THE UNIVERSITY OF QUEENSLAND

Lattice theory, circular statistics and polynomial phase signals

By

Robby G. McKilliam

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in December 2010
School of Information Technology and Electrical Engineering

Typeset in LaTeX 2_ε.

—That's why it's always worth having a few philosophers around the place. One minute it's all Is Truth Beauty and Is Beauty Truth, and Does A Falling Tree in the Forest Make A Sound if There's No one There to Hear It, and then just when you think they're going to start dribbling one of 'em says, Incidentally, putting a thirty-foot parabolic reflector on a high place to shoot the rays of the sun at an enemy's ships would be a very interesting demonstration of optical principles.

Terry Pratchett on the many and varied advantages of philosophy.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis. I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award. I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the Copyright Act 1968. I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material.

_____

Robby G. McKilliam

Lattice theory, circular statistics and polynomial phase signals

Robby G. McKilliam

December 12, 2010

Acknowledgments

First and foremost thanks go to my supervisor, Vaughan Clarkson, who has been tremendously supportive throughout the course of my study. I honestly can't imagine a better supervisor. I must also thank my associate supervisor Barry Quinn who has been generous both with his time and his mathematical rigor. If I have learnt any statistics during the course of my PhD it is due to Barry. I must also thank my other associate supervisor, Iain Collings, for giving his time and also for recommending that I apply for a scholarship with the CSIRO. Financially, this has been incredibly helpful. I would like to thank Daniel Ryan for having me work with him in at NTNU and also for being good company at a number of conferences.

There are a number of people whom I must thank for accommodating me at various times. Firstly Robert Dallinger and Peter Fertl for having me in Vienna, Glen Maddern and Ben Hoskings for having me in Melbourne, Matt McKay for accommodating me in Hong Kong and Doug Cochran for accommodating me in Phoenix. Doug deserves a special mention for accompanying me on a motorbike ride from his home in Phoenix to the ICASSP conference in Las Vegas.

I must also thank Tom Wallis, Michael Kearney, Chris Nolan, Warrick Roseboom and Michael Poole for being regulars at the University of Queensland staff club on a Friday evening. Friday evenings have been significantly less fun since the departure of Tom (to Harvard University) and Michael (to Melbourne University). I hope to have a beer with Tom and Michael again soon. Thanks also to good friends, Michael Lane, David Muller, Daniel Manderson, Tim Veitch and Julia Keenan who are a constant source fun and interesting conversation.

Finally I acknowledge the unwavering support and love of my family. My parents Rob and Lisa, and my brother and sister, Andy and Nikki. Last, but certainly not least I am immensely grateful to my loving girlfriend Pip whose patience, particularly in recent months, has been much appreciated.

Abstract

This thesis studies connections between two fields, **lattice theory** and **circular statistics**. We focus on the estimation and theoretical analysis of **polynomial phase signals**. These signals have a vast array of applications in science, in particular in astronomy, optics, biology, geology, geography and meteorology and also in engineering, particularly in communications and radar. Despite this, we find that the theoretical tools for analysing these signals are lacking. We discover a special family of **lattices**, called $V^*_{n/m}$, that are particularly useful for studying polynomial phase signals. Using these lattices we are able to close a number of the theoretical gaps that exist in the literature, and also produce some remarkably accurate estimators.

We firstly describe some new results in the field of lattice theory. The most significant result is the discovery of a fast **nearest point algorithm** for the lattice A^*_n and also a related family of lattices called the **Coxeter lattices**. The new algorithms all require a linear number of operations in the dimension of the lattice. This is significantly faster than previous algorithms that require, in the worst case, a quadratic number of operations. We then study the lattices $V^*_{n/m}$. We describe a number of their properties and devise a nearest point algorithm that requires at most a polynomial number of operations in the dimension of the lattice. This is an improvement over the fastest nearest point algorithms for *random* lattices that require an exponential number of operations.

We then consider polynomial phase signals and their estimation. For polynomial phase signals of order zero the estimation problem is equivalent to a fundamental problem in circular statistics, that of estimating the **mean direction** of a set of circular data. A standard approach to mean direction estimation is to compute the **sample circular mean**. In this thesis we consider an alternative estimator called the **angular least squares estimator**, and we discover that it can be computed rapidly by finding a nearest lattice point in the lattice A^*_n, a problem we have solved. In some scenarios the angular least squares estimator is statistically more accurate and also computationally simpler than the sample circular mean. Therefore the results of this thesis potentially have implications for the wide variety of fields in science, engineering and statistics that currently use the sample circular mean.

For higher order polynomial phase signals the estimation problem is equivalent to **single frequency estimation** (when the order is equal to one) and **polynomial phase estimation** (when the order is greater than one). These problems are common to radar, sonar, astronomy and telecommunications. We find that a very accurate estimator results from computing a nearest point in the lattice $V^*_{n/m}$ and derive the asymptotic properties of this estimator. We show that the estimator is

strongly consistent and describe its central limit theorem. For polynomial phase signals of order greater than one these theoretical results are the first of their kind. While deriving these statistical results, we produce a number of new theorems that describe the **aliasing** properties of polynomial phase signals. These results can be viewed as higher order versions of the Nyquist sampling theorem. Lattice theory is crucial in the description of these aliasing properties. These results will be of great value to engineers, scientists and statisticians studying polynomial phase signals.

Keywords

Lattice theory, circular statistics, polynomial phase signal, mean direction estimation, single frequency estimation, polynomial phase estimation, nearest lattice point problem

Australian and New Zealand standard research classifications

100% 090609 Signal Processing

Contents

List of Figures

List of Algorithms

—Many physical phenomena exhibit some form of periodicity. From the ticking of a clock to the quantisation of energy, they pervade the physical world. This thesis has been motivated by the need to understand the interactions between periodic processes with differing periods and to estimate the periods of infrequently observed periodic processes. That this should lead to the study of integers is not surprising, for the purest representation of a periodic process is the embedding of the integers in the continuum.

I. Vaughan L. Clarkson

1 Introduction

This thesis studies two seemingly unrelated fields, **lattice theory** (a subset of the theory of numbers) and the statistics of circular data, otherwise called **circular statistics**. These fields have found substantial application in their own right. Lattice theory has proved particularly useful in communications engineering, chemistry and cryptography. Circular statistics has a vast array of applications in science, particularly in astronomy, optics, biology, geology, geography and meteorology and also in engineering, particularly in telecommunications and radar.

We will use lattice theory and circular statistics to study **polynomial phase signals** and in particular we consider the problem of estimating the polynomial coefficients of a polynomial phase signal. For polynomial phase signals of order zero this problem is equivalent to that of estimating the **mean direction** of a set of circular data. Mean direction estimation is of fundamental importance in engineering and science. For example, if you listen to the weather report you are often told (an estimate of) the direction of the wind. Obtaining an accurate estimate requires a method for accurately estimating the mean wind direction from a number of observations of the wind direction.

When the polynomial phase signal has order one, the problem is equivalent to a well studied problem called **frequency estimation** and has application to, for example, radar, sonar, telecommunications, astronomy and medicine [Quinn and Hannan, 2001]. For higher order polynomial phase signals the problem is called **polynomial phase estimation**. In radar and sonar applications these signals occur when acquiring radial velocity and acceleration (and higher order motion descriptors) of a target from a reflected signal and also in continuous wave and low probability of intercept radar [Levanon and Mozeson, 2004; Wiley, 1982]. In biology, polynomial phase signals can be used to describe the sounds emitted by bats and dolphins for echo location [Suga et al., 1975; Thomas et al., 2005; Peleg and Friedlander, 1995].

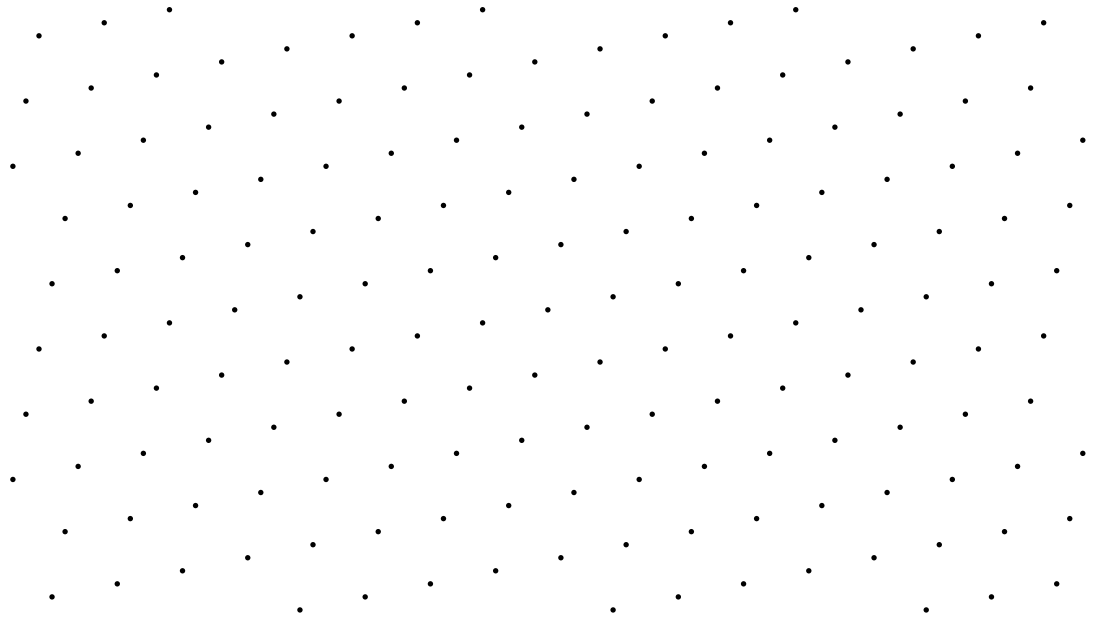


FIGURE 1.1: A lattice in 2 dimensions. The lattice points extend indefinitely in the plane so we have only drawn a subset of the lattice.

Lattice theory studies sets called **lattices** that consist of highly structured arrangements of points. Figure 1.1 is a picture of a part of a lattice in two dimensions. Common questions about a lattice are:

- What is the shortest distance between any two points in the lattice?
- What is the 'density' of the lattice points or equivalently how many lattice points are there per unit volume?
- Given an arbitrary point not necessarily in the lattice, which lattice point is nearest to this point?

The final dot point describes a fundamental problem in lattice theory called the **nearest lattice point problem** that is of significant importance in this thesis.

The study of lattices was originally developed by Minkowski [1910] and Voronoi [1908] as a part of another topic in number theory called **quadratic forms**. Modern treatises on the subject are given by Conway and Sloane [1998] and Martinet [2003]. At least one reason why lattices are mathematically interesting is that they produce excellent solutions to the **sphere packing problem**: the problem of packing as many non-intersecting, equally sized, n-dimensional spheres into the smallest possible volume. Lattices also produce excellent solutions to the **kissing number problem**: the problem of placing as many non-intersecting, equally sized, n-dimensional spheres so that they all just touch, or *kiss*, a central sphere placed at the origin. From a more 'practical' point of view lattices have found substantial application in the field of information theory where they can be used to produce excellent quantisers and codes used for storing and transmitting information [Erez and Zamir, 2004; Erez et al., 2005]. They are also extensively used in cryptography [Goldreich et al.,

1997; Gentry, 2009a,b] and steganography [Cox et al., 2008]. In three dimensions lattices play a central role in crystallography [Hammond, 2001; Belov, 1965]. More recently lattices have found applications in communications systems featuring multiple antennas [Peel et al., 2005; Ryan et al., 2008]. Put simply, your mobile phone, computer and the internet would not work as well without lattices (and this is only considering the telecommunications applications!)

In this context any new results in the field of lattice theory are valuable in their own right. This thesis contains a number of such results. A significant result is the discovery of a fast algorithm to compute a nearest lattice point in a long studied and important lattice called A_n^*. The algorithm requires a linear number of operations in the dimension of the lattice. It was not that long ago that the fastest nearest point algorithm for this lattice required a quadratic number of operations [Conway and Sloane, 1982]. We also find very fast algorithms for another family of lattices called the Coxeter lattices. These algorithms also require a number of operations that is linear in the dimension of the lattice.

We study a new family of lattices related to A_n^* that we call $V_{n/m}^*$. We derive some important properties of this family of lattices and also develop a nearest lattice point algorithm that requires a number of operations that is polynomial in the dimension of the lattice. This is an improvement over the fastest nearest point algorithms for *random* lattices that require an exponential number of operations. The motivation for studying both A_n^* and $V_{n/m}^*$ stems from the fact that these lattices have application to polynomial phase signals and also circular statistics.

Circular statistics aims to describe the nature of data that is measured in angles or 2-dimensional unit vectors or complex numbers on the unit circle. Practical examples of such data are wind direction as measured from a weather vane (commonly used in meteorology) or the direction of flight taken by a bird (commonly used in ornithology). A slightly more subtle example is the *phase* of a periodic signal as commonly observed in radar, sonar, telecommunications, the global positioning system and many other useful devices. Circular data is naturally displayed on a circle as in Figure 1.2. The figure depicts a fundamental problem in circular statistics, that of estimating the **mean direction** of a set of circular data. The figure contains two plots of 100 circular data points. On the right is a plot of data that appears uniformly distributed on the circle. On the left the data appears (roughly) clustered around $\pi/6$. For example, the samples might represent 100 measurements of the direction of the wind taken over the course of a day. If the measurements looked like the plot on the left we would likely conclude that the wind blew in a north easterly direction that day. What should we conclude if the measurements looked like the plot on the right? We consider these problems in Chapters 5 and 6 and provide two definitions of the mean direction, the **circular mean**, that is common in the literature, and the **unwrapped mean**, that although uncommon, is of great utility in this thesis. We find that both of these means, although not equal in general, conform well with our intuitive notion of mean direction.

We describe methods for estimating these means from a set of circular data and derive the statistical properties of the estimators. The circular mean can be estimated by averaging a set of complex numbers, an approach that is common in the literature. We find that the unwrapped mean can be estimated by computing

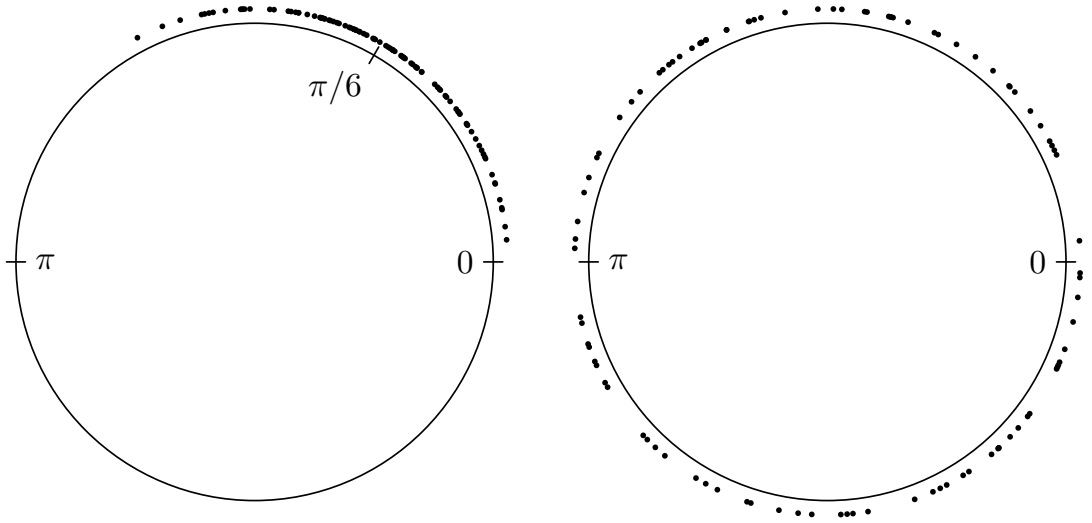


FIGURE 1.2: Plots of 100 circular data points. On the right is a plot of data that appears uniformly distributed on the circle. On the left the data appears (roughly) clustered around $\pi/6$. For example, the samples might represent 100 measurements of the direction of the wind taken over the course of a day. If the measurements looked like the plot on the left we would likely conclude that the wind blew in a north easterly direction that day.

a nearest point in the lattice A_n^*. As we found a fast algorithm to achieve this, the unwrapped mean can be estimated very computationally efficiently.

The problem of estimating the mean direction is equivalent to estimating the single coefficient of a polynomial phase signal of order zero. For higher order polynomial phase signals an excellent estimator arises from finding a nearest point in the lattice $V_{n/m}^*$. We call the resulting estimator the **angular least squares estimator**. Any estimator of polynomial phase signals must take account of the effect of **aliasing** that occurs when the signals are sampled uniformly. For polynomial phase signals of order one the aliasing effect corresponds with the **Nyquist sampling theorem** and is well understood. We completely describe these aliasing properties for polynomial phase signals of any order. We find that a number of estimators in the literature, particularly those based on **polynomial phase transforms** fail to properly account for aliasing. This adversely affects their statistical performance. However, the angular least squares estimator does not suffer from this. This is shown, both by Monte-Carlo simulation, and by theoretically deriving the estimators asymptotic statistical properties. For polynomial phase signals of order greater than one these theoretical results are the first of their kind.

1.1 Organisation of this thesis

The thesis is divided into three parts. Part I concentrates on lattice theory, and in particular, some important lattices called A_n, A_n^* and $V_{n/m}$ and $V_{n/m}^*$. Part II concentrates on circular statistics, with a focus on the definition and estimation of the **mean direction** of a set of circular data. Part III focuses on polynomial phase signals.

In Chapter 2 we describe some introductory concepts from lattice theory. Most of the material is derived from two books, *Sphere Packings, Lattices and Groups* by Conway and Sloane [1998] and *Perfect Lattices in Euclidean Spaces* by Martinet [2003]. Our focus is on **tessellating regions**, the **Voronoi cell**, the **nearest lattice point problem**, **dual lattices**, **sublattices**, **quotient groups** and also the properties of lattices generated by intersection with or projection into a subspace. These will be the most useful concepts for describing the lattices A_n, A_n^*, $V_{n/m}$ and $V_{n/m}^*$. This focus is different from *Sphere Packings, Lattices and Groups* which largely focuses on the **packing** and **covering** properties of lattices, and is also different from *Perfect Lattices in Euclidean Spaces* which focuses on a lattice property called **perfection**.

In Chapter 3 we consider the lattices A_n, A_n^* and also a related family called the **Coxeter lattices**. We describe a number of properties of these lattices and then develop fast nearest lattice point algorithms. The algorithms require only a linear number of operations in the dimension of the lattice n and exploit some peculiar properties of the Voronoi cell of these lattices. These new algorithms are the fastest known and, up to order, fastest possible.

In Chapter 4 we describe some of the properties of the $V_{n/m}$ and $V_{n/m}^*$ lattices. We find that $V_{n/m}$ can be constructed by taking the intersection of the **integer lattice** with a subspace of dimension $m+1$ and that the lattice $V_{n/m}^*$ can be constructed by projecting the integer lattice into this subspace. We provide a convenient way to enumerate the elements in a special finite group called the **dual quotient group** and this leads to a nearest lattice point algorithm for $V_{n/m}^*$ that requires a number of operations that is polynomial in the dimension n. This is an improvement over the fastest nearest point algorithms for *random* lattices that require an exponential number of operations.

In Chapter 5 we give a brief overview of circular statistics. We describe **circular random variables** and their **probability density functions**. We show how the standard definition of the mean in terms of the expected value does not map well to our intuitive notion of **mean direction** and to solve this we consider two different definitions, the **circular mean** and the **unwrapped mean**. We find that both of these means map well to our intuitive notion of mean direction. The circular mean and the unwrapped mean are not always equal and, in fact for some probability distributions they do not even exist, but for many useful distributions they always exist and are equal. We call such distributions **unimean** and we describe a set of criteria that guarantee a distribution is unimean. We then consider a number of popular circular distributions from the literature, the von Mises distribution, the wrapped normal distribution, the wrapped uniform distribution and the projected normal distribution and we describe some conditions under which these are unimean. These distributions are used for modeling noise processes in Chapters 6, 9 and 10.

In Chapter 6 we consider approaches to estimating the mean direction from a number, say N, of observations of a circular random variable. This problem is of fundamental importance in circular statistics and estimates of the mean direction are used in a wide variety of applications in engineering and science. The first estimator we describe is the **sample circular mean** that is common in the literature.

Following the work of Quinn [2010] we describe how the sample circular mean converges to the circular mean of a random variable as the number of observations increases. The sample circular mean can be computed efficiently by averaging N complex numbers. We then describe an alternative estimator called the **angular least squares estimator** and we show that it converges to the unwrapped mean of a circular random variable as the number of observations increases. The angular least squares estimator can be computed very efficiently by finding a nearest point in the lattice A_n^*, a problem we solve in Chapter 3. In the remainder of the chapter we compare the performance of the sample circular mean and the angular least squares estimators for a number of problems in signal processing and communications engineering, these being **phase estimation**, **noncoherent detection** and **delay estimation**. We find that in some scenarios the sample circular mean is more accurate, and in other scenarios angular least squares is more accurate. Both estimators require a linear number of operations in the number of observations N, however, a potentially large computational advantage of the angular least squares estimator is that it can avoid performing trigonometric operations and we discuss this in Section 6.7.

Estimating the mean direction of a circular random variable is equivalent to estimating the phase of a polynomial phase signal of order zero, otherwise called a **constant phase signal**. In Chapters 7, 8, 9 and 10 we generalise this concept to polynomial phase signals of arbitrary order. Before we describe methods for estimating polynomial phase signals we consider some interesting phenomena that occur when polynomial phase signals are sampled in Chapters 7. It turns out that two (or more) distinct polynomial phase signals can sometimes *take exactly the same values* when they are sampled. We call such signals **aliases** and we completely describe how the aliasing occurs using some ideas from lattice theory. For polynomial phase signals of order one, this aliasing effect it equivalent to the Nyquist criterion. Ängeby [2000a] and Abatzoglou [1986] have described the effect of this aliasing for polynomial phase signals of order two, but here we describe the effect for polynomial phase signal of any order. In practice, we typically want to estimate the coefficients of a polynomial phase signal from a set of observations and an understanding of these aliasing properties is required to ensure the **identifiability** of any estimator of the coefficients. We describe a convenient method of ensuring identifiability by restricting the polynomial coefficients to a tessellating region of particular lattice. We call the chosen region the **identifiable region**. We show how aliasing can be resolved by computing a **nearest point** in this lattice.

We then consider the problem of estimating the coefficients of a polynomial phase signal from a number, say N, of observations. In Chapter 8 we consider a direct analogue of the angular least squares estimator that we used for mean direction estimation. We show how this estimator can be computed by finding a nearest lattice point in the lattice $V_{n/m}^*$. Using the nearest point algorithm we described in Chapter 4 we find that the angular least squares estimator can be computed in a number of operations that is polynomial in N. We then derive the statistical asymptotic properties of this estimator showing that it is **strongly consistent** and that it obeys a **central limit theorem**. For polynomial phase signals of order larger than one, these are the first results of their kind.

In Chapter 9 we consider the special case of estimating the two coefficients of a polynomial signal of order one. This is equivalent to a well studied problem called **frequency estimation** and has application to, for example, radar, sonar, telecommunications, astronomy and medicine [Quinn and Hannan, 2001]. We consider three estimators that exist in the literature, the **periodogram estimator**, the **Quinn-Fernandes estimator** and **Kay's unwrapping estimator**. We also discuss the angular least squares estimator that can be computed by finding a nearest point in the lattice $V^*_{N-1/1}$. We could use the exact nearest point algorithm described in Chapter 4 but we find that it is quite slow. Instead we describe a simple approximate nearest point algorithm that is much faster, and for frequency estimation, has almost identical statistical performance to the exact nearest point algorithm. We compare the estimators by Monte-Carlo simulation and it is found that the angular least squares estimator is very accurate, but is computationally more expensive than the other estimators.

In Chapter 10 we consider estimating the $m+1$ coefficients of a polynomial phase signal of order m. We first consider the standard **least squares estimator** and describe a practical method for its computation. The least squares estimator is very computationally intensive and for this reason many authors have considered methods to reduce the computational complexity. We consider two such estimators from the literature, the **discrete polynomial phase transform** [Peleg and Friedlander, 1995] and **Kitchen's unwrapping estimator** [Kitchen, 1994]. It turns out that both of these estimators only work correctly for coefficients in a subset of the identifiable region. The problem is particularly acute with the discrete polynomial phase transform that only works for coefficients in a *very* small subset of the identifiable region, and moreover, the subset shrinks rapidly as the number of observations N increases. We consider how this problem might be overcome by increasing the rate at which observations are acquired (the **sample rate**), but we show that increasing the sample rate comes with inevitable statistical penalties.

We also consider the angular least squares estimator of the polynomial coefficients. We could use the exact nearest point algorithm for the lattice $V^*_{n/m}$ that we described in Chapter 4 to compute the estimator in a number of operations that is polynomial in N but, we find that this is very slow in practice. We instead consider alternative algorithms, the **sphere decoder**, the **K-best algorithm** and **Babai's nearest plane algorithm**, to compute or approximate the nearest point. We find these approaches are feasible, but they are still computationally more expensive than the discrete polynomial phase transform and Kitchen's unwrapping estimator.

In Section 10.5 we use Monte-Carlo simulation to compare the performance of the estimators in practice. We find that the angular least squares estimators and the least squares estimators are both very accurate. Both of these estimators work correctly for polynomial coefficients anywhere in the identifiable region. Kitchen's unwrapping estimator and the DPT are less accurate. The DPT suffers from the fact that it operates very poorly on a large range of coefficients inside the identifiable region. We also discuss some computational properties of the various estimators.

1.2 Who should read this thesis and how

This thesis is written with two audiences in mind. Firstly, number theorists who are interested in lattice theory, in particular the lattice A_n^*, and secondly, engineers and statisticians who are interested in circular statistics, particularly regarding mean direction estimation, frequency estimation and polynomial phase estimation.

To accommodate both audiences the thesis is separated into three parts, the first part introduces lattice theory (Chapter 2), the lattices A_n and A_n^* (Chapter 3), and the related lattices $V_{n/m}$ and $V_{n/m}^*$ (Chapter 4). In these chapters we make little reference to circular statistics and polynomial phase signals so that the reader more interested in lattice theory can read these chapters unhindered. In the second and third parts we consider the statistical problems of mean direction estimation, frequency estimation and polynomial phase estimation. Here, it is sometimes necessary to refer to the material on lattice theory from the previous chapters. We have taken care to reference the required material in such a way that the reader interested only in circular statistics and polynomial phase signals can begin reading at Part II (page 67) and need refer to the lattice theory material only sparingly.

