Polynomial phase estimation by phase unwrapping 2: Asymptotic normality and simulations

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 second part we prove that the estimator is asymptotically normally distributed. The results of Monte Carlo simulations are presented and these support our asymptotic theory.

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

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.

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 uniformly sampled polynomial-phase signal of order m is a signal of the form

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

where n is an integer and

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

is a polynomial of order m. 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

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

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.

An obvious estimator of the unknown coefficients is the least squares estimator (LSE). This is also the maximum likelihood estimator (MLE) when the noise sequence $\{X_n\}$ is white and Gaussian. When $m = 0$ (phase estimation) or $m = 1$ (frequency estimation) the LSE is an effective approach, being both computationally efficient and statistically accurate [**? ? ? ?**][**?** , Sec. 6.4 and 9.1]. When $m \geq 2$ the computational complexity of the LSE is large [**?** , Sec. 10.1][**?**]. For this reason many authors have considered alternative approaches to polynomial phase estimation. These can loosely be grouped into two classes, estimators based on polynomial phase transforms, such as the discrete polynomial phase transform (DPT) [**? ? ?**] and the high order phase function [**?**], and estimators based on phase unwrapping, such as Kitchen's unwrapping estimator [**?**], the estimator of Djuric and Kay [**?**], and Morelande's Bayesian unwrapping estimator [**?**].

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 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 papers we derive the asymptotic statistical properties of the LSU estimator. The first part in the series [**?**] asserts the strong consistency of the LSU estimator under some conditions on the distribution on the noise $\{X_n\}$. In this second part we prove, under similar conditions on $\{X_n\}$, that the LSU estimator is asymptotically normally distributed.

The proof of asymptotic normality is complicated by the fact that the objective function corresponding with the LSU estimator is not differentiable everywhere. Empirical process techniques [**? ? ? ?**] and results from the literature on hyperplane arrangements [**? ?**] become useful here. 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 the LSU estimator and states a theorem asserting the estimator to be asymptotically normally distributed. The theorem is proved in Section III. Section IV describes the results of Monte Carlo simulations that compare the performance of the LSU estimator with some existing estimators. These simulations agree with the derived asymptotic properties. This paper is self contained and can be read independently of the first part [**?**]. Those results required from [**?**] are referenced here. However, it is intended (and recommended) that the papers be read in order.

II. The least squares unwrapping estimator

As in [**?**] the least squares unwrapping (LSU) estimator of the polynomial coefficients is the minimiser

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}\in B} SS(\boldsymbol{\mu}) \tag{2}$$

where $B = \prod_{k=0}^{m}\left[-\frac{0.5}{k!}, \frac{0.5}{k!}\right)$ is a subset of $\mathbb{R}^{m+1}$ defined in [**?** , Sec. 3] called the *identifiable region*, and where the objective function

$$SS(\boldsymbol{\mu}) = \sum_{n=1}^{N}\left\langle \Theta_n - \sum_{k=0}^{m}\mu_k n^k \right\rangle^2, \tag{3}$$

where

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

is the phase (divided by 2π) of the observations $Y_1, \dots, Y_N$, and

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

are random variables representing the *phase noise* observed at the receiver [**?**]. We denote by $\angle z$ the complex argument of the complex number z, and by $\langle x \rangle = x - \lceil x \rfloor$ the (centered) fractional part of the real number x, and by $\lceil x \rfloor$ the nearest integer to x. We have chosen to round half integers up.

As in [? , Sec. 4] the phase noise $\Phi_1, \dots, \Phi_N$ are *circular* random variables with support on $[-1/2, 1/2)$ [? ? ? ?]. The *intrinsic mean* or *Fréchet mean* of Φ_n is defined as [? ? ?],

$$\mu_{\text{intr}} = \arg \min_{\mu \in [-1/2, 1/2)} \mathbb{E} \langle \Phi_n - \mu \rangle^2 , \tag{5}$$

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. Observe the following property of circular random variables with zero intrinsic mean.

Proposition 1. *Let Φ be a circular random variable with intrinsic mean $\mu_{intr} = 0$ and intrinsic variance σ^2. Then Φ has zero mean and variance σ^2, that is, $\mathbb{E}\Phi = 0$ and $\mathbb{E}(\Phi - \mathbb{E}\Phi)^2 = \sigma^2$.*

Proof: Assume the proposition is false and that $\mu = \mathbb{E}\Phi \neq 0$. But, then

$$\begin{aligned} \sigma^2 &= \mathbb{E} \langle \Phi - \mu_{\text{intr}} \rangle^2 = \mathbb{E} \langle \Phi \rangle^2 \\ &= \mathbb{E}\Phi^2 > \mathbb{E}(\Phi - \mu)^2 \geq \mathbb{E} \langle \Phi - \mu \rangle^2 , \end{aligned}$$

violating the fact that $\mu_{\text{intr}} = 0$ is the minimiser of (5). ■

The next theorem describes the asymptotic distribution of the LSU estimator. The theorem statement makes use of the function $\operatorname{dealias}(\cdot)$ that is defined in [? , Sec. 3].

Theorem 1. *(Asymptotic normality) Let $\widehat{\boldsymbol{\mu}}$ be defined by (2) and put $\widehat{\boldsymbol{\lambda}}_N = \operatorname{dealias}(\tilde{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}})$. Denote the elements of $\widehat{\boldsymbol{\lambda}}_N$ by $\widehat{\lambda}_{0,N}, \dots, \widehat{\lambda}_{m,N}$. Suppose $\Phi_1, \dots, \Phi_N$ are independent and identically distributed with zero intrinsic mean, intrinsic variance σ^2, and pdf f such that $f(\langle x \rangle)$ is continuous at $x = -1/2$ and $f(-1/2) < 1$. Then the distribution of*

$$\begin{bmatrix} \sqrt{N}\widehat{\lambda}_{0,N} & N\sqrt{N}\widehat{\lambda}_{1,N} & \dots & N^m\sqrt{N}\widehat{\lambda}_{m,N} \end{bmatrix}'$$

converges to the normal with zero mean and covariance

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

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

A proof of this theorem is given in the next section. As in [?] 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 $\tilde{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}$. This makes sense in view of the identifiability conditions described in [**?** , Sec. 3]

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 $f(-1/2) \leq 1$ [**?** , 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 these exceptional cases other expressions for the asymptotic variance can be found (similar to [**?** , Theorem 3.1]), but this comes at a substantial increase in complexity and for this reason we have omitted them.

III. Proof of asymptotic normality

Let $\boldsymbol{\psi}$ be the vector with kth component $\psi_k = N^k \lambda_k$, $k = 0, \dots, m$ and let

$$T_N(\boldsymbol{\psi}) = S_N(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{n=1}^{N} \left\langle \Phi_n + \sum_{k=0}^{m} (\tfrac{n}{N})^k \psi_k \right\rangle^2 .$$

Let $\widehat{\boldsymbol{\psi}}_N$ be the vector with elements $\widehat{\psi}_{k,N} = N^k \widehat{\lambda}_{k,N}$ so that $\widehat{\boldsymbol{\psi}}_N$ is the minimiser of T_N. Because each of $N^k \widehat{\lambda}_{k,N}$ converges almost surely to zero as $N \to \infty$ [**?** , Theorem 2], it follows that $\widehat{\boldsymbol{\psi}}_N$ converges almost surely to $\mathbf{0}$ as $N \to \infty$. We want to find the asymptotic distribution of

$$\sqrt{N} \widehat{\boldsymbol{\psi}}_N = \begin{bmatrix} \sqrt{N} \widehat{\psi}_{0,N} \\ \sqrt{N} \widehat{\psi}_{1,N} \\ \vdots \\ \sqrt{N} \widehat{\psi}_{m,N} \end{bmatrix} = \begin{bmatrix} \sqrt{N} \widehat{\lambda}_{0,N} \\ N\sqrt{N} \widehat{\lambda}_{1,N} \\ \vdots \\ N^m \sqrt{N} \widehat{\lambda}_{m,N} \end{bmatrix} .$$

The proof is complicated by the fact that T_N is not differentiable everywhere as $\langle x \rangle^2$ is not differentiable when $\langle x \rangle = \frac{1}{2}$. This precludes the use of "standard approaches" to proving asymptotic normality that are based on the mean value theorem [**? ? ? ? ?**]. However, we show in Lemma 1 that all the partial derivatives $\frac{\partial T_N}{\partial \psi_\ell}$ for $\ell = 0, \dots, m$ exist, and are equal to zero, at the minimiser $\widehat{\boldsymbol{\psi}}_N$. Thus, putting

$$W_n = \left\lfloor \Phi_n + \sum_{k=0}^{m} (\tfrac{n}{N})^k \widehat{\psi}_{k,N} \right\rceil, \tag{6}$$

we have, for each $\ell = 0, \dots, m$,

$$\begin{aligned} 0 &= \frac{\partial T_N}{\partial \psi_\ell}(\widehat{\boldsymbol{\psi}}_N) \\ &= \frac{2}{N} \sum_{n=1}^{N} (\tfrac{n}{N})^\ell \left(\Phi_n - W_n + \sum_{k=0}^{m} (\tfrac{n}{N})^k \widehat{\psi}_{k,N} \right), \end{aligned}$$

so that

$$D_{\ell,N} = K_{\ell,N},$$

where

$$D_{\ell,N} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tfrac{n}{N})^{\ell} \Phi_n,$$

and

$$K_{\ell,N} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tfrac{n}{N})^{\ell} \left(W_n - \sum_{k=0}^{m} (\tfrac{n}{N})^{k} \widehat{\psi}_{k,N} \right). \tag{7}$$

Lemma 3 shows that, for all $\ell = 0, \dots m$,

$$K_{\ell,N} = (h-1)\sqrt{N} \sum_{k=0}^{m} \widehat{\psi}_{k,N} \big(C_{\ell k} + o_P(1) \big) + o_P(1) \tag{8}$$

where $C_{\ell k} = \frac{1}{\ell+k+1}$, and $h = f(-1/2)$, and $o_P(1)$ denotes a random variable converging in probability to zero as $N \to \infty$.

