[20] C. Pimentel, T. H. Falk, and L. Lisbôa, "Finite-state Markov modeling of correlated Rician channels," in *Proc. Canadian Workshop Inform. Theory*, Waterloo, ON, Canada, May 2003, pp. 16–19.
[21] M. J. Ruf and J. W. Modestino, "Rate-distortion performance for joint source and channel coding of images," in *Proc. Int. Conf. Image Process.*, vol. II, Washington, DC, Oct. 1995, pp. 77–80.
[22] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, Jun. 1996.
[23] P. G. Sherwood and K. Zeger, "Progressive image coding on noisy channels," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1997, pp. 72–81.
[24] P. G. Sherwood, X. Tian, and K. Zeger, "Channel code block length and rate optimization for progressive image transmission," in *Proc. Wireless Commun. Networking Conf.*, New Orleans, LA, Sep. 1999, pp. 978–982.
[25] K. P. Subbalakshmi and J. Vaisey, "Optimal decoding of entropy-coded Markov sources over channels with memory," in *Proc. 33rd Annu. Conf. Inf. Sci. Syst.*, vol. 2, Mar. 1999, pp. 624–629.
[26] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," *IEEE J. Sel. Areas Commun.*, vol. 10, pp. 926–943, Jun. 1992.
[27] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Wireless image transmission using turbo codes and optimal unequal error protection," in *Proc. Int. Conf. Image Process.*, vol. I, Washington, DC, Sep. 2003, pp. 73–76.
[28] V. Vaishampayan and N. Farvardin, "Optimal block cosine transform image coding for noisy channels," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 327–336, Mar. 1990.
[29] K. Zeger, J. Vaisey, and A. Gersho, "Globally optimal vector quantizer design by stochastic relaxation," *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 310–322, Feb. 1992.

Quasi-ML Period Estimation From Incomplete Timing Data

Nicholas D. Sidiropoulos, Ananthram Swami, and Brian M. Sadler

***Abstract*—Given a noisy sequence of (possibly shifted) integer multiples of a certain period, it is often of interest to accurately estimate the period. With known integer regressors, the problem is classical linear regression. In many applications, however, the regressors are *unknown* integers, and only loose bounds on the period are available. Examples include hop period and timing estimation, wherein hops may be missed at the output of the frequency discriminator or the emitter may hop out of band; Pulse Repetition Interval (PRI) analysis; and passive rotating-beam radio scanning. We study several pertinent period estimators. Our emphasis is on a Quasi-Maximum Likelihood approach developed herein and an earlier method based on the Fourier Transform of a Dirac delta train representation of the data. Surprisingly, both are capable of attaining the *clairvoyant* Cramér–Rao Bound at moderate signal-to-noise ratios (SNRs), even for short (e.g., 10) samples. We carefully address parameter identifiability issues and corroborate our findings with extensive simulations.**

Manuscript received July 10, 2003; revised February 23, 2004. This work was prepared through collaborative participation in the Collaborative Technology Alliance for Communications and Networks sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0011. The associate editor coordinating the review of this paper and approving it for publication was Dr. Chong-Yung Chi.

N. D. Sidiropoulos is with the Departent of Elelctrical and Computer Engineering, Technical University of Crete, 73100 Chania, Crete, Greece, and also with the Departent of Elelctrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@telecom.tuc.gr).

A. Swami and B. Sadler are with the Army Research Laboratory, Adelphi, MD 20783-1197 USA (e-mail: aswami@arl.army.mil; bsadler@arl.army.mil).

Digital Object Identifier 10.1109/TSP.2004.840761

***Index Terms*—Fourier transform, frequency estimation, missing data, period estimation, pulse repetition interval analysis, synchronization, timing estimation.**

I. Introduction

Consider the following observation model:

$$\tau(n) = \phi + \kappa(n)T + w(n), \quad n = 1, \ldots, N \tag{1}$$

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 $\{\tau(n)\}$. 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."

The model in (1) is reminiscent of two well-known problems. In the special case that $\kappa(n) = n, n = 1, \ldots N$, 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 nonstandard 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 yields

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

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 signal-to-noise ratio (SNR) or moderate samples and above. Interestingly, Tretter [13] 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 AWG noise, and the problems become essentially equivalent [13]. Another related approach to the problem of frequency estimation involves working with zero-crossings or higher order zero crossings of the observation [6], [11].

The harmonic retrieval problem with missing samples has also been considered [8]. Early approaches were periodogram-based (the periodogram often works reasonably well with mild multiplicative noise), but parametric techniques have also been developed [9]. In most cases, a simple Bernoulli miss model is adopted [9], [12]; otherwise it is assumed that missing samples occur periodically with known outage period. Harmonic retrieval in multiplicative noise has been dealt with (see, e.g., [4]), 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 Fogel and Gavish [3], Sadler and Casey [1], [10], and Clarkson *et al.* [2], who also considered period estimation from the model in (1) with missing observations.

