HIGH RESOLUTION NEARLY-ML ESTIMATION OF SINUSOIDS IN NOISE USING A FAST FREQUENCY DOMAIN APPROACH

Dr Malcolm D. Macleod
Cambridge University Engineering Department, Trumpington Street,
CAMBRIDGE, CB2 1PZ, UK Tel: +44 1223 332671; fax: +44 1223 332662
e-mail: mdm@eng.cam.ac.uk

ABSTRACT

Estimating the frequencies, amplitudes and phases of sinusoids in noise is a problem which arises in many applications. The aim of the methods in this paper is to achieve computational efficiency and near-ML performance (i.e. low bias, variance and threshold SNR), in problems such as vibration or audio analysis where the number of tones may be large (e.g. > 20). An approach has recently been published for resolved tones [4]. This paper extends that frequency domain approach to the high-resolution problem.

1 INTRODUCTION

The estimation of the frequencies, amplitudes and phases of sinusoids in noise is important in many applications, including radar, sonar, instrumentation, and audio analysis. In such applications, many of the tones may not be resolved (i.e. their frequency separations may be $< 4\pi/N$ rad/sample, where N is the block-length). In cases like these, especially where the number of tones is large, Maximum Likelihood (ML) estimation is usually rejected [1] because it requires computationally expensive non-linear optimisation. A recent algorithm [3,4] uses frequency-domain interpolators, coupled with a simple non-linear optimisation strategy, to obtain nearly-ML estimates in the case of resolved multiple tones (frequencies separated by at least $4\pi/N$ rad/sample). This paper extends these results to the high-resolution case, giving very nearly ML estimates with much reduced computation.

1.1 Problem definition

The observed discrete signal y_n is modelled as $y_n = x_n + z_n$, where x_n is the sum of M *cisoids* (complex sinusoids) and z_n is zero-mean complex noise [1] of variance σ^2, with independent real and imaginary parts, each of variance $\sigma^2/2$. If the cisoids have amplitudes a_i, phases ϕ_i, and frequencies ω_i radians per sample, x_n can be written as

$$x_n = \sum_{i=1}^{M} b_i \exp(j\omega_i n), \qquad (1)$$

where $b_i = a_i \exp(\phi_i)$ is the *complex amplitude* of the i^{th} cisoid. Let the N-sample data blocks be written as column vectors, $\mathbf{y} = [y_0, y_1, ..., y_{N-1}]^T$, etc., and define parameter vectors $\mathbf{b} = [b_1, b_2, ..., b_M]^T$ and $\boldsymbol{\omega} = [\omega_1, \omega_2, ..., \omega_M]^T$. The problem, assuming M is known, is to estimate $\mathbf{b}$ and $\boldsymbol{\omega}$, given $\mathbf{y} = \mathbf{x} + \mathbf{z}$. (Pure real signals, and the problem of estimating M, are considered later).

Let matrix $\mathbf{G}$ have columns which are the cisoid basis functions at frequencies $\omega_1 ... \omega_M$:

$\mathbf{G}(\boldsymbol{\omega}) = [\mathbf{e}(\omega_1), \mathbf{e}(\omega_2), ..., \mathbf{e}(\omega_M)]$, where $\mathbf{e}(\omega) = [1, \exp(j\omega), ..., \exp(j(N-1)\omega)]^T$. Then the signal model (1) can be written $\mathbf{x} = \mathbf{G}(\boldsymbol{\omega})\mathbf{b}$. When the noise $\mathbf{z}$ is white (i.i.d.) Gaussian, the Maximum Likelihood (ML) estimate of the parameters ($\hat{\mathbf{b}}, \hat{\boldsymbol{\omega}}$) is the one [1] which minimises $S = ||\mathbf{y} - \mathbf{G}(\hat{\boldsymbol{\omega}})\hat{\mathbf{b}}||^2$, the sum of squared errors (SSE) between the estimated signal $\hat{\mathbf{x}} = \mathbf{G}(\hat{\boldsymbol{\omega}})\hat{\mathbf{b}}$ and the observed signal $\mathbf{y}$. For any given estimate $\hat{\boldsymbol{\omega}}$, the ML estimate of $\hat{\mathbf{b}}$ is given by [1]

