QUASI-ML HOP PERIOD ESTIMATION FROM INCOMPLETE DATA

N.D. Sidiropoulos[*]

Dept. of ECE, Univ. of Minnesota
Minneapolis, MN 55455 USA
& Tech. Univ. of Crete, 73100 Greece
nikos@ece.umn.edu

A. Swami, B.M. Sadler

Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197
(aswami,bsadler)@arl.army.mil

ABSTRACT

Given a noisy sequence of (possibly shifted) integer multiples of a certain period, it is often of interest to estimate the period (and offset). With known integer regressors, the problem is classical linear regression. In many applications, however, the actual regressors are unknown; only categorical information (i.e., the regressors are integers) and perhaps loose bounds are available. Examples include hop timing estimation, Pulse Repetition Interval (PRI) analysis, and passive rotating-beam radio scanning. With unknown regressors, this seemingly simple problem exhibits many surprising twists. Even for small sample sizes, a Quasi-Maximum Likelihood approach proposed herein essentially meets the clairvoyant CRB at moderately high SNR - the latter assumes knowledge of the unknown regressors. This is quite unusual, and it holds despite the fact that our algorithm ignores noise color. We outline analogies and differences between our problem and classical linear regression and harmonic retrieval, and corroborate our findings with careful simulations.

Keywords: Period estimation, timing offset, harmonic analysis, synchronization, hop timing, Pulse Repetition Interval (PRI) analysis, deinterleaving

1. INTRODUCTION

Consider the following observation model

$$\tau(n) = \phi + \kappa(n)T + w(n), \quad n = 1, \cdots, N, \qquad (1)$$

where ϕ is an unknown shift, $\kappa(n) \in \mathbb{Z}$ is a generally unknown sequence of ordered integers, T is the unknown period, and $w(n)$ is additive white Gaussian (AWG) noise, with variance σ_w^2. The problem is to estimate ϕ and T from $\{\tau(n)\}_{n=1}^N$. In practice, there are many situations wherein the only information that can be assumed about the regressors is that $\kappa(n) \in \mathbb{Z}$, and perhaps also loose upper and lower bounds on T, or qualitative information of the type "lengthy gaps are rather rare".

1.1. Overview

The model in (1) is reminiscent of two well-known problems. In the special case that $\kappa(n) = n, \; n = 1, \cdots N$, the problem is classical line regression; if the integers $\{\kappa(n)\}_{n=1}^N$ are known, then a standard linear regression problem appears. If the regressors $\{\kappa(n)\}_{n=1}^N$ are unknown integers, then a non-standard regression problem emerges.

On the other hand, the problem in (1) is closely related to harmonic retrieval. That is, raising the data in (1) to the exponent

$$x(n) := e^{j\tau(n)} = e^{jw(n)}e^{j(\phi+\kappa(n)T)}, \quad n = 1, \cdots, N,$$

which is a *harmonic retrieval problem with missing samples in non-Gaussian multiplicative noise*. Note, however, that raising the data to the exponent is not a reversible operation, hence the problems are generally not equivalent.

The classical (single-) harmonic retrieval problem has been thoroughly investigated in the literature, including optimal (periodogram) and suboptimal linear-complexity solutions. The latter achieve near-optimal performance at moderate SNR or moderate samples and above. Interestingly, Tretter [11, 4] has shown that a computationally attractive solution can be obtained by casting the frequency estimation problem as a line regression problem in the phase domain. At high SNR, phase noise can be approximated by additive white Gaussian noise, and the problems become essentially equivalent [11, 4]. Another related approach to the problem of frequency estimation involves working with zero-crossings or higher-order zero-crossings of the observation [5].

The harmonic retrieval problem with missing samples has also been considered [7]. Early approaches were periodogram-based (the periodogram often works reasonably well with mild multiplicative noise) but parametric techniques were also developed [8]. In most cases, a simple Bernoulli miss model is adopted [10, 8], or else it is assumed that missing samples occur periodically with known outage period. Harmonic retrieval in multiplicative noise has been dealt with, (see, e.g., [3]) but, to the best of our knowledge, harmonic retrieval in multiplicative noise and a deterministic unknown model for the missing samples has not been addressed in the literature.

