Frequency Estimation by Phase Unwrapping

Robby G. McKilliam, Barry G. Quinn, I. Vaughan L. Clarkson, and Bill Moran

***Abstract*—Single frequency estimation is a long-studied problem with application domains including radar, sonar, telecommunications, astronomy and medicine. One method of estimation, called *phase unwrapping*, attempts to estimate the frequency by performing linear regression on the phase of the received signal. This procedure is complicated by the fact that the received phase is 'wrapped' modulo 2π and therefore must be 'unwrapped' before the regression can be performed. In this paper, we propose an estimator that performs phase unwrapping in the least squares sense. The estimator is shown to be strongly consistent and its asymptotic distribution is derived. We then show that the problem of computing the least squares phase unwrapping is related to a problem in algorithmic number theory known as the nearest lattice point problem. We derive a polynomial time algorithm that computes the least squares estimator. The results of various simulations are described for different values of sample size and SNR.**

***Index Terms*—Central limit theorem, frequency estimation, lattices, nearest lattice point problem, number theory, phase unwrapping.**

I. Introduction

ESTIMATION of the frequency of a single noisy sinusoid is a long studied problem with applications including radar, sonar, telecommunications, astronomy, and medicine [1], [2]. In this paper, a single frequency signal is modelled as a complex sinusoid of the form

$$A\exp\left(2\pi j(f_0 n+\theta_0)\right) \tag{1}$$

where f_0 is the frequency, θ_0 is the phase, $n \in \{1, 2\ldots, N\}$, A is the signal amplitude and $j=\sqrt{-1}$. The aim is to estimate the parameters f_0 and θ_0 from the signal

$$v_n = A\exp\left(2\pi j(f_0 n+\theta_0)\right)+s_n \tag{2}$$

where the sequence $s_1, s_2, \ldots$ is a complex noise process. We shall assume, in this paper, that the random variables s_n are independent and identically distributed, and that the distribution of $e^{j\alpha}s_n$ does not depend on α. This will occur exactly when the distribution of s_n depends only on $|s_n|$. To ensure identifiability we assume that f_0 and θ_0 are in $[-0.5, 0.5)$.

The maximum likelihood estimator of frequency under Gaussian white assumptions is known to be very closely approximated by the frequency that maximizes the *periodogram* (squared magnitude of the Fourier transform) of the v_n [1]. We refer to this estimator as the *periodogram estimator*; its asymptotic properties have been known for some time [1], [3], [4].

Rife and Boorstyn have suggested a practical method for computing the periodogram estimator, by using the fast Fourier transform to obtain the value of the periodogram at the Fourier frequencies $f = -\frac{1}{2}, \frac{1}{N}-\frac{1}{2}, \frac{2}{N}-\frac{1}{2}, \ldots, \frac{1}{2}-\frac{1}{N}$ [2]. The Fourier frequency that maximizes the periodogram is found and this estimate is then further refined by a numerical approach such as Newton's method. A problem with this approach is that the numerical procedure can fail to locate the correct maximizer [5]. To avoid the problem, Rife and Boorstyn suggested zero padding the signal to the length $4N$ before performing the fast Fourier transform. This has recently been shown to work by Quinn *et al.*, who also show that applying Newton's method to the derivative of certain monotonic functions of the periodogram, rather than to the periodogram itself, ensures that Newton's method will succeed even without any zero padding [6].

Regardless of these implementation difficulties, the periodogram estimator is widely seen as the best method for single frequency estimation. It provides the most accurate results and using the fast Fourier transform can be computed in only $O(N\log N)$ arithmetic operations. Nevertheless, many other methods for single frequency estimation exist.

One alternative method is *phase unwrapping* [7]. Phase unwrapping estimators appear to have been first suggested by Tretter [8], who utilized the fact that the phase of a complex sinusoid is a linear function that is 'wrapped' modulo 2π. If the phase could be 'unwrapped' then f_0 and θ_0 could be estimated by linear regression. Many phase unwrapping estimators have since been suggested [9]–[13]. A common approach is to compute the first differences of the arguments of the v_n. The resulting signal then resembles a moving average process, whose parameters can be estimated by standard linear techniques. This approach was first suggested by Kay [9], [10]. A significant advantage of this approach is that the moving average process has enough structure for the estimates to be computed with only $O(N)$ arithmetic operations. The estimator also appears to perform well when the signal-to-noise ratio is large. The major drawback of Kay's estimator is that it thresholds at a relatively

Manuscript received July 09, 2009; accepted February 10, 2010. Date of publication March 15, 2010; date of current version May 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ta-Hsin Li. The work of R. McKilliam was supported by a scholarship from the Wireless Technologies Laboratory, CSIRO ICT Centre, Sydney, Australia.

R. G. McKilliam and I. V. L. Clarkson are with the School of Information Technology and Electrical Engineering, The University of Queensland, Qld., 4072, Australia (e-mail: robertm@itee.uq.edu.au; v.clarkson@itee.uq.edu.au).

B. G. Quinn is with the Department of Statistics, Macquarie University, Sydney, NSW, 2109, Australia (e-mail: bquinn@efs.mq.edu.au).

B. Moran is with the Department of Electrical Engineering and Computer Science, Melbourne Systems Lab, Department of Electrical and Electronic Engineering, University of Melbourne, Vic. 3010, Australia (e-mail: wmoran@unimelb.edu.au).

Digital Object Identifier 10.1109/TSP.2010.2045786

high SNR and also performs poorly at moderate SNRs [14]. In fact, the estimator has been shown to be inconsistent, i.e., converges almost surely as $N \to \infty$ to the wrong frequency [1].

Another approach is to unwrap the phase in the least squares sense. We call the resulting estimator the least squares phase unwrapping estimator (LSPUE). Clarkson has shown how the LSPUE is related to a problem in algorithmic number theory known as the nearest lattice point problem [11], [15]. He has also shown that the LSPUE is closely related to the periodogram estimator. Quinn has derived the asymptotic properties of a LSPUE in the simpler case of phase estimation, i.e. when $f_0 = 0$ and the aim is to estimate θ_0 [16].

In this paper we derive the asymptotic properties of the LSPUE for single frequency estimation (Section III). We show that the estimator is strongly consistent and derive its central limit theorem. Following Clarkson [11] we discuss methods to compute the LSPUE using a nearest lattice point approach (Section IV). While Clarkson suggested solving the nearest lattice point problem using the approximate algorithm of Babai [17], we describe a polynomial-time algorithm that computes the estimator exactly. In Section V we provide simulations that display the statistical performance of the LSPUE alongside the periodogram estimator and some other single frequency estimators. The LSPUE proves to be only marginally less accurate than the periodogram estimator and significantly more accurate than other estimators based on phase unwrapping. We also use the simulations to confirm the asymptotic results derived in Section III.

II. The Least Squares Phase Unwrapping Estimator

The argument of the v_n, denoted $\angle v_n$, is given by

$$\angle v_n = 2\pi(f_0 n + \theta_0 + X_n) \pmod{2\pi}$$

where X_n is noise in the range $[-0.5, 0.5)$ and has a distribution depending only on A and the probability density function (pdf) of the s_n. When s_n is complex Gaussian noise the distribution of the X_n is known as the projected normal distribution and has been studied by Mardia and Jupp [18, p. 46]. The distribution has also been discussed by Quinn [16] and Tretter [8]. Other circular noise distributions may be used, for example, the wrapped normal, or von Mises distributions [18], [19]. In this paper we assume that the X_n are continuous, independent and identically symmetrically distributed, with cumulative distribution function (cdf) $F_X(x)$ and pdf $f_X(x)$. We denote by σ^2 the common variance of the X_n. We also assume that $f_X(x)$ is unimodal, with mode at 0. These assumptions are satisfied by a wide range of circular distributions including the projected normal, wrapped normal and von Mises distributions [18], [19].

Let $Y_n = \angle v_n/(2\pi)$. Then

$$Y_n = \langle f_0 n + \theta_0 + X_n \rangle \tag{3}$$

where $\langle x \rangle = x - \lfloor x \rceil$ is the fractional part of x and $\lfloor x \rceil$ denotes the nearest integer to x.[1]

[1]The direction of rounding for half-integers is not important, as long as it is consistent. The authors have chosen to round up half-integers here.

We may write (3) as

$$Y_n = f_0 n + \theta_0 + X_n + U_n \tag{4}$$

where $U_n = -\lfloor f_0 n + \theta_0 + X_n \rceil$. By considering the U_n as nuisance parameters, where $U_n \in \mathbb{Z}$, we may derive the sum of squares function

$$\sum_{n=1}^{N} (Y_n - \theta - fn - U_n)^2. \tag{5}$$

For fixed f and θ, (5) is minimized when $U_n = -\lfloor Y_n - \theta - fn \rceil$. Substituting this into (5) we obtain

$$SS(f,\theta) = \sum_{n=1}^{N} \langle Y_n - \theta - fn \rangle^2. \tag{6}$$

The LSPUE returns the f and θ that minimise $SS(f,\theta)$.