It is now convenient to write in vector form. Let

$$\mathbf{k}_N = \begin{bmatrix} K_{0,N} \\ \vdots \\ K_{m,N} \end{bmatrix} \qquad \text{and} \qquad \mathbf{d}_N = \begin{bmatrix} D_{0,N} \\ \vdots \\ D_{m,N} \end{bmatrix}. \tag{9}$$

From (8),

$$\mathbf{d}_N = \mathbf{k}_N = \sqrt{N}(h-1)(\mathbf{C} + o_P(1))\widehat{\boldsymbol{\psi}}_N + o_P(1)$$

where $o_P(1)$ here means a vector or matrix of the appropriate dimension with every element converging in probability to zero as $N \to \infty$. Thus $\sqrt{N}\widehat{\boldsymbol{\psi}}_N$ has the same asymptotic distribution as $(h-1)^{-1}\mathbf{C}^{-1}\mathbf{d}_N$. Lemma 5 shows that $\mathbf{d}_N$ is asymptotically normally distributed with zero mean and covariance matrix $\sigma^2\mathbf{C}$. Hence $\sqrt{N}\widehat{\boldsymbol{\psi}}_N$ is asymptotically normal with zero mean and covariance matrix

$$\frac{\sigma^2 \mathbf{C}^{-1}\mathbf{C}(\mathbf{C}^{-1})'}{(1-h)^2} = \frac{\sigma^2\mathbf{C}^{-1}}{(1-h)^2}.$$

It remains to prove Lemmas 1, 3 and 5.

Lemma 1. *For all $\ell = 0, \dots, m$ the partial derivatives $\frac{\partial T_N}{\partial \psi_\ell}$ exist, and are equal to zero, at the minimiser $\widehat{\boldsymbol{\psi}}_N$. That is,*

$$\frac{\partial T_N}{\partial \psi_\ell}(\widehat{\boldsymbol{\psi}}_N) = 0 \qquad \textit{for each } \ell = 0, \dots, m.$$

Proof: The function $\langle x \rangle^2$ is differentiable everywhere except if $\langle x \rangle \neq -1/2$. Recalling that

$$T_N(\boldsymbol{\psi}) = \frac{1}{N} \sum_{n=1}^{N} \left\langle \Phi_n + \sum_{k=0}^{m} (\tfrac{n}{N})^k \psi_k \right\rangle^2$$

we see that T_N will be differentiable with respect to $\boldsymbol{\psi}$ at $\widehat{\boldsymbol{\psi}}_N$ if

$$\left\langle \Phi_n + \sum_{k=0}^{m} (\tfrac{n}{N})^k \widehat{\psi}_{k,N} \right\rangle \neq -1/2 \qquad \text{for all } n = 1, \dots, N.$$

This is proved in Lemma 2. So the partial derivatives $\frac{\partial T_N}{\partial \psi_\ell}$ exist for all $\ell = 0, \dots, m$ at $\widehat{\boldsymbol{\psi}}_N$. That each of the partial derivatives is equal to zero at $\widehat{\boldsymbol{\psi}}_N$ follows immediately from the fact that $\widehat{\boldsymbol{\psi}}_N$ is a minimiser of T_N. ■

Lemma 2. $|\langle \Phi_n + \sum_{k=0}^{m} (n/N)^k \widehat{\psi}_{k,N} \rangle| \leq \frac{1}{2} - \frac{1}{2N}$ *for all* $n = 1, \dots, N$.

Proof: To simplify our notation let

$$B_n = \Phi_n + \sum_{k=1}^{m} (n/N)^k \widehat{\psi}_{k,N}$$

so that we now require to prove $|\langle B_n + \widehat{\psi}_{0,N} \rangle| \leq \frac{1}{2} - \frac{1}{2N}$ for all $n = 1, \dots, N$. From (6), $W_n = \lceil B_n + \widehat{\psi}_{0,N} \rfloor$, and

$$\begin{aligned} T_N(\widehat{\boldsymbol{\psi}}_N) &= \frac{1}{N} \sum_{n=1}^{N} \langle B_n + \widehat{\psi}_{0,N} \rangle^2 \\ &= \frac{1}{N} \sum_{n=1}^{N} (B_n + \widehat{\psi}_{0,N} - W_n)^2. \end{aligned}$$

Since $\widehat{\psi}_{0,N}$ is the minimiser of the quadratic above,

$$\widehat{\psi}_{0,N} = -\frac{1}{N} \sum_{n=1}^{N} (B_n - W_n). \tag{10}$$

The proof now proceeds by contradiction. Assume that for some k,

$$\langle B_k + \widehat{\psi}_{0,N} \rangle > \frac{1}{2} - \frac{1}{2N}. \tag{11}$$

Let $F_n = W_n$ for all $n \neq k$ and $F_k = W_k + 1$, and let

$$\phi = -\frac{1}{N} \sum_{n=1}^{N} (B_n - F_n) = \widehat{\psi}_{0,N} + \frac{1}{N}.$$

Then,

$$\begin{aligned} (B_k + \phi - F_k)^2 &= (B_k + \phi - W_k - 1)^2 \\ &= (B_k + \phi - W_k)^2 - 2(B_k + \phi - W_k) + 1 \\ &= (B_k + \phi - W_k)^2 - 2(B_k + \widehat{\psi}_{0,N} - W_k) + 1 - \frac{2}{N} \\ &= (B_k + \phi - W_k)^2 - 2\langle B_k + \widehat{\psi}_{0,N} \rangle + 1 - \frac{2}{N} \\ &< (B_k + \phi - W_k)^2 - \frac{1}{N}, \end{aligned} \tag{12}$$

where the inequality in the last line follows from (11). Let

$$\mathbf{b} = \left[\begin{array}{cccc} \phi & \widehat{\psi}_{1,N} & \cdots & \widehat{\psi}_{m,N} \end{array} \right]'$$

be the vector of length $m+1$ with components $b_0 = \phi$ and $b_\ell = \widehat{\psi}_{\ell,N}$ for $\ell = 1, \dots m$. Now,

$$NT_N(\mathbf{b}) = \sum_{n=1}^{N} \langle B_n + \phi \rangle^2 \leq \sum_{n=1}^{N} (B_n + \phi - F_n)^2,$$

and using the inequality from (12),

$$\begin{aligned} NT_N(\mathbf{b}) &< -\frac{1}{N} + \sum_{n=1}^{N} (B_n + \phi - W_n)^2 \\ &= -\frac{1}{N} + \sum_{n=1}^{N} (B_n + \widehat{\psi}_{0,N} + \frac{1}{N} - W_n)^2 \\ &= \sum_{n=1}^{N} (B_n + \widehat{\psi}_{0,N} - W_n)^2 + \frac{2}{N} \sum_{n=1}^{N} (B_n + \widehat{\psi}_{0,N} - W_n) \\ &= NT_N(\widehat{\boldsymbol{\psi}}_N), \end{aligned}$$

because $\frac{2}{N} \sum_{n=1}^{N} (B_n + \widehat{\psi}_{0,N} - W_n) = 0$ as a result of (10). But, now $T_N(\mathbf{b}) < T_N(\widehat{\boldsymbol{\psi}}_N)$ violating the fact that $\widehat{\boldsymbol{\psi}}_N$ is a minimiser of T_N. So (11) is false by contradiction.

If $\langle B_k + \widehat{\psi}_{0,N} \rangle < -\frac{1}{2} + \frac{1}{2N}$ for some k, we set $F_k = W_k - 1$ and using the same procedure as before obtain $T_N(\mathbf{b}) < T_N(\widehat{\boldsymbol{\psi}}_N)$ again. The proof follows. ■

Lemma 3. *With $K_{\ell,N}$ defined as in* (7) *and $h = f(-1/2)$,*

$$K_{\ell,N} = (h-1)\sqrt{N} \sum_{k=0}^{m} \widehat{\psi}_{k,N} \big(C_{\ell k} + o_P(1) \big) + o_P(1)$$

for all $\ell = 0, \dots, m$, where $C_{\ell k} = \frac{1}{\ell + k + 1}$.

