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 2π 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

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

where t is a real number usually representing time and

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

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

$$s_n = s(n) = e^{2\pi j y(n)} \qquad n \in \mathbb{Z}.$$

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

$$Y_n = \rho s_n + X_n \qquad n = 1, \dots, N, \tag{1}$$

where ρ 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 may exist for these special lattices, that are yet to be discovered. In this paper we apply another general purpose approximate nearest point algorithm called the K-best algorithm [33–35] for approximating the

Robby McKilliam and Badri Vellambi are with the Institute for Telecommunications Research, The University of South Australia, SA, 5095. Barry Quinn is with the Department of Statistics, Macquarie University, Sydney, NSW, 2109, Australia. Vaughan Clarkson is with the School of Information Technology & Electrical Engineering, The University of Queensland, QLD., 4072, Australia. B. Moran is with the Defense Science Institute, University of Melbourne, Vic. 3010, Australia.

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 $\{1, 2, \dots, N\}$ [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 $\frac{2}{3}\pi$ for low SNR. Related results have been presented in [6]. This factor of $\frac{2}{3}\pi$ 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.

The paper is organised in the following way. Section II describes some required concepts from lattice theory. In Section III we describe 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.

II. Lattice Theory

A *lattice*, Λ, is a discrete subset of points in $\mathbb{R}^n$ for which

$$\Lambda = \{\mathbf{x} = \mathbf{B}\mathbf{u} \;;\; \mathbf{u} \in \mathbb{Z}^d\}$$

where $\mathbf{B} \in \mathbb{R}^{n\times d}$ has rank d and is called the generator matrix. If $n = d$ the lattice is said to be of full rank. Lattices are discrete Abelian groups under vector addition. They are subgroups of the Euclidean group $\mathbb{R}^n$. Lattices naturally give rise to tessellations of $\mathbb{R}^n$ by the specification of a set of coset representatives for the quotient $\mathbb{R}^n/\Lambda$. One choice for a set of coset representatives is a fundamental parallelepiped, the parallelepiped generated by the columns of a generator matrix. Another choice is based on the Voronoi cell, those points from $\mathbb{R}^n$ nearest (with respect to the Euclidean norm in this paper) to the lattice point at the origin. It is always possible

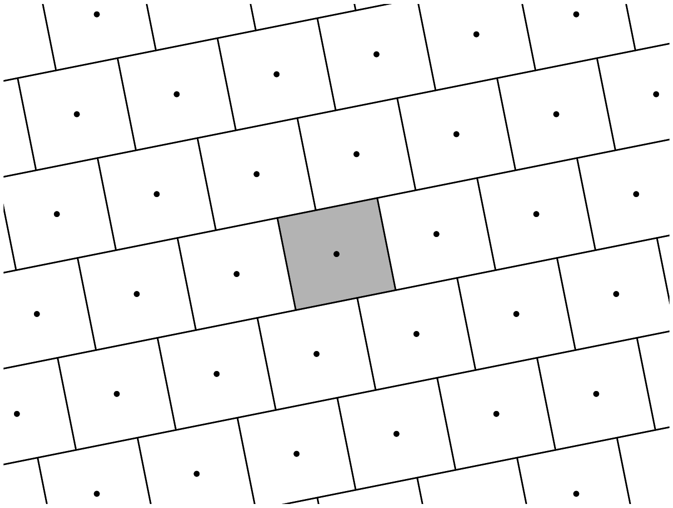

Fig. 1. Rectangular tessellation constructed according to Proposition 1 where Λ is a 2 dimensional lattice with generator matrix having columns $[1, 0.2]'$ and $[0.2, 1]'$. Any one of the boxes is a rectangular set of coset representatives for $\mathbb{R}^2/\Lambda$. The shaded box centered at the origin is the one given by Proposition 1.

to construct a rectangular set of representatives, as the next proposition will show. We will use these rectangular regions to describe the aliasing properties of polynomial phase signals in Section III. These rectangular regions will be important for the derivation of the asymptotic properties of the LSU estimator.

Proposition 1. *Let Λ be an n dimensional lattice and $\mathbf{B} \in \mathbb{R}^{n\times n}$ be a generator matrix for Λ. Let $\mathbf{B} = \mathbf{QR}$ where $\mathbf{Q}$ is orthonormal and $\mathbf{R}$ is upper triangular with elements r_{ij}. Then the rectangular prism $\mathbf{Q}P$ where*

$$P = \prod_{k=1}^{n} \Big[-\frac{r_{kk}}{2}, \frac{r_{kk}}{2} \Big)$$

is a set of coset representatives for $\mathbb{R}^n/\Lambda$.

Proof: This result is well known [50, Chapter IX, Theorem IV] [10, Proposition 2.1]. ∎

A fundamental problem in lattice theory is the *nearest lattice point problem* [29]: Given $\mathbf{y} \in \mathbb{R}^n$ and some lattice $\Lambda \subset \mathbb{R}^n$, find a lattice point $\mathbf{x} \in \Lambda$ such that the Euclidean distance between $\mathbf{y}$ and $\mathbf{x}$ is minimised, that is, find $\arg\min_{\mathbf{x}\in\Lambda} \|\mathbf{y}-\mathbf{x}\|^2$. The nearest lattice point problem is known to be NP-hard 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

$$s(t) = e^{2\pi j y(t)} \qquad \text{and} \qquad r(t) = e^{2\pi j z(t)}$$

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

$$\begin{aligned} s(n) &= e^{2\pi j y(n)} = e^{2\pi j y(n)} e^{2\pi j p(n)} \\ &= e^{2\pi j (y(n)+p(n))} = e^{2\pi j z(n)} = r(n) \end{aligned}$$

because $p(n)$ is always an integer and therefore $e^{2\pi j p(n)} = 1$ 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 2π) of two distinct polynomial phase signals is plotted on the left, and the principal component of the phase (also divided by 2π) 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

$$p_k(x) = \binom{x}{k} = \frac{x(x-1)(x-2)\dots(x-k+1)}{k!},$$

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

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

$$c_0 p_0 + c_1 p_1 + \dots + c_m p_m, \tag{2}$$

where the $c_i \in \mathbb{Z}$.

Proof: See [58, p. 2] or [5]. ■

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

$$\operatorname{coef}(g) = \begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_m \end{bmatrix}',$$

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

$$L_{m+1} = \{\operatorname{coef}(p) \; ; \; p \in \mathcal{Z}\}$$

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

$$\begin{aligned} L_{m+1} &= \{\operatorname{coef}(c_0 p_0 + c_1 p_1 + \dots + c_m p_m) \; ; \; c_i \in \mathbb{Z}\} \\ &= \{c_0 \operatorname{coef}(p_0) + \dots + c_m \operatorname{coef}(p_m) \; ; \; c_i \in \mathbb{Z}\}. \end{aligned}$$

Let

$$\mathbf{P} = \begin{bmatrix} \operatorname{coef}(p_0) & \operatorname{coef}(p_1) & \dots & \operatorname{coef}(p_m) \end{bmatrix}$$

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

$$L_{m+1} = \{\mathbf{x} = \mathbf{P}\mathbf{u} \; ; \; \mathbf{u} \in \mathbb{Z}^{m+1}\}$$

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 $\mathbf{P}$. 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 $\operatorname{coef}(y)$ and $\operatorname{coef}(z)$ 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 $\tilde{\boldsymbol{\mu}}$, 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*.

