Optimizing Correlation Based Frequency Estimators

Björn Völcker and Peter Händel
Department of Signals, Sensors and Systems
Royal Institute of Technology, Stockholm, Sweden
bjorn.volcker@s3.kth.se; ph@s3.kth.se

ABSTRACT

As a complement to the periodogram, low-complexity frequency estimators are of interest. One such estimator is based on Prony's method and rely on phase information of the auto-correlations. Both performance and computational complexity are functions of the choice of correlations used in the estimator and often we have a trade off situation. In this paper, frequency estimation from *an arbitrary set* of estimated auto-correlations is studied. We further introduce a design strategy by optimizing a performance criterion given a predetermined computational constraint. We illustrate this by numerical examples.

1 Introduction

Estimation of the parameters of a noise corrupted sinusoidal model is a frequently addressed problem in the signal processing literature. Starting with an observed sample $\{y(0),\ldots,y(N-1)\}$ where N is the number of data points, there exist numerous methods which can be used to estimate the sought parameters. Often, the estimation of the frequencies is of particular interest. It is well known that in most applications, excellent estimates of the sought frequencies are easily obtained by peak-picking the Periodogram of data [1]. However, it is not applicable when the real-time constraints on numerical complexity requires low-complexity methods.

There have been a large amount of papers on low-complexity estimators. Basically, they can be divided into two classes, that is data based and correlation based. In this paper, we concentrate on the single tone case and correlation based frequency estimation. Consider the single tone model

$$y(n) = ae^{\mathrm{i}\omega n} + e(n), \qquad n = 0,\ldots,N-1$$

where $a = |a|e^{\mathrm{i}\phi}$ is a complex-valued amplitude. The noise $\{e(n)\}$ is zero mean complex-valued circular white Gaussian with variance σ^2. For notational brevity and without loss of generality we let $\omega \in [0, 2\pi)$ be the normalized (angular) frequency. The parameters ($|a|$, ϕ, ω, σ^2) are all unknown, but the frequency ω is the parameter of main interest.

For correlation based estimators, an estimate of the frequency is obtained by the information of one or several estimated entries of the auto-correlation sequence of $y(n)$

$$r(m) = \mathrm{E}[y(n)y^*(n-m)] = |a|^2 e^{\mathrm{i}\omega m} + \sigma^2\delta_{m,0} \qquad (1)$$

where $\mathrm{E}[\cdot]$ denotes statistical expectation. Further, $\delta_{m,0}$ is the Kronecker delta and $(^*)$ denotes complex conjugate. From data, we can form the sample correlation sequence $\{\hat{r}(0),\ldots,\hat{r}(N-1)\}$, where $\hat{r}(m)$ is, for example, the unbiased estimator

$$\hat{r}(m) = \frac{1}{N-m}\sum_{n=m}^{N-1} y(n)y^*(n-m). \qquad (2)$$

From (1) it is evident that information about the frequency is gathered in the phase angle of $r(m)$, that is, for $m \neq 0$,

$$m\omega = \angle[r(m)] + 2\pi\ell \qquad (3)$$

for some integer ℓ satisfying $0 \le \ell < m$. Here $\angle[\cdot]$ denotes the phase angle in $[0, 2\pi)$.

For $m = 1$ it is evident from (3) that $\ell = 0$, and we can form an estimate of the frequency as $\hat{\omega} = \angle[\hat{r}(1)]$. Lank, Reed and Pollon showed in [2] that the performance of the linear prediction estimator can be increased by using a different correlation lag $m = L \neq 1$. A disadvantage with using $m = L > 1$ is the introduction of ambiguities to the frequency estimates. If this ambiguity cannot be resolved, the frequency range is reduced to $|\omega| < \pi/L$. To resolve this ambiguity a sequence of correlations is required, e.g. the full sequence $r(1),\ldots,r(N-1)$. This is not computationally efficient though and we have to rely on a much smaller set of $K \ll N$ correlations. Clearly, one can use the truncated sequence $r(1),\ldots,r(K)$ and fit the unwrapped phase to a straight line [3].

Both the threshold level and the statistical efficiency at high SNR, for this class of methods, vary with the number of correlations, the choice of correlations, weighting of the phases as well as the weighting of the sample correlation calculations.

