Polynomial phase estimation by phase unwrapping 1: Identifiability and strong consistency

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 two part series of papers we consider an estimator that performs phase unwrapping in a least squares manner. In this first part we describe conditions for the identifiability of polynomial phase signals and we prove the strong consistency of our unwrapping estimator.**

***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 [**?**]. In radar and sonar applications polynomial phase signals arise when acquiring radial velocity and acceleration (and higher order motion descriptors) of a target from a reflected signal, and also in continuous wave radar and low probability of intercept radar [**?**]. In biology, polynomial phase signals are used to describe the sounds emitted by bats and dolphins for echo location [**? ?**].

A polynomial phase signal of order m is a function of the form

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

where $j = \sqrt{-1}$, and t is a real number, often 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 practice the signal is typically sampled at discrete points in 'time', t. In this paper we only consider uniform sampling, where the gap between consecutive samples is constant. In this case we can always consider the samples to be taken at some set of consecutive integers and our sampled polynomial phase signal looks like

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

where n is an integer. 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

$$Y_n = \rho s_n + X_n, \tag{1}$$

where ρ is a positive real number representing the (usually unknown) signal amplitude and $\{X_n, n \in \mathbb{Z}\}$ is a sequence of complex noise variables. 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*. It was shown in [**?**] that an identifiable region tessellates a particular $m+1$ dimensional lattice. We discuss this in Section III.

One estimation 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 two part series of papers we consider the estimator that results from unwrapping the phase in a least squares manner. We call this the *least squares unwrapping estimator* (LSU) [**? ?**][**?** , Chap. 8]. It was shown in [**? ?**] that the LSU estimator can be represented as a *nearest lattice point problem*, and Monte-Carlo simulations were used to show the LSU estimator's favourable statistical performance. In this two part series of we derive the asymptotic statistical properties of the LSU estimator showing, under some assumptions on the noise $X_1, \ldots, X_N$, that it is strongly consistent and asymptotically normally distributed. Strong consistency is proved in this paper, while asymptotic normality is proved in the second part [**?**]. Similar results were stated without a complete proof in [**?**]. Here, we give a proof. The results here are also more general than in [**?**], allowing for a wider class of noise distributions.

An interesting property is that the estimator of the kth polynomial phase coefficient converges to $\tilde{\mu}_k$ at rate $o(N^{-k})$. This is perhaps not surprising, since it is the same rate observed in polynomial regression. However, asserting that convergence at this rate occurs in the polynomial phase setting is not trivial. For this purpose we make use of an elementary result about the number of arithmetic progressions contained inside subsets of $\{1, 2, \ldots, N\}$ [**? ? ?**]. We are hopeful that the proof techniques developed here will 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 [**? ?**].

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 also de-

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 Department of Electrical Engineering and Computer Science, Melbourne Systems Lab, Dept of Elec & Electronic Eng, Uni of Melbourne, Vic. 3010, Australia.

rived in [**?**]. These identifiability results are required in order to properly understand the statistical properties of polynomial phase estimators. Section IV describes the LSU estimator and states a theorem asserting the estimator to be strongly consistent under some assumptions on the noise $X_1, \ldots, X_N$. The theorem is proved in Section V. In the second paper [**?**] we prove the asymptotic normality of the LSU estimator and describe the results of Monte Carlo simulations. These simulations agree with our derived asymptotic statistical properties.

II. Lattice Theory

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

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

where $\mathbf{B} \in \mathbb{R}^{n\times d}$ is an $n \times d$ matrix of rank d, called the generator matrix. If $n = d$ the lattice is said to be 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 to construct a rectangular set of representatives, as the next proposition will show. We will use these rectangular regions for describing 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 $\mathbf{B} = \mathbf{Q}\mathbf{R}$ 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 [**?** , Chapter IX, Theorem IV] [**?** , Proposition 2.1]. This result is for lattices with full rank. A result in the general case can be obtained similarly, but is not required here. ■

III. Identifiability and aliasing

As discussed in the introduction, a polynomial phase signal of order m is a complex valued function of the form $s(t) = e^{2\pi j y(t)}$ where t is a real number and $y(t)$ is a polynomial of order m. We will often drop the (t) and just write the polynomial as y and the polynomial phase signal as s whenever there is no chance of ambiguity. Aliasing can occur when polynomial-phase signals are sampled. That is, two or more distinct polynomial-phase signals can take exactly the same values at the sample points. These aliasing results are also given in [**?**], but the presentation here is different, and is better suited to studying the asymptotic properties of the LSU estimator.