III. Asymptotic Properties

In this section the LSPUE is shown to be strongly consistent and its central limit theorem is derived. The main result is Theorem 1, the proof of which is given at the end of this section. In what follows, order notation will always refer to behavior as $N \to \infty$, and the notation $O_P(\cdot)$ will mean order in probability as $N \to \infty$.

Theorem 1: Let $(\widehat{f}_N, \widehat{\theta}_N)$ be the minimizer of (6) over $[-0.5, 0.5)^2$. Then, $N(\widehat{f}_N - f_0)$ and $\widehat{\theta}_N - \theta_0$ converge almost surely to 0 as $N \to \infty$, and the distribution of

$$\left[N^{3/2}(\widehat{f}_N - f_0) \quad N^{1/2}(\widehat{\theta}_N - \theta_0) \right]'$$

converges to the normal with mean 0 and covariance matrix

$$\frac{12\sigma^2}{(1-h)^2} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

where $\sigma^2 = \mathrm{var} X_n$ and $h = f_X(-1/2)$.

Substituting (3) into $SS(f,\theta)$ we obtain

$$\begin{aligned} &\sum_{n=1}^{N} \langle \langle \theta_0 + f_0 n + X_n \rangle - \theta - fn \rangle^2 \\ &\quad = \sum_{n=1}^{N} \langle X_n + (f_0 - f)n + (\theta_0 - \theta) \rangle^2 \\ &\quad = \sum_{n=1}^{N} \langle X_n + \lambda n + \phi \rangle^2 \\ &\quad = N S_N(\lambda, \phi) \end{aligned}$$

say, where $\lambda = f_0 - f$ and $\phi = \theta_0 - \theta$. Note that S_N is periodic with period 1 in both λ and ϕ. We may thus assume that $(\lambda, \phi) \in B$ where $B = [-0.5, 0.5)^2$. Let $\widehat{\lambda}_N$ and $\widehat{\phi}_N$ be the minimizers of $S_N(\lambda, \phi)$. We shall show that $N\widehat{\lambda}_N \to 0$ and $\widehat{\phi}_N \to 0$ almost surely as $N \to \infty$. Put

$$\begin{aligned} V_N(\lambda,\phi) &= S_N(\lambda,\phi) - E S_N(\lambda,\phi) \\ &= \frac{1}{N} \sum_{n=1}^{N} \left(\langle X_n + \lambda n + \phi \rangle^2 - E\langle X_n + \lambda n + \phi \rangle^2 \right). \end{aligned}$$

Let (λ_j, ϕ_k) denote the point

$$(\lambda_j, \phi_k) = \left(\frac{j}{N^{b+1}} - \frac{1}{2}, \frac{k}{N^b} - \frac{1}{2} \right),$$

for some $b > 0$ and let

$$B_{jk} = \left\{ (x,y); \frac{j}{N^{b+1}} \le x + \frac{1}{2} < \frac{j+1}{N^{b+1}}, \right. \\ \left. \frac{k}{N^b} \le y + \frac{1}{2} < \frac{k+1}{N^b} \right\}.$$

Then

$$\begin{aligned} &\sup_{(\lambda,\phi)\in B} |V_N(\lambda,\phi)| \\ &\quad = \sup_{j,k} \sup_{(\lambda,\phi)\in B_{jk}} |V_N(\lambda_j,\phi_k) + V_N(\lambda,\phi) - V_N(\lambda_j,\phi_k)| \\ &\quad \le \sup_{j,k} |V_N(\lambda_j,\phi_k)| \\ &\qquad + \sup_{j,k} \sup_{(\lambda,\phi)\in B_{jk}} |V_N(\lambda,\phi) - V_N(\lambda_j,\phi_k)| \end{aligned}$$

where $j = 0, 1, \ldots, N^{b+1} - 1$ and $k = 0, 1, \ldots, N^b - 1$.

Lemma 1: $\sup_{j,k} |V_N(\lambda_j, \phi_k)| \to 0$, almost surely as $N \to \infty$.

Proof: For any $\varepsilon > 0$

$$P\left(\sup_{j,k} |V_N(\lambda_j,\phi_k)| > \varepsilon \right) \le \sum_{j,k} P\left(|V_N(\lambda_j,\phi_k)| > \varepsilon \right) \\ \le \sum_{j,k} \frac{E\left(V_N^{2\beta}(\lambda_j,\phi_k) \right)}{\varepsilon^{2\beta}}$$

by Markov's inequality, for any $\beta > 0$. Let

$$E\left(V_N^{2\beta}(\lambda,\phi) \right) = \frac{1}{N^{2\beta}} E\left(\sum_{n=1}^{N} Z_n \right)^{2\beta}$$

where each of

$$Z_n = \langle X_n + \lambda n + \phi \rangle^2 - E\langle X_n + \lambda n + \phi \rangle^2$$

has mean 0. In Lemma 9 in the Appendix we show, for integers β, that

$$E\left(\sum_{n=1}^{N} Z_n \right)^{2\beta} = O(N^{\beta}). \tag{7}$$

Hence

$$P\left(\sup_{j,k} |V_N(\lambda_j,\phi_k)| > \varepsilon \right) = O(N^{2b+1-\beta})$$

and, since for any $b > 0$, we may choose β so that $2b + 1 - \beta < -1$, it follows that

$$\sum_{N=1}^{\infty} P\left(\sup_{j,k} |V_N(\lambda_j,\phi_k)| > \varepsilon \right) < \infty$$

and consequently from the Borel-Cantelli lemma that

$$\sup_{j,k} |V_N(\lambda_j,\phi_k)| \to 0$$

almost surely as $N \to \infty$. ■

Lemma 2: $\sup_{j,k} \sup_{(\lambda,\phi)\in B_{jk}} |V_N(\lambda,\phi) - V_N(\lambda_j,\phi_k)| \to 0$, almost surely as $N \to \infty$.

Proof: Let $a_n = X_n + \lambda n + \phi$ and $b_n = X_n + \lambda_j n + \phi_k$. For $(\lambda, \phi) \in B_{jk}$

$$\langle a_n \rangle = \langle b_n + \nu_n \rangle$$

where

$$0 \le \nu_n = a_n - b_n = (\lambda - \lambda_j)n + (\phi - \phi_k) \le 2N^{-b}.$$

Thus

$$\langle a_n \rangle = \langle b_n \rangle + \nu_n - \delta_n$$

where

$$\delta_n = \begin{cases} 0; & \langle b_n \rangle + \nu_n < 0.5 \\ 1; & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} \langle a_n \rangle^2 - \langle b_n \rangle^2 &= (\nu_n - \delta_n)^2 + 2(\nu_n - \delta_n)\langle b_n \rangle \\ &= \nu_n^2 + 2\nu_n \langle b_n \rangle - \delta_n \left[2\left(\langle b_n \rangle + \nu_n \right) - 1 \right] \end{aligned}$$

since $\delta_n^2 = \delta_n$. Now, if $\delta_n = 1$,

$$0 \le 2\left(\langle b_n \rangle + \nu_n \right) - 1 \le 2\nu_n \le 4N^{-b}.$$

Thus

$$0 \le \frac{1}{N} \sum_{n=1}^{N} \delta_n \left[2\left(\langle b_n \rangle + \nu_n \right) - 1 \right] \le 4N^{-b}$$

and so

$$0 \le \frac{1}{N} E \sum_{n=1}^{N} \delta_n \left[2\left(\langle b_n \rangle + \nu_n \right) - 1 \right] \le 4N^{-b}.$$

Also

$$\begin{aligned} \left| \nu_n^2 + 2\nu_n \langle b_n \rangle - E\left[\nu_n^2 + 2\nu_n \langle b_n \rangle \right] \right| &\le 2\nu_n \left| \langle b_n \rangle \right| + 2E\nu_n \left| \langle b_n \rangle \right| \\ &\le 4N^{-b}. \end{aligned}$$

Hence

$$\sup_{(\lambda,\phi)\in B_{jk}} |V_N(\lambda,\phi) - V_N(\lambda_j,\phi_k)| \le 12N^{-b}$$

and the result follows as the bound does not depend on j or k. ■

Theorem 2: $\sup_{\lambda,\phi} |V_N(\lambda,\phi)| \to 0$, almost surely as $N \to \infty$.

Proof: The proof follows immediately from the two previous lemmas. ■

Lemma 3: Let $g : [-0.5, 0.5) \to \mathbb{R}$ be given by

$$g(x) = E\langle X_n + x \rangle^2 - \sigma^2.$$

Then $g(x) \geq 0$, with equality if and only if $x = 0$.

