Polynomial phase estimation by least squares phase unwrapping

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

Abstract—Estimating the coefficients of a noisy polynomial phase signal is important in fields including radar, biology and radio communications. One approach attempts to perform polynomial regression on the phase of the signal. This is complicated by the fact that the phase is wrapped modulo $2\pi$ and must be unwrapped before regression can be performed. In this paper we consider an estimator that performs phase unwrapping in a least squares manner. We call this the least squares unwrapping (LSU) estimator. The LSU estimator can be computed in a reasonable amount of time for data sets of moderate size using existing general purpose algorithms from algebraic number theory. Under mild conditions on the distribution of the noise we describe the asymptotic properties of this estimator, showing that it is strongly consistent and asymptotically normally distributed. A key feature is that the LSU estimator is accurate over a far wider range of parameters than many popular existing estimators. Monte-Carlo simulations support our theoretical results and demonstrate the excellent statistical performance of the LSU estimator when compared with existing state-of-the-art estimators.

Index Terms—Polynomial phase signals, phase unwrapping, asymptotic properties, nearest lattice point problem

I. INTRODUCTION

Polynomial phase signals arise in fields including radar, sonar, geophysics, biology, and radio communication [1–4]. A polynomial phase signal of order $m$ is a function of the form

$$s(t) = e^{2\pi jy(t)},$$

where $t$ is a real number usually representing time and

$$y(t) = \tilde{\mu}_0 + \tilde{\mu}_1 t + \tilde{\mu}_2 t^2 + \ldots + \tilde{\mu}_m t^m$$

is a polynomial of order $m$. In this paper we assume that the signal is sampled at the integers and the sampled polynomial phase signal takes the form

$$s_n = s(n) = e^{2\pi jy(n)} \quad n \in \mathbb{Z}.$$

Of practical importance is the estimation of the coefficients $\tilde{\mu}_0, \ldots, \tilde{\mu}_m$ from a number, say $N$, of observations of the noisy sampled signal

$$Y_n = \rho s_n + X_n \quad n = 1, \ldots, N,$$

where $\rho$ is a positive number representing the (usually unknown) signal amplitude and $\{X_n, n \in \mathbb{Z}\}$ is a sequence of complex random variables representing noise. In order to ensure identifiability it is necessary to restrict the $m + 1$ coefficients to a region of $m + 1$ dimensional Euclidean space $\mathbb{R}^{m+1}$ called an identifiable region [5]. We discuss this further in Section III.

An obvious estimator of the unknown coefficients is the least squares estimator. This is also the maximum likelihood estimator when the noise sequence $\{X_n\}$ is white and Gaussian. When $m = 0$ (phase estimation) or $m = 1$ (frequency estimation) the least squares estimator is an effective approach, being both computationally efficient and statistically accurate [6–10]. When $m \geq 2$ the computational complexity of the least squares estimator is large [10–12]. For this reason many authors have considered alternative approaches to polynomial phase estimation. These can loosely be grouped into two classes: estimators based on `multilinear transforms', such as the high order ambiguity function (HAF) [13–18], the product HAF (PHAF) [19], and cubic phase function (CPF) [11, 20–22]; and estimators based on phase unwrapping, such as Kitchen's unwrapping estimator [23], the estimator of Djuric and Kay [24], and Morelande's Bayesian unwrapping estimator [25].

In this paper we consider the estimator that results from unwrapping the phase in a least squares manner. We call this the least squares unwrapping (LSU) estimator [10, 26–28]. It was shown in [26, 27] that the LSU estimator can be represented as a nearest lattice point problem [29], and Monte-Carlo simulations were used to show the LSU estimator's favourable statistical performance. A drawback of the LSU estimator is that computing a nearest lattice point is, in general, computationally difficult. In [26], two standard techniques were considered, the sphere decoder [29–31] that exactly computes a nearest lattice point, and Babai's nearest plane algorithm [32] that only produces an approximation. The sphere decoder was observed to have excellent statistical performance but can only be computed efficiently for $N$ less than about 60. Babai's algorithm is computationally simple, but its statistical performance has been observed to be comparatively poor at low signal-to-noise ratio (SNR) [26]. Of interest is that the lattices considered have significant structure, and may admit fast nearest point algorithms. This has been studied in [10] where polynomial time algorithms were found that compute the nearest point exactly. Although polynomial time, the algorithms are still computationally demanding in practice. Fast (exact or approximate) algorithms, yet to be discovered, may exist for these special lattices. In this paper we apply
another general purpose approximate nearest point algorithm called the K-best algorithm [33–35] for approximating the LSU estimator. We show that this provides near sphere decoder (exact) performance, but can be computed in a reasonable amount of time if \( N \) is not too large (approximately \( N \) less than 1000).

In addition to the above computational results a major purpose of this paper is to study the asymptotic properties of the LSU estimator. Under mild assumptions about the distribution of the noise we prove that the LSU estimator is strongly consistent, that is, the estimated coefficients converge almost surely to the true coefficients as the number of observations, \( N \), grows. The proof makes use of an elementary result about the number of arithmetic progressions contained inside subsets of \( \{ 1, 2, \ldots, N \} \) [36–38]. This proof technique appears to be novel and may be useful for purposes other than polynomial phase estimation, and in particular other applications involving data that is ‘wrapped’ in some sense. Potential candidates are the phase wrapped images observed in modern radar and medical imaging devices such as synthetic aperture radar and magnetic resonance imaging [39–41].

The analyses of most existing polynomial phase estimators make what is known as a ‘high SNR’ assumption [11, 18, 19, 21, 24, 42]. This assumption is used to justify what is called a ‘perturbation analysis’ to obtain expressions for the mean square error of the estimator in question. The high SNR assumption is usually supported by Monte-Carlo simulations showing the mean square error predicted by the perturbation analysis to be accurate provided that the signal-to-noise ratio (SNR) is larger than a ‘threshold’ value (see the simulations in Section V for examples). Often the threshold is observed to occur at smaller values of SNR as \( N \) increases. A problem with the high SNR assumption is that it is not known if the threshold occurs at smaller SNR with increasing \( N \), or if it instead converges to some finite (but perhaps small) value of SNR as \( N \) increases. The consequence of the latter is that the estimator will be inaccurate if the SNR is smaller than this finite value, regardless of the number of observations \( N \). Our analysis of the LSU estimator requires no high SNR assumption and guarantees that no such finite value of SNR exists for the LSU estimator. As a result the LSU estimator is accurate for signals with arbitrarily small SNR provided that the number of observations is sufficiently large.

An analysis of the HAF estimator under the assumption that the noise process \( \{ X_n \} \) is white and Gaussian and without the high SNR assumption has been conducted in [15] (see Lemma 1 and the remark that follows in [15]). To date, the literature does not contain a corresponding analysis of any estimator based on phase unwrapping. This has apparently led to the misconception that phase unwrapping estimators are only accurate when the noise is ‘small’ in some sense. This misconception may have been motivated by early papers such as [43, 44] where it was observed that the noise occurring in the phase of the signal only looks ‘Gaussian’ when the complex noise producing it has small variance. The results in this paper make it clear that phase unwrapping estimators can be accurate at arbitrarily low SNR. The information contained in the amplitude is not essential for accurate estimation of polynomial phase signals, and can be discarded with only small loss in accuracy. For example, in the case that \( \{ X_n \} \) is white and Gaussian, our results show that, asymptotically in \( N \), the ratio of the variance of the LSU estimator to the Cramer-Rao lower bound (CRB) converges to 1 for high SNR, and to \( \frac{\pi}{2} \) for low SNR. Related results have been presented in [6]. This factor of \( \frac{\pi}{2} \) for low SNR is better than popular estimators such as the HAF and CPF, for which the ratio of the variance to the CRB diverges as the SNR becomes small [11, 15, 21].

A key property of the LSU estimator is that it works for polynomial phase parameters contained anywhere inside the identifiable region described in Section III and [5, 45]. This is the maximum possible range of parameters. By contrast popular estimators such as the HAF and CPF work only for parameters in a range much smaller than the identifiable region. This property has been studied previously [5, 45–47] and can be a critical limitation for some applications, for example, interference suppression in the presence of a jammer [46, 48]. Approaches are known that extend the range of parameters suitable for the HAF [46, 47], but the resulting estimators typically have poor statistical performance. For this reason, the LSU estimator may be the only known estimator that is viable for certain applications, such as interference suppression. The theoretical results given in Section IV prove that the LSU estimator works over the entire identifiable region. To our knowledge this is the first time a result of this type has been obtained for any polynomial phase estimator.

One drawback of the phase unwrapping approach is that it does not appear to generalise well to the problem of estimating the parameters in a signal that contains the sum of multiple polynomial phase signals. This appears to be a negative feature of all polynomial phase estimators based on phase unwrapping. However, the LSU estimator does generalise in other ways. For example, the LSU estimator can be extended to deal with polynomial phase signals in more dimensions than one. Our algorithms and analysis would carry over to this case with only minor modification. Even more generally, the ideas behind the LSU estimator might apply to other types of phase unwrapping problems such as observed in synthetic aperture radar and magnetic resonance imaging [39–41]. This is beyond the scope of our present paper, but may be the subject of future investigation.

The paper is organised in the following way. Section II describes some required concepts from lattice theory and Section III describes the identifiable region that was derived in [5]. These identifiability results are required in order to understand the statistical properties of the LSU estimator. Section IV describes the LSU estimator and states a theorem asserting the estimator to be strongly consistent and asymptotically normally distributed under some mild assumptions on the noise process \( \{ X_n \} \). The proof of strong consistency is given in the appendix. Space restrictions make it impossible to give a proof of asymptotic normality in this paper. A complete proof of this is available in [49]. Section V describes the results of Monte Carlo simulations that compare the LSU estimator with existing estimators including the HAF, PHAF and CPF. These simulations agree with the asymptotic properties.
under certain conditions when the lattice itself is considered as an additional input parameter [51]. On the other hand, fast algorithms exist for specific lattices [52–55]. Even in the general case, where the lattice has no known structure, there are algorithms that compute a nearest lattice point in reasonable time if the dimension is small [29–31]. These algorithms typically operate using a tree search to enumerate all lattice points inside a ball of radius sufficiently large to contain a nearest point. The algorithms vary in the order that branches in the tree are searched. A popular approach due to Schnorr and Euchner [29, 56] searches branches in a greedy order. We refer to this algorithm as the sphere decoder in this paper, although other variants of the algorithm also go by this name. With modern computers, the sphere decoder is computationally viable if the dimension of the lattice is less than approximately 60.

Algorithms that approximate a nearest point have also been studied. One example is Babai’s nearest plane algorithm [32]. Another pragmatic approximate approach is to employ the sphere decoder, but prune branches in the search tree to ensure that there are at most a finite number, say $K$, remaining branches at any time. This approach has been studied previously [33–35] and resembles what is called the sequential $M$-algorithm [57] from coding theory. The approaches vary in how branches are pruned. In our implementation we prune branches using the same greedy metric as that used by the Schnorr and Euchner sphere decoder [29, 56] for deciding the order in which branches are searched. Following [34] we refer to this as the $K$-best algorithm. The $K$-best algorithm requires $O(NK^2 \log K)$ arithmetic operations. The maximum number of branches $K$ is free to be chosen. For polynomial phase estimation we have found that setting $K = 20N$ yields excellent results, and in this case the algorithm requires $O(N^3 \log N)$ operations. Simulations in Section V show that this algorithm produces excellent estimates of polynomial phase coefficients and can be computed in a reasonable amount of time for $N$ less than about 1000.

III. IDENTIFIABILITY AND ALIASING

Aliasing can occur when polynomial phase signals are sampled [5]. That is, two or more distinct polynomial phase signals can take exactly the same values at the sample points. Let $\mathcal{Z}$ be the set of polynomials of order at most $m$ that take integer values when evaluated at integers. That is, $\mathcal{Z}$ contains all polynomials $p$ such that $p(n)$ is an integer whenever $n$ is an integer. Let $y$ and $z$ be two distinct polynomials such that $z = y + p$ for some polynomial $p$ in $\mathcal{Z}$. The two polynomial phase signals

$$s(t) = e^{2\pi jy(t)} \quad \text{and} \quad r(t) = e^{2\pi jz(t)}$$

are distinct because $y$ and $z$ are distinct, but if we sample $s$ and $r$ at the integers

$$s(n) = e^{2\pi jy(n)} = e^{2\pi jy(n)} e^{2\pi jp(n)} = e^{2\pi jy(n) + p(n)} = e^{2\pi jz(n)} = r(n)$$

because $p(n)$ is always an integer and therefore $e^{2\pi jp(n)} = 1$ for all $n \in \mathbb{Z}$. The polynomial phase signals $s$ and $r$ are
equal at the integers, and although they are distinct, they are indistinguishable from their samples. We call such polynomial phase signals aliases and immediately obtain the following lemma.

**Lemma 1.** Two polynomial phase signals \( s(t) = e^{2\pi j y(t)} \) and \( r(t) = e^{2\pi j z(t)} \) are aliases if and only if the polynomials that define their phase, \( y \) and \( z \), differ by a polynomial from the set \( \mathcal{Z} \), that is, \( y - z \in \mathcal{Z} \).

The above is illustrated by Figures 2 and 3, where the phase (divided by \( 2\pi \)) of two distinct polynomial phase signals is plotted on the left, and the principal component of the phase (also divided by \( 2\pi \)) on the right. The circles display the samples at the integers. Note that the samples of the principal components intersect. The corresponding polynomial phase signals are aliases.

We can derive an analogue of the theorem above in terms of the coefficients of the polynomials \( y \) and \( z \). This will be useful when we consider estimating the coefficients in Section IV. We first need the following family of polynomials.

**Definition 1.** (Integer valued polynomials)

The integer valued polynomial of order \( k \), denoted by \( p_k \), is

\[
p_k(x) = \binom{x}{k} = \frac{x(x-1)(x-2) \cdots (x-k+1)}{k!},
\]

where we define \( p_0(x) = 1 \).

**Lemma 2.** The integer valued polynomials \( p_0, \ldots, p_m \) form an integer basis for \( \mathcal{Z} \). That is, every polynomial in \( \mathcal{Z} \) can be uniquely written as

\[
c_0 p_0 + c_1 p_1 + \cdots + c_m p_m,
\]

where the \( c_i \in \mathbb{Z} \).

**Proof:** See [58, p. 2] or [5].

Given a polynomial \( g(x) = a_0 + a_1 x + \cdots + a_m x^m \), let

\[
\text{coef}(g) = [ a_0 \ a_1 \ a_2 \ \cdots \ a_m ]^t,
\]

denote the column vector of length \( m + 1 \) containing the coefficients of \( g \), where \( ^t \) indicates the transpose operation. If \( y \) and \( z \) differ by a polynomial in \( \mathcal{Z} \) then we can write \( y = z + p \) where \( p \in \mathcal{Z} \) and \( \text{coef}(y) = \text{coef}(z) + \text{coef}(p) \).

Consider the set

\[
L_{m+1} = \{ \text{coef}(p) : p \in \mathcal{Z} \}
\]

containing the coefficient vectors corresponding to the polynomials in \( \mathcal{Z} \). Since the integer valued polynomials form a basis for \( \mathcal{Z} \),

\[
L_{m+1} = \{ \text{coef}(c_0 p_0 + c_1 p_1 + \cdots + c_m p_m) : c_i \in \mathbb{Z} \} = \{ c_0 \text{coef}(p_0) + \cdots + c_m \text{coef}(p_m) : c_i \in \mathbb{Z} \}.
\]

Let

\[
P = \begin{bmatrix}
\text{coef}(p_0) & \text{coef}(p_1) & \cdots & \text{coef}(p_m)
\end{bmatrix}
\]

be the \( m + 1 \) by \( m + 1 \) matrix with columns given by the coefficients of the integer valued polynomials. Then,

\[
L_{m+1} = \{ Pu : u \in \mathbb{Z}^{m+1} \}
\]

and it is clear that \( L_{m+1} \) is an \( m + 1 \) dimensional lattice. That is, the set of coefficients of the polynomials from \( \mathcal{Z} \) forms a lattice with generator matrix \( P \). We can restate Lemma 1 as:

**Corollary 1.** Two polynomial phase signals \( s(t) = e^{2\pi j y(t)} \) and \( r(t) = e^{2\pi j z(t)} \) are aliases if and only if \( \text{coef}(y) \) and \( \text{coef}(z) \) differ by a lattice point in \( L_{m+1} \).

For the purpose of estimating the coefficients of a polynomial phase signal we must (in order to ensure identifiability) restrict the set of allowable coefficients so that no two polynomial phase signals are aliases of each other. In consideration of Corollary 1 we require that the coefficients of \( y(t) \), written in vector form \( \mu \), are contained in a set of coset representatives for the quotient \( \mathbb{R}^{m+1}/L_{m+1} \). We call the chosen set of representatives the identifiable region.

As an example consider the polynomial phase signal of order zero \( e^{2\pi j \mu_0} \). Since \( e^{2\pi j \mu_0} = e^{2\pi j (\mu_0 + k)} \) for any integer \( k \) we must, in order to ensure identifiability, restrict \( \mu_0 \) to some interval of length 1. A natural choice is the interval \([-1/2, 1/2)\). The lattice \( L_1 \) is the 1-dimensional integer lattice \( \mathbb{Z} \) and the interval \([-1/2, 1/2)\) is the Voronoi cell of \( L_1 \). When \( m = 1 \) it turns out that a natural choice of identifiable region is the square box \([-1/2, 1/2)^2\) in accordance with the Nyquist criterion. The lattice \( L_2 \) is \( \mathbb{Z}^2 \) and \([-1/2, 1/2)^2\) is the Voronoi cell of \( L_2 \). When \( m > 1 \) the identifiable region becomes more complicated and \( L_{m+1} \neq \mathbb{Z}^{m+1} \).

In general there are infinitely many choices for the identifiable region. A natural choice is the Voronoi cell of \( L_{m+1} \) as used in [5]. Another potential choice is a fundamental parallelepiped of \( L_{m+1} \). In this paper we will use the rectangular set constructed using Proposition 1. Observe that the matrix
We will make use of this region when deriving the statistical properties of the LSU estimator in the next section.

Given \( x \) and \( y \) in \( \mathbb{R}^{n+1} \), we say that \( x \equiv y \mod L_{m+1} \) if \( x \) and \( y \) differ by a lattice point in \( L_{m+1} \). We define the function \( \text{dealias}(x) \) to take \( x \) to its coset representative inside \( B \). That is, \( \text{dealias}(x) = z \in B \) where \( x - z \in L_{m+1} \). When \( m = 0 \) or 1 then \( \text{dealias}(x) = \langle x \rangle \) where \( \langle x \rangle = x - \lfloor x \rfloor \) denotes the (centered) fractional part and \( \lfloor x \rfloor \) denotes the nearest integer to \( x \) and both \( \langle \cdot \rangle \) and \( \lfloor \cdot \rfloor \) operate on vectors elementwise. The direction of rounding for half-integers is not important so long as it is consistent. We have chosen to round up half-integers here. For \( m \geq 2 \) the function \( \text{dealias}(x) \) can be computed by a simple sequential algorithm [10, Sec. 7.2.1].

IV. The Least Squares Unwrapping Estimator

We now describe the least squares unwrapping (LSU) estimator of the polynomial coefficients. Recall that we desire to estimate \( \mu_0, \ldots, \mu_m \) from the observations \( Y_1, \ldots, Y_N \) defined in (1). Let

\[
\Theta_n = \frac{\angle Y_n}{2\pi} = \langle \Phi_n + y(n) \rangle
\]

where \( \angle z \) denotes the argument (or angle) of the complex number \( z \), and

\[
\Phi_n = \frac{1}{2\pi} \angle (1 + \rho^{-1}s^{-1}X_n)
\]

are random variables representing the phase noise induced by the \( X_n \) [6, 43]. If the distribution of \( X_n \) is circularly symmetric (i.e., the angle \( \angle X_n \) is uniformly distributed on \([-\pi, \pi]\) and is independent of the magnitude \( |X_n| \)) then the distribution of \( \Phi_n \) is the same as the distribution of \( \frac{1}{2\pi} \angle (1 + \rho^{-1}X_n) \). If the \( X_1, \ldots, X_N \) are circularly symmetric and identically distributed, then \( \Phi_1, \ldots, \Phi_N \) are also identically distributed.

Let \( \mu \) be the vector \([\mu_0, \mu_1, \ldots, \mu_m]\) and let

\[
SS(\mu) = \sum_{n=1}^{N} \left( \Theta_n - \sum_{k=0}^{m} \mu_k n^k \right)^2 .
\]

The LSU estimator \( \hat{\mu} \) is the minimiser of \( SS(\mu) \) over the identifiable region \( B \), that is,

\[
\hat{\mu} = \arg \min_{\mu \in B} SS(\mu).
\]

This minimisation problem can be posed as a nearest lattice point problem [10, 26]. Write \( SS \) as

\[
SS(\mu) = \sum_{n=1}^{N} \left( \Theta_n - W_n - \sum_{k=0}^{m} \mu_k n^k \right)^2 ,
\]

where \( W_n = [\Theta_n - \sum_{k=0}^{m} \mu_k n^k] \) for \( n = 1, \ldots, N \) are integers called wrapping variables. If we consider \( W_1, \ldots, W_N \) as nuisance parameters to be estimated then \( SS \) can be considered as a function of both \( \mu_0, \ldots, \mu_m \) and \( W_1, \ldots, W_N \), the minimiser with respect to \( W_1, \ldots, W_N \) being \( SS(\mu) \).

The LSU estimator can be found by jointly minimising over \( \mu_0, \ldots, \mu_m \) and \( W_1, \ldots, W_N \). This joint minimisation problem can be solved by computing a nearest point in a lattice. To see this write \( SS \) as a function of both \( \mu_0, \ldots, \mu_m \) and \( W_1, \ldots, W_N \), that is,

\[
SS(\mu, w) = \| \theta - X\mu - w \|^2 ,
\]

where \( \| \cdot \|^2 \) denotes the squared Euclidean norm, \( X \) is the rectangular Vandermonde matrix

\[
X = \begin{bmatrix}
1 & 1 & 1 & \cdots & 1 \\
1 & 2 & 4 & \cdots & 2^m \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & N & N^2 & \cdots & N^m
\end{bmatrix}
\]

and the column vectors \( w = [W_1, \ldots, W_N]' \) and \( \theta = [\Theta_1, \ldots, \Theta_N]' \). For fixed \( w \) the minimiser of \( SS(\mu, w) \) with respect to \( \mu \) is

\[
\mu(w) = (X'X)^{-1}X'(\theta - w) = X^+ (\theta - w),
\]

where \( X^+ = (X'X)^{-1}X' \) is the Moore-Penrose pseudoinverse of \( X \). We may thus compute \( \hat{\mu} \) as \( \hat{\mu} = \mu(\hat{w}) \), where \( \hat{w} \) is the minimiser of

\[
SS(w) = SS(\mu(w), w) = \| Q\theta - Qw \|^2 ,
\]

where \( Q = I - XX^+ \) and \( I \) is the \( N \times N \) identity matrix.1

Let \( \Lambda \) be the lattice generated by \( Q \), that is \( \Lambda = \{ Qx : x \in \mathbb{Z}^N \} \). It follows that \( Qw \) is a lattice point in \( \Lambda \) and that minimising \( SS(w) \) is equivalent to finding the nearest lattice point in \( \Lambda \) to \( Q\theta \). Denote this point by \( Q\hat{w} \). Now the estimate \( \hat{\mu} \) is given by substituting \( \hat{w} \) for \( w \) in (9). After this procedure it is possible that the \( \hat{\mu} \) obtained is not in the identifiable region but is instead an aliased version of the desired estimate. This can be resolved by computing \( \text{dealias}(\hat{\mu}) \). There exist algorithms that can compute a nearest point in \( \Lambda \) using a number of operations that is polynomial in \( N \) [10, Sec 4.3]. Although polynomial in complexity, these algorithms are not fast in practice. The existence of practically fast nearest point algorithms for these lattices is an interesting open problem.

The next theorem describes the asymptotic properties of this estimator. Before we give the proof it is necessary to understand some of the properties of the phase noise \( \Phi_1, \ldots, \Phi_N \), which are circular random variables with support on \([-1/2, 1/2]\) [9, 10, 59, 60]. Circular random variables are often considered modulo \( 2\pi \) and therefore have support \([-\pi, \pi]\) with \(-\pi \) and \( \pi \) being identified as equivalent. Here we instead consider circular random variables modulo 1 with support \([-1/2, 1/2]\) and with \(-1/2 \) and \( 1/2 \) being equivalent. This is nonstandard but it allows us to use notation such as \( \lfloor \cdot \rfloor \) for rounding and \( \langle \cdot \rangle \) for the centered fractional part in a convenient way.

1We have slightly abused notation here by reusing \( SS \). This should not cause any confusion as \( SS(\mu) \) and \( SS(\mu, w) \) and \( SS(w) \) are easily told apart by their inputs.

2The \( N \times N \) projection matrix \( Q \) does not have full rank, it has rank \( N - m - 1 \), and for this reason it is not strictly a generator matrix. However, a generator can be constructed by removing \( m + 1 \) consecutive columns from \( Q \). In our implementation the last \( m + 1 \) columns are removed.
The intrinsic mean (or Fréchet mean) of \( \Phi_n \) is defined as [9, 61],
\[
\mu_{\text{intr}} = \arg \min_{\mu \in [-1/2, 1/2]} \mathbb{E} (\Phi_n - \mu)^2,
\]
and the intrinsic variance is
\[
\sigma_{\text{intr}}^2 = \mathbb{E} (\Theta - \mu_{\text{intr}})^2 = \min_{\mu \in [-1/2, 1/2]} \mathbb{E} (\Phi_n - \mu)^2,
\]
where \( \mathbb{E} \) denotes expected value. Depending on the distribution of \( \Phi_n \), the argument that minimises (10) may not be unique. The set of minima is often called the Fréchet mean set [61, 62]. If the minimiser is not unique we say that \( \Phi_n \) has no intrinsic mean. We are now equipped to state the asymptotic properties of the LSU estimator.

**Theorem 1.** Let \( \hat{\mu} \) be defined by (6) and put \( \hat{\lambda}_N = \text{dealias}(\hat{\mu} - \bar{\mu}) \). Denote the elements of \( \hat{\lambda}_N \) by \( \hat{\lambda}_{0,N}, \ldots, \hat{\lambda}_{m,N} \). Suppose \( \Phi_1, \ldots, \Phi_N \) are independent and identically distributed with zero intrinsic mean, intrinsic variance \( \sigma^2 \), and pdf \( f \), then:

1. (Strong consistency) \( N^k \hat{\lambda}_{k,N} \) converges almost surely to 0 as \( N \to \infty \) for all \( k = 0, 1, \ldots, m \).
2. (Asymptotic normality) If \( f(x) \) is continuous at \( x = -1/2 \) and \( f(-1/2) < 1 \) then the distribution of
\[
\left[ \sqrt{N} \hat{\lambda}_{0,N} N \sqrt{N} \hat{\lambda}_{1,N} \ldots N^m \sqrt{N} \hat{\lambda}_{m,N} \right]'
\]
converges to the normal with zero mean and covariance matrix
\[
\frac{\sigma^2}{(1 - f(-1/2))^2} C^{-1},
\]
where \( C \) is the \( m + 1 \) by \( m + 1 \) Hilbert matrix with elements \( C_{i,k} = 1/(i + k + 1) \) for \( i, k \in \{0, 1, \ldots, m\} \).

A proof of the first part of this theorem (strong consistency) is given in Appendix A. Space restrictions make it impossible to prove asymptotic normality in this paper. A complete proof of normality is available in [49]. Proofs for the case of \( m = 0 \) were given in [9] and \( m = 1 \) in [27]. The proof here takes a similar approach, but requires new techniques. A statement similar to Theorem 1 without a complete proof was given in [28]. Here, we give a proof. The results here are also more general than in [28], allowing for a wider class of noise distributions.

The theorem makes statements about the dealiased difference dealias(\( \mu - \bar{\mu} \)) between the true coefficients \( \mu \) and the estimated coefficients \( \hat{\mu} \) rather than directly about the difference \( \mu - \bar{\mu} \). To see why this makes sense, consider the case when \( m = 0 \), \( \hat{\mu}_0 = -0.5 \) and \( \bar{\mu}_0 = 0.49 \), so that \( \hat{\mu}_0 - \bar{\mu}_0 = -0.99 \). However, the two phases are obviously close, since the phases \( \pm 0.5 \) are actually the same. In this case
\[
\text{dealias}(\hat{\mu}_0 - \bar{\mu}_0) = (\hat{\mu}_0 - \bar{\mu}_0) = 0.01.
\]
The same reasoning holds for \( m > 0 \).

The requirement that \( \Phi_1, \ldots, \Phi_N \) be identically distributed will typically hold only when the complex random variables \( X_1, \ldots, X_N \) are identically distributed and circularly symmetric. It would be possible to drop the assumption that \( \Phi_1, \ldots, \Phi_N \) be identically distributed, but this complicates the theorem statement and the proof. In the interest of simplicity we only consider the case when \( \Phi_1, \ldots, \Phi_N \) are identically distributed here. If \( X_n \) is circularly symmetric with pdf nonincreasing with magnitude \( |X_n| \), then, the corresponding \( \Phi_n \) necessarily has zero intrinsic mean [10, Theorem 5.2]. Thus, our theorem covers commonly used distributions for \( X_1, \ldots, X_N \), such as the normal distribution. No assumptions about the magnitude of \( X_1, \ldots, X_N \) are required, that is, we do not require the noise to be ‘small’ in any sense. For example, if \( X_1, \ldots, X_N \) are normally distributed with variance \( \sigma^2 \), then the theorem holds regardless of how large \( \sigma^2 \) is, and as a result the LSU estimator will be accurate provided \( N \) is sufficiently large.

The proof of asymptotic normality places requirements on the pdf \( f \) of the phase noise. The requirement that \( \Phi_1, \ldots, \Phi_N \) have zero intrinsic mean implies that \( f(-1/2) \leq 1 \) [9, Lemma 1], so the only case not handled is when equality holds, i.e., if \( f(-1/2) = 1 \) or when \( f(x) \) is discontinuous at \( x = -1/2 \). In these exceptional cases other expressions for the asymptotic variance can be found (similar to [63, Theorem 3.1]), but this comes at a substantial increase in complexity and for this reason we have omitted them.

V. **Simulations**

This section describes the results of Monte-Carlo simulations with the least squares unwrapping (LSU) estimator, the high order ambiguity function (HAF) estimator [13], the product high order ambiguity function (PHAF) estimator [19], and the cubic phase function (CPF) estimator [11]. Simulation are also performed with two recently proposed estimators that combine the HAF and CPF [21]. The standard version of this estimator we refer to using the abbreviation CPF-HAF [21, Section 3] and the product form of the estimator we abbreviate by PCH [21, Section 5.A]. The CPF estimator only applies to polynomial phase signals of order \( m = 3 \) and when the number of observations \( N \) is odd. In all simulations the unknown amplitude is \( \rho = 1 \) and \( N = 199 \). The \( X_1, \ldots, X_N \) are pseudorandomly generated independent and identically distributed circularly symmetric complex Gaussian with variance \( \mathbb{E} |X_1|^2 = \sigma^2 \). The corresponding signal-to-noise ratio (SNR) is \( \rho^2 \sigma_c^2 = \sigma_c^2 \). The number of replications of each Monte-Carlo experiment is \( T = 2000 \) to obtain estimates \( \hat{\mu}_1, \ldots, \hat{\mu}_T \) and the corresponding dealiased errors \( \hat{\lambda}_t = \text{dealias}(\hat{\mu}_t - \hat{\mu}_0) \) are computed. The sample mean square error (MSE) of the \( k \)th coefficient is computed according to
\[
\frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_{k,t}^2 \]
where \( \hat{\lambda}_{k,t} \) is the \( k \)th element of \( \hat{\lambda}_t \). In all simulations four lag sets are used for the PHAF and PCH estimators [19, 21] and the LSU estimator is approximated using the \( K \)-best method described in Section II. The maximum number of branches allowed is \( K = 20N \). The arithmetic complexity of the LSU estimator is \( O(N^3 \log N) \) in this case.

Figure 4 shows the sample MSE of the highest order coefficient \( \mu_3 \) for a polynomial phase signal of order \( m = 3 \). All the estimators perform poorly at low SNR until a ‘threshold’ is reached, after which the variance is close to the Cramer-Rao bound (CRB) [64] indicated by the solid line. The asymptotic variance of the LSU estimator predicted by Theorem 1 is displayed by the dashed line. Provided the SNR is large enough...
can be shown to be

The volume of the coefficients suitable for the CPF estimator

is known to diverge as the SNR becomes small [11, 15, 21]. In this figure

the true coefficients are

where

and

The threshold for the estimator is

Figure 6 shows the performance of the Zhou and

region

B

a method based on the extended Euclidean algorithm that

previously [5, 25, 45–47, 65] and can be a critical limitation for

the HAF, PHAF and CPF estimators. When

the algorithm from [27]. In this case the value at which the

LSU estimator reaches a threshold is similar to the maximum

likelihood periodogram estimator [7, 8, 66]. This property was

observed in [27]. Both the LSU and periodogram estimators

are strongly consistent, and so, the threshold values can be

made arbitrarily small by choosing N large. This is indicated

in the bottom row of Table I by the −∞ symbols. When

the LSU estimator can be exactly computed using the sphere

decoder [29, 31] in a reasonable amount of time when

When

we use the approximate K-best algorithm. We

observe that the K-best algorithm produces estimates with the

same threshold as the exactly computed LSU estimator when

N = 19 and 49. As

→ ∞ the asymptotic

threshold of the HAF estimator has been studied in [15] and these values

are indicated in the bottom rows of Tables II and III. By

Theorem 1 the LSU estimator is strongly consistent and the

threshold can be made as small as desired by choosing N large. Whether the LSU estimator computed approximately

using the K-best algorithm is strongly consistent is an open

theoretical question. That the K-best algorithm provides a
close approximation of the exact LSU estimator for small N

is an encouraging observation.

Figure 9 shows the running time of the LSU estimator
TABLE I
Estimated SNR threshold in decibels with \( m = 1 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>LSU</th>
<th>( K )-best</th>
<th>CPF</th>
<th>HAF</th>
<th>ZW</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>-2</td>
<td>-4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>-6</td>
<td>-8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td>-14</td>
<td>-14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \infty )</td>
<td>( -\infty )</td>
<td>( -\infty )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE II
Estimated SNR threshold in decibels with \( m = 3 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>LSU</th>
<th>( K )-best</th>
<th>CPF</th>
<th>HAF</th>
<th>ZW</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>49</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>29</td>
</tr>
<tr>
<td>99</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>1</td>
<td>-1</td>
<td>5</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>( \infty )</td>
<td>( -\infty )</td>
<td>( 3.77 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

computed using the \( K \)-best method versus the HAF estimator for polynomial phase signals of order 3. The HAF estimator is computed using the fast Fourier transform and has complexity \( O(N \log N) \). With \( K = 20N \) the LSU estimator computed using the \( K \)-best method requires \( O(N^3 \log N) \) operations in the worst case. The \( K \)-best method tends to be faster when the SNR is large. This is due to the algorithm rapidly finding a nearest point, and subsequently terminating branches in its search tree early. On the other hand, when the SNR is small, all branches in the tree tend to be searched. Interestingly, the LSU estimator computed using the \( K \)-best algorithm is comparable with the HAF when the SNR is large and when \( N \) is approximately less than 100. The LSU estimator can be computed in a reasonable amount of time (a few minutes) when \( N = 1000 \). The algorithm becomes prohibitively expensive for large \( N \). Benchmarks have been performed with polynomial phase signals of order larger than 3. The results are omitted due to space constraints. The running time of the LSU estimator does not appear to change significantly as \( m \) increases. This is expected since a nearest point in a lattice of approximately the same dimension \( N - m - 1 \) must be computed. The complexity of the HAF estimator grows linearly with \( m \).

VI. CONCLUSION

This paper has considered the estimation of the coefficients of a noisy polynomial phase signal by least squares phase unwrapping (LSU). The LSU estimator was shown to be strongly consistent and asymptotically normally distributed under mild conditions on the distribution of the noise. The estimator can be computed by finding a nearest point in a lattice. Polynomial time nearest point algorithms for these lattices exist [10, Sec 4.3], but these algorithms are not fast in practice. We instead make use of the approximate \( K \)-best algorithm [33–35]. This algorithm results in an accurate estimator and can be computed in a reasonable amount of time for moderately large numbers of observations \( N \).

A key property of the LSU estimator is that it works for polynomial phase parameters contained anywhere inside
the identifiable region described in Section III. This is the maximum possible range of parameters. By contrast popular estimators such as the HAF and CPF work only for parameters in a range much smaller than the identifiable region. For this reason, the LSU estimator may be the only presently existing estimator that is viable for certain applications, such as interference suppression [46, 48].

The major outstanding question is whether faster nearest point algorithms exist for these specific lattices, particularly for the case when $N$ is large. Considering the excellent statistical performance (both theoretically and practically) of the LSU estimator, even fast approximate nearest point algorithms are likely to prove useful for the estimation of polynomial phase signals.

APPENDIX

A. Proof of strong consistency

Substituting (4) into $SS$ we obtain

$$SS(\mu) = \sum_{n=1}^{N} \left( \Phi_n + \sum_{k=0}^{m} \tilde{\mu}_k n^k \right) - \sum_{k=0}^{m} \mu_k n^k \right)^2$$

$$= \sum_{n=1}^{N} \left( \Phi_n + \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k) n^k \right)^2.$$ 

Let $\lambda = \text{dealias}(\tilde{\mu} - \mu) = \tilde{\mu} - \mu - p$ where $p$ is a lattice point from $L_{m+1}$. From the definition of $L_{m+1}$ we have $p_0 + p_1 n + \cdots + p_m n^m$ an integer whenever $n$ is an integer, and so

$$\left( \sum_{k=0}^{m} \lambda_k n^k \right) = \left( \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k - p_k) n^k \right)$$

$$= \left( \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k) n^k \right).$$

Let

$$SS(\mu) = \sum_{n=1}^{N} \left( \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right)^2 = NS_N(\lambda).$$

From the definition of the $\text{dealias}(\cdot)$ function $\lambda \in B$ so the elements of $\lambda$ satisfy

$$-\frac{0.5}{k!} \leq \lambda_k < \frac{0.5}{k!}. \quad (11)$$

Now $\tilde{\lambda}_N = \text{dealias}(\tilde{\mu} - \tilde{\mu})$ is the minimiser of $S_N$ in $B$. Let

$$V_N(\lambda) = E_S(\lambda) = \frac{1}{N} \sum_{n=1}^{N} \left( \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right)^2.$$
Using techniques described in [69, 70] one can show that
\[ \sup_{\lambda \in B} |S_N(\lambda) - V_N(\lambda)| \to 0 \] (12)
almost surely as \( N \to \infty \). This type of result is called a uniform law of large numbers [69, 70]. A full proof of (12) in given in [9]. We now concentrate attention on the minimiser of \( V_N \). Because \( \Phi_n \) has zero intrinsic mean
\[ \mathbb{E} \langle \Phi_n + z \rangle^2 \] (13)
is minimised uniquely at \( z = 0 \) for \( z \in [-1/2, 1/2) \). Since the intrinsic variance of \( \Phi_n \) is \( \sigma^2 \), when \( z = 0 \),
\[ \mathbb{E} \langle \Phi_1 + z \rangle^2 = \mathbb{E} \langle \Phi_1 \rangle^2 = \sigma^2, \] (14)
and so the minimum attained value is \( \sigma^2 \).

**Lemma 3.** The minimum value of \( V_N(\lambda) \) over \( \lambda \in B \) is \( \sigma^2 \), uniquely attained at the \( \lambda = 0 \) where \( 0 \) is the origin.

**Proof:** Put \( z(n) = \lambda_0 + \lambda_1 n + \ldots + \lambda_m n^m \). Then
\[ V_N(\lambda) = \frac{1}{N} \mathbb{E} \sum_{n=1}^{N} \left( \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right)^2 \]
\[ = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E} \langle \Phi_n + \langle z(n) \rangle \rangle^2. \]
From above \( \mathbb{E} \langle \Phi_n + \langle z(n) \rangle \rangle^2 \) has minimum value \( \sigma^2 \), uniquely when \( \langle z(n) \rangle = 0 \). But, \( \langle z(n) \rangle \) is zero for all integers \( n \) if and only if \( z \in \mathbb{Z} \), or equivalently if \( \text{coeff}(z) \) is a lattice point in \( L_{m+1} \). By definition \( B \) contains precisely one lattice point from \( L_{m+1} \), namely \( 0 \). Therefore \( V_N \) is minimised uniquely at \( 0 \), at which point it takes the value \( \sigma^2 \). \( \Box \)

**Lemma 4.** \( |V_N(\hat{\lambda}_N) - \sigma^2| \to 0 \) almost surely as \( N \to \infty \).

**Proof:** Since \( \hat{\lambda}_N = \arg \min_{\lambda \in B} S_N(\lambda) \),
\[ 0 \leq S_N(0) - S_N(\hat{\lambda}_N). \]
Also, because \( V_N \) is minimised at \( 0 \), it follows that
\[ 0 \leq V_N(\hat{\lambda}_N) - V_N(0) \]
\[ \leq V_N(\hat{\lambda}_N) - V_N(0) + S_N(0) - S_N(\hat{\lambda}_N) \]
\[ \leq |V_N(\hat{\lambda}_N) - S_N(\hat{\lambda}_N)| + |S_N(0) - V_N(0)| \]
which converges almost surely to zero as \( N \to \infty \) as a result of (12).

We have now shown that \( V_N \) is uniquely minimised at \( 0 \), that \( V_N(0) = \sigma^2 \), and that \( V_N(\hat{\lambda}_N) \) converges almost surely to \( \sigma^2 \). We could now proceed to prove that \( \hat{\lambda}_N \) converges almost surely to \( 0 \), but this would tell us nothing about the superior rates of convergence stated in Theorem 1. To prove these stronger properties we need some preliminary results about arithmetic progressions, and from the calculus of finite differences.

Let \( W = \{1, 2, \ldots, N\} \) and let \( K \) be a subset of \( W \). For any integer \( h \), let
\[ A(h, K) = \{ n : n + ih \in K \ \forall \ i \in \{0, 1, \ldots, m\} \} \] (15)
be the set containing all integers \( n \) such that the arithmetic progression \( n, n + h, n + 2h, \ldots, n + mh \) of length \( m+1 \) is contained in the subset \( K \). As the size of the subset \( K \) decreases it is possible that \( A(h, K) \) might be empty. However, the next two lemmas and the following corollary will show that if \( K \) is sufficiently large then it always contains at least one arithmetic progression (for all sufficiently small \( h \)) and therefore \( A(h, K) \) is not empty. We do not wish to claim any novelty here. The study of arithmetic progressions within subsets of \( W \) has a considerable history [36–38]. In particular, Gowers [38, Theorem 1.3] gives a result far stronger than we require here. Denote by \( K \setminus \{r\} \) the set \( K \) with the element \( r \) removed.

**Lemma 5.** Let \( r \in K \). For any \( h \), removing \( r \) from \( K \) removes at most \( m+1 \) arithmetic progressions \( n, n + h, \ldots, n + mh \) of length \( m+1 \). That is,
\[ |A(h, K) \setminus \{r\}| \geq |A(h, K)| - (m+1). \]

**Proof:** The proof follows because there are at most \( m+1 \) integers, \( n \), such that \( n + ih = r \) for some \( i \in \{0, 1, \ldots, m\} \). That is, there are at most \( m+1 \) arithmetic progressions of type \( n, n + h, \ldots, n + mh \) that contain \( r \).

**Lemma 6.** \( |A(h, K)| \geq N - mh - (N - |K|)(m+1). \)

**Proof:** Note that \( |A(h, W)| = N - mh \). The proof follows by starting with \( A(h, W) \) and applying Lemma 5 precisely \( |W| - |K| = N - |K| \) times. That is, \( K \) can be constructed by removing \( N - |K| \) elements from \( W \) and this removes at most \( (N - |K|)(m+1) \) arithmetic progressions from \( A(h, W) \).

**Corollary 2.** Let \( K \subseteq W \) such that \( |K| > \frac{2m+1}{2m+2} N \). For all integers \( h \) such that \( 1 \leq h \leq \frac{N}{m^2} \) the set \( K \) contains at least one arithmetic progression \( n, n + h, \ldots, n + mh \) of length \( m+1 \). That is, \( |A(h, K)| > 0 \).

**Proof:** By substituting the bounds \( |K| > \frac{2m+1}{2m+2} N \) and \( h \leq \frac{N}{m^2} \) into the inequality from Lemma 6 we immediately obtain \( |A(h, K)| > 0 \).

The next result we require comes from the calculus of finite differences. For any function \( d : \mathbb{R} \to \mathbb{R} \), let
\[ \Delta_h^k d(n) = d(n + h) - d(n) \]
denote the first difference with interval \( h \), and let
\[ \Delta_h^r d(n) = \Delta_h^{r-1} d(n + h) - \Delta_h^{r-1} d(n) \]
\[ = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} d(n + kh) \] (16)
denote the \( r \)th difference with interval \( h \). Since \( \sum_{k=0}^{r} \binom{r}{k} = 2^r \) it follows that \( \Delta_h^r d(n) \) can be represented by adding and subtracting
\[ d(n), d(n + h), \ldots, d(n + mh) \]
precisely \( 2^r \) times.
The operator $\Delta_h^r$ has special properties when applied to polynomials. If $d(n) = a_rn^r + \cdots + a_0$ is a polynomial of order $r$ then

$$\Delta_h^r d(n) = h^r r! a_r.$$  

(17)

Thus, the $r$th difference of $d(n)$ is a constant depending on $h$, $r$ and the $r$th coefficient $a_r$. We can now continue the proof of strong consistency. The next lemma is a key result.

**Lemma 7.** Suppose $\{\lambda_n\}$ is a sequence from $B$ with $V_N(\lambda_N) - \sigma^2 \to 0$ as $N \to \infty$. Then the elements $\lambda_{0,N}, \ldots, \lambda_{m,N}$ of $\lambda_N$ satisfy $N^k \lambda_k,N \to 0$ as $N \to \infty$.

**Proof:** Define the function

$$g(z) = E(\Phi_1 + z)^2 - \sigma^2$$

(18)

which is continuous in $z$. Because of (13) and (14), $g(z) \geq 0$ with equality only when $z = 0$. Now

$$V_N(\lambda_N) - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} g\left(\left(\sum_{k=0}^{m} n^k \lambda_{k,N}\right)\right) \to 0$$

as $N \to \infty$. Let

$$z_N(n) = \lambda_{0,N} + \lambda_{1,N}n + \cdots + \lambda_{m,N}n^m$$

so that

$$V_N(\lambda_N) - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} g\left(\langle z_N(n) \rangle\right) \to 0$$

as $N \to \infty$. Choose constants

$$c = \frac{2m + 1}{2m + 2} \quad \text{and} \quad 0 < \delta < \frac{1}{2^{2m+1}}$$

and define the set $K_N = \{n \leq N : |\langle z_N(n) \rangle| < \delta\}$. There exists $N_0$ such that for all $N > N_0$ the number of elements in $K_N$ is at least $cN$. Too see this, suppose that $|K_N| < cN$, and let $\gamma$ be the minimum value of $g$ over $[-1/2, -\delta] \cup [\delta, 1/2]$. Because $g(0) = 0$ is the unique minimiser of $g$, then $\gamma$ is strictly greater than 0 and

$$V_N(\lambda_N) - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} g\left(\langle z_N(n) \rangle\right)$$

$$\geq \frac{1}{N} \sum_{n \in K_N} \gamma = (1 - c)\gamma,$$

contradicting the fact that $V_N(\lambda_N) - \sigma^2$ converges to zero as $N \to \infty$. We will assume that $N > N_0$ in what follows.

From Corollary 2 it follows that for all $h$ satisfying $1 \leq h \leq \frac{N}{2m}$ the set $A(h, K_N)$ contains at least one element, that is, there exists $n' \in A(h, K_N)$ such that all the elements from the arithmetic progression $n', n'+h, \ldots, n'+mh$ are in $K_N$ and therefore

$$|\langle z_N(n') \rangle|, |\langle z_N(n' + h) \rangle|, \ldots, |\langle z_N(n' + mh) \rangle|$$

are all less than $\delta$. Because the $n$th difference is a linear combination of $2^m$ elements (see (16)) from

$$\langle z_N(n') \rangle, \langle z_N(n' + h) \rangle, \ldots, \langle z_N(n' + mh) \rangle$$

all with magnitude less than $\delta$ we obtain, from Lemma 8,

$$|\langle \Delta_h^n z_N(n') \rangle| \leq |\Delta_h^n \langle z_N(n') \rangle| < 2^m \delta.$$  

(19)

From (17) it follows that the left hand side is equal to a constant involving $h$, $m$ and $\lambda_{m,N}$ giving the bound

$$|\langle h^m m! \lambda_{m,N} \rangle| = |\langle \Delta_h^m z_N(n') \rangle| < 2^m \delta$$

(20)

for all $h$ satisfying $1 \leq h \leq \frac{N}{2m}$. Setting $h = 1$ and recalling from (11) that $\lambda_{m,N} \in [-\frac{0.5}{m!}, \frac{0.5}{m!}]$, we have

$$|\langle m! \lambda_{m,N} \rangle| = |m! \lambda_{m,N}| < 2^m \delta.$$  

Now, because we chose $\delta < \frac{1}{2^{2m+1}}$ it follows that

$$|\lambda_{m,N}| < \frac{2^m}{m!} \delta < \frac{1}{m! 2^{m+1}}.$$  

Thus, when $h = 2$,

$$|\langle 2^m m! \lambda_{m,N} \rangle| = |2^m m! \lambda_{m,N}| < 2^m \delta$$

because $2^m m! \lambda_{m,N} \in [-0.5, 0.5)$. Therefore

$$|\lambda_{m,N}| < \frac{1}{m!} \delta < \frac{1}{m! 2^{2m+1}}.$$  

Now, with $h = 4$, we similarly obtain

$$|\langle 4^m m! \lambda_{m,N} \rangle| = |4^m m! \lambda_{m,N}| < 2^m \delta$$

and iterating this process we eventually obtain

$$|\lambda_{m,N}| < \frac{2^m}{2^m m!} \delta$$

(21)

where $2^m$ is the largest power of 2 less than or equal to $\frac{N}{2m}$. By substituting $2^{m+1} > \frac{N}{2m}$ it follows that

$$N^m |\lambda_{m,N}| < \frac{2^m m! m^m}{m!} \delta$$

for all $N > N_0$. As $\delta$ is arbitrary, $N^m \lambda_{m,N} \to 0$ as $N \to \infty$.

We have now shown that the highest order coefficient $\lambda_{m,N}$ converges as required. The remaining coefficients will be shown to converge by induction. Assume that $N^k \lambda_{k,N} \to 0$ for all $k = r + 1, r + 2, \ldots, m$. That is, assume that the $m - r$ highest order coefficients all converge as required by the lemma. Let

$$z_{r,N}(n) = \lambda_{0,N} + \lambda_{1,N}n + \cdots + \lambda_{r,N}n^r.$$  

Because the $m - r$ highest order coefficients converge we can write $z_{r,N}(n) = z_{r,N}(n) + \gamma_N(n)$ where

$$\sup_{n \in \{1, \ldots, N\}} |\gamma_N(n)| \to 0 \quad \text{as} \quad N \to \infty.$$  

Now the bound from (19), but applied using the $r$th difference, gives

$$|\langle \Delta_h^n z_{r,N}(n') \rangle| = |\langle \Delta_h^n (\gamma_N(n') + \Delta_h^n z_{r,N}(n')) \rangle|$$

$$= |(\epsilon + h^r r! \lambda_{r,N})| < 2^r \delta,$$

(22)

where

$$\epsilon = \Delta_h^r \gamma_N(n') \leq 2^r \sup_{n \in \{1, \ldots, N\}} |\gamma_N(n)| \to 0$$
as $N \to \infty$. Choose $\delta$ and $\epsilon$ such that $2^r \delta < \frac{1}{4}$ and $|\epsilon| < \frac{1}{2}$. Then, from (22) and from Lemma 9,

$$|\langle h^r \hat{\tau} \lambda_r, N \rangle| < 2^r \delta + |\epsilon|$$

for all $h$ such that $1 \leq h \leq \frac{N}{2m}$. Choosing $2^r \delta + |\epsilon| < 2^{-2r-1}$ and using the same iterative process as for the highest order coefficient $\lambda_{m,H}$ (see (20) to (21)) we find that $N^r \lambda_{r,N} \to 0$ as $N \to \infty$. The proof now follows by induction.

Lemma 8. Let $a_1, a_2, \ldots, a_r$ be $r$ real numbers for which $|\langle a_n \rangle| < \delta$ for all $n = 1, 2, \ldots, r$. Then $|\langle \sum_{n=1}^r a_n \rangle| < r \delta$.

Proof: If $\delta > \frac{1}{2}$ the proof is trivial as $|\langle \sum_{n=1}^r a_n \rangle| \leq \frac{1}{2}$ for all $a_n \in \mathbb{R}$. If $\delta \leq \frac{1}{2}$ then $\langle \sum_{n=1}^r a_n \rangle = \sum_{n=1}^r \langle a_n \rangle$ and

$$\left| \sum_{n=1}^r \langle a_n \rangle \right| = \sum_{n=1}^r |\langle a_n \rangle| \leq r \delta.$$ 

Lemma 9. Let $|\langle a + \epsilon \rangle| < \delta$ where $|\epsilon| < \frac{1}{4}$ and $0 < \delta < \frac{1}{4}$. Then $|\langle a \rangle| < \delta + |\epsilon|$. 

Proof: By supposition $n - \delta < a + \epsilon < n + \delta$ for some $n \in \mathbb{Z}$. Since $-\delta - \epsilon < -\frac{1}{2}$ and $\delta - \epsilon < \frac{1}{2}$, it follows that $n - \frac{1}{2} < n - \delta - \epsilon < a < n + \delta - \epsilon < n + \frac{1}{2}$. Hence $\langle a \rangle = a$ and so

$$-\delta - |\epsilon| \leq -\delta - \epsilon < \langle a \rangle < \delta - \epsilon \leq \delta + |\epsilon|$$

and $|\langle a \rangle| \leq \delta + |\epsilon|$.

We are now in a position to complete the proof of strong consistency. Let $\Omega$ be the sample space on which the random variables $\{X_n\}$ and $\{\Phi_n\}$ are defined. Let $A$ be the subset of $\Omega$ on which $V_{2N}(\hat{\alpha}_N) - \sigma^2 \rightarrow 0$ as $N \rightarrow \infty$. Now $\Pr\{A\} = 1$ as a result of Lemma 4. Let $A'$ be the subset of $\Omega$ on which $N^{k} \lambda_{k,N} \rightarrow 0$ for $k = 0, \ldots, m$ as $N \rightarrow \infty$. As a result of Lemma 7, $A \subseteq A'$, and so $\Pr\{A'\} = \Pr\{A\} = 1$. Strong consistency follows.

References


[29] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point


Robby G. McKilliam is at the Institute for Telecommunications Research at the University of South Australia, Adelaide. His fields of interest are number theory, estimation theory, statistical signal processing and communications. He admires Lamport’s structured proofs and is a fan of reproducible research. In December 2010 he completed his PhD at the University of Queensland, Brisbane, Australia. In 2006 he completed undergraduate studies in computer systems engineering and mathematics at the University of Queensland.

Barry G. Quinn received a B.A. Hons degree in Pure Mathematics, Applied Mathematics and Statistics in 1978 and a PhD in Statistics in 1981, both from the Australian National University. He is currently Professor of Statistics at Macquarie University, Sydney. He has previously held Professorships at UMIST (now part of the University of Manchester) and Goldsmiths College, London. He has also held lectureships at the Universities of Wollongong and Queensland and a Senior Lectureship at the University of Newcastle, and was a Principal Research Scientist at DSTO, Adelaide. His interests are in Time Series Analysis and Signal Processing, and especially frequency estimation.

I. Vaughan L. Clarkson received a B.Sc. in Mathematics and a B.E. (Hons. I) in Computer Systems Engineering from The University of Queensland, Australia, in 1989 and 1990, respectively, and a Ph.D. in Systems Engineering from The Australian National University in 1997.

Starting in 1988, he was employed by the Defence Science and Technology Organisation, Australia, first as a Cadet, later as a Professional Officer, and finally as a Research Scientist. From 1998 to 2000, he was a Lecturer at The University of Melbourne, Australia. From 2000 to 2008, he was a Senior Lecturer in the School of Information Technology and Electrical Engineering at The University of Queensland. In 2008, he was promoted to Reader. He was a Visiting Professor at The University of British Columbia, Canada, in 2005 and at the Vienna University of Technology, Austria, in 2012. His research interests include statistical signal processing for communications and defence, image processing, information theory and lattice theory.

Bill Moran Professor Bill Moran is the Research Director of Defence Science Institute at University of Melbourne, where he has been a Professor of Electrical Engineering since 2001. Previously he was Professor of Mathematics ('76–'91), Head of the Department of Pure Mathematics ('77–79, '84–'86), Dean of Mathematical and Computer Sciences ('81, '82, '89) at the University of Adelaide, and Head of the Mathematics Discipline at the Flinders University of South Australia (91–95’). He was a Chief Investigator ('92–'95), and Head of the Medical Signal Processing Program (’95–’99) in the Cooperative Research Centre for Sensor Signal and Information Processing. He was elected to the Fellowship of the Australian Academy of Science in 1984. He holds a Ph.D. in Pure Mathematics from the University of Sheffield, UK (’68), and a First Class Honours B.Sc. in Mathematics from the University of Birmingham (’65).

He has been a Principal Investigator on numerous research grants and contracts, in areas spanning pure mathematics to radar development, from both Australian and US Research Funding Agencies, including DARPA, AFOSR, AFRL, Australian Research Council (ARC), Australian Department of Education, Science and Training, and DSTO. He was a member of the Australian Research Council College of Experts from 2007 to 2009.

His main areas of research interest are in signal processing both theoretically and in applications to radar, waveform design and radar theory, sensor networks, sensor management, and information geometry. He also works in various areas of mathematics including harmonic analysis, representation theory, and number theory.

Badri N. Vellambi Badri N. Vellambi (M09) received the B.Tech. in electrical engineering from the Indian Institute of Technology-Madras in 2002. He received his M.S. (Electrical), M.S. (Mathematics), and Ph.D. degrees (Electrical Engineering) from the Georgia Institute of Technology, Atlanta, in 2005, 2008, and 2008, respectively.

Since 2008, he has been a research fellow at the Institute for Telecommunications Research (ITR) in the University of South Australia (UniSA). His current research interests include Shannon theory, channel coding, multi-user information theory and communications theory.