 + j\tfrac{c_o-c}{2}n^2 \right\} \right|^2$$

by applying (13)

$$g(c,d) = \left(\left(\tfrac{d_o^2+d^2}{2dd_o} \right)^2 + d_o^2d^2(c_o-c)^2 \right)^{-1/2}$$

which has a unique maxima at $c = c_o$ and $d = d_o$. □

Theorem 1. *The function*

$$h(t,\omega,c,d) = \left| \langle s(n;t_o,\omega_o,c_o,d_o), s(n;t,\omega,c,d) \rangle \right|^2$$

has a unique maximum at $t = t_o$, $\omega = \omega_o$, $c = c_o$, and $d = d_o$.

Proof. Lemmas one and two. □

[7] The model is mis-specified since at each iteration we are looking for one chirplet in noise, where in reality the signal consists of multiple chirplets or is something completely different.

II. Ambiguity Function

Let $y(t) = x(t)\, e^{jct^2/2}$ and $r_y(\tau) = y(t)\, y^*(t+\tau)$, then

$$\int |A_x(cr, r)|^2 \, dr = \iint |A_x(\theta + c\tau, \tau)|^2 \, \delta(\theta) \, d\theta \, d\tau =$$
$$\iint |A_y(\theta, \tau)|^2 \, \delta(\theta) \, d\theta \, d\tau = \int |A_y(0, \tau)|^2 \, d\tau =$$
$$\int |r_y(\tau)|^2 \, d\tau = \int |Y(\omega)|^4 \, d\omega.$$

Intermediate steps use well known properties of the ambiguity function [6].

References

[1] S. Chen and D. Donoho. Atomic decomposition by basis pursuit. *Siam Journal on Scientific Computing*, 20(1):33–61, January 1999.

[2] I. Gorodnitsky and B.D. Rao. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. on Signal Processing*, 45(3):600–616, March 1997.

[3] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3415, December 1993.

[4] B.D. Rao. Signal processing with the sparseness constraint. In *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, 1998.

[5] S. Jaggi, W.C. Karl, S. Mallat, and A. Willsky. Feature extraction through high-resolution pursuit. *Applied and Computational Harmonic Analysis*, 5(4):428–449, October 1998.

[6] P. Flandrin. *Time-Frequency/Time-Scale Analysis.* Academic Press, San Diego, 1999.

[7] M.L. Brown. *Optimal representation of transient biological signals using the adaptive Gabor transform.* PhD thesis, Univ. of Michigan, 1994.

[8] A. Bultan. A four-parameter atomic decomposition of chirplets. *IEEE Trans. on Signal Processing*, 47(3):731–745, March 1999.

[9] S. Qian and D. Chen. Signal representation using adaptive normalized Gaussian functions. *Signal Processing*, pages 1–11, March 1994.

[10] S. Qian and D. Chen. Adaptive chirplet based signal approximation. In *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, 1998.

[11] P. Djuric and S. Godsill. Paremetric modeling and estimation of time-varying spectra. In *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, 1998.

[12] E.J. Zalubas, J.C. O'Neill, W.J. Williams, and A.O. Hero III. Shift and scale invariant detection. In *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, volume 5, pages 3637–3640, 1997.

[13] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[14] L. Cohen. *Time-Frequency Analysis.* Prentice Hall, Englewood Cliffs, NJ, 1995.

[15] R.L. Hudson. When is the Wigner quasi-probability density non negative? *Rep. Math. Phys.*, 6:249–252, 1974.

[16] S. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1993.

[17] G.A.F. Seber and C.J. Wild. *Nonlonear Regression.* Wiley, 1989.

[18] J.C. O'Neill and P. Flandrin. Cramér-Rao lower bounds for atomic decomposition. In *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, volume 3, pages 1581–1584, 1999.

[19] S. Golden and B. Friedlander. Maximum likelihood estimation, analysis, and applications of exponential polynomial phase signals. *IEEE Trans. on Signal Processing*, 47(6), June 1999.