2 Frequency Estimation from Arbitrary Sets of Correlations

Starting with the problem of estimating the frequency from phase information of K correlations according to (3), a system of K equations and $K+1$ unknowns $(\omega, \ell_1, \ldots, \ell_K)$ follows

$$\mathbf{L}\omega = \boldsymbol{\varphi} + 2\pi\boldsymbol{\ell}$$

where, $\mathbf{L} = [L_1 \dots L_K]^{\mathrm{T}}$, $\boldsymbol{\varphi} = [\varphi_1 \dots \varphi_K]^{\mathrm{T}}$ and $\boldsymbol{\ell} = [\ell_1 \dots \ell_K]^{\mathrm{T}}$. Further, $\varphi_k = \angle[r(L_k)]$ and $\{L_k\}$ is the set of correlation lags (strictly positive integers). Note that ℓ_k is a non-negative integer less than L_k, hence equation k in (2) gives L_k possible values of ω. In an ideal case (no noise) only one ω satisfies all the K equations if there is no common divisor among $\{L_k\}$. With noisy measurements we can solve for ω in a least squares sense for every combination of $\{\ell_k\}$ and pick the best one. The number of combinations is $\bar{L} = \prod_{k=1}^{K} L_k$, but only a subset of these are feasible in practice and the search is reduced.

With the phase estimates $\hat{\boldsymbol{\varphi}} = [\angle[\hat{r}(L_1)] \dots \angle[\hat{r}(L_K)]]^{\mathrm{T}}$ a weighted least squares (WLS) problem can be formulated [4] and given $\boldsymbol{\ell}$ and a constant weight $\mathbf{W}$, a closed form expression for $\hat{\omega}$ yields

$$\hat{\omega}(\boldsymbol{\ell}) = \frac{\mathbf{L}^{\mathrm{T}}\mathbf{W}(\hat{\boldsymbol{\varphi}} + 2\pi\boldsymbol{\ell})}{\mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{L}}. \tag{4}$$

This render the concentrated WLS problem

$$\hat{\boldsymbol{\ell}} = \arg\min_{\boldsymbol{\ell} \in \mathcal{L}_\ell} (\hat{\boldsymbol{\varphi}} + 2\pi\boldsymbol{\ell})^{\mathrm{T}} \boldsymbol{\Pi}_{\mathbf{L}}^{\perp} (\hat{\boldsymbol{\varphi}} + 2\pi\boldsymbol{\ell}) \tag{5}$$

where $\boldsymbol{\Pi}_{\mathbf{L}}^{\perp} = \mathbf{W} - \frac{\mathbf{W}\mathbf{L}\mathbf{L}^{\mathrm{T}}\mathbf{W}}{\mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{L}}$ and $\mathcal{L}_\ell$ is the set of feasible combinations. The original frequency estimation problem is now separated into two subproblems. First, phase unwrapping, i.e., determining the unknown set $\{\ell_k\}$ in (5). Second, frequency estimation from the unwrapped phase, i.e., (4). Despite the joint nature of the problem, phase unwrapping and frequency estimation are often treated separately in the literature. In a high SNR scenario though, the probability of an incorrect phase unwrapping is negligible, which justifies frequency estimation from an unwrapped phase.

The phase unwrapping in (5) is identifiable if there is no common divisor among the entries in $\mathbf{L}$. If $\{L_k\}$ are all relatively prime, the complexity can be reduced. In [4] an alternative approach to the phase unwrapping in (5) is introduced. That method is computationally more efficient, but yields the same performance. We assume that $\boldsymbol{\ell}$ is known or correctly estimated, and consider the frequency estimation problem given a set of K sample correlations.

2.1 Frequency Estimator

Let $\hat{r}(L_1), \dots, \hat{r}(L_K)$ (ordered such that $L_1 < \dots < L_K$) be K sample correlations. Any frequency estimator based on phase information can be formed as a weighted average of the unwrapped phase, i.e.,

$$\hat{\omega}_{\boldsymbol{\alpha}}(\mathbf{L}) = \boldsymbol{\alpha}^{\mathrm{T}}(\hat{\boldsymbol{\varphi}} + 2\pi\boldsymbol{\ell}) \tag{6}$$

where $\boldsymbol{\alpha}$ is a weighting vector with $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{L} = 1$ for unbiased estimates. For instance, the estimator in (4) is a special case of (6). For clarification we explicitly state that the estimate is a function of $\mathbf{L}$.