$$\hat{\mathbf{b}}(\hat{\boldsymbol{\omega}}) = (\mathbf{G}(\hat{\boldsymbol{\omega}})^H \mathbf{G}(\hat{\boldsymbol{\omega}}))^{-1} \mathbf{G}(\hat{\boldsymbol{\omega}})^H \mathbf{y} \qquad (2)$$

and the joint ML estimate of $\mathbf{b}$ and $\boldsymbol{\omega}$ is found by maximising

$$\lambda(\hat{\boldsymbol{\omega}}) = \mathbf{y}^H \mathbf{G}(\hat{\boldsymbol{\omega}})(\mathbf{G}(\hat{\boldsymbol{\omega}})^H \mathbf{G}(\hat{\boldsymbol{\omega}}))^{-1} \mathbf{G}(\hat{\boldsymbol{\omega}})^H \mathbf{y} \qquad (3)$$

by searching over the M-dimensional $\hat{\boldsymbol{\omega}}$. For the single-tone case (M=1), $\mathbf{G}(\hat{\boldsymbol{\omega}}) = \mathbf{e}(\hat{\omega})$, and the ML estimate of the scalar $\hat{\omega}$ is obtained [1] by maximising the *periodogram* of the signal $\mathbf{y}$,

$$\begin{aligned} P(\hat{\omega}) &= \mathbf{y}^H \mathbf{e}(\hat{\omega})(\mathbf{e}(\hat{\omega})^H \mathbf{e}(\hat{\omega}))^{-1} \mathbf{e}(\hat{\omega})^H \mathbf{y} \\ &= (1/N)||\mathbf{e}(\hat{\omega})^H \mathbf{y}||^2, \end{aligned}$$

since $\mathbf{e}(\hat{\omega})^H \mathbf{e}(\hat{\omega}) = N$. From (2), the ML estimate of b is $\hat{b}(\hat{\omega}) = (1/N)Y^o(\omega)$, where $Y^o(\omega) = \mathbf{e}(\hat{\omega})^H \mathbf{y}$ is the DTFT of $\mathbf{y}$. The DTFT of a single cisoid at frequency ω_A is $\mathbf{e}(\omega)^H \mathbf{e}(\omega_A) = D_N(\omega - \omega_A)$, where

$$\begin{aligned} D_N(\omega) &= \sum_{k=0}^{N-1} \exp(-jk\omega) \\ &= \exp(-j\omega((N-1)/2)) \frac{\sin(\omega N/2)}{\sin(\omega/2)} \qquad (4) \end{aligned}$$

is a form of the Dirichlet kernel. It has the properties: $D_N(0) = N$; $D_N(k\omega_0) = 0$ if $k \neq 0$; and $D_N(\omega) \ll N$, if $\omega \gg \omega_0$. A traditional way to estimate $\hat{\omega}$ is perform a coarse search for the periodogram peak, using a zero-padded DFT, and then refine the estimate by optimisation. A more efficient approach [3,4] is to locate the peak in the standard DFT and estimate $\hat{\omega}$ using a closed-form interpolator in the discrete frequency domain.

For $M > 1$ the non-linear search over $\hat{\boldsymbol{\omega}}$ is in general computationally intensive. The elements of the matrix $\mathbf{T} = \mathbf{G}(\boldsymbol{\omega})^H\mathbf{G}(\boldsymbol{\omega})$, whose inverse appears in (3), are

$$T_{mn} = \mathbf{e}(\omega_m)^H\mathbf{e}(\omega_n) = D_N(\omega_m - \omega_n). \quad (5)$$

From (4), the diagonal elements $T_{mm} = N$, and $T_{mn} = T^*_{nm}$, so $\mathbf{T}$ is Hermitian.

1.2 Low resolution multiple tone ML analysis

If $\omega_m - \omega_n \gg \omega_0$ (where $\omega_0 = 2\pi/N$), the off-diagonal elements T_{mn} are much smaller than the diagonal elements, and the m^{th} and n^{th} tones produce resolved peaks in the periodogram. Simple application of a single-tone estimator to each peak gives biased estimates [2], caused by the non-zero off-diagonal elements of $\mathbf{T}$. Provided the tone frequencies are separated by at least $2\omega_0$ (2 'bins'), the bias may be removed [3,4] by a computationally simple iterative optimisation procedure which converges rapidly.

