FAST STATISTICALLY EFFICIENT ALGORITHMS FOR SINGLE FREQUENCY ESTIMATION

Jeffrey D. Klein

Mindspeed Inc.
4000 MacArthur Blvd.
Newport Beach, CA 92660

ABSTRACT

In this article three new estimators of the frequency of a single complex sinusoid are presented. The "rotate-add-decimate" (RAD) method of Crozier is first modified to more closely approach the Cramer-Rao Bound, with the same computation. A second estimator almost achieves the CRB above an SNR threshold approximately 1dB above that of RAD. It can be shown to achieve the CRB for high SNR using $\log_2 N$ arctangents and $2N$ MACs. A third method matches the SNR threshold of RAD and achieves the CRB at high SNR with $\log_2 N$ arctangents and $3N$ MACs.

1. INTRODUCTION

Estimation of the frequency of a single complex sinusoid arises in many contexts, among them carrier acquisition in communications systems, radar and other signal processing disciplines. The data may be modeled by

$$x_n = Ae^{j(\omega n+\phi)} + \epsilon_n, \quad n = 0, ..., N-1 \tag{1}$$

where ω, the angular frequency to estimated, satisfies $-\pi \leq \omega < \pi$, A and ϕ are unknown constants, and the complex white Gaussian noise ϵ_n has variance σ^2. Hence the signal-to-noise-ratio (SNR) Λ is $\frac{A^2}{\sigma^2}$.

The Cramer-Rao bound (CRB) for the estimation of ω is given in [1]. This is the lowest mean-square-error (MSE) that can be achieved by any unbiased estimator. Hence the CRB is the standard to which any single-frequency estimator (SFE) is compared.

The most theoretically appealing SFE is the maximum likelihood estimator (MLE). It is given by the frequency of the peak magnitude of the discrete Fourier transform (DFT). The estimator is unbiased and yields an MSE equal to the CRB for all SNR above an "SNR threshold" that depends on the number of data points N [1]. Thus it is efficient in the statistical sense [2] for sufficiently high SNR. In addition, its performance is uniform over frequency.

However, the MLE's high computational cost has led to a search for alternative methods that approach its performance, but with less computation. The signal processing literature on these techniques is extensive (e.g. see [3] and the introduction in [4]). All such methods exhibit a threshold effect like the MLE, but at a higher SNR. They may or may not achieve the CRB at SNR's above threshold. Also, most have frequency-dependent performance.

The important metrics of SFE methods are computational complexity, SNR threshold and how closely the estimator approaches the CRB near the threshold (and to a lesser extent in the limit as SNR approaches infinity). It is desirable to minimize all of these, but tradeoffs between them are inevitable. The choice of method will depend upon the particular application.

In this paper three new methods are introduced which achieve low values for all three of these metrics that perform well at all frequencies. Two of them are of particular interest as they have low SNR thresholds and achieve the CRB for moderate SNR. They both require $\log_2 N$ arctangents. In addition, $2N$ or $3N$ complex multiply-add operations are required. The latter method has as low an SNR threshold as any SFE method except the MLE.

Regarding notation, $\vec{x}$ denotes a column vector x and $\vec{x}'$ is its transpose.

E-mail: jeffrey.klein@mindspeed.com.

2. THE RAD METHOD

As the methods introduced in this paper have the most in common with the RAD method of [5], it will be discussed in some detail in preparation for presentation of new results based upon it.

The RAD method makes an initial frequency estimate, ω_0, using the LP method [6] with lag = 1. The RAD iteration is as follows, starting with m = 1. An auxiliary vector $\vec{v}_0$ is initialized to contain the data $\{x_n\}$. At the mth iteration, a new correction estimation vector $\vec{v}_m$ is generated by 1) filtering $\vec{v}_{m-1}$ with a 2-point bandpass filter centered at the $(m-1)$th correction frequency ω_{m-1} (determined by application of LP to $\vec{v}_{m-1}$) and 2) decimating the result by 2. The effect is to frequency-shift the filter at each stage. The LP method is applied to the resulting $\frac{N}{2^m}$-long modified vector $\vec{v}_m$ to generate an mth-order correction to the $(m-1)$th-order frequency estimate ω_{m-1}, yielding ω_m to be used in the next iteration. $\vec{v}_m$ is referred to as the "mth correction estimation vector".