Fogel and Gavish [3] considered Maximum Likelihood (ML) period estimation from incomplete data for a certain convenient choice of the noise probability density function (pdf) that explicitly depends on the sought period. The said pdf is compactly supported, and hence, non-

Gaussian, but can approximate a Gaussian at high SNR. For this particular noise pdf, Fogel and Gavish [3] claimed that the Fourier Transform (FT) of the Dirac delta train representation of the data $\sum_{n=1}^{N}\delta(t-\tau(n))$ yields the ML estimate of the sought period. Proof of this claim was not included in [3]. The claim reappeared in [10, App.], but the argument therein is incomplete.

The work of Sadler and Casey [1], [10] is based on modifications of the Euclidean algorithm for the computation of the greatest common divisor. Following Sadler and Casey, Clarkson *et al.* [2] also considered the same problem. Clarkson's joint ML method requires multidimensional nonlinear optimization, which is computationally demanding and subject to local minima. Relative efficiency has not been considered in [2], which, however, sheds some light into the properties of the associated joint likelihood function. In particular, [2] showed that without a lower bound on the sought period, there are either infinitely many ML estimates or none.

Several years after Fogel and Gavish [3], Clarkson *et al.* [2] also proposed the FT of the delta train representation of the data as a means of period estimation. Section 5 of [2] shows that there exist finitely many points of the aforementioned spectrum that approximate, in a certain sense, respective values of the ML criterion for Gaussian noise. The link, as exposed in [2], is weak: It ignores the denominator of the ML criterion, and what happens at other points is left open. In [2], no claim is made regarding efficiency of the FT method by analysis or simulation[1]. In [2] (and [3]), parameter identifiability issues were not considered.

Our contributions can be summarized as follows. We study the model in (1) with respect to parameter identifiability, develop a new quasi-ML period estimation algorithm, and benchmark it against several earlier algorithms and the pertinent Cramér–Rao Bound (CRB). Our simulations show that both the new quasi-ML algorithm and the algorithm proposed by Fogel and Gavish [3], *when used with a proper lower bound on the search range*, guarantee identifiability and are capable of attaining the *clairvoyant* CRB. This is remarkable because the clairvoyant CRB assumes knowledge of the unknown regressors.

Our particular motivating application is hop period estimation in the context of Frequency-Hopped (FH) radios. Therein, one may often observe only part of the spread FH bandwidth because the true FH band may be unknown or noncontiguous, as well as 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 exact scan period may be unknown. Another situation wherein deterministic unknown modeling of missed observations may be appropriate could be a multitasked best-effort surveillance processor.

II. Basic Algorithm

Given estimates of T and $\{\kappa(n)\}_{n=1}^{N}$, the estimation of the shift ϕ is straightforward. As a first step toward simplifying the problem, we may take pairwise differences, which eliminates the offset ϕ. 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 nonoverlapping 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 3–dB loss in terms of noise amplification (note that in this case, the noise is still white): $t(n)=k(n)T+v(n), n=1,\ldots,M:=\lfloor(N/2)\rfloor$, where $t(n):=\tau(2n)-\tau(2n-1), k(n):=\kappa(2n)-\kappa(2n-1)\in\mathbb{Z}$, and $v(n):=w(2n)-w(2n-1)$. 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 least-squares (LS) problem $\min_{\tilde{T}\in\mathbb{R}_+,\mathbf{k}\in\mathbb{Z}^M}\|\mathbf{t}-\mathbf{k}\tilde{T}\|_2^2$. This problem is linearly separable, and the cost function can be concentrated with respect to $\mathbf{k}$; this yields

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

This latter minimization can be accomplished via simple line search over T.

It is important to note that, 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 yields finer granularity, which results in increasingly deeper minima 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 L on $\tilde{T}$, just like we need to impose an upper bound on frequency in order to prevent aliasing in the context of frequency estimation. This lower bound must clearly satisfy $(T/2)<L<T$; this yields $T<2L<2T$, and thus, we can search for

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

which can be solved by line search over $(L,2L)$. We will refer to this as the Separable Least Squares Line Search (SLS2) algorithm.

III. Identifiability

In the noiseless case, $\mathbf{t}=\mathbf{k}T$, and the question is whether or not one can uniquely determine $T>0$ (positivity is needed to avoid the trivial "reflection" ambiguity) and $\mathbf{k}\in\mathbb{Z}^M$ from $\mathbf{t}$. We have already seen that a bound L with $(T/2)<L<T$ is needed for identifiability, and this means that we search for solutions in the range $(L,2L)\subset((T/2),2T)$. Therefore, let $\boldsymbol{\ell}T'=\mathbf{k}T$ with $\boldsymbol{\ell}\in\mathbb{Z}^M$ and $(T/2)<T'<2T$ be another factorization of the data. It follows that $(\ell(i)/k(i))=(T/T'), i=1,\ldots,M$, and therefore, $(1/2)<(\ell(i))/(k(i))<2$ and $k(i)\ell(j)-k(j)\ell(i)=0$, for all i and j in $\{1,\ldots,M\}$. Consider, for example

$$k(1)\ell(2)-k(2)\ell(1)=0$$

which is a *homogeneous linear Diophantine equation* in two integer variables $(\ell(1),\ell(2))$, whose complete set of solutions is given by (e.g., [7])

$$\ell(1)=\frac{rk(1)}{\gcd(k(1),k(2))},\quad \ell(2)=\frac{rk(2)}{\gcd(k(1),k(2))}$$

parameterized by $r\in\mathbb{Z}$; here, $\gcd(k(1),k(2))$ stands for the *greatest common divisor* of $k(1)$ and $k(2)$. If $k(1)$ and $k(2)$ are relatively prime, then it follows that $\ell(1)=rk(1), \ell(2)=rk(2), r\in\mathbb{Z}$. For $r=1$, we recover the desired solution (recall that $(\ell(i))/(k(i))=$

[1]Recall that efficiency implies ML but not vice-versa.

