

CRAMÉR-RAO LOWER BOUNDS FOR ATOMIC DECOMPOSITION

Jeffrey C. O'Neill

Patrick Flandrin

Electrical and Computer Engineering
Boston University

Laboratoire de Physique
Ecole Normale Supérieure de Lyon

ABSTRACT

In a previous paper [1] we presented a method for atomic decomposition with chirped, Gabor functions based on maximum likelihood estimation. In this paper we present the Cramér-Rao lower bounds for estimating the seven chirp parameters, and the results of a simulation showing that our sub-optimal, but computationally tractable, estimators perform well in comparison to the bound at low signal-to-noise ratios. We also show that methods based on signal dictionaries will require much higher computations to perform well in low signal-to-noise ratios.

1. INTRODUCTION

Given a signal, $x(n)$, our goal is to find a sparse decomposition of the signal as a weighted sum of chirped, Gabor functions

$$\mathbf{x} = x(n) = \sum_{i=1}^M A_i e^{j\phi_i} s(n; t_i, \omega_i, c_i, d_i),$$

where

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

The parameters t , ω , c , and d represent, respectively, the location in time, the location in frequency, the chirp rate, and the duration, and $s()$ is defined such that

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

In [1] we presented a method for obtaining this decomposition based on the principles of maximum likelihood estimation which we will briefly review in the next section. We will then present the Cramér-Rao lower bounds along with estimates of the mean and variance of the estimators proposed in [1]. Finally we will compare our methods with those that use signal dictionaries and investigate the issue of greed.

2. ESTIMATION OF CHIRP PARAMETERS

In [1] we presented the maximum likelihood estimator (MLE) for the following signal model

$$x(n) = A e^{j\phi} s(n; t, \omega, c, d) + w(n) \quad n = 1, \dots, N, \quad (1)$$

where $w(n)$ is complex, white, gaussian noise (CWGN) with a mean of zero and a variance of σ^2 , and the seven unknown parameters are

$$\boldsymbol{\theta} = [\sigma^2 \quad A \quad d \quad c \quad t \quad \omega \quad \phi].$$

The MLE can be formulated as follows

$$[\hat{t}_{ml} \quad \hat{\omega}_{ml} \quad \hat{c}_{ml} \quad \hat{d}_{ml}] = \arg \max_{t,\omega,c,d} |\langle \mathbf{x}, \mathbf{s}_{t,\omega,c,d} \rangle|^2 \quad (2a)$$

$$z = \langle \mathbf{x}, \mathbf{s}_{\hat{t}_{ml}, \hat{\omega}_{ml}, \hat{c}_{ml}, \hat{d}_{ml}} \rangle \quad (2b)$$

$$\hat{\phi}_{ml} = -\angle z \quad (2c)$$

$$\hat{A}_{ml} = |z| \quad (2d)$$

$$\hat{\sigma}_{ml}^2 = (\|\mathbf{x}\|^2 - |z|^2) / 2N. \quad (2e)$$

We have shown that, in the absence of noise, equation 2a is a uni-modal function. However in the presence of noise, solving equation 2a is computationally expensive, so we will approximate it with the following sub-optimal estimator based on the results in [1].

1. Estimate, globally, the chirp rate and duration with the ambiguity function.
2. Using the current estimate of the chirp rate and the duration, estimate the location in time and frequency with the spectrogram.
3. Estimate, locally, the chirp rate and the duration using the methods proposed in [1].
4. Go to step 2 until the parameters converge.
5. Find the nearest local maximum of the likelihood function by using a quasi-Newton procedure.

The estimated chirp is then subtracted from the signal, and the procedure repeated to find the next element in the decomposition.

3. CRAMÉR-RAO LOWER BOUNDS

Here we will present examples that provide intuition for understanding the Cramér-Rao lower bounds (CRLBs), followed by the actual derivation of the CRLBs.