This thesis assumes that the reader has no knowledge of lattice theory and we will provide all of the required preliminary results about lattices in Chapter 2. However, in Parts II and III some knowledge of statistics and in particular estimation theory is assumed. For example, familiarity is assumed with the concept of a **random variable**, the **Cràmer Rao lower bound** and also the various forms of statistical convergence, i.e., **convergence in distribution**, **convergence in probability** and **convergence almost surely**. For an introduction to these topics the reader is referred to an introductory text on estimation theory, for example Sage and Melsa [1971] or van der Vaart [1998].

1.3 Original contributions

The major original contributions can be summarised as follows:

- In Chapter 3 we derive linear-time nearest point algorithms for A_n^* and the Coxeter lattices A_n^m. These are the fastest known nearest point algorithms for these lattices and also, up to order of complexity, the fastest possible. These results are also available in McKilliam et al. [2008a,b, 2010c].

- In Chapter 4 we explore the $V_{n/m}$ and $V_{n/m}^*$ lattices. These lattices do not appear elsewhere in the literature and we describe a number of their properties in Chapter 4. We find nearest point algorithms for $V_{n/m}^*$ that require at most a polynomial number of operations in the dimension of the lattice n.

- In Chapter 6 we describe the angular least squares estimator of the **mean direction** of a circular random variable. Using the fast nearest lattice point algorithm for the lattice A_n^* this estimator can be computed in linear-time. We have applied this estimator to the problem of noncoherent detection of phase-shift-keyed digital communications symbols and also delay estimation

from sparse, noisy timing data. Related work was presented in McKilliam et al. [2009b] and McKilliam and Clarkson [2008].

- A complete description of the effect of aliasing that occurs in polynomial phase signals is described in Chapter 7. This material is published in McKilliam and Clarkson [2009] but the presentation here is more thorough. Ängeby [2000a] and Abatzoglou [1986] have independently described the effect of this aliasing for polynomial phase signals of order two, but here we describe the effect for polynomial phase signal of any order.

- In Chapter 8 we describe a new approach to polynomial phase estimation, the angular least squares estimator, that is based on finding a nearest lattice point in $V^*_{n/m}$. We also provide a thorough analysis of the statistical properties of this estimator. For the case of polynomial phase signals of order one (frequency estimation), this material has been published in McKilliam et al. [2010a]. We extend these results to polynomial phase signals of arbitrary order in Chapter 8.

1.4 Notation

We write vectors and matrices in bold, with vectors in lower case and matrices in upper case. So, $\mathbf{x}$ is a vector and $\mathbf{X}$ is a matrix. The ith element in a vector is denoted by a subscript, as in x_i, and the element in the ith row and jth column of a matrix, say $\mathbf{X}$ is given by $x_{i,j}$. The transpose or Hermitian transpose of a vector or matrix is indicated by superscript $\dagger$, i.e. $\mathbf{x}^\dagger$. We denote by $\mathbf{1}$ a column vector of ones, by $\mathbf{0}$ a column vector of zeros and by $\mathbf{e}_i$ a column vector of zeros with a one in the ith position. The inner (or scalar) product of the vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathbf{x} \cdot \mathbf{y}$ and the mean of a vector is denote by a bar, that is $\bar{y} = \frac{\mathbf{y}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}$ denotes the mean of the elements in $\mathbf{y}$. The determinant of a matrix $\mathbf{M}$ is denoted $\det(\mathbf{M})$.

We use $\lfloor x \rfloor$ to denote the largest integer less than or equal to x (the *floor* of x) and $\lceil x \rceil$ to denote the smallest integer greater than or equal to x (the *ceiling* of x), and we use $\lceil \cdot \rfloor$ to denote rounding to the nearest integer. The direction of rounding for half-integers is not important so long as it's consistent. In this thesis we have chosen to round up half-integers. Also, $\langle x \rangle = x - \lceil x \rfloor$ denotes the *centered* fractional part of x. For a vector $\mathbf{x}$ the functions $\lfloor \mathbf{x} \rfloor$, $\lceil \mathbf{x} \rceil$, $\lceil \mathbf{x} \rfloor$ and $\langle \mathbf{x} \rangle$ all operate element wise.

We use capital letters such as X, Y, or Z to denote random variables. When describing estimators we use a tilde, as in $\tilde{\mu}$ to denote the *true* value of a parameter and a hat, as in $\hat{\mu}$, to denote the estimate of $\tilde{\mu}$. A common notation in the statistics literature is to use a subscript 0, as in μ_0, to denote the *true* value of a parameter, but we do not use this notation because it is too easily confused with the zeroth element in the vector $\boldsymbol{\mu}$.

We use order notation $O(\cdot)$ and $o(\cdot)$ is the standard way and we use $O_P(\cdot)$ and $o_P(\cdot)$ to denote convergence in probability. For example, $X_N = o_p(N^{-1})$ means that for any $\epsilon > 0$ the probability $\text{Prob}(|NX_N| > \epsilon)$ converges to zero as $N \to \infty$ and $X_N = O_p(N^{-1})$ means that NX_N converges in distribution as $N \to \infty$.

Finally, we denote sets using a capital and the number of elements in the set S (the cardinality) is denoted by $|S|$.

Part I

Lattice theory

—*Lattices are everywhere.*
Ram Zamir

2
An introduction to lattice theory

This chapter introduces **lattice theory**. The primary purpose of this chapter is to give sufficient background so that the properties of the particular lattices A_n^*, A_n and A_n^m and the lattices $V_{n/m}^*$, $V_{n/m}$ and $V_{n/m}^\perp$ can be investigated in Chapters 3 and 4. We will use these lattice to estimate and analyse polynomial phase signals in Parts II and III of this thesis.

We begin by defining some fundamental properties of a lattice such as the **generator matrix** and the **fundamental parallelepiped** in Section 2.1. Lattices naturally give rise to tessellations of n-dimensional space and we consider this in Section 2.2. We consider a simple type of *rectangular* tessellation that exists for all lattices. We will have use of these rectangular tessellations when describing the **aliasing** of polynomial phase signals in Chapters 7.

In Section 2.3 we consider a particularly important tessellation called the **Voronoi cell** that describes the region of space that is closest to a lattice point. We then consider some traditional problems of lattice theory that are related to the Voronoi cell: the problems of **packing**, **covering** and the **kissing number**. In Section 2.4 we consider the properties of **sublattices** (lattices that are a subset of another lattice) and in Section 2.5 we consider an important lattice called the **dual lattice**. The lattice A_n^* is the dual lattice of A_n and the lattice $V_{n/m}^*$ is the dual lattice of $V_{n/m}$ so the properties of dual lattices will be very important in later chapters.

In Section 2.6 we consider lattices that are generated by *intersection* with a subspace and *projection* into a subspace. We find that lattices generated by intersection and projection have some intriguing relationships with dual lattices and also with sublattices. These relationships will be useful for describing the properties of the lattices A_n^*, A_n and A_n^m and the lattices $V_{n/m}^*$, $V_{n/m}$ and $V_{n/m}^\perp$ in Chapters 3 and 4.

In the final section of this chapter we consider a fundamental problem in lattice theory called the **nearest lattice point problem**. This problem has found substantial application in information theory, integer programming, cryptography and a host of other problems. We give a brief account of these applications and also

consider some of the standard algorithms that exist for computing a nearest lattice point. In Chapter 5 we discover very fast algorithms for finding a nearest lattice point in A_n^*, A_n and A_n^m. These algorithms make use of some special properties of the Voronoi cell of A_n^*, A_n and A_n^m. In Chapter 4 we will describe an algorithm that can find a nearest point in the lattice $V_{n/m}^*$. This algorithm makes use of some general properties of dual lattices and also some results about lattices generated by intersections and projections that we describe in Section 2.6. We will use these nearest point algorithms to solve estimation problems involving polynomial phase signals in Parts II and III.

The material in this chapter is mostly derived from two books, *Sphere packings, lattices and groups* by Conway and Sloane [1998] and *Perfect lattices in Euclidean spaces* by Martinet [2003]. Some material is also taken from the book *Lattice points* by Erdös et al. [1989]. Although almost all of the material from this introduction can be found in these books, this introduction approaches lattice theory from a somewhat different angle and with a different purpose. *Sphere packings, lattices and groups* is probably the most thorough account of lattice theory available, but, as the title suggests, it places much emphasis on the **packing** properties of lattices and also on traditional problems such as the covering problem and the **kissing number problem**. On the other hand *Perfect lattices in Euclidean spaces* places significant emphasis on a lattice property called **perfection** which we will not talk about in this thesis. This thesis has little use for many of these traditional problems and therefore this introduction will only touch on these briefly. Instead we are more interested in the properties of sublattices, dual lattices, the nearest lattice point problem and also the properties of lattices generated by intersections and projections. Because these topics are our primary concern this allows our introduction to be substantially more focused than these books are. This chapter is not claimed to contain any original material. All of the results presented here are known. However, in many places we have found it easier to rederive the simple results we need, rather than map them to a specific (and likely less well fitting) result in the literature.

Another property of the books is a general assumption that the reader is reasonably well versed in group theory. This is not surprising as lattices are infinite discrete abelian groups and many concepts from lattice theory are naturally and elegantly described by concepts well known to group theorists. However, for the reader not knowledgeable in group theory these book can sometimes be difficult to read. Because at least some of the intended audience of this thesis is engineers and statisticians we have made no assumption about the readers familiarity with group theory and all of the concepts needed will be described in an elementary fashion. We do make use of a number of ideas from group theory in this thesis, particularly when describing sublattices in Section 2.4, but we have attempted to introduce these concepts in a gentle manner and to give intuitive and visual representations as much as possible. It is hoped that this effort has been worthwhile and that even the reader with no prior knowledge of groups will leave this introduction with an intuitive understanding of the few simple concepts from group theory that we require in this thesis.

2.1 Basic properties of lattices

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

$$\Lambda = \{\mathbf{x} = \mathbf{B}\mathbf{u} \mid \mathbf{u} \in \mathbb{Z}^n\}$$

where $\mathbf{B}$ is an $m \times n$ matrix of rank n called the **generator matrix** or **basis matrix** or simply generator or basis. If the generator is square, i.e. $m = n$, then the lattice points span $\mathbb{R}^n$ and we say that the lattice is **full rank**. If $\mathbf{B}$ has more rows than columns, i.e. $m > n$, then the lattice points lie in a n-dimensional subspace of $\mathbb{R}^m$. The set of integers $\mathbb{Z}^n$ is also a lattice (a generator is the identity matrix) and we call this the **integer lattice**. We will often abbreviate the definition of a lattice above to

$$\Lambda = \mathbf{B}\mathbb{Z}^n$$

meaning that the lattice Λ contains the points from the integer lattice transformed by the generator matrix $\mathbf{B}$.

The generator matrix for a lattice is not unique. Let $\mathbf{M}$ be an $n \times n$ matrix with integer elements such that $\det \mathbf{M} = \pm 1$. Then both $\mathbf{B}$ and $\mathbf{BM}$ are generator matrices for Λ and $\mathbf{M}$ is called a **unimodular matrix**. Lattices are considered equivalent under scaling, rotation and reflection so a lattice Λ and a lattice $\hat{\Lambda}$ are called equivalent, or *isomorphic*, if and only if

$$\Lambda = \alpha \mathbf{R} \hat{\Lambda}$$

where $\alpha > 0$ is real and $\mathbf{R}$ is an orthogonal matrix. We write $\Lambda \simeq \hat{\Lambda}$ to denote lattice isomorphism. If a generator matrix $\mathbf{B}$ is not square then we can always find a square matrix that generates an isomorphic lattice to $\mathbf{B}\mathbb{Z}^n$ that is full rank. On a number of occasions in this chapter it will be convenient to assume that the generator matrix is square and it is important to realise that this can be done without any loss of generality.

The column vectors of a generator matrix are called **basis vectors** for the lattice. A **fundamental parallelepiped** of the lattice is the parallelepiped constructed from any set of basis vectors. As the generator matrix is not unique, neither is the fundamental parallelepiped. Two examples of fundamental parallelepiped for a lattice in $\mathbb{R}^2$ with generator matrix

$$\begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \tag{2.1.1}$$

are shown in Figure 2.1. We will use this lattice for many of the examples throughout this chapter.

Given a lattice Λ with generator matrix $\mathbf{B}$ the matrix

$$\mathbf{A} = \mathbf{B}^\dagger \mathbf{B} \tag{2.1.2}$$

is called the **Gram matrix**, where $\mathbf{B}^\dagger$ denotes the transpose of $\mathbf{B}$. The **determinant** of Λ, denoted $\det \Lambda$, is defined as the determinant of the Gram matrix.

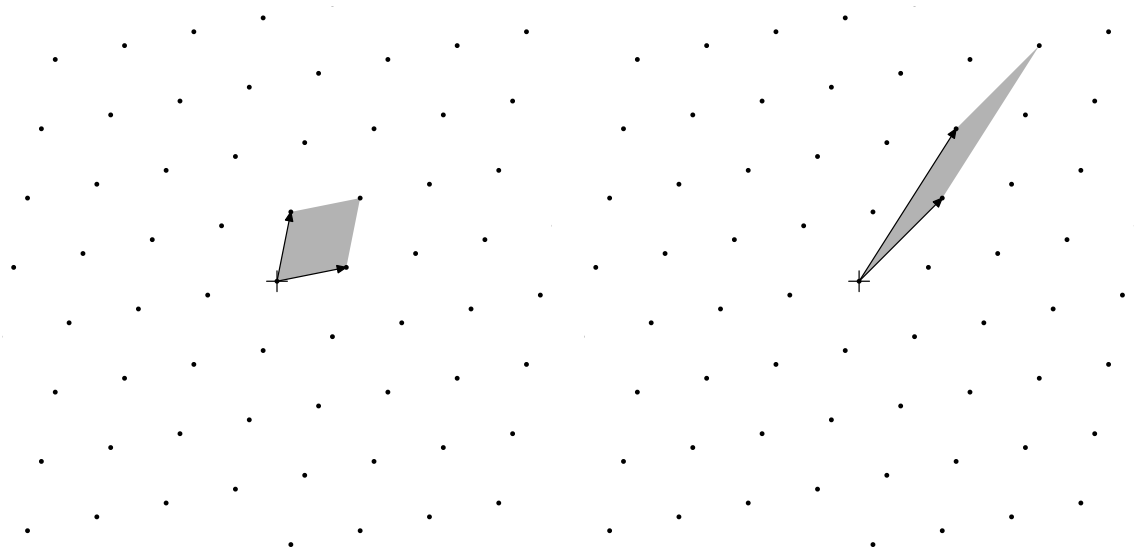


FIGURE 2.1: Two examples of fundamental parallelepiped (shaded region) for a lattice with generator matrix from (2.1.1). For the left plot the basis vectors are $[1, 0.2]^\dagger$ and $[0.2, 1]^\dagger$. For the right plot the basis vectors are $[1.2, 1.2]^\dagger$ and $[1.4, 2.2]^\dagger$. The edges marked by the black arrows are closed, and the other edges are open.

The fundamental parallelepiped of an n-dimensional lattice is n-dimensional and has n-volume given by

$$\sqrt{\det \mathbf{A}} = \sqrt{\det \Lambda}. \tag{2.1.3}$$

A lattice is called **integral** if and only if the inner products of its lattice points are integers. That is, a lattice is integral if the inner product $\mathbf{x} \cdot \mathbf{y}$ is an integer for every pair of lattice points $\mathbf{x}$ and $\mathbf{y}$ from Λ. Equivalently, a lattice is integral if its Gram matrix is integral (i.e. the matrix has only integer entries). A lattice is called **unimodular** if its Gram matrix is unimodular. A unimodular lattice has determinant 1 and the volume of its fundamental parallelepiped is also 1. It is easy to see that the integer lattice $\mathbb{Z}^n$ is unimodular.

2.2 Tessellating regions

Given some subset $S \subset \mathbb{R}^m$ and some point $\mathbf{x} \in \mathbb{R}^m$ and a transformation $T : \mathbb{R}^m \mapsto \mathbb{R}^m$ we will use the notation $TS + \mathbf{x}$ to denote the set of points $\{Ts + \mathbf{x} \mid s \in S\}$. We say that the set R *tessellates on the lattice* Λ if and only if the intersection of two copies of R translated by distinct lattice points is empty, and also the union of R translated over all lattice points is all of $\mathbb{R}^m$. That is,

$$(R + \mathbf{x}) \cap (R + \mathbf{y}) = \emptyset \ \ \forall \mathbf{x}, \mathbf{y} \in \Lambda, \, \mathbf{x} \neq \mathbf{y} \tag{2.2.1}$$

$$\bigcup_{\mathbf{x} \in \Lambda} (R + \mathbf{x}) = \mathbb{R}^m. \tag{2.2.2}$$

We call a set that tessellates on the lattice a **tessellating region**. Two examples of tessellating regions are given in Figures 2.2 and 2.3. Figure 2.2 shows a disconnected

tessellating region constructed using two triangles. Figure 2.3 shows a connected tessellating region constructed using a rectangle. Rectangular tessellating regions are particularly useful in this thesis and we will describe them in Proposition 2.1.

We desire a fundamental parallelepiped to tessellate on Λ. To ensure this we require to define half of the faces of the parallelepiped to be closed and half to be open. There are many ways to do this, but here we will define the faces that intersect with the origin to be closed and the remaining faces to be open (see Figure 2.1). If the lattice has full rank then the volume of the tesselating region is equal to that of the fundamental parallelepiped, i.e. $\sqrt{\det \Lambda}$. If the lattice is not full rank, but is of dimension $n < m$, then the tesselating regions have infinite volume, but if we take the intersection of the tessellating region with the n-dimensional subspace spanned by the lattice then the n-volume of this intersection is also $\sqrt{\det \Lambda}$.

Theorem 2.1. *Let the set R tesselate on the lattice Λ. Then there is precisely one lattice point from Λ in R.*

Proof. There is at least one lattice point in R, otherwise $\{R + \mathbf{x} \mid \mathbf{x} \in \Lambda\}$ would not contain any points from L, violating (2.2.2). If there is more than one lattice point in R, say $\mathbf{x}$ and $\mathbf{y}$, then $(R + \mathbf{x}) \cap (R + \mathbf{y}) \neq \emptyset$, violating (2.2.1). □

In view of this theorem we see that $\sqrt{\det \Lambda}$ is the average number of lattice points per unit volume in $\mathbb{R}^n$. An interesting account of the following corollary in the context of structural chemistry is given by Belov [1965].

Corollary 2.1. *No lattice point lies in the interior of a fundamental parallelepiped.*

Proof. Let F be a fundamental parallelepiped. The lattice point at the origin lies on the boundary of F and is contained in F. Due to Theorem 2.1 there can be no other lattice point in F and therefore no lattice point lies in the interior of F. □

Tessellating regions can have very complicated structures in general, far more complicated than a fundamental parallelepiped. They do not have to be polytopes and they do not even need to be connected (see for example Figure 2.2). In the next section we consider a very important tessellating region called the **Voronoi cell** that, for some lattices, has a very complicated structure. However, before we describe the Voronoi cell we will describe a particularly simple *rectangular* tessellating region that exists for any lattice. In fact each basis for a lattice describes such a rectangular tessellating region as we shall see in the next proposition.

Proposition 2.1. *Let Λ be an n dimensional lattice and $\mathbf{B}$ be a generator matrix for Λ. Let $\mathbf{B} = \mathbf{QR}$ where $\mathbf{Q}$ is orthonormal and $\mathbf{R}$ is upper triangular with elements $r_{i,j}$. Then the rectangular prism $\mathbf{Q}P$ where*

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

tessellates on Λ.

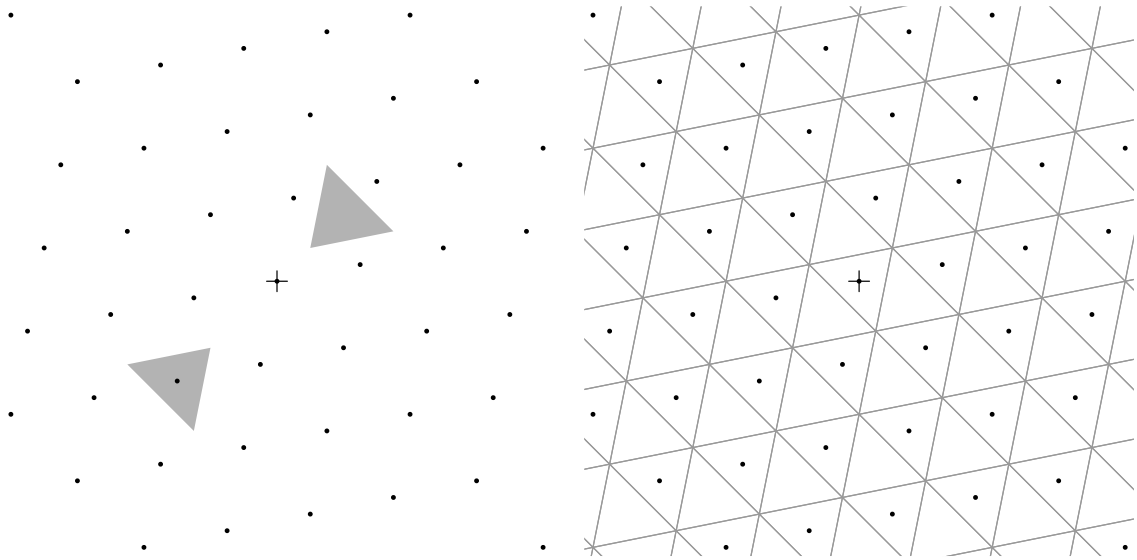


FIGURE 2.2: A disconnected tessellating region where Λ has generator matrix from (2.1.1).

Proof. Let L be the lattice generated by $\mathbf{R}$. It is sufficient to show that P tessellates on L as this clearly implies that $\mathbf{Q}P$ tessellates on $\Lambda = \mathbf{Q}L$. We need to prove that (2.2.1) and (2.2.2) hold for the region P and lattice L. We will prove (2.2.1) first. Let $\mathbf{x}$ and $\mathbf{y}$ be distinct points in L. We may write

$$x_i = \sum_{j=i}^{n} r_{i,j}u_j \qquad \text{and} \qquad y_i = \sum_{j=i}^{n} r_{i,j}v_j$$

where u_i and v_i are integers and $u_i \neq v_i$ for at least one i. Let k be the largest integer such that $u_k \neq v_k$ and let

$$t = \sum_{j=k+1}^{n} r_{k,j}u_j = \sum_{j=k+1}^{n} r_{k,j}v_j.$$

Then

$$\left(\left[-\frac{r_{k,k}}{2}, \frac{r_{k,k}}{2}\right) + r_{k,k}u_k + t\right) \cap \left(\left[-\frac{r_{k,k}}{2}, \frac{r_{k,k}}{2}\right) + r_{k,k}v_k + t\right) = \emptyset$$

because $|u_k - v_k| \geq 1$ and (2.2.1) follows immediately.

We prove (2.2.2) by contradiction. Assume that (2.2.2) is false. Then there exists a point $\mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{z} \notin (P + \mathbf{x})$ for all $\mathbf{x} \in L$. Then for some element of $\mathbf{z}$, say z_k, it is true that

$$z_k \notin \left[-\frac{r_{k,k}}{2}, \frac{r_{k,k}}{2}\right) + r_{k,k}u_k + t$$

for any integer u_k where $t = \sum_{j=k+1}^{n} r_{j,k}u_j$ is a real number. However, setting $u_k = \lceil (x_k - t)/r_{k,k} \rfloor$ immediately yields $z_k \in [-r_{k,k}/2, r_{k,k}/2) + r_{k,k}u_k + t$, a contradiction. □

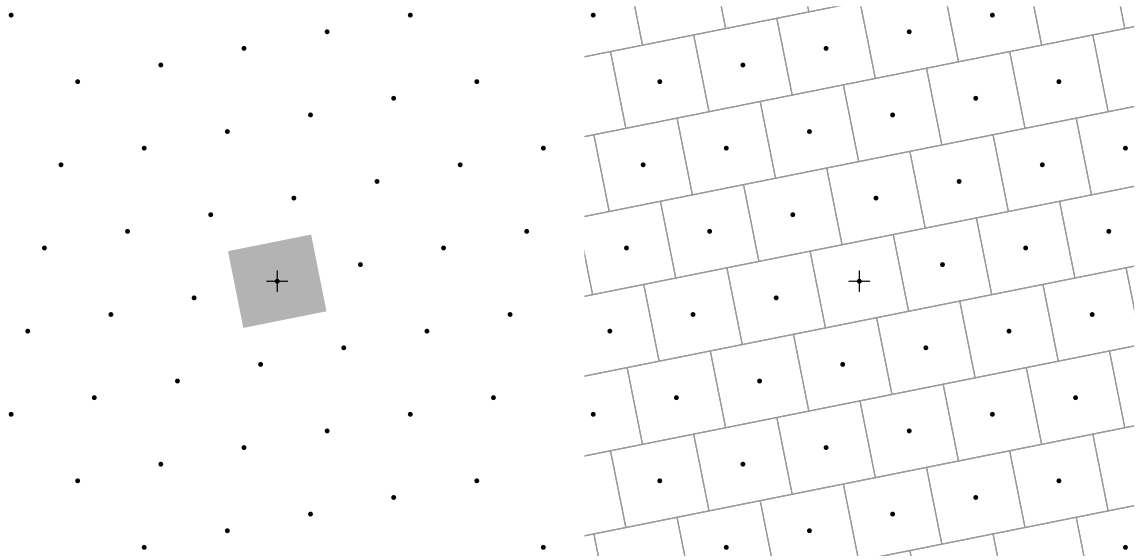


Figure 2.3: Rectangular tessellating region constructed according to Proposition 2.1 where Λ has generator matrix from (2.1.1).

2.3 The Voronoi cell

The (open) **Voronoi cell**, denoted $\mathrm{Vor}(\Lambda)$, of a lattice Λ in $\mathbb{R}^n$ is the subset of $\mathbb{R}^n$ containing all points nearer, with respect to a given norm, the lattice point at the origin than any other lattice point. The Voronoi cell is an n-dimensional convex polytope that is symmetric about the origin. In this thesis we will always assume the Euclidean norm (or 2-norm), and therefore $\mathrm{Vor}(\Lambda)$ contains those points nearest in Euclidean distance to the origin. If $\mathbf{x} \in \Lambda$ it follows that $\mathrm{Vor}(\Lambda) + \mathbf{x}$ is the subset of $\mathbb{R}^n$ that is nearer to $\mathbf{x}$ than any other lattice point in Λ. Figure 2.4 is an example of the Voronoi cell.

Equivalently the Voronoi cell can be defined as the intersection of the half spaces

$$H_{\mathbf{v}} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{v} < \tfrac{1}{2}\mathbf{v} \cdot \mathbf{v}\}$$

for all $\mathbf{v} \in \Lambda \backslash \{\mathbf{0}\}$. It is not necessary to consider *all* $\mathbf{v}$. The minimal set of lattice vectors $\mathcal{R}$ such that $\mathrm{Vor}(\Lambda) = \cap_{\mathbf{v} \in \mathcal{R}} H_{\mathbf{v}}$ is called the set of **Voronoi relevant vectors** or simply **relevant vectors** [Voronoi, 1908]. Figure 2.4 depicts the relevant vectors of the lattice with generator given by (2.1.1). A lattice point in $\mathcal{R}$ is called **relevant**.