$(\ell(1))/(k(1))$, and it thus follows that $\ell(i) = k(i), \forall i)$. All other possibilities for r are precluded by virtue of the inequality $(1/2) < (\ell(i))/(k(i)) < 2$. There is nothing special about the choice of the first two sample indices; we therefore arrive at the following result.

Theorem 1: If $\mathbf{k}$ contains a pair of relatively prime elements, then there is a unique factorization $\mathbf{t} = \boldsymbol{\ell} T'$ with $(T/2) < T' < 2T$, namely, $\mathbf{k}T$.

Note that the condition provided in the above result is sufficient but, in general, not necessary for uniqueness. To arrive at a necessary and sufficient condition, we must jointly consider all $\binom{M}{2}$ equations implied by the relation $k(i)\ell(j) - k(j)\ell(i) = 0, i, j \in \{1, \ldots, M\}$. This yields a homogeneous linear system of Diophantine equations in M variables, and identifiability is tantamount to uniqueness of solution of this system under the inequality constraints $(1/2) < (\ell(i))/(k(i)) < 2$. The system under consideration has special structure' in the sense that the system matrix is sparse (every row has two nonzero entries). A lot of progress has been made on various aspects of problems of this type, including algorithms for computing a basis that generates all solutions, checking for bounds, etc. This path leads straight into number theory and is beyond the scope of this paper, particularly in light of the fact that the sufficient condition already provided will likely be satisfied in all practical scenarios of interest. Note that if $1 \in \{k(i)\}_{i=1}^{M}$, then the condition is trivially satisfied. In addition, the probability that two integers picked at random are relatively prime is $6/\pi^2$ (e.g., see [1] and references therein); hence, even for $M = 10$, the probability that the sufficient condition will be violated is very small (strictly less than 10^{-2}).

IV. Further Algorithmic Aspects

Smoothing Considerations: If overlapping pair differences are used to extend the available sample size, 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 $\mathbf{t} = \mathbf{W}\bar{\mathbf{k}}T + \text{AWGN}$, where $\mathbf{W}$ is the whitening matrix, and the abbreviation AWGN stands for Additive White Gaussian Noise. 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 $\bar{\mathbf{k}}$. We therefore opt to ignore noise color and use plain LS. This will not matter much at high SNR [5].

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 $\hat{\phi}_{\text{CLS}} = \arg\min_{\phi} \|((\boldsymbol{\tau} - \phi\mathbf{1})/(\hat{T})) - \text{round}((\boldsymbol{\tau} - \phi\mathbf{1})/(\hat{T}))\|_2^2$, 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 an 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 $(\epsilon/M)\text{sum}(\mathbf{k})$, where M is the length of $\mathbf{k}$. Without missing observations, this is already equal to $((M+1)/2)\epsilon$; 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 estimates are required. For this reason, we do not pursue this further.

V. Simulations

Choosing Pairs: SLS2 can be applied to nonoverlapping adjacent-pair differences (getting rid of ϕ but reducing the sample size by one half), 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 $O(((U - L)/\Delta)M)$, 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 will 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 was implemented in two steps. The first was a "coarse" uniform grid search over 10^4 points spanning lower bound L to upper bound U [see (2)]. L was set to $0.55T$ or an estimate thereof, as noted on each experiment. U is less important. We could have used $U = 2L$ (since $L > T/2$ implies $2L > T$), but we have observed that setting U to the maximum value in the input sample does not degrade performance. The reason is that the likelihood function is roughly increasing at the macroscopic level. This means that a higher U typically does not throw off the estimate, whereas a lower L opens up the possibility of selecting T divided by some integer. The first coarse localization search was 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) was 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. [1], [10]): $\text{SNR} := 20\log_{10}(T/(\sigma_w))$, 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 $((3\sigma_w)/T) \times 100\%$ because this measures the essential support of the error density over the period to be estimated.