In Figure 1 we present the Wigner distributions of six chirps. The first three have a chirp rate of zero, and the only difference between them is the value of the duration parameter. The last three have a chirp rate of $1/64$, and again the only difference between them is the value of the duration parameter. Using these examples, we present our intuitions about estimating the location in time and frequency, the chirp rate, and the duration. (The other three parameters are more obvious.)

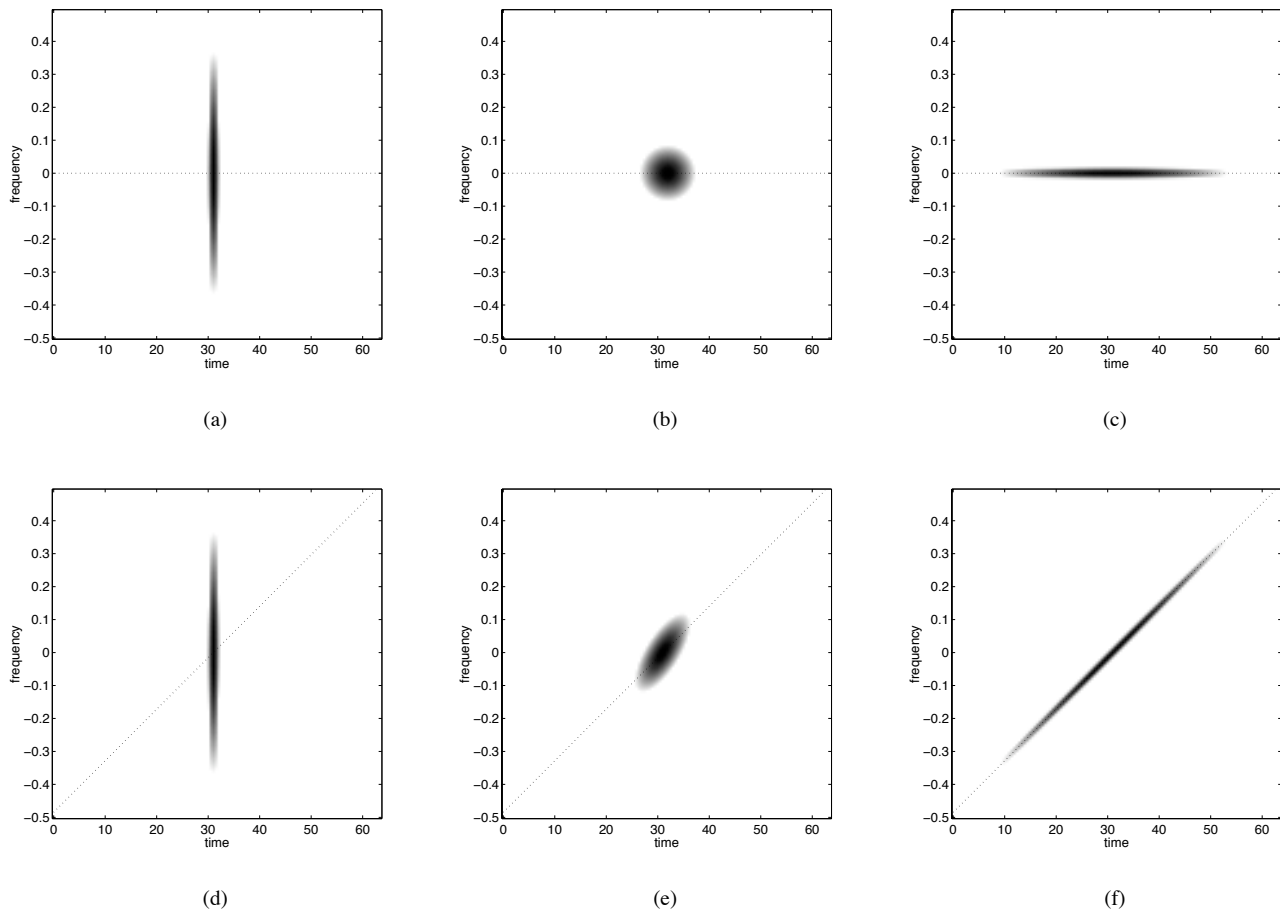


Figure 1: Figures (a), (b), and (c) have a chirp rate of 0, while figures (d), (e), and (f) all have a chirp rate of $1/64$. The chirp rates are indicated by the dotted lines, and the only difference within each row is the duration of the chirp.