Proof: Care must be taken since $\widehat{\boldsymbol{\psi}}_N$ depends on the sequence $\{\Phi_n\}$. For $n = 1, \dots, N$ and positive N, let

$$p_{nN}(\boldsymbol{\psi}) = \sum_{k=0}^{m} \left(\tfrac{n}{N}\right)^k \psi_k, \tag{13}$$

and put

$$q_n(x) = \lceil \Phi_n + x \rfloor, \qquad Q(x) = \mathbb{E} q_n(x) = \mathbb{E} q_1(x).$$

Let

$$G_N(\boldsymbol{\psi}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(\tfrac{n}{N}\right)^\ell \big(q_n(p_{nN}(\boldsymbol{\psi})) - Q(p_{nN}(\boldsymbol{\psi})) \big). \tag{14}$$

and put

$$\widehat{p}_{nN} = p_{nN}(\widehat{\boldsymbol{\psi}}_N) = \sum_{k=0}^{m} \left(\tfrac{n}{N}\right)^k \widehat{\psi}_{k,N}. \tag{15}$$

Observe that G_N depends on ℓ and we could write $G_{\ell,N}$ but have suppressed the subscript ℓ for notational simplicity. Now W_n from (6) can be written as $W_n = \lceil \Phi_n + \widehat{p}_{nN} \rfloor = q_n(\widehat{p}_{nN})$ and $K_{\ell,N}$ from (7) can be written as

$$\begin{aligned}
K_{\ell,N} &= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tfrac{n}{N})^\ell \big(q_n(\widehat{p}_{nN}) - \widehat{p}_{nN}\big) \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tfrac{n}{N})^\ell \big(q_n(\widehat{p}_{nN}) - \widehat{p}_{nN} + Q(\widehat{p}_{nN}) - Q(\widehat{p}_{nN})\big) \\
&= G_N(\widehat{\boldsymbol{\psi}}_N) + H_{\ell,N},
\end{aligned}$$

where

$$H_{\ell,N} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tfrac{n}{N})^\ell \big(Q(\widehat{p}_{nN}) - \widehat{p}_{nN}\big). \tag{16}$$

Lemma 6 in the Appendix shows that for any $\delta > 0$ and $\nu > 0$ there exists an $\epsilon > 0$ such that

$$\Pr\left\{ \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| > \delta \right\} < \nu$$

for all positive integers N, where $\|\boldsymbol{\psi}\|_\infty = \sup_k |\psi_k|$. Since $\widehat{\boldsymbol{\psi}}_N$ converges almost surely to zero, it follows that for any $\epsilon > 0$,

$$\lim_{N\to\infty} \Pr\left\{ \|\widehat{\boldsymbol{\psi}}_N\|_\infty \geq \epsilon \right\} = 0$$

and therefore $\Pr\{\|\widehat{\boldsymbol{\psi}}_N\|_\infty \geq \epsilon\} < \nu$ for all sufficiently large N. Now

$$\begin{aligned}
&\Pr\left\{ \left|G_N(\widehat{\boldsymbol{\psi}}_N)\right| > \delta \right\} \\
&\quad = \Pr\left\{ \left|G_N(\widehat{\boldsymbol{\psi}}_N)\right| > \delta \,,\; \|\widehat{\boldsymbol{\psi}}_N\|_\infty < \epsilon \right\} \\
&\qquad + \Pr\left\{ \left|G_N(\widehat{\boldsymbol{\psi}}_N)\right| > \delta \,,\; \|\widehat{\boldsymbol{\psi}}_N\|_\infty \geq \epsilon \right\} \\
&\quad \leq \Pr\left\{ \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| > \delta \right\} + \Pr\left\{ \|\widehat{\boldsymbol{\psi}}_N\|_\infty \geq \epsilon \right\} \\
&\quad \leq 2\nu
\end{aligned}$$

for all sufficiently large N. Since ν and δ can be chosen arbitrarily small, it follows that $G_N(\widehat{\boldsymbol{\psi}}_N)$ converges in probability to zero as $N \to \infty$, and therefore $K_{\ell,N} = H_{\ell,N} + o_P(1)$. Lemma 4 shows that

$$H_{\ell,N} = (h-1)\sqrt{N} \sum_{k=0}^{m} \widehat{\psi}_{k,N} \big(C_{\ell k} + o_P(1)\big).$$

The proof follows. ∎

Lemma 4. *With $H_{\ell,N}$ defined in* (16) *and $\widehat{p}_{nN}$ defined in* (15)*, and with $h = f(-1/2)$,*

$$H_{\ell,N} = (h-1)\sqrt{N}\sum_{k=0}^{m}\widehat{\psi}_{k,N}\big(C_{\ell k} + o_P(1)\big),$$

where $C_{\ell k} = \frac{1}{\ell+k+1}$.

Proof: If $|x| < 1$, then

$$q_n(x) = \lceil \Phi_n + x \rfloor = \begin{cases} 1, & \Phi_n + x \geq 1/2 \\ -1, & \Phi_n + x < -1/2 \\ 0, & \text{otherwise}, \end{cases}$$

and,

$$Q(x) = E q_1(x) = \begin{cases} \int_{1/2 - x}^{1/2} f(t)\,dt, & x \geq 0 \\ -\int_{-1/2}^{-1/2 - x} f(t)\,dt, & x < 0. \end{cases}$$

Because $f(\langle x \rangle)$ is continuous at $-1/2$ it follows that

$$Q(x) = x\big(h + \zeta(x)\big),$$

where $\zeta(x)$ is a function that converges to zero as x converges to zero. Observe that $|\widehat{p}_{nN}| \leq \sum_{k=0}^{m}|\widehat{\psi}_{k,N}|$ and, since each of the $\widehat{\psi}_{k,N} \to 0$ almost surely as $N \to \infty$, it follows that $\widehat{p}_{nN} \to 0$ almost surely uniformly in $n = 1,\dots,N$ as $N \to \infty$. Thus $\zeta(\widehat{p}_{nN}) \to 0$ almost surely (and therefore also in probability) uniformly in $n = 1,\dots,N$ as $N \to \infty$. Now,

$$\begin{aligned} Q(\widehat{p}_{nN}) - \widehat{p}_{nN} &= \widehat{p}_{nN}\big(h - 1 + \zeta(\widehat{p}_{nN})\big) \\ &= \widehat{p}_{nN}\big(h - 1 + o_P(1)\big), \end{aligned}$$

and so,

$$\begin{aligned} H_{\ell,N} &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(\tfrac{n}{N})^{\ell}\widehat{p}_{nN}\big(h-1+o_P(1)\big) \\ &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(\tfrac{n}{N})^{\ell}\sum_{k=0}^{m}(\tfrac{n}{N})^{k}\widehat{\psi}_{k,N}\big(h-1+o_P(1)\big) \\ &= \sqrt{N}\sum_{k=0}^{m}\widehat{\psi}_{k,N}\frac{1}{N}\sum_{n=1}^{N}\frac{n^{\ell+k}}{N^{\ell+k+1}}\big(h-1+o_P(1)\big). \end{aligned}$$

The Riemann sum
$$\frac{1}{N}\sum_{n=1}^{N}\frac{n^{\ell+k}}{N^{\ell+k+1}} = \int_0^1 x^{k+\ell+1}dx + o_P(1),$$
and since the integral above evaluates to $\frac{1}{k+\ell+1}$, we have
$$H_{\ell,N} = (h-1)\sqrt{N}\sum_{k=0}^{m}\widehat{\psi}_{k,N}\left(\frac{1}{k+\ell+1} + o_P(1)\right)$$
as required. ■

Lemma 5. *The distribution of the vector* $\mathbf{d}_N$*, defined in* (9)*, converges to the multivariate normal with zero mean and covariance matrix* $\sigma^2\mathbf{C}$*.*

Proof: For any constant vector $\boldsymbol{\alpha}$, let
$$z_N = \boldsymbol{\alpha}'\mathbf{d}_N = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\Phi_n\sum_{\ell=0}^{m}\alpha_\ell\left(\frac{n}{N}\right)^\ell.$$
Observe that Φ_n has zero mean and variance σ^2 as a result of Proposition 1. By Lypanov's central limit theorem $\mathbf{z}_N$ is asymptotically normally distributed with zero mean and variance
$$\lim_{N\to\infty}\sigma^2\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{\ell=0}^{m}\alpha_\ell\left(\frac{n}{N}\right)^\ell\right)^2 = \sigma^2\boldsymbol{\alpha}'\mathbf{C}\boldsymbol{\alpha}.$$
By the Cramér-Wold theorem it follows that $\mathbf{d}_N$ is asymptotically normally distributed with zero mean and covariance $\sigma^2\mathbf{C}$. ■

IV. Simulations

This section describes the results of Monte-Carlo simulations with the least squares unwrapping (LSU) estimator, Kitchen's unwrapping estimator [**?**], and the discrete polynomial phase transform (DPT) estimator of Peleg and Porat [**?**]. The sample sizes considered are $N = 10, 50, 200$ and the unknown amplitude is $\rho = 1$. The $X_1, \dots, X_N$ are pseudorandomly generated independent and identically distributed circularly symmetric complex Gaussian random variables with variance σ_c^2. The coefficients $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_0, \dots, \tilde{\mu}_m]$ are distributed uniformly randomly in the identifiable region B [**?** , Sec. 3]. The number of replications of each experiment is $T = 2000$ to obtain estimates $\widehat{\boldsymbol{\mu}}_1, \dots, \widehat{\boldsymbol{\mu}}_T$ and the corresponding dealiased errors $\widehat{\boldsymbol{\lambda}}_t = \operatorname{dealias}(\widehat{\boldsymbol{\mu}}_T - \tilde{\boldsymbol{\mu}})$ 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 $\widehat{\boldsymbol{\lambda}}_t$.