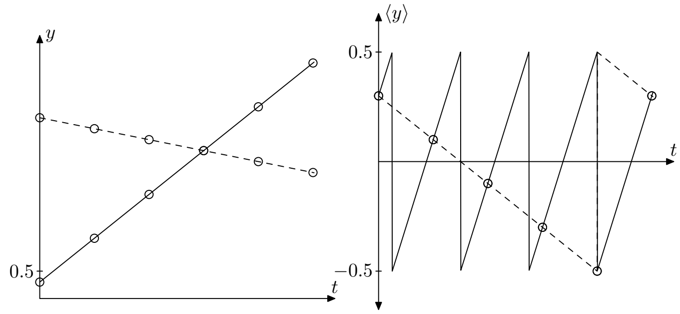

Fig. 2. The first order polynomials $\frac{1}{10}(3+8t)$ (solid) and $\frac{1}{10}(33-2t)$ (dashed line).

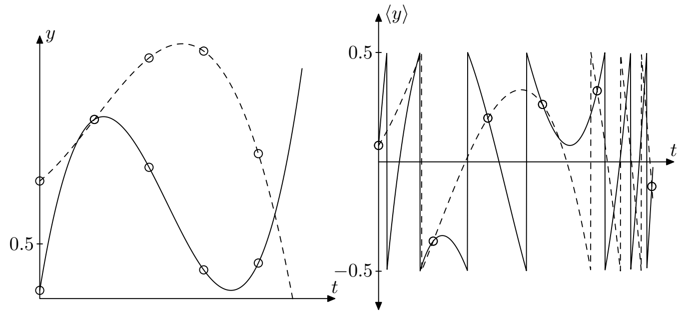

Fig. 3. The cubic polynomials $\frac{1}{160}(174+85t-118t^2+40t^3)$ (solid line) and $\frac{1}{48}(84+19t+12t^2-4t^3)$ (dashed line).

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 μ_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 *Nyqvist 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 $\mathbf{P}$ is upper triangular with kth diagonal element equal to $\frac{1}{k!}$. So this rectangular region is

$$B = \prod_{k=0}^{m} \left[-\frac{0.5}{k!}, \frac{0.5}{k!}\right). \tag{3}$$

We will make use of this region when deriving the statistical properties of the LSU estimator in the next section.

Given $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^{m+1}$ we say that $\mathbf{x} \equiv \mathbf{y} \bmod L_{m+1}$ if $\mathbf{x}$ and $\mathbf{y}$ differ by a lattice point in L_{m+1}. We define the function $\operatorname{dealias}(\mathbf{x})$ to take $\mathbf{x}$ to its coset representative inside B. That is, $\operatorname{dealias}(\mathbf{x}) = \mathbf{z} \in B$ where $\mathbf{x} - \mathbf{z} \in L_{m+1}$. When $m = 0$ or 1 then $\operatorname{dealias}(\mathbf{x}) = \langle \mathbf{x} \rangle$ where $\langle \mathbf{x} \rangle = \mathbf{x} - \lceil \mathbf{x} \rfloor$ denotes the (centered) fractional part and $\lceil \mathbf{x} \rfloor$ denotes the nearest integer to $\mathbf{x}$ and both $\langle\cdot\rangle$ and $\lceil\cdot\rfloor$ 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 $\operatorname{dealias}(\mathbf{x})$ 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 $\tilde{\mu}_0, \dots, \tilde{\mu}_m$ from the observations $Y_1, \dots, Y_N$ defined in (1). Let

$$\Theta_n = \frac{\angle Y_n}{2\pi} = \langle \Phi_n + y(n) \rangle \tag{4}$$

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

$$\Phi_n = \frac{1}{2\pi} \angle (1 + \rho^{-1} s_n^{-1} X_n)$$

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 Φ_n is the same as the distribution of $\frac{1}{2\pi}\angle(1 + \rho^{-1} X_n)$. If the $X_1, \dots, X_N$ are circularly symmetric and identically distributed, then $\Phi_1, \dots, \Phi_n$ are also identically distributed.

Let $\boldsymbol{\mu}$ be the vector $[\mu_0, \mu_1, \dots, \mu_m]$ and put,

$$SS(\boldsymbol{\mu}) = \sum_{n=1}^{N} \left\langle \Theta_n - \sum_{k=0}^{m} \mu_k n^k \right\rangle^2. \tag{5}$$

The LSU estimator $\widehat{\boldsymbol{\mu}}$ is the minimiser of SS over the identifiable region B, that is,

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu} \in B} SS(\boldsymbol{\mu}). \tag{6}$$

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

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

where $W_n = \lceil \Theta_n - \sum_{k=0}^{m} \mu_k n^k \rfloor$ for $n = 1, \dots, N$ are integers called *wrapping variables*. If we consider $W_1, \dots, W_N$ as nuisance parameters to be estimated then SS can be considered as a function of both $\mu_0, \dots, \mu_m$ and $W_1, \dots, W_N$, the minimiser with respect to $W_1, \dots, W_N$ being $SS(\boldsymbol{\mu})$. The LSU estimator can be found by jointly minimising over $\mu_0, \dots, \mu_m$ and $W_1, \dots, W_N$. 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 $\mu_0, \dots, \mu_m$ and $W_1, \dots, W_N$,

$$SS(\boldsymbol{\mu}, \mathbf{w}) = \| \boldsymbol{\theta} - \mathbf{X}\boldsymbol{\mu} - \mathbf{w} \|^2, \tag{7}$$

where $\|\cdot\|^2$ denotes the squared Euclidean norm, $\mathbf{X}$ is the rectangular Vandermonde matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & 2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & N & N^2 & \cdots & N^m \end{bmatrix}, \tag{8}$$

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

$$\boldsymbol{\mu}(\mathbf{w}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\theta} - \mathbf{w}) = \mathbf{X}^{+}(\boldsymbol{\theta} - \mathbf{w}), \tag{9}$$

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

$$SS(\mathbf{w}) = SS(\boldsymbol{\mu}(\mathbf{w}), \mathbf{w}) = \| \mathbf{Q}\boldsymbol{\theta} - \mathbf{Q}\mathbf{w} \|^2,$$

where $\mathbf{Q} = \mathbf{I} - \mathbf{X}\mathbf{X}^{+}$ and $\mathbf{I}$ is the $N \times N$ identity matrix[1].