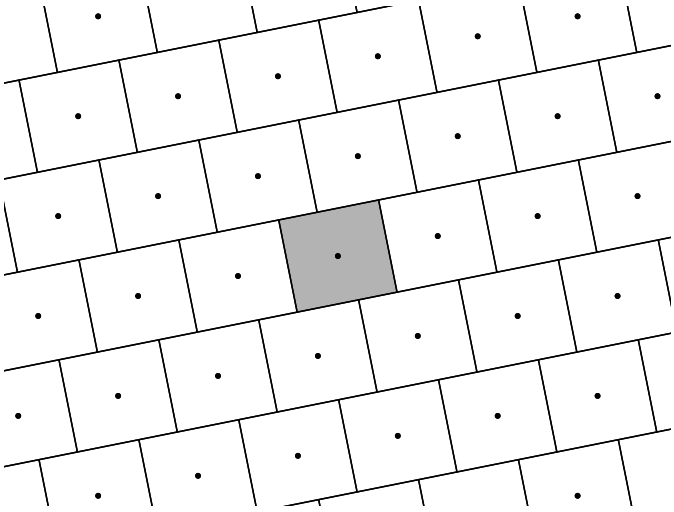

Fig. 1. Rectangular tessellation constructed according to Proposition 1 where Λ is a 2 dimensional lattice with generator matrix having columns $[1, 0.2]'$ and $[0.2, 1]'$. Any one of the boxes is a rectangular set of coset representatives for $\mathbb{R}^2/\Lambda$. The shaded box centered at the origin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

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 theorem.

Theorem 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, $y - z \in \mathcal{Z}$.*

It may be helpful to observe Figures 2 to 5. In these, the phase (divided by 2π) of two distinct polynomial phase signals is plotted on the left, and on the right the principal component of the phase (also divided by 2π) is plotted. 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 1. *The integer valued polynomials $p_0, \dots, p_m$ are an integer basis for $\mathcal{Z}$. That is, every polynomial in $\mathcal{Z}$ can be uniquely written as*

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

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

Proof: See [**?** , p. 2] or [**?**]. ∎

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

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

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

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

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

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

Let

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

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 Theorem 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 $\mathrm{coef}(y)$ and $\mathrm{coef}(z)$ 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 $\boldsymbol{\mu}$, 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*.

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

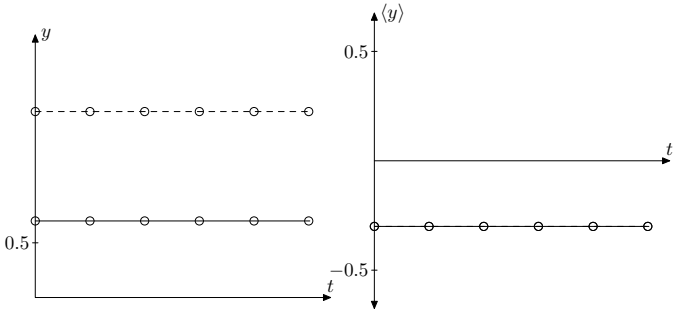

Fig. 2. The zeroth order polynomials $\frac{7}{10}$ (solid line) and $\frac{17}{10}$ (dashed line).

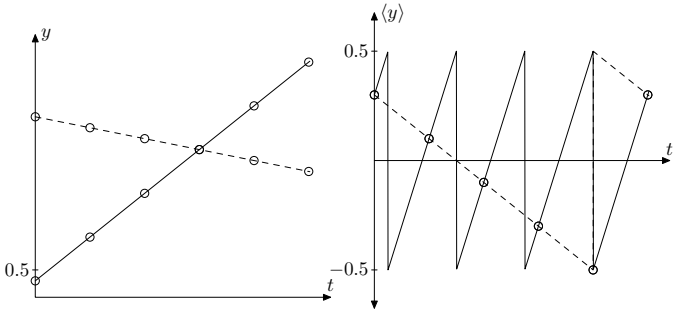

Fig. 3. The first order polynomials $\frac{1}{10}(3 + 8t)$ (solid) and $\frac{1}{10}(33 - 2t)$ (dashed line).