The RAD method as presented in [5] was specifically for the case $N = 3*2^J$. This was done to approach the CRB as closely as possible at high SNR, since it naturally leads to an estimate with $l = \frac{2N}{3}$, the lag value which was shown in [7] to minimize the error variance.

However, this requirement is unnecessarily restrictive for what follows. RAD may be modified to handle other input lengths.

The improved RAD algorithm (IRAD) is given below for $N = K \cdot 2^J$, where $K \leq 4$:
step 0:

$$v_n^{(0)} = x_n, \quad n = 0, ..., N-1 \tag{2}$$

with $m = 0$.

step 1:

$$\theta_m = \arg\left[\sum_{n=0}^{N/2^m-2} v_{n+1}^{(m)} v_n^{(m)*}\right] \quad (3)$$

step 2: if $\frac{N}{2^m} \leq 4$, go to END. Else
step 3 ("reduction"):

$$v_n^{(m+1)} = v_{2n}^{(m)} + e^{-j\theta_m} v_{2n+1}^{(m)}, \quad n = 0, ..., N/2^{m+1} - 1 \quad (4)$$

set $m = m + 1$ and go to step 1.
END: Thus after J reduction steps the final IRAD frequency estimate is given by

$$\hat{\omega}_{\mathrm{I}} = \vec{c_{\mathrm{I}}}'\vec{\Delta} \quad (5)$$

where

$$\vec{c_{\mathrm{I}}} = \left[1 \;\; \frac{1}{2} \;\; \frac{1}{4} \;\; \cdots \;\; \frac{1}{2^J}\right]' \quad (6)$$

and

$$\vec{\Delta} = \mathrm{princ}\left((I-2Z)\vec{\theta}\right) = \begin{bmatrix} \theta_0 \\ \mathrm{princ}(\theta_1 - 2\theta_0) \\ \vdots \\ \mathrm{princ}(\theta_J - 2\theta_{J-1}) \end{bmatrix} \quad (7)$$

Here $\mathrm{princ}(x) = (x+\pi)_{\mathrm{mod}2\pi} - \pi$ denotes the principal value of x. I is the identity matrix and Z is the lower shift matrix. The use of $\vec{\Delta}$ instead of $\vec{\theta}$ eliminates the need for any phase unwrapping.

Note that θ_m is an estimate of $\mathrm{princ}(2^m\omega)$, and they are combined in (5) to obtain the final ω estimate. Each reduction (4) decreases the number of data points by a factor of 2, but increases the pointwise SNR of $v^{(m+1)}$ by roughly 3dB over that of $v^{(m)}$, compensating for the halving of the number of points with which to estimate θ_{m+1}. Thus each θ_m will have approximately the same MSE, at least for sufficiently high SNR.

IRAD does not achieve the CRB for any SNR. For $N = 3*2^J$, it approaches it within 0.51dB at high SNR. For $N = 2^M$, it comes within 0.74dB.

As noted in [5], this deficiency may be circumvented by applying an MLE-based fine search to the v subsequence defined in (4) which contains between 8 and 16 points. Three DFT values centered on the final frequency estimate of (5) are first computed using v. The refined frequency estimate is determined from the peak location of a quadratic fit to these values. These additional steps have little effect on the total computation for large N because they operate on the short subsequence v, not on the original data. The IRAD method including this modification will be denoted IRADF in what follows.

Using the ideas of [8] it is possible to derive new versions of the IRAD algorithm with better performance. This is done in the next two sections for the case $N = 2^M$.