[20] S. Barbarossa. Analysis of multicomponent LFM signals by a combined Wigner-Hough transform. *IEEE Trans. on Signal Processing*, 43(6):1511–1515, June 1995.

[21] F.S. Cohen, S. Kadambe, and G.F. Boudreaux-Bartels. Tracking of unknown nonstationary chirp signals using unsupervised clustering in the Wigner distribution space. *IEEE Trans. on Signal Processing*, 41(11):3085–3101, November 93.

[22] P. Djurić and S.M. Kay. Parameter estimation of chirp signals. *IEEE Trans. on Signal Processing*, 38(12):2118–2126, December 1990.

[23] B. Friedlander and J.M. Francos. Estimation of amplitude and phase parameters of multicomponent signals. *IEEE Trans. on Signal Processing*, 43(4):917–926, April 1995.

[24] F. Gini and G.B. Giannakis. Hybrid FM-polynomial phase signal modeling: Parameter estimation and Cramér-Rao lower bounds. *IEEE Trans. on Signal Processing*, 47(2):363–377, February 1999.

[25] S. Golden and B. Friedlander. Estimation and statistical analysis of exponential polynomial signals. *IEEE Trans. on Signal Processing*, 46(11):3127–3131, November 1998.

[26] M.Z. Ikram, K. Abed-Meriam, and Y. Hua. Estimating the parameters of chirp signals: An iterative approach. *IEEE Trans. on Signal Processing*, 46(12):3436–3441, December 1998.

[27] S. Kay and G.F. Boudreaux-Bartels. On the optimality of the Wigner distribution for detection. In *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, pages 1017–1020, 1985.

[28] H. Li and P.M. Djurić. MMSE estimation of nonlinear parameters of multiple linear/quadratic chirps. *IEEE Trans. on Signal Processing*, 46(3):796–800, March 1998.

[29] S. Peleg and B. Porat. Estimation and classification of polynomial phase signals. *IEEE Trans. on Information Theory*, 37(2):422–430, March 1991.

[30] W. Wang, A.K. Chan, and C.K. Chui. Linear frequency-modulated signal detection using Radon-ambiguity transform. *IEEE Trans. on Signal Processing*, 46(3):571–586, March 1998.

[31] J.C. Wood and D.T. Barry. Radon transformation of time-frequency distributions for analysis of multicomponent signals. *IEEE Trans. on Signal Processing*, 42(11):3166–3177, November 1994.

[32] G. Zhou, G.B. Giannakis, and A. Swami. On polynomial phase signals with time-varying amplitudes. *IEEE Trans. on Signal Processing*, 44(4):848–861, April 1996.

[33] J.C. O'Neill and W.J. Williams. Shift covariant time-frequency distributions of discrete signals. *IEEE Trans. on Signal Processing*, 47(1):133–146, January 1999.

[34] H.L. Van Trees. *Detection, estimation, and modulation theory, part III.* Wiley, New York, 1971.

[35] B. Ristic and B. Boashash. Kernel design for time-frequency signal analysis using the Radon transform. *IEEE Trans. on Signal Processing*, 41(5):1996–2008, May 1993.

[36] J.C. Wood and D.T. Barry. Tomographic time-frequency analysis and its application toward time-varying filtering and adaptive kernel design for multicomponent linear-FM signals. *IEEE Trans. on Signal Processing*, 42(8):2094–2104, August 1994.

[37] J.C. O'Neill and W.J. Williams. A function of time, frequency, lag, and doppler. *IEEE Trans. on Signal Processing*, 47(3):789–799, March 1999.

[38] J.C. O'Neill and P. Flandrin. Chirp hunting. In *Proc. of the IEEE Int. Symp. on Time-Frequency and Time-Scale Analysis*, pages 425–428, 1998.