Let Λ be the lattice generated by $\mathbf{Q}$, that is $\Lambda = \{\mathbf{Q}\mathbf{x} \,;\, \mathbf{x} \in \mathbb{Z}^N\}$[2]. It follows that $\mathbf{Q}\mathbf{w}$ is a lattice point in Λ and that minimising $SS(\mathbf{w})$ is equivalent to finding the nearest lattice point in Λ to $\mathbf{Q}\boldsymbol{\theta}$. Denote this point by $\mathbf{Q}\widehat{\mathbf{w}}$. Now the estimate $\widehat{\boldsymbol{\mu}}$ is given by substituting $\widehat{\mathbf{w}}$ for $\mathbf{w}$ in (9). After this procedure it is possible that the $\widehat{\boldsymbol{\mu}}$ obtained is not in the identifiable region but is instead an aliased version of the desired estimate. This can be resolved by computing $\operatorname{dealias}(\widehat{\boldsymbol{\mu}})$. There exist algorithms that can compute a nearest point in Λ 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 $\Phi_1, \dots, \Phi_N$, which are *circular* random variables with support on $[-1/2, 1/2)$ [9, 10, 59, 60]. Circular random variables are often considered modulo 2π and therefore have support $[-\pi, \pi)$ with $-\pi$ and π 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 $\lceil\cdot\rfloor$ for rounding and $\langle\cdot\rangle$ for the centered fractional part in a convenient way.

The *intrinsic mean* (or *Fréchet mean*) of Φ_n is defined as [9, 61],

$$\mu_{\text{intr}} = \arg\min_{\mu \in [-1/2, 1/2)} \mathbb{E} \langle \Phi_n - \mu \rangle^2, \tag{10}$$

and the *intrinsic variance* is

$$\sigma_{\text{intr}}^2 = \mathbb{E} \langle \Theta - \mu_{\text{intr}} \rangle^2 = \min_{\mu \in [-1/2, 1/2)} \mathbb{E} \langle \Phi_n - \mu \rangle^2,$$

where $\mathbb{E}$ denotes expected value. Depending on the distribution of Φ_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 Φ_n has no intrinsic mean. We are now equipped to state the asymptotic properties of the LSU estimator.

Theorem 1. *Let $\widehat{\boldsymbol{\mu}}$ be defined by (6) and put $\widehat{\boldsymbol{\lambda}}_N = \operatorname{dealias}(\tilde{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}})$. Denote the elements of $\widehat{\boldsymbol{\lambda}}_N$ by*

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

[2] The $N \times N$ projection matrix $\mathbf{Q}$ 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 $\mathbf{Q}$. In our implementation the last $m + 1$ columns are removed.

$\widehat{\lambda}_{0,N}, \dots, \widehat{\lambda}_{m,N}$. *Suppose $\Phi_1, \dots, \Phi_N$ are independent and identically distributed with zero intrinsic mean, intrinsic variance σ^2, and pdf f, then:*

1) *(Strong consistency) $N^k \widehat{\lambda}_{k,N}$ converges almost surely to 0 as $N \to \infty$ for all $k = 0, 1, \dots, m$.*
2) *(Asymptotic normality) If $f(\langle x \rangle)$ is continuous at $x = -1/2$ and $f(-1/2) < 1$ then the distribution of*

$$\left[\begin{array}{cccc} \sqrt{N}\widehat{\lambda}_{0,N} & N\sqrt{N}\widehat{\lambda}_{1,N} & \dots & N^m \sqrt{N}\widehat{\lambda}_{m,N} \end{array} \right]'$$

converges to the normal with zero mean and covariance matrix

$$\frac{\sigma^2}{(1 - f(-1/2))^2} \mathbf{C}^{-1},$$

where $\mathbf{C}$ is the $(m+1) \times (m+1)$ Hilbert matrix with elements $C_{ik} = 1/(i+k+1)$ for $i, k \in \{0, 1, \dots, m\}$.

A proof of the first part of this theorem (strong consistency) is given in Appendix A. Space restrictions make it impossible 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 $\operatorname{dealias}(\tilde{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}})$ between the true coefficients $\tilde{\boldsymbol{\mu}}$ and the estimated coefficients $\widehat{\boldsymbol{\mu}}$ rather than directly on the difference $\tilde{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}$. To see why this makes sense, consider the case when $m = 0$, $\tilde{\mu}_0 = -0.5$ and $\widehat{\mu}_0 = 0.49$, so that $\tilde{\mu}_0 - \widehat{\mu}_0 = -0.99$. However, the two phases are obviously close, since the phases ± 0.5 are actually the same. In this case

$$\operatorname{dealias}(\tilde{\mu}_0 - \widehat{\mu}_0) = \langle \tilde{\mu}_0 - \widehat{\mu}_0 \rangle = 0.01.$$

The same reasoning holds for $m > 0$.

The requirement that $\Phi_1, \dots, \Phi_N$ be identically distributed will typically hold only when the complex random variables $X_1, \dots, X_N$ are identically distributed and circularly symmetric. It would be possible to drop the assumption that $\Phi_1, \dots, \Phi_N$ be identically distributed, but this complicates the theorem statement and the proof. In the interest of simplicity we only consider the case when $\Phi_1, \dots, \Phi_N$ are identically distributed here. If X_n is circularly symmetric with pdf nonincreasing with magnitude $|X_n|$, then, the corresponding Φ_n necessarily has zero intrinsic mean [10, Theorem 5.2]. Thus, our theorem covers commonly used distributions for $X_1, \dots, X_N$, such as the normal distribution. No assumptions about the magnitude of $X_1, \dots, X_N$ are required, that is, we do not require the noise to be 'small' in any sense. For example, if $X_1, \dots, X_N$ are normally distributed with variance σ_c^2, then the theorem holds regardless of how large σ_c^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 $\Phi_1, \dots, \Phi_N$ 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., when $f(-1/2) = 1$ or when $f(\langle x \rangle)$ is discontinuous at $x = -1/2$. In this exceptional case 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], the cubic phase function (CPF) estimator [11] and the estimator that results from combining the HAF and CPF [21]. 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 $X_1, \dots, X_N$ are pseudorandomly generated independent and identically distributed circularly symmetric complex Gaussian with variance $\mathbb{E}|X_1|^2 = \sigma_c^2$. 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 $\widehat{\boldsymbol{\mu}}_1, \dots, \widehat{\boldsymbol{\mu}}_T$ and the corresponding dealiased errors $\widehat{\boldsymbol{\lambda}}_t = \operatorname{dealias}(\tilde{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}_t)$ are computed. The sample mean square error (MSE) of the kth coefficient is computed according to $\frac{1}{T}\sum_{t=1}^{T} \widehat{\lambda}_{k,t}^2$ where $\widehat{\lambda}_{k,t}$ is the kth element of $\widehat{\boldsymbol{\lambda}}_t$. In all simulations four *lag sets* are used for the PHAF estimator [19] and the LSU estimator is approximated using the K-best method described in Section II. The maximum number of branches allowed is $K = 20N$.

Figure 4 shows the sample MSE of the highest order coefficient μ_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 (so that the 'threshold' is avoided) the sample MSE of the LSU estimator is close to that predicted by Theorem 1. Using Theorem 1 we can predict that, asymptotically in N, the ratio of the variance of the LSU estimator to the CRB will approach 1 for high SNR and $\frac{2}{3}\pi$ for low SNR (see Lemma 3 and the subsequent discussion in [6]). Contrastingly, the ratio of the variance of the HAF, PHAF and CPF estimators to the CRB is known to diverge as the SNR becomes small [11, 15, 21].