3. IMPROVING THE EFFICIENCY OF THE IRAD ALGORITHM

The improvement to IRAD proposed here uses a weighting in (5) that is different from (6). To derive this for $N = 2^M$, note first that the IRAD method estimates the quantities $\{\theta_m\}$ such that

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{M-1} \end{bmatrix} = \mathrm{princ}\left(\begin{bmatrix} 1 \\ 2 \\ \vdots \\ 2^{M-1} \end{bmatrix}\omega + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_{M-1} \end{bmatrix}\right) \quad (8)$$

or

$$\vec{\theta} = \mathrm{princ}(\vec{k}\omega + \vec{\epsilon}) \approx \mathrm{princ}(\vec{k}\omega) + \vec{\epsilon} \quad (9)$$

where $\vec{\epsilon}$ is a real, approximately Gaussian [8] random vector with covariance R_θ.

From (7), we see that

$$\vec{\Delta} = \vec{e_1}\omega + (I - 2Z)\vec{\epsilon} \quad (10)$$

where $\vec{e_m}$ is the mth unit M-vector.

The MLE of ω based on this data model [1] is found by minimizing

$$Q = (\vec{\Delta} - \omega\vec{e_1})' R_\Delta^{-1} (\vec{\Delta} - \omega\vec{e_1}) \quad (11)$$

where R_Δ is the covariance matrix of $(I - 2Z)\vec{\epsilon}$,

$$R_\Delta = (I - 2Z) R_\theta (I - 2Z)' \quad (12)$$

The solution is given by ([2], chapter 4)

$$\hat{\omega} = \frac{\vec{e_1}' R_\Delta^{-1} \vec{\Delta}}{\vec{e_1}' R_\Delta^{-1} \vec{e_1}} = \frac{\vec{k}' R_\theta^{-1} (I-2Z)^{-1}\vec{\Delta}}{\vec{k}' R_\theta^{-1}\vec{k}} \quad (13)$$

since $\vec{k} = (I - 2Z)^{-1}\vec{e_1} = \sum_{m=0}^{M-1} 2^m \vec{e_m}$.

The variance of this new "weighted IRAD" (WIRAD) estimator is given by

$$\mathrm{var}(\hat{\omega}) = \frac{1}{\vec{k}' R_\theta^{-1}\vec{k}} \quad (14)$$

and the weights are given by

$$\vec{c_{\mathrm{W}}}' = \frac{\vec{k}' R_\theta^{-1}(I-2Z)^{-1}}{\vec{k}' R_\theta^{-1}\vec{k}} \quad (15)$$

It remains only to determine R_θ to obtain the MLE of ω based on $\vec{\Delta}$. It is shown in [9] that, for the IRAD method with $N = 2^M$ and sufficiently high SNR, the kl-th element of the covariance of $\vec{\theta}$ is approximately

$$R_\theta^{(kl)} = \frac{2^{\min(k,l)}}{D_{kl}}\left(1 + \frac{B_{kl}}{D_{kl}} + \frac{B_{kk}}{D_{kk}} + \frac{B_{ll}}{D_{ll}}\right) \quad (16)$$

where

$$B_{kl} = 2^{M+1} - 2^{\max(k,l)+1} - 2^{\min(k,l)} \quad (17)$$

$$D_{kl} = \Lambda(N - 2^k)(N - 2^l) \quad (18)$$

and Λ is the SNR, $\frac{A^2}{\sigma^2}$.

Note that using this relation in (13) and (14) gives weights and variance that depend upon SNR. Since SNR is assumed not to be known a priori, the weights are chosen to approach the CRB for high SNR. This has been found to give good performance regardless of the actual SNR.

In this case we have

$$R_\theta^{(kl)} \approx \frac{2^{\min(k,l)}}{\Lambda(N-2^k)(N-2^l)} \quad (19)$$

and $\vec{c}_W$ is independent of Λ. Using (14) for $N = 32$ and 128, we find that the MSE approaches the CRB within 0.15dB for large SNR ($\geq$ 20dB). This is a 0.59dB improvement over the unweighted IRAD method. This comes at the additional computational cost of one more arctangent, and requires using the weight vector from (15) rather than that of (6). Depending upon the application, this may be good enough to use WIRAD instead of the more complex IRADF.

[1] Note that this is not the true MLE of ω, only an approximation.

4. EFFICIENT CORRELATION-BASED ESTIMATION

It is possible to obtain an estimator of the θ_m which is efficient by using a different θ_m estimation strategy, which is described in the following. The new method will be referred to as ERAD ("efficient RAD").