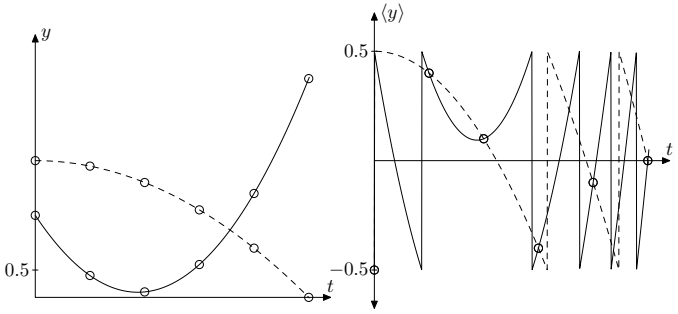

Fig. 4. The quadratic polynomials $\frac{1}{10}(15 - 15t + 4t^2)$ (solid line) and $\frac{1}{10}(25 - t^2)$ (dashed line).

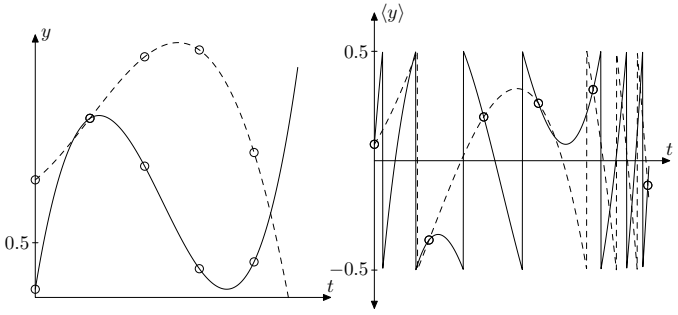

Fig. 5. The cubic polynomials $\frac{1}{160}(174 + 85t - 118t^2 + 40t^3)$ (solid line) and $\frac{1}{48}(84 + 19t + 12t^2 - 4t^3)$ (dashed line).

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 $[-1/2, 1/2)$ corresponds to the Voronoi cell of L_1. When $m = 1$ it turns out that a natural choice of identifiable region is the square box $[-1/2, 1/2)^2$. This corresponds with the *Nyquist criterion*. The lattice L_2 is equal to $\mathbb{Z}^2$ so the box $[-1/2, 1/2)^2$ corresponds with 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} used in [**?**]. 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 $\mathbf{P}$ 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 vectors $\mathbf{x}$ 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 $\mathbf{x}$ with half integers rounded upwards and both $\langle \cdot \rangle$ and $\lceil \cdot \rfloor$ operate on vectors elementwise. For $m \geq 2$ the function $\operatorname{dealias}(\mathbf{x})$ can be computed by a simple sequential algorithm [**?** , 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 the coefficients $\tilde{\mu}_0, \ldots, \tilde{\mu}_m$ from the noisy samples $Y_1, \ldots, Y_N$ given in (1). We take the complex argument of the Y_n and divide by 2π to obtain

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

where $\angle z$ denotes the complex argument 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 [**? ?**]. 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 $X_1, \ldots, X_N$ are circularly symmetric and identically distributed, then $\Phi_1, \ldots, \Phi_n$ are also identically distributed.

Let $\boldsymbol{\mu}$ be the vector $[\mu_0, \mu_1, \ldots, \mu_m]$ 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 least squares unwrapping estimator is defined as those coefficients $\widehat{\mu}_0, \ldots, \widehat{\mu}_m$ that minimise SS over the identifiable region B. That is, the LSU estimator is,

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

It is shown in [**?** , Sec 8.1][**?**] how this minimisation problem can be posed as that of computing a nearest lattice point in a particular lattice. Polynomial time algorithms that compute the nearest point are described in [**?** , 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. In this paper we focus on the asymptotic statistical properties of the LSU estimator, rather than computational aspects.