- When the duration is short, 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.
- 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 long durations (e.g. Figure 1(f)). In fact, the signal will be the most concentrated in frequency for some intermediate duration (e.g. Figure 1(e)). Thus for non-zero chirp rates, the estimation of the location in frequency should be difficult for both short and long durations with a minimum somewhere in between.
- The estimation of the duration should be analogous to the estimation of the location in time.
- When the duration is short, the chirp rate should be very difficult to estimate (compare Figures 1(a) and (d)). As the duration increases, it should become easier to estimate the chirp rate.

The calculation of the Fisher information matrix is tedious,

though straightforward [2] and the inverse is

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

where N is the number of samples of $x(n)$. The CRLBs are the diagonal elements of the matrix. The derivation is based on the following approximations

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

which are valid as long as the signal has been sampled appropri-

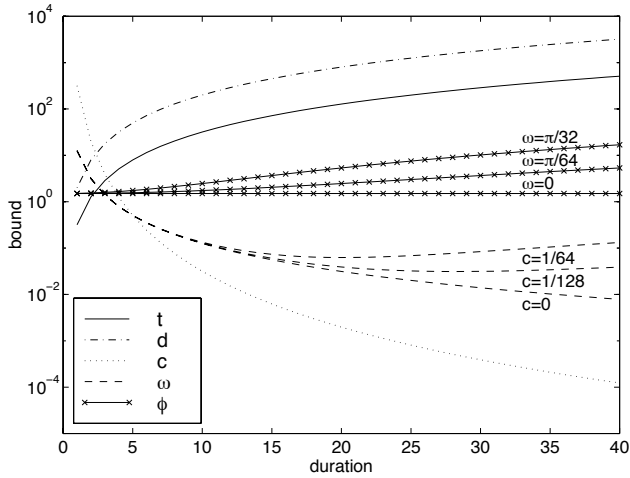


Figure 2: The Cramér-Rao lower bounds as a function of the independent variables.

ately. The bounds are plotted in Figure 2 as a function of the independent variables. It is straightforward to see that the bounds are in agreement with the intuition presented above.

4. SIMULATION RESULTS

To evaluate the performance of our sub-optimal estimators, we ran simulations corresponding to equation 1 with $N = 64$, $t = 32$, $\omega = 0$, $c = 2\pi/64 \approx 0.0982$, $d = 5$, $A = 1$, and $\phi = 0$. Four noise levels were chosen $\sigma = 0.05, 0.08, 0.10$, and 0.12 which correspond to signal-to-noise ratios (SNRs)¹ of 4.9, 0.9, -1.1, and -2.7 dB, respectively. We ran 10000 trials for each noise level and applied the standard sample mean and sample variance formulae. The results are shown in Table 1 in the columns corresponding to e_1 .

Since the signal model in equation 1 is not a linear function of the unknown parameters, the MLE will not be efficient and thus not meet the CRLBs with equality (except asymptotically). Thus, the CRLBs are not realistic bounds for our sub-optimal estimators. To provide a more realistic bound, we ran another simulation consisting of finding the local maxima of the likelihood function closest to the true value of the parameters. This is obviously an unrealistic estimator, since it requires the actual values of the parameters, but this estimator should have roughly the same performance as the MLE for high SNR (since the closest local maximum is very likely to be the global maximum) and should perform better than the MLE for low SNR (since the global maximum is not necessarily the closest local maximum). The results are shown in Table 1 in the columns corresponding to e_2 .

Table 1 confirms that the MLE is not unbiased and not efficient. As a result, we can not expect our sub-optimal estimators to be unbiased or efficient. However, the sub-optimal estimator performs well in comparison to the unrealistic estimator, described above, and thus performs well in comparison to the MLE.