The baseline for the present research is mostly the work of Sadler and Casey [2, 9], who also considered period estimation from the model in (1) with missing observations. Their work is based on modifications of the Euclidean algorithm for the computation of the greatest common divisor. Relative to that baseline, our work offers a quasi-ML algorithm that attains much improved performance (particularly for small sample sizes), plus additional insights into the model and its properties.

A word about applications is in order. Our particular motivating application is hop period and timing offset estimation in the context of Frequency-Hopped (FH) radios. Therein, one may often observe only part of the spread FH bandwidth, because the

[*] This work was supported by the ARL Communications & Networks CTA.

true FH band may be unknown or non-contiguous; and also for noise considerations (opening up the receiver bandwidth includes more noise energy). Another application with missing observations is passive rotating-beam radio scanning. In this case observations (period multiples) are periodically missing only if the scan period is harmonically related to the sought period. In addition, the correct scan period may be unknown. Another situation wherein deterministic unknown modeling of missed observations may be appropriate could be a multi-tasked best-effort surveillance processor.

2. ALGORITHM

Given estimates of T and $\{\kappa(n)\}_{n=1}^{N}$, the estimation of the shift ϕ is easy. As a first step towards simplifying the problem, we may take pairwise differences. We can take up to $\binom{N}{2}$ such differences; this produces many more data points, at the expense of coloring the noise sequence, which is analogous to smoothing for line spectrum estimation. Here, we begin with simple adjacent-sample differences of non-overlapping pairs of samples. This yields a model that is independent of the shift ϕ, at the cost of halving the available sample size and a 3dB loss in terms of noise amplification (note that in this case the noise is still white). This yields

$$t(n) = k(n)T + v(n), \quad n = 1, \cdots, M := \lfloor \frac{N}{2} \rfloor,$$

where $t(n) := \tau(2n) - \tau(2n-1), k(n) := \kappa(2n) - \kappa(2n-1) \in \mathbb{Z}, v(n) := w(2n) - w(2n-1)$. In vector form and with obvious notation,

$$\mathbf{t} = \mathbf{k}T + \mathbf{v}.$$

If the only assumption on $\{k(n)\}_{n=1}^{M}$ is that $k(n) \in \mathbb{Z}$, and the noise *after taking differences* is AWG, then the maximum-likelihood (ML) principle yields the following least-squares (LS) problem

$$\min_{T \in \mathbb{R}_+,\, \mathbf{k} \in \mathbb{Z}^M} ||\mathbf{t} - \mathbf{k}T||_2^2.$$

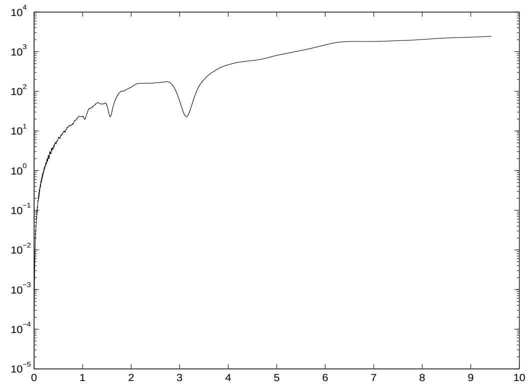
This problem is linearly separable, and the cost function can be concentrated with respect to $\mathbf{k}$; this yields

$$\min_{T \in \mathbb{R}_+} ||\mathbf{t} - T\,\mathrm{round}\,(\mathbf{t}/T)||_2^2.$$

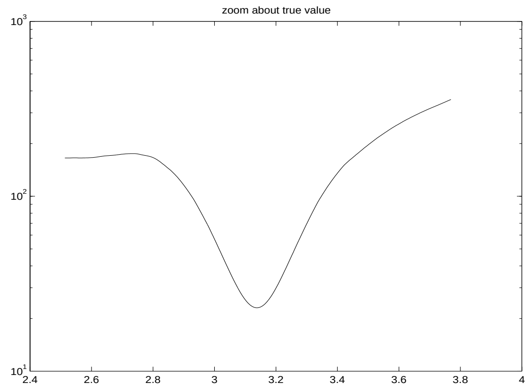
This latter minimization can be accomplished via simple line search on T. However, Fig. 1, which depicts a typical plot of the concentrated cost function, immediately points to two pitfalls:

- In the noiseless case, if a certain $\hat{T}$ is a zero-cost solution, so is $\hat{T}$ divided by an arbitrary integer. This is because one may counter-scale the $k(n)$ sequence. This yields infinitely many solutions to the problem in the noiseless case.
- In the noisy case, smaller $\hat{T}$ corresponding to T divided by a large integer yield finer granularity, which in turn allows expressing *almost anything* as a multiple of small enough $\hat{T}$: in the noisy case, these rescaled minima of the cost function are increasingly deep as one moves closer to zero.

These issues arise because the problem is not well-posed. The ambiguity is analogous to aliasing in the context of frequency estimation, or scale ambiguity in blind system identification. In order to have a well-posed problem, we need to impose a lower bound



(a) Cost(T).



(b) Zoom about true T $(= \pi)$.

Fig. 1. Typical plot of concentrated cost(T) at moderate SNR.

on T, just like we need to impose an upper bound on frequency in order to prevent aliasing in the context of frequency estimation. A coarse upper bound is also needed to limit the search, hence the problem becomes

$$\min_{L<T<U} ||\mathbf{t} - T\,\mathrm{round}\,(\mathbf{t}/T)||_2^2, \tag{2}$$

which can be solved by line search over (L, U). We will refer to this as the SLS2 algorithm, which stands for *Separable Least Squares Line Search*.

2.1. Smoothing and Noise Whitening

If overlapping pairs or smoothing is used to extend the available sample size after differencing, then noise whitening is needed in order to maintain ML optimality of LS. Note that if the original additive noise is AWG, then the color of the noise after differencing is known, because it is induced by our processing of the data. Hence we end up with a modified data model after differencing and whitening, which reads

$$\mathbf{t} = \mathbf{W}\bar{\mathbf{k}}T + \ \mathrm{AWGN}.$$

where $\mathbf{W}$ is the whitening matrix. For this model, however, the integer vector parameter $\bar{\mathbf{k}}$ *is no longer* linearly separable, due to the premultiplication by the whitening matrix $\mathbf{W}$; hence the cost function cannot be concentrated with respect to $\bar{\mathbf{k}}$. Although it is possible to adopt an iterative LS approach, in which T is first estimated and then $\bar{\mathbf{k}}$ is updated in a conditional LS fashion using enumeration or more sophisticated "almost optimal" integer LS solvers (such as sphere decoding [1], or semidefinite relaxation [6]), this would be computationally very demanding for the application at hand. The recommended work-around is to simply ignore noise color at this stage, and use plain LS. This will not matter much at high SNR.

3. ESTIMATION OF ϕ AND ITERATIVE LEAST SQUARES

Once T has been estimated, we can go back to the original data and estimate ϕ via another LS line search. Specifically, conditioned on a given estimate $\hat{T}$, the conditional LS estimate of ϕ is given by

$$\hat{\phi}_{CLS} = \arg\min_{\phi} \left\| \frac{\mathbf{t} - \phi\mathbf{1}}{\hat{T}} - \text{round}\left(\frac{\mathbf{t} - \phi\mathbf{1}}{\hat{T}}\right) \right\|_2^2,$$