[39] J.C. O'Neill and P. Flandrin. Virtues and vices of quartic time-frequency distributions. *to apear in IEEE Trans. on Signal Processing*, 2000. `http://mdsp.bu.edu/jeffo`.

[40] W. Li. Wigner distribution method equivalent to dechirp method for detecting a chirp signal. *IEEE Trans. on Acoust., Speech, and Signal Processing*, 35(8):1210–1211, August 1987.

[41] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood estimation from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Ser. B (methodological)*, 39:1–38, 1977.

[42] M. Feder and E. Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. on Acoust., Speech, and Signal Processing*, 36(4):477–489, April 1988.

[43] G.J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions.* Wiley, 1997.

[44] H. Akaike. A new look at statistical model identification. *IEEE Trans. Autom. Control*, AC-19:716–723, dec 1974.

[45] D.L. Jones and R.G. Baraniuk. An adaptive optimal kernel time-frequency representation. *IEEE Trans. on Signal Processing*, 43(10):2361–2371, October 1995.

[46] Stephen Greineder. personal communication, 1999.

[47] H. White. *Estimation, Inference, and Specification Analysis.* Cambridge University Press, 1996.

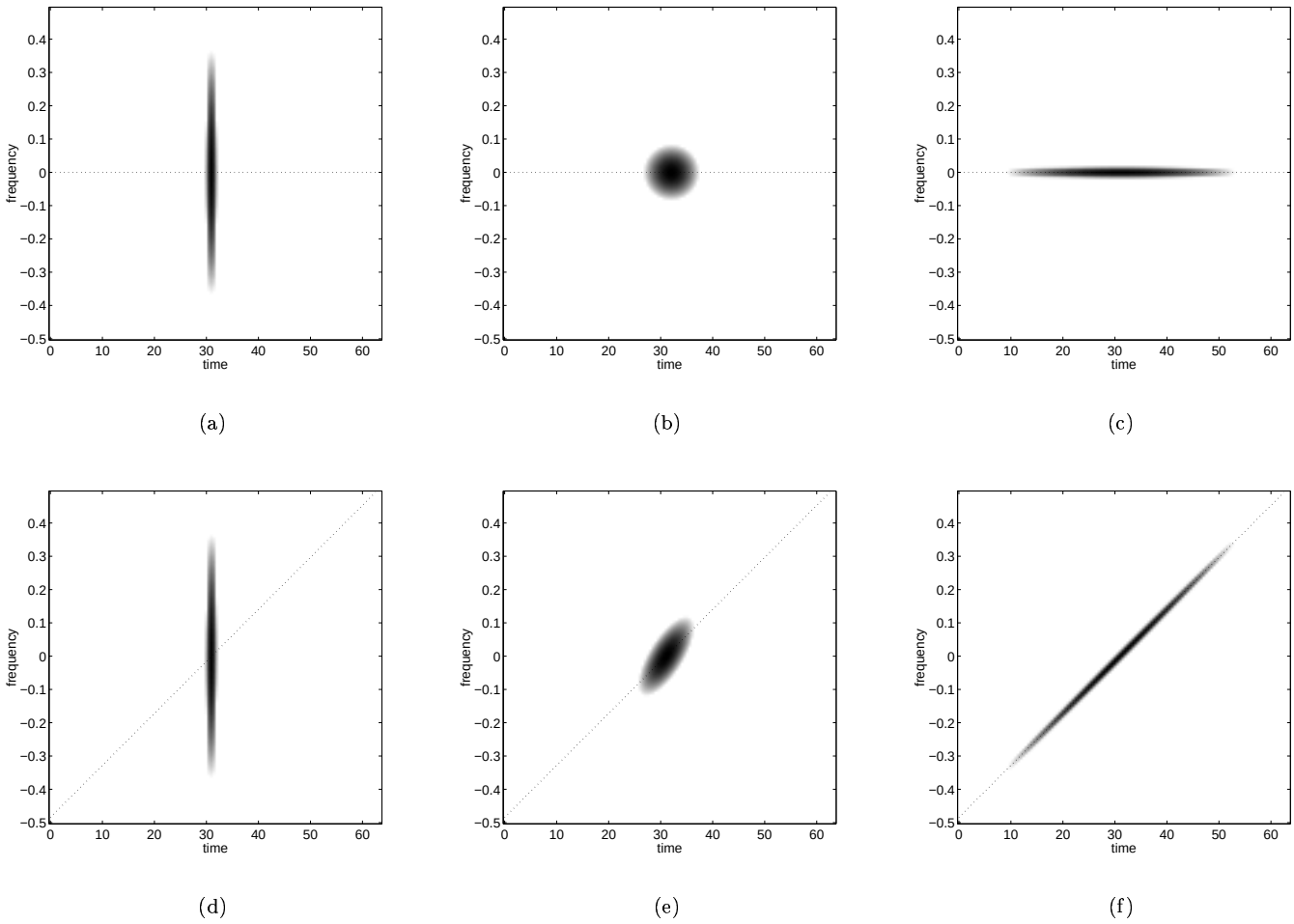


Fig. 1. Figures (a), (b), and (c) have a chirp rate of 0, while figures (d), (e), and (f) all have a chirp rate of 1/64. The chirp rates are indicated by the dotted lines, and the only difference within each row is the duration of the chirp.

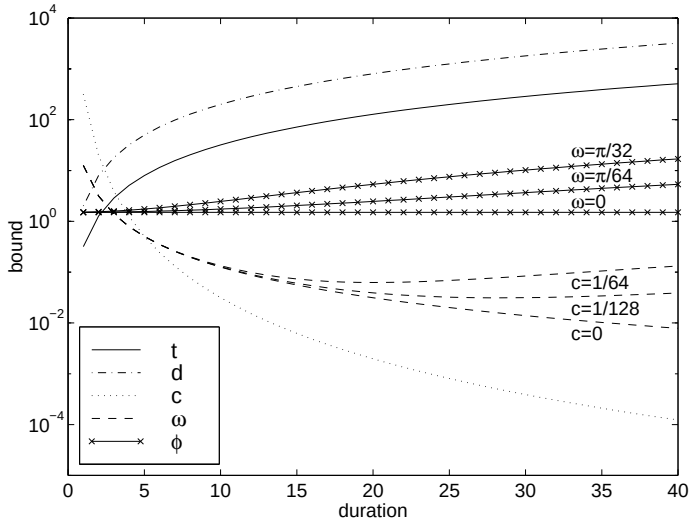


Fig. 2. The Cramér-Rao lower bounds as a function of the independent variables.

TABLE I
COMPUTATIONS

Step	Computations
Global estimate of c	$O(MN \log N)$
Global estimate of d	$O(MN \log N)$
Estimate of t and ω	$O(NM \log M)$
Local estimate of c	$O(M^2 \log M)$
Local estimate of d	$O(M^2)$
Quasi-Newton Maximization	$O(M)$