Clairvoyant CRB: Consider the vector model $\boldsymbol{\tau} = \boldsymbol{\kappa}T + \phi\mathbf{1} + \mathbf{w}$ of (1), with obvious notation. The conditional CRB for T, assuming that $\boldsymbol{\kappa}$ is *known* (ϕ is unknown) is [5] $\text{CRB}(T) = ((\sigma_w^2)/(N\bar{\sigma}_\kappa^2))$, 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 remark that adjacent-sample differencing *does not* affect this CRB. To see this, let $\mathbf{D}$ denote the $(N-1) \times N$ matrix with ones on the diagonal and -1 on the super-diagonal. Then, $\mathbf{t} = \mathbf{D}\boldsymbol{\tau}$ represents the differencing operation. Based on $\mathbf{t}$, the LS estimator of T, with prewhitening, (i.e., ML based on $\mathbf{t}$) has variance given by $\text{var}(\hat{T}_{\mathbf{t}}) = [\boldsymbol{\kappa}'\mathbf{D}'(\mathbf{D}\mathbf{R}_{\mathbf{w}}\mathbf{D}')^{-1}\mathbf{D}\boldsymbol{\kappa}]^{-1}$, where $\mathbf{R}_w$

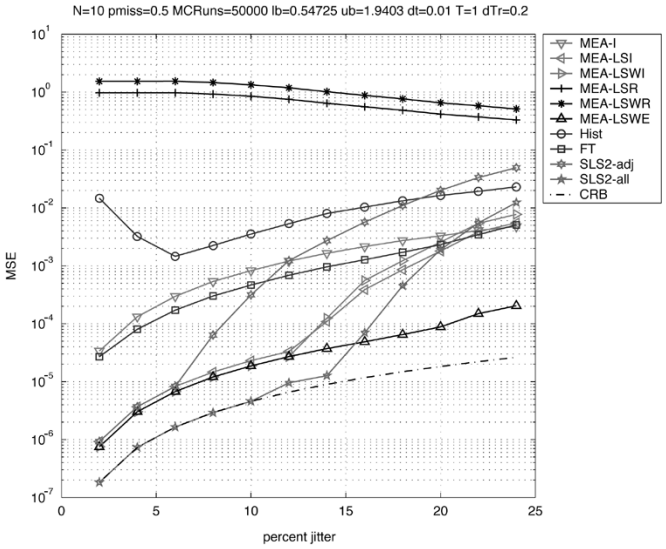

Fig. 1. Simulation results for $N = 10$: MSE.

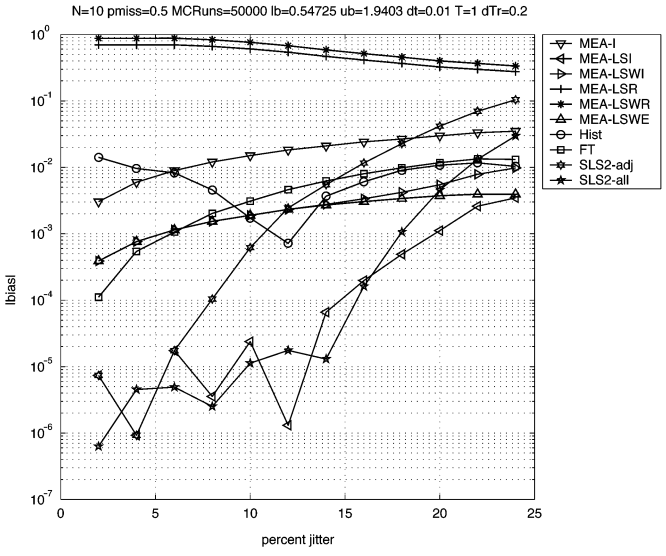

Fig. 2. Simulation results for $N = 10$: Absolute Bias.

is the noise covariance matrix for the model in (1). For AWGN as in (1), $\mathbf{R}_w = \sigma_w^2 \mathbf{I}$, and we have

$$\begin{aligned}\mathrm{var}(\hat{T}_{\mathbf{y}}) &= \sigma_w^2 [\boldsymbol{\kappa}' \mathbf{D}' (\mathbf{D}\mathbf{D}')^{-1} \mathbf{D} \boldsymbol{\kappa}]^{-1} \\ &= \sigma_w^2 \left[\boldsymbol{\kappa}' \left(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}' \right) \boldsymbol{\kappa} \right]^{-1} \\ &= \sigma_w^2 \left[\boldsymbol{\kappa}' \boldsymbol{\kappa} - \frac{1}{N} (\mathbf{1}' \boldsymbol{\kappa})^2 \right]^{-1} = \frac{\sigma_w^2}{N \bar{\sigma}_\kappa^2}\end{aligned}$$

which is the same as the CRB(T) in the joint estimation of (T, ϕ) based on the original data.

Comprehensive Monte Carlo (MC) Experiments: In all of our MC simulations, $\boldsymbol{\kappa}$ was drawn from a simple Bernoulli miss model with miss probability $p_{\mathrm{miss}} = 0.5$; it was drawn once and remained fixed for the entire MC simulation. The numerical results depend on the particular realization of $\boldsymbol{\kappa}$, but qualitative conclusions remain valid for different $\boldsymbol{\kappa}$, as verified by further simulation.