Define a lower-complexity version of the IRAD $\vec{\theta}$ estimate (3) in which the dot product is between the even and odd points of $\vec{v_m}$, unlike the version in (3):

$$\theta_m = \arg\left[\sum_{n=0}^{N/2^{m+1}-1} v_{2n+1}^{(m)} v_{2n}^{(m)*}\right] \quad (20)$$

It is shown in [9] that, for this choice of θ_m, $R_\theta \approx \frac{2}{N\Lambda} I$ at high enough SNR Λ. It is also shown in [9] that the associated estimator is efficient in this situation. This is a consequence from the lack of correlation between the ERAD θ_m estimates.

The corresponding weight vector $\vec{c_{\mathrm{E}}}$ is then obtained from (15)

$$\vec{c_{\mathrm{E}}}' = \frac{\vec{k}'(I-2Z)^{-1}}{\vec{k}'\vec{k}} \quad (21)$$

It can be shown that the ERAD frequency estimate can be written as

$$\hat{\omega} = \vec{c_{\mathrm{E}}}'\vec{\Delta} = \frac{N^2}{N^2-1}\sum_{m=0}^{M-1} \Delta_m \left(\frac{1}{2^m} - \frac{1}{2^{2M-m}}\right) \quad (22)$$

where Δ_m is the mth element of $\vec{\Delta}$. Thus the weighting can be done using only shifts, adds and one real multiply by $\frac{N^2}{N^2-1}$.

Equations (2), (20), (4) and (22) define the ERAD algorithm. As we shall see from the simulations in the next section, this estimator effectively achieves the CRB for SNR only slightly above the threshold.

The disadvantage of ERAD is that its SNR threshold is slightly higher that that of WIRAD. This is because the WIRAD dot products in (3) are roughly twice as long as those of ERAD in (20). This results in less noise in the WIRAD θ_m estimates. However, it introduces correlation between them, reducing WIRAD efficiency at high SNR.

A modification to ERAD is introduced in [9] that combines the lower SNR threshold of WIRAD with the statistical efficiency of ERAD. The idea is to make this modified ERAD (MERAD) behave like ERAD at medium-to-high SNR and like WIRAD at low SNR. This is done by computing both of the dot products used in the ERAD and WIRAD θ_m estimates at each stage. The one which should give better results is used to compute θ_m used in (4) and (7). See [9] for details.

5. SIMULATIONS

In this section we present simulations for comparison of the new methods with some existing SFE methods. Here we define the SNR threshold as the lowest SNR at which 1) the MSE is within 1 dB of the CRB and 2) the MSE is never more than 1 dB from the CRB at higher SNR.

Figure 1 shows the performance of the new methods in comparison to IRAD and IRADF for N = 32 with f = 0.2Hz. Abscissa values represent the data SNR, Λ, in dB. The ordinate values of these plots are $10\log_{10}$(1/(mean-square estimation error)) for the number of realizations simulated, N_r. These plots will be referred to as "MSE-vs-SNR" plots. Note that the CRB appears as a unity-slope line in this log-log format. N_r is 10^7 in all cases.

As expected, the MLE performs best, with an SNR threshold of -0.15dB, and is within 0.1dB of the CRB for SNR above 2dB. IRAD never approaches the CRB within less than 0.74dB, but approaches within 0.1dB of this asymptote for SNR above 10dB.

IRADF is the best performer of the non-MLE methods, with a threshold of 1.0dB. It is within 0.1dB of the CRB for SNR above 4dB and comes within 0.03dB of the CRB at high SNR, as noted in [5].

Considering the new methods, WIRAD has a threshold of 1.4dB and is within 0.2dB of the CRB for SNR above 12dB. However, it approaches the CRB only within 0.15dB at high SNR, as predicted in section III.

ERAD has an SNR threshold of 1.75dB, 0.35dB higher than WIRAD due to the smaller length of the dot products in (20) relative to (3). However, comes within 0.1dB of the CRB for SNR above 7dB and is efficient at high SNR.