Proof: If $x \geq 0$

$$\begin{aligned} g(x) &= \int_{-1/2}^{1/2-x} (x+y)^2 f_X(y)dy \\ &\quad + \int_{1/2-x}^{1/2} (x+y-1)^2 f_X(y)dy - \int_{-1/2}^{1/2} y^2 f_X(y)dy \\ &= x^2 + \int_{1/2-x}^{1/2} (1-2x-2y) f_X(y)dy \end{aligned}$$

since $E(X) = 0$. Similarly, when $x < 0$

$$g(x) = x^2 + \int_{-1/2}^{-1/2-x} (1+2x+2y) f_X(y)dy = g(-x)$$

and therefore g is even. Now, since $f_X(y)$ is even, when $x \geq 0$

$$\begin{aligned} g'(x) &= 2x - 2\left[1 - F_X\left(\frac{1}{2} - x\right)\right] \\ &= 2x - 2F_X\left(x - \frac{1}{2}\right). \end{aligned}$$

Since f_X is symmetric and unimodal with mode at 0, $F_X(-0.5) = 0$, $F_X(0) = 1/2$ and $F_X(x - 0.5)$ is strictly convex on [0, 0.5). It follows that g is monotonically increasing on [0, 0.5) and, being even, is monotonically decreasing on $[-0.5, 0)$. Also $g(0) = 0$. Thus $g(x) \geq 0$, with equality if and only if $x = 0$. ∎

Let

$$(\widehat{\lambda}_N, \widehat{\phi}_N) = \arg\min S_N(\lambda, \phi) \tag{8}$$

and put $\tau_N(\lambda, \phi) = ES_N(\lambda, \phi)$. Now

$$\tau_N(\lambda, \phi) - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} g\left(\langle \lambda n + \phi \rangle\right) \geq 0.$$

Thus, since

$$S_N(\widehat{\lambda}_N, \widehat{\phi}_N) \leq S_N(0, 0)$$

we have

$$V_N(\widehat{\lambda}_N, \widehat{\phi}_N) + \tau_N(\widehat{\lambda}_N, \widehat{\phi}_N) \leq V_N(0, 0) + \sigma^2$$

and so

$$0 \leq \tau_N(\widehat{\lambda}_N, \widehat{\phi}_N) - \sigma^2 \leq -V_N(\widehat{\lambda}_N, \widehat{\phi}_N) + V_N(0, 0).$$

However, from Theorem 2, the right hand side converges almost surely to 0 as $N \to \infty$. Hence

$$\tau_N(\widehat{\lambda}_N, \widehat{\phi}_N) - \sigma^2 \to 0 \tag{9}$$

almost surely as $N \to \infty$.

Lemma 4: Let K be a subset of the integers $W_N = \{1, 2, \ldots, N\}$ for which $\#K > 3N/4$ where $\#K$ is the cardinality of K. Then[2]

$$W_{N/2} \subset K - K = \{k - k'; k, k' \in K\}.$$

Proof: Suppose this is not the case. Then there is some $r \in W_{N/2}$ for which $r \notin K - K$, so that

$$K \cap (K + r) = \emptyset.$$

Let $K_{N/2} = K \cap W_{N/2}$. Then $\#K_{N/2} > N/4$, and the same is true of $K \cap [r+1, r+N/2]$. Since both $r + K_{N/2}$ and $K \cap [r+1, r+N/2]$ are subsets of $[r+1, r+N/2]$, are disjoint, and each has more than $N/4$ elements, we have a contradiction. ∎

Lemma 5: Let $|\langle a \rangle| < \delta$ and $|\langle b \rangle| < \delta$. Then $|\langle a + b \rangle| < 2\delta$.

Proof: If $\delta > 1/4$, the proof is trivial as $|\langle a + b \rangle| \leq 1/2$ for all $a, b \in \mathbb{R}$. If $\delta \leq 1/4$ then $\langle a \rangle + \langle b \rangle = \langle a + b \rangle$ and by the triangle inequality

$$|\langle a + b \rangle| = |\langle a \rangle + \langle b \rangle| \leq |\langle a \rangle| + |\langle b \rangle| < 2\delta.$$

∎

Lemma 6: Suppose (λ_N) and (ϕ_N) are sequences with $(\lambda_N, \phi_N) \in B$ and with

$$\frac{1}{N} \sum_{n=1}^{N} g\left(\langle n\lambda_N + \phi_N \rangle\right) \to 0$$

where $g(x)$ is continuous and even, and $g(x) \geq 0$, with equality if and only if $\langle x \rangle = 0$. Then $N\lambda_N \to 0$ and $\phi_N \to 0$.

Proof: For any $\delta > 0$, there exists N_0 such that if $N > N_0$ and $K_N = \{n \leq N; |\langle n\lambda_N + \phi_N| < \delta\}$ then $\#K_N > 3N/4$. Choose $\delta < 1/8$. By Lemma 4

$$\{1, 2, \ldots, N/2\} \subset K_N - K_N.$$

Let $m, n \in K_N$. Then

$$\begin{aligned} |\langle n\lambda_N + \phi_N \rangle| &< \delta, \\ |\langle m\lambda_N + \phi_N \rangle| &< \delta \end{aligned}$$

and from Lemma 5

$$|\langle (m-n)\lambda_N \rangle| < 2\delta.$$

Thus, for $k \in \{1, 2, \ldots, N/2\}$

$$|\langle k\lambda_N \rangle| < 2\delta.$$

In particular, $|\lambda_N| < 2\delta < 1/4$, and so $2\lambda_N = \langle 2\lambda_N \rangle$. Putting $k = 2$ above, we therefore have $|2\lambda_N| < 2\delta$, and so $|\lambda_N| < \delta$. Hence $|4\lambda_N| < 4\delta < 1/2$, and so $4\lambda_N = \langle 4\lambda_N \rangle$. Continuing in this way, we have $|2^r \lambda_N| < 2\delta$, for all r such that $2^r \leq N/2$. Let r be such that $2^r \leq N/2$ and $2^{r+1} > N/2$. Then

$$|N\lambda_N| < |2^{r+2}\lambda_N| < 8\delta.$$

Since δ is arbitrary, it follows that $N\lambda_N \to 0$ as $N \to \infty$.

[2]The notation $K - K$ should not be confused with set subtraction.

Let $n \in K_N$. From the above argument $n\lambda_N \to 0$ and $|\langle n\lambda_N + \phi_N\rangle| < \delta$. Consequently $|\phi_N| < \delta$ and, therefore, $\phi_N \to 0$. ∎

Theorem 3: Let $\widehat{\lambda}_N$ and $\widehat{\phi}_N$ be given by (8). Then $N\widehat{\lambda}_N \to 0$ and $\widehat{\phi}_N \to 0$, almost surely as $N \to \infty$.

Proof: Let A be the subset of the sample space Ω on which $\tau_N(\widehat{\lambda}_N, \widehat{\phi}_N) \to \sigma^2$, and let (λ_N, ϕ_N) be $(\widehat{\lambda}_N, \widehat{\phi}_N)$ at some point in A. Note that $P(A) = 1$. Lemma 6, applied to the function g defined in Lemma 3 shows that $N\lambda_N \to 0$ and $\phi_N \to 0$. The proof follows. ∎

We are now in a position to complete the derivation of the asymptotic properties.

Proof: (Theorem 1) Strong consistency follows directly from Theorem 3. The proof of the central limit theorem is along different lines from usual proofs, as the second derivatives of (6) do not exist everywhere in a neighbourhood of the true values. Let

$$T_N(\psi, \phi) = S_N(\psi/N, \phi) = \frac{1}{N}\sum_{n=1}^{N} \langle X_n + n\psi/N + \phi\rangle^2.$$

In view of the strong consistency, it is only the local behavior near $(\psi, \phi) = (0, 0)$ which is relevant. We shall assume in what follows that $|\psi| + |\phi| < 1$. It is shown in Lemma 10 in the Appendix that

$$\begin{bmatrix} \frac{\partial T_N}{\partial \psi} \\ \frac{\partial T_N}{\partial \phi} \end{bmatrix} = \frac{2}{N}\sum_{n=1}^{N} \begin{bmatrix} n/N \\ 1 \end{bmatrix} X_n + 2\left(1 - h + o(1)\right)\mathbf{A}\begin{bmatrix} \psi \\ \phi \end{bmatrix} + \left(|\psi| + |\phi|\right) O_P(N^{-1/2}). \quad (10)$$

where $h = f_X(-0.5)$ and $\mathbf{A}$ is the 2×2 matrix

$$\mathbf{A} = N^{-1}\sum_{n=1}^{N} \begin{bmatrix} (n/N)^2 & n/N \\ n/N & 1 \end{bmatrix}.$$

Put

$$\begin{bmatrix} \psi_N \\ \phi_N \end{bmatrix} = -\frac{1}{N(1-h)}\mathbf{A}^{-1}\sum_{n=1}^{N} \begin{bmatrix} n/N \\ 1 \end{bmatrix} X_n.$$