2 HIGH RESOLUTION ANALYSIS

The key to the new high resolution approach is to recognise that in typical multi-tone high-resolution problems, *some* of the tones will be resolved, while others will be in 'clusters' with frequency separations $< 4\pi/N$. Assume that the frequencies are indexed so that $\omega_1 < \omega_2 < ... < \omega_M$. Define a 'cluster' of L tones, with frequencies ω_j ... ω_{j+L-1}, by the property that the frequency separation between any tone in the cluster and any tone not in the cluster is much greater than ω_0. That is, for any ω_m, $j \leq m \leq j+L-1$, and ω_n, $n < j$ or $n > j+L-1$, we have $|\omega_m - \omega_n| \gg \omega_0$, hence $|D_N(\omega_m - \omega_n)| \ll N$.

Assume that in a given case there are K clusters. If matrix elements with magnitudes $\ll N$ are regarded as negligible, the matrix $\mathbf{T}$ has approximately the following structure (illustrated for the example of $K = 3$ 'clusters'):

$$\mathbf{T} = \left[\begin{array}{c|c|c} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_2 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}_3 \end{array}\right] \quad (6)$$

in which the square sub-matrices $\mathbf{A}_1$ - $\mathbf{A}_3$ correspond to the clusters, and have non-negligible off-diagonal elements. The overall maximisation in (3) can then be achieved by independently maximising, for each cluster $k = 1, ..., K$, the function

$$\lambda_k(\hat{\boldsymbol{\omega}}_k) = \mathbf{y}^H\mathbf{A}_k(\boldsymbol{\omega}_k)(\mathbf{A}_k(\boldsymbol{\omega}_k)^H\mathbf{A}_k(\boldsymbol{\omega}_k))^{-1}\mathbf{A}_k(\boldsymbol{\omega}_k)^H\mathbf{y}, \quad (7)$$

where $\boldsymbol{\omega}_k = [\omega_j, ..., \omega_{j+L-1}]^T$ contains the frequencies of the L tones in cluster k. Typically many of the 'clusters' will be single isolated tones, so the maximisation (7) associated with the corresponding submatrix of size 1x1 will be achieved by fast single tone estimation [4].

This reduces the number of *parameters* in each minimisation, but computation of $\mathbf{A}_k(\boldsymbol{\omega}_k)^H\mathbf{y}$ in (7) still requires LN multiplications and additions. A substantial further improvement can be obtained by extending the frequency domain approach proposed in [3,4].

2.1 Frequency domain computation

Since the DFT is a linear transform, the ML estimation task can be formulated equivalently in the discrete frequency domain. Specifically, $\lambda(\hat{\boldsymbol{\omega}})$ in (3) can be shown to be equal to

$$\lambda(\hat{\boldsymbol{\omega}}) = (1/N^3)\mathbf{Y}^H\boldsymbol{\Gamma}(\hat{\boldsymbol{\omega}})(\boldsymbol{\Gamma}(\hat{\boldsymbol{\omega}})^H\boldsymbol{\Gamma}(\hat{\boldsymbol{\omega}}))^{-1}\boldsymbol{\Gamma}(\hat{\boldsymbol{\omega}})^H\mathbf{Y} \quad (8)$$

where $\mathbf{Y}$ is the DFT of $\mathbf{y}$ and $\boldsymbol{\Gamma}(\hat{\boldsymbol{\omega}})$ is the column-by-column DFT of $\mathbf{G}(\hat{\boldsymbol{\omega}})$. Similarly, $\lambda_k(\hat{\boldsymbol{\omega}}_k)$ in (7) has a frequency domain equivalent of the form of (8). The i^{th} column of $\boldsymbol{\Gamma}(\hat{\boldsymbol{\omega}})$ in (8), being the DFT, $\mathbf{E}(\omega_i)$, of the cisoid $\mathbf{e}(\omega_i)$, is of the form $D_N(k\omega_0 - \omega_i)$. The majority of the "energy" (sum squared modulus) of $\mathbf{E}(\omega_i)$ is contained in only a few samples centred around the frequency ω_i. We showed in [3,4] that for single tones, the use of only 5 DFT samples gives estimates very close to the true ML estimates; this reduces computation in the ratio $5/N$, which is very significant for large N. The size of window required for multi-tone clusters is discussed below.