Figure 1 shows the sample MSEs obtained for a polynomial phase signal of order $m = 3$. The LSU estimator can be computed by finding a nearest lattice point in a particular lattice [**? ?**]. When $N = 10$

and 50 the LSU estimator can be computed exactly using a general purpose algorithm for finding nearest lattice points called the sphere decoder [**? ? ?**]. This is displayed by the circles in the figures. When $N = 200$ the sphere decoder is computationally intractable and we instead use an approximate nearest point algorithm called the K-best method [**?**]. This is displayed by the dots. For the purpose of comparison we have also plotted the results for the K-best method when $N = 10$ and 50. The asymptotic variance predicted in Theorem 1 is displayed by the dashed line. Provided the noise variance is small enough (so that the 'threshold' is avoided) the sample MSE of the LSU estimator is close to that predicted by Theorem 1. The Cramér-Rao lower bound for the variance of unbiased polynomial phase estimators in Gaussian noise is also plotted using the solid line [**?**]. When the noise variance is small the asymptotic variance of the LSU estimator is close to the Cramér-Rao lower bound.

Kitchen's unwrapping estimator and the DPT estimator perform very poorly in Figure 1. The reason is that both estimators do not work correctly for all parameters in the identifiable region B. For this reason Figure 2 plots the sample MSEs for the estimators with the true coefficients fixed to $\tilde{\boldsymbol{\mu}} = [0.1, 0.2, 10^{-3}, 10^{-5}]$. This is within the range suitable for the DPT estimator [**?**]. Both Kitchen's unwrapping estimator and the DPT estimator perform better in this case. Only results for the zeroth and third order parameters $\tilde{\mu}_0$ and $\tilde{\mu}_3$ are displayed. The results for the other parameters are similar.

V. Conclusion

This series of papers has considered the estimation of the coefficients of a noisy polynomial phase signal by least squares phase unwrapping (LSU). Under some assumptions on the distribution of the noise, it has been shown that the LSU estimator is strongly consistent and asymptotically normally distributed. The results of Monte Carlo experiment were described and these support the asymptotic analysis.

It is shown in [**? ?**] how the LSU estimator can be computed by finding a nearest lattice point in a particular lattice. Polynomial time nearest point algorithms for these lattices exist [**?** , Sec 4.3], but these algorithms are not fast in practice. The major outstanding question is whether faster nearest point algorithms exist for these specific lattices. 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.

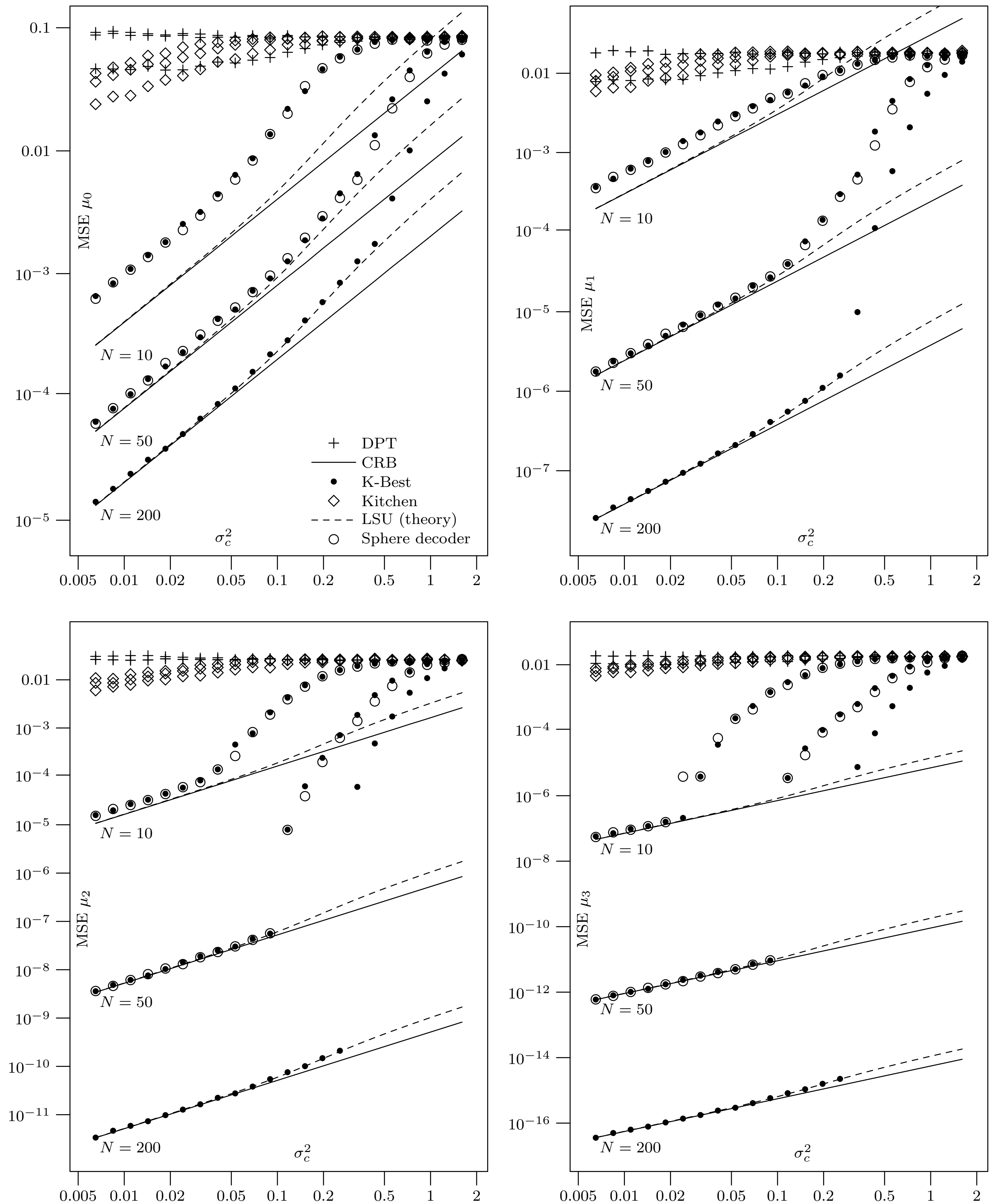

Fig. 1. Mean square error (MSE) of polynomial phase estimators. The true coefficients $\tilde{\boldsymbol{\mu}}$ are uniformly distributed in the identifiable region B. (Top left) MSE in the phase coefficient μ_0. (Top right) MSE in the frequency coefficient μ_1. (Bottom left) MSE in the quadratic coefficient μ_2. (Bottom right) MSE in the cubic coefficient μ_3.

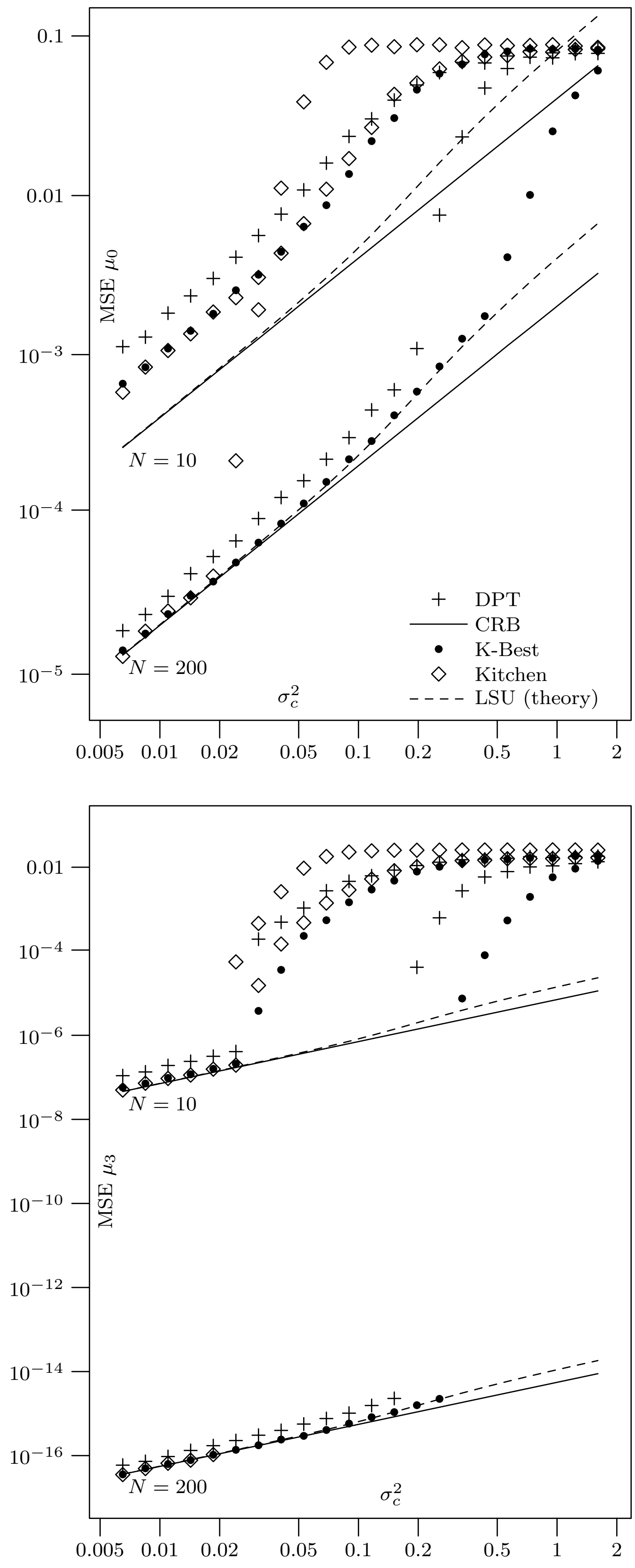

Fig. 2. Mean square error (MSE) of polynomial phase estimators. The true coefficients are fixed to $\tilde{\boldsymbol{\mu}} = [0.1, 0.2, 10^{-3}, 10^{-5}]$. (Top) MSE in the phase coefficient $\widehat{\mu}_0$. (Bottom right) MSE in the cubic coefficient $\widehat{\mu}_3$.