Now

$$N^{-1/2}\sum_{n=1}^{N} \begin{bmatrix} n/N \\ 1 \end{bmatrix} X_n$$

is asymptotically normal with mean 0 and covariance matrix

$$\sigma^2 \lim_{N\to\infty} \mathbf{A} = \sigma^2 \begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

where $\sigma^2 = \mathrm{var} X_n$. Thus $N^{1/2}[\psi_N \ \phi_N]'$ is asymptotically normal with mean 0 and covariance matrix

$$\frac{\sigma^2}{(1-h)^2}\begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix}^{-1} = \frac{12\sigma^2}{(1-h)^2}\begin{bmatrix} 1 & -1/2 \\ -1/2 & 1/3 \end{bmatrix}.$$

Hence, since ψ_N and ϕ_N are $O_P(N^{-1/2})$, and

$$\mathbf{A}^{-1}\begin{bmatrix} \frac{\partial T_N}{\partial \psi} \\ \frac{\partial T_N}{\partial \phi} \end{bmatrix} = -2(1-h)\left(\begin{bmatrix} \psi_N \\ \phi_N \end{bmatrix} - \begin{bmatrix} \psi \\ \phi \end{bmatrix}\right) + 2o(1)\begin{bmatrix} \psi \\ \phi \end{bmatrix} + \left(|\psi| + |\phi|\right) O_P(N^{-1/2})$$

it follows that

$$\begin{bmatrix} \frac{\partial T_N}{\partial \psi} \\ \frac{\partial T_N}{\partial \phi} \end{bmatrix} = 0 \quad (11)$$

only if

$$\begin{bmatrix} \psi \\ \phi \end{bmatrix} = \begin{bmatrix} \psi_N \\ \phi_N \end{bmatrix}\left(1 + o_P(1)\right).$$

The result follows, since $N\widehat{\lambda}_N = N(f_0 - \widehat{f}_N)$ and $\widehat{\phi}_N = \theta_0 - \widehat{\theta}_N$ and since (11) holds at $(\psi, \phi) = (N\widehat{\lambda}_N, \widehat{\phi}_N)$ by Lemma 11 contained in the Appendix. ∎

IV. Least Squares Unwrapping and the Nearest Lattice Point Problem

In this section, we describe methods to compute the LSPUE. We find that the computational problem can be transformed into a nearest lattice point problem [15] in an $N-2$ dimensional lattice to be specified shortly. We first require some concepts from lattice theory.

The set L is said to be a lattice, with *generator* or *basis* matrix $\mathbf{B}$ if [20]

$$L = \{\mathbf{p} \in \mathbb{R}^n | \mathbf{p} = \mathbf{B}\mathbf{w}, \mathbf{w} \in \mathbb{Z}^n\}.$$

Vectors and matrices are written in bold font and $'$ is used to denote transpose. A fundamental problem in lattice theory is the *nearest lattice point problem*. The nearest lattice point problem is, given $\mathbf{q} \in \mathbb{R}^n$ and some lattice L whose lattice points lie in $\mathbb{R}^n$, to find that lattice point $\mathbf{p} \in L$ for which the Euclidean distance between $\mathbf{q}$ and $\mathbf{p}$ is minimized. The notation $\mathrm{NPt}(\mathbf{q}, L)$ is used to denote the nearest point in L to $\mathbf{q}$. We assume that, when two or more lattice points are of equal distance to $\mathbf{q}$, $\mathrm{NPt}(\mathbf{q}, L)$ selects one of the lattice points in a systematic manner.

The nearest lattice point problem is known to be NP-hard under certain conditions when the lattice itself, or rather a basis thereof, is considered as an additional input parameter [21]–[23]. Nevertheless, algorithms exist that can compute the nearest lattice point in reasonable time if the dimension is small [15], [24], [25]. One such algorithm introduced by Pohst [25] in 1981 was popularized in the signal processing and communications fields by Viterbo and Boutros [24] and has since been called the *sphere decoder*. Approximate algorithms for computing the nearest point have also been studied. One example is Babai's nearest plane algorithm [17], which requires $O(n^4)$ arithmetic operations in the worst case. For specific lattices where the generator matrix is known *a priori*, many fast nearest point algorithms are known [26]–[31].

We now show how the LSPUE can be represented as a nearest lattice point problem in a lattice determined by N. Define the N-dimensional vectors $\mathbf{n} = [1\ 2\ \cdots\ N]'$, $\mathbf{1} = [1\ 1\ \cdots\ 1]'$, $\mathbf{y} = [Y_1\ Y_2\ \cdots\ Y_N]'$, $\mathbf{x} = [X_1\ X_2\ \cdots\ X_N]'$ and $\mathbf{u} = [U_1\, U_2\ \cdots\ U_N]'$. The sum of squares function (5) can be written in vector form as

$$\|\mathbf{y} - f\mathbf{n} - \theta\mathbf{1} - \mathbf{u}\|^2. \tag{12}$$

Define the matrix $\mathbf{M} = [\mathbf{n}\ \mathbf{1}]$. The least squares estimator is then

$$\begin{bmatrix}\widehat{f}_n\\ \widehat{\theta}_n\end{bmatrix} = \arg \min_{(f,\theta)\in[-0.5,0.5)^2} \min_{\mathbf{u}\in\mathbb{Z}^N} \left\|\mathbf{y} - \mathbf{M}\begin{bmatrix}f\\ \theta\end{bmatrix} - \mathbf{u}\right\|^2. \tag{13}$$

Given $\mathbf{u}$, the least squares estimators of f and θ are obtained by the usual linear regression formulae and are given by

$$\begin{bmatrix}\widehat{f}_n\\ \widehat{\theta}_n\end{bmatrix} = \mathbf{M}^+(\mathbf{y} - \mathbf{u}) \tag{14}$$

where $\mathbf{M}^+ = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$. Substituting (14) into (13) and rearranging, the least squares estimator of $\mathbf{u}$, given f and θ, is

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}\in\mathbb{Z}^N} \|\mathbf{B}(\mathbf{y} - \mathbf{u})\| \tag{15}$$

where $\mathbf{B} = \mathbf{I} - \mathbf{M}\mathbf{M}^+$ and $\mathbf{I}$ is the $N \times N$ identity matrix.

Let Λ be the lattice with generator matrix $\mathbf{B}$. The least squares phase unwrapping is then the $\hat{\mathbf{u}}$ for which $\mathbf{B}\hat{\mathbf{u}}$ is the nearest lattice point in Λ to $\mathbf{B}\mathbf{y}$. The least squares estimate (13) can be computed by first computing $\mathrm{NPt}(\mathbf{B}\mathbf{y}, \Lambda)$, producing both $\mathbf{B}\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}$. $\widehat{f}_N$ and $\widehat{\theta}_N$ can then be computed using (14).

The most difficult part of the procedure is finding the nearest lattice point. One possible solution is to use the sphere decoder. Unfortunately the sphere decoder has worst case exponential complexity and therefore is only computationally feasible when N is small. However, the lattice Λ is not random and has a generator matrix with significant structure. Using this structure we can construct a polynomial-time algorithm to compute the nearest point.

Note that the nearest point is given by $\mathbf{B}\hat{\mathbf{u}}$ where

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}\in\mathbb{Z}^N} \min_{(f,\theta)\in[-0.5,0.5)^2} \|\mathbf{y} - f\mathbf{n} - \theta\mathbf{1} - \mathbf{u}\|^2. \tag{16}$$

Fixing both f and $\mathbf{u}$ and minimising with respect to θ we obtain

$$\theta = \frac{\mathbf{1}'(\mathbf{y} - f\mathbf{n} - \mathbf{u})}{N}. \tag{17}$$

Substituting this into (16) we find that

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}\in\mathbb{Z}^N} \min_{f\in[-0.5,0.5)^2} \|\mathbf{Q}\mathbf{y} - f\mathbf{Q}\mathbf{n} - \mathbf{Q}\mathbf{u}\|^2 \tag{18}$$

where $\mathbf{Q}$ is the projection matrix

$$\mathbf{Q} = \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{\mathbf{1}'\mathbf{1}}\right). \tag{19}$$

The matrix $\mathbf{Q}$ is the generator matrix for a well studied lattice called A^*_{N-1} [20]. Numerous nearest point algorithms exist for the lattice A^*_N [26], [28], [30], [32]. The fastest known algorithm was described by McKilliam *et al.* and requires $O(N)$ arithmetic operations [30].