Four compound plots are presented in Figs. 1 and 2 and in Figs. 3 and 4 for $N = 10$ and $N = 30$ samples, respectively. These depict

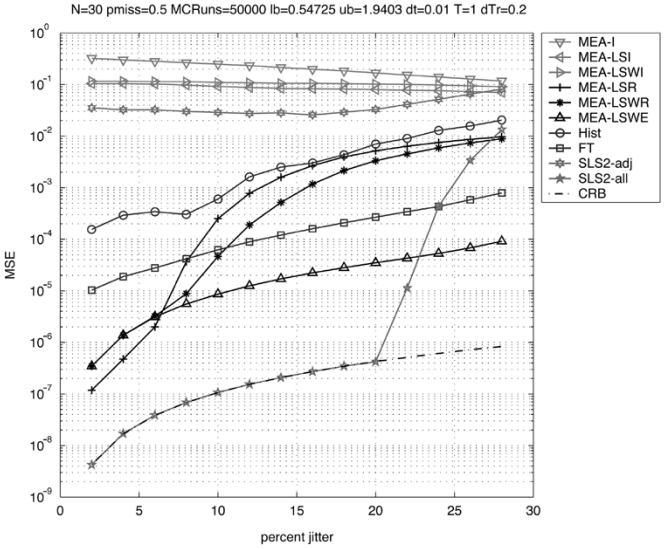

Fig. 3. Simulation results for $N = 30$: MSE.

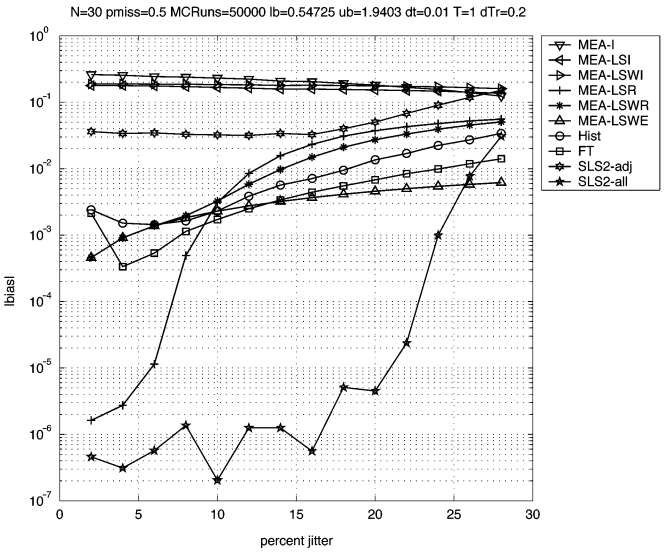

Fig. 4. Simulation results for $N = 30$: Absolute Bias.

mean squared error (MSE) and absolute bias results for the estimation of T. 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 in percent jitter, going from smaller to higher jitters (higher SNR to lower SNR).

We compared SLS2-ALL and SLS2-ADJ with several other benchmark algorithms and the clairvoyant CRB. Six variants of the Modified Euclidean Algorithm (MEA) [1], [10] were evaluated. The suffixes I, R, and E denote internal initialization, random initialization, and exact initialization, respectively. LS and LSW denote the least-squares and least-squares with whitening solutions. Internal MEA initialization was via the gradient/clustering procedure described in [10], yielding $\hat{T}_{\mathrm{MEA}}$. The random initialization was based on $T_{\mathrm{init}} := T(1 + 0.2 \times \mathrm{sign}(\mathrm{randn}))$. For SLS2-ADJ, $L = 0.55 \times \hat{T}_{\mathrm{MEA}}$ was used; for SLS2-ALL, $L = 0.55T$ was used. Recall that $L > T/2$ is *necessary* for identifiability to avoid aliasing. Although not reported here, we have observed from other simulations that SLS2-ADJ with $L = 0.55T$ performs better than with $L = 0.55 \times \hat{T}_{\mathrm{MEA}}$ but still considerably worse than SLS2-ALL with $L = 0.55T$.

We also included in the comparison an empirical histogram-based (Hist) method and a Fourier transform (FT) method. These are described next.

Consider $z_\tau(n) := \text{rem}(t(n), \tau)/\tau$, where $t(n) = k(n)T + v(n)$, and $\text{rem}(t(n), \tau)$ stands for the remainder of the division of $t(n)$ by τ. When T/τ is integer and the noise is small, $z_\tau(n)$ will be clustered around integer values; otherwise, its distribution will be spread over $[0, \tau)$. In the latter case, the associated empirical distribution (obtained via the histogram) will have relatively high entropy. Thus, the τ that yields minimum entropy can be taken as an estimate of T. The FT method[2] is related to the histogram approach. The FT algorithm computes[3]

$$S(\tau) = \sum_{n=1}^{M} e^{-j2\pi t(n)/\tau} = \sum_{n=1}^{M} e^{-j2\pi(k(n)T/\tau + v(n)/\tau)}$$

and maximizes $|S(\tau)|$ via a one-dimensional grid search. Note that $e^{-j2\pi t(n)/\tau} = e^{-j2\pi z_\tau(n)}$; hence, $S(\tau)$ is another statistic based on the $z_\tau(n)$'s. Again, when T/τ is integer (and noise is small), $S(\tau)$ will be large; for other values of τ, the phase will be uniformly distributed, and the expected value of $S(\tau)$ will be zero.