Let $\mathbf{R}$ be the covariance matrix of $\hat{\boldsymbol{\varphi}}$. Then the variance of the weighted estimator $\hat{\omega}_{\boldsymbol{\alpha}}$, as given in (6), is $\mathrm{var}[\hat{\omega}_{\boldsymbol{\alpha}}] = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\boldsymbol{\alpha}$. With use of the Gauss-Markov Theorem the optimal (minimal variance) weighting scheme, for a given SNR is

$$\boldsymbol{\alpha}_{\mathrm{opt}}(\mathbf{L}) = \frac{\mathbf{R}^{-1}\mathbf{L}}{\mathbf{L}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{L}}$$

with the corresponding estimator $\hat{\omega}_{\mathrm{opt}}(\mathbf{L}) = \boldsymbol{\alpha}_{\mathrm{opt}}^{\mathrm{T}}(\mathbf{L})(\hat{\boldsymbol{\varphi}} + 2\pi\boldsymbol{\ell})$. Note that this coincides with the WLSE when the weighting matrix is $\mathbf{W} = \mathbf{R}^{-1}$. The variance of this estimator is

$$\mathrm{var}[\hat{\omega}_{\mathrm{opt}}(\mathbf{L})] = \frac{1}{\mathbf{L}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{L}} \tag{7}$$

and may serve as a lower bound on the performance of this class of frequency estimators, given the correlation lags $\mathbf{L}$. This bound is tighter than the CRB given by [1]. An asymptotic expression for $\mathbf{R}$ as $\mathrm{SNR} \to \infty$ is explicitly given by [4]

$$[\mathbf{R}]_{k,l} = \frac{1}{\mathrm{SNR}} \left(\frac{\min(L_k, N - L_l)}{(N - L_k)(N - L_l)} + \frac{\delta_{k,l}}{2\,\mathrm{SNR}(N - L_k)} \right) \tag{8}$$

where $l \geq k$. Hence, we have an explicit expression of the asymptotic (as $\mathrm{SNR} \to \infty$) performance $\lim_{\mathrm{SNR}\to\infty} \mathrm{SNR} \cdot \mathrm{var}[\hat{\omega}_{\mathrm{opt}}]$, that is a function of $\mathbf{L}$. Recall that this implies the use of a suboptimal weighting

$$\boldsymbol{\alpha}_{\infty}(\mathbf{L}) = \lim_{\mathrm{SNR}\to\infty} \boldsymbol{\alpha}_{\mathrm{opt}}(\mathbf{L})$$

which is feasible to calculate for any $\mathbf{L}$.

It is well known that frequency estimation suffers from threshold effects. Algorithms that rely on phase data have a higher threshold due to their use of phase unwrapping. An incorrect phase unwrapping gives a dramatic error in the frequency estimate, which is the main contribution to the threshold effect.

As the SNR decreases a correct analysis must incorporate both the probability of an incorrect phase unwrapping as well as a variance expression. For SNR values below the threshold the phases tend to be uncorrelated and uniformly distributed over $[0, 2\pi)$ and the frequency cannot be determined. In this case the weighting has no effect on the performance. Thus, if the frequency estimator is to operate in a low SNR environment one has to choose a correlation lag constellation $\mathbf{L}$ that gives a low SNR threshold. This is in general achieved for small L_k [4].

3 Design Strategy

From a designing point of view we seek the best estimator given a computational complexity. The aim is to minimize the variance of the frequency estimator in (6). We consider the high SNR case. To compensate for the threshold effect we put the additional term $||\mathbf{L}||^2$ to the variance. This gives the cost function

$$V(\mathbf{L}) = (1 - \kappa) \left(\lim_{\mathrm{SNR}\to\infty} \mathrm{SNR} \cdot \mathrm{var}[\hat{\omega}_{\mathrm{opt}}] \right) + \kappa ||\mathbf{L}||^2 \tag{9}$$

where κ $(0 \leq \kappa \leq 1)$ is a trade-off parameter between the asymptotic variance and the threshold level. The asymptotic variance is a function of $\mathbf{L}$ and is given by (7) together with (8).

Empirical studies show that the threshold is not a convex function of $\mathbf{L}$ and a relaxation of the problem is not straightforward. We know that the probability of an incorrect phase unwrapping (the reason for the threshold behavior) increases with increasing values in $\mathbf{L}$ [4]. Therefore, $||\mathbf{L}||^2$ is a good approximation and it is also convex in $\mathbf{L}$.

Recall from (8) that the covariance matrix includes the min–function, which in this case causes the cost function (9) to be non convex, but convex within subregions (defined by the possible outcomes of this min–function) [4]. For each k there are three inequalities, defining the subregion

$$L_k + 1 \leq L_{k+1}$$

and

$$\begin{cases} L_k + L_{k+1} \leq N \\ \quad\text{or} \\ L_k + L_{k+1} \geq N \end{cases} \qquad \begin{cases} 1 \leq L_k \leq \frac{N}{2} \\ \quad\text{or} \\ \frac{N}{2} \leq L_k \leq N-1 \end{cases} .$$

There are only $2K$ feasible regions, but some redundancy is included for a clearer presentation. Every region q is a polyhedra which can be written on the form $\mathbf{A}(q)\mathbf{L} \leq \mathbf{c}(q)$, where $\mathbf{A}(q)$ and $\mathbf{c}(q)$ are constructed from the inequalities above.