Let $\mathbf{z} = \mathbf{Q}\mathbf{y}$, $\mathbf{g} = \mathbf{Q}\mathbf{n}$ and $\boldsymbol{\zeta} = \mathbf{Q}\mathbf{u}$. Given f, (18) is minimized when

$$\boldsymbol{\zeta} = \mathrm{NPt}\left(\mathbf{z} - f\mathbf{g}, A^*_{N-1}\right).$$

Conversely, given $\boldsymbol{\zeta}$, (18) is minimized by putting

$$f = \frac{\mathbf{g}'(\mathbf{z} - \boldsymbol{\zeta})}{\mathbf{g}'\mathbf{g}}.$$

If we define the set

$$S = \left\{\mathrm{NPt}\left(\mathbf{z} - f\mathbf{g}, A^*_{N-1}\right); f \in [-0.5, 0.5)\right\} \tag{20}$$

then the nearest point is given by $\mathbf{B}\hat{\mathbf{u}}$ where $\hat{\mathbf{u}}$ satisfies

$$\mathbf{Q}\hat{\mathbf{u}} = \arg \min_{\boldsymbol{\zeta}\in S} h(\boldsymbol{\zeta}) \tag{21}$$

and $h : \mathbb{R}^N \mapsto \mathbb{R}$ is defined by

$$h(\boldsymbol{\zeta}) = \left\|\mathbf{z} - \frac{\mathbf{g}'(\mathbf{z} - \boldsymbol{\zeta})}{\mathbf{g}'\mathbf{g}}\mathbf{g} - \boldsymbol{\zeta}\right\|^2.$$

The set S contains the lattice points in A^*_{N-1} that are nearest to some point on the line segment defined by $\mathbf{z} - f\mathbf{g}$ where $f \in [-0.5, 0.5)$.

Note that $\#S$ is finite. We require an algorithm to find the lattice points in S. The nearest point can then be found by testing each lattice point and returning the minimizer according to (21). Computing S directly proves to be difficult, and, in fact, it is easier to compute a superset of S. To show this we require some properties of the lattice A^*_{N-1}. The generator matrix for A^*_{N-1} is the matrix $\mathbf{Q}$ defined in (19), which is the projection matrix into the hyperplane orthogonal to $\mathbf{1}$. Let H denote this hyperplane. Then A^*_{N-1} consists of the lattice points in $\mathbb{Z}^N$ projected orthogonally onto H. That is

$$A^*_{N-1} = \{\mathbf{Q}\mathbf{w} | \mathbf{w} \in \mathbb{Z}^N\}.$$

Another way of representing A^*_{N-1} is as a union of N translates of its dual lattice $A_{N-1} = \mathbb{Z}^N \cap H$ [20]. That is

$$\begin{aligned} A^*_{N-1} &= \bigcup_{i=0}^{n-1} ([i] + A_{N-1}) \\ &= \bigcup_{i=0}^{n-1} \left([i] + \mathbb{Z}^N\right) \bigcap H \qquad (22)\\ &\subset \bigcup_{i=0}^{n-1} \left([i] + \mathbb{Z}^N\right) \qquad (23)\end{aligned}$$

where the $[i]$ are known as *glue vectors* and are defined in this case as

$$[i] = \frac{1}{N}[\underbrace{i, \ldots, i}_{j \text{ times}}, \underbrace{-j, \ldots, -j}_{i \text{ times}}] \tag{24}$$

for $i = 0, 1, \ldots, N-1$ with $i + j = N$ [20, p. 109].

There exists a fast algorithm to find the points in the set

$$K(i) = \left\{\mathrm{NPt}\left(\mathbf{z} - f\mathbf{g}, [i] + \mathbb{Z}^N\right) | f \in \left[-\frac{1}{2}, \frac{1}{2}\right)\right\}$$

due to Ryan *et al.* [33], [34]. We use this algorithm for each $i = 0, 1, \ldots, N-1$. Let

$$S^+ = \bigcup_{i=0}^{n-1} K(i). \tag{25}$$

The proof of the following lemma is not difficult and is omitted.

Lemma 7: The minimizer of h over S is equal to the minimizer of h over S^+. That is

$$\mathbf{Q}\hat{\mathbf{u}} = \arg\min_{\boldsymbol{\zeta}\in S} h(\boldsymbol{\zeta}) = \arg\min_{\boldsymbol{\zeta}\in S^+} h(\boldsymbol{\zeta}).$$

It is thus sufficient to use the set S^+ rather than S to find the nearest lattice point. Since S^+ is the union of the $K(i)$,

$$\mathbf{Q}\hat{\mathbf{u}} = \arg\min_{\boldsymbol{\zeta}\in K(i)} \min_{i=0,1,\ldots,N-1} h(\boldsymbol{\zeta}).$$

For each i the minimization can be computed using the algorithm in [33], [34]. The number of operations required is $O(\#S^+ \log N)$.

Lemma 8: For $i = 0, 1, \ldots, N$

$$\#K(i) = O(N^2).$$

Proof: The 'first' lattice point in $K(i)$ corresponds to $f = -1/2$, and is given by

$$\mathbf{v} = \left\lfloor \mathbf{z} + \frac{\mathbf{g}}{2} - [i] \right\rceil + [i]$$

while the "last" lattice point in $K(i)$ corresponds to $f = 1/2$, and is given by

$$\mathbf{w} = \left\lfloor \mathbf{z} - \frac{\mathbf{g}}{2} - [i] \right\rceil + [i].$$

Consecutive elements in $K(i)$ satisfy

$$\mathbf{v}_{\text{next}} = \mathbf{v}_{\text{previous}} + \mathtt{sign}(g_j)\mathbf{e}_j$$

for some j, where $\mathbf{e}_j$ is a vector conistsing of zeros apart from a one in the jth position and $\mathtt{sign}(g_j)$ is 1 if $g_j > 0$, -1 if $g_j < 0$ and 0 if $g_j = 0$. Thus

$$\#K(i) = \sum_{n=1}^{N} |v_n - w_n| \le \sum_{n=1}^{N} |w_n| + |v_n|.$$

Since z_n and $[i]_n \in [-0.5, 0.5)$, and $|g_n| \le N$, $|v_n|$, $|w_n| \le N+2$ for all $n = 1, 2, \ldots, N$. Hence

$$\#K(i) \le \sum_{n=1}^{N} (2N+4) O(N^2).$$

■

As a consequence of the lemma, $\#S^+ = O(N^3)$ and the algorithm requires $O(N^3 \log N)$ arithmetic operations.

V. Simulations

We have compared the performance of five estimators: the periodogram estimator [2]; the LSPUE; the parabolic, smoothed central finite difference estimator (PSCFD) [12]; Kay's window estimator [9]; and the Quinn-Fernandes estimator [35], [36].

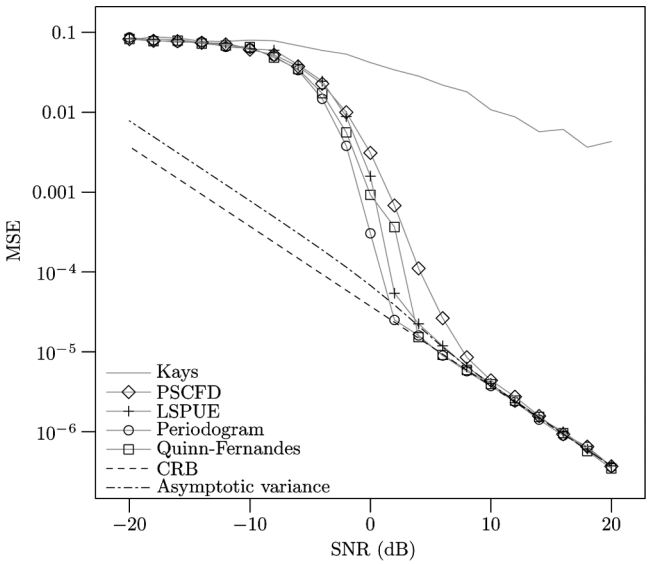

Fig. 1. MSE in frequency versus SNR when $N = 16$.

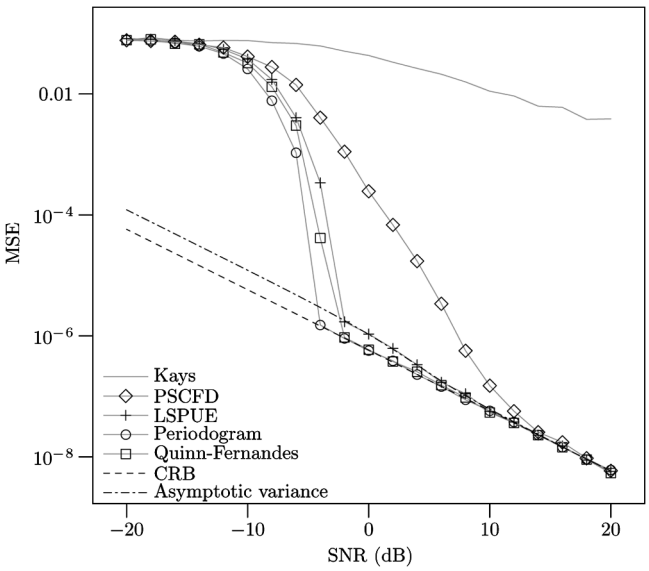

Fig. 2. MSE in frequency versus SNR when $N = 64$.

Those estimators based on phase unwrapping are the LSPUE, Kay's estimator and the PSCFD estimator. Five simulations were run with $N = 16$, $N = 64$, $N = 256$, $N = 512$, and $N = 1024$ (Figs. 1 to 5, respectively), each with SNR varied between -20 and 20 dB, and 1000 trials were run for each SNR value. The value of (f_0, θ_0) was varied uniformly in the range $[-0.5, 0.5)^2$. The distribution of the s_n was assumed to be complex i.i.d. and Gaussian with variance σ^2. This gave an SNR of $10\log_{10}(A^2/2\sigma^2)$ dB.