In Figure 4 the CPF estimator reaches a threshold at 1 dB, the LSU and PHAF estimators at 3 dB and the HAF estimator at 5 dB. In Figure 4 the true coefficients are $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8N}, \frac{1}{24N^2}\right]$. These are specifically chosen to satisfy the requirements of the HAF, PHAF and CPF estimators. The HAF and PHAF estimators require that the coefficients satisfy $\tilde{\mu}_0, \tilde{\mu}_1 \in [-0.5, 0.5)$ and

$$\tilde{\mu}_k \in \left[-\frac{0.5}{k!\tau^{k-1}}, \frac{0.5}{k!\tau^{k-1}}\right) \qquad k = 2, 3, \dots, m,$$

where τ is the *lag* and we choose $\tau = \lfloor N/m \rfloor$ as suggested in [13, 19]. The CPF estimator requires similar, but slightly more complicated conditions on the coefficients [11]. It is informative to compare the volume of the set of coefficients suitable for the HAF, PHAF and CPF estimators with the volume of the identifiable region B from (3). The identifiable region has volume $\operatorname{vol}(B) = \prod_{k=0}^{m} \frac{1}{k!}$. The volume of coefficients suitable for the HAF and PHAF estimators is

$$\prod_{k=2}^{m} \frac{1}{k!\tau^{k-1}} = \operatorname{vol}(B) \prod_{k=2}^{m} \frac{1}{\tau^{k-1}} \leq \operatorname{vol}(B) \prod_{k=2}^{m} \left(\frac{m}{N}\right)^{k-1}.$$

The volume of the coefficients suitable for the CPF estimator can be shown to be $\frac{3}{2N^3} = 18 \operatorname{vol}(B) N^{-3}$. It is now worth reconsidering the results in Figure 4. Substituting $m = 3$ and $N = 199$ in the formula above, we see that the HAF, PHAF and CPF estimators search a space of parameters that is less than 10^{-5} the size of the identifiable region searched by the LSU estimator. In this context we might expect the HAF, PHAF and CPF estimators to have thresholds at lower SNR than the LSU estimator. It is encouraging that the LSU estimator has comparable performance. To put these results further into context, Figure 5 repeats the simulations in Figure 4, but where the true coefficients are $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{24}\right]$. The HAF, PHAF and CPF estimators fail consistently, as expected. The performance of the LSU estimator is unchanged, as expected in view of Theorem 1.

The comparatively small set of coefficients suitable for the HAF, PHAF and CPF estimators has been studied previously [5, 25, 45–47, 65] and can be a critical limitation for some applications, for example, interference suppression in the presence of a jammer [46, 48]. Zhou and Wang [47] describe a method based on the extended Euclidean algorithm that enables the HAF estimator to operate on the entire identifiable region B. Unfortunately this comes at the price of statistical accuracy. Figure 6 shows the performance of the Zhou and Wang estimator when the coefficients are $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{24}\right]$, the same as in Figure 5. The threshold for the estimator is approximately 29 dB. There does not appear to exist (at least in the literature) a similar method for increasing the range of coefficients for the PHAF and CPF estimators.

Figure 7 displays the sample MSE of the highest order coefficient μ_5 for a polynomial phase signal of order $m = 5$. In this case the LSU and PHAF estimators reach a threshold at 12 dB, the HAF estimator at 18 dB, and the estimator that combines the CPF and HAF at 10 dB. In this figure the true coefficients are $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8N}, \frac{1}{24N^2}, \frac{1}{96N^3}, \frac{1}{480N^4}\right]$ and these are chosen specifically to satisfy the requirement of the HAF, PHAF and CPF-HAF estimators. As in the case where $m = 3$, the region of coefficients suitable for the HAF, PHAF and CPF-HAF estimators is much smaller that the identifiable region. The effect becomes more severe as m increases. When $m = 5$ and $N = 199$ the space of coefficients searched by these estimators is approximately 10^{-16} the size of the identifiable region. This is a dramatic reduction in the range of detectable polynomial phase signals. Figure 8 displays results under the same conditions as Figure 7 but with $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{24}, \frac{1}{96}, \frac{1}{480}\right]$. The conclusion is similar to that from Figure 5. The performance of the LSU estimator remains unchanged, but the HAF, PHAF and CPF-HAF estimators fail completely. The Zhou and Wang estimator can again be used to increase the range of parameters for the HAF, but simulations indicate that the threshold for the estimator is approximately 103 dB.

Tables I and II ... BLERG

	LSU	CPF	PHAF	HAF	ZW
$N = 49$					
$N = 99$	6	3	5	7	27
$N = 199$	3	1	3	5	23
$N = 299$	2	-1		4	
$N \to \infty$	$-\infty$	?	?	3.77	?

TABLE I
ESTIMATED SNR THRESHOLD IN DECIBELS WITH $m = 3$

	LSU	CPF-HAF	PHAF	HAF	ZW
$N = 49$					
$N = 99$	15	12	13	19	> 100
$N = 199$	12	10	12	18	103
$N = 299$					
$N \to \infty$	$-\infty$	≈ 6.5	?	16.6	?

TABLE II
ESTIMATED SNR THRESHOLD IN DECIBELS WITH $m = 5$

Figure 9 shows the running time of the LSU estimator 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 $O(N \log N)$. Wtih $K = 20N$ 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 lattice 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

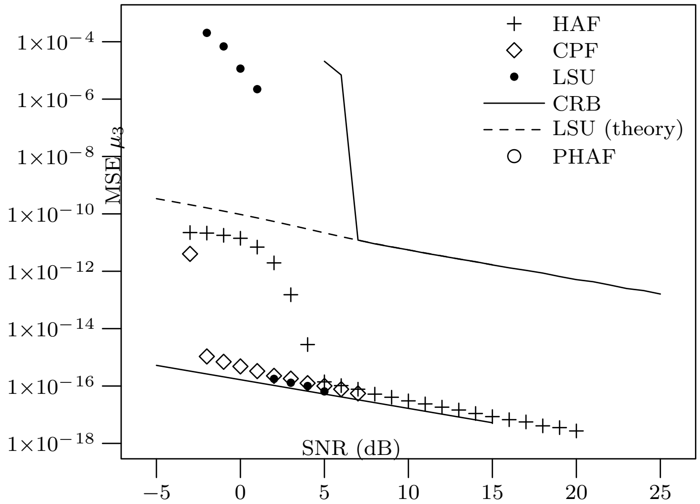

Fig. 4. Sample MSE for coefficient $\tilde{\mu}_3$ for a polynomial phase signal of order 3 with $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8N}, \frac{1}{24N^2}\right]$ and $N = 199$.