The number of computations is mainly determined by the calculation of the sample correlations. To calculate $\hat{r}(L_k)$ in (2) it requires $4(N - L_k) + 1$ real valued multiplications and $4(N - L_k) - 1$ real valued additions. Add to this the number of computations needed in the phase unwrapping [4] followed by the weighted average in (6)

$$\begin{aligned} c_{\text{add}} &= 4NK - 4\mathbf{1}^{\mathrm{T}}\mathbf{L} + K - 1 \\ c_{\text{mult}} &= 4NK - 4\mathbf{1}^{\mathrm{T}}\mathbf{L} + 4K \end{aligned}$$

where $\mathbf{1} = [1 \ \dots \ 1]^{\mathrm{T}}$. In addition, $c_{\text{phase}} = K$ phase calculations are performed. The number of computations must be kept below a given maximum $c_{\max}$. This can be treated separated in three different constraints, for additions, multiplications and phase calculations respectively, or weighted together into a total number of computations as

$$c_{\text{tot}} = t_{\text{a}} c_{\text{add}} + t_{\text{m}} c_{\text{mult}} + t_{\text{p}} c_{\text{phase}} \leq c_{\max}. \tag{10}$$

This, as well as the feasible region, is an affine constraint. Hence, we have the following set of $2K$ convex optimization problems

$$\begin{aligned} \min_{\substack{q=1,\dots,2K \\ \mathbf{L}}} \quad & V(\mathbf{L}) \\ \text{subject to} \quad & \mathbf{A}(q)\mathbf{L} \leq \mathbf{c}(q) \\ & c_{\text{tot}} \leq c_{\max} \end{aligned} \tag{11}$$

where the first constraint corresponds to the feasible region q and the second to the numerical complexity. The cost function $V(\mathbf{L})$ is given in (9). The maximum number of computations $c_{\max}$ is chosen by the designer. The complexity constraint is easily separated into, for example, number of multiplications, additions and phase calculations as mentioned above.

Strictly, an optimization with respect to $\mathbf{L}$ is subject to the condition that $\mathbf{L}$ is an integer vector. In case of non-integer values, we choose the closest ones. If N is large this quantization effect is negligible.

What is required for a feasible solution to exist? First, note that the phase unwrapping requires at least $K = 2$. Second, the minimum possible number of computations is attained for $\mathbf{L} = [N-2, N-1]^{\mathrm{T}}$, which gives $c_{\text{tot}} = 37$ if all computational operations are weighted equally. Hence, $c_{\max} \geq 37$.

In Table 1 the design algorithm is summarized.

Table 1: A design strategy subject to a given computational complexity.

given	$0 \leq \kappa \leq 1$, $K = 2$ and $c_{\max} \geq 37$. Set $V = \infty$.
repeat	Solve optimization problem (11).
	if feasible solution, Store the constellation $\mathbf{M}$ and the minimum value $V(\mathbf{M})$ obtained from solving (11). *if* $V(\mathbf{M}) < V$, $\mathbf{L} := \mathbf{M}$ $V := V(\mathbf{M})$ $K := K + 1$ *else* $K := N$ (to exit the repeat loop)
until	$2K + 1 \geq N$

4 Design Examples

All simulations are evaluated over 5000 trials and the frequency is, from trial to trial, drawn from a uniform distribution in the interval $[\pi/4, 3\pi/4]$. The number of snapshots is $N = 33$ and we assume additions and multiplications are equally efficiently implemented ($t_{\text{a}} = t_{\text{m}} = 1$), but phase calculations are six times as expensive ($t_{\text{p}} = 6$).

Example 1. Compared to the WLSE in (4)–(5) we can save computations by using the phase unwrapping in [4]. For the WLSE we set the following parameters: $\mathbf{W} = \mathbf{R}^{-1}, K = 2$. The correlation lag constellation $\mathbf{L} = [7, \ 13]^{\mathrm{T}}$ is used, which follows the suboptimal strategy proposed in [4]. We choose the number of computations ($c_{\max} = 526$) required by the WLSE as the constraint in our design algorithm. We put a threshold penalty of $\kappa = 0.2$. The design algorithm returns (with lags rounded to closest integer) $\mathbf{L} = [5, \ 12]^{\mathrm{T}}$. The performance is compared in Figure 1. By using the low complexity phase unwrapping we can allocate computational resources to smaller values in $\mathbf{L}$ without losing in performance.