Standard behavior was observed for the "nonlinear" estimators. The mean square error (MSE) was large below a particular threshold SNR. Above that threshold, the estimators appeared to converge to the Cramer-Rao lower bound (CRB) [2] depicted by the dashed line. It is clear that the periodogram estimator produces the most statistically accurate results, with the LSPUE and Quinn-Fernandes estimators only marginally worse. Kay's

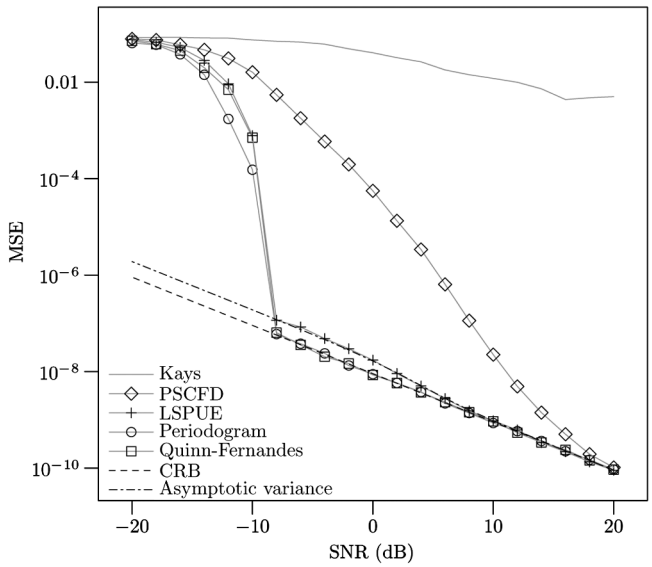

Fig. 3. MSE in frequency versus SNR when $N = 256$.

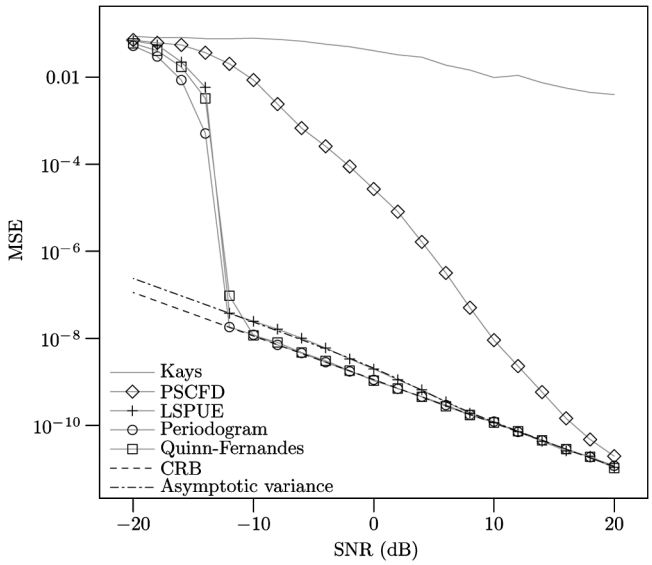

Fig. 4. MSE in frequency versus SNR when $N = 512$.

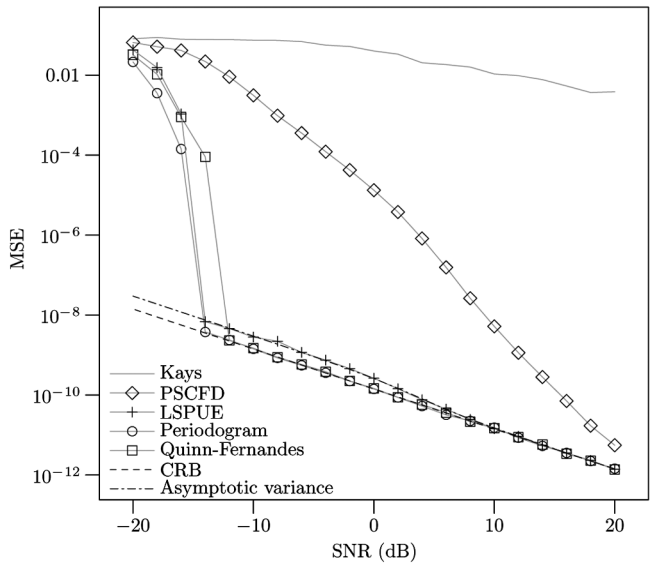

Fig. 5. MSE in frequency versus SNR when $N = 1024$.

TABLE I
COMPUTATION TIME IN SECONDS FOR 10^5 TRIALS

Algorithm	n=16	n=64	n=256	n=512	n=1024
Kay	0.156	0.625	2.406	4.813	9.641
PSCFD	0.141	0.547	2.094	4.187	8.375
Quinn-Fernandes	0.438	1.297	4.469	8.985	21.141
Periodogram	0.437	1.656	7.578	17.828	49.157
LSPUE	5.578	346.2	$> 10^4$	$> 10^5$	$> 10^6$

estimator and the PSCFD perform comparatively poorly, particularly for large N. Kay's estimator performs particularly poorly as it fails to correctly estimate f_0 when it is near 0.5, i.e., when f_0 is near the branch cut on the unit circle [11]. As noted elsewhere [10], the performance of Kay's estimator is improved if f_0 is bounded away from 0.5.

The dash-dotted line is the asymptotic variance of the LSPUE derived in Theorem 1. It can be seen that, provided the SNR is high enough to avoid the threshold effect, the performance of the LSPUE closely agrees with the asymptotic results. Note that the asymptotic variance of the LSPUE is larger than the CRB. This performance loss can be overcome by using a numerical optimization procedure, such as Newton's method, starting at the estimate given by the LSPUE. In order to show the correctness of our asymptotic theory we have not displayed these results here.

Table I shows the computation time for 10^5 trials of each estimator for $N = 16, 64, 256, 512, 1024$. The computer used is a 2.13-Ghz Intel Core2. As expected the LSPUE is significantly slower than other estimators. The computational complexity of our LSPUE algorithm (Section IV) has order $O(N^3 \log N)$ whereas the other estimators have complexity $O(N)$ or $O(N \log N)$. For this reason, the periodogram or Quinn-Fernandes estimators are to be preferred in practice. Nevertheless, it may be that significantly faster algorithms exist to compute the LSPUE.

VI. CONCLUSION

We have discussed single frequency estimation via least squares phase unwrapping. This estimator has been shown to be strongly consistent and its central limit theorem derived. The problem of computing the least squares phase unwrapping has been demonstrated to be related to a problem in algorithmic number theory known as the nearest lattice point problem. We have derived an algorithm that computes the least squares estimate in $O(N^3 \log N)$ arithmetic operations where N is the sample size. The complexity is high when compared with other single frequency estimators and arises from the need to solve the nearest lattice point problem in the lattice Λ derived in Section IV. One possible algorithm has been described here. However, it may be that much faster nearest point algorithms exist for this specific lattice.

We have compared the performance of the LSPUE and the periodogram estimator [2] by Monte Carlo simulation. It was

found that the LSPUE is marginally less accurate than the periodogram estimator and significantly more accurate than other estimators based on phase unwrapping. The simulations agree with the theoretical central limit theorem derived in Section III.

APPENDIX

Lemma 9: Let $Z_1, \ldots, Z_N$ be independent, zero-mean random variables with $|Z_j| \leq 1$. Then, for any integer $\beta > 0$, as $N \to \infty$

$$S = E(Z_1 + \cdots + Z_n)^{2\beta} = O(N^\beta).$$

Proof: We can write S according to the multinomial expansion

$$S = \sum_{k_1+\cdots+k_n=2\beta} \binom{2\beta}{k_1, \ldots, k_n} \prod_{j=1}^{N} E\left(Z_j^{k_j}\right) \tag{26}$$

where we make use of the multinomial coefficients

$$\binom{2\beta}{k_1, \ldots, k_N} = \frac{(2\beta)!}{k_1!, \ldots, k_N!}.$$

Now, because the Z_j have zero mean, the product in (26) is zero if any $k_j = 1$. Accordingly, we define the set

$$\mathcal{K} = \left\{ \mathbf{k} \in \mathbb{Z}^n | k_j \geq 0, k_j \neq 1, \sum_{j=1}^{N} k_j = 2\beta \right\} \subset [0, 2\beta]^N.$$

In view of the fact that the Z_j are bounded with $|Z_j| \leq 1$, we then have

$$|S| \leq \sum_{\mathbf{k} \in \mathcal{K}} \binom{2\beta}{\mathbf{k}} \tag{27}$$