Computation of $\mathbf{E}(\omega_i)$ is /em not carried out by computing the DFT of $\mathbf{e}(\omega_i)$, but by the much more efficient direct evaluation of $D_N(k\omega_0 - \omega_i)$ using (4). Other advantages of the frequency domain approach [4] are that it remains near-optimal in non-Gaussian input noise $\mathbf{z}$, and/or coloured noise, for typical large values of N. Pure real (as opposed to complex) signals are handled by a simple extension of the above procedure [4]. Only the parameters of positive frequency tones are estimated, and corresponding negative frequencies are inferred.

2.2 Size of frequency domain window

For multi-tone clusters, the number of terms of $\mathbf{E}(\omega_i)^H\mathbf{Y}$ needed to achieve accurate estimates can be determined by extending the Cramer-Rao bound (CRB) calculation approach outlined in [4]. For example, consider the case of two tones. The solid line in Figure 1 shows the CRB for frequency estimation of one of the tones, normalised to the single-tone CRB for that tone, and plotted against the frequency difference between the two tones, for the worst case relative phase between the two tones (as shown in [2]).

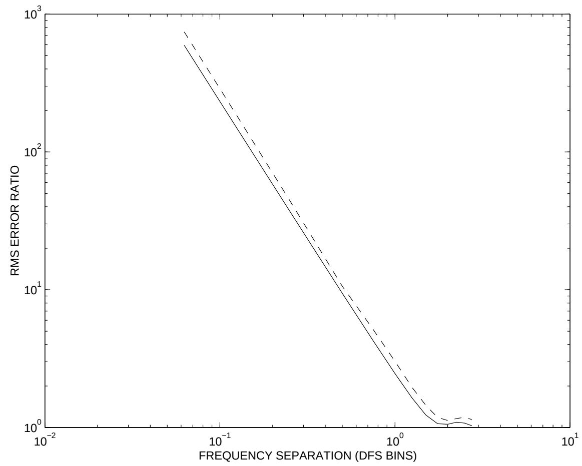

Fig 1. Two tone case. Bound on frequency estimate rms error (i.e. $\sqrt{\text{CRB}}$) normalised to single-tone CRB, plotted against frequency separation in 'bins'. Solid line: CRB of full ML estimator; Dashed line: CRB of frequency domain estimator using only 9 DFS samples.

The dashed line in Fig. 1 shows the CRB of the frequency domain estimator using only 9 terms of $\mathbf{E}(\omega_i)^H\mathbf{Y}$, centred on the frequency of one of the two tones. The estimator variance is increased by only 1.3 dB at a frequency separation of 0.0625 bins, falling to less than 0.75 dB for frequency separations of 2 bins or more. If for example 11 terms are used, these impairments are reduced to 1.06 dB and 0.63 dB respectively.

2.3 Full algorithm

The full algorithm is :

- Compute the DFT $\mathbf{Y}$.
- Repeatedly detect the largest local peak with amplitudes above a detection threshold, and apply single-tone estimation to the new peak (as in [4]).
- Apply a single tone bias estimation heuristic [4] and, for close tones, re-estimate the frequencies by iteration [4].
- Test the residual error over the 5 samples centred on each peak. If this is sufficiently small for all peaks, finish.
- For all peaks with large residual errors, increase the cluster size L by 1 and re-estimate the L frequencies and amplitudes of the cluster.
- If there are other tones or clusters close enough to be affected, re-estimate their frequencies.
- Test the residual errors for each cluster. If they are now all small, finish; otherwise continue to increase the cluster size L (up to a suitable limit) for clusters with large residual errors, and repeat.

Model order estimation is an intrinsic part of this algorithm. The initial estimate of model order (number of tones) is simply the number of detected peaks. This is then increased whenever a cluster size is increased.

3 CONTINUOUS ESTIMATION

The approach described in this paper is being used for musical audio analysis, where typical blocklengths are N=2048 with M = 20-50 tones. In applications such as this a further requirement is to combine estimates from sequential (perhaps overlapping) blocks optimally. This requires knowledge of the estimate variance which, for a nearly-ML estimator, is approximately equal to the CRB. However, the CRBs depend strongly on the relative phase of the tones, and only the CRBs for worst case phase were published in [2]. A closed-form expression for the CRBs is desirable. We will consider the two tone case because it is the most commonly occurring, and in any case estimator variance increases rapidly as further close tones are added.