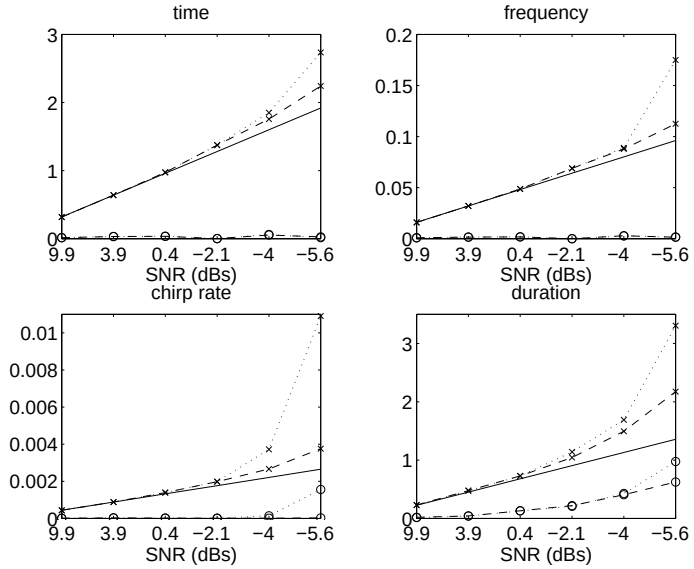
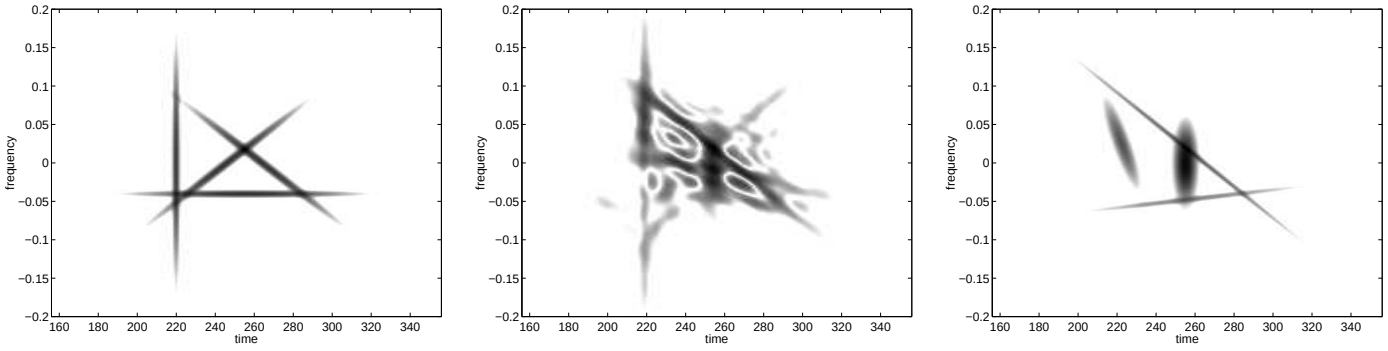
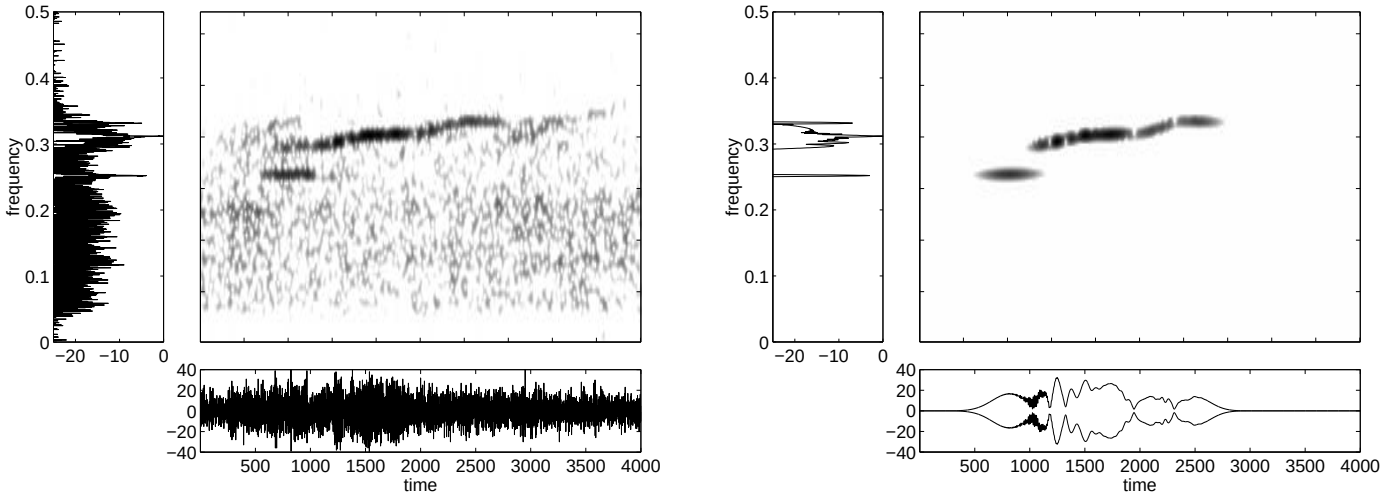