So far we have assumed that the faces of the Voronoi cell are open. It is convenient to modify the definition of the Voronoi cell slightly so that it tessellates on the lattice. To ensure this we require that if a face of $\mathrm{Vor}(\Lambda)$ is open, then we define its opposing face to be closed. Specifically, if $\mathbf{x} \in \mathrm{Vor}(\Lambda)$ is on the boundary of $\mathrm{Vor}(\Lambda)$ then $-\mathbf{x} \notin \mathrm{Vor}(\Lambda)$. We wont specifically define which opposing face is open and which is closed. The results in this thesis hold for any choice of open and closed opposing faces. When the lattice has full rank the volume of $\mathrm{Vor}(\Lambda)$ is equal to the volume of a fundamental parallelepiped, i.e. $\mathrm{vol}(\mathrm{Vor}(\Lambda)) = \sqrt{\det \Lambda}$. If the lattice is not fuull rank then the intersection of $\mathrm{Vor}\,\Lambda$ with the n-dimensional subspace spanned by the lattice has n-volume equal to $\sqrt{\det \Lambda}$.

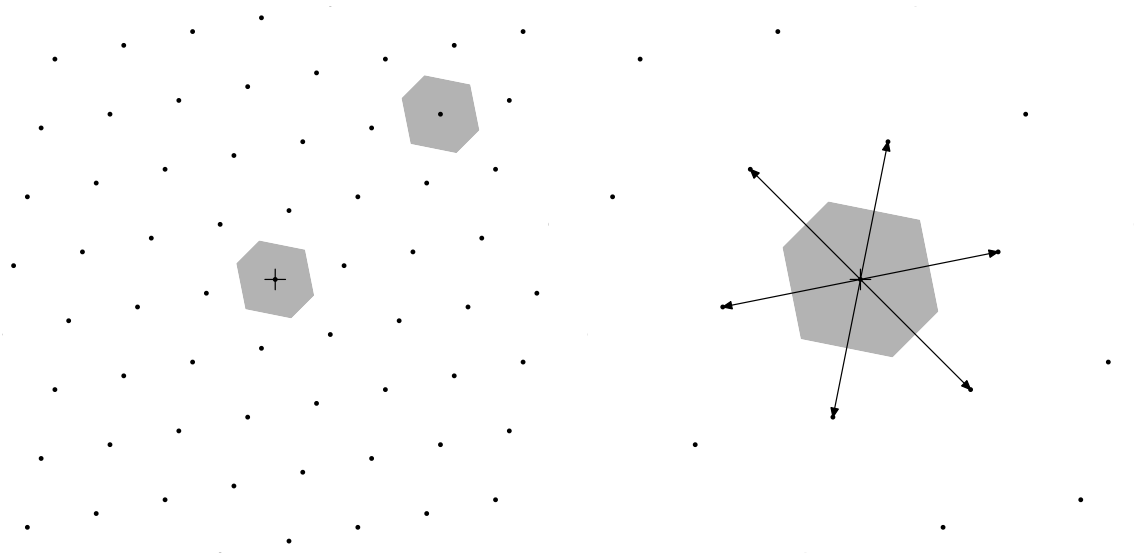


FIGURE 2.4: (Left) The Voronoi cells $\mathrm{Vor}(\Lambda)$ and $\mathrm{Vor}(\Lambda) + [2.4, 2.4]^\dagger$ (shaded) where Λ has generator matrix from (2.1.1). (Right) The Voronoi cell and relevant vectors. The relevant vectors are indicated by arrows.

Remark 2.1. . *Let* $\mathbf{v}$ *be a point in the lattice* Λ *other than the origin. If* $\mathbf{y} \in \mathrm{Vor}(\Lambda)$ *then*

$$\mathbf{y} \cdot \mathbf{v} \leq \tfrac{1}{2}\mathbf{v} \cdot \mathbf{v}.$$

Packing and covering

Two interesting properties of a lattice are its **inradius**, denoted ρ, and its **sphere packing density**, denoted Δ. Imagine a sphere placed around every lattice point such that no two spheres intersect. This is depicted in Figure 2.5. The radius of the spheres is called the **inradius** or the **packing radius** of the lattice. The inradius is half the length of the shortest vector in the lattice. The squared Euclidean norm of this vector is generally called the **norm** of the lattice, so the norm of a lattice is twice the inradius squared. The sphere packing density is the ratio between the volume of all the spheres and the volume of the entire space. This is equivalent to the ratio between the volume of one sphere and the volume of the Voronoi cell. That is,

$$\Delta = \frac{\rho^n S_n}{\sqrt{\det \Lambda}} \tag{2.3.1}$$

where

$$S_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$$

is the volume of the n-dimensional sphere with unit radius and $\Gamma(\cdot)$ is the gamma function. A large portion of lattice theory focuses on finding lattices that yield very dense packings. This is called the **sphere packing problem**.

Another important property of a lattice is its **covering radius**, denoted R. Here we wish to place the smallest possible sphere around each lattice point so that all of

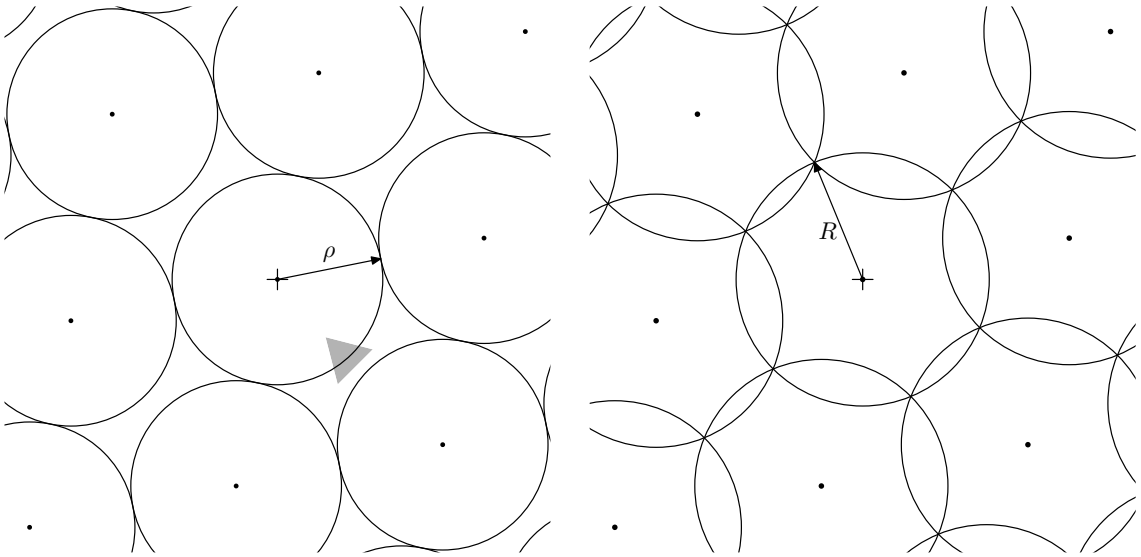


Figure 2.5: A sphere packing and the inradius ρ (left) and a sphere covering and covering radius R (right) for the lattice with generator matrix from (2.1.1).

the space is covered. The radius of the sphere is the covering radius of the lattice. This is depicted in Figure 2.5. The covering radius is the distance to the vertex of the Voronoi cell that is furthest from the origin. A lattice property related to the covering radius is the **thickness**, denoted Θ. This is the ratio between the volume of the covering sphere and the volume of the Voronoi cell

$$\Theta = \frac{R^n S_n}{\sqrt{\det \Lambda}}. \tag{2.3.2}$$

A lattice with a small value of Θ is referred to as **thin**.

Minimal vectors and the kissing number

The **minimal vectors** of a lattice are all those lattice points with length equal to twice the inradius. The minimal vectors are sometimes also called **short vectors**. The number of minimal vectors is called the **kissing number**. It is not hard to see that the angle between any two minimal vectors from a lattice must be greater than $\frac{\pi}{3}$ (otherwise there would exist a lattice point of shorter length than twice the inradius).

2.4 Sublattices and quotient groups

Let Λ be a lattice. A **sublattice** of Λ, denoted $\Lambda' \subseteq \Lambda$, is a subset of points from Λ that also forms a lattice. A simple example of a sublattice is $k\Lambda$ for any integer k. More complicated examples exist where Λ and Λ' are not isomorphic. For example consider the rectangular lattice with basis vectors $[1, 0]^\dagger$ and $[0, 2]^\dagger$ which is clearly a sublattice of $\mathbb{Z}^2$. If $\Lambda' \subseteq \Lambda$ then an obvious property of their Voronoi cells is

$$\mathrm{Vor}(\Lambda) \subseteq \mathrm{Vor}(\Lambda'). \tag{2.4.1}$$

A lattice Λ can be considered as a discrete abelian group[1] and a sublattice Λ' can be considered as a (normal) subgroup of Λ. We may then define the **quotient group** Λ/Λ' which operates on the **cosets** given by translates of the lattice Λ' by lattice points in Λ. Figure 2.6 is a pictographic depiction of the cosets. In the figure the lattice is given by the dots and the sublattice is given by the circles. There are five coset, each depicted by the five different *shapes*, the circle, triangle, square, hexagon and star. For example, one coset is all of the hexagons and another coset is all of the triangles.

Let C be a set, of smallest possible size, containing points from Λ such that the union of translates of the sublattice Λ' by the points in C is equal to the superlattice Λ, that is,

$$\Lambda = \bigcup_{\mathbf{x}\in C} \mathbf{x} + \Lambda'. \tag{2.4.2}$$

The elements in C are called **coset representatives**. For example, from Figure 2.6, any set containing one lattice point of each shape is a set of coset representatives. The number of lattice points from Λ per unit volume is given by $\sqrt{\det \Lambda}$ and the number of lattice points from Λ' per unit volume is $\sqrt{\det \Lambda'}$ so the number of elements in C, i.e. the number of coset representatives, is

$$|C| = \sqrt{\frac{\det \Lambda'}{\det \Lambda}}. \tag{2.4.3}$$

This number is called the **order** of Λ/Λ' and is denoted $|\Lambda/\Lambda'|$. As $|C|$ is an integer then $\sqrt{\det \Lambda'}$ is always a multiple of $\sqrt{\det \Lambda}$.

The set C of coset representatives is not unique as we can replace any element in C with itself plus a lattice point from Λ'. This is clearly evident from consideration of the shapes in Figure 2.6 as *any* set of distinct shapes is a set of coset representatives. Another way to see this is to take a tessellating region, R, for Λ', then a set of coset representatives is given by intersecting Λ with R (see Figure 2.7). That is, we can choose,

$$C = \{\mathbf{x} \in \Lambda \mid \mathbf{x} \in R\} = \Lambda \cap R.$$

It is common in the literature to assume that the coset representatives are given by the points from Λ contained in the Voronoi cell of Λ', but we will not make this assumption here. Two example sets of coset representatives are depicted in Figure 2.7. It is easy to verify the following remark.

Remark 2.2. *A set C is a set of coset representatives for the quotient Λ/Λ' if and only if $|C| = |\Lambda/\Lambda'|$ and for every pair $\mathbf{c}_1, \mathbf{c}_2 \in C$ the difference $\mathbf{c}_1 - \mathbf{c}_2$ is not a point in Λ'.*

[1] A lattice is closed under vector addition of its points and it always contains the origin (identity). If $\mathbf{x}$ is a lattice point then $-\mathbf{x}$ is also a lattice point (inverse) and vector addition is associative and commutative.

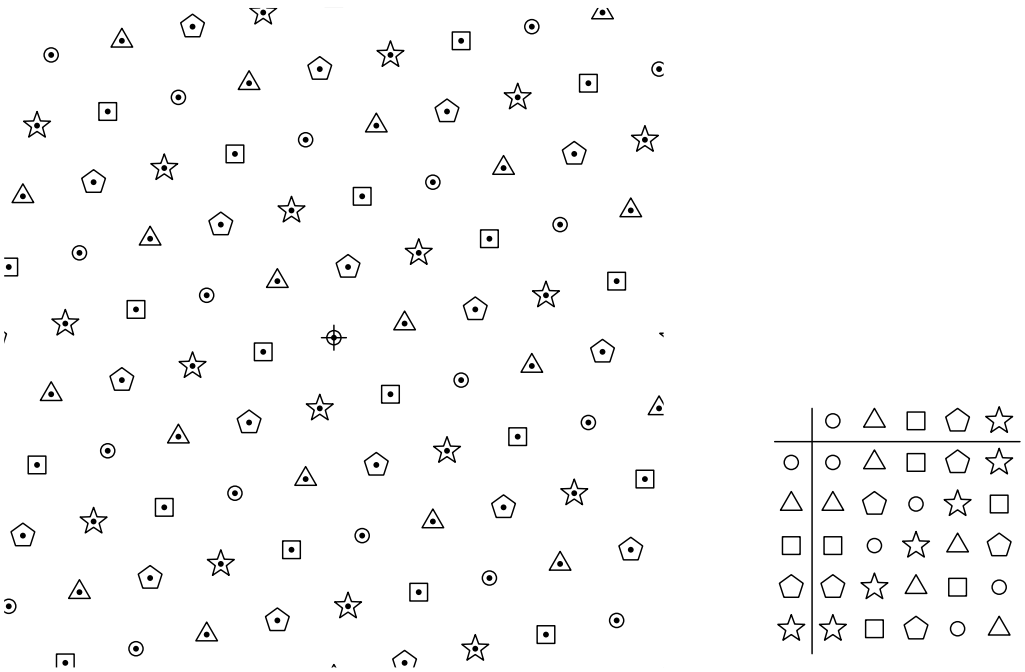


Figure 2.6: Lattice Λ with generator given by (2.1.1) (dots) and a sublattice Λ' with basis vectors $[1.8, -0.6]^\dagger$ and $[1.4, 2.2]^\dagger$ (circles). The cosets are given by the different *shapes*. On the right is the group table for the quotient group Λ/Λ' acting on the cosets. This figure is inspired by one given by Conway [1997, page 63].

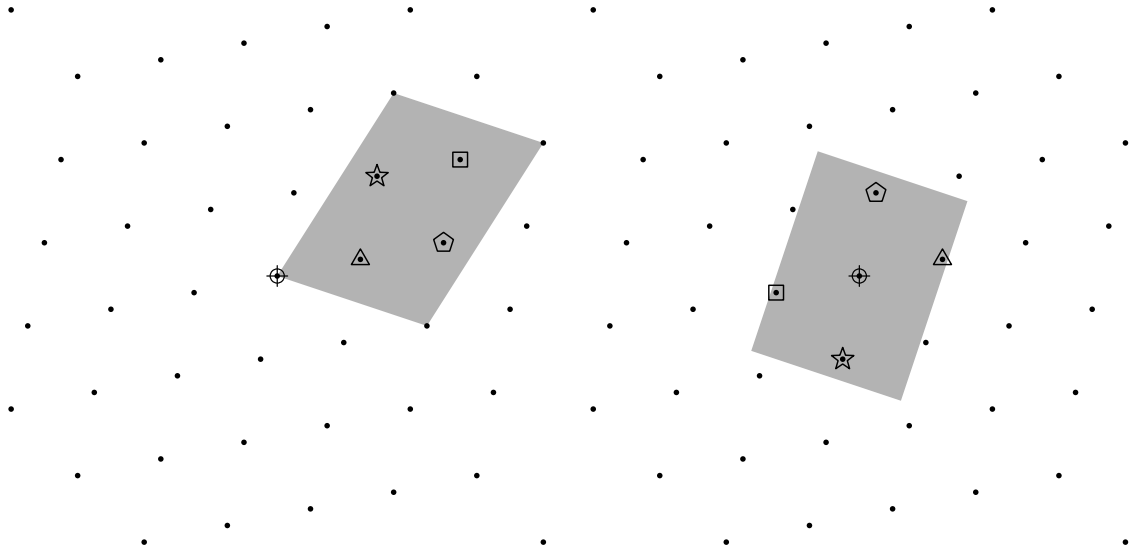


Figure 2.7: Two examples of the coset representatives for the quotient group Λ/Λ' of the lattices from Figure 2.6. The representatives are marked using the shapes. On the left the representatives are chosen to be those points from Λ' that intersect a fundamental parallelepiped of Λ'. On the right the coset representatives are chosen to be those points from Λ that intersect the rectangular tessellating region constructed using Proposition 2.1.

2.4.1 Enumerating coset representatives

It will be important in later sections that we are able to enumerate a set of coset representatives for a quotient Λ/Λ'. For this we make use of the **Hermite decomposition** of an integral matrix (a matrix with integer elements) [Cohen, 1993, page 69]. Given a $n \times n$ integral matrix $\mathbf{M}$ the Hermite decomposition of $\mathbf{M}$ is given by $\mathbf{M} = \mathbf{UR}$ where $\mathbf{U}$ is unimodular and $\mathbf{R}$ is an upper triangular integral matrix[2]. In a sense, the Hermite decomposition is an *integer* analogue of the QR-decomposition from linear algebra. The Hermite decomposition always exists when the columns of $\mathbf{M}$ are independent and it is also possible to compute the decomposition when $\mathbf{M}$ is not square, but we will not have use of this here. Numerous algorithms exist to compute the Hermite decomposition and the fastest require only a polynomial number of operations in n. See, for example [Cohen, 1993, Algorithm 2.4.5] or [Kannan and Bachem, 1979] or [Micciancio and Warinschi, 2001].

Let $\mathbf{A}$ be a generator for Λ and $\mathbf{B}$ be a generator for the sublattice Λ' then we can always find a square integral matrix $\mathbf{M}$ so that

$$\mathbf{B} = \mathbf{AM}.$$

We can compute $\mathbf{M}$ as $\mathbf{M} = \mathbf{A}^{-1}\mathbf{B}$ if $\mathbf{A}$ is square and $\mathbf{M} = \mathbf{A}^{+}\mathbf{B}$ if $\mathbf{A}$ is rectangular where $\mathbf{A}^{+} = (\mathbf{A}^{\dagger}\mathbf{A})^{-1}\mathbf{A}^{\dagger}$ is the pseudoinverse of $\mathbf{A}$. It is also straightforward to see that the determinant of $\mathbf{M}$ is equal to the order of the quotient group $|\Lambda/\Lambda'|$ because

$$\det \mathbf{M} = \sqrt{\frac{\det(\mathbf{B}^{\dagger}\mathbf{B})}{\det(\mathbf{A}^{\dagger}\mathbf{A})}} = \sqrt{\frac{\det \Lambda'}{\det \Lambda}} = |\Lambda/\Lambda'|.$$

Because $\mathbf{M}$ is integral we can compute its Hermite decomposition and we will use this to compute a set of coset representatives for the quotient Λ/Λ'. The process is explained in the next proposition.

Proposition 2.2. *Let $\mathbf{A}$ be a generator for the n-dimensional lattice Λ and let $\mathbf{B}$ be a generator for the sublattice Λ'. Let $\mathbf{M}$ be the integral matrix such that $\mathbf{B} = \mathbf{AM}$ and let $\mathbf{M} = \mathbf{UR}$ be the Hermite decomposition of $\mathbf{M}$. A set of coset representatives for the quotient Λ/Λ' is given by $\mathbf{AUt}$ for all vectors $\mathbf{t}$ with elements $t_i = 0, 1, 2, \ldots, r_{i,i} - 1$ where $r_{i,i}$ is the ith diagonal of $\mathbf{R}$.*

Proof. It will be convenient to define the set T to contain all vectors $\mathbf{t}$ with elements $t_i = 0, 1, 2, \ldots, r_{i,i} - 1$. Notice that because $\mathbf{R}$ is upper triangular and $\mathbf{U}$ is unimodular then the number of such vectors $\mathbf{t}$ is precisely $\det \mathbf{M} = |\Lambda/\Lambda'|$ so the proposition does define the correct number of coset representatives. Following Remark 2.2 it is now sufficient to show that for no two distinct vectors $\mathbf{t}$, $\mathbf{t}' \in T$ do we have $\mathbf{AU}(\mathbf{t} - \mathbf{t}')$ being a lattice point in Λ. The proof is now by contradiction.

[2]Cohen [1993] actually defines the Hermite decomposition by $\mathbf{M} = \mathbf{RU}$ with the unimodular matrix on the right. We find it more convenient to put the unimodular matrix on the left here. The existence of both the left and right version is guaranteed. In fact, the **Smith decomposition** guarantees the existence of two unimodular matrices $\mathbf{U}$ and $\mathbf{V}$ and a diagonal integral matrix $\mathbf{D}$ such that $\mathbf{M} = \mathbf{UDV}$ [Cohen, 1993, page 75].

Assume that $\mathbf{AU}(\mathbf{t}-\mathbf{t}') \in \Lambda$ then we can write $\mathbf{AU}(\mathbf{t}-\mathbf{t}') = \mathbf{Bw}$ for some $\mathbf{w} \in \mathbb{Z}^n$. Multiplying both sides by the inverse (or psuedoinverse) of $\mathbf{AU}$ we obtain

$$\mathbf{t}-\mathbf{t}' = \mathbf{Rw}.$$

Notice from the definition of the $\mathbf{t}$ that the elements must satisfy

$$|t_i - t_i'| < r_{k,k}$$

and as $\mathbf{t}$ and $\mathbf{t}'$ are distinct we can let k be the largest integer such that $t_k - t_k' \neq 0$. Now because $t_i - t_i = 0$ for all $i > k$ then $w_i = 0$ for all $i > k$ and therefore

$$t_k - t_k' = r_{k,k} w_k$$

because $\mathbf{R}$ is upper triangular. But this is impossible because $0 < |t_i - t_i'| < r_{k,k}$ and w_k is an integer, hence the proposition is true by contradiction. □

To make the process of coset enumeration completely clear we will give the following example using the lattices from Figure 2.6 with generators

$$\mathbf{A} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} 1.8 & 1.4 \\ -0.6 & 2.2 \end{bmatrix}.$$

The quotient $\mathbf{A}\mathbb{Z}^2/\mathbf{B}\mathbb{Z}^2$ has order $\det \mathbf{B}/\det \mathbf{A} = 5$. In order to compute the coset representatives we first compute the integral matrix

$$\mathbf{M} = \mathbf{A}^{-1}\mathbf{B} = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

and compute its Hermite decomposition

$$\mathbf{M} = \mathbf{UR} = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 5 \end{bmatrix}.$$

The coset representatives are then given as

$$C = \left\{ \mathbf{AU} \begin{bmatrix} 0 \\ t \end{bmatrix} \mid t = 0, 1, 2, 3, 4 \right\}.$$

In Chapter 4 we will be interested in enumerating a set of coset representatives involving the particular lattices $V_{n/m}$ and $V_{n/m}^*$ both of dimension n. For these special lattices, it is actually possible to enumerate the cosets in a (significantly) lower dimensional space than the lattice and this leads to some computational gains. We will explain this technique in a general setting in Section 2.6.

2.5 The dual lattice

Let Λ be a lattice. Its **dual lattice**, denoted Λ^*, contains those points that have integral inner product with all points from Λ, that is

$$\Lambda^* = \{\mathbf{x} \mid \forall \mathbf{y} \in \Lambda \ , \ \mathbf{x} \cdot \mathbf{y} \in \mathbb{Z}\}.$$

It is not hard to show that if $\mathbf{B}$ is a square generator for Λ then $(\mathbf{B}^{-1})^\dagger$ is a square generator for the dual lattice Λ^*. Similarly if $\mathbf{B}$ is a rectangular generator for Λ then $(\mathbf{B}^+)^\dagger$ is a rectangular generator for Λ^*. It is also easy to show that if $\mathbf{A}$ is a Gram matrix for Λ then $\mathbf{A}^{-1}$ is a Gram matrix for Λ^* and it follows that the determinants of a lattice and its dual are reciprocals, that is

$$\det \Lambda = \frac{1}{\det \Lambda^*}. \tag{2.5.1}$$

These results can be found in, for example, Conway and Sloane [1998, p. 10].

The dual lattice has special properties if Λ is integral. This is because (as we shall see in the next proposition) an integral lattice is a sublattice of its dual and we can therefore define a quotient group. Such a quotient group is given the special name **dual quotient group**.

Proposition 2.3. *If Λ is an integral lattice of dimension n then:*

1. *Λ is a sublattice of its dual lattice, i.e. $\Lambda \subseteq \Lambda^*$.*
2. *The dual quotient group Λ^*/Λ has order equal to $\det \Lambda$.*
3. *Λ is* self-dual *(i.e. $\Lambda = \Lambda^*$) if and only if Λ is unimodular.*

Proof. Statement (1) follows by letting $\mathbf{x} \in \Lambda$ and noticing that $\mathbf{x} \cdot \mathbf{y} \in \mathbb{Z}$ for any point $\mathbf{y} \in \Lambda$ and therefore $\mathbf{x} \in \Lambda^*$. Statement (2) follows from (2.5.1) and (2.4.3). Statement (3) is true because if Λ is unimodular then $\det \Lambda = 1$ and therefore $|\Lambda^*/\Lambda| = 1$ which occurs if and only if $\Lambda = \Lambda^*$. The argument in reverse shows that if Λ is self-dual then it is also unimodular. □

2.6 Lattices generated by intersections and projections

We will have extensive use of the following results, much of which can be found in Martinet [2003, Section 1.3]. In this section we let H be an r dimensional subspace of $\mathbb{R}^n$ and let $H^\perp$ denote the $n - r$ dimensional subspace orthogonal to H (the complementary space). We denote by p the orthogonal projection into H and by $p^\perp$ the orthogonal projection into $H^\perp$. For suitable choices of H it is the case that $\Lambda \cap H$ is an r dimensional sublattice of Λ. This sublattice and its dual have some interesting properties.