The FT algorithm can be further motivated as follows. The data $\{t(n)\}_{n=1}^{M}$ can be represented by the Dirac delta train

$$\Delta(t) = \sum_{n=1}^{M} \delta(t - t(n))$$

whose (continuous-time) Fourier transform is given by

$$\Delta(\omega) = \int_{t=-\infty}^{\infty} \Delta(t) e^{-j\omega t}\, dt = \sum_{n=1}^{M} e^{-j\omega t(n)}$$

or, setting $\omega = 2\pi/\tau$

$$S(\tau) := \Delta(2\pi/\tau) = \sum_{n=1}^{M} e^{-j2\pi t(n)/\tau}.$$

Note that when using the delta train representation, additive noise manifests as timing jitter (noise is no longer additive on the signal part of $\Delta(t)$). For both of the above empirical estimators, the search range for τ was limited to (0.547 25, 1.9403), and a fixed step size (0.01) was used for the grid search. Peak picking was followed by quadratic interpolation, but that did not improve performance.

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 Figs. 5 and 6, which depict relative efficiency (RE) with respect to the clairvoyant CRB (CRB divided by estimator variance) for the two setups in Figs. 1–4, respectively.

VI. FT, Revisited

During the second-round review of our manuscript, one reviewer remarked that the FT method based on the Dirac delta train representation

[2]Note that this FT-based estimator is not the same as peak-picking the Fourier transform of the data. The latter is clearly not appropriate in our present context.

[3]Here, $1/\tau =: f$ plays the role of the frequency variable.

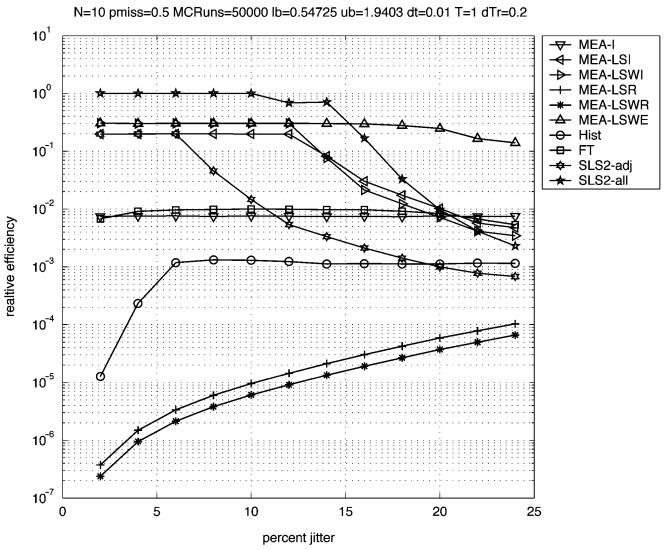

Fig. 5. Relative efficiency wrt clairvoyant CRB plots: $N = 10$.

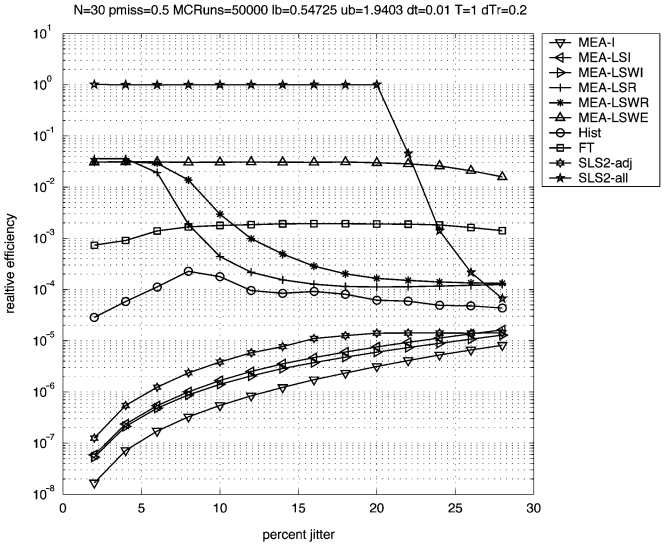

Fig. 6. Relative efficiency wrt clairvoyant CRB plots: $N = 30$.

of the original data (instead of the pair-wise differenced data), as advocated in [2] (cf. [3]), should also provide excellent performance.

> Γηράσκω αεί διδασκόμενος (*I age, perpetually being taught*, Solon, Seventh century B.C.)

Indeed, this is correct, and we are grateful to the reviewer for pointing it out. We have simulated the FT method using the Dirac delta train representation of the original data. The results are presented in Figs. 7–9. Sample length $N = 30$ and MC $= 30\,000$ runs were used. Two versions of FT were tested against SLS2-ALL. The first (FTwID) used the same identifiability-induced lower bound $(0.55T)$ on the search range as SLS2-ALL, whereas the second (FTnID) ignored identifiability considerations. The grid search step size was set to 10^{-4} for all three methods. Otherwise, the setup is as in the earlier simulations. Summarizing the results in Figs. 7–9, we have the following.