MERAD has an SNR threshold of 1.2dB, 0.2dB below WIRAD. It has almost the same MSE as ERAD above the ERAD SNR threshold, as it was designed to do. However, it doesn't approach the CRB as closely as IRADF until the SNR is above 8dB, at which point it is within 0.05dB of the CRB. Like ERAD, it is efficient at high SNR.

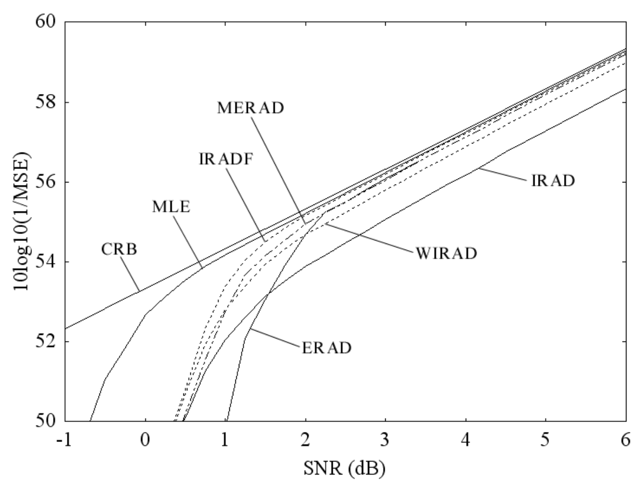

Fig. 1. $10\log_{10}(1/\mathrm{MSE})$ vs. SNR, N= 32, $\frac{\omega}{2\pi}$ = 0.20 Hz

Figure 2 shows the threshold-vs.-frequency performance of the same methods as in figure 1. The observed SNR threshold (as defined above) is plotted versus frequency. The threshold is estimated by linearly interpolating the MSE-vs-SNR curves (sampled every 0.25dB) to estimate the SNR at which the MSE differs from the CRB by 1dB.

The threshold-vs.-frequency results are shown only for positive f since the performance for each algorithm considered here was found to be symmetric about $f = 0$.

Note that the relative performance of these algorithms is roughly constant versus frequency, though the absolute thresholds worsen slightly as f approaches 0.5 for all methods, even the MLE.

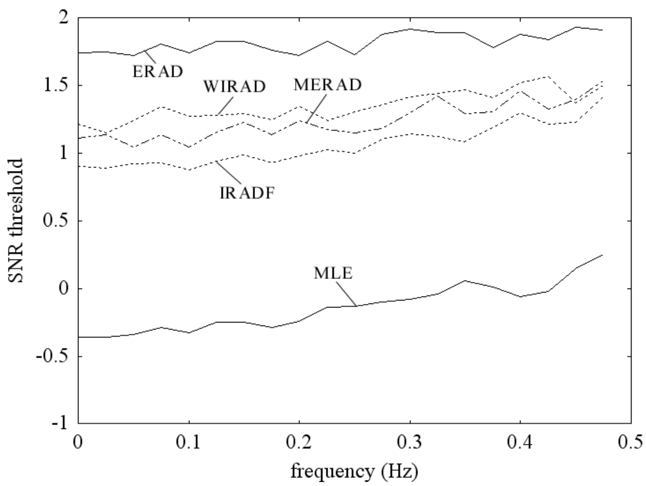

Fig. 2. SNR threshold (dB) vs. frequency, N= 32

6. COMPLEXITY-PERFORMANCE TRADEOFFS

Details of the computational requirements of these methods and others are discussed in [9]. Of those in the figures, only IRADF has a sizeable constant overhead that is not a function of N. This is due to the fine search, which must operate on a v subsequence of length 16 in order to approach the CRB within 0.03dB for high SNR [5]. Using a shorter v for the fine search is possible, but will result in not as close an approach to the CRB, reducing the performance advantage of IRADF over other methods.

However, since the fine search overhead is constant, IRADF is preferable to IRAD and WIRAD when N is large (see [9]), since it is closer to the CRB, has a lower threshold and has essentially the same complexity. In fact, IRADF seems to be the best non-MLE algorithm for large N if the user does not insist upon actually reaching the CRB at high SNR. If efficiency is required, MERAD is most suitable.