where $\mathbf{1}$ is a vector of unit entries. Having obtained an estimate of ϕ, T can now be re-estimated via line search from the original data (without differencing). This procedure can then be repeated till convergence. Convergence in fit is assured, because each step is a LS line search. However, we note that the first estimate of ϕ is *very* sensitive to mismatch in the original estimate of T. This can be appreciated by considering the much simpler case wherein $\mathbf{k}$ is known. Given an estimate $\hat{T}$ of T, define $\epsilon := T - \hat{T}$. Given $\hat{T}$ and $\mathbf{k}$, for AWGN ϕ is estimated using $\hat{\phi} = \text{mean}(\mathbf{t} - \mathbf{k}\hat{T})$. This yields a systematic error term (bias) equal to $\frac{\epsilon}{M}\text{sum}(\mathbf{k})$, where M is the length of $\mathbf{k}$. Without missing observations, this is already equal to $\frac{M+1}{2}\epsilon$; the situation is further aggravated with missing observations, for then $\text{sum}(\mathbf{k})$ grows faster with M. When $\mathbf{k}$ is unknown, the task of estimating ϕ is further compounded. The net result is that iterative LS estimation only makes sense in terms of improving the quality of the estimates at very high SNR. In that regime, the improvement is probably not worth the associated complexity, unless very accurate synchronization is required. For this reason, we do not pursue this further here.

4. SIMULATIONS

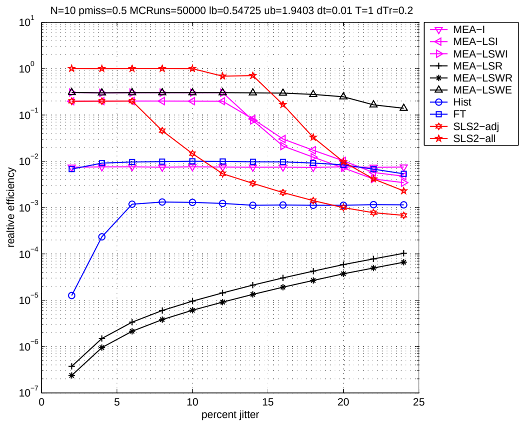
Choosing pairs: SLS2 can be applied to non-overlapping adjacent-pair differences (getting rid of ϕ but reducing the sample size by one half), or overlapping adjacent-pair differences (preserving the sample size), or even to data comprising all $\binom{N}{2}$ pairwise differences that can be extracted from the available sample - thus quadratically expanding the sample size. We will refer to these three options as SLS2-NOVLP, SLS2-ADJ, and SLS2-ALL, respectively. The SLS2 line search has complexity $O(\frac{U-L}{\Delta}M)$, where M is the SLS2-input sample size and Δ is the desired step-size accuracy. Aside from noise coloring considerations, SLS2-ALL has $M = O(N^2)$, which makes complexity quadratic in the original sample size. As we shall see, SLS2-ALL is well-worth this additional computational effort, especially for small sample sizes - which is the norm in many applications. As a sneak preview, we note that SLS2-ALL vastly outperforms SLS2-ADJ, which in turn outperforms SLS2-NOVLP, despite ignoring noise color.

Implementation of line search: Throughout our experiments, the SLS2 line search is implemented in two steps. The first is a "coarse" uniform grid search over 10^4 points spanning lower bound L to upper bound U (see (2)). L is set to $0.55T$ or an estimate thereof, as noted on each experiment. U is less critical (it does not affect identifiability), and is always set to the maximum value in the input sample. One could also use $U = 2L$ to further limit the search, since it is assumed that $L > T/2$, hence also $2L > T$. This first coarse localization search is followed by a refined search in the optimum bin. In this second step, quadratic interpolation of the cost function (finely sampled over 10^4 equispaced bin points) is used to localize the minimum.

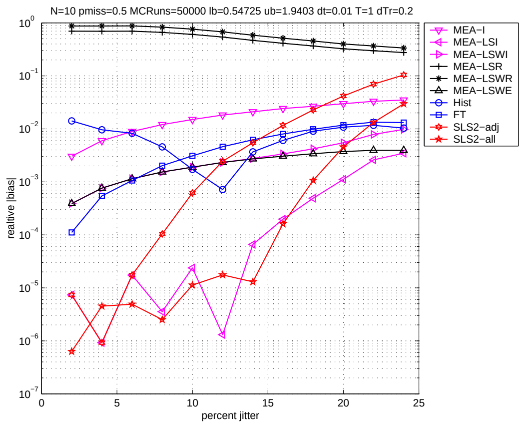
SNR considerations: Defining an appropriate measure of SNR turns out to be unexpectedly complicated for the simple model in (1). We skip the details due to space considerations, and state our chosen measure (cf. [2, 9]): $SNR := 20\log_{10}\frac{T}{\sigma_w}$, which is a measure of "jitter". Here, σ_w^2 denotes the variance of the AWGN in (1). In the simulations, we have chosen to parameterize performance via percent jitter, which is defined as $\frac{3\sigma_w}{T} \times 100\%$ because this measures the essential support of the error density over the period to be estimated.