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 $[-1/2, 1/2)$ [**? ? ? ?**]. 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 [**? ? ?**],

$$\mu_{\text{intr}} = \arg \min_{\mu \in [-1/2, 1/2)} \mathbb{E} \langle \Phi_n - \mu \rangle^2, \tag{7}$$

and the *intrinsic variance* is

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

where $\mathbb{E}$ denotes the expected value. Depending on the distribution of Φ_n the argument that minimises (7) may not be unique. The set of minima is often called the *Fréchet mean set* [**? ?**]. If the minimiser is not unique we say that Φ_n has no intrinsic mean. We are now equipped to state the main result of this paper.

Theorem 2. *(Strong consistency) 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 $\widehat{\lambda}_{0,N}, \ldots, \widehat{\lambda}_{m,N}$. Suppose $\Phi_1, \ldots, \Phi_N$ are independent and identically distributed with zero intrinsic mean and intrinsic variance σ^2, then $N^k \widehat{\lambda}_{k,N}$ converges almost surely to* 0 *as $N \to \infty$ for all $k = 0, 1, \ldots, m$.*

A proof of this theorem is given in the next section. Proofs for the case when $m = 0$ was given in [**?**] and for the case when $m = 1$ was given in [**?**]. The proof here takes a similar approach, but requires new techniques. The theorem gives conditions on 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 phase ± 0.5 are actually

the same. In this case

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

The same reasoning holds for $m > 0$.

The requirement that $\Phi_1, \ldots, \Phi_N$ 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 density function nonincreasing with magnitude $|X_n|$, then, the corresponding Φ_n necessarily has zero intrinsic mean [**?** , Theorem 5.2, page 78]. Thus, our theorem covers commonly used distributions for $X_1, \ldots, X_N$, such as the normal distribution.

Although we will not prove it here the assumption that $\Phi_1, \ldots, \Phi_N$ have zero intrinsic mean is not only sufficient, but also necessary, for if $\Phi_1, \ldots, \Phi_N$ have intrinsic mean $x \in [-1/2, 1/2)$ with $x \neq 0$ then $\langle \widehat{\lambda}_{0,N} - x \rangle \to 0$ almost surely as $N \to \infty$, and so $\widehat{\lambda}_{0,N}$ does not converge to zero. On the other hand if $\Phi_1, \ldots, \Phi_N$ do not have an intrinsic mean then $\widehat{\lambda}_{0,N}$ will not converge.

V. 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 $\boldsymbol{\lambda} = \text{dealias}(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \tilde{\boldsymbol{\mu}} - \boldsymbol{\mu} - \mathbf{p}$ 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, 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 $\text{dealias}(\cdot)$ function $\boldsymbol{\lambda} \in B$ so the elements of $\boldsymbol{\lambda}$ satisfy

$$-\frac{0.5}{k!} \leq \lambda_k < \frac{0.5}{k!}. \tag{8}$$

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

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

It will follow that

$$\sup_{\boldsymbol{\lambda} \in B} |S_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda})| \to 0 \tag{9}$$

almost surely as $N \to \infty$. This type of result has been called a *uniform law of large numbers* and follows from standard techniques [**?**]. We give a full proof of (9) in Appendix A. We now concentrate attention on the minimiser of V_N. Because Φ_n has zero intrinsic mean

$$\mathbb{E} \langle \Phi_n + z \rangle^2 \tag{10}$$

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

$$\mathbb{E} \langle \Phi_1 + z \rangle^2 = \mathbb{E} \langle \Phi_1 \rangle^2 = \sigma^2, \tag{11}$$

and so the minimum attained value is σ^2.

Lemma 2. *For $\boldsymbol{\lambda} \in B$ the function $V_N(\boldsymbol{\lambda})$ is minimised uniquely at $\mathbf{0}$, the vector of all zeros. At this minimum $V_N(\mathbf{0}) = \sigma^2$.*

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

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

We know that $\mathbb{E} \langle \Phi_n + \langle z(n) \rangle \rangle^2$ is minimised uniquely when $\langle z(n) \rangle = 0$ at which point it takes the value σ^2. Now $\langle z(n) \rangle$ is equal to zero for all integers n if and only if $z \in \mathcal{Z}$, or equivalently if $\text{coef}(z)$ is a lattice point in L_{m+1}. By definition B contains precisely one lattice point from L_{m+1}, this being the origin $\mathbf{0}$. Therefore V_N is minimised uniquely at $\mathbf{0}$, at which point it takes the value σ^2. ∎