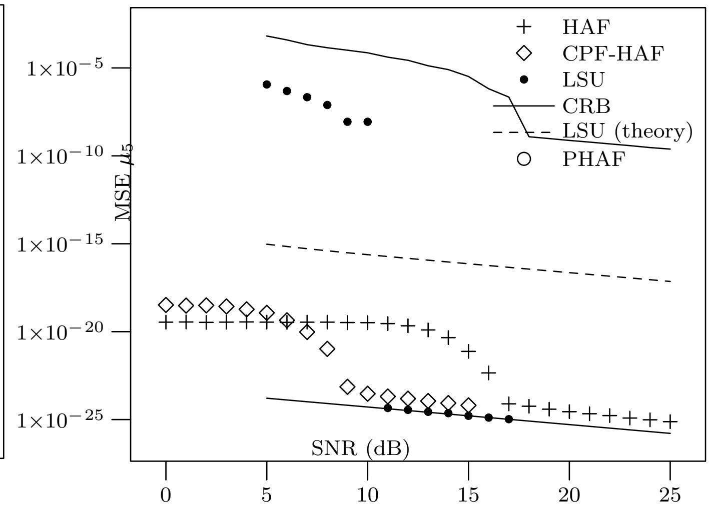

Fig. 7. Sample MSE for the coefficient $\tilde{\mu}_5$ for a polynomial phase signal of order 5 with $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8N}, \frac{1}{24N^2}, \frac{1}{96N^3}, \frac{1}{480N^4}\right]$ and $N = 199$.

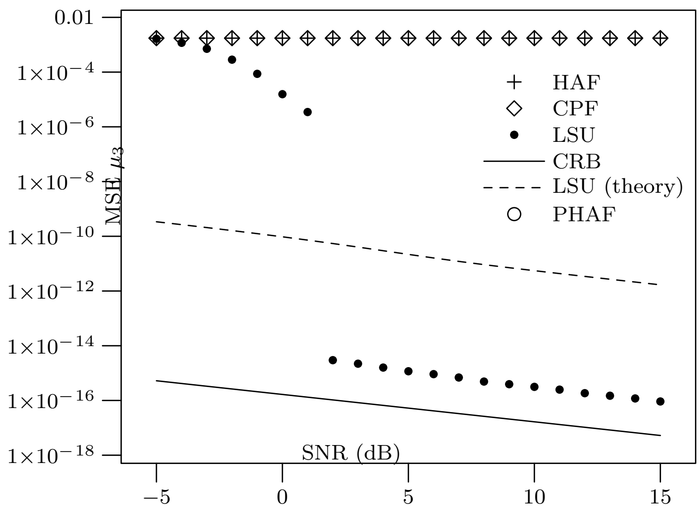

Fig. 5. Sample MSE for the coefficient $\tilde{\mu}_3$ for a polynomial phase signal of order 3 with $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{24}\right]$ and $N = 199$.

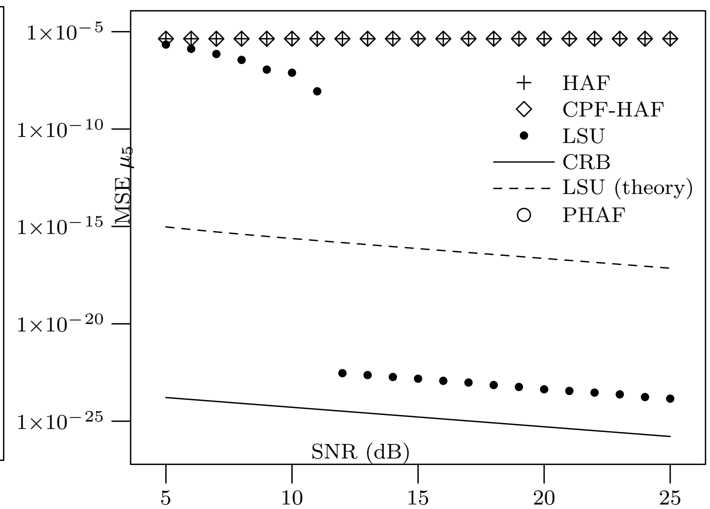

Fig. 8. Sample MSE for the coefficient $\tilde{\mu}_5$ for a polynomial phase signal of order 5 with $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{24}, \frac{1}{96}, \frac{1}{480}\right]$ and $N = 199$.

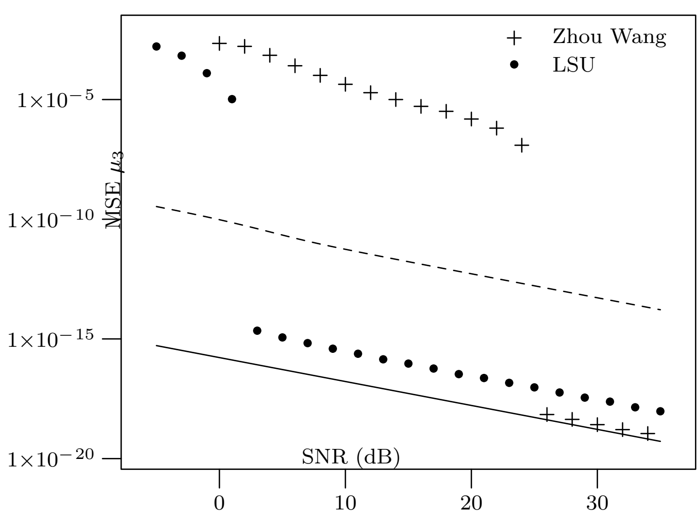

Fig. 6. Sample MSE for the coefficient $\tilde{\mu}_3$ for a polynomial phase signal of order 3 with $\tilde{\boldsymbol{\mu}} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{24}\right]$ and $N = 199$.

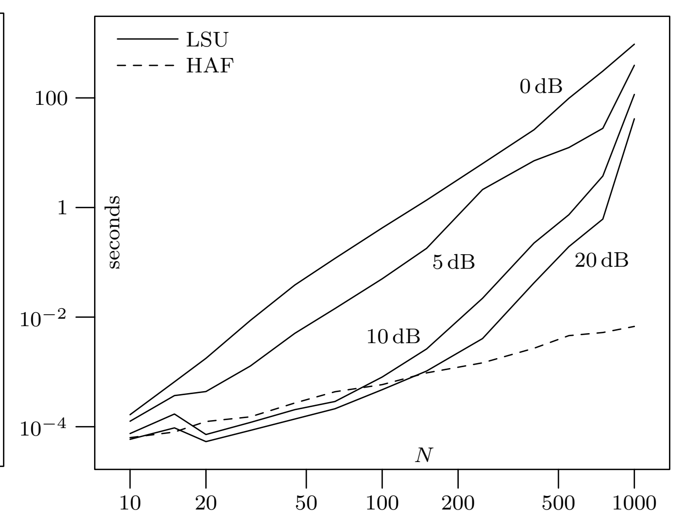

Fig. 9. Running time in seconds with the LSU and HAF estimators. The LSU estimator is benchmarked at SNR equal to 0 dB, 5 dB, 10 dB and 20 dB.