Proposition 2.4. [Martinet, 2003] *Let Λ be an n dimensional lattice. Then $\Lambda \cap H$ is an r dimensional lattice if and only if $\Lambda^* \cap H^\perp$ is an $n - r$ dimensional lattice. Furthermore if $\Lambda \cap H$ is an r dimensional lattice then:*

1. *The dual of $\Lambda \cap H$ is the orthogonal projection of Λ^* onto H. That is*

$$(\Lambda \cap H)^* = p\Lambda^*$$

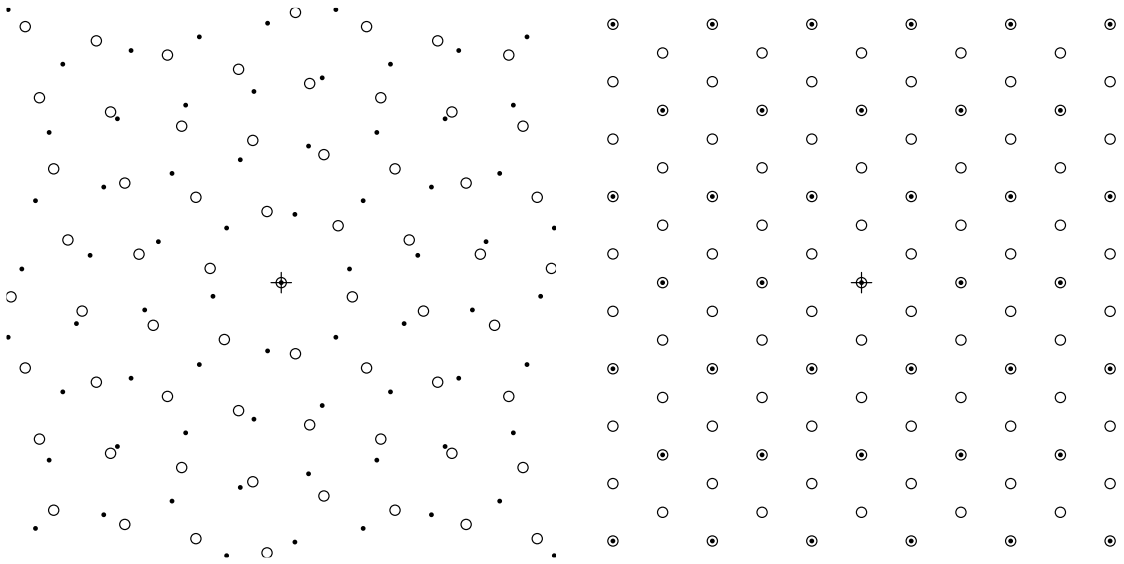


Figure 2.8: (Left) The lattice with generator matrix (2.1.1) (dots) and it's dual lattice (circles). (Right) The hexagonal lattice A_2 (dots) and it's dual A_2^* (circles). A_2 is an integral lattice and therefore A_2 is a sublattice of A_2^*. Notice that in this case A_2 is isomorphic to A_2^*, but this is not in general true for integral lattices and their duals.

2. The determinants of Λ, $\Lambda \cap H$ *and* $\Lambda^* \cap H^\perp$ *are related by*

$$\det(\Lambda)\det(\Lambda^* \cap H^\perp) = \det(\Lambda \cap H).$$

Proof. The proofs are given by Martinet [2003, Section 1.3]. The basic approach is to consider a basis matrix $\mathbf{B}$ for Λ such that the first r columns of $\mathbf{B}$ are in H and therefore are a basis for $\Lambda \cap H$. Projecting the columns of $\mathbf{B}$ by p and by $p^\perp$ leads to the proofs.

□

The next lemma relates the Voronoi cell of a lattice with the Voronoi cell of a lattice generated by an intersection. We will have particular use of this lemma when considering some properties of the Voronoi cell of the lattice A_n in Chapter 3.

Lemma 2.1. *Let* Λ *be an* n *dimensional lattice. If* $\Lambda \cap H$ *is an* r *dimensional lattice then the projection of the Voronoi cell of* Λ *into* H *is a subset of intersection of the Voronoi cell of* $\Lambda \cap H$ *with* H*. That is*

$$p\,\mathrm{Vor}(\Lambda) \subseteq H \cap \mathrm{Vor}(\Lambda \cap H).$$

Proof. Let $\mathbf{y} \in \mathrm{Vor}(\Lambda)$. Decompose $\mathbf{y}$ into orthogonal components so that $\mathbf{y} = \mathbf{t} + p\mathbf{y}$ for some $\mathbf{t} \in H^\perp$. Then $p\mathbf{y} \in p\,\mathrm{Vor}(\Lambda)$. Assume that $p\mathbf{y} \notin H \cap \mathrm{Vor}(\Lambda \cap H)$. Then there exists some $\mathbf{x} \in \Lambda \cap H$ such that

$$\begin{aligned} \|\mathbf{x} - p\mathbf{y}\|^2 < \|\mathbf{0} - p\mathbf{y}\|^2 &\Rightarrow \|\mathbf{x} - \mathbf{y} + \mathbf{t}\|^2 < \|\mathbf{y} - \mathbf{t}\|^2 \\ &\Rightarrow \|\mathbf{x} - \mathbf{y}\|^2 + 2\mathbf{x}\cdot\mathbf{t} < \|\mathbf{y}\|^2. \end{aligned}$$

By definition $\mathbf{x} \cdot \mathbf{t} = 0$ and so $\|\mathbf{x} - \mathbf{y}\|^2 < \|\mathbf{y}\|^2$. This violates that $\mathbf{y} \in \mathrm{Vor}(\Lambda)$ and hence $p\mathbf{y} \in H \cap \mathrm{Vor}(\Lambda \cap H)$. □

We will be particularly interested in the case when Λ is unimodular and we will make this assumption for the remainder of this section. Recall that unimodular lattices have determinant 1 and are integral and self dual, i.e. $\Lambda = \Lambda^*$ (see Proposition 2.3). Applying this to Proposition 2.4 we obtain the following corollary.

Corollary 2.2. *Let Λ be an n dimensional unimodular lattice. Then $\Lambda \cap H$ is an r dimensional lattice if and only if $\Lambda \cap H^\perp$ is an $n-r$ dimensional lattice and:*

1. *The lattice $p\Lambda$ is the dual of $\Lambda \cap H$ and the lattice $p^\perp\Lambda$ is the dual of $\Lambda \cap H^\perp$. That is*
$$p\Lambda = (\Lambda \cap H)^* \qquad \text{and} \qquad p^\perp\Lambda = (\Lambda \cap H^\perp)^*.$$
2. *The determinants of $\Lambda \cap H$ and $\Lambda \cap H^\perp$ are equal. That is*
$$\det(\Lambda \cap H) = \det(\Lambda \cap H^\perp).$$

Clearly $\Lambda \cap H$ and $\Lambda \cap H^\perp$ are integral lattices because they are subsets of the integral lattice Λ. So from Proposition 2.3 we can define the dual quotient groups $p\Lambda/\Lambda \cap H$ and $p^\perp\Lambda/\Lambda \cap H^\perp$. It follows from the above corollary that the order of these quotient groups is the same and is given by

$$|p\Lambda/\Lambda \cap H| = \det(\Lambda \cap H) = \det(\Lambda \cap H^\perp) = |p^\perp\Lambda/\Lambda \cap H^\perp|.$$

It turns out that the groups $p\Lambda/\Lambda \cap H$ and $p^\perp\Lambda/\Lambda \cap H^\perp$ are **isomorphic groups**. This means that they are essentially the same group, but they operate on different elements. We will not prove this (although the proof is quite easy) firstly because it is not the intention of this thesis to involve too much group theory and secondly because we only have use of the following related result connecting a set of coset representatives of $p\Lambda/\Lambda \cap H$ to a set of coset representatives of $p^\perp\Lambda/\Lambda \cap H^\perp$. We present this result in Theorem 2.2 but we first require the following simple lemma.

Lemma 2.2. *Let $\mathbf{x}$ be a lattice point in Λ. Then $p\mathbf{x}$ is a lattice point in $\Lambda \cap H$ if and only if $p^\perp\mathbf{x}$ is a lattice point in $\Lambda \cap H^\perp$.*

Proof. Notice that $\mathbf{x} = p\mathbf{x} + p^\perp\mathbf{x}$ so clearly $p\mathbf{x}$ is a lattice point in Λ if and only if $p^\perp\mathbf{x}$ is a lattice point in Λ. The proof follows because $p\mathbf{x} \in H$ and $p^\perp\mathbf{x} \in H^\perp$. □

Theorem 2.2. *Let K be a set of vectors from Λ and denote by pK and $p^\perp K$ sets containing the points from K projected into H and $H^\perp$ repectively. Then pK is a set of coset representatives for $p\Lambda/\Lambda \cap H$ if and only if $p^\perp K$ is a set of coset representatives for $p^\perp\Lambda/\Lambda \cap H^\perp$.*

Proof. If pK is a set of coset representatives for $p\Lambda/\Lambda \cap H$ then from Remark 2.2 it follows that for all pairs $\mathbf{k}_1, \mathbf{k}_2 \in K$ then $p(\mathbf{k}_1 - \mathbf{k}_2)$ is not a lattice point in $\Lambda \cap H$. From Lemma 2.2 it follows that $p^\perp(\mathbf{k}_1 - \mathbf{k}_2)$ is not a lattice point in $\Lambda \cap H^\perp$ and, because $|K| = |p^\perp\Lambda/\Lambda \cap H^\perp|$, then $p^\perp K$ is a set of coset representatives for $p^\perp\Lambda/\Lambda \cap H^\perp$. The converse follows using a similar argument. □

The utility of Theorem 2.2 is that we can obtain a set of coset representatives for both $p\Lambda/\Lambda \cap H$ and $p^{\perp}\Lambda/\Lambda \cap H^{\perp}$ by enumerating a set of coset representatives for just one of them. Notice that the dimension of $\Lambda \cap H$ is r and the dimension of $\Lambda \cap H^{\perp}$ is s. In some situations s is much smaller than r (or the opposite) and it is computationally easier to enumerate the coset representatives for $p^{\perp}\Lambda/\Lambda \cap H^{\perp}$ than it is for $p\Lambda/\Lambda \cap H$ (or the opposite). This result is useful in Chapter 4 when enumerating a set of coset representatives for the quotient $V^*_{n/m}/V_{n/m}$. In this case r is often very large (hundreds or even thousands) whereas s is usually quite small (we will not consider s larger than 4). Enumerating a set of coset representatives requires computing the Hermite decomposition (see Section 2.4) and for high dimensional matrices (when $r > 100$) this can be burdensome. It is convenient to instead enumerate the coset representatives in the low dimensional (i.e. $s < 4$) lattice. Moreover when s is sufficiently small we will be able to manipulate the Hermite decomposition by hand and this will lead to closed-form expressions for the coset representatives in a number of cases.

2.7 The nearest lattice point problem

Given a point $\mathbf{y} \in \mathbb{R}^n$ and a lattice Λ contained in $\mathbb{R}^n$, the **nearest lattice point problem** is to find the point $\mathbf{x} \in \Lambda$ such that the distance, with respect to a given norm, between $\mathbf{y}$ and $\mathbf{x}$ is minimised. Here, unless otherwise stated, the Euclidean norm will be assumed. We use the notation $\text{NearestPt}(\mathbf{y}, \Lambda)$ to denote the nearest point to $\mathbf{y}$ in the lattice Λ. It follows from the definition of the Voronoi cell that

$$\mathbf{x} = \text{NearestPt}(\mathbf{y}, \Lambda) \Leftrightarrow \mathbf{y} \in \text{Vor}(\Lambda) + \mathbf{x}.$$

Solutions to the nearest lattice point problem have numerous applications. For example, if a lattice is used as a quantiser then the nearest lattice point corresponds to the minimum distortion point. If the lattice is used as a code for a Gaussian channel, then the nearest lattice point corresponds to what is called **lattice decoding** and has been shown to yield arbitrarily good codes by Erez and Zamir [2004]. In communications systems featuring multiple antennas (MIMO) the problems of minimum mean square error decoding [El Gamal et al., 2004], vector perturbation [Peel et al., 2005; Hochwald et al., 2005; Ryan et al., 2008] and limited feedback beamforming [Ryan et al., 2007b; Love et al., 2004] all involve solving the nearest lattice point problem. The unwrapping of phase data for location estimation can also be posed as a nearest lattice point problem and this has been applied by Teunissen [1995, 2006] and Hassibi and Boyd [1998] to the global positioning system. Numerous cryptographic schemes also require solving the nearest lattice point problem [Goldreich et al., 1997; Gentry, 2009a,b].

The nearest lattice point problem is known to be NP-hard under certain conditions when the lattice itself, or rather a basis thereof, is considered as an additional input parameter [Micciancio, 2001; Ajtai, 1998; Dinur et al., 2003; Jalden and Ottersten, 2005]. Nevertheless, algorithms exist that can compute the nearest lattice point in reasonable time if the dimension is small. One such algorithm introduced by Pohst [1981] was popularised in the signal processing and communications fields by

Viterbo and Boutros [1999] and has since been called the **sphere decoder**. Kannan [1987] suggested a different approach that is known to be asymptotically faster than the sphere decoder. A good overview of these techniques is given by Agrell et al. [2002].

Approximate algorithms for computing the nearest point have also been studied. A classic example is **Babai's nearest plane algorithm** [Babai, 1986], which requires $O(n^4)$ arithmetic operations in the worst case where n is the dimension of the lattice and only $O(n^2)$ if the lattice basis is **Lovász reduced** [Lenstra et al., 1982]. Recently, various approximate techniques for solving the nearest lattice point problem have been motivated by applications to MIMO communications. An example is the K**-best algorithm** [Guo and Nilsson, 2006] that works similarly to the sequential M-algorithm [Anderson and Mohan, 1984] used in coding theory. Yet another example is the **fixed sphere decoder** [Jalden et al., 2009; Barbero and Thompson, 2008]. In this thesis we will make use of Babai's nearest plane algorithm, the sphere decoder and the K-best algorithm. We will not detail the workings of these algorithms as excellent descriptions already exist in the literature cited above.

Fast nearest point algorithms are known for specific lattices where the generator matrix is known a priory [Conway and Sloane, 1982, 1986; Clarkson, 1999a; McKilliam et al., 2008a,b, 2010c; Vardy and Be'ery, 1993]. In Chapter 3 we derive fast algorithms to find the nearest point in the lattice A_n^* and the related lattices A_n and A_n^m. In Chapter 5 we show how this algorithm can be used to estimate the **mean direction** of circular data. In Chapter 7 we show how the problem of polynomial phase estimation can be represented as a nearest lattice point problem in the lattice $V_{n/m}^*$ and in Chapter 4 we derive nearest point algorithms for $V_{n/m}^*$ that require a number of operations that is polynomial in the dimension of the lattice.

2.7.1 The nearest point in a superlattice

The following general approach can be useful to find the nearest point in a superlattice. We will use this approach later in Sections 3.5.4 and 4.3. Assume we have a lattice Λ and a sublattice $\Lambda' \subset \Lambda$ and that a nearest point algorithm for Λ' is known. That is, we assume it is easy to compute $\text{NearestPt}(\mathbf{y}, \Lambda')$. Then a simple nearest point algorithm for Λ can be constructed by iterating the nearest point algorithm Λ' for each of the coset representatives of the quotient Λ/Λ'. Pseudocode is given in Algorithm 2.1. This type of algorithm has previously been suggested by Conway and Sloane [1982]. The number of operations required is dependent on the order of the quotient group. In some cases the order can be large and this type of algorithm is not very efficient. We will use a similar approach to this in Chapter 4 to construct a polynomial time nearest point algorithm for $V_{n/m}^*$ using the the coset representatives of $V_{n/m}^*/V_{n/m}$.

2.7.2 Decoding into a rectangular tessellating region

Recall that in Proposition 2.1 we described how a lattice has a rectangular tessellating region (in fact there are many). It will be useful in later chapters to be able compute the lattice point that lies at the centre of a particular rectangular region.

Input: $\mathbf{y} \in \mathbb{R}^n$
$D = \infty$
$C =$ a set of coset representatives for $V^*_{n/m}/V_{n/m}$
foreach $\mathbf{g} \in C$ **do**
  $\mathbf{x} = \text{NearestPt}(\mathbf{y} - \mathbf{g}, \Lambda')$
  if $\|\mathbf{x} - \mathbf{y}\| < D$ **then**
    $\mathbf{x}_{\text{NP}} = \mathbf{x} + \mathbf{g}$
    $D = \|\mathbf{x} - \mathbf{y}\|$
return $\mathbf{x}_{NP}$

ALGORITHM 2.1: Computing the nearest point in a superlattice using coset representatives of the quotient group Λ/Λ' given by the set C.

That is, given a point $\mathbf{y} \in \mathbb{R}^n$ and an n-dimensional lattice Λ with generator matrix $\mathbf{B}$ and its associated rectangular tessellating region, we want to compute the lattice point $\mathbf{x} \in \Lambda$ that is at the centre of the rectangular region that contains $\mathbf{y}$. In effect we have replaced the Voronoi cell in the standard nearest lattice point problem with the rectangular tessellating region. Algorithm 2.2 describes how to achieve this. The algorithm requires $O(n^3)$ operations due to the QR-decomposition. If the QR-decomposition can be computed in advance then the remainder of the algorithm requires only $O(n^2)$ operations. If the generator matrix $\mathbf{B}$ is Lovász reduced then this algorithm is equivalent to Babai's nearest plane algorithm and can be used as an approximation to the nearest lattice point. We will make use of Babai's nearest plane algorithm for estimating polynomial phase signals in Chapter 10.

Input: $\mathbf{y} \in \mathbb{R}^m, \mathbf{B} \in \mathbb{R}^{m \times n}$
$[\mathbf{Q}, \mathbf{R}] = \text{QR}(\mathbf{B})$
$\mathbf{y}^* = \mathbf{Q}^T\mathbf{y}$
for $k = n$ **to** 1 **do**
  $r = \sum_{i=k+1}^{n} r_{k,i}u_i$
  $u_k = \lceil (y^*_k - r)/r_{k,k} \rfloor$
return Bu

ALGORITHM 2.2: Decoding into a rectangular tessellating region. If the basis matrix is Lovász reduced then this algorithm is equivalent to Babai's nearest plane algorithm which can be used as an approximation to the nearest lattice point.

2.8 Summary

In this chapter we have given a brief overview of lattice theory. We showed how a lattice is generated by multiplying the integer lattice $\mathbb{Z}^n$ by a matrix called a **generator matrix** and defined a **fundamental parallelepiped** as the parallelepiped constructed from the columns of a generator matrix. We also defined the **determinant** of a lattice as the square of the n-volume of the fundamental parallelepiped or equivalently as the determinant of the **Gram matrix**.

In Section 2.2 we showed how a fundamental parallelepiped tessellates on the lattice and we also considered another *rectangular* tessellating region that exists for any lattice (Proposition 2.1). We will find that the rectangular tessellating regions are very useful in Chapter 7 when we describe some of the aliasing properties of polynomial phase signals. In Section 2.3 we described a special tessellation called the **Voronoi cell** that describes the region of space nearest to a lattice point. The Voronoi cell is of crucial importance when considering the **nearest lattice point problem**.

In Section 2.4 we introduced **sublattices**, the **quotient group** and the associated **cosets** and **coset representatives**. We described how a set of **coset representatives** can be computed using the **Hermite decomposition**. This will be useful in Chapter 4 when we derive an algorithm to compute a nearest lattice point in the lattice $V_{n/m}^*$. We also discussed **dual lattices** in Section 2.5 and considered some of the special properties of the dual of an **integral lattice**. We defined the **dual quotient group** and derived some of its properties.

In Section 2.6 we considered lattices that are generated by *projection* and *intersection* with a subspace of $\mathbb{R}^n$. Elegant relationships emerge when we consider the projection and intersection of a **unimodular lattice**. The projected lattice is then the dual of the intersected lattice and some interesting properties exist relating the corresponding dual quotient groups. These results were summarised in Theorem 2.2.

In the next chapter we will show that the lattice A_n is an intersection of the integer lattice $\mathbb{Z}^{n+1}$ with a hyperplane and that the dual lattice A_n^* is the projection of the integer lattice into to the hyperplane. This leads to some very interesting geometric properties regarding A_n and A_n^* and inspires some very fast nearest lattice point algorithms. In Chapter 4 it turns out that the lattice $V_{n/m}$ is the intersection of the integer lattice with a subspace of dimension $m+1$ and $V_{n/m}^*$ is the projection of the integer lattice into the subspace. Corollary 2.2 asserts that these lattices are dual and leads to convenient formula for the determinant and also to descriptions of generator matrices for these lattices. By applying Theorem 2.2 we obtain simple methods for enumerating a set of coset representatives for the dual quotient group $V_{n/m}^*/V_{n/m}$. This leads to the discovery of a nearest point algorithm for $V_{n/m}^*$ that requires a polynomial number of operations in the dimension of the lattice n.

Finally, in Section 2.7, we discussed the **nearest lattice point problem**. In general the problem is known to be NP-hard. We described some standard approaches to computing or approximating a nearest point in any lattice. These are the **sphere decoder**, **Babai's nearest plane algorithm** and the **K-best algorithm**. We will consider using these algorithms for the estimation of polynomial phase signals in Chapter 10. The specific algorithms we discover for A_n^*, A_n and A_n^m in the next chapter are much faster than these general purpose algorithms.

—Science is what we understand well enough to explain to a computer. Art is everything else we do.

Donald Knuth

3

The lattices A_n, A_n^* and A_n^m

In this chapter we introduce the lattice A_n and its dual, the important lattice A_n^*. We also introduce the related lattices **Coxeter lattices**, A_n^m, that *lie between* A_n and A_n^*. In Chapter 6 we will show that an excellent estimator for the mean direction of a circular random variable can be computed by finding a nearest lattice point in the lattice A_n^*. For this reason the primary goal of this chapter is to derive fast algorithms for computing a nearest lattice point in the lattice A_n^*. As a by product we also find fast nearest point algorithms for the related lattices A_n and A_n^m. The algorithm we describe for A_n has appeared previously in the literature due to A. M. Odlyzko (see Conway and Sloane [1998, page 448]) but the algorithms we describe for A_n^* and for A_n^m are new and are the fastest known (and up to order the fastest possible).

In order to describe these algorithms we require an overview of some of the interesting properties of these lattices and in the first section we define A_n and A_n^* using *intersections* and *projections* of the integer lattice $\mathbb{Z}^{n+1}$. We can then use the results derived in Section 2.6 to easily find the determinant of A_n and A_n^*. Sections 3.2, 3.3 and 3.4 describe some more specific properties of the individual lattices A_n, A_n^* and A_n^m, the majority of which can be found in Conway and Sloane [1998, pp. 108-117] and Martinet [2003, Section 4.2 and 5.2]. In Section 3.5 we derive a number of nearest point algorithms for these lattices, the fastest of which require only a linear number of operations in the dimension of the lattice.

The algorithms for A_n and A_n^m do not feature again in this thesis and the reader more interested in circular statistics could skip Sections 3.4 and 3.5.3. That said, the derivation of the nearest point algorithms for each lattice are so neatly related that it is appropriate to include all of them. It is hoped that the fast nearest point algorithms for the A_n^m lattices will find their own applications in the future.

3.1 Definition of A_n and A_n^*

Let H be the hyperplane orthogonal to the all ones vector of length $n+1$, denoted by $\mathbf{1}$, that is

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^\dagger.$$

Any vector in H has the property that the sum (and therefore the mean) of its elements is zero and for this reason H is often referred to as the **zero-sum plane** or the **zero-mean plane**. Let $H^\perp$ be the subspace spanned by $\mathbf{1}$, i.e the subspace orthogonal to H. We define the lattice A_n to be the intersection of the integer lattice $\mathbb{Z}^{n+1}$ with H, that is

$$A_n = \mathbb{Z}^{n+1} \cap H = \big\{\mathbf{x} \in \mathbb{Z}^{n+1} \mid \mathbf{x} \cdot \mathbf{1} = 0\big\}. \tag{3.1.1}$$

Equivalently, A_n consists of all of those points in $\mathbb{Z}^{n+1}$ with coordinate sum equal to zero. Let $\mathbf{Q}$ be the orthogonal projection into H then, because $\mathbb{Z}^{n+1}$ is self-dual, it follows from Corollary 2.2 that the lattice constructed by projecting $\mathbb{Z}^{n+1}$ into H is the dual lattice of A_n, that is

$$A_n^* = \mathbf{Q}\mathbb{Z}^{n+1} = \big\{\mathbf{x} - \bar{x}\mathbf{1} \mid \mathbf{x} \in \mathbb{Z}^{n+1}\big\}$$

where $\bar{x} = \frac{1}{n+1}\mathbf{x} \cdot \mathbf{1}$ is the mean of the elements of $\mathbf{x}$. Corollary 2.2 also shows that the determinants are given by

$$\det(A_n) = \det(A_n^*)^{-1} = \det(\mathbb{Z}^{n+1} \cap H^\perp) = n+1$$

where $\mathbb{Z}^{n+1} \cap H^\perp$ is the lattice of points $\{c\mathbf{1} \mid c \in \mathbb{Z}\}$ which clearly has determinant $\mathbf{1} \cdot \mathbf{1} = n+1$.

3.2 Properties of A_n

A generator matrix for A_n is the $(n+1) \times n$ matrix

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}.$$

The lattice has kissing number $\tau = n(n+1)$, packing radius $\rho = 1/\sqrt{2}$, norm 2 and covering radius

$$R = \sqrt{\frac{a(n+1-a)}{n+1}}$$

where $a = \lfloor (n+1)/2 \rfloor$. The minimal vectors are given by $\mathbf{e}_i - \mathbf{e}_j$ where $i, j \in \{1, 2, \ldots, n+1\}$ and $i \neq j$. The minimal vectors are also the relevant vectors [Conway and Sloane, 1998, p. 108]. The Voronoi cell of A_n is closely related to the $n+1$ dimensional hypercube $\mathrm{Vor}(\mathbb{Z}^{n+1})$ as we shall show in the next theorem.

Theorem 3.1. *The projection of* $\mathrm{Vor}(\mathbb{Z}^{n+1})$ *into* H *is equal to* $H \cap \mathrm{Vor}(A_n)$*. That is*

$$H \cap \mathrm{Vor}(A_n) = \mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1}).$$

Proof. The n-volume of $H \cap \mathrm{Vor}(A_n)$ is given by the square root of the determinant of A_n, that is

$$\sqrt{\det A_n} = \sqrt{n+1}.$$

From Burger et al. [1996] we find that the n-volume of the projected hypercube $\mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1})$ is equal to $\sqrt{n+1}$ also. It follows from Lemma 2.1 that $\mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1}) \subseteq H \cap \mathrm{Vor}(A_n)$, so, because the volumes are the same, and because $H \cap \mathrm{Vor}_H(A_n)$ and $\mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1})$ are polytopes, we have $H \cap \mathrm{Vor}(A_n) = \mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1})$. □

3.3 Properties of A_n^*

The generator matrix for A_n^* is any n rows of the $(n+1)\times(n+1)$ projection matrix

$$\mathbf{Q} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\dagger}{n+1} = \frac{1}{n+1}\begin{bmatrix} n & -1 & -1 & -1 & \cdots & -1 & -1 \\ -1 & n & -1 & -1 & \cdots & -1 & -1 \\ -1 & -1 & n & -1 & \cdots & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & -1 & \cdots & -1 & n \end{bmatrix} \tag{3.3.1}$$

where $\mathbf{I}$ is the $(n+1)\times(n+1)$ identity matrix. The lattice A_n^* has determinant $(n+1)^{-1}$, kissing number $\tau = 2$ when $n = 1$ and $\tau = 2n+2$ when $n \geq 2$, norm $n/(n+1)$ and packing radius and covering radius given by

$$\rho = \frac{1}{2}\sqrt{\frac{n}{n+1}} \qquad R = \rho\sqrt{\frac{n+2}{3}}$$

[Conway and Sloane, 1998, p. 115].

An interesting property of A_n^* is that it gives the thinnest known covering in dimensions 2 through to 5. Until recently A_n^* was the thinnest known lattice covering in dimensions 2 to 23. Recent advances in computational geometry have enabled the discovery of thinner coverings in dimensions above 6 [Sikirič et al., 2008]. It turns out that many of the better coverings are given by the Coxeter lattices A_n^m that we discuss in the next section. The lattice A_n^* has also found application in a number of estimation problems including period and delay estimation from sparse timing data [Clarkson, 2008; McKilliam and Clarkson, 2008], frequency estimation [Clarkson, 1999b; McKilliam et al., 2010a] and phase estimation [Quinn, 2007]. We will discuss a number of these applications and more in Chapters 6.

The dual quotient group A_n^*/A_n has order $\det(A_n) = n+1$ and the coset representatives are given by Conway and Sloane [1998, p. 109] as

$$\mathbf{Q}\big[\underbrace{1, 1\ldots, 1}_{i \text{ times}}, 0, 0, \ldots, 0\big]^\dagger \tag{3.3.2}$$

for all $i \in \{0, \ldots, n\}$. An alternative definition is

$$i\mathbf{Q}\mathbf{e}_1 = i\left(\mathbf{e}_1 - \frac{\mathbf{1}}{n+1}\right) \tag{3.3.3}$$

for all $i \in \{0, \ldots, n\}$. Both of these definitions can be justified using Theorem 2.2, but we wont detail this here as it will become apparent when we discuss the coset representatives in more detail in the next chapter. An alternative generator matrix for A_n^* follows from (3.3.3) and (2.4.2) as

$$\begin{bmatrix} np & 0 & 0 & \cdots & 0 & 0 \\ -p & 1 & 0 & \cdots & 0 & 0 \\ -p & -1 & 1 & \cdots & 0 & 0 \\ -p & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -p & 0 & 0 & \cdots & -1 & 1 \\ -p & 0 & 0 & \cdots & 0 & -1 \end{bmatrix} \tag{3.3.4}$$

where $p = 1/(n+1)$.

Theorem 3.2. *The relevant vectors of A_n^* are given by the* $\mathbf{Q}\mathbf{u}$ *where*

$$\mathbf{u} = \sum_{i \in I} \mathbf{e}_i$$

and where $I \subset \{1, 2, \ldots, n+1\}$.

This result has been known since Voronoi [1908] and we wont provide a proof here. Perhaps the most accessible derivation is given by Conway and Sloane [1992] who make use of the concept of an **obtuse superbasis**. We will have use of the following trivial corollary when deriving a linear time nearest point algorithm for A_n^* in Section 3.5.2.

Corollary 3.1. *The lattice points $\mathbf{Q}\mathbf{e}_i \in A_n^*$ for $i = 1, 2, \ldots n+1$ are relevant vectors in A_n^*.*

3.4 Properties of the Coxeter lattices A_n^m

The lattices A_n^m are a family first described by Coxeter [1951]. The family is typically referred to as the **Coxeter lattices** [Martinet, 2003] and defined as

$$A_n^m = \left\{\mathbf{Q}\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}^{n+1}, \mathbf{x} \cdot \mathbf{1} \equiv 0 \bmod m\right\}. \tag{3.4.1}$$

That is, A_n^m consists of all those points from $\mathbb{Z}^{n+1}$ that have coordinate sum equal to zero modulo m. If m does not divide $n+1$ then $A_n^m = A_n^1$. Hence, in the sequel, we assume that m divides $n+1$. The Coxeter lattices are closely related to A_n and A_n^* because, when $m = 1$

$$A_n^1 = A_n^* = \{\mathbf{Q}\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}^{n+1}\}, \tag{3.4.2}$$

and when $m = n + 1$

$$A_n^{n+1} = A_n = \left\{ \mathbf{x} \in \mathbb{Z}^{n+1} \mid \mathbf{x} \cdot \mathbf{1} = 0 \right\}. \tag{3.4.3}$$

It is also easy to see that $A_n \subseteq A_n^m \subset A_n^k \subseteq A_n^*$ whenever $k < m$ and therefore

$$\mathrm{Vor}(A_n^*) \subseteq \mathrm{Vor}(A_n^k) \subset \mathrm{Vor}(A_n^m) \subseteq \mathrm{Vor}(A_n). \tag{3.4.4}$$

The quotient group A_n^m/A_n^k has order k/m. In particular the quotient group A_n^m/A_n has order $q = (n+1)/m$ and it follows from (2.4.3) that the determinant of A_n^m is given by

$$\det(A_n^m) = q^2 \det(A_n) = \frac{m^2}{n+1}.$$

The coset representatives are given by

$$im\mathbf{Q}\mathbf{e}_1 = im\left(\mathbf{e}_1 - \frac{\mathbf{1}}{n+1}\right)$$

where $i \in \{0, 1, \ldots, q-1\}$. A generator matrix for A_n^m follows immediately as (3.3.4) where $p = m/(n+1)$. The next corollary describes how the Voronoi cell of A_n^m is related to the $n+1$ dimensional hypercube.

Corollary 3.2. *The projection of* $\mathrm{Vor}(\mathbb{Z}^{n+1})$ *into* H *is a superset of* $H \cap \mathrm{Vor}(A_n^m)$. *That is*

$$H \cap \mathrm{Vor}(A_n^m) \subseteq \mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1}).$$

Proof. Follows directly from Theorem 3.1 and (3.4.4). □

3.5 Computing the nearest point

In this section we derive nearest point algorithms for A_n, A_n^* and A_n^m. We find that linear-time algorithms exist for all of these lattices. We will make use of the following definitions. Given two sets A and B we let $A + B$ be their Minkowski sum. That is, $x \in A + B$ if and only if $x = a + b$ where $a \in A$ and $b \in B$. We will also write $\mathbf{1}\mathbb{R}$ to denote the line of points $\mathbf{1}r$ for all $r \in \mathbb{R}$. Then $H \cap \mathrm{Vor}(A_n^m) + \mathbf{1}\mathbb{R}$ is an infinite cylinder with cross-section $H \cap \mathrm{Vor}(A_n^m)$. Because A_n^m is contained in the subspace orthogonal to $\mathbf{1}$ we see that

$$H \cap \mathrm{Vor}(A_n^m) + \mathbf{1}\mathbb{R} = \mathrm{Vor}(A_n^m).$$

Lemma 3.1. *If* $\mathbf{x} = \mathbf{Q}\mathbf{k}$ *is a closest point in* A_n^m *to* $\mathbf{y} \in \mathbb{R}^{n+1}$ *then there exists some* $\lambda \in \mathbb{R}$ *for which* $\mathbf{k}$ *is a closest point in* $\mathbb{Z}^{n+1}$ *to* $\mathbf{y} + \lambda\mathbf{1}$.

Proof. As $\mathbf{Q}\mathbf{k}$ is the nearest point to $\mathbf{y}$ then $\mathbf{y} \in \mathrm{Vor}(A_n^m) + \mathbf{Q}\mathbf{k}$, so for all $\mu \in \mathbb{R}$

$$\mathbf{y} + \mathbf{1}\mu \in \mathrm{Vor}(A_n^m) + \mathbf{Q}\mathbf{k} = H \cap \mathrm{Vor}(A_n^m) + \mathbf{k} + \mathbf{1}\mathbb{R}.$$

From Corollary 3.2 we have $H \cap \mathrm{Vor}(A_n^m) \subseteq \mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1})$ so

$$H \cap \mathrm{Vor}(A_n^m) + \mathbf{k} + \mathbf{1}\mathbb{R} \subseteq \mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1}) + \mathbf{k} + \mathbf{1}\mathbb{R}.$$

Then $\mathbf{y} + \mathbf{1}\mu \in \mathbf{Q}\,\mathrm{Vor}(\mathbb{Z}^{n+1}) + \mathbf{k} + \mathbf{1}\mathbb{R}$ and for some $\lambda \in \mathbb{R}$

$$\mathbf{y} + \mathbf{1}\lambda \in \mathrm{Vor}(\mathbb{Z}^{n+1}) + \mathbf{k}.$$

The proof now follows from the definition of the Voronoi cell. □

Now consider the function $\mathbf{f} : \mathbb{R} \mapsto \mathbb{Z}^{n+1}$ defined so that

$$\mathbf{f}(\lambda) = \lfloor \mathbf{y} + \lambda\mathbf{1} \rceil \tag{3.5.1}$$

That is, $\mathbf{f}(\lambda)$ gives a nearest point in $\mathbb{Z}^{n+1}$ to $\mathbf{y} + \lambda\mathbf{1}$ as a function of λ. Observe that $\mathbf{f}(\lambda + 1) = \mathbf{f}(\lambda) + \mathbf{1}$. Hence,

$$\mathbf{Q}\mathbf{f}(\lambda + 1) = \mathbf{Q}\mathbf{f}(\lambda). \tag{3.5.2}$$

Lemma 3.1 implies there exists some $\lambda \in \mathbb{R}$ such that $\mathbf{x} = \mathbf{Q}\mathbf{f}(\lambda)$ is a closest point in A_n^m to $\mathbf{y}$. Furthermore, we see from (3.5.2) that λ can be found within an interval of length 1. Hence, if we define the set

$$S = \{\mathbf{f}(\lambda) \mid \lambda \in [0, 1)\}$$

then the set $\mathbf{Q}S$ contains a closest point in A_n^m to $\mathbf{y}$. By setting $m = 1$ and $m = n+1$ it is clear that $\mathbf{Q}S$ also contains a nearest point in A_n^* and A_n to $\mathbf{y}$. The principle of all of the algorithms discussed here, with the exception of Section 3.5.4, is to search the set $\mathbf{Q}S$ for the nearest point. In order to evaluate the elements in S we require the following function.

Definition 3.1. (sort indices)
We define the function

$$\mathbf{s} = \mathrm{sortindices}(\mathbf{z})$$

to take a vector $\mathbf{z}$ *of length* $n + 1$ *and return a vector* $\mathbf{s}$ *of indices such that*

$$z_{s_1} \geq z_{s_2} \geq z_{s_3} \geq \cdots \geq z_{s_{n+1}}.$$

Computing sortindices($\mathbf{z}$) for a vector of length n requires $O(n \log n)$ arithmetic operations [Knuth, 1998]. Let

$$\mathbf{s} = \mathrm{sortindices}(\langle \mathbf{y} \rangle)$$

where $\langle g \rangle = g - \lceil g \rfloor$ denotes the centered fractional part of $g \in \mathbb{R}$ and we define $\langle \cdot \rangle$ to operate on vectors by taking the centered fractional part of each element in the vector. Observe that S contains at most $n + 2$ vectors, i.e.,

$$S \subseteq \left\{ \lceil \mathbf{y} \rfloor, \lceil \mathbf{y} \rfloor + \mathbf{e}_{s_1}, \lceil \mathbf{y} \rfloor + \mathbf{e}_{s_1} + \mathbf{e}_{s_2}, \ldots, \lceil \mathbf{y} \rfloor + \mathbf{e}_{s_1} + \cdots + \mathbf{e}_{s_{n+1}} \right\}. \tag{3.5.3}$$

It can be seen that the last vector listed in the set is simply $\lceil \mathbf{y} \rfloor + \mathbf{1}$ and so, once multiplied by $\mathbf{Q}$, the first and the last vectors are identical.

3.5.1 Algorithms for A_n

From the previous discussion we see that the nearest point in A_n to $\mathbf{y}$ is contained in $\mathbf{Q}S$. Noting that A_n may be defined as

$$A_n = A_{n/n+1} = \{\mathbf{Q}\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}^{n+1}, \mathbf{x} \cdot \mathbf{1} \equiv 0 \bmod n+1\}$$

we see that only those elements $\mathbf{Q}\mathbf{x}$, $\mathbf{x} \in S$ that satisfy $\mathbf{x} \cdot \mathbf{1} \equiv 0 \bmod n+1$ are candidates for the nearest point. Moreover, it is clear that there is exactly one element in $\mathbf{Q}S$ that satisfies this criterion, this element being

$$\mathbf{Q}\left(\lceil \mathbf{y} \rfloor + \mathbf{e}_1 + \mathbf{e}_2 + \ldots \mathbf{e}_\gamma\right)$$

where $\gamma = (n+1 - \lceil \mathbf{y} \rfloor \cdot \mathbf{1}) \operatorname{rem}(n+1)$ where $a \operatorname{rem} b$ indicates the remainder after division of a by b. Algorithm 3.1 follows.

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
1 $\gamma = (n+1 - \lceil \mathbf{y} \rfloor \cdot \mathbf{1}) \operatorname{rem}(n+1)$
2 $\mathbf{s} = \text{sortindices}(\langle \mathbf{y} \rangle)$
3 $\mathbf{u} = \lceil \mathbf{y} \rfloor$
4 **foreach** $i = 1$ **to** γ **do**
5 $\quad u_{s_i} = u_{s_i} + 1$
6 $\mathbf{x} = \mathbf{Q}\mathbf{u}$
7 **return** $\mathbf{x}$

Algorithm 3.1: Algorithm to find a nearest lattice point in A_n to $y \in \mathbb{R}^n$ that requires $O(n \log n)$ operations.

The operations on lines 1 and 3 and the loop on line 4 all require at most $O(n)$ operations. The matrix operation on line 6 can be computed in $O(n)$ operations by projecting $\mathbf{x}$ orthogonal to $\mathbf{1}$, that is $\mathbf{Q}\mathbf{x} = \mathbf{x} - \mathbf{x} \cdot \mathbf{1}/(n+1)$. The most computationally intensive operation is the sortindices$(\cdot)$ function that requires $O(n \log n)$ operations. Notice that the order in which the loop on line 4 iterates over the s_i for $i = 1, 2, \ldots, \gamma$ is unimportant. We can exploit this and improve the algorithm to require only $O(n)$ operations. We require the following function.

Definition 3.2. (quick partition)
We define the function

$$\mathbf{b} = \text{quickpartition}(\mathbf{z}, c)$$

to take a vector $\mathbf{z}$ *of length* $n+1$ *and integer* $c = 1, \ldots, n+1$ *and return a vector* $\mathbf{b}$ *of indices such that for* $i = 1, \ldots, c-1$ *and* $t = c+1, \ldots, n+1$

$$z_{b_i} \geq z_{b_c} \geq z_{b_t}.$$

Note that quickpartition$(\mathbf{z}, c)$ locates the cth largest element in $\mathbf{z}$. Somewhat surprisingly quickpartition$(\mathbf{z}, c)$ can be implemented such that the required number of operations is $O(n)$. This is facilitated by the **Rivest-Tarjan selection algorithm** [Blum et al., 1973; Floyd and Rivest, 1975a,b; Knuth, 1997]. A linear-time algorithm for A_n can be constructed from Algorithm 3.1 by replacing the sortindices$(\cdot)$ function with quickpartition$(\cdot)$ (see Algorithm 3.2). This algorithm has previously been suggested by A. M. Odlyzko [Conway and Sloane, 1998, page 448].

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
1 $\gamma = (n + 1 - \lceil \mathbf{y} \rfloor \cdot \mathbf{1}) \operatorname{rem}(n + 1)$
2 $\mathbf{b} = \text{quickpartition}(\langle \mathbf{y} \rangle, \gamma)$
3 $\mathbf{u} = \lceil \mathbf{y} \rfloor$
4 **for** $i = 1$ **to** γ **do**
5 $\quad u_{b_i} = u_{b_i} + 1$
6 $\mathbf{x} = \mathbf{Q}\mathbf{u}$
7 **return** $\mathbf{x}$

ALGORITHM 3.2: Algorithm to find a nearest lattice point in A_n to $y \in \mathbb{R}^n$ that requires $O(n)$ operations.

3.5.2 Algorithms for A_n^*

From Lemma 3.1 and the subsequent discussion we know that the nearest point in A_n^* to $\mathbf{y}$ is contained in $\mathbf{Q}S$. We desire to find the point $\mathbf{x} \in S$ such that $\|\mathbf{Q}\mathbf{x} - \mathbf{y}\|^2$ is minimised. That is, the nearest point in A_n^* to $\mathbf{y}$ is given by $\mathbf{Q}\mathbf{x}$ where

$$\mathbf{x} = \arg\min_{\mathbf{x} \in S} \|\mathbf{Q}\mathbf{x} - \mathbf{y}\|^2.$$

We can compute the elements in S using the sortindices$(\cdot)$ function. A naive approach would be to compute the distance $\|\mathbf{Q}\mathbf{x} - \mathbf{y}\|^2$ for each $\mathbf{x} \in S$ individually. This would require $O(n^2)$ operations. It is possible to compute each distance efficiently. Label the elements of S according to the order given in (3.5.3). That is, set $\mathbf{u}_0 = \lceil \mathbf{y} \rfloor$ and, for $i = 1, \ldots, n$, we can consecutively compute the elements in S as

$$\mathbf{u}_i = \mathbf{u}_{i-1} + \mathbf{e}_{s_i}. \tag{3.5.4}$$

Let $\mathbf{z}_i = \mathbf{y} - \mathbf{u}_i$. Clearly, $\mathbf{z}_0 = \langle \mathbf{y} \rangle$. Decompose $\mathbf{y}$ into orthogonal components $\mathbf{Q}\mathbf{y}$ and $t\mathbf{1}$ for some $t \in \mathbb{R}$. The squared distance between $\mathbf{Q}\mathbf{u}_i$ and $\mathbf{y}$ is

$$\|\mathbf{y} - \mathbf{Q}\mathbf{u}_i\|^2 = d_i + t^2(n + 1) \tag{3.5.5}$$

where d_i is defined as

$$d_i = \|\mathbf{Q}\mathbf{z}_i\|^2 = \left\| \mathbf{z}_i - \frac{\mathbf{z}_i \cdot \mathbf{1}}{n+1}\mathbf{1} \right\|^2 = \mathbf{z}_i \cdot \mathbf{z}_i - \frac{(\mathbf{z}_i \cdot \mathbf{1})^2}{n+1} = \beta_i - \frac{\alpha_i^2}{n+1}. \tag{3.5.6}$$

We know that the nearest point to $\mathbf{y}$ is that $\mathbf{Q}\mathbf{u}_i$ which minimises (3.5.5). Since the term $t^2(n + 1)$ is independent of the index i, it can be ignored. That is, it is sufficient to minimise d_i, $i = 0, \ldots, n$. The d_i can be calculated inexpensively using the following recursion. From (3.5.4),

$$\alpha_i = \mathbf{z}_i \cdot \mathbf{1} = (\mathbf{z}_{i-1} - \mathbf{e}_{s_i}) \cdot \mathbf{1} = \alpha_{i-1} - 1 \tag{3.5.7}$$

and

$$\beta_i = \mathbf{z}_i \cdot \mathbf{z}_i = (\mathbf{z}_{i-1} - \mathbf{e}_{s_i}) \cdot (\mathbf{z}_{i-1} - \mathbf{e}_{s_i}) = \beta_{i-1} - 2z_{s_i} + 1. \tag{3.5.8}$$

Algorithm 3.3 now follows. The main loop beginning at line 7 calculates the α_i and β_i recursively. There is no need to retain their previous values, so the subscripts

are dropped. The variable D maintains the minimum value of the (implicitly calculated values of) d_i so far encountered, and k the corresponding index. Each line of the main loop requires $O(1)$ arithmetic computations so the loop (and that on line 13) requires $O(n)$ in total. The vector operations on lines 2–4, 1 and 15 all require $O(n)$ operations. The computational cost of the algorithm is dominated by the sortindices function and is therefore $O(n \log n)$.

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
1 $\mathbf{u} = \lceil \mathbf{y} \rfloor$
2 $\mathbf{z} = \mathbf{y} - \mathbf{u}$
3 $\alpha = \mathbf{z} \cdot \mathbf{1}$
4 $\beta = \mathbf{z} \cdot \mathbf{z}$
5 $\mathbf{s} = \text{sortindices}(\mathbf{z})$
6 $D = \infty$
7 **for** $i = 1$ **to** $n + 1$ **do**
8 **if** $\beta - \frac{\alpha^2}{n+1} < D$ **then**
9 $D = \beta - \frac{\alpha^2}{n+1}$
10 $k = i - 1$
11 $\alpha = \alpha - 1$
12 $\beta = \beta - 2z_{s_i} + 1$
13 **for** $i = 1$ **to** k **do**
14 $u_{s_i} = u_{s_i} + 1$
15 $\mathbf{x} = \mathbf{Qu}$
16 **return** $\mathbf{x}$

Algorithm 3.3: Algorithm to find a nearest lattice point in A_n^* to $\mathbf{y} \in \mathbb{R}^{n+1}$ that requires $O(n \log n)$ arithmetic operations.

The linear-time algorithm

It is possible to improve this algorithm so that only $O(n)$ operations are required. We will show in Theorem 3.3 that only some of the vectors in $\mathbf{Q}S$ are candidates for the nearest point. This fact allows us to avoid using the sortindices$(\cdot)$ function. Instead a partial sorting operation called a **bucket sort** can be used [Cormen et al., 2001].

Lemma 3.2. *Let* $\mathbf{Qf}(\lambda_0)$ *be the closest point in* A_n^* *to* $\mathbf{y} \in \mathbb{R}^{n+1}$. *If* $\mathscr{I}$ *is defined as the interval containing* λ_0 *on which* $\mathbf{f}(\lambda)$ *is constant then the length of the interval is not less than* $1/(n+1)$.

Proof. Observe that $\mathbf{f}(\lambda)$ is piecewise constant, by virtue of the rounding operation. $\mathscr{I}$ will be open at one end and closed at the other. Which end is open and which is closed depends on the direction that half-integers are rounded. Let the endpoints of the interval be $\lambda_{\min} \leq \lambda_{\max}$.

Consider $\lambda \in \mathscr{I}$. For all such λ, $\mathbf{f}(\lambda)$ is constant. Let its value be $\mathbf{k}$ and let $\mathbf{x} = \mathbf{Qk}$. It is clear that $\mathbf{y} \in \text{Vor}(A_n^*) + \mathbf{x}$ and so $\mathbf{y} + \lambda\mathbf{1} \in \text{Vor}(A_n^*) + \mathbf{k}$. Also,

$\mathbf{y}+\lambda\mathbf{1} \in \mathrm{Vor}(\mathbb{Z}^{n+1})+\mathbf{k}$. With $\mathbf{z} = \mathbf{y}-\mathbf{k}$, it follows that $\mathbf{z}+\lambda\mathbf{1} \in \mathrm{Vor}(A_n^*)\cap\mathrm{Vor}(\mathbb{Z}^{n+1})$. The fact that $\mathbf{z}+\lambda\mathbf{1} \in \mathrm{Vor}(A_n^*)$ does not immediately yield any information on the length of the interval $\mathscr{I}$ since this Voronoi cell is an infinite cylinder whose central axis is in the direction of the vector $\mathbf{1}$. On the other hand, $\mathbf{z}+\lambda\mathbf{1} \in \mathrm{Vor}(\mathbb{Z}^{n+1})$ implies that $|z_i+\lambda| \leq \frac{1}{2}$. If we set $\ell = \arg\max_i z_i$ and $t = \arg\min_i z_i$, it is clear that $\lambda_{\max} = \frac{1}{2} - z_\ell$ and $\lambda_{\min} = -\frac{1}{2} - z_t$. Hence, the length of the interval is

$$\lambda_{\max} - \lambda_{\min} = 1 - z_\ell + z_t. \tag{3.5.9}$$

From Corollary 3.1 we see that both $\mathbf{Qe}_\ell$ and $\mathbf{Qe}_t$ are relevant vectors of A_n^*. From Remark 2.1, it follows that

$$\mathbf{z}\cdot(\mathbf{Qe}_\ell) \leq \frac{1}{2}\|\mathbf{Qe}_\ell\|^2$$

which implies that

$$z_\ell - \bar{z} \leq \frac{n}{2(n+1)} \tag{3.5.10}$$

where $\bar{z} = \mathbf{1}\cdot\mathbf{z}/(n+1)$. On the other hand, we must also have that

$$\mathbf{z}\cdot(-\mathbf{Qe}_t) \leq \frac{1}{2}\|\mathbf{Qe}_t\|^2$$

which implies that

$$z_t - \bar{z} \geq -\frac{n}{2(n+1)}. \tag{3.5.11}$$

Combining (3.5.9), (3.5.10) and (3.5.11), we find that the length of the interval $\mathscr{I}$ conforms to the lower bound

$$\lambda_{\max} - \lambda_{\min} \geq \frac{1}{n+1}.$$

□

Theorem 3.3. *If* $\mathbf{Qk}$ *is the nearest point in* A_n^* *to* $\mathbf{y} \in \mathbb{R}^{n+1}$ *then*

$$\mathbf{k} = \mathbf{f}\left(\frac{i-1}{n+1}\right)$$

for some $i \in \{1, \cdots, n+1\}$.

Proof. Assume that the lemma is false. Then $\mathbf{k} = \mathbf{f}(\lambda)$ for some $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ such that

$$\frac{i-1}{n+1} < \lambda < \frac{i}{n+1}$$

for some $i \in \{1, \cdots, n+1\}$. However then

$$\lambda_{\max} - \lambda_{\min} < \frac{i}{n+1} - \frac{i-1}{n+1} = \frac{1}{n+1}$$

contradicting Lemma 3.2. □

From Theorem 3.3 we see that only the lattice points

$$\mathbf{Qf}\left(\frac{i-1}{n+1}\right)$$

for $i \in \{1, \cdots, n+1\}$ are candidates for the nearest point. We will show how these points can be found in linear time. Define $n+1$ sets

$$B_i = \left\{ j \mid 0.5 - \langle y_j \rangle \in \left(\frac{i-1}{n+1}, \frac{i}{n+1} \right] \right\}$$

for $i \in \{1, \cdots, n+1\}$. Then it follows that

$$\begin{aligned}
\mathbf{f}\left(\frac{0}{n+1}\right) &= \lceil \mathbf{y} \rfloor \\
\mathbf{f}\left(\frac{1}{n+1}\right) &= \lceil \mathbf{y} \rfloor + \sum_{j \in B_1} \mathbf{e}_j \\
\mathbf{f}\left(\frac{2}{n+1}\right) &= \lceil \mathbf{y} \rfloor + \sum_{j \in B_1} \mathbf{e}_j + \sum_{j \in B_2} \mathbf{e}_j
\end{aligned}$$

and in general

$$\mathbf{f}\left(\frac{i}{n+1}\right) = \mathbf{f}\left(\frac{i-1}{n+1}\right) + \sum_{j \in B_i} \mathbf{e}_j \tag{3.5.12}$$

Let

$$k(i) = |B_1| + |B_2| + \cdots + |B_i|.$$

Then

$$\mathbf{u}_{k(i)} = \mathbf{f}\left(\frac{i}{n+1}\right) \tag{3.5.13}$$

and

$$\mathbf{z}_{k(i)} = \mathbf{y} - \mathbf{u}_{k(i)} \tag{3.5.14}$$

The nearest point is now given by $\mathbf{Qu}_{k(i)}$ where

$$i = \arg \min_{i=0,1,\ldots,n} \|\mathbf{Qu}_{k(i)} - \mathbf{y}\|^2.$$

We can again compute these distances efficiently in an identical manner to (3.5.5) and (3.5.6) where the recursive formula for the $d_{k(i)}$ is

$$\begin{aligned}
\alpha_{k(i)} = \mathbf{z}_{k(i)} \cdot \mathbf{1} &= \left(\mathbf{z}_{k(i-1)} - \sum_{j \in B_i} \mathbf{e}_j \right) \cdot \mathbf{1} \\
&= \alpha_{k(i-1)} - |B_i|
\end{aligned}$$

and

$$\begin{aligned}
\beta_{k(i)} = \mathbf{z}_{k(i)} \cdot \mathbf{z}_k(i) &= \left\| \mathbf{z}_{k(i-1)} - \sum_{j \in B_i} \mathbf{e}_j \right\|^2 \\
&= \beta_{k(i-1)} + |B_i| - 2 \sum_{j \in B_i} z_{s_i}.
\end{aligned}$$

Algorithm 3.4 now follows. Lines 4–7 calculate the sets B_i. This is the bucket sort operation [Cormen et al., 2001]. The remainder of the algorithm functions similarly to Algorithm 3.3. The vector operations on lines 1–3 and 21 all require $O(n)$ operations. Provided that the set operations on lines 4, 7, 11 and 19 can be performed in constant time the loops on lines 4, 5, 10 and 18 require only $O(n)$ operations. The overall computational complexity of the algorithm is then $O(n)$.

Naive implementation of the set operations may lead to poor performance. For this reason we have provided a second version of the pseudocode (Algorithm 3.6 on page 51) that hides the set notation but demonstrates how to efficiently implement the algorithm in practice. The sets, B_i, are replaced by two arrays ***bucket*** and ***link***, both of length $n+1$.

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
1 $\mathbf{z} = \langle \mathbf{y} \rangle$
2 $\alpha = \mathbf{z} \cdot \mathbf{1}$
3 $\beta = \mathbf{z} \cdot \mathbf{z}$
4 **for** $i = 1$ to $n+1$ **do** $B_i = \emptyset$
5 **for** $t = 1$ to $n+1$ **do**
6 $\quad i = n + 1 - (n+1)\lfloor z_t + 0.5 \rfloor$
7 $\quad B_i = B_i \cup \{t\}$
8 $D = \beta - \frac{\alpha^2}{n+1}$
9 $k = 0$
10 **for** $i = 1$ to $n+1$ **do**
11 $\quad$ **forall** $t \in B_i$ **do**
12 $\quad\quad \alpha = \alpha - 1$
13 $\quad\quad \beta = \beta - 2z_t + 1$
14 $\quad$ **if** $\beta - \frac{\alpha^2}{n+1} < D$ **then**
15 $\quad\quad D = \beta - \frac{\alpha^2}{n+1}$
16 $\quad\quad k = i$
17 $\mathbf{u} = \lceil \mathbf{y} \rfloor$
18 **for** $i = 1$ to k **do**
19 $\quad$ **forall** $t \in B_i$ **do**
20 $\quad\quad u_t = u_t + 1$
21 $\mathbf{x} = \mathbf{Q}\mathbf{u}$
22 **return x**

ALGORITHM 3.4: Algorithm to find a nearest lattice point in A_n^* to $\mathbf{y} \in \mathbb{R}^{n+1}$ that requires $O(n)$ arithmetic operations.

3.5.3 Algorithms for A_n^m

Define the set $W \subseteq S$ such that

$$W = \{\mathbf{x} \in S \mid \mathbf{x} \cdot \mathbf{1} \equiv 0 \bmod m\}. \tag{3.5.15}$$

It follows from Lemma 3.1 and the subsequent discussion that $\mathbf{Q}W$ contains the nearest point in A_n^m to $\mathbf{y}$. Algorithm 3.5 is now easily derived from Algorithm 3.3. The only difference being the variable γ used to ensure that the lattice points tested are in A_n^m, i.e. $\mathbf{Qw} \in \mathbf{Q}S$ is tested if $\mathbf{w} \cdot \mathbf{1} \equiv 0 \bmod m$. The number of operations required is dominated by the sortindices$(\cdot)$ function and is therefore $O(n \log n)$.

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
$\mathbf{u} = \lceil \mathbf{y} \rfloor$
$\mathbf{z} = \mathbf{y} - \mathbf{u}$
$\alpha = \mathbf{z} \cdot \mathbf{1}$
$\beta = \mathbf{z} \cdot \mathbf{z}$
$\gamma = \mathbf{u} \cdot \mathbf{1} \operatorname{rem} m$
$\mathbf{s} = \text{sortindices}(\mathbf{z})$
$D = \infty$
for $i = 1$ **to** $n + 1$ **do**
  if $\beta - \frac{\alpha^2}{n+1} < D$ **and** $\gamma = 0$ **then**
    $D = \beta - \frac{\alpha^2}{n+1}$
    $k = i - 1$
  $\alpha = \alpha - 1$
  $\beta = \beta - 2z_{s_i} + 1$
  $\gamma = (\gamma + 1) \operatorname{rem} m$
for $i = 1$ **to** k **do**
  $u_{s_i} = u_{s_i} + 1$
$\mathbf{x} = \mathbf{Qu}$
return $\mathbf{x}$

Algorithm 3.5: Algorithm to find a nearest lattice point in A_n^m to $\mathbf{y} \in \mathbb{R}^{n+1}$ that requires $O(n \log n)$ arithmetic operations.

The linear-time algorithm

We will show that some of the elements of $\mathbf{Q}W$ can be immediately excluded from consideration. This property leads to a nearest point algorithm that requires at most $O(n)$ arithmetic operations. The algorithm requires both a bucket sort and the quickpartition$(\cdot)$ function (Definition 3.2). It is interesting to note the similarities between the log-linear time algorithms for A_n, A_n^* and A_n^m. All make similar use of the sortindices$(\cdot)$ function which dominates the computational complexity. For A_n the key to enable a linear-time algorithm is the quickpartition$(\cdot)$ function implemented using the Rivest-Tarjan algorithm. For A_n^* the key is the bucket sort. It is somewhat satisfying that the linear-time algorithm for A_n^m should require both of these techniques.

Lemma 3.3. *Suppose, for some integers $i, m > 0, k \geq 2$, that*

$$\langle y_{s_i} \rangle - \langle y_{s_{i+km}} \rangle \leq \frac{m}{n+1}. \tag{3.5.16}$$

Then the minimum of the d_{i+cm} (3.5.6)*, over* $c = 0, \ldots, k$*, occurs at* $c = 0$ *or* $c = k$*.*

Proof. The proof proceeds by contradiction. Suppose, to the contrary, that

$$d_{i+cm} < d_i \qquad \text{and} \qquad d_{i+cm} < d_{i+km}.$$

Observe that

$$d_{i+cm} - d_i = \frac{2\alpha_i cm - (cm)^2}{n+1} + \sum_{j=1}^{cm}(1 - 2\left\langle y_{s_{i+j}}\right\rangle).$$

Now, since $\left\langle y_{s_{i+j}}\right\rangle \leq \langle y_{s_i}\rangle$, it follows that

$$d_{i+cm} - d_i \geq \frac{2\alpha_i cm - (cm)^2}{n+1} + cm(1 - 2\langle y_{s_i}\rangle)$$

and with the assumption that $d_{i+cm} - d_i < 0$, we have that

$$\frac{2\alpha_i - cm}{n+1} < 2\langle y_{s_i}\rangle - 1. \tag{3.5.17}$$

Similarly, observe that

$$d_{i+km} - d_{i+cm} = \frac{2\alpha_i(k-c)m - (k^2-c^2)m^2}{n+1} + \sum_{j=cm+1}^{km}(1 - 2\left\langle y_{s_{i+j}}\right\rangle).$$

Since $\left\langle y_{s_{i+j}}\right\rangle \geq \left\langle y_{s_{i+km}}\right\rangle$, it follows that

$$d_{i+km} - d_{i+cm} \leq \frac{2\alpha_i(k-c)m - (k^2-c^2)m^2}{n+1} + (k-c)m\big(1 - 2\left\langle y_{s_{i+km}}\right\rangle\big)$$

and with the assumption that $d_{i+km} - d_{i+cm} > 0$, we have that

$$\frac{2\alpha_i - cm}{n+1} > \frac{km}{n+1} - 1 + 2\left\langle y_{s_{i+km}}\right\rangle. \tag{3.5.18}$$

Equations (3.5.17) and (3.5.18) together imply that

$$\langle y_{s_i}\rangle - \left\langle y_{s_{i+km}}\right\rangle > \frac{km}{2(n+1)},$$

which contradicts (3.5.16) because $k \geq 2$. □

From S we can construct the following $q = (n+1)/m$ subsets

$$U_j = \left\{ \mathbf{u}_i \mid 0.5 - \langle y_{s_i}\rangle \in \left(\frac{m(j-1)}{n+1}, \frac{mj}{n+1} \right] \right\} \tag{3.5.19}$$

where $j = 1, \cdots, q$. Note that $\mathbf{Q}S = \mathbf{Q}\bigcup_{j=1}^{q} U_j$. We are interested in the $\mathbf{u}_i \in U_j$ such that $\mathbf{u}_i \cdot \mathbf{1} \equiv 0 \bmod m$, i.e., the elements in $U_j \cap W$. Let g be the smallest

integer such that $\mathbf{u}_g \in U_j \cap W$. Let p be the largest integer such that $\mathbf{u}_p \in U_j \cap W$. It follows that $p = g + km$ for some $k \in \mathbb{Z}$. Also, from (3.5.19)

$$\langle y_{s_g} \rangle - \langle y_{s_p} \rangle \leq \frac{m}{n+1}.$$

It then follows from Lemma 3.3 that (3.5.5) is minimised either by $\mathbf{u}_g$ or $\mathbf{u}_p$ and not by any $\mathbf{u}_i \in U_j \cap W$ where $g < i < p$. We see that for each set $\mathbf{Q}U_j$ there are at most two elements that are candidates for the nearest point. An algorithm can be constructed as follows: test the (at most two) candidates in each set $\mathbf{Q}U_j$ and return the closest one to $\mathbf{y}$. We will now show how this can be achieved in linear time.

We construct q sets

$$B_i = \left\{ j \mid 0.5 - \langle y_j \rangle \in \left(\frac{m(i-1)}{n+1}, \frac{mi}{n+1} \right] \right\}. \tag{3.5.20}$$

and let $k(i) = |B_1| + |B_2| + \cdots + |B_i|$. It follows that

$$\begin{aligned} \mathbf{u}_0 &= \lceil \mathbf{y} \rfloor \\ \mathbf{u}_{k(1)} &= \lceil \mathbf{y} \rfloor + \sum_{t \in B_1} \mathbf{e}_t \\ \mathbf{u}_{k(2)} &= \lceil \mathbf{y} \rfloor + \sum_{t \in B_1} \mathbf{e}_t + \sum_{t \in B_2} \mathbf{e}_t \end{aligned}$$

and in general

$$\mathbf{u}_{k(i)} = \mathbf{u}_{k(i-1)} + \sum_{t \in B_i} \mathbf{e}_t.$$

Let g be the smallest integer such that $1 \leq g \leq |B_i|$ and

$$\mathbf{1} \cdot \mathbf{u}_{k(i-1)+g} \equiv 0 \bmod m$$

and let p be the largest integer such that $1 \leq p \leq |B_i|$ and

$$\mathbf{1} \cdot \mathbf{u}_{k(i-1)+p} \equiv 0 \bmod m.$$

From the previous discussion the only candidates for the nearest point out of the elements

$$\mathbf{Q} \left\{ \mathbf{u}_{k(i-1)+1}, \ldots, \mathbf{u}_{k(i-1)+|B_i|} \right\} = \mathbf{Q}U_j$$

are $\mathbf{Q}\mathbf{u}_{k(i-1)+g}$ and $\mathbf{Q}\mathbf{u}_{k(i-1)+p}$. We can compute these elements quickly using the following function.

Definition 3.3. *We define the function*

$$\mathbf{b} = \text{quickpartition2}(\mathbf{z}, B_i, g, p)$$

to return the vector $\mathbf{b}$ *containing integers from* B_i *so that for* $j = 1, \ldots, g-1$ *and* $t = g+1, \ldots, p-1$ *and* $c = p+1, \ldots, |B_i|$

$$z_{b_j} \geq z_{b_g} \geq z_{b_t} \geq z_{b_p} \geq z_{b_c}.$$

Notice that quickpartition2$(\cdot)$ can be performed by two consecutive iterations of the Rivest-Tarjan algorithm and therefore requires $O(|B_i|)$ operations. We can compute

$$\mathbf{b} = \text{quickpartition2}(\mathbf{z}, B_i, g, p), \tag{3.5.21}$$

then the first candidate nearest point is

$$\mathbf{u}_{k(i-1)+g} = \mathbf{u}_{k(i-1)} + \sum_{t=1}^{g} \mathbf{e}_{b_t} \tag{3.5.22}$$

and the second candidate nearest point is

$$\mathbf{u}_{k(i-1)+p} = \mathbf{u}_{k(i-1)} + \sum_{t=1}^{p} \mathbf{e}_{b_t}. \tag{3.5.23}$$

Algorithm 3.7 stated on page 52 follows. Lines 2-5 construct the sets B_i using a bucket sort similar to that in Algorithm 3.4. The main loop on line 12 then computes the values of g and p for each B_i. The $d_{k(i)+g}$ and $d_{k(i)+p}$ are computed within the loop on line 16 and the index of the nearest lattice point is stored using the variable k^*. The concatenate$(\mathbf{w}, \mathbf{b})$ function on line 24 adds the elements of $\mathbf{b}$ to the end of the array $\mathbf{w}$. This can be performed in $O(|B_i|)$ operations. Lines 25-27 recovers the nearest lattice point using $\mathbf{w}$ and k^*.

In practice the B_i can be implemented as a list so that the set insertion operation on line 5 can be performed in constant time. Then the loops on lines 2 and 3 require $O(n)$ arithmetic operations. The operations inside the main loop on line 12 require $O(|B_i|)$ operations. The complexity of these loops is then

$$\sum_{i=1}^{(n+1)/m} O(|B_i|) = O(n)$$

The remaining lines require $O(n)$ or less operations. The algorithm then requires $O(n)$ arithmetic operations in total.

3.5.4 Algorithms based on the quotient group A_n^m/A_n

Given that a nearest point algorithm exists for the lattice A_n we can use the approach suggested in Section 2.7 to construct a nearest point algorithm for A_n^m by using a set of coset representatives for A_n^m/A_n. Recall from Section 3.4 and (3.3.3) that the coset representatives are given by $mi\mathbf{Q}\mathbf{e}_1$ where $i = 1, 2, \ldots, q$ and $q = (n+1)/m$. Algorithm 3.8 stated on page 52 follows. The running time depends on the order of A_n^m/A_n which is equal to q. If q is constant then the algorithm requires $O(n)$ operations. However, q may grow with n. For example $q = n + 1$ in the case of $A_n^1 = A_n^*$ and the algorithm requires $O(n^2)$ operations. This algorithm, applied to the lattice A_n^*, has previously been suggested by Conway and Sloane [1982].

3.5.5 Run-time analysis

Here we tabulate some practical computation times attained with the nearest point algorithms described in this chapter. The dimension of the lattices was set to $n+1 = 16i + 4$ for $i = 1$ to 33 and 10^5 trials were run for each value of n. The algorithms were written in Java and the computer used is a 900 MHz Intel Celeron M. The results are displayed in Figure 3.1. It is evident that the linear-time algorithms are faster than the log-linear-time algorithms in practice, at least when n is sufficiently large (approximately $n > 40$). Notice from Figure 3.1(d) how the algorithm based on the quotient group is comparable to the other algorithms for A_n^m where $m = (n+1)/4$. This behaviour is expected as the order of the quotient group is only $q = m/(n+1) = 4$ and therefore the complexity is $O(n)$. Contrastingly the quotient group algorithm is very slow for A_n^4 and A_n^*, where it has quadratic complexity (Figures 3.1(c) and 3.1(b)).

3.6 Summary

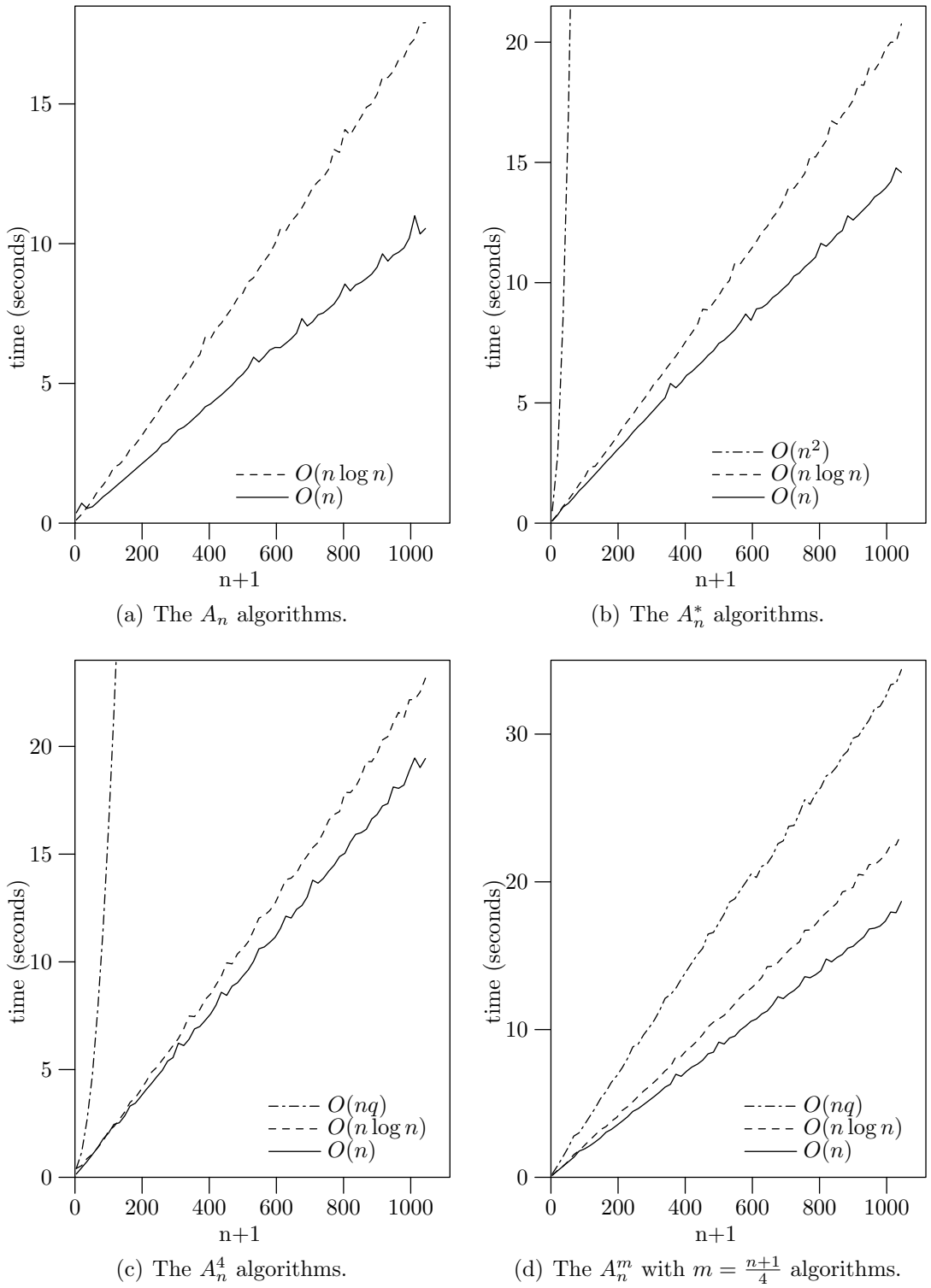
In this chapter we have derived linear time nearest point algorithms for the lattice A_n, its dual A_n^* and the family of Coxeter lattices A_n^m that *lie between* A_n and A_n^*. We showed how A_n can be constructed as the intersection of the integer lattice $\mathbb{Z}^{n+1}$ with the subspace, denoted H, that is orthogonal to the all ones vector $\mathbf{1}$, and how A_n^* can be constructed as the projection of $\mathbb{Z}^{n+1}$ into this subspace.

We discovered that the Voronoi cell of the lattice A_n is equivalent to the convex polytope that results from projecting the $n+1$ dimensional hypercube into H. It follows that the Voronoi cell of the Coxeter lattices and A_n^* is a subset of the projected hypercube. We use this fact in Lemma 3.1 to show that a nearest lattice point in A_n, A_n^* or A_n^m is inside a particular set, the we denoted by S, that is of size $n+1$.

Only one lattice point in S is also in A_n and therefore a nearest point in A_n results from locating this single point. We derived two algorithms to locate this point, Algorithms 3.1 and 3.2. The first algorithm requires sorting $n+1$ elements and therefore has complexity $O(n \log n)$. The second algorithm uses the **Rivest-Tarjan selection algorithm** and requires at most a linear number of operations.

Every element in S is a lattice point in A_n^* and we developed an efficient recursion that can test each lattice point in S in order to find the nearest point. This results in Algorithm 3.3 that requires $O(n \log n)$ arithmetic operations due to a sorting operation. We show in Lemma 3.2 and Theorem 3.3 that *only some* of the points in S can be nearest lattice points and that a full sort is not required, only a partial **bucket sort** is required. This leads to Algorithms 3.4 and 3.6 that require at most $O(n)$ operations.

For the Coxeter lattices A_n^m we can again produce an algorithm that uses a sort to compute a nearest point in $O(n \log n)$ operations. However, by employing both the Rivest-Tarjan selection algorithm and a bucket sort we showed how a nearest lattice point could be found in only $O(n)$ operations.



(a) The A_n algorithms.

(b) The A_n^* algorithms.

(c) The A_n^4 algorithms.

(d) The A_n^m with $m = \frac{n+1}{4}$ algorithms.

FIGURE 3.1: Computation time in seconds for 10^5 trials for the nearest point algorithms for A_n, A_n^* and A_n^m.

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
$\mathbf{z} = \langle \mathbf{y} \rangle$
$\alpha = \mathbf{z} \cdot \mathbf{1}$
$\beta = \mathbf{z} \cdot \mathbf{z}$
$\boldsymbol{bucket} = \mathbf{0}$
for $t = 1$ to $n + 1$ **do**
    $i = n + 1 - (n + 1) \lfloor z_t + 0.5 \rfloor$
    $link_t = bucket_i$
    $bucket_i = t$
$D = \beta - \frac{\alpha^2}{n+1}$
$k = 0$
for $i = 1$ to $n + 1$ **do**
    $t = bucket_i$
    while $t \neq 0$ **do**
        $\alpha = \alpha - 1$
        $\beta = \beta - 2z_t + 1$
        $t = link_t$
    if $\beta - \frac{\alpha^2}{n+1} < D$ **then**
        $D = \beta - \frac{\alpha^2}{n+1}$
        $k = i$
$\mathbf{u} = \lceil \mathbf{y} \rfloor$
for $i = 1$ to k **do**
    $t = bucket_i$
    while $t \neq 0$ **do**
        $u_t = u_t + 1$
        $t = link_t$
$\mathbf{x} = \mathbf{Q}\mathbf{u}$
return x

Algorithm 3.6: Algorithm to find a nearest lattice point in A_n^* to $\mathbf{y} \in \mathbb{R}^{n+1}$ that requires $O(n)$ arithmetic operations. This pseudocode indicates how to implement the algorithm in practice using two arrays.

Input: $\mathbf{y} \in \mathbb{R}^{n+1}$
$\mathbf{z} = \mathbf{y} - \lceil \mathbf{y} \rfloor$
for $j = 1$ **to** q **do** $B_j = \emptyset$
for $i = 1$ **to** $n+1$ **do**
  $j = q - \lfloor q(z_i + 1/2) \rfloor$
  $B_j = B_j \cup i$
$\mathbf{u} = \lceil \mathbf{y} \rfloor$
$\alpha = \mathbf{z} \cdot \mathbf{1}$
$\beta = \mathbf{z} \cdot \mathbf{z}$
$\gamma = \mathbf{u} \cdot \mathbf{1} \operatorname{rem} m$
$k = 1$
$D = \infty$
for $j = 1$ **to** q **do**
  $g = m - \gamma$
  $p = |B_j| - (|B_j| + \gamma) \operatorname{rem} m$
  $\mathbf{b} = \text{quickpartition2}(\mathbf{z}, B_j, g, p)$
  for $i = 1$ **to** $|B_j|$ **do**
    $\alpha = \alpha - 1$
    $\beta = \beta - 2z_{b_i} + 1$
    $\gamma = (\gamma + 1) \operatorname{rem} m$
    if *(*$i = g$ **or** $i = p$*)* **and** $\beta - \alpha^2/(n+1) < D$ **then**
      $D = \beta - \alpha^2/(n+1)$
      $k^* = k$
    $k = k + 1$
  concatenate$(\mathbf{w}, \mathbf{b})$
for $i = 1$ **to** k^* **do**
  $u_{w_i} = u_{w_i} + 1$
$\mathbf{x} = \mathbf{Q}\mathbf{u}$
return $\mathbf{x}$

ALGORITHM 3.7: Algorithm to find a nearest lattice point in A_n^m to $\mathbf{y} \in \mathbb{R}^{n+1}$ that requires $O(n)$ arithmetic operations.

Input: $\mathbf{y} \in \mathbb{R}^n$
$D = \infty$
$\mathbf{g} = \mathbf{Q}\mathbf{e}_1$
for $i = 0$ **to** $q - 1$ **do**
  $[\mathbf{x}, \mathbf{u}] = \text{NearestPt}(\mathbf{y} - im\mathbf{g}, A_n) + im\mathbf{g}$
  if $\|\mathbf{x} - \mathbf{y}\| < D$ **then**
    $\mathbf{x}_{\text{NP}} = \mathbf{x}$
    $D = \|\mathbf{x} - \mathbf{y}\|$
return $\mathbf{x}_{\text{NP}}$

ALGORITHM 3.8: Nearest point algorithm for A_n^m using a set of coset representatives for A_n^m / A_n.

–Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law.

Douglas Hofstadter

4

The lattices $V_{n/m}$, $V_{n/m}^*$ and $V_{n/m}^\perp$

In this chapter we describe the lattices $V_{n/m}$, $V_{n/m}^*$ and $V_{n/m}^\perp$. The lattice $V_{n/m}^*$ will be used in polynomial phase estimation in Chapters 7, 9 and 10. We find that an excellent estimator for a polynomial phase signal of order m is given by finding a nearest lattice point in $V_{n/m}^*$. Because of these applications the primary goal of this chapter is to derive a nearest point algorithm for $V_{n/m}^*$.

Like A_n and A_n^* the lattices $V_{n/m}$, $V_{n/m}^*$ and $V_{n/m}^\perp$ can be described using *projections* and *intersections* of the integer lattice. The lattice $V_{n/m}$ is the intersection of the integer lattice with an $m+1$ dimensional subspace and the dual lattice $V_{n/m}^*$ is the projection of the integer lattice into this subspace. The lattice $V_{n/m}^\perp$ is the $m+1$ dimensional lattice formed by intersecting the integer lattice with the orthogonal subspace (the complementary space) and for this reason we call $V_{n/m}^\perp$ the **complementary lattice** of $V_{n/m}$.

After defining the lattices using projections and intersections we derive generator matrices for both $V_{n/m}^\perp$ and $V_{n/m}^*$ in Section 4.2. These derivations are aided by knowledge of two special families of discrete polynomials, the **integer valued polynomials** and the **discrete Legendre polynomials** and we define these at the start of the section. Using properties of these polynomials we find a closed form formula for the determinant of $V_{n/m}^\perp$, $V_{n/m}$ and $V_{n/m}^*$ and also the order of the quotient group $V_{n/m}^*/V_{n/m}$. The integer valued polynomials will also play a key role in Chapter 7 for describing the phenomenon of **aliasing** that occurs when polynomial phase signals are sampled.

In Section 4.3 we consider how to compute a nearest point in $V_{n/m}^*$. We take an approach similar to Algorithm 2.1, that is we compute a nearest point for each coset in the quotient $V_{n/m}^*/V_{n/m}$ and return the closest point found over all cosets. We show how this yields a nearest point algorithm that requires a number of operations that is polynomial in the dimension of the lattice n. This is a substantial improvement over the fastest algorithms for *random* lattices, such as the sphere decoder, that

require a number of operations that is exponential in the dimension of the lattice (see in particular Jalden and Ottersten [2005] who show that even the *expected* complexity of the sphere decoder is exponential in n). The algorithm requires a method for enumerating a set of coset representatives for the quotient $V^*_{n/m}/V_{n/m}$ and we describe a convenient method that is based on Theorem 2.2 and also the Hermite decomposition that was introduced in Section 2.4.

4.1 Definition of $V_{n/m}$, $V^*_{n/m}$ and $V^\perp_{n/m}$

We will make use of the following notation for mapping polynomials to vectors. Given a polynomial $g(x) = a_0 + a_1 x + \cdots + a_m x^m$ we use $\mathrm{vec}(g)$ to denote the column vector generated by evaluating $g(x)$ at the integers $x = 1, 2, \ldots, n+m+1$. That is,

$$\mathrm{vec}(g) = \begin{bmatrix} g(1) & g(2) & g(3) & \ldots & g(N) \end{bmatrix}^\dagger$$

where $N = n+m+1$. We use $\mathrm{coef}(g)$ to denote the column vector of length $m+1$ containing the coefficients of g, that is,

$$\mathrm{coef}(g) = \begin{bmatrix} a_0 & a_1 & a_2 & \ldots & a_m \end{bmatrix}^\dagger .$$

Let $H^\perp$ be the subspace of dimension $m+1$ spanned by the $m+1$ vectors $\mathrm{vec}(x^0), \mathrm{vec}(x^1), \ldots, \mathrm{vec}(x^m)$ and let H be the n-dimensional subspace orthogonal to $H^\perp$. The subspace $H^\perp$ is sometimes called the (discrete) **space of polynomials** because it describes the vector space spanned by the $\mathrm{vec}(\cdot)$ of any set of $m+1$ linearly independent polynomials of order m. We define the lattice $V^\perp_{n/m}$ as the intersection of the N-dimensional integer lattice with $H^\perp$, that is

$$V^\perp_{n/m} = \mathbb{Z}^N \cap H^\perp. \tag{4.1.1}$$

Clearly $V^\perp_{n/m}$ is a lattice in $m+1$ dimensions because each of the $\mathrm{vec}(x^k) \in \mathbb{Z}^N$. Noticing that $\mathbb{Z}^N$ is unimodular it follows from Corollary 2.2 that

$$V_{n/m} = \mathbb{Z}^N \cap H \tag{4.1.2}$$

is an n dimensional lattice and that

$$\det V_{n/m} = \det V^\perp_{n/m}. \tag{4.1.3}$$

Let p denote the orthogonal projection onto H, then projecting the integer lattice orthogonally into H produces the n-dimensional lattice

$$V^*_{n/m} = p\mathbb{Z}^N \tag{4.1.4}$$

that is the dual of $V_{n/m}$ and $\det V^*_{n/m} = \big(\det V_{n/m}\big)^{-1}$. When $m = 0$ we find that these lattices are equal to $\mathbb{Z}$, A_n and A^*_n, that is

$$V^\perp_{n/0} = \mathbb{Z} \qquad V_{n/0} = A_n \qquad V^*_{n/0} = A^*_n. \tag{4.1.5}$$

The $V_{n/m}$ family of lattices can therefore be seen as an extension of the A_n lattices.

4.2 Generator matrices

We will describe generator matrices for $V_{n/m}^{\perp}$ and $V_{n/m}^{*}$ in this section. It is also reasonably easy to find a generator for $V_{n/m}$ but we do not require it in this thesis. We will have use of two important families of discrete polynomials.

Definition 4.1. (**Integer valued polynomials**)
The integer valued polynomial of order k, denoted by p_k, is given as

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

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

Definition 4.2. (**Discrete Legendre polynomials**)
The discrete Legendre polynomial of order k, denoted by l_k, is given as

$$l_k(x) = \frac{k!}{\binom{2k}{k}} \sum_{s=0}^{k} (-1)^{s+k} \binom{s+k}{s} \binom{N-s-1}{N-k-1} p_s(x-1).$$

where p_s is the integer valued polynomial of order s.

The discrete Legendre polynomials (as we have defined them) are monic, i.e. the coefficient of the highest order term is equal to one, and the kth discrete Legendre polynomial l_k has order k. The l_k are also orthogonal in the sense that

$$l_k \cdot l_j = \sum_{x=1}^{N} l_k(x) l_j(x) = \begin{cases} 0 & , k \neq j \\ (k!)^2 \binom{2k}{k}^{-1} \binom{N+k}{2k+1} & , k = j \end{cases} \tag{4.2.1}$$

where $l_k \cdot l_j = \mathrm{vec}(l_k) \cdot \mathrm{vec}(l_j)$ is the (discrete) polynomial inner product [Szego, 1975; Eisinberg et al., 2001; Eisinberg and Fedele, 2007].

It is clear that the $\mathrm{vec}(x^k)$, the $\mathrm{vec}(l_k)$ and the $\mathrm{vec}(p_k)$ for $k = 0, 1, \dots, m$ all span the space of polynomials $H^{\perp}$. Define the $N \times (m+1)$ matrices $\mathbf{X}$, $\mathbf{L}$ and $\mathbf{P}$ to be the matrices with column vectors given by the the $\mathrm{vec}(x^k)$, the $\mathrm{vec}(l_k)$ and the $\mathrm{vec}(p_k)$ respectively. For example $\mathbf{X}$ is the Vandermonde matrix

$$\mathbf{X} = \begin{bmatrix} \mathrm{vec}(x^0) & \mathrm{vec}(x^1) & \dots & \mathrm{vec}(x^m) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & 2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & N & N^2 & \cdots & N^m \end{bmatrix}. \tag{4.2.2}$$

Notice that the orthogonality of the discrete Legendre polynomials implies that the columns of $\mathbf{L}$ are orthogonal.

Also define the $(m+1) \times (m+1)$ matrices $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{P}}$ to be the matrices with column vectors given by the coefficient vectors $\mathrm{coef}(x^k)$, the $\mathrm{coef}(l_k)$ and the

coef(p_k) respectively. For example $\boldsymbol{\mathcal{X}}$ is the identity matrix. Also, when $m = 3$, the first four integer valued polynomial are

$$\begin{aligned} p_0 &= 1 \\ p_1 &= n \\ p_2 &= \frac{x^2}{2} - \frac{x}{2} \\ p_3 &= \frac{n^3}{6} - \frac{n^2}{2} + \frac{n}{3} \end{aligned}$$

and therefore the matrix

$$\boldsymbol{\mathcal{P}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1/2 & 1/3 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 0 & 1/6 \end{bmatrix}. \tag{4.2.3}$$

It is easy to see that $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{P}}$ are upper triangular. Because the discrete Legendre polynomials are monic the diagonal elements of $\boldsymbol{\mathcal{L}}$ are all 1 and $\det(\boldsymbol{\mathcal{L}}) = 1$. The diagonal elements of $\boldsymbol{\mathcal{P}}$ are $1, 1, 1/2, 1/3!, \ldots, 1/m!$ and therefore

$$\det(\boldsymbol{\mathcal{P}}) = \prod_{k=0}^{m} \frac{1}{k!}. \tag{4.2.4}$$

It is also not hard to check that the following relationships between $\mathbf{X}$, $\mathbf{L}$ and $\mathbf{P}$ hold

$$\mathbf{P} = \mathbf{X}\boldsymbol{\mathcal{P}} \tag{4.2.5}$$

$$\mathbf{L} = \mathbf{X}\boldsymbol{\mathcal{L}} \tag{4.2.6}$$

$$\mathbf{P} = \mathbf{L}\boldsymbol{\mathcal{L}}^{-1}\boldsymbol{\mathcal{P}}. \tag{4.2.7}$$

4.2.1 A generator for $V^\perp_{n/m}$

Let $\mathcal{Z}$ denote the set of polynomials of order at most m that are integer valued when evaluated at integers. That is, $\mathcal{Z}$ contains all polynomials p, of order at most m, such that $p(x)$ is an integer whenever x is an integer. Then

$$V^\perp_{n/m} = \mathbb{Z}^N \cap H^\perp = \{\text{vec}(p) \mid p \in \mathcal{Z}\}. \tag{4.2.8}$$

Lemma 4.1. [Cahen and Chabert, 1997, p. 2] *The integer valued polynomials p_k for $k = 0, 1, \ldots, m$ are an integer basis for $\mathcal{Z}$. That is, every element of $\mathcal{Z}$ can be uniquely written as*

$$c_0 p_0 + c_1 p_1 + \cdots + c_m p_m \tag{4.2.9}$$

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

Proof. Note that $x(x-1)(x-2)\dots(x-k+1)$ is divisible by all integers $1, 2, \dots, k$ and so p_k takes integer values for all $x \in \mathbb{Z}$. Then any polynomial generated as in (4.2.9) is an element in $\mathcal{Z}$. The proof proceeds by induction. Consider any polynomial $f \in \mathcal{Z}$. Let $d < n$ and assume that $c_i \in \mathbb{Z}$ for all $i \leq d$. Let g be the polynomial

$$g = f - \sum_{k=0}^{d} c_k p_k$$

and note that $g \in \mathcal{Z}$. Then

$$g = c_{d+1}p_{d+1} + \cdots + c_m p_m. \tag{4.2.10}$$

Now $p_{d+1}(d+1) = 1$ and $p_k(d+1) = 0$ for all $k > d+1$. Then $g(d+1) = c_{d+1}p_{d+1}(d+1)$ and therefore $c_{d+1} = g(d+1) \in \mathbb{Z}$. The proof follows by induction because $f(0) = c_0 \in \mathbb{Z}$. □

From the above lemma and (4.2.8) we see that

$$\begin{aligned} V_{n/m}^{\perp} &= \{\operatorname{vec}(c_0 p_0 + c_1 p_1 + \cdots + c_m p_m) \mid c_i \in \mathbb{Z}\} \\ &= \{c_0 \operatorname{vec}(p_0) + c_1 \operatorname{vec}(p_1) + \cdots + c_m \operatorname{vec}(p_m) \mid c_i \in \mathbb{Z}\} \end{aligned}$$

and from the definition of the matrix $\mathbf{P}$ we have

$$V_{n/m}^{\perp} = \{\mathbf{P}\mathbf{c} \mid \mathbf{c} \in \mathbb{Z}^{m+1}\}$$

and it follows that $\mathbf{P}$ is a generator matrix for $V_{n/m}^{\perp}$. We can also find the following formula for the determinant

$$\begin{aligned} \det(V_{n/m}^{\perp}) &= \det(\mathbf{P}^{\dagger}\mathbf{P}) \\ &= \det((\mathbf{L}\boldsymbol{\mathcal{L}}^{-1}\boldsymbol{\mathcal{P}})^{\dagger}\mathbf{L}\boldsymbol{\mathcal{L}}^{-1}\boldsymbol{\mathcal{P}}) \\ &= \det(\mathbf{L}^{\dagger}\mathbf{L}) \prod_{k=0}^{m} \frac{1}{(k!)^2} \\ &= \prod_{k=0}^{m} \binom{2k}{k}^{-1} \binom{N+k}{2k+1}. \end{aligned} \tag{4.2.11}$$

This follows from (4.2.7), (4.2.4) and the fact that the columns of $\mathbf{L}$ are orthogonal. This immediately gives formula for the determinant of $V_{n/m}$ and $V_{n/m}^{*}$ because $\det V_{n/m} = \det V_{n/m}^{\perp}$ and $\det V_{n/m}^{*} = (\det V_{n/m}^{\perp})^{-1}$. We also obtain a formula for the order of the dual quotient group because $|V_{n/m}^{*}/V_{n/m}| = \det V_{n/m}$ (see Proposition 2.3).

4.2.2 A generator for $V_{n/m}^{*}$

A generator matrix for $V_{n/m}^{*}$ is easily derived as any n columns of the $N \times N$ orthogonal projection matrix

$$\mathbf{Q} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\dagger}\mathbf{X})^{-1}\mathbf{X}^{\dagger}. \tag{4.2.12}$$

4.3 Computing the nearest point in $V^*_{n/m}$

When $m = 0$ a nearest point in $V^*_{n/0} = A^*_n$ can be computed in $O(n)$ operations using the algorithm described in Chapter 3. When $m = 1$ a nearest point in $V^*_{n/1}$ can be computed using a 'line searching' algorithm described in McKilliam et al. [2010a]. We will not describe this algorithm here and instead we describe an algorithm that works for all m in a number of operations that is polynomial in the dimension of the lattice n. The algorithm is a small modification of Algorithm 2.1 (page 31). That is, we compute a nearest point for each coset in the quotient $V^*_{n/m}/V_{n/m}$ and return the overall closest point found. Direct application of Algorithm 2.1 requires a nearest lattice point algorithm for $V_{n/m}$, something we do not have. However, we will show that rather than computing a nearest point in $V_{n/m}$ it is sufficient to instead compute a nearest point in the lattice $\mathbb{Z}^N$, which is easily achieved by rounding each element in a vector to its nearest integer.

Let C be a set of coset representatives for $V^*_{n/m}/V_{n/m}$. We will describe a convenient method for enumerating the set C in the next section. For now note that, given C, we have

$$V^*_{n/m} = \bigcup_{\mathbf{g}\in C} V_{n/m} + \mathbf{g} = V_{n/m} + C$$

where $V_{n/m}+C$ denotes the Minkowski sum of $V_{n/m}$ and C, that is $V_{n/m}+C$ contains the elements $\mathbf{x}+\mathbf{y}$ for all $\mathbf{x} \in V_{n/m}$ and $\mathbf{y} \in C$. As $V_{n/m} = \mathbb{Z}^N \cap H$ we can write

$$V^*_{n/m} = (\mathbb{Z}^N \cap H) + C = (\mathbb{Z}^N + C) \cap H. \tag{4.3.1}$$

The set $\mathbb{Z}^N + C$ is a lattice in N dimensions. Its intersection with H is the lattice $V^*_{n/m}$ and it is not hard to show that its intersection with $H^\perp$ is the dual lattice of $V^\perp_{n/m}$. It is convenient to visualise $\mathbb{Z}^N + C$ as being built up of *layers* of $V^*_{n/m}$ that lie in hyperplanes parallel to H. In this context it is not surprising that the nearest point in $V^*_{n/m}$ is closely related to the nearest point in $\mathbb{Z}^N + C$ as the next theorem will show.

Theorem 4.1. *Let $\mathbf{y} \in \mathbb{R}^N$ be a point in H. The nearest point to $\mathbf{y}$ in $\mathbb{Z}^N + C$ is equal to the nearest point to $\mathbf{y}$ in $V^*_{n/m}$. That is,*

$$\text{NearestPt}(\mathbf{y}, V^*_{n/m}) = \text{NearestPt}(\mathbf{y}, \mathbb{Z}^N + C).$$

Proof. From (4.3.1) we see that $V^*_{n/m}$ contains all of those points from $\mathbb{Z}^N + C$ that are contained in H. Therefore, to prove the theorem, it is sufficient to show that the nearest point to $\mathbf{y}$ in $\mathbb{Z}^N + C$ is contained in H. Let $\mathbf{x}$ be the nearest point in $\mathbb{Z}^N + C$ to $\mathbf{y}$ and assume that $\mathbf{x} \notin H$. Let $\mathbf{x}' = \mathbf{Q}\mathbf{x}$ be the orthogonal projection of $\mathbf{x}$ into H. It follows from the definition of $V^*_{n/m}$ and $\mathbb{Z}^N + C$ that $\mathbf{x}'$ is also a lattice point in $\mathbb{Z}^N + C$. We can write $\mathbf{x} = \mathbf{x}' + \mathbf{t}$ where $\mathbf{t}$ is a nonzero vector in $H^\perp$ and then

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}' + \mathbf{t} - \mathbf{y}\|^2 = \|\mathbf{x}' - \mathbf{y}\|^2 + \|\mathbf{t}\|^2.$$

because $\mathbf{y}$ and $\mathbf{x}'$ are orthogonal to $\mathbf{t}$. As $\mathbf{t}$ is nonzero then

$$\|\mathbf{x} - \mathbf{y}\|^2 > \|\mathbf{x}' - \mathbf{y}\|^2$$

contradicting that $\mathbf{x}$ is a nearest point to $\mathbf{y}$. □

Algorithm 4.1 follows as a result of the previous theorem. The point $\mathbf{y}$ is projected orthogonally into H on line 1. The remainder of the algorithm operates similarly to Algorithm 2.1 except for line 6 which computes a nearest point in $\mathbb{Z}^N$ rather that the lattice $V_{n/m}$. The number of cosets is

$$|C| = |V^*_{n/m}/V_{n/m}| = \det V_{n/m} = \prod_{k=0}^{m} \binom{2k}{k}^{-1} \binom{N+k}{2k+1} = O\left(N^{(m+1)^2}\right)$$

and the coset representatives can be computed in an amount of time that is proportional to $|C|$ as we will show in the next section. For each coset representative $O(N)$ operations are required so the algorithm requires $O(N^{(m+1)^2+1})$ operations in total. Although polynomial in N, for large N this algorithm is very slow, even when m is quite small. For this reason we will consider other, approximate, approaches to computing the nearest point in Chapter 7.

Input: $\mathbf{y} \in \mathbb{R}^N$
1 $\mathbf{y}' = \mathbf{Q}\mathbf{y}$
2 $D = \infty$
3 **foreach** $\mathbf{g} \in C$ **do**
4 $\quad \mathbf{x} = \lceil \mathbf{y}' - \mathbf{g} \rfloor$
5 $\quad$ **if** $\|\mathbf{x} - \mathbf{y}'\| < D$ **then**
6 $\quad\quad \mathbf{x}_{\text{NP}} = \mathbf{x} + \mathbf{g}$
7 $\quad\quad D = \|\mathbf{x} - \mathbf{y}'\|$
8 **return** $\mathbf{x}_{NP}$

Algorithm 4.1: Computing the nearest point in $V^*_{n/m}$ using a set of coset representatives for the lattice $V^*_{n/m}/V_{n/m}$ given by the set C.

4.4 Coset representatives for $V^*_{n/m}/V_{n/m}$

In this section we will describe some efficient methods for enumerating a set of coset representatives for the dual quotient group $V^*_{n/m}/V_{n/m}$. We will have use of the dual lattice of $V^{\perp}_{n/m}$ which we could denote by $V^{\perp *}_{n/m}$ but to avoid the awkward superscript we will simply denote it by D. We will make use of Theorem 2.2 (page 28) that provides a connection between the dual quotient $D/V^{\perp}_{n/m}$ and the dual quotient $V^*_{n/m}/V_{n/m}$. We first consider computing a set of coset representatives for $D/V^{\perp}_{n/m}$. We showed in Section 4.2.1 that a generator for $V^{\perp}_{n/m}$ is the N by $m+1$ matrix $\mathbf{P}$ constructed using the integer valued polynomials. A generator for the dual lattice D is then given by taking the transpose of the pseudoinverse of $\mathbf{P}$, i.e. $(\mathbf{P}^+)^\dagger = \mathbf{P}(\mathbf{P}^\dagger\mathbf{P})^{-1}$ (see Section 2.5), and we will denote this matrix by $\mathbf{D}$. Using the approach suggested in Section 2.4.1 we can enumerate the coset representatives by taking the Hermite decomposition of the $m+1$ by $m+1$ integral matrix

$$\mathbf{D}^+\mathbf{P} = \left(\mathbf{P}^+)^\dagger\right)^+ \mathbf{P} = \mathbf{P}^\dagger\mathbf{P}$$

which is a Gram matrix for $V^\perp_{n/m}$. In Proposition 2.2 we showed how the Hermite decomposition can be used to compute a set K of vectors from $\mathbb{Z}^{m+1}$ such that $\mathbf{D}K$ is a set of coset representatives for the quotient $D/V^\perp_{n/m}$. Using Theorem 2.2 we will show how this set of coset representatives can be converted into a set of coset representatives for $V^*_{n/m}/V_{n/m}$.

At this stage it is not immediately clear how we apply Theorem 2.2 because using this theorem requires a set of coset representatives that have been constructed by projecting vectors from $\mathbb{Z}^N$ into $H^\perp$. We will show how the coset representatives $\mathbf{D}K$ can be considered in this way. First consider the $m+1$ by $m+1$ matrix $\mathbf{S}$ that consists of the first $m+1$ columns of $\mathbf{P}^\dagger$. Using the definition of the integer valued polynomials it is easy to check that $\mathbf{S}$ is integral, upper triangular and with diagonal elements all equal to 1. So $\det \mathbf{S} = 1$ and $\mathbf{S}$ is unimodular and its inverse is therefore also integral and unimodular. Now construct the N by $m+1$ matrix $\mathbf{B}$ by appending an n by $m+1$ matrix of zeros below the inverse of $\mathbf{S}$, that is

$$\mathbf{B} = \begin{bmatrix} \mathbf{S}^{-1} \\ \mathbf{0}_{n,m+1} \end{bmatrix} \tag{4.4.1}$$

where $\mathbf{0}_{n,m+1}$ denotes the n by $m+1$ matrix of zeros. Notice that because $\mathbf{S}^{-1}$ is integral then $\mathbf{B}$ is integral.

It is easy to see that $\mathbf{P}^\dagger\mathbf{B}$ is the $m+1$ by $m+1$ identity matrix and therefore the sets

$$\mathbf{D}K = \mathbf{D}\mathbf{P}^\dagger\mathbf{B}K$$

are equal. So the set $\mathbf{D}\mathbf{P}^\dagger\mathbf{B}K$ is a set of coset representatives for $D/V^\perp_{n/m}$. Now observe that the matrix $\mathbf{D}\mathbf{P}^\dagger = \mathbf{P}(\mathbf{P}^\dagger\mathbf{P})^{-1}\mathbf{P}^\dagger$ is an orthogonal projection into $H^\perp$ and therefore we may write

$$\mathbf{D}K = \mathbf{P}(\mathbf{P}^\dagger\mathbf{P})^{-1}\mathbf{P}^\dagger\mathbf{B}K = p^\perp\mathbf{B}K$$

and it follows that $p^\perp\mathbf{B}K$ is a set of coset representatives for $D/V^\perp_{n/m}$. Because $\mathbf{B}$ is integral the set $\mathbf{B}K$ contains vectors from $\mathbb{Z}^N$. We are now in a position to utilise Theorem 2.2 from which it follows that projecting the set $\mathbf{B}K$ orthogonally into the subspace H produces a set of coset representatives for the quotient $V^*_{n/m}/V_{n/m}$. We will now summarise this process of enumerating the cosets representative of $V^*_{n/m}/V_{n/m}$.

1. Compute the $m+1$ by $m+1$ integral Gram matrix $\mathbf{P}^\dagger\mathbf{P}$ consisting of inner products of the integer valued polynomials and also compute its Hermite decomposition consisting of an upper triangular integral matrix $\mathbf{R}$ and unimodular matrix $\mathbf{U}$. So we obtain

$$\mathbf{P}^\dagger\mathbf{P} = \begin{bmatrix} p_0 \cdot p_0 & p_0 \cdot p_1 & \cdots & p_0 \cdot p_m \\ p_1 \cdot p_0 & p_1 \cdot p_1 & \cdots & p_1 \cdot p_m \\ \vdots & \vdots & \ddots & \vdots \\ p_m \cdot p_0 & p_m \cdot p_1 & \cdots & p_m \cdot p_m \end{bmatrix} = \mathbf{U}\mathbf{R}.$$

2. Compute the set K of vectors from $\mathbb{Z}^{m+1}$ consisting of the points $\mathbf{Ut}$ for all $\mathbf{t}$ with elements $t_i = 0, 1, \ldots, r_{i,i} - 1$ where $r_{i,i}$ is the ith diagonal element of the matrix $\mathbf{R}$.

3. Compute the $m+1$ by $m+1$ submatrix $\mathbf{S}$ consisting of the first $m+1$ columns of $\mathbf{P}^\dagger$ and then compute the N by $m+1$ matrix $\mathbf{B}$ by appending a matrix of zeros to the inverse of $\mathbf{S}$ according to (4.4.1).

4. A set of coset representatives for $V^*_{n/m}/V_{n/m}$ is now given by the projecting the set $\mathbf{B}K$ orthogonally into H, i.e. the set $p\mathbf{B}K$.

For the purpose of Algorithm 4.1 it is not necessary to store the entire set K or the set of coset representatives $p\mathbf{B}K$. It is only necessary to store the $m+1$ by $m+1$ matrices $\mathbf{U}$ and $\mathbf{S}^{-1}$ and the $m+1$ diagonal elements of $\mathbf{R}$. Each coset representative can then be computed in turn using only $O(N)$ operations from these matrices. We will now consider a number of examples when m is small. Closed form expressions can often be found for the matrices $\mathbf{R}$ and $\mathbf{U}$ and also the matrix $\mathbf{S}^{-1}$ for some small m and in effect this gives a closed form representation for the coset representatives.

Coset representatives when $m = 0$

When $m = 0$ then lattice $V^\perp_{n/0} = \mathbf{1}\mathbb{Z}$ and the Gram matrix $\mathbf{P}^\dagger\mathbf{P}$ is the 1×1 matrix with single element equal to $N = n+1$. The Hermite decomposition is trivial so by Proposition 2.2 the coset representatives of $D/V^\perp_{n/0} = \mathbb{Z}/(n+1)\mathbb{Z}$ are the integers $0, 1, 2, \ldots, n$. The matrix $\mathbf{B}$ is the N by 1 vector with the first element equal to one and remaining elements equal to zero, i.e. the vector $\mathbf{e}_1$. So the coset representatives for $V^*_{n/0}/V_{n/0} = A^*_n/A_n$ are the projections of $0\mathbf{e}_1, 1\mathbf{e}_1, 2\mathbf{e}_1, \ldots, (n+1)\mathbf{e}_1$ into the zero-mean plane, i.e. projected orthogonally to the all ones vector $\mathbf{1}$. Explicitly the $n+1$ coset representatives are

$$k\left(\mathbf{e}_1 - \frac{\mathbf{1}}{n+1}\right)$$

for $k = 0, 1, \ldots n$. This is precisely the definition we gave in (3.3.3) when discussing some of the properties of A^*_n in Section 3.3.

Coset representatives when $m = 1$

When $m = 1$ we find that the Gram matrix of $V^\perp_{n/1}$ is given by

$$\mathbf{P}^\dagger\mathbf{P} = \begin{bmatrix} p_0 \cdot p_0 & p_0 \cdot p_1 \\ p_1 \cdot p_0 & p_1 \cdot p_1 \end{bmatrix} = \begin{bmatrix} N & \frac{1}{2}N(N-1) \\ \frac{1}{2}N(N-1) & \frac{1}{6}N(N+1)(2N+1) \end{bmatrix}.$$

In this case we can compute the Hermite decomposition by hand to obtain

$$\mathbf{P}^\dagger\mathbf{P} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2}(N-1) & 1 \end{bmatrix} \begin{bmatrix} N & \frac{1}{2}N(N-1) \\ 0 & \frac{1}{12}N(N^2-1) \end{bmatrix}$$

when N is odd and

$$\mathbf{P}^\dagger\mathbf{P} = \begin{bmatrix} 2 & 1 \\ N-1 & \frac{1}{2}N \end{bmatrix} \begin{bmatrix} \frac{1}{2}N & -\frac{1}{12}N(N^2+9N+2) \\ 0 & \frac{1}{6}N(N^2+12N-1) \end{bmatrix}$$

where N is even. The matrix $\mathbf{S}$ and its inverse are equal to

$$\mathbf{S} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{S}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

and the matrix $\mathbf{B}$ is constructed by appending an $n \times 2$ matrix of zeros to $\mathbf{S}^{-1}$. The coset representatives for $V^*_{n/1}/V_{n/1}$ can be computed readily from these matrices.

For larger m we can continue the hand calculation and obtain more closed form expressions, but it is easier not too. The Hermite decompositions for any particular n and m are easily obtained by computer using one of the many available algorithms, see for example [Cohen, 1993, Algorithm 2.4.5] or Kannan and Bachem [1979] or Micciancio and Warinschi [2001]. Moreover because the matrices are only $m+1$ by $m+1$ in size the computation is very fast.

4.5 Summary and discussion

In this chapter we have derived a number of properties of the lattices $V_{n/m}$, $V^*_{n/m}$ and $V^\perp_{n/m}$. We showed how these lattices are generated by taking intersections and projections of the integer lattice $\mathbb{Z}^{n+m+1}$ and used the results we derived in Section 2.6 to derive some relationships between these lattices. In Section 4.2 we derived generator matrices for $V^\perp_{n/m}$ and $V^*_{n/m}$ and found closed-form formulas for the determinants and the order of the dual quotient group $V^*_{n/m}/V_{n/m}$.

In Part II we will find that an accurate estimator of the coefficients of a polynomial phase signal of order m is given by computing a nearest point in the lattice $V^*_{n/m}$. For this reason the primary focus of this chapter has been to derive an algorithm to compute the nearest point in $V^*_{n/m}$. The algorithm we developed is based on Algorithm 2.1 and requires a set of coset representatives for the quotient $V^*_{n/m}/V_{n/m}$. Using ideas developed in Section 2.6 and in particular Theorem 2.2 we found a very convenient way to describe the coset representatives using the Hermite decomposition of the $m+1$ by $m+1$ Gram matrix of the lattice $V^\perp_{n/m}$. This greatly reduces the complexity of enumerating the coset representatives when n is large and m is small, which is the most common case for the estimation problems we consider later. Moreover we obtained closed form representations for the coset representatives when $m = 0$ and $m = 1$.

The nearest point algorithm requires $O(n^{(m+1)^2+1})$ operations in total which is polynomial in the lattice dimension n but is exponential in the *projection* parameter m. Unfortunately we will find that this algorithm is too slow for practical use in the later chapters and will therefore consider some approximate approaches to computing the nearest lattice point in Chapters 9 and 10. However, we think that the construction of this *exact* polynomial-time nearest point algorithm is very important because it proves that the problem of finding a nearest point in $V^*_{n/m}$ is not in

the complexity class NP-hard, which is the case for *random* lattices. In fact, for many lattice families with *known* structure, for example the famous **Barnes-Wall lattices**, the fastest known nearest point algorithms require a number of operations that increases exponentially in the dimension. Our hope is that substantially faster nearest point algorithms for $V_{n/m}^*$ exist. As we shall see in the next chapters, even fast *approximate* nearest point algorithms for $V_{n/m}^*$ would prove exceedingly useful for the estimation of polynomial phase signals.

Because the focus of this thesis is on circular statistics and polynomial phase signals we have omitted a number of interesting results about the $V_{n/m}$ and $V_{n/m}^*$ lattices that we will briefly mention now. It turns out that the lattices $V_{n/m}$ are sublattices of a known family called **Craig's difference lattices** which produces the densest known sphere packings in some dimensions (see Conway and Sloane [1998, page 222] and also Martinet [2003, page 163]). The $V_{n/m}$ lattices appear to have inherited at least some of these density properties and it is possible to show that the norm of $V_{n/m}$ is at least $2(m+1)$. This result follows quite readily from consideration of a problem in number theory called the **Prouchet-Tarry-Escott problem**. A good description of this problem is given by Hardy and Wright [2008, p. 435] and also Borwein and Ingalls [1994].

Another result we have omitted is a description of the relevant vectors for the lattice $V_{n/m}^*$. These can be described using the concept of an **obtuse superbasis** similarly to how the relevant vectors for A_n^* are described by Conway and Sloane [1992]. It is possible that combining knowledge of the relevant vectors will lead to lattice properties such as the covering radius or the inradius. It is also possible that knowledge of the relevant vectors will help in the discovery of faster nearest lattice point algorithms.

Part II

Circular statistics

—When in danger or in doubt, run in circles, scream and shout.

Herman Wouk

5 Circular statistics

In Part II of this thesis we consider the problem of estimating the mean direction of a set of circular data. We will find that an accurate estimator for the mean direction is given by computing a nearest point in the lattice A_n^*. In Part III we will extend this idea and consider the estimation of polynomial phase signals. We will construct accurate estimators for polynomial phase signals that work by computing a nearest lattice point in $V_{n/m}^*$. However, before we begin to discuss any of these topics we must understand some fundamental concepts from the field of **circular statistics**.

Circular statistics aims to describe the nature of data that is measured in angles or 2-dimensional unit vectors or complex numbers on the unit circle. Such data occur frequently in science, particularly in astronomy, biology, geology, geography and meteorology and also in engineering, particularly in communications and radar. A meteorological example is the direction of the wind, and a biological example is the direction of flight taken by a bird. The field of circular statistics is in some sense an ancient one, probably begun when mankind first started recording the motion of the sun, the moon and the stars. A thorough historical account of the subject is given by Fisher [1993, Chapter 1].

There are two seminal texts on the subject, *Directional Statistics* by Mardia and Jupp [2000] and *The Statistical Analysis of Circular Data* by Fisher [1993]. The reader wishing to a obtain a thorough background knowledge of this field is referred to either of these books. That said, this chapter contains some differences, both in presentation and content, to these books and therefore even the reader who is well versed in the circular statistics literature should not be tempted to skip this chapter.

In Section 5.2 we define **circular random variables** and their associated **probability density functions** (pdf). We consider two different ways to plot the pdf of a circular random variable, the **unwrapped distribution plot** and the **circular distribution plot**. We will make extensive use of these plots throughout this chapter.

The intuitive notions of 'mean' and 'variance' is less rigidly defined for circular

random variables than for random variables on the real line and we consider two different definitions of the mean and variance in Section 5.2. The first are called the **circular mean** and the **circular variance**. These are the most common definitions used in the literature. The second definitions we call the **unwrapped mean** and the **unwrapped variance**. These definitions are less common in the literature. However, in this thesis we will find them just as, if not more, useful. In later chapters the unwrapped mean and variance are used to describe the performance of estimators for polynomial phase signals that are based on the lattices A_n^* and $V_{n/m}^*$.

The circular mean and unwrapped mean are not equal in general. In fact, for some circular distributions the means do not even *exist*, in a way we will define. We consider these anomalies in Section 5.2.3 and we find that for a large class of sufficiently symmetric distributions the circular mean and the unwrapped mean are always defined and moreover, they are equal. We call these distributions **unimean**. It so happens that many of the circular distributions considered in the literature are unimean and that all of the circular distributions we will use in the later chapters are also unimean.

In the remaining sections we describe a number of circular distribution functions that are common in the literature. These are the **von Mises distribution**, the **projected normal distribution**, the **wrapped Gaussian distribution** and the **wrapped uniform distribution**. We describe some conditions under which these distributions are unimean. In the later chapters we will use these distributions to model noise processes.

5.1 Circular random variables and probability density functions

As the purpose of circular statistics is to describe the nature of *angles* it is common in the literature to define a circular random variable to take values on $[0, 2\pi)$ or $[-\pi, \pi)$. In this thesis we will find it more convenient to define circular random variables to take values on the interval $[-1/2, 1/2)$. So, when we refer to an *angle* we mean a real number in the interval $[-1/2, 1/2)$. This is a somewhat nonstandard but it will allow us to use notation such as $\lceil \cdot \rfloor$ for rounding and $\langle \cdot \rangle$ for the centered fractional part in a convenient way, and will also lead to close ties between circular statistics and the lattices A_n^* and $V_{n/m}^*$. We will typically write random variables with a capital, such as Y, X and Z and we will write circular random variables using the capital Greek letters Θ or Φ.

It is common in the literature to define a special *circular* probability density function f to be periodic with period 1 (or 2π) so that $f(\theta + k) = f(\theta)$ for any integer k and the integral $\int_T f(\theta) d\theta = 1$ where T is any interval of length one. We *will not* use this definition. In this thesis a circular random variable is just a random variable with pdf that has support on $[-1/2, 1/2)$. The utility of this is that we have not separated *circular* random variables from *regular* random variables in any way. Sometimes it will be convenient to think of a circular random variable as simply a random variable that returns values in $[-1/2, 1/2)$ and other times it will be more natural to think of a circular random variable as describing angles wrapped around a

circle. By not making any severe distinction between circular random variables and regular random variables we are able to switch between these two mindsets freely without any notational baggage. If we want to consider f as a periodic function we will use $f(\langle x \rangle)$ that is clearly periodic with period one for $x \in \mathbb{R}$.

So, a circular random variable comes with all the properties of a *regular* random variable. For example, if Θ is a circular random variable with pdf f then its expectation is given in the usual way by

$$E[\Theta] = \int_{-\infty}^{\infty} \theta f(\theta) d\theta = \int_{-1/2}^{1/2} \theta f(\theta) d\theta$$

and the expected value of a function $g(\Theta)$ of Θ is given in the usual way by

$$E[g(\Theta)] = \int_{-\infty}^{\infty} g(\theta) f(\theta) d\theta = \int_{-1/2}^{1/2} g(\theta) f(\theta) d\theta.$$

This leads to the usual definitions of mean and variance for a circular random variable as

$$E[\Theta] \quad \text{and} \qquad \text{var}[\Theta] = E[\Theta^2] - E[\Theta]^2.$$

A little thought must be given here. The mean $E[\Theta]$ does not necessarily correspond to the **mean direction** of Θ in the sense the reader might expect. This is because the mean $E[\Theta]$ ignores the fact that, for example, the angles -0.49 and 0.49 are actually *close* to one another on the circle. In Section 5.2 and 5.2.2 we will consider different quantities called the **circular mean** and the **unwrapped mean** that do correspond with our intuitive notion of mean direction.

In this thesis we will quite often use the mean and variance of a circular random variable i.e. $E[\Theta]$ and $\text{var}[\Theta]$. When we want one of the other notions, i.e. the circular mean and circular variance, or the unwrapped mean and unwrapped variance we will always state these names in full. This is in contrast to much of the circular statistics literature that uses the terms mean and variance to refer to the circular mean and circular variance.

Plotting a circular probability density function

We will consider two ways of plotting the pdf of a circular random variable, one called an **unwrapped distribution plot** and another called a **circular distribution plot**. Both of these plots are displayed in Figure 5.1. On the left is the unwrapped distribution plot. This is the *usual* way to plot a pdf on the real line. The value of the pdf is displayed on the vertical axis and the pdf takes nonzero values only on the interval $[-1/2, 1/2)$. It is important to remember that $-1/2$ and $1/2$ are *connected* and this notion is lost in the unwrapped distribution plot. This problem is amended by the circular distribution plot displayed on the right. Here the value of the pdf is given by the distance of the curve from the origin. In both plots the two dotted lines display the minimum and maximum values of the pdf.

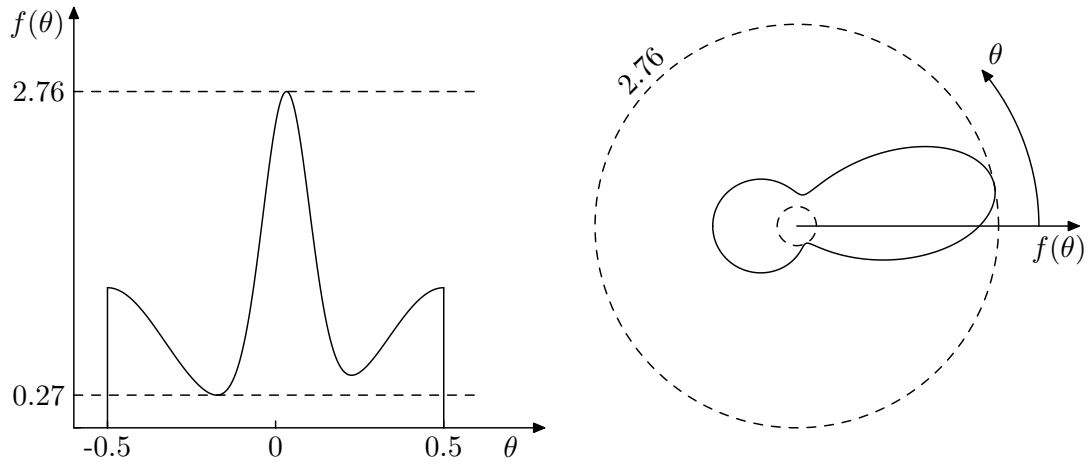


FIGURE 5.1: Two ways to plot a circular distribution. On the left is the unwrapped distribution plot and on the right the circular distribution plot. The pdf in this figure is bimodal.

5.2 The 'mean' and 'variance' of a circular random variable

The intuitive notion of a 'mean' and a 'variance' for a circular random variable is not as neatly defined as it is for regular random variables on the real line. To see why this might be the case, consider Figure 5.2. Here we have plotted 100 data points on a circle. The plot on the left has points bunched around 1/6 and we would likely conclude that the *average* or *mean* direction of the data points is about 1/6. However, the plot on the right shows points that are roughly uniformly spread around the circle. What is the mean direction of these data points? Should we conclude that these points have no mean? We will consider two different definitions, the **circular mean** and the **unwrapped mean** that provide some answers to these questions. We will also consider analogous notions of *variance* that measure the *spread* of the data points, these are the **circular variance** and **unwrapped variance**.

5.2.1 The circular mean and circular variance

Given a circular random variable with pdf $f(\theta)$ the most common analogue of 'mean' and 'variance' in the literature is the **circular mean** given by

$$\mu_{\text{circ}} = \frac{\angle c}{2\pi}$$

and the **circular variance** given by

$$\nu = 1 - |c|$$

where c is a complex number given by the integral

$$c = \int_{-1/2}^{1/2} e^{2\pi j\theta} f(\theta) d\theta, \tag{5.2.1}$$

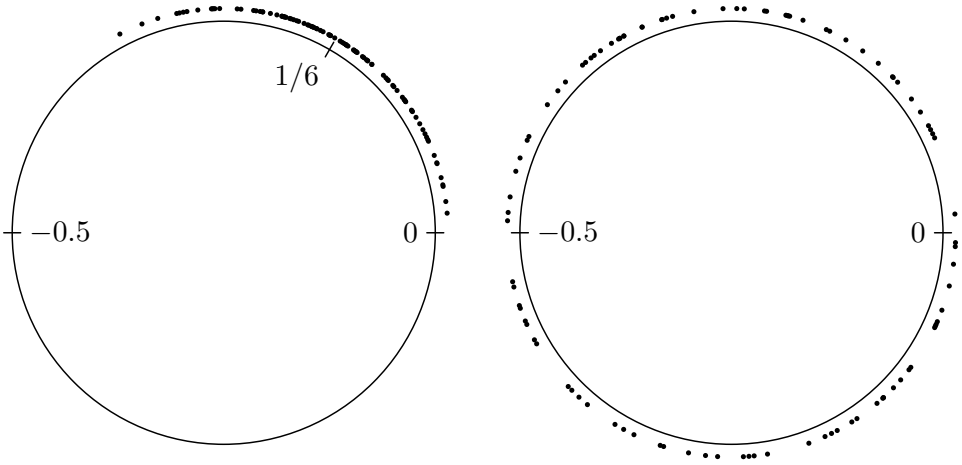


Figure 5.2: Two plots of 100 data points on a circle. For the plot on the left the points are bunched around 1/6 and we would likely conclude that the points have a mean direction of 1/6. However, the points on the right the right appear roughly uniformly spread and we would likely conclude that the points have no clear mean direction.

and $j = \sqrt{-1}$ and $\angle c$ and $|c|$ are respectively the complex argument and the magnitude of c. It is interesting to consider the case when $c = 0$. In this case the circular variance is equal to 1 but the circular mean is not well defined because we are not sure what the complex argument of zero should be. In this thesis we will simply say that the distribution *has no circular mean* when $c = 0$.

For example, consider again the uniformly spaced data points displayed on the right in Figure 5.2. These points have been generated using the **circular uniform distribution** which has pdf as displayed in Figure 5.3. From the symmetry of this distribution it can be seen that the value of c given by the integral from (5.2.1) will be zero and we therefore conclude that this distribution has *no circular mean.* This result conforms well with our intuition because we were hesitant to prescribe a circular mean to the uniformly spread data points.

The circular uniform distribution is not the only case where the circular mean is undefined. Consider, for example, the bimodal pdf displayed in Figure 5.4. The symmetry of this bimodal pdf clearly ensures that $c = 0$. It is tempting to conclude that this distribution has *two* circular means, and this definition is potentially possible. However, in this thesis we will say that this distribution has *no circular mean.* To stress this point we make the following formal definition of the circular mean.

Definition 5.1. (**Circular mean**) *Let Θ be a circular random variable with probability density function f and let c be given by the integral from* (5.2.1)*. Then Θ has circular mean equal to $\mu_{circ} = \frac{\angle c}{2\pi}$ if and only if $c \neq 0$. Otherwise, if $c = 0$, we say that Θ has no circular mean.*

5.2.2 The unwrapped mean and unwrapped variance

Alternatives to the circular mean and variance are the **unwrapped mean** and **unwrapped variance**. Before we define these note that if Θ is a circular random

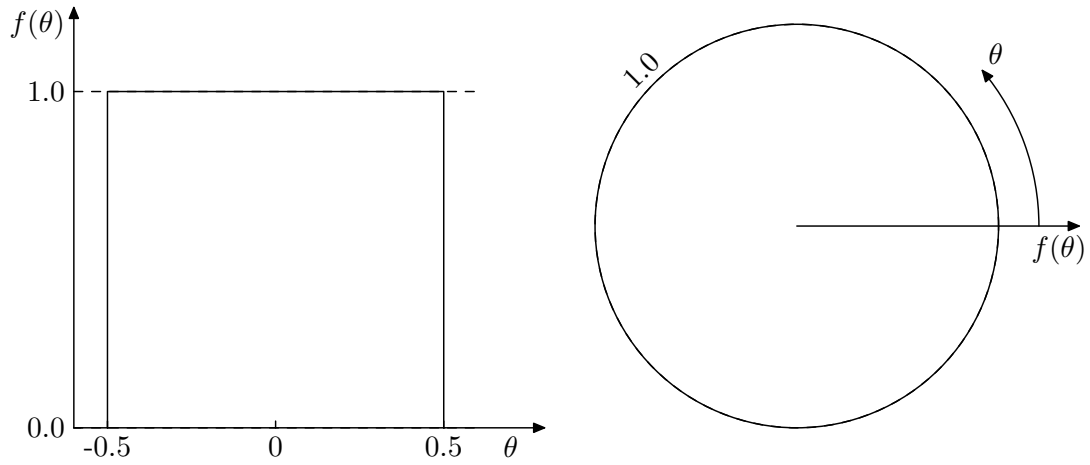


FIGURE 5.3: The circular uniform distribution which has no circular mean and no unwrapped mean.

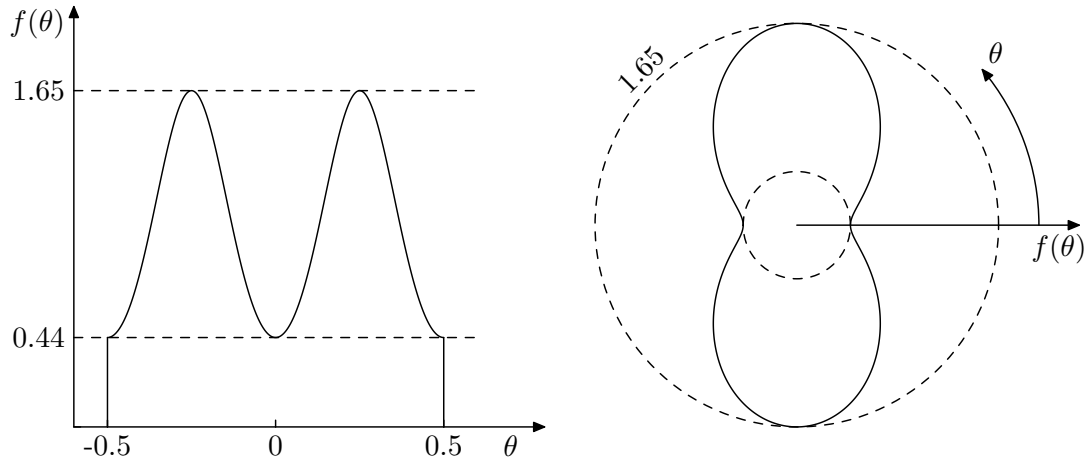


FIGURE 5.4: A bimodal distribution with no circular mean and no unwrapped mean.

variable then for some $\phi \in \mathbb{R}$ the random variable given by $\langle \Theta + \phi \rangle$ is a *rotated* version of Θ. That is, if $f(\theta)$ is the pdf of Θ then $f(\langle \theta - \phi \rangle)$ is the pdf of $\langle \Theta + \phi \rangle$ and if we displayed these pdfs using a circular distribution plot we would see that $f(\theta - \phi)$ is a *rotated* version of $f(\theta)$. As an example, consider Figure 5.5 which displays the pdf of a random variable Θ and the rotated random variable $\langle \Theta + 1/4 \rangle$.

Let Θ be a circular random variable. The unwrapped mean of Θ is defined as the angle μ_{unwrap} such that the rotated random variable $\langle \Theta - \mu_{\text{unwrap}} \rangle$ has minimum variance (in the usual sense), that is, the unwrapped mean is defined as

$$\begin{aligned} \mu_{\text{unwrap}} &= \arg \min_{\mu \in [-1/2, 1/2)} \operatorname{var} \langle \Theta - \mu \rangle \\ &= \arg \min_{\mu \in [-1/2, 1/2)} \int_{-1/2}^{1/2} \langle \theta - \mu \rangle^2 f(\theta) d\theta. \end{aligned} \tag{5.2.2}$$

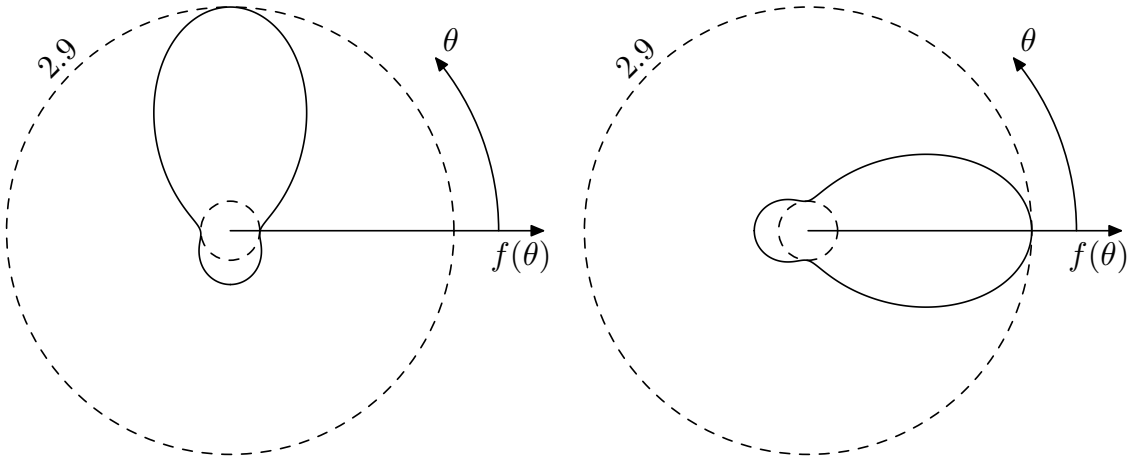


Figure 5.5: The pdf of a circular random variable Θ (left) and the pdf of the *rotated* random variable $\langle \Theta + \frac{1}{4} \rangle$.

The unwrapped variance of Θ is the variance of $\langle \Theta - \mu_{\text{unwrap}} \rangle$, that is

$$\sigma^2 = \text{var} \langle \Theta - \mu_{\text{unwrap}} \rangle = \int