Lemma 3. *$|V_N(\widehat{\boldsymbol{\lambda}}_N) - \sigma^2| \to 0$ almost surely as $N \to \infty$.*

Proof: By definition $\widehat{\boldsymbol{\lambda}}_N = \arg\min_{\boldsymbol{\lambda} \in B} S_N(\boldsymbol{\lambda})$ so

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

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 (9). ∎

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. These results are enough to show that $\widehat{\boldsymbol{\lambda}}_N$ converges almost surely to zero. However, this tells us nothing about the rate at which the components of $\widehat{\boldsymbol{\lambda}}_N$ approach zero as required by Theorem 2. 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

$$A(h, K) = \left\{ n \,;\, n + ih \in K \;\forall\, i \in \{0, 1, \ldots, m\} \right\} \tag{12}$$

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. If K is a small subset of W then $A(h,K)$ 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 [**?** **?** **?**]. In particular, Gower's [**?** , 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 4. *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,*

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

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

Lemma 5. $|A(h,K)| \geq N - mh - (N-|K|)(m+1)$.

Proof: Note that $|A(h,W)| = N-mh$. The proof follows by starting with $A(h,W)$ and applying Lemma 4 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 h such that $1 \leq h \leq \frac{N}{2m}$ 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 $h \leq \frac{N}{2m}$ into the inequality from Lemma 5 we immediately obtain $|A(h,K)| > 0$. ∎

The next result we require comes from the calculus of finite differences. For any function $d(n)$ mapping $\mathbb{R}$ to $\mathbb{R}$, let

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

denote the first difference with interval h, and let

$$\begin{aligned}\Delta_h^r d(n) &= \Delta_h^{r-1} d(n+h) - \Delta_h^{r-1} d(n) \\ &= \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} d(n+kh)\end{aligned} \tag{13}$$

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

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

precisely 2^r times.

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

$$\Delta_h^r d(n) = h^r r! a_r. \tag{14}$$

So, the rth difference of the polynomial is a constant depending on h, r and the rth coefficient a_r [**?** , page 51]. We can now continue the proof of strong consistency. The next lemma is a key result.

Lemma 6. *Suppose $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots$ is a sequence of vectors from B with $V_N(\boldsymbol{\lambda}_N) - \sigma^2 \to 0$ as $N \to \infty$. Then the elements $\lambda_{0,N}, \ldots \lambda_{m,N}$ of $\boldsymbol{\lambda}_N$ satisfy $N^k \lambda_{k,N} \to 0$ as $N \to \infty$.*

Proof: Define the function

$$g(z) = \mathbb{E}\left\langle \Phi_1 + z \right\rangle^2 - \sigma^2 \tag{15}$$

which is continuous in z. Because of (10) and (11), $g(z) \geq 0$ with equality only at $z=0$ for $z \in [-1/2, 1/2)$. 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

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

so that

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

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 $K_N = \{n \leq N\ ;\ |\left\langle z_N(n) \right\rangle| < \delta\}$. 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

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

violating that $V_N(\boldsymbol{\lambda}_N) - \sigma^2$ converges to zero as $N \to \infty$. We will assume $N > N_0$ 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

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

are all less than δ. Because the mth difference is a linear combination of 2^m elements (see (13)) from

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

all with magnitude less than δ we obtain, from Lemma 7,

$$|\left\langle \Delta_h^m z_N(n') \right\rangle| \leq |\Delta_h^m \left\langle z_N(n') \right\rangle| < 2^m \delta. \tag{16}$$

From (14) it follows that the left hand side is equal to a constant involving h, m and $\lambda_{m,N}$ giving the bound

$$|\langle h^m m!\lambda_{m,N}\rangle| = |\langle \Delta_h^m z_N(n')\rangle| < 2^m\delta \tag{17}$$

for all h satisfying $1 \le h \le \frac{N}{2m}$. Setting $h = 1$ and recalling from (8) 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

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

So, 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

$$N^m|\lambda_{m,N}| < \frac{2^{2m+m}m^m}{m!}\delta \tag{18}$$

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 $k = r+1, r+2, \ldots, m$, that is, assume that the $m-r$ highest order coefficients all converge as required. Let