Clairvoyant CRB: The conditional CRB for T assuming that $\{\kappa\}_{n=1}^N$ in (1) is *known* (ϕ is unknown) is $\text{CRB}(T) = \frac{\sigma_w^2}{N\bar{\sigma}_\kappa^2}$, where $\bar{\sigma}_\kappa^2 := \frac{1}{N}\sum_{n=1}^N \kappa^2(n) - \left(\frac{1}{N}\sum_{n=1}^N \kappa(n)\right)^2$ is the "sample variance of κ". We have shown that adjacent-sample differencing *does not* affect this CRB; proof of this claim is omitted due to space considerations.

Comprehensive Monte Carlo (MC) experiments: In all of our simulations, $\mathbf{k}$ is drawn from a simple Bernoulli miss model with miss probability 0.5; it is drawn once and remains fixed for the entire MC simulation. The numerical results depend on the particular realization of $\mathbf{k}$, but qualitative conclusions remain valid for different $\mathbf{k}$, as verified by further simulation. Throughout, $T = 1$ is used for the true value of the period. We used $50,000$ MC runs per datum reported, and the x-axis is percent jitter, going from smaller to higher jitters. Four compound plots are presented in Figures 2(a-b), and 3(a-b) for $N = 10$, and $N = 30$ samples, respectively. These depict relative efficiency (RE - estimation variance measured with respect to (wrt) the (clairvoyant) CRB for the given $\mathbf{k}$), and absolute bias. We compare SLS2-ALL and SLS2-ADJ with several other benchmark algorithms and the clairvoyant CRB. Six variants of MEA [2, 9] are evaluated. The suffix I, R, E denotes internal initialization, random initialization, and exact initialization. LS and LSW denote the least-squares and least-squares with whitening solutions. Internal MEA initialization was via the gradient/clustering procedure described in Sadler-Casey, yielding $\hat{T}_{MEA}$. The random initialization was based on $T_{init} := T(1 + 0.2 \times \text{sign}(\text{randn}))$. We also include a Fourier transform (FT) method as yet another baseline. For SLS2-ADJ, $L = 0.55 \times \hat{T}_{MEA}$ was used; for SLS2-ALL, $L = 0.55T$ was used. Recall that $L > T/2$ is *necessary* for identifiability, to avoid aliasing. Further note that SLS2-ADJ with $L = 0.55T$ performs better than with $L = 0.55 \times \hat{T}_{MEA}$, but still considerably worse than SLS2-ALL with $L = 0.55T$. SLS2-ALL clearly outperforms all other algorithms by a significant margin; for $N = 30$, it attains the clairvoyant CRB for jitter $\leq 20\%$. The SLS2-ALL efficiency breakpoint is a function of sample size N - it shifts to the right (higher jitter) with increasing N, as seen by comparing the results in Figures 2(a), and 3(a). We have verified via other simulation results (not included here) that SLS2 has RE close to unity for all values of N (with jitter below threshold). We note that for $N > 50$ MEA-LSWI also approaches RE=1 without requiring careful fine-tuning of internal initialization parameters.



(a) Relative Efficiency wrt clairvoyant CRB.



(b) Absolute Bias.