The situation is different for smaller N, where the MLE threshold is closer to those of several competing lower-complexity methods. For $N = 32$ the fine search overhead of IRADF is about 40 percent of its complexity. Hence, if a slight decrease in performance is acceptable, MERAD may be used instead of IRADF to eliminate the fine search and save significant computation.

More savings are possible if a slight increase in SNR threshold is permissible. In this case ERAD may be used instead of IRADF or MERAD to save another $\frac{1}{3}$ of the computation, while maintaining statistical efficiency at high SNR.

7. CONCLUSION

In this paper three new versions of the RAD SFE algorithm of [5] have been introduced. After a discussion of an improved RAD (IRAD) method, it is shown how to modify it so that it more closely approaches the CRB for $N = 2^M$. This is done by deriving a weighting of its intermediate frequency estimates. A technique similar to that used in [8] to derive the "phase average" algorithm was used here to obtain a new more efficient phase-based SFE. The new algorithm, WIRAD, has essentially the same SNR threshold as IRAD, similar computational complexity, and only a change of multiplicative constants in (6) is required.

It is also demonstrated with simulations that IRAD can be modified to actually achieve the CRB for SNR sufficiently above the threshold. This is done by using an alternative definition of the intermediate frequency estimates. In [9] the resulting method (ERAD) is shown theoretically to be efficient for high SNR. It reduces computation by 1/3 relative to IRAD. However, ERAD raises the SNR threshold slightly since it uses dot products shorter than IRAD and IRADF, resulting in noisier estimates of the $\{2^m\omega\}$.

Finally, the heuristic MERAD algorithm has been developed (see [9]) that combines the best features of ERAD and IRAD. It is designed to behave like the former at high SNR (i.e. achieve the CRB) but like the latter at low SNR (i.e. to have the lowest possible SNR threshold). This is done for almost the same computational load as IRAD and, unlike IRADF, requires no fine search.

It should be noted that the ERAD and MERAD algorithms, though best suited for $N = 2^M$, can also be derived for $N = K2^J$ using the approach outlined in [10]. These algorithms have low SNR thresholds and are efficient at high SNR as well.

Note that the new algorithms may be recast in on-line forms which process each sample as it arrives. See [5] for details for the $N = 3 * 2^{M-2}$ case.

8. REFERENCES

[1] D. C. Rife and R. R. Boorstyn, "Single-Tone Parameter Estimation from Discrete-Time Observations," *IEEE Trans. on Information Theory*, Vol. IT-20, No. 5, pp. 591-598, September 1974.

[2] S. Kay, *Fundamentals of Statistical Signal Processing, Volume I; Estimation Theory*, Prentice Hall, 1993.

[3] M. Fowler, "Phase-based Frequency Estimation: A Review," *Digital Signal Processing*, Vol. 12, pp. 590-615, 2002.

[4] T. Brown and M. M. Wang, "An Iterative Algorithm for Single-Frequency Estimation," *IEEE Trans. on Signal Processing*, Vol. 50, No. 11, pp. 2671-2682, November 2002.

[5] S. Crozier and K. Gracie, "Low Complexity Frequency Estimation with Close-to-Optimum Performance for QPSK and Offset-QPSK Modulated Signals," *Proc. 9th International Conference on Wireless Communications (Wireless97)*, Calgary, Alberta, Canada, pp. 294-306, July 9-11, 1997.

[6] L. B. Jackson, D. W. Tufts, F. K. Soong and R. M. Rao, "Frequency Estimation by Linear Prediction," *Proceedings of ICASSP-1978*, pp. 352-356, 1978.

[7] G. W. Lank, I. S. Reed and G. E. Pollon, "A Semicoherent Detection and Doppler Estimation Statistic," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-9, pp. 151-165, 1973.

[8] S. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 37, pp. 1987-1990, December 1989.

[9] J. D. Klein, "Fast Algorithms For Single Frequency Estimation," submitted to *IEEE Trans. on Signal Processing*, July 2004.

[10] J. D. Klein, "Recursive Single Frequency Estimation," *Proceedings of ICASSP-2001*, Vol.5, pp. 3077-3080, May 2001.