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

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

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

Now the bound from (16), but applied using the rth difference, gives

$$\begin{aligned} |\langle \Delta_h^r z_N(n')\rangle| &= |\langle \Delta_h^r\gamma_N(n') + \Delta_h^r z_r(n')\rangle| \\ &= |\langle \epsilon + h^r r!\lambda_{r,N}\rangle| < 2^r\delta, \end{aligned} \tag{19}$$

where

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

as $N \to \infty$. Choose δ and ϵ such that $2^r\delta < \frac{1}{4}$ and $|\epsilon| < \frac{1}{4}$. Then, from (19) and from Lemma 8,

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

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 (17) to (18)) we find that $N^r\lambda_{r,N} \to 0$ as $N \to \infty$. The proof now follows by induction. ∎

Lemma 7. *Let $a_1, a_2, \ldots, a_r$ be r real numbers such that $|\langle a_n\rangle| < \delta$ for all $n = 1, 2, \ldots, r$. Then $|\langle\sum_{n=1}^r a_n\rangle| < r\delta$.*

Proof: If $\delta > \frac{1}{2r}$ 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 8. *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. As is customary, let $(\Omega, \mathcal{F}, \Pr)$ be the probability space over which the random variables $\{X_i\}$ and $\{\Phi_i\}$ are defined. Let A be the subset of the sample space 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 3. Let A' be the subset of Ω on which $N^k\widehat{\lambda}_{k,N} \to 0$ for $k = 0, \ldots, m$ as $N \to \infty$. As a result of Lemma 6, $A \subseteq A'$, and so $\Pr\{A'\} \ge \Pr\{A\} = 1$. Strong consistency follows.

VI. Conclusion

This paper has considered the estimation of the coefficients of a noisy polynomial phase signal by least squares phase unwrapping (LSU). We have derived conditions under which the polynomial phase estimation problem is identifiable. Under these conditions, and under some assumptions on the distribution of the noise, the LSU estimator is shown to be strongly consistent. In the second paper in this series [**?**] we show the LSU estimator to be asymptotically normally distributed and present the results of Monte Carlo simulations that support our asymptotic results.

Appendix

A. A uniform law of large numbers

During the proof of strong consistency we made use of the fact that

$$\sup_{\boldsymbol{\lambda}\in B} |S_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda})| \to 0 \tag{20}$$

almost surely as $N \to \infty$, where $V_N(\boldsymbol{\lambda}) = \mathbb{E}S_N(\boldsymbol{\lambda})$. We prove this result here. Put

$$D_N(\boldsymbol{\lambda}) = S_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda}).$$

Now, for any $\epsilon > 0$,

$$\sum_{N=1}^{\infty} \Pr\left\{\sup_{\boldsymbol{\lambda}\in B} |D_N(\boldsymbol{\lambda})| > \epsilon\right\} < \infty$$

by Lemma 9. So (20) follows from the Borel-Cantelli lemma. In what follows we use order notation in the standard way, that is, for functions h and g, we write $h(N) = O(g(N))$ to mean that there exists a constant $K > 0$ and a finite N_0 such that $h(N) \le Kg(N)$ for all $N > N_0$.

Lemma 9. *For any $\epsilon > 0$ and $c < 2$,*

$$\Pr\left\{\sup_{\boldsymbol{\lambda}\in B} |D_N(\boldsymbol{\lambda})| > \epsilon\right\} = O(e^{-c\epsilon^2 N}).$$

Proof: Consider a rectangular grid of points spaced over the identifiable region B. We use $\boldsymbol{\lambda}[\mathbf{r}]$, where $\mathbf{r} \in \mathbb{Z}^{m+1}$, to denote the grid point

$$\boldsymbol{\lambda}[\mathbf{r}] = \left[\frac{r_0}{N^b} - \frac{1}{2},\ \frac{r_1}{N^{b+1}} - \frac{1}{2},\ \ldots,\ \frac{r_m}{m!N^{b+m}} - \frac{1}{2(m!)}\right]$$

for some constant $b > 0$. Adjacent grid points are separated by $\frac{1}{N^b}$ in the zeroth coordinate, $\frac{1}{N^{b+1}}$ in the first coordinate and $\frac{1}{k!N^{b+k}}$ in the kth coordinate. Let

$$B[\mathbf{r}] = \left\{\mathbf{x} \in \mathbb{R}^{m+1}; \frac{r_k}{N^{b+k}} \le x_k + \frac{1}{2(k!)} < \frac{r_k + 1}{N^{b+k}}\right\}.$$

and let G be the finite set of grid points

$$G = \left\{\mathbf{x} \in \mathbb{Z}^{m+1}\ ;\ x_k = 0, 1, 2 \ldots, N^{b+k} - 1\right\}.$$