The CRB for frequency estimation of tone i can be approximated by three asymptotes. The first is the single-tone CRB,

$$\text{var}\left(\frac{\omega_i}{\omega_0}\right) \approx \frac{6\sigma^2}{4\pi^2 N a_i^2}. \tag{9}$$

This is an absolute lower bound. The second is

$$\text{var}\left(\frac{\omega_i}{\omega_0}\right) \approx \frac{6\sigma^2}{4\pi^2 N a_i^2}\frac{2\pi}{(\delta F)^4} \tag{10}$$

where δF is the frequency separation of the tones in bins: $\delta F = (\omega_i - \omega_j)/(2\pi)$. This asymptote meets the single tone bound at $\delta F \approx 1.6$ bins. The third asymptote depends on relative phase. Define $\Delta\Phi = \phi_i - \phi_j + \pi\delta F(N-1)/N$; this equals the phase difference at the block centre (half way between sample $N/2-1$ and sample $N/2$). The third asymptote is

$$\text{var}\left(\frac{\omega_i}{\omega_0}\right) \approx \frac{6\sigma^2}{4\pi^2 N a_i^2}\frac{0.5\pi}{\sin^2(\Delta\Phi)(\delta F)^2}. \tag{11}$$

Note that this becomes infinite as $\Delta\Phi \to 0$ or π. The complete estimate for the CRB is as follows; it is max[(9), min[(10), (11)]]. Hence max[(9), (10)] is the bound for worst case phase ($\Delta\Phi = 0$ or π), as first shown in [2].

To confirm the above model, Fig. 2 shows the actual CRBs and the above asymptotic fit for two tone estimation, for $\Delta\Phi = 0, \pi/16, \pi/8, \pi/4, \pi/2$, as functions of frequency.

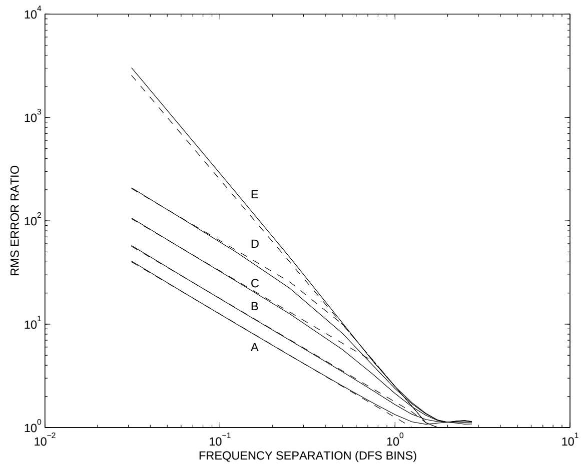

Fig 2. Two tone case. Bound on frequency estimate rms error (i.e. $\sqrt{\mathrm{CRB}}$) normalised to single-tone CRB, plotted against frequency separation in 'bins'. A: $\Delta\Phi = \pi/2$; B: $\Delta\Phi = \pi/4$; C: $\Delta\Phi = \pi/8$; D: $\Delta\Phi = \pi/16$; E: $\Delta\Phi = 0$.

This closed-form expression makes it possible to combine the estimates from successive blocks with the appropriate weighting to reflect the (potentially very different) variances of the estimates from the different blocks.

4 CONCLUSIONS

The frequency domain approach described in section 2 achieves high resolution estimation of sinusoids in white or coloured noise, with performance very close to ML. It is computationally efficient, particularly for problems such as audio analysis where there may be many tones, many of them resolved.

The CRB model described in section 3 illustrates the nature of the two-tone CRB more fully than in [2], and permits fast approximate calculation of the two-tone CRB. This is of value in continuous estimation of frequencies from successive blocks.

REFERENCES

1. Kay, S.M., *Modern Spectral Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
2. Rife, D.C., Boorstyn, R.R., "Multiple Tone Parameter Estimation from Discrete-Time Observations", *BSTJ*, Vol 55, No 9, Nov 1976, pp.1389 - 1410.
3. Macleod, M. D., "Fast high accuracy estimation of multiple cisoids in noise", *Signal Processing V*, (Proc. Eusipco 90), Elsevier, 1990, pp. 333-336.
4. Macleod, M. D., "Fast Nearly-ML Estimation of the Parameters of Real or Complex Single Tones or Resolved Multiple Tones",*IEEE Trans SP*, 1998, Vol.46, No. 1, pp.141-148 (Jan 98).