- SLS2-ALL and FTwID are both efficient at high SNR (including very high SNR).
- FTnID is not efficient at *very high* SNR. This can be understood as follows. In the noiseless case, the criterion $|\sum_{n=1}^{N} e^{-j2\pi\tau(n)/\tau}|$ is maximized at all τ for which T/τ is integer. Without a bound

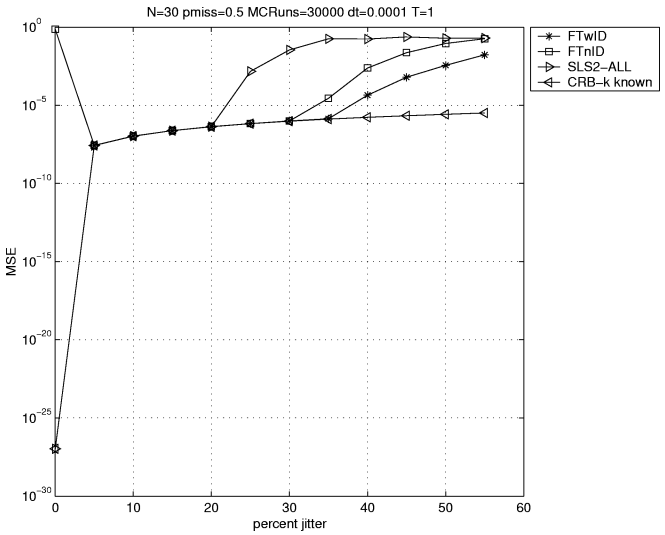

Fig. 7. Simulation results for FT operating on the original data: MSE.

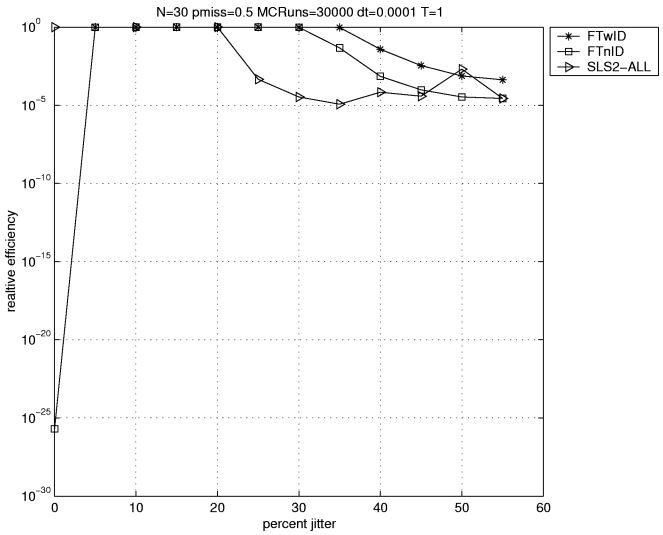

Fig. 9. Relative efficiency wrt clairvoyant CRB: FT operating on original data.

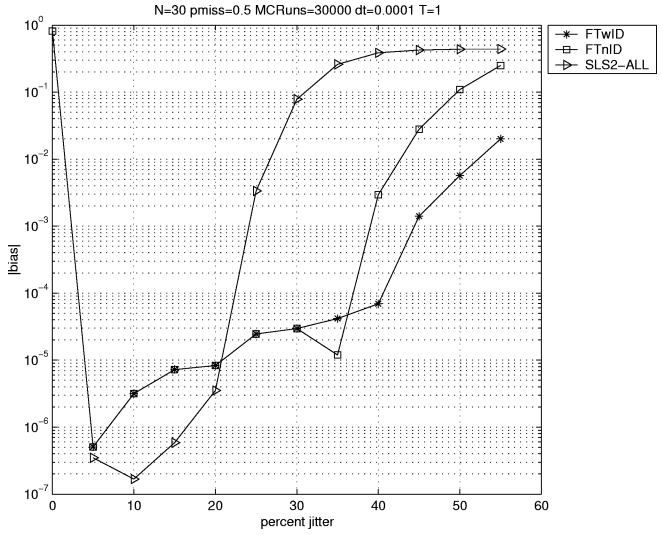

Fig. 8. Simulation results for FT operating on the original data: Absolute Bias.

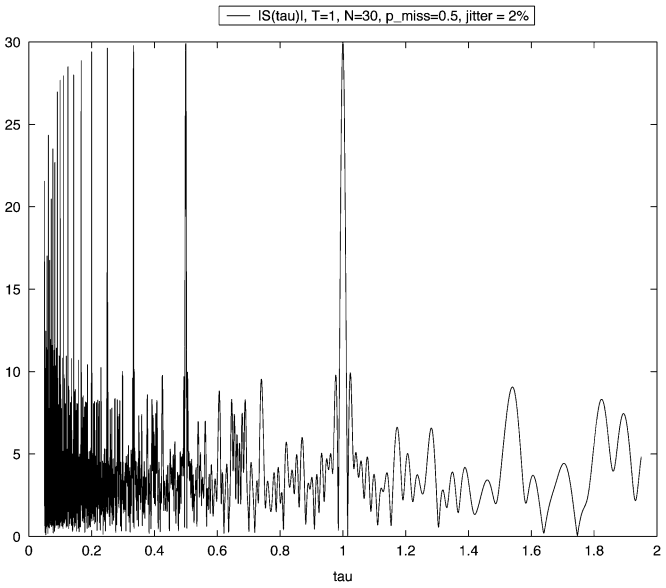

Fig. 10. Typical $|S(\tau)|$ spectrum at 2% jitter. Notice that noise dithers the artifacts at $T/k, k$ integer > 1.

that is greater than $T/2$, artifact modes at $T/k, k$ integer > 1 often prevail.