Let $c(\mathbf{k})$ be the number of nonzero elements in the vector $\mathbf{k}$. Since $k_j \neq 1$ and the k_j sum to 2β, it follows that $c(\mathbf{k}) \leq \beta$ for all $\mathbf{k} \in \mathcal{K}$. Clearly, in addition, $c(\mathbf{k}) \geq 1$, and so from (27) we have

$$|S| \leq \sum_{d=1}^{\beta} \sum_{\substack{\mathbf{k} \in \mathcal{K} \\ c(\mathbf{k})=d}} \binom{2\beta}{\mathbf{k}} \leq \sum_{d=1}^{\beta} \sum_{\substack{\mathbf{k} \in \mathcal{K} \\ c(\mathbf{k})=d}} \frac{(2\beta)!}{2^d} \leq \sum_{d=1}^{\beta} \sum_{\substack{\mathbf{k} \in [0,2\beta]^N \\ c(\mathbf{k})=d}} \frac{(2\beta)!}{2^d}$$

Consider the number of integer vectors $\mathbf{k}$ that satisfy the conditions of the innermost sum, i.e., that $\mathbf{k} \in [0, 2\beta]^N$ and $c(\mathbf{k}) = d$. There are $\binom{N}{d}$ ways of selecting the indices of nonzero k_j so that $c(\mathbf{k}) = d$ and then 2β possibilities for the value of each such nonzero k_j. Therefore, there are $(2\beta)^d \binom{N}{d}$ possible vectors in total. Hence

$$\begin{aligned} |S| &\leq (2\beta)! \sum_{d=1}^{\beta} \binom{N}{d} \beta^d \leq (2\beta)! \sum_{d=1}^{\beta} (N\beta)^d \\ &\leq (2\beta)! \beta (N\beta)^\beta = O(N^\beta). \end{aligned}$$

■

Lemma 10: Assume that $|\psi| + |\phi| < 1$. Then (10) holds.

Proof:

$$\begin{aligned} \begin{bmatrix} \frac{\partial T_N}{\partial \psi} \\ \frac{\partial T_N}{\partial \phi} \end{bmatrix} &= \frac{2}{N} \sum_{n=1}^{N} \begin{bmatrix} n/N \\ 1 \end{bmatrix} (X_n + n\psi/N + \phi - I_{n,N}) \\ &= \frac{2}{N} \sum_{n=1}^{N} \begin{bmatrix} n/N \\ 1 \end{bmatrix} [X_n + n\psi/N + \phi - E(I_{n,N}) \\ &\qquad + E(I_{n,N}) - I_{n,N}] \end{aligned}$$

where since $\forall n \ |n\psi/N + \phi| < 1$

$$I_{n,N} = \begin{cases} 1; & X_n + n\psi/N + \phi \geq \frac{1}{2} \\ -1; & X_n + n\psi/N + \phi < -\frac{1}{2} \\ 0; & \text{otherwise.} \end{cases}$$

Now, if $n\psi/N + \phi > 0$,

$$\begin{aligned} E(I_{n,N}) &= P\left(X_n + n\psi/N + \phi \geq \frac{1}{2}\right) \\ &= \int_{\frac{1}{2} - (n\psi/N + \phi)}^{\frac{1}{2}} f_X(x) dx \\ &= (n\psi/N + \phi) f_X(\xi_{n,N}) \end{aligned} \tag{28}$$

where $1/2 - (n\psi/N + \phi) \leq \xi_{n,N} < 1/2$, while, if $n\psi/N + \phi < 0$

$$\begin{aligned} E(I_{n,N}) &= -P(X_n + n\psi/N + \phi < -1/2) \\ &= (n\psi/N + \phi) f_X(\xi_{n,N}) \end{aligned} \tag{29}$$

where $-1/2 \leq \xi_{n,N} \leq -(1/2) - (n\psi/N + \phi)$. Thus, if $n\psi/N + \phi > 0$,

$$\text{var} I_{n,N} = E(I_{n,N}) - [E(I_{n,N})]^2 < (n\psi/N + \phi) f_X(\xi_{n,N})$$

while, if $n\psi/N + \phi < 0$,

$$\text{var} I_{n,N} < -(n\psi/N + \phi) f_X(\xi_{n,N}).$$

Hence

$$\begin{aligned} \text{var}\, N^{-\frac{1}{2}} \sum_{n=1}^{N} [I_{n,N} - E(I_{n,N})] &< \frac{4}{N} \sum_{n=1}^{N} |n\psi/N + \phi| f_X(\xi_{n,N}) \\ &< \frac{4}{N} (|\psi| + |\phi|) \sum_{n=1}^{N} f_X(\xi_{n,N}). \end{aligned}$$

Similarly

$$\begin{aligned} &\text{var}\, N^{-\frac{1}{2}} \sum_{n=1}^{N} \frac{n}{N} [I_{n,N} - E(I_{n,N})] \\ &< \frac{4}{N} \sum_{n=1}^{N} \left(\frac{n}{N}\right)^2 |n\psi/N + \phi| f_X(\xi_{n,N}) \\ &< \frac{4}{N} (|\psi| + |\phi|) \sum_{n=1}^{N} f_X(\xi_{n,N}). \end{aligned}$$

Since f_X is symmetric and unimodal, with mode at 0, there exists a unique ξ, for which $f_X(x) < 1$ when $|x| > \xi$. Thus, as long as $|\psi| + |\phi| < (1/2) - \xi$

$$\frac{4}{N}(|\psi|+|\phi|)\sum_{n=1}^{N} f_X(\xi_{n,N}) < 4(|\psi|+|\phi|)$$

and so

$$N^{-1}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix}[I_{n,N}-E(I_{n,N})] = (|\psi|+|\phi|)\,O_P(N^{-1/2}).$$

But, using (28) and (29), we obtain

$$\begin{aligned}
&\frac{2}{N}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix}[X_n+n\psi/N+\phi-E(I_{n,N})] \\
&\quad=\frac{2}{N}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix}[X_n+n\psi/N+\phi-(n\psi/N+\phi)f_X(\xi_{n,N})] \\
&\quad=\frac{2}{N}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix}[X_n+(n\psi/N+\phi)(1-f_X(\xi_{n,N}))]
\end{aligned}$$

and so

$$\begin{aligned}
\begin{bmatrix} \frac{\partial T_N}{\partial \psi} \\ \frac{\partial T_N}{\partial \phi} \end{bmatrix} &= \frac{2}{N}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix} X_n \\
&\quad - \frac{2}{N}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix}[I_{n,N}-E(I_{n,N})] \\
&\quad + \frac{2}{N}\sum_{n=1}^{N}(1-f_X(\xi_{n,N}))\begin{bmatrix} (n/N)^2 & n/N \\ n/N & 1 \end{bmatrix}\begin{bmatrix} \psi \\ \phi \end{bmatrix} \\
&= \frac{2}{N}\sum_{n=1}^{N}\begin{bmatrix} n/N \\ 1 \end{bmatrix} X_n + (|\psi|+|\phi|)\,O_P(N^{-1/2}) \\
&\quad + \frac{2}{N}\sum_{n=1}^{N}(1-f_X(-1/2)+o(1)) \\
&\quad \times \begin{bmatrix} (n/N)^2 & n/N \\ n/N & 1 \end{bmatrix}\begin{bmatrix} \psi \\ \phi \end{bmatrix}
\end{aligned}$$

where, because $|n\psi/N + \phi| \to 0$ as $N \to \infty$, $\xi_{n,N}$ approaches either 1/2 or $-1/2$ for all n and, therefore, $f_X(\xi_{n,N}) \to f_X(-1/2)$. ■

Lemma 11: The partial derivatives of (6) are zero with probability 1 at $(\widehat{f}_N, \widehat{\theta}_N)$, the minimizer of (6) over B.

Proof: Let

$$Z(f,\theta) = \sum_{n=1}^{N}\left(Y_n - nf - \theta - \lfloor Y_n - n\widehat{f}_N - \widehat{\theta}_N \rceil\right)^2.$$

Let (f', θ') be the minimizer of $Z(f,\theta)$. Observe that Z is quadratic in (f,θ) and so the partial derivatives of Z at the unique minimizer (f', θ') are 0. Now

$$Z(f',\theta') \le Z(\widehat{f}_N, \widehat{\theta}_N) = SS(\widehat{f}_N, \widehat{\theta}_N) \le SS(f',\theta').$$

Thus $Z(f',\theta') \le SS(f',\theta')$. However, $\forall n$

$$\left|Y_n - nf' - \theta' - \lfloor Y_n - n\widehat{f}_N - \widehat{\theta}_N \rceil\right| \ge \langle Y_n - nf' - \theta' \rangle.$$

Hence $Z(f',\theta') \ge SS(f',\theta')$, and so $Z(f',\theta') = SS(f',\theta')$ and $Z(f',\theta') = Z(\widehat{f}_N, \widehat{\theta}_N)$. Consequently $f' = \widehat{f}_N$ and $\theta' = \widehat{\theta}_N$. The partial derivatives of Z and SS are thus identical whenever the latter's exist, and since the derivatives of SS exist everywhere except on $\cup_n\{(f,\theta); \langle Y_n - nf - \theta\rangle = -1/2\}$, which has probability zero, the result follows.[3] ■

REFERENCES

[1] B. G. Quinn and E. J. Hannan, *The Estimation and Tracking of Frequency*. New York: Cambridge Univ. Press, 2001.

[2] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. 20, pp. 591–598, Sep. 1974.

[3] E. J. Hannan, "The estimation of frequency," *J. Appl. Probabil.*, vol. 10, no. 3, pp. 510–519, Sep. 1973.

[4] A. M. Walker, "On the estimation of a harmonic component in a time series with stationary independent residuals," *Biometrika*, vol. 58, pp. 21–36, 1971.

[5] J. A. Rice and M. Rosenblatt, "On frequency estimation," *Biometrika*, vol. 75, no. 3, pp. 477–484, Sep. 1988.

[6] B. G. Quinn, R. G. McKilliam, and I. V. L. Clarkson, "Maximizing the periodogram," *IEEE GLOBECOM*, pp. 1–5, Dec. 2008.

[7] M. L. Fowler, "Phase-based frequency estimation: A review," *Digit. Signal Process.*, vol. 12, no. 4, pp. 590–615, 2002.

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

[9] S. M. Kay, "A fast and accurate single frequency estimator," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 1989, vol. 37, pp. 1987–1990.

[10] I. V. L. Clarkson, P. J. Kootsookos, and B. G. Quinn, "Analysis of the variance threshold of Kay's weighted linear predictor frequency estimator," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2370–2379, Sep. 1994.

[11] I. V. L. Clarkson, "Frequency estimation, phase unwrapping and the nearest lattice point problem," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Mar. 1999, vol. 3, pp. 1609–1612.

[12] B. C. Lovell, P. J. Kootsookos, and R. C. Williamson, "The circular nature of discrete-time frequency estimates," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Apr. 1991, vol. 5, pp. 3369–3372.

[13] B. C. Lovell and R. C. Williamson, "The statistical performance of some instantaneous frequency estimators," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1708–1723, Jul. 1992.

[14] I. V. L. Clarkson, "Efficient single frequency estimators," in *Proc. Int. Symp. Signal Process. Appl.*, Aug. 1992, pp. 327–330.

[15] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[16] B. G. Quinn, "Estimating the mode of a phase distribution," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 587–591.

[17] L. Babai, "On Lovasz lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, pp. 1–13, 1986.

[18] K. V. Mardia and P. Jupp, *Directional Statistics*, 2nd ed. New York: Wiley, 2000.

[19] N. I. Fisher, *Statistical Analysis of Cirular Data*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[20] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed. New York: Springer, 1998.

[21] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1212–1215, 2001.

[22] M. Ajtai, "The shortest vector problem in l_2 is NP-hard for randomized reductions," in *Proc. 30th ACM Symp. Theory of Comput.*, May 1998, pp. 10–19.

[23] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[3]It is possible to remove the 'with probability zero' statement by appealing to some concepts in lattice theory that we describe in Section IV. Due to space restrictions we have not included this result. Theorem 1 holds in any case.

[24] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.
[25] M. Pohst, "On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications," *SIGSAM Bull.*, vol. 15, no. 1, pp. 37–44, 1981.
[26] J. H. Conway and N. J. A. Sloane, "Fast quantizing and decoding and algorithms for lattice quantizers and codes," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 227–232, Mar. 1982.
[27] J. H. Conway and N. J. A. Sloane, "Soft decoding techniques for codes and lattices, including the Golay code and the Leech lattice," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 41–50, Jan. 1986.
[28] I. V. L. Clarkson, "An algorithm to compute a nearest point in the lattice A_n^*," in *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, M. Fossorier, H. Imai, S. Lin, and A. Poli, Eds. New York: Springer, 1999, vol. 1719, Lecture Notes in Comput. Sci., pp. 104–120.
[29] R. G. McKilliam, I. V. L. Clarkson, and B. G. Quinn, "An algorithm to compute the nearest point in the lattice A_n^*," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4378–4381, Sep. 2008.
[30] R. G. McKilliam, I. V. L. Clarkson, W. D. Smith, and B. G. Quinn, "A linear-time nearest point algorithm for the lattice A_n^*," in *Proc. Int. Symp. Inf. Theory Appl.*, Dec. 2008.
[31] R. G. McKilliam, W. D. Smith, and I. V. L. Clarkson, "Linear-time nearest point algorithms for Coxeter lattices," *IEEE Trans. Inf. Theory*, 2009, accepted for publication.
[32] R. G. McKilliam, D. J. Ryan, I. V. L. Clarkson, and I. B. Collings, "An improved algorithm for optimal noncoherent QAM detection," in *Proc. Austral. Commun. Theory Workshop*, Jan. 2008, pp. 64–68.
[33] D. J. Ryan, I. V. L. Clarkson, and I. B. Collings, "GLRT-optimal noncoherent lattice decoding," *IEEE Trans. Signal Process.*, vol. 55, pp. 3773–3786, July 2007.
[34] R. G. McKilliam and I. V. L. Clarkson, "Maximum-likelihood period estimation from sparse, noisy timing data," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Mar. 2008, pp. 3697–3700.
[35] B. G. Quinn and J. M. Fernandes, "A fast efficient technique for the estimation of frequency," *Biometrika*, vol. 78, no. 3, pp. 489–497, Sep. 1991.
[36] B. G. Quinn, "Recent advances in rapid frequency estimation," *Digit. Signal Process.*, Apr. 2008.

Robby G. McKilliam was born in Brisbane, Queensland, Australia in 1983. He received the B.Sc. degree in mathematics and the B.E. degree (Hons. I) in computer systems engineering from the University of Queensland, Brisbane, Australia, in 2006.

He is pursuing the Ph.D. degree at the University of Queensland. His fields of interest are lattice theory, number theory, and signal processing and communications. He currently receives funding from an Australian postgraduate award and the CSIRO ICT Centre in Sydney, Australia.

Barry G. Quinn received the B.A. Hons. degree in pure mathematics, applied mathematics, and statistics in 1978 and the Ph.D. degree in statistics in 1981, both from the Australian National University, Canberra.

He is currently Professor of Statistics and Head of the Statistics Department with Macquarie University, Sydney. He has previously held Professorships at UMIST (now part of the University of Manchester, U.K.) and Goldsmiths College, London, U.K. He has also held lectureships with the Universities of Wollongong, Queensland, and Newcastle, and was a Principal Research Scientist with DSTO, Adelaide. His interests are in time series analysis and signal processing, and especially frequency estimation.

I. Vaughan L. Clarkson was born in Brisbane, Queensland, Australia, in 1968. He received the B.Sc. degree in mathematics and the B.E. degree (Hons. I) in computer systems engineering from The University of Queensland, Brisbane, in 1989 and 1990, respectively, and the Ph.D. degree in systems engineering from The Australian National University, Canberra, in 1997.

Beginning in 1988, he was with the Defence Science and Technology Organisation, Adelaide, Australia, first as a Cadet, later as a Professional Officer, and finally as a Research Scientist. From 1998 to 2000, he was a Lecturer with The University of Melbourne, Melbourne, Australia. From 2000 to 2008, he was a Senior Lecturer with the School of Information Technology and Electrical Engineering, The University of Queensland. In 2008, he was promoted to Reader. In 2005, he was a Visiting Professor with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada. His research interests include statistical signal processing for communications and defense, image processing, information theory, and lattice theory.

Bill Moran received First Class Honours B.Sc. degree in mathematics from the University of Birmingham, U.K., in 1965, and the Ph.D. degree in pure mathematics from the University of Sheffield, U.K., in 1968.

He is with the Department of Electrical and Electronic Engineering, University of Melbourne, Australia, where he is the Managing Complexity Theme Leader for National ICT Australia, and Research Director of Melbourne Systems Laboratory (MSL). In 2001, he became a Professor of Electrical Engineering, MSL. Previously, he was Professor of Mathematics (1976–1991), Head of the Department of Pure Mathematics (1977–1979, 1984–1986), Dean of Mathematical and Computer Sciences (1981, 1982, 1989) at the University of Adelaide, and Head of the Mathematics Discipline, Flinders University of South Australia (1991–1995). He was a Chief Investigator (1992–1995) and Head of the Medical Signal Processing Program (1995–1999) in the Cooperative Research Centre for Sensor Signal and Information Processing. His main areas of research interest are in signal processing, both theoretically and in applications to radar, waveform design and radar theory, sensor networks, and sensor management. He also works in various areas of mathematics including harmonic analysis, representation theory, and number theory.

Dr. Moran was elected to the Fellowship of the Australian Academy of Science in 1984. He has been a Principal Investigator on numerous research grants and contracts, in areas spanning pure mathematics to radar development, from both Australian and U.S. Research Funding Agencies, including DARPA, AFOSR, AFRL, Australian Research Council (ARC), Australian Department of Education, Science and Training, and DSTO. He is a currently a member of the Australian Research Council College of Experts.