The total number of grid points is $|G| = N^{(m+1)(2b+m)/2}$, and the $B[\mathbf{r}]$ partition B, that is, $B = \cup_{\mathbf{r}\in G} B[\mathbf{r}]$. Now

$$\begin{aligned}
&\sup_{\boldsymbol{\lambda}\in B} |D_N(\boldsymbol{\lambda})| \\
&= \sup_{\mathbf{r}\in G} \sup_{\boldsymbol{\lambda}\in B[\mathbf{r}]} |D_N(\boldsymbol{\lambda}[\mathbf{r}]) + D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])| \\
&\le \sup_{\mathbf{r}\in G} |D_N(\boldsymbol{\lambda}[\mathbf{r}])| + \sup_{\mathbf{r}\in G} \sup_{\boldsymbol{\lambda}\in B[\mathbf{r}]} |D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])|. \quad (21)
\end{aligned}$$

From Lemma 10 it will follow that

$$\Pr\left\{\sup_{\mathbf{r}\in G} |D_N(\boldsymbol{\lambda}[\mathbf{r}])| > \frac{\epsilon}{2}\right\} = O(e^{-c\epsilon^2 N})$$

for any $\epsilon > 0$ and $c < 2$. In Lemma 12 we show that

$$\sup_{\mathbf{r}\in G} \sup_{\boldsymbol{\lambda}\in B[\mathbf{r}]} |D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])| < 2\frac{m+1}{N^b}.$$

Combining these results with (21), we obtain

$$\Pr\left(\sup_{\boldsymbol{\lambda}\in B} |D_N(\boldsymbol{\lambda})| > \frac{\epsilon}{2} + \frac{2(m+1)}{N^b}\right) = O(e^{-c\epsilon^2 N}),$$

and for sufficiently large N, we have $\epsilon/2 + \frac{2(m+1)}{N^b} < \epsilon$ completing the proof. It remains to prove Lemmas 10 and 12. ∎

Lemma 10. *For any $\epsilon > 0$ and $c < 8$,*

$$\Pr\left\{\sup_{\mathbf{r}\in G} |D_N(\boldsymbol{\lambda}[\mathbf{r}])| > \epsilon\right\} = O(e^{-c\epsilon^2 N}).$$

Proof: Fix $\boldsymbol{\lambda}$ and write

$$D_N(\boldsymbol{\lambda}) = \bar{Z} = \frac{1}{N}\sum_{n=1}^{N} Z_n,$$

where

$$Z_n = \left\langle \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right\rangle^2 - \mathbb{E}\left\langle \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right\rangle^2$$

are independent with zero mean and $|Z_n| \le \frac{1}{4}$. It follows from Hoeffding's inequality [**?**] that,

$$\Pr\{|D_N(\boldsymbol{\lambda})| > \epsilon\} \le 2e^{-8\epsilon^2 N},$$

and so,

$$\begin{aligned}
\Pr\left\{\sup_{\mathbf{r}\in G} |D_N(\boldsymbol{\lambda}[\mathbf{r}])| > \epsilon\right\} &\le \sum_{\mathbf{r}\in G} \Pr\left\{|D_N(\boldsymbol{\lambda}[\mathbf{r}])| > \epsilon\right\} \\
&= 2|G|e^{-8\epsilon^2 N} = O(e^{-c\epsilon^2 N}),
\end{aligned}$$

where c is any real number less than 8, since $|G| = N^{(m+1)(2b+m)/2}$ is polynomial in N. ∎

Before proving Lemma 12 we need the following result.