Appendix

A. A tightness result

During the proof of asymptotic normality in Lemma 3 we made use of the following result regarding the function,

$$G_N(\boldsymbol{\psi}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(\tfrac{n}{N}\right)^{\ell} \big(q_n(p_{nN}(\boldsymbol{\psi})) - Q(p_{nN}(\boldsymbol{\psi}))\big),$$

where the functions q_n, Q and p_{nN} are defined above (14) and $\ell \in \{0, 1, \dots, m\}$.

Lemma 6. *For any $\delta > 0$ and $\nu > 0$ there exists $\epsilon > 0$ such that*

$$\Pr\left\{ \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| > \delta \right\} < \nu$$

for all positive integers N.

This result is related to what is called *tightness* or *asymptotic continuity* in the literature on empirical processes and weak convergence on metric spaces [**? ? ? ?**]. The lemma is different from what is usually proved in the literature because the function

$$p_{nN}(\boldsymbol{\psi}) = \sum_{k=0}^{m} \left(\tfrac{n}{N}\right)^{k} \psi_k$$

depends on n. Nevertheless, the methods of proof from the literature can be used if we include a known result about hyperplane arrangements [**?** , Ch. 5][**?** , Ch. 6]. Our proof is based on a technique called *symmetrisation* and another technique called *chaining* (also known as *bracketing*) [**? ?**].

Proof: Define the function

$$\begin{aligned} f_{nN}(\boldsymbol{\psi}, \Phi_n) &= \left(\tfrac{n}{N}\right)^{\ell} q_n(p_{nN}(\boldsymbol{\psi})) \\ &= \left(\tfrac{n}{N}\right)^{\ell} \left[\Phi_n + \sum_{k=0}^{m} \left(\tfrac{n}{N}\right)^{k} \psi_k \right] \end{aligned}$$

so that G_N can be written as

$$G_N(\boldsymbol{\psi}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \big(f_{nN}(\boldsymbol{\psi}, \Phi_n) - \mathbb{E} f_{nN}(\boldsymbol{\psi}, \Phi_n)\big).$$

Let $\{g_n\}$ be a sequence of independent standard normal random variables, independent of the phase noise sequence $\{\Phi_n\}$. Lemma 7 shows that

$$\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| \leq \sqrt{2\pi}\, \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})|,$$

where

$$Z_N(\boldsymbol{\psi}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} g_n f_{nN}(\boldsymbol{\psi}, \Phi_n), \tag{17}$$

and where $\mathbb{E}$ runs over both $\{g_n\}$ and $\{\Phi_n\}$. Conditionally on $\{\Phi_n\}$, the process $\{Z_N(\boldsymbol{\psi}), \boldsymbol{\psi} \in \mathbb{R}^{m+1}\}$ is a *Gaussian process*, and numerous techniques exist for its analysis. Lemma 8 shows that for any $\kappa > 0$ there exists an $\epsilon > 0$ such that

$$\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})| < \kappa.$$

It follows that,

$$\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| < \sqrt{2\pi}\, \kappa,$$

and by Markov's inequality,

$$\Pr\left\{ \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| > \delta \right\} \leq \sqrt{2\pi}\, \frac{\kappa}{\delta},$$

for any $\delta > 0$. The proof follows with $\nu = \sqrt{2\pi}\kappa/\delta$. It remains to prove Lemmas 7 and 8.

■

The proof of the next lemma is based on a technique called *symmetrisation* [**?** **?**] and the proof we give closely follows that of Pollard [**?** , Section 4].

Lemma 7. *(Symmetrisation)*

$$\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| \leq \sqrt{2\pi}\, \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})|\,.$$

Proof: Let $\{\Phi'_n\}$ be a sequence of random variables distributed identically to $\{\Phi_n\}$ and independent of both $\{\Phi_n\}$ and $\{g_n\}$. Write $\mathbb{E}_\Phi$ to denote expectation conditional on $\{\Phi_n\}$. Since Φ_n and Φ'_n are identically distributed,

$$\mathbb{E} f_{nN}(\boldsymbol{\psi}, \Phi_n) = \mathbb{E}_\Phi f_{nN}(\boldsymbol{\psi}, \Phi'_n),$$

so that,

$$\begin{aligned} G_N(\boldsymbol{\psi}) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \big(f_{nN}(\boldsymbol{\psi}, \Phi_n) - \mathbb{E}_\Phi f_{nN}(\boldsymbol{\psi}, \Phi'_n) \big) \\ &= \mathbb{E}_\Phi \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (f_{nN} - f'_{nN}), \end{aligned}$$

where, for notational convenience, we put

$$f_{nN} = f_{nN}(\boldsymbol{\psi}, \Phi_n) \qquad \text{and} \qquad f'_{nN} = f_{nN}(\boldsymbol{\psi}, \Phi'_n).$$

Taking absolute values followed by supremums,

$$\begin{aligned}\sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} |G_N(\boldsymbol{\psi})| &= \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \mathbb{E}_\Phi \sum_{n=1}^{N} (f_{nN} - f'_{nN}) \right| \\ &\leq \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \mathbb{E}_\Phi \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (f_{nN} - f'_{nN}) \right|,\end{aligned}$$

the upper bound following from Jensen's inequality. Since $\sup \mathbb{E} |\dots| \leq \mathbb{E} \sup |\dots|$,

$$\sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} |G_N(\boldsymbol{\psi})| \leq \mathbb{E}_\Phi \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (f_{nN} - f'_{nN}) \right|. \tag{18}$$

Applying $\mathbb{E}$ to both sides gives the inequality

$$\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} |G_N(\boldsymbol{\psi})| \leq \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (f_{nN} - f'_{nN}) \right|, \tag{19}$$

since $\mathbb{E}\mathbb{E}_\Phi$ is equivalent to $\mathbb{E}$. Let $\sigma_n = g_n / |g_n|$, and put $\sigma_n = 1$ if $g_n = 0$. The σ_n are thus independent with $\Pr\{\sigma_n = -1\} = \frac{1}{2}$ and $\Pr\{\sigma_n = 1\} = \frac{1}{2}$. The symmetry of the distribution of g_n implies that g_n and σ_n are independent. As Φ_n and Φ'_n are independent and identically distributed, the random variable $f_{nN} - f'_{nN}$ is symmetrically distributed about zero, and is therefore distributed identically to $\sigma_n(f_{nN} - f'_{nN})$. Thus,

$$\begin{aligned}&\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (f_{nN} - f'_{nN}) \right| \\ &\qquad = \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n (f_{nN} - f'_{nN}) \right|.\end{aligned}$$

By the triangle inequality

$$\left| \sum_{n=1}^{N} \sigma_n (f_{nN} - f'_{nN}) \right| \leq \left| \sum_{n=1}^{N} \sigma_n f_{nN} \right| + \left| \sum_{n=1}^{N} \sigma_n f'_{nN} \right|,$$

and it follows from (19), that

$$\begin{aligned}\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} |G_N(\boldsymbol{\psi})| &\leq \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n f_{nN} \right| \\ &\qquad + \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n f'_{nN} \right| \\ &= 2\, \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n f_{nN} \right|.\end{aligned}$$

Write $\mathbb{E}_{\Phi\sigma}$ to denote expectation conditional on $\{\Phi_n\}$ and $\{\sigma_n\}$. Since g_n is a standard normal random variable, $\mathbb{E}\,|g_n| = \mathbb{E}_{\Phi\sigma}\,|g_n| = \sqrt{2/\pi}$, and

$$\begin{aligned}
&\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n f_{nN} \right| \\
&\qquad = \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n f_{nN} \sqrt{\frac{\pi}{2}} \mathbb{E}_{\Phi\sigma}\,|g_n| \right| \\
&\qquad \leq \sqrt{\frac{\pi}{2}} \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sigma_n\,|g_n|\, f_{nN} \right|,
\end{aligned}$$

the last line following from Jensen's inequality and by moving $\mathbb{E}_{\Phi\sigma}$ outside the supremum similarly to (18). As $g_n = \sigma_n\,|g_n|$, it follows that

$$\begin{aligned}
\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |G_N(\boldsymbol{\psi})| &\leq 2\sqrt{\frac{\pi}{2}} \mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} g_n f_{nN} \right| \\
&= \sqrt{2\pi}\,\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})|
\end{aligned}$$

as required. ∎

Lemma 8. *For any $\kappa > 0$ there exists $\epsilon > 0$ such that*

$$\mathbb{E} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})| < \kappa,$$

where $Z_N(\boldsymbol{\psi})$ is defined by (17).