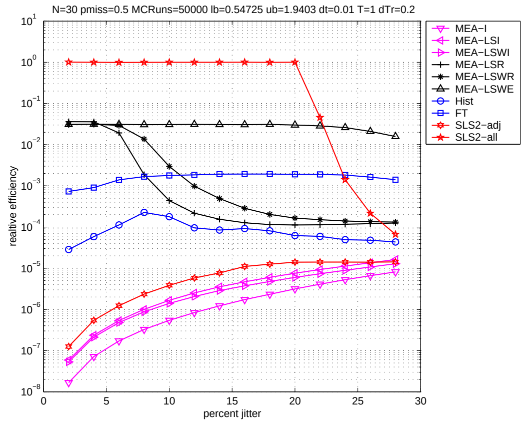
Fig. 2. Simulation results for $N = 10$.

5. CONCLUSIONS

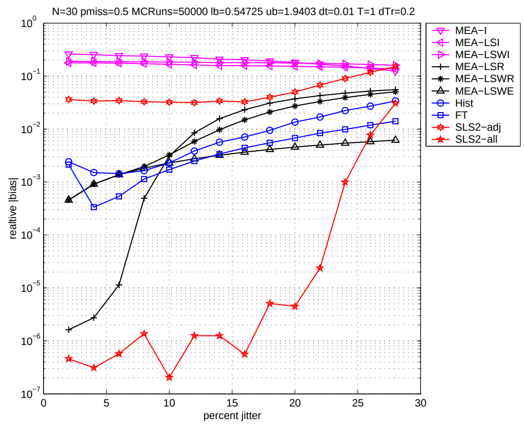
For moderate SNR and above, and even small sample sizes (e.g., $N = 10$), SLS2-ALL turns out to be "super-efficient", in the sense that it achieves the clairvoyant CRB - quite a surprise. We are currently investigating pertinent Barankin bounds and associated asymptotic analysis. In the presence of outliers, robust regression techniques can also be adopted. Perhaps the simplest way to endow SLS2 with a measure of robustness to outliers is to switch from ℓ_2 to ℓ_1 regression. This can be derived from the joint ML principle for Laplacian noise, and it simply changes the norm that is involved in the line search. Results will be reported elsewhere.

6. REFERENCES

[1] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channel", *IEEE Trans. Info. Theory*, 45(7), 1639–42, July 1999.



(a) Relative Efficiency wrt clairvoyant CRB.



(b) Absolute Bias.

Fig. 3. Simulation results for $N = 30$.

[2] S.D. Casey and B.M. Sadler, "Modifications of the Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements", *IEEE Transactions on Signal Processing*, 44(9):2260–2272, 1996.

[3] M. Ghogho, A. Swami and B. Garel, "Performance analysis of cyclic statistics for the estimation of harmonics in multiplicative and additive noise", *IEEE Trans. Sig. Proc.*, 47(12), 3235-49, Dec 1999.

[4] S. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. ASSP*, Vol. 37, No. 12, pp. 1987–1990, Dec. 1989.

[5] B. Kedem, "Spectral analysis and discrimination by zero-crossings," *Proc. IEEE*, Vol. 74, No. 11, pp. 1477–1493, Nov. 1986.

[6] M. Wing-Kin , T.N. Davidson, K.M. Wong, Z-Q Luo, C. Pak-Chung, "Quasi-ML multiuser detection using semi-definite relaxation with application to synchronous CDMA", *IEEE Trans. Sig. Proc.*, 50(4):912 -922, Apr. 2002.

[7] E. Parzen, "On spectral analysis with missing observations and amplitude modulation," *Sankhya, Ser. A*, Vol. 25, pp. 383–392, 1963.

[8] Y. Rosen, and B. Porat, "The second-order moments of the sample covariances for time series with missing observations," *IEEE Transactions on Information Theory*, Vol. 35, No. 2, pp. 334–341, 1989.

[9] B.M. Sadler and S.D. Casey, "On periodic pulse interval analysis with outliers and missing observations," *IEEE Trans. SP*, Vol. 46, No. 11, pp. 2990–3002, November 1998.

[10] P.A. Scheinok, "Spectral analysis with randomly missed observations: the binomial case," *Ann. Math. Statistics*, Vol. 36, pp. 971–977, 1965.

[11] S.A. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Transactions on Information Theory*, Vol. IT-31, pp. 832–835, 1985.