- Interestingly, FTnID quickly becomes efficient as SNR decreases (jitter increases). Noise dithers the artifacts at T/k in a way that consistently favors the mode at T. This is illustrated in Fig. 10. To explain this phenomenon, consider the spectrum

$$|S(\tau)| = \left| \sum_{n=1}^{N} e^{-j2\pi\tau(n)/\tau} \right|$$

at points $\tau = T/m, m \in \mathbb{Z}$. With $\tau(n) = \kappa(n)T + w(n), \kappa(n) \in \mathbb{Z}$

$$|S(T/m)| = \left| \sum_{n=1}^{N} e^{-j2\pi \frac{m}{T} w(n)} \right| = \left| \hat{\Psi}_w \left(-2\pi \frac{m}{T} \right) \right|$$

where $\Psi_w(\Omega) := E[e^{j\Omega w}]$ is the characteristic function of the noise pdf, and $\hat{\Psi}_w(\Omega)$ is its sample estimate. This sample estimate is consistent; hence, for sufficiently large N, it will approximate $\Psi_w(\Omega)$. For all noise pdfs that have a unimodal characteristic function (including the Gaussian), $\Psi_w(\Omega)$ is a decreasing function of the magnitude of its argument. For this reason, $|S(T/m)| = |\hat{\Psi}_w(-2\pi(m/T))|$ is a decreasing[4] function of $m \in \mathbb{Z}$ for large enough N. When SNR is very high (or N is very low, roughly under $N = 10$), numerical accuracy (respectively, inconsistency) problems emerge and throw off the estimate. This also happens irrespective of N, when noise is zero, cf. the earlier discussion on identifiability.
- All three methods depart from the clairvoyant CRB at low-enough SNR.
- FTwID uniformly outperforms the other two estimators, remaining essentially on the clairvoyant CRB for a wide range of SNR.
- SLS2-ALL appears to have slightly higher relative efficiency than the other two at 50% jitter, but this is compensated by higher bias,

[4]If $\Psi_w(-2\pi(1/T)) = 0$, then the mode of interest is also suppressed. This is not the case for the Gaussian.

as is evident by the MSE plots. The SLS2-ALL bias at low SNR can be attributed to noise coloring.

Even in light of [2] and [3] (see also the discussion in Section I), some of the above findings are puzzling. We have verified that qualitative conclusions are consistent across various N and p_{miss}.

VII. Conclusion

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. This is quite unexpected, given that SLS2-ALL *jointly* detects $\mathbf{k}$ and estimates T and further ignores noise color. This superefficiency can be partially attributed to using all $\binom{N}{2}$ pairwise differences and quantization effects due to the round$(\cdot)$ operator. These very reasons are also responsible for complicating performance analysis: Due to the nonlinearity and discontinuity of the round$(\cdot)$ operator and the dependence introduced by taking all $\binom{N}{2}$ pairwise differences, theoretical analysis of SLS2-ALL performance proved elusive, even in the asymptotic regime.

The FT of the delta train representation of the data, proposed by Fogel and Gavish [3], when used in conjunction with the proper identifiability-induced lower bound on the search range, outperforms all other estimators and attains the clairvoyant Gaussian CRB for a very wide range of SNR. This is remarkable and yet far from being understood to our satisfaction. Even without the identifiability-induced bound, the said FT method works very well, except at very high and very low SNR. The mechanism behind this behavior of FTnID at very high SNR has been exposed.

References

[1] S. D. Casey and B. M. Sadler, "Modifications of the Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements," *IEEE Trans. Signal Process.*, vol. 44, no. 9, pp. 2260–2272, Sep. 1996.

[2] I. V. L. Clarkson, S. D. Howard, and I. M. Y. Mareels, "Estimating the period of a pulse train from a set of sparse, noisy measurements," in *Proc. 4th Int. Symp. Signal Process. Applicat.*, vol. 2, Gold Coast, Australia, Aug. 1996, pp. 885–888.

[3] E. Fogel and M. Gavish, "Parameter estimation of quasiperiodic sequences," in *Proc. ICASSP*, vol. 4, 1988, pp. 2348–2351.

[4] M. Ghogho, A. Swami, and B. Garel, "Performance analysis of cyclic statistics for the estimation of harmonics in multiplicative and additive noise," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3235–3249, Dec. 1999.

[5] S. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*, NJ: Prentice-Hall, 1993.

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

[7] L. J. Mordell, *Diophantine Equations*. New York: Academic, 1969.

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

[9] Y. Rosen and B. Porat, "The second-order moments of the sample covariances for time series with missing observations," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 334–341, Mar. 1989.

[10] B. M. Sadler and S. D. Casey, "On periodic pulse interval analysis with outliers and missing observations," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2990–3002, Nov. 1998.

[11] ——, "Sinusoidal frequency estimation via sparse zero crossings," *J. Franklin Inst.*, vol. 337, pp. 131–145, 2000.

[12] P. A. Scheinok, "Spectral analysis with randomly missed observations: The binomial case," *Ann. Math. Statist.*, vol. 36, pp. 971–977, 1965.

[13] S. A. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 832–835, Nov. 1985.