Proof: Without loss of generality, assume that $\epsilon < \frac{1}{m+1}$. Lemma 9 shows that

$$\mathbb{E}_{\Phi} \sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})| \leq K_1 \sqrt{C_\epsilon(\{\Phi_n\})} \tag{20}$$

where K_1 is a finite, positive constant, and $C_\epsilon(\{\Phi_n\})$ is the average number of times $|\Phi_1|, \dots, |\Phi_N|$ is greater than or equal to $1/2 - (m+1)\epsilon$. That is,

$$C_\epsilon(\{\Phi_n\}) = \frac{1}{N} \sum_{n=1}^{N} I_\epsilon(|\Phi_n|) \tag{21}$$

where $I_\epsilon(|\Phi_n|)$ is 1 when $|\Phi_n| \geq 1/2 - (m+1)\epsilon$ and zero otherwise. Recall that f is the probability density function of Φ_n, and (by assumption in Theorem 1) that $f(\langle x \rangle)$ is continuous at $x = -1/2$. Because of this, the expected value of $C_\epsilon(\{\Phi_n\})$ is small when ϵ is small, since

$$\begin{aligned}\mathbb{E}C_\epsilon(\{\Phi\}) &= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}I_\epsilon(|\Phi_n|) \\ &= \Pr\left\{|\Phi_1| \geq 1/2 - (m+1)\epsilon\right\} \\ &= \int_{-1/2}^{-1/2+(m+1)\epsilon} f(\phi)d\phi + \int_{1/2-(m+1)\epsilon}^{1/2} f(\phi)d\phi \\ &= \int_{-1/2+(m+1)\epsilon}^{1/2-(m+1)\epsilon} f(\langle\phi\rangle)d\phi \\ &= 2(m+1)\epsilon(f(-1/2) + o(1))\end{aligned}$$

where $o(1)$ goes to zero as ϵ goes to zero. Since $\sqrt{\cdot}$ is a concave function on the positive real line, and since $C_\epsilon(\{\Phi_n\})$ is always positive,

$$\mathbb{E}\sqrt{C_\epsilon(\{\Phi_n\})} \leq \sqrt{\mathbb{E}C_\epsilon(\{\Phi_n\})} < \sqrt{K_2\epsilon}$$

by Jensen's inequality, for some constant K_2. Applying $\mathbb{E}$ to both sides of (20),

$$\mathbb{E}\sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon}|Z_N(\boldsymbol{\psi})| \leq K_1\sqrt{\mathbb{E}C_\epsilon(\{\Phi_n\})} < K_1\sqrt{K_2\epsilon}.$$

Choosing $\epsilon = \kappa^2/(K_1^2K_2)$ completes the proof. It remains to prove Lemma 9. ■

The proofs of Lemmas 9 and 11 are based on a technique called *chaining* (or *bracketing*) [**?** **?** **?** **?** **?**]. The proofs here follow those of Pollard [**?**]. In the remaining lemmas we consider expectation conditional on $\{\Phi_n\}$ and consequently treat $\{\Phi_n\}$ as a fixed realisation. We also use the abbreviations

$$C_\epsilon = C_\epsilon(\{\Phi_n\}) \qquad \text{and} \qquad f_{nN}(\boldsymbol{\psi}) = f_{nN}(\boldsymbol{\psi}, \Phi_n)$$

As in Lemma 8 we assume, without loss of generality, that $\epsilon < \frac{1}{m+1}$.

Lemma 9. *There exists a positive constant* K_1 *such that*

$$\mathbb{E}_\Phi \sup_{\|\boldsymbol{\psi}\|_\infty<\epsilon}|Z_N(\boldsymbol{\psi})| \leq K_1\sqrt{C_\epsilon}.$$

Proof: Let

$$B_\epsilon = \{\mathbf{x} \in \mathbb{R}\,;\ \|\mathbf{x}\|_\infty < \epsilon\}.$$

For each non negative integer k, let $T_\epsilon(k)$ be a discrete subset of $\mathbb{R}^{m+1}$ with the property that for every $\boldsymbol{\psi} \in B_\epsilon$ there exists some $\boldsymbol{\psi}^* \in T_\epsilon(k)$ such that the pseudometric

$$d(\boldsymbol{\psi}, \boldsymbol{\psi}^*) = \sum_{n=1}^{N}\left(f_{nN}(\boldsymbol{\psi}) - f_{nN}(\boldsymbol{\psi}^*)\right)^2 \leq 2^{-k}C_\epsilon N.$$

We specifically define $T_\epsilon(0)$ to contain a single point, that being the origin $\mathbf{0}$. Defined this way, $T_\epsilon(0)$ satisfies the inequality above because

$$d(\boldsymbol{\psi}, \mathbf{0}) = \sum_{n=1}^{N} f_{nN}(\boldsymbol{\psi})^2 \leq C_\epsilon N,$$

for all $\boldsymbol{\psi} \in B_\epsilon$, as a result of Lemma 10.

The existence of $T_\epsilon(k)$ for each positive integer k will be proved in Lemma 13. It is worth giving some intuition regarding $T_\epsilon(k)$. If we place a 'ball' of radius $2^{-k} C_\epsilon N$ with respect to the pseudometric $d(\cdot,\cdot)$ around each point in $T_\epsilon(k)$, then, by definition, the union of these balls is a superset of B_ϵ. The balls are said to *cover* B_ϵ and $T_\epsilon(k)$ is said to form a *covering* of B_ϵ [**?** , Section 1.2]. The minimum number of such balls required to cover B_ϵ is called a *covering number* of B_ϵ. In Lemma 13 we show that no more than $K_3 2^{(m+1)k}$ balls of radius $2^{-k} C_\epsilon N$ are required to cover B_ϵ, where K_3 is a constant, independent of N and ϵ. We make use of this upper bound in Lemma 11.

We now continue the proof. For all $\boldsymbol{\psi} \in \mathbb{R}^{m+1}$,

$$f_{nN}(\boldsymbol{\psi}) = \left(\tfrac{n}{N}\right)^{\ell} \lceil \Phi_n + p_{nN}(\boldsymbol{\psi}) \rfloor$$

is a multiple of $N^{-\ell}$ and so $d(\boldsymbol{\psi}, \boldsymbol{\psi}^*)$ is a multiple of $N^{-2\ell}$. When $2^k > C_\epsilon N^{1+2\ell}$ we have

$$0 \leq d(\boldsymbol{\psi}, \boldsymbol{\psi}^*) < 2^{-k} C_\epsilon N < N^{-2\ell},$$

and so we must have $d(\boldsymbol{\psi}, \boldsymbol{\psi}^*) = 0$. Consequently $f_{nN}(\boldsymbol{\psi}) = f_{nN}(\boldsymbol{\psi}^*)$ for every $n = 1, \dots, N$, and $Z_N(\boldsymbol{\psi}) = Z_N(\boldsymbol{\psi}^*)$. Thus,

$$\sup_{\|\boldsymbol{\psi}\|_\infty < \epsilon} |Z_N(\boldsymbol{\psi})| = \sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(\boldsymbol{\psi})|$$

for all k large enough that $2^k > C_\epsilon N^{1+2\ell}$. So, to analyse the supremum of $Z_N(\boldsymbol{\psi})$ over the continuous interval B_ϵ it is enough to analyse the supremum over the discrete set $T_\epsilon(k)$ for large k. Lemma 11 shows that

$$\mathbb{E}_\Phi \sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(\boldsymbol{\psi})| \leq \sqrt{C_\epsilon} \sum_{i=1}^{k} \frac{\sqrt{iA_1 + A_2}}{2^{i/2}} < \infty$$

for every positive integer k, where A_1 and A_2 are the constants,

$$A_1 = 18(m+1)\log 2 \qquad \text{and} \qquad A_2 = 18 \log K_3 \tag{22}$$

and $\log(\cdot)$ is the natural logarithm. The lemma holds with $K_1 = \sum_{i=1}^{\infty} 2^{-i/2} \sqrt{iA_1 + A_2}$. ■

Lemma 10. *For $\epsilon < \frac{1}{m+1}$ and all $\boldsymbol{\psi} \in B_\epsilon$ and $n = 1, \dots, N$,*

$$f_{nN}(\boldsymbol{\psi})^2 \leq |f_{nN}(\boldsymbol{\psi})| \leq I_\epsilon(|\Phi_n|)$$

and consequently,

$$\sum_{n=1}^{N} f_{nN}(\boldsymbol{\psi})^2 \leq \sum_{n=1}^{N} |f_{nN}(\boldsymbol{\psi})| \leq N C_\epsilon. \tag{23}$$