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

$$\begin{aligned} SS(\boldsymbol{\mu}) &= \sum_{n=1}^{N} \left\langle \left\langle \Phi_n + \sum_{k=0}^{m} \tilde{\mu}_k n^k \right\rangle - \sum_{k=0}^{m} \mu_k n^k \right\rangle^2 \\ &= \sum_{n=1}^{N} \left\langle \Phi_n + \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k) n^k \right\rangle^2 . \end{aligned}$$

Let $\boldsymbol{\lambda} = \operatorname{dealias}(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \tilde{\boldsymbol{\mu}} - \boldsymbol{\mu} - \mathbf{p}$ where $\mathbf{p}$ 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

$$\begin{aligned} \left\langle \sum_{k=0}^{m} \lambda_k n^k \right\rangle &= \left\langle \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k - p_k) n^k \right\rangle \\ &= \left\langle \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k) n^k \right\rangle . \end{aligned}$$

Let

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

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

$$-\frac{0.5}{k!} \leq \lambda_k < \frac{0.5}{k!}. \tag{11}$$

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

$$V_N(\boldsymbol{\lambda}) = \mathbb{E} S_N(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E} \left\langle \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right\rangle^2 .$$

Using techniques described in [68, 69] one can show that

$$\sup_{\boldsymbol{\lambda} \in B} |S_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda})| \to 0 \tag{12}$$

almost surely as $N \to \infty$. This type of result is called a *uniform law of large numbers* [68, 69]. A full proof of (12) in given in [9]. We now concentrate attention on the minimiser of V_N. Because Φ_n has zero intrinsic mean

$$\mathbb{E} \langle \Phi_n + z \rangle^2 \tag{13}$$

is minimised uniquely at $z = 0$ for $z \in [-1/2, 1/2)$. Since the intrinsic variance of Φ_n is σ^2, when $z = 0$,

$$\mathbb{E} \langle \Phi_1 + z \rangle^2 = \mathbb{E} \langle \Phi_1 \rangle^2 = \sigma^2, \tag{14}$$

and so the minimum attained value is σ^2.

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

Proof: Put $z(n) = \lambda_0 + \lambda_1 n + \cdots + \lambda_m n^m$. Then

$$\begin{aligned} V_N(\boldsymbol{\lambda}) &= \frac{1}{N} \mathbb{E} \sum_{n=1}^{N} \left\langle \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right\rangle^2 \\ &= \frac{1}{N} \sum_{n=1}^{N} \mathbb{E} \langle \Phi_n + \langle z(n) \rangle \rangle^2 . \end{aligned}$$

From above $\mathbb{E} \langle \Phi_n + \langle z(n) \rangle \rangle^2$ has minimum value σ^2, uniquely when $\langle z(n) \rangle = 0$. But, $\langle z(n) \rangle$ is zero for all integers n if and only if $z \in \mathcal{Z}$, or equivalently if $\operatorname{coef}(z)$ is a lattice point in L_{m+1}. By definition B contains precisely one lattice point from L_{m+1}, namely $\mathbf{0}$. Therefore V_N is minimised uniquely at $\mathbf{0}$, at which point it takes the value σ^2. ■

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

Proof: Since $\widehat{\boldsymbol{\lambda}}_N = \arg\min_{\boldsymbol{\lambda} \in B} S_N(\boldsymbol{\lambda})$,

$$0 \leq S_N(\mathbf{0}) - S_N(\widehat{\boldsymbol{\lambda}}_N).$$

Also, because V_N is minimised at $\mathbf{0}$, it follows that

$$\begin{aligned} 0 &\leq V_N(\widehat{\boldsymbol{\lambda}}_N) - V_N(\mathbf{0}) \\ &\leq V_N(\widehat{\boldsymbol{\lambda}}_N) - V_N(\mathbf{0}) + S_N(\mathbf{0}) - S_N(\widehat{\boldsymbol{\lambda}}_N) \\ &\leq |V_N(\widehat{\boldsymbol{\lambda}}_N) - S_N(\widehat{\boldsymbol{\lambda}}_N)| + |S_N(\mathbf{0}) - V_N(\mathbf{0})| \end{aligned}$$

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 $\mathbf{0}$, that $V_N(\mathbf{0}) = \sigma^2$, and that $V_N(\widehat{\boldsymbol{\lambda}}_N)$ converges almost surely to σ^2. We could now proceed to prove that $\widehat{\boldsymbol{\lambda}}_N$ converges almost surely to $\mathbf{0}$, 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

$$A(h, K) = \{ n \, ; \, n + ih \in K \ \forall \, i \in \{0, 1, \ldots, m\} \} \tag{15}$$

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 $K\backslash\{r\}$ 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 $n, n+h, \dots n+mh$ of length $m+1$. That is,*

$$|A(h, K\backslash\{r\})| \geq |A(h,K)| - (m+1).$$

Proof: The proof follows because there are at most $m+1$ integers, n, such that $n + ih = r$ for some $i \in \{0,1,\dots,m\}$. That is, there are at most $m+1$ arithmetic progressions of type $n, n+h, \dots n+mh$ 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 $1 \leq h \leq \frac{N}{2m}$ the set K contains at least one arithmetic progression $n, n+h, \dots, n+mh$ of length $m+1$. That is, $|A(h,K)| > 0$.*

Proof: By substituting the bounds $|K| > \frac{2m+1}{2m+2}N$ and $h \leq \frac{N}{2m}$ 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 $d(n)\colon \mathbb{R} \to \mathbb{R}$, let

$$\Delta_h^1 d(n) = d(n+h) - d(n)$$

denote the first difference with interval h, and let

$$\begin{aligned} \Delta_h^r d(n) &= \Delta_h^{r-1} d(n+h) - \Delta_h^{r-1} d(n) \\ &= \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} d(n+kh) \end{aligned} \tag{16}$$

denote the rth 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 the

$$d(n),\ d(n+h),\ \dots,\ d(n+kh)$$

precisely 2^r times.

The operator Δ_h^r has special properties when applied to polynomials. If $d(n) = a_r n^r + \dots + a_0$ is a polynomial of order r then

$$\Delta_h^r d(n) = h^r r! a_r. \tag{17}$$

Thus, the rth difference of $d(n)$ is a constant depending on h, r and the rth coefficient a_r [70, page 51]. We can now continue the proof of strong consistency. The next lemma is a key result.

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

Proof: Define the function

$$g(z) = \mathbb{E}\langle \Phi_1 + z \rangle^2 - \sigma^2 \tag{18}$$

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

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

as $N \to \infty$. Let

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

so that

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

as $N \to \infty$. Choose constants

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

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 γ 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 γ is strictly greater than 0 and

$$\begin{aligned} V_N(\boldsymbol{\lambda}_N) - \sigma^2 &= \frac{1}{N}\sum_{n=1}^{N} g\left(\langle z_N(n) \rangle\right) \\ &\geq \frac{1}{N}\sum_{n \in K_N} \gamma = (1-c)\gamma, \end{aligned}$$

contradicting the fact that $V_N(\boldsymbol{\lambda}_N) - \sigma^2$ 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 $n', n'+h, \dots, n'+mh$ are in K_N and therefore

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

are all less than δ. Because the mth difference is a linear combination of 2^m elements (see (16)) from

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

all with magnitude less than δ we obtain, from Lemma 8,

$$|\langle \Delta_h^m z_N(n') \rangle| \leq |\Delta_h^m \langle z_N(n') \rangle| < 2^m \delta. \tag{19}$$

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

$$|\langle h^m m! \lambda_{m,N} \rangle| = |\langle \Delta_h^m z_N(n') \rangle| < 2^m \delta \tag{20}$$

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 $\delta < \frac{1}{2^{2m}}$ 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