Fig. 3. Simulation Results, CRLB (solid line), MLE (dashed line), approximate MLE (dotted line). x corresponds to the standard deviation and o corresponds to the bias.



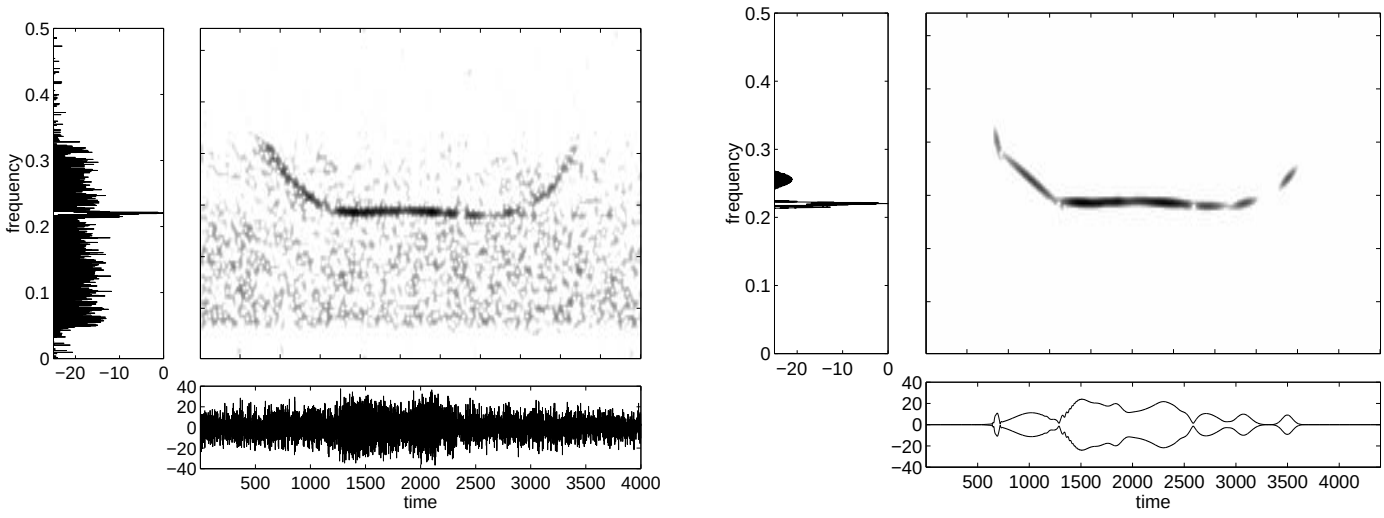
(a) (b) (c)

Fig. 4. (a) is the sum of the Wigner distributions of the four chirplets. (b) is the AOK TFD of the four chirplets. (c) is the sum of the Wigner distributions of the estimated chirplets without the EM refinement step.



(a) Whale Whistle (b) Approximation

Fig. 5. First example. (a) is the time-series, spectrum, and spectrogram of a whale whistle. (b) is the envelope of the time-series, the spectrum, and spectrogram of the approximation with seven chirplets.



(a) Whale Whistle (b) Approximation

Fig. 6. Second example. On the left is the time-series, spectrum, and spectrogram of a whale whistle. On the right is the envelope of the time-series, the spectrum, and spectrogram of the approximation with eight chirplets.