Proof: Because $|\psi_i| < \epsilon$ for all $i = 0, \dots, m$,

$$|p_{nN}(\boldsymbol{\psi})| = \left| \sum_{i=0}^{m} \left(\tfrac{n}{N}\right)^i \psi_i \right| \leq (m+1)\epsilon < 1. \tag{24}$$

Since $\Phi_n \in [-1/2, 1/2)$, it follows that $f_{nN}(\boldsymbol{\psi})$ equals either $-(\frac{n}{N})^\ell$, $(\frac{n}{N})^\ell$ or 0 and so

$$f_{nN}(\boldsymbol{\psi})^2 \leq |f_{nN}(\boldsymbol{\psi})| \leq 1. \tag{25}$$

Whenever $f_{nN}(\boldsymbol{\psi}) \neq 0$ we must have

$$|\Phi_n| \geq 1/2 - |p_{nN}(\boldsymbol{\psi})| \geq 1/2 - (m+1)\epsilon$$

and $I_\epsilon(|\Phi_n|) = 1$. Thus, for all $n = 1, \dots, N$,

$$f_{nN}(\boldsymbol{\psi})^2 \leq |f_{nN}(\boldsymbol{\psi})| \leq I_\epsilon(|\Phi_n|).$$

Summing the terms in this inequality over $n = 1, \dots, N$ and using (21) gives (23). ∎

Lemma 11. *(Chaining) For all positive integers* k*,*

$$\mathbb{E}_\Phi \sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(\boldsymbol{\psi})| \leq \sqrt{C_\epsilon} \sum_{i=1}^{k} \frac{\sqrt{iA_1 + A_2}}{2^{i/2}},$$

where the constants A_1 *and* A_2 *are defined in* (22)*.*

Proof: Let b_k be a function that maps each $\boldsymbol{\psi} \in T_\epsilon(k)$ to $b_k(\boldsymbol{\psi}) \in T_\epsilon(k-1)$ such that $d(\boldsymbol{\psi}, b_k(\boldsymbol{\psi})) \leq 2^{1-k} C_\epsilon N$. The existence of the function b_k is guaranteed by the definition of $T_\epsilon(k)$. By the triangle inequality,

$$|Z_N(\boldsymbol{\psi})| \leq |Z_N(b_k(\boldsymbol{\psi}))| + |Z_N(\boldsymbol{\psi}) - Z_N(b_k(\boldsymbol{\psi}))|,$$

and by taking supremums on both sides,

$$\begin{aligned} &\sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(\boldsymbol{\psi})| \\ &\leq \sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(b_k(\boldsymbol{\psi}))| + \sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(\boldsymbol{\psi}) - Z_N(b_k(\boldsymbol{\psi}))| \\ &\leq \sup_{\boldsymbol{\psi} \in T_\epsilon(k-1)} |Z_N(\boldsymbol{\psi})| + \sup_{\boldsymbol{\psi} \in T_\epsilon(k)} |Z_N(\boldsymbol{\psi}) - Z_N(b_k(\boldsymbol{\psi}))|, \end{aligned} \tag{26}$$

the last line following since $b_k(\boldsymbol{\psi}) \in T_\epsilon(k-1)$ and so

$$\sup_{\boldsymbol{\psi}\in T_\epsilon(k)} |Z_N(b_k(\boldsymbol{\psi}))| \leq \sup_{\boldsymbol{\psi}\in T_\epsilon(k-1)} |Z_N(\boldsymbol{\psi})| \, .$$

Conditional on $\Phi_1, \Phi_2, \dots$ the random variable

$$X(\boldsymbol{\psi}) = Z_N(\boldsymbol{\psi}) - Z_N(b_k(\boldsymbol{\psi}))$$

has zero mean, and is normally distributed with variance

$$\begin{aligned}
\sigma_X^2 &= \mathbb{E}_\Phi \frac{1}{N}\sum_{n=1}^{N} g_n^2 (f_{nN}(\boldsymbol{\psi}) - f_{nN}(b_k(\boldsymbol{\psi})))^2 \\
&= \frac{1}{N}\sum_{n=1}^{N} (f_{nN}(\boldsymbol{\psi}) - f_{nN}(b_k(\boldsymbol{\psi})))^2 \\
&= d(\boldsymbol{\psi}, b_k(\boldsymbol{\psi})) \leq 2^{1-k} C_\epsilon,
\end{aligned}$$

since $\mathbb{E}_\Phi g_n^2 = 1$. Using Lemma 12,

$$\begin{aligned}
\mathbb{E}_\Phi \sup_{\boldsymbol{\psi}\in T_\epsilon(k)} |X(\boldsymbol{\psi})| &\leq 3\sqrt{2^{1-k} C_\epsilon \log |T_\epsilon(k)|} \\
&\leq \sqrt{C_\epsilon} \frac{\sqrt{kA_1 + A_2}}{2^{k/2}}
\end{aligned}$$

because $\log |T_\epsilon(k)| \leq k(m+1)\log 2 + \log K_3$. Taking expectations on both sides of (26),

$$\begin{aligned}
\mathbb{E}_\Phi \sup_{\boldsymbol{\psi}\in T_\epsilon(k)} |Z_N(\boldsymbol{\psi})| \leq \mathbb{E}_\Phi &\sup_{\boldsymbol{\psi}\in T_\epsilon(k-1)} |Z_N(\boldsymbol{\psi})| \\
&+ \sqrt{C_\epsilon} \frac{\sqrt{kA_1 + A_2}}{2^{k/2}},
\end{aligned}$$

which involves a recursion in k. By unravelling the recursion, and using the fact $T_\epsilon(0)$ contains only the origin, and therefore

$$\mathbb{E}_\Phi \sup_{\boldsymbol{\psi}\in T_\epsilon(0)} |Z_N(\boldsymbol{\psi})| = \mathbb{E}_\Phi |Z_N(0)| = 0,$$

we obtain

$$\mathbb{E}_\Phi \sup_{\boldsymbol{\psi}\in T_\epsilon(k)} |Z_N(\boldsymbol{\psi})| < \sqrt{C_\epsilon} \sum_{i=1}^{k} \frac{\sqrt{kA_1 + A_2}}{2^{k/2}}$$

as required. ∎

Lemma 12. *(Maximal inequalities) Suppose $X_1, \dots, X_N$ are zero mean Gaussian random variables each with variance less than some positive constant K, then*

$$\mathbb{E} \sup_{n=1,\dots,N} |X_n| \leq 3\sqrt{K \log N}$$

where $\log N$ *is the natural logarithm of* N.

Proof: This result is well known, see for example [**?** , Section 3] ∎

Lemma 13. *(Covering numbers) For* $k \in \mathbb{Z}$ *there exists a discrete set* $T_\epsilon(k) \subset \mathbb{R}^{m+1}$ *with the property that, for every* $\boldsymbol{\psi} \in B_\epsilon$*, there is a* $\boldsymbol{\psi}^* \in T_\epsilon(k)$ *such that,*

$$d(\boldsymbol{\psi}, \boldsymbol{\psi}^*) = \sum_{n=1}^{N} \big(f_{nN}(\boldsymbol{\psi}) - f_{nN}(\boldsymbol{\psi}^*) \big)^2 \leq \frac{C_\epsilon N}{2^k}.$$

The number of elements in $T_\epsilon(k)$ *is no more than* $K_3 2^{(m+1)k}$ *where* K_3 *is a positive constant, independent of* N*,* ϵ *and* k*.*

Before we give the proof of this lemma we need some results from the literature on hyperplane arrangements and what are called ϵ-*cuttings* [**? ?**]. Let H be a set of m-dimensional affine hyperplanes lying in $\mathbb{R}^{m+1}$. By *affine* it is meant that the hyperplanes need not pass through the origin. For each hyperplane $h \in H$ let $D(h)$ and its complement $\bar{D}(h)$ be the corresponding half spaces of $\mathbb{R}^{m+1}$. For a point $\mathbf{x} \in \mathbb{R}^{m+1}$, let

$$b(h, \mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in D(h) \\ 0 & \mathbf{x} \in \bar{D}(h). \end{cases} \tag{27}$$

Note that $b(h, \mathbf{x})$ is piecewise constant in $\mathbf{x}$. Two points $\mathbf{x}$ and $\mathbf{y}$ from $\mathbb{R}^{m+1}$ are in the same halfspace of h if and only if $b(h, \mathbf{x}) = b(h, \mathbf{y})$. Then the pseudometric

$$\sigma(\mathbf{x}, \mathbf{y}) = \sum_{h \in H} |b(h, \mathbf{x}) - b(h, \mathbf{y})|$$

is the number of hyperplanes in H that pass between the points $\mathbf{x}$ and $\mathbf{y}$.

The next theorem considers the partitioning of $\mathbb{R}^{m+1}$ into subsets so that not too many hyperplanes intersect with any subset. Proofs can be found in Theorem 5.1 on page 206 of [**?**] and also Theorem 6.5.3 on page 144 of [**?**].