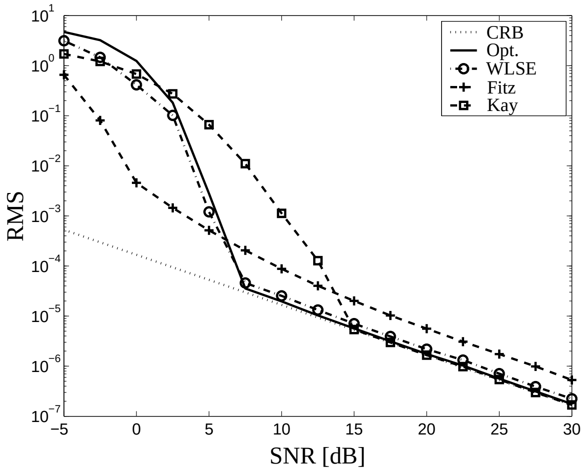

Figure 1: Design examples 1 and 2. The WLSE $(\circ)$, the Fitz $(+)$ estimator and Kay's estimator $(\square)$ are used as references for computational complexity. The performances of the estimators proposed by the design algorithm are the same in both examples, and is shown with a solid line.

Example 2. The same setup as in example 1, but with the Fitz estimator with $K = 2$ as a computational complexity reference ($c_{\max} = 519$). The design algorithm returns $\mathbf{L} = [5,\ 12]^{\mathrm{T}}$. Note that we obtain the same correlation lag constellation as in the previous example, and therefore plot the performance curves in Figure 1 as well. For high SNR we have lowered the variance, but we have to pay in a higher threshold.

Example 3. Kay's frequency estimator [5] is a well known low complexity estimator. Computationally it requires less additions and multiplications, but more phase calculations, an operation that in general is more expensive. With the more expensive phases the maximum number of operations yield ($c_{\max} = 447$). Together with $\kappa = 0.2$ the design algorithm returns $\mathbf{L} = [5,\ 12]^{\mathrm{T}}$. In Figure 1 it is seen that we have lowered the threshold level to a small cost in asymptotic performance.

Next we investigate how sensitive the trade-off between a low threshold level and a small asymptotic variance is to variations in κ. The performance curves for different κ are shown in Figure 2 and it is seen that the threshold level decreases with increasing κ, as expected. The trade-off is not too sensitive to variations in κ and we may conclude that the norm of $\mathbf{L}$ is a good representation of the threshold.

5 Conclusions

We have analyzed a weighted average frequency estimator based on phase information of the sample correlations. This

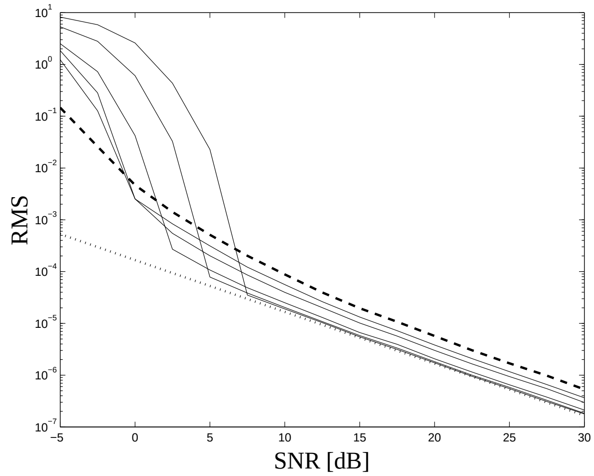

Figure 2: The performance curves for the proposed estimator (solid lines) that correspond to $\kappa = [0.04,\ 0.46,\ 0.77,\ 0.92,\ 0.95]$ starting with the rightmost one. In addition the CRB (dotted line) and the Fitz estimator (dashed line) are included.

estimator includes many of the ones in this class. An optimal weighting scheme was derived.

We have taken a design approach and formulated a convex optimization problem to solve for the best constellation. The objective is to minimize the asymptotic variance subject to a given computational complexity. A punish can be put to the incorrect phase unwrapping through a design parameter κ in the objective function and thereby keep a low SNR threshold. The algorithm can outperform most correlation based frequency estimators and we gave three design examples to illustrate this.

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, 1974.

[2] 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.

[3] Michael P. Fitz, "Further results in the fast estimation of a single frequency," *IEEE Trans. on Communications*, vol. 42, no. 2/3/4, pp. 862–864, February/March/April 1994.

[4] Björn Völcker and Peter Händel, "Frequency estimation from proper sets of correlations," *IEEE Trans. on Signal Processing*, 2001, To appear.

[5] Steven M. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 37, no. 12, pp. 1987–1990, 1989.