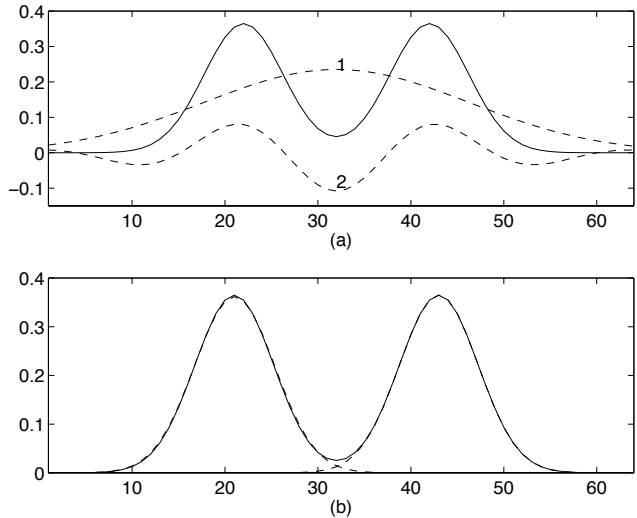


Figure 3: Two examples showing the “greediness” of the estimator.

5. DICTIONARY METHODS

Another approach to performing atomic decomposition is based on signal dictionaries [3]. To do this, one creates a signal dictionary matrix, \mathbf{D} , where the columns are chirped, Gabor functions with different parameter values. Since the number of dictionary elements, M , will generally be much larger than the signal length, N , one approach is to solve the under-determined, linear, inverse problem $\mathbf{D}\mathbf{b} = \mathbf{x}$, where \mathbf{b} provides the weighting of each dictionary element. The Moore-Penrose inverse provides the minimum ℓ^2 norm solution, but this solution will not be sparse. Some methods which attempt to find sparse solutions to the above linear, inverse problem are Matching Pursuit [4, 5, 6], Basis Pursuit [7], and FOCUSS [8].

If one discretizes the parameters of t , ω , c , and d to $\mathcal{O}(N)$ values, then the dictionary will contain $\mathcal{O}(N^4)$ elements and all of the above dictionary methods will be extremely expensive computationally. To circumvent this problem, one can discretize the parameters at lower rates to decrease the size of the dictionary. For example in [4, 5], the location in time was discretized at a rate of $\mathcal{O}(\sqrt{N})$, but since the minimum distance between dictionary elements in time will be $\mathcal{O}(\sqrt{N})$, Matching Pursuit will likely find a local maximum, rather than the global maximum, as N increases. Thus, we expect the performance to greatly deteriorate as N increases.

With dictionary methods, it does not seem possible to have a computationally feasible algorithm that will reliably find the global maximum. In contrast, the performance of the method that we present here and in [1] is independent of N and requires a reasonable $\mathcal{O}(N^2 \log N)$ computations.

6. GREED

Both the method presented here and Matching Pursuit attempt to maximize the objective function in equation 2a, though the meth-

¹Calculated according to $10 \log_{10} \frac{A^2}{2N\sigma^2}$.