Lemma 11. *For real numbers x and delta,*

$$\langle x\rangle^2 - |\delta| \le \langle x + \delta\rangle^2 \le \langle x\rangle^2 + |\delta|.$$

Proof: Since $|\delta| \le |n + \delta|$ for all $\delta \in [-1/2, 1/2)$ and $n \in \mathbb{Z}$, the result will follow if we can show that it holds when both x and δ are in $[-1/2, 1/2)$. Also, for reasons of symmetry, we need only show that it holds when $\delta \ge 0$. Now

$$\langle x + \delta\rangle^2 - x^2 = \begin{cases} 2x\delta + \delta^2, & x \in [-1/2, 1/2 - \delta) \\ 2x(\delta - 1) + (\delta - 1)^2, & x \in [1/2 - \delta, 1/2) \end{cases}$$

But, when $x \in [-1/2, 1/2 - \delta)$,

$$-1 \le -1 + \delta \le 2x + \delta < 1 - \delta \le 1,$$

and so

$$-\delta \le (-1 + \delta)\delta \le (2x + \delta)\delta < (1 - \delta)\delta \le \delta.$$

Also, when $x \in [1/2 - \delta, 1/2)$ we have $-\delta \le 2x + \delta - 1 < \delta$, and consequently

$$-\delta \le -\delta(1 - \delta) \le (2x + \delta - 1)(1 - \delta) \le \delta(1 - \delta) \le \delta.$$

∎

Lemma 12. *For all positive integers N,*

$$\sup_{\mathbf{r}\in G} \sup_{\boldsymbol{\lambda}\in B[\mathbf{r}]} |D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])| < 2\frac{m+1}{N^b}.$$

Proof: Put

$$b_n = \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \quad \text{and} \quad a_n = \Phi_n + \sum_{k=0}^{m} \lambda[\mathbf{r}]_k n^k,$$

where $\lambda[\mathbf{r}]_k$ denotes the kth element of the grid point $\boldsymbol{\lambda}[\mathbf{r}]$. For $\boldsymbol{\lambda} \in B[\mathbf{r}]$ we have $b_n = a_n + \delta_n$, where

$$|\delta_n| \le \sum_{k=0}^{m} \frac{n^k}{k!N^{b+k}} \le \frac{m+1}{N^b}.$$

From Lemma 11 it follows that

$$-|\delta_n| \leq \langle x + b_n \rangle^2 - \langle x + a_n \rangle^2 \leq |\delta_n|,$$

and consequently

$$|\, \langle x + b_n \rangle^2 - \langle x + a_n \rangle^2 \,| \leq \tfrac{m+1}{N^b}$$

for all $x \in \mathbb{R}$. Now

$$S_N(\boldsymbol{\lambda}) - S_N(\boldsymbol{\lambda}[\mathbf{r}]) = \frac{1}{N} \sum_{n=1}^{N} \left(\langle \Phi_n + b_n \rangle^2 - \langle \Phi_n + a_n \rangle^2 \right)$$

and therefore

$$|S_N(\boldsymbol{\lambda}) - S_N(\boldsymbol{\lambda}[\mathbf{r}])| \leq \frac{m+1}{N^b}$$

for all $\boldsymbol{\lambda} \in B[\mathbf{r}]$. As this bound is independent of $\Phi_1 \ldots \Phi_N$, we have

$$|V_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda}[\mathbf{r}])| \leq \mathbb{E}|S_N(\boldsymbol{\lambda}) - S_N(\boldsymbol{\lambda}[\mathbf{r}])| \leq \frac{m+1}{N^b}$$

by Jensen's inequality. Therefore,

$$\begin{aligned}
&|D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])| \\
&= |S_N(\boldsymbol{\lambda}) - S_N(\boldsymbol{\lambda}[\mathbf{r}]) + V_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda}[\mathbf{r}])| \\
&\leq |S_N(\boldsymbol{\lambda}) - S_N(\boldsymbol{\lambda}[\mathbf{r}])| + |V_N(\boldsymbol{\lambda}) - V_N(\boldsymbol{\lambda}[\mathbf{r}])| \\
&\leq 2\frac{m+1}{N^b},
\end{aligned}$$

for all $\boldsymbol{\lambda} \in B[\mathbf{r}]$. The lemma follows because this bound is independent of $\mathbf{r}$. ■