Theorem 2. *There exists a constant* K*, independent of the set of hyperplanes* H*, such that for any positive real number* r*, we can partition* $\mathbb{R}^{m+1}$ *into* Kr^{m+1} *generalised* $(m+1)$*-dimensional simplices with the property that no more than* $|H|/r$ *hyperplanes from* H *pass through the interior of any simplex.*

By the phrase 'There exists a constant K, independent of the set of hyperplanes H', it is meant that the constant K is valid for every possible set of hyperplanes in $\mathbb{R}^{m+1}$, regardless of the number of hyperplanes or their position and orientation. A *generalised* $(m+1)$-dimensional simplex is the region defined by the intersection of $m+2$ half spaces in $\mathbb{R}^{m+1}$. Note that a generalised simplex (unlike an ordinary simplex) can be unbounded. For our purposes Theorem 2 is important because of the following corollary.

Corollary 1. *There exists a constant K, independent of the set of hyperplanes H, such that for every positive real number r there is a discrete subset $T \subset \mathbb{R}^{m+1}$ containing no more than Kr^{m+1} elements with the property that for every $\mathbf{x} \in \mathbb{R}^{m+1}$ there exists $\mathbf{y} \in T$ with $\sigma(\mathbf{x}, \mathbf{y}) \leq |H|/r$.*

Proof: Let C be the set of generalised simplices constructed according to Theorem 2. Define T as a set containing precisely one point from the interior of each simplex in C. Let $\mathbf{x} \in \mathbb{R}^{m+1}$. Since $b(h, \mathbf{x})$ is piecewise constant for each $h \in H$, and since C partitions $\mathbb{R}^{m+1}$, there must exist a simplex $c \in C$ with a point $\mathbf{z}$ in its interior such that $b(h, \mathbf{z}) = b(h, \mathbf{x})$ for all $h \in H$, and correspondingly $\sigma(\mathbf{z}, \mathbf{x}) = 0$. Let $\mathbf{y}$ be the element from T that is in the interior of c. Since at most $|H|/r$ hyperplanes cross the interior of c there can be at most $|H|/r$ hyperplanes between $\mathbf{z}$ and $\mathbf{y}$, and so $\sigma(\mathbf{z}, \mathbf{y}) \leq |H|/r$. Now,

$$\sigma(\mathbf{x}, \mathbf{y}) < \sigma(\mathbf{z}, \mathbf{x}) + \sigma(\mathbf{z}, \mathbf{y}) \leq \frac{|H|}{r}$$

follows from the triangle inequality. ∎

The previous corollary ensures that we can *cover* $\mathbb{R}^{m+1}$ using Kr^{m+1} 'balls' of radius $|H|/r$ with respect to the pseudometric $\sigma(\mathbf{x}, \mathbf{y})$. The balls are placed at the positions defined by points in the set T, and these points could be anywhere in $\mathbb{R}^{m+1}$. The next corollary asserts that we can cover a subset of $\mathbb{R}^{m+1}$, by placing the balls at points only within this subset.

Corollary 2. *Let B be a subset of $\mathbb{R}^{m+1}$. There exists a constant K, independent of the set of hyperplanes H, such that for every positive real number r there is a discrete subset $T_B \subset B$ containing no more than Kr^{m+1} elements with the property that for every $\mathbf{x} \in B$ there is a $\mathbf{y} \in T_B$ with $\sigma(\mathbf{x}, \mathbf{y}) \leq |H|/r$.*

Proof: Let C be the set of generalised simplices constructed according to Theorem 2 and let C_B be the subset of those indices that intersect B. Let T_B contain a point from $c \cap B$ for each simplex $c \in C_B$. The proof now follows similarly to Corollary 1. ∎

We are now ready to prove Lemma 13.

Proof: (Lemma 13) Put

$$g_{nN}(\boldsymbol{\psi}) = (\tfrac{N}{n})^{\ell} f_{nN}(\boldsymbol{\psi}) = \lceil \Phi_n + p_{nN}(\boldsymbol{\psi}) \rfloor,$$

and let

$$d_g(\boldsymbol{\psi}, \boldsymbol{\psi}^*) = \sum_{n=1}^{N} \big(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*)\big)^2.$$

We have $d(\boldsymbol{\psi}, \boldsymbol{\psi}^*) \leq d_g(\boldsymbol{\psi}, \boldsymbol{\psi}^*)$, and so it suffices to prove the Lemma with d replaced by d_g.

From (24),

$$|p_{nN}(\boldsymbol{\psi})| \leq (m+1)\epsilon < 1.$$

Since $\Phi_n \in [-1/2, 1/2)$, when $\Phi_n \geq 0$,

$$g_{nN}(\boldsymbol{\psi}) = \begin{cases} 1 & p_{nN}(\boldsymbol{\psi}) \geq 1/2 - \Phi_n \\ 0 & \text{otherwise,} \end{cases}$$

and when $\Phi_n < 0$,

$$g_{nN}(\boldsymbol{\psi}) = \begin{cases} -1 & p_{nN}(\boldsymbol{\psi}) < -1/2 - \Phi_n \\ 0 & \text{otherwise.} \end{cases}$$

Thus $(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*))^2$ is either equal to one when $g_{nN}(\boldsymbol{\psi}) \neq g_{nN}(\boldsymbol{\psi}^*)$ or zero when $g_{nN}(\boldsymbol{\psi}) = g_{nN}(\boldsymbol{\psi}^*)$. Now $g_{nN}(\boldsymbol{\psi}) \neq 0$ only if

$$|\Phi_n| \geq 1/2 - |p_{nN}(\boldsymbol{\psi})| \geq 1/2 - (m+1)\epsilon,$$

that is, only if $I_\epsilon(\Phi_n) = 1$. Let

$$A = \{n \in \{1, \dots, N\} \; ; \; I_\epsilon(\Phi_n) = 1\}$$

be the subset of the indices where $I_\epsilon(\Phi_n) = 1$. By definition the number of elements in A is $C_\epsilon N$ (see (21)). If both $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$ are in B_ϵ, then

$$(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*))^2 \neq 0$$

only if $n \notin A$. Thus,

$$\begin{aligned} d_g(\boldsymbol{\psi}, \boldsymbol{\psi}^*) &= \sum_{n=1}^{N} \big(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*)\big)^2 \\ &= \sum_{n \in A} \big(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*)\big)^2. \end{aligned}$$

We now use Corollary 2. Let h_n be the m dimensional hyperplane in $\mathbb{R}^{m+1}$ satisfying

$$p_{nN}(\boldsymbol{\psi}) = \sum_{i=0}^{m} \left(\tfrac{n}{N}\right)^i \psi_i = \tfrac{1}{2}\operatorname{sgn}(\Phi_n) - \Phi_n$$

where $\operatorname{sgn}(\Phi_n)$ is equal to 1 when $\Phi_n \geq 0$ and -1 otherwise. The hyperplane h_n divides $\mathbb{R}^{m+1}$ into two halfspaces, $D(h_n)$ and its complement $\bar{D}(h_n)$. If $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$ are in the same halfspace, then

$|b(h_n, \boldsymbol{\psi}) - b(h_n, \boldsymbol{\psi}^*)| = 0$ and $g_{nN}(\boldsymbol{\psi}) = g_{nN}(\boldsymbol{\psi}^*)$, and therefore $(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*))^2 = 0$. Otherwise, if $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$ are in different halfspaces, then $|b(h_n, \boldsymbol{\psi}) - b(h_n, \boldsymbol{\psi}^*)| = 1$ and $g_{nN}(\boldsymbol{\psi}) \neq g_{nN}(\boldsymbol{\psi}^*)$, and therefore $(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*))^2 = 1$. Thus,

$$(g_{nN}(\boldsymbol{\psi}) - g_{nN}(\boldsymbol{\psi}^*))^2 = |b(h_n, \boldsymbol{\psi}) - b(h_n, \boldsymbol{\psi}^*)|$$

for $n = 1, \dots, N$. Let H be the finite set of hyperplanes $\{h_n, n \in A\}$ and observe that the number of hyperplanes is $|H| = |A| = C_\epsilon N$. When both $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$ are inside B_ϵ, d_g can be written as

$$\begin{aligned} d_g(\boldsymbol{\psi}, \boldsymbol{\psi}^*) &= \sum_{n \in A} |b(h_n, \boldsymbol{\psi}) - b(h_n, \boldsymbol{\psi}^*)| \\ &= \sum_{h \in H} |b(h, \boldsymbol{\psi}) - b(h, \boldsymbol{\psi}^*)| = \sigma(\boldsymbol{\psi}, \boldsymbol{\psi}^*). \end{aligned}$$

That is, when both $\boldsymbol{\psi}, \boldsymbol{\psi}^* \in B_\epsilon$, $d_g(\boldsymbol{\psi}, \boldsymbol{\psi}^*)$ is the number of hyperplanes from H that pass between the points $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$.

It follows from Corollary 2 that for any positive r there exists a finite subset T_B of B_ϵ containing at most $K_3 r^{m+1}$ elements, such that for every $\boldsymbol{\psi} \in B_\epsilon$ there is a $\boldsymbol{\psi}^* \in T_B$ with

$$d_g(\boldsymbol{\psi}, \boldsymbol{\psi}^*) = \sigma(\boldsymbol{\psi}, \boldsymbol{\psi}^*) \leq \frac{|H|}{r} = \frac{|A|}{r} = \frac{C_\epsilon N}{r}.$$

Putting $r = 2^k$ and choosing $T_\epsilon(k) = T_B$ completes the proof. ■