| | noise (σ) | $E(e_1)$ | $E(e_2)$ | value | $Var(e_1)$ | $Var(e_2)$ | CRLB |
|----------|--------------------|----------|----------|--------|------------|------------|---------|
| t | 0.05 | 32.00 | 32.00 | 32 | 0.254 | 0.254 | 0.250 |
| | 0.08 | 32.00 | 32.00 | 32 | 0.687 | 0.687 | 0.640 |
| | 0.10 | 32.01 | 32.01 | 32 | 1.211 | 1.194 | 1.000 |
| | 0.12 | 32.02 | 32.01 | 32 | 2.150 | 1.864 | 1.440 |
| ω | 0.05 | 0.000 | 0.000 | 0 | 0.0025 | 0.0025 | 0.0025 |
| | 0.08 | 0.000 | 0.000 | 0 | 0.0069 | 0.0069 | 0.0064 |
| | 0.10 | 0.001 | 0.001 | 0 | 0.0121 | 0.0120 | 0.0100 |
| | 0.12 | 0.003 | 0.001 | 0 | 0.0216 | 0.0187 | 0.0145 |
| c | 0.05 | 0.0981 | 0.0981 | 0.0982 | 0.89e-5 | 0.89e-5 | 0.80e-5 |
| | 0.08 | 0.0981 | 0.0981 | 0.0982 | 2.62e-5 | 2.62e-5 | 2.05e-5 |
| | 0.10 | 0.0981 | 0.0982 | 0.0982 | 5.59e-5 | 4.80e-5 | 3.20e-5 |
| | 0.12 | 0.0972 | 0.0979 | 0.0982 | 24.6e-5 | 9.56e-5 | 4.61e-5 |
| d | 0.05 | 5.04 | 5.04 | 5 | 0.136 | 0.136 | 0.125 |
| | 0.08 | 5.11 | 5.11 | 5 | 0.412 | 0.412 | 0.320 |
| | 0.10 | 5.22 | 5.21 | 5 | 0.809 | 0.779 | 0.500 |
| | 0.12 | 5.47 | 5.36 | 5 | 1.840 | 1.467 | 0.720 |

Table 1: Simulation Results

ods for doing so are quite different. As a result, our method, like Matching Pursuit, will be a greedy algorithm and will not be able to resolve atoms that are closely spaced.

The signal model in equation 1 for the maximum likelihood estimator is a single chirped, Gabor function in CWGN. Thus when presented with a two-component signal, like the solid line in Figure 3a, the most likely answer is the chirped, Gabor function that lies in between (indicated by the dashed line marked with a “1”), and the discrepancies are attributed to noise. The second term in the decomposition has to compensate for the poor initial choice (the dashed line marked “2”).

If the two components of the signal are a little more separated, as in Figure 3b, then the most likely answer corresponds to one of the components, and the other component is attributed to a bizarre noise realization. Here, the second term in the decomposition corresponds precisely with the other signal component. Thus it seems that both algorithms will resolve components that are sufficiently separated, though we have not yet quantified “sufficiently”.

Given the formulation presented here, an obvious choice for overcoming the greediness of the algorithm is to use a signal model consisting of multiple chirped, Gabor functions in CWGN. The likelihood function for this model will no longer be a uni-modal function in the absence of noise, thus finding the global maximum is a much more difficult task. One can try to maximize this likelihood function with a quasi-Newton procedure, though this becomes increasing difficult as the number of parameters is increased. Initial results with the Expectation-Maximization (EM) algorithm [10, 11, 12] are promising, though the rate of convergence and the convergence to local, rather than global, maxima needs to be investigated.

Software implementing the decomposition described here is available at <http://www.eecs.umich.edu/~jefffo>.

7. REFERENCES

- [1] J. C. O’Neill and P. Flandrin. Chirp hunting. In *Proc. of the IEEE Int. Symp. on Time-Frequency and Time-Scale Analysis*, 1998.
- [2] S. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice Hall, 1993.
- [3] B.D. Rao. Signal processing with the sparseness constraint. In *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, 1998.
- [4] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3415, December 1993.
- [5] A. Bultan. A four-parameter atomic decomposition of chirplets. *to appear in IEEE Trans. on Signal Processing*, 1998.
- [6] S. Qian and D. Chen. Signal representation using adaptive normalized Gaussian functions. *Signal Processing*, pages 1–11, March 1994.
- [7] S. Chen and D. Donoho. Atomic decomposition by basis pursuit. *to appear in Siam Journal on Scientific Computing*, 1998.
- [8] 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.
- [9] M.L. Brown. *Optimal representation of transient biological signals using the adaptive Gabor transform*. PhD thesis, Univ. of Michigan, 1994.
- [10] 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.
- [11] R.A. Redner and H.F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, April 1984.
- [12] 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.