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

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

$$N^m|\lambda_{m,N}| < \frac{2^{2m+m}m^m}{m!}\delta \tag{21}$$

for all $N > N_0$. As δ 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 $k = r+1, r+2, \dots, m$, that is, assume that the $m-r$ highest order coefficients all converge as required. Let

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

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

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

Now the bound from (19), but applied using the rth difference, gives

$$\begin{aligned}|\langle\Delta_h^r z_N(n')\rangle| &= |\langle\Delta_h^r\gamma_N(n') + \Delta_h^r z_r(n')\rangle| \\ &= |\langle\epsilon + h^r r!\lambda_{r,N}\rangle| < 2^r\delta,\end{aligned} \tag{22}$$

where

$$\epsilon = \Delta_h^r\gamma_N(n') \le 2^r \sup_{n\in\{1,\dots,N\}} |\gamma_N(n)| \to 0$$

as $N \to \infty$. Choose δ and ϵ such that $2^r\delta < \frac{1}{4}$ and $|\epsilon| < \frac{1}{4}$. Then, from (22) and from Lemma 9,

$$|\langle h^r r!\lambda_{r,N}\rangle| < 2^r\delta + |\epsilon|$$

for all h such that $1 \le h \le \frac{N}{2m}$. Choosing $2^r\delta + |\epsilon| < 2^{-2r-1}$ and using the same iterative process as for the highest order coefficient $\lambda_{m,N}$ (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 $a_1, a_2, \dots, a_r$ be r real numbers for which $|\langle a_n\rangle| < \delta$ for all $n = 1, 2, \dots, r$. Then $|\langle\sum_{n=1}^r a_n\rangle| < r\delta$.*

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

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

■

Lemma 9. *Let $|\langle a+\epsilon\rangle| < \delta$ where $|\epsilon| < 1/4$ and $0 < \delta < 1/4$. 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

$$n - \tfrac{1}{2} < n - \delta - \epsilon < a < n + \delta - \epsilon < n + \tfrac{1}{2}.$$

Hence $\langle a\rangle = a - n$ and so

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

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

■

We are now in a position to complete the proof of strong consistency. Let Ω be the sample space on which the random variables $\{X_i\}$ and $\{\Phi_i\}$ are defined. Let A be the subset of Ω on which $V_N(\widehat{\boldsymbol{\lambda}}_N) - \sigma^2 \to 0$ as $N \to \infty$. Now $\Pr\{A\} = 1$ as a result of Lemma 4. Let A' be the subset of Ω on which $N^k\widehat{\lambda}_{k,N} \to 0$ for $k = 0, \dots, m$ as $N \to \infty$. As a result of Lemma 7, $A \subseteq A'$, and so $\Pr\{A'\} \ge \Pr\{A\} = 1$. Strong consistency follows.

References

[1] F. Hlawatsch and G. F. Boudreaux-Bartels, “Linear and quadratic time-frequency signal representations,” *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, Apr. 1992.

[2] M. D. Ridley, *Speech time-frequency representations*. Boston: Kluwer Academic Publishers, 1989.

[3] N. Suga, J. A. Simmons, and P. H. Jen, “Peripheral specialization for fine analysis of doppler-shifted echoes in the auditory system of the CF-FM bat Pteronotus Parnellii,” *Journal of Experimental Biology*, vol. 63, pp. 161–192, 1975.

[4] J. Ängeby, “Estimating signal parameters using the nonlinear instantaneous least squares approach,” *IEEE Trans. Sig. Process.*, vol. 48, pp. 2721–2732, Oct. 2000.

[5] R. G. McKilliam and I. V. L. Clarkson, “Identifiability and aliasing in polynomial-phase signals,” *IEEE Trans. Sig. Process.*, vol. 57, no. 11, pp. 4554–4557, Nov. 2009.

[6] B. G. Quinn, “Phase-only information loss,” *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, pp. 3982–3985, Mar. 2010.

[7] E. J. Hannan, “The estimation of frequency,” *Journal of Applied Probability*, vol. 10, no. 3, pp. 510–519, Sep 1973.

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

[9] R. G. McKilliam, B. G. Quinn, and I. V. L. Clarkson, “Direction estimation by minimum squared arc length,” *IEEE Trans. Sig. Process.*, vol. 60, no. 5, pp. 2115–2124, May 2012.

[10] R. G. McKilliam, “Lattice theory, circular statistics and polynomial phase signals,” Ph.D. dissertation, University of Queensland, Australia, December 2010.

[11] P. O’Shea, “A fast algorithm for estimating the parameters of a quadratic FM signal,” *IEEE Trans. Sig. Process.*, vol. 52, no. 2, pp. 385 – 393, Feb. 2004.

[12] T. Abatzoglou, “Fast maximum likelihood joint estimation of frequency and frequency rate,” *IEEE Trans. Aerospace Elec. Systems*, vol. 22, no. 6, pp. 708–715, Nov. 1986.

[13] S. Peleg and B. Friedlander, "The discrete polynomial-phase transform," *IEEE Trans. Sig. Process.*, vol. 43, no. 8, pp. 1901–1914, Aug. 1995.

[14] S. Peleg and B. Porat, "Estimation and classification of polynomial-phase signals," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 422–430, Mar. 1991.

[15] B. Porat and B. Friedlander, "Asymptotic statistical analysis of the high-order ambiguity function for parameter estimation of polynomial-phase signals," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 995–1001, May 1996.

[16] M. Farquharson, P. O'Shea, and G. Ledwich, "A computationally efficient technique for estimating the parameters of polynomial-phase signals from noisy observations," *IEEE Trans. Sig. Process.*, vol. 53, no. 8, pp. 3337–3342, Aug. 2005.

[17] B. Boashash and P. O'Shea, "Polynomial wigner-ville distributions and their relationship to time-varying higher order spectra," *IEEE Trans. Sig. Process.*, vol. 42, no. 1, pp. 216–220, Jan. 1994.

[18] S. Barbarossa and V. Petrone, "Analysis of polynomial-phase signals by the integrated generalized ambiguity function," *IEEE Trans. Sig. Process.*, vol. 45, pp. 316–327, Feb. 1997.

[19] S. Barbarossa, A. Scaglione, and G. B. Giannakis, "Product high-order ambiguity function for multicomponent polynomial-phase signal modeling," *Signal Processing, IEEE Transactions on*, vol. 46, no. 3, pp. 691–708, Mar. 1998.

[20] P. O'Shea, "A new technique for instantaneous frequency rate estimation," *IEEE Signal Processing Letters*, vol. 9, no. 8, pp. 251 –252, Aug. 2002.

[21] I. Djurović, M. Simeunovic, S. Djukanović, and P. Wang, "A hybrid CPF-HAF estimation of polynomial-phase signals: Detailed statistical analysis," *IEEE Trans. Sig. Process.*, vol. 60, no. 10, pp. 5010–5023, Oct. 2012.

[22] P. O'Shea and R. Wiltshire, "A new class of multilinear functions for polynomial phase signal analysis," *IEEE Trans. Sig. Process.*, vol. 57, no. 6, pp. 2096–2109, june 2009.

[23] J. Kitchen, "A method for estimating the coefficients of a polynomial phase signal," *Signal Processing*, vol. 37, no. 1, pp. 463–470, Jun. 1994.

[24] P. M. Djuric and S. M. Kay, "Parameter estimation of chirp signals," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 12, pp. 2118–2126, Dec. 1990.

[25] M. R. Morelande, "Parameter estimation of phase-modulated signals using Bayesian unwrapping," *IEEE Trans. Sig. Process.*, vol. 57, no. 11, pp. 4209–4219, Nov. 2009.

[26] R. G. McKilliam, I. V. L. Clarkson, B. G. Quinn, and B. Moran, "Polynomial-phase estimation, phase unwrapping and the nearest lattice point problem," in *Forty-Third Asilomar Conference on Signals, Systems and Computers*. Asilomar, California: IEEE, Nov. 2009, pp. 493–495.

[27] R. G. McKilliam, B. G. Quinn, I. V. L. Clarkson, and B. Moran, "Frequency estimation by phase unwrapping," *IEEE Trans. Sig. Process.*, vol. 58, no. 6, pp. 2953–2963, June 2010.

[28] ——, "The asymptotic properties of polynomial phase estimation by least squares phase unwrapping," in *Proc. Internat. Conf. Acoust. Spe. Sig. Process.* Prauge, Czech Republic: IEEE, May 2011, pp. 3592–3595.

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

[30] 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.

[31] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.

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

[33] S. Baro, J. Hagenauer, and M. Witzke, "Iterative detection of MIMO transmission using a list-sequential (LISS) detector," in *IEEE International Conference on Communications, 2003. ICC '03*, vol. 4, May 2003, pp. 2653–2657.

[34] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491 – 503, March 2006.

[35] S. Mondal, A. M. Eltawil, and K. N. Salama, "Architectural optimizations for low-power k-best MIMO decoders," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3145–3153, Sep. 2009.

[36] P. Erdös and P. Turan, "On some sequences of integers," *J. London Math. Soc.*, vol. 11, pp. 261–264, 1936.

[37] E. Szemerédi, "On sets of integers containing no k elements in arithmetic progression," *Acta Arith.*, vol. 27, pp. 299–345, 1975.

[38] W. T. Gowers, "A new proof of Szemerédi's theorem," *Geom. Func. Anal.*, vol. 11, p. 465588, 2001.

[39] V. Katkovnik, J. Astola, and K. Egiazarian, "Phase local approximation (PhaseLa) technique for phase unwrap from noisy data," *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 833–846, June 2008.

[40] G. Nico, G. Palubinskas, and M. Datcu, "Bayesian approaches to phase unwrapping: theoretical study," *IEEE Trans. Sig. Process.*, vol. 48, no. 9, pp. 2545–2556, Sep. 2000.

[41] B. Friedlander and J. Francos, "Model based phase unwrapping of 2-D signals," *IEEE Trans. Sig. Process.*, vol. 44, no. 12, pp. 2999–3007, Dec. 1996.

[42] S. Peleg and B. Porat, "Linear fm signal parameter estimation from discrete-time observations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 607 – 616, Jul. 1991.

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

[44] S. M. Kay, "A fast and accurate single frequency estimator," *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, vol. 37, pp. 1987–1990, 1989.

[45] X. G. Xia, "Dynamic range of the detectable parameters for polynomial phase signals using multiple-lag diversities in high-order ambiguity functions," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1378–1384, May 2001.

[46] S. Djukanović and I. Djurović, "Aliasing detection and resolving in the estimation of polynomial-phase signal parameters," *Signal Processing*, vol. 92, pp. 235–239, 2012.

[47] G. T. Zhou and Y. Wang, "Exploring lag diversity in the high-order ambiguity function for polynomial phase signals," *IEEE Signal Processing Letters*, vol. 4, no. 8, pp. 240–242, Aug. 1997.

[48] S. Djukanović, V. Popović, M. Daković, and L. Stanković, "A parametric method for non-stationary interference suppression in direct sequence spread-spectrum systems," *Signal Process.*, vol. 91, no. 6, pp. 1425–1431, Jun. 2011.

[49] R. G. McKilliam, B. G. Quinn, I. V. L. Clarkson, B. Moran, and B. N. Vellambi, "Polynomial phase estimation by phase unwrapping," Nov. 2012. [Online]. Available: http://arxiv.org/abs/1011.1262v2

[50] J. W. S. Cassels, *An Introduction to the Geometry of Numbers*. Springer, 1997.

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

[52] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, 3rd ed. New York: Springer, 1998.

[53] ——, "Fast quantizing and decoding and algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 227–232, Mar. 1982.

[54] R. G. McKilliam, W. D. Smith, and I. V. L. Clarkson, "Linear-time nearest point algorithms for Coxeter lattices," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1015–1022, Mar. 2010.

[55] 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. Inform. Theory*, vol. 54, no. 9, pp. 4378–4381,

Sep. 2008. [Online]. Available: http://arxiv.org/abs/0801.1364
[56] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Programming*, vol. 66, pp. 181–191, 1994.
[57] J. Anderson and S. Mohan, "Sequential coding algorithms: A survey and cost analysis," *IEEE Trans. Commun.*, vol. 32, no. 2, pp. 169 – 176, feb 1984.
[58] P. J. Cahen and J. L. Chabert, *Integer-valued Polynomials*. Providence RI: American Mathematical Society, July 1997.
[59] K. V. Mardia and P. Jupp, *Directional Statistics*, 2nd ed. John Wiley & Sons, 2000.
[60] N. I. Fisher, *Statistical analysis of circular data*. Cambridge University Press, 1993.
[61] R. N. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds I," *Annals of Statistics*, vol. 31, no. 1, pp. 1–29, 2003.
[62] ——, "Large sample theory of intrinsic and extrinsic sample means on manifolds II," *Annals of Statistics*, vol. 33, no. 3, pp. 1225–1259, 2005.
[63] T. Hotz and S. Huckemann, "Intrinsic means on the circle: Uniqueness, locus and asymptotics," 2011, http://arxiv.org/abs/1108.2141.
[64] S. Peleg and B. Porat, "The Cramèr-Rao lower bound for signals with constant amplitude and polynomial phase," *IEEE Trans. Sig. Process.*, vol. 39, no. 3, pp. 749–752, Mar. 1991.
[65] J. Ängeby, "Aliasing of polynomial-phase signal parameters," *IEEE Trans. Sig. Process.*, vol. 48, no. 5, pp. 1488–1491, May 2000.
[66] P. O'Shea, "On refining polynomial phase signal parameter estimates," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 3, pp. 978–987, July 2010.
[67] A. Amar, A. Leshem, and A. J. van der Veen, "A low complexity blind estimaotor of narrowband polynomial phase signals," *ieeetsp*, vol. 58, no. 9, pp. 4674–4683, Sep. 2010.
[68] D. Pollard, *Convergence of Stochastic Processes*. New York: Springer-Verlag, 1984.
[69] S. A. van de Geer, *Empirical Processes in M-Estimation*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2009.
[70] C. Jordan, *Calculus of finite differences*. New York, N.Y.: Chelsea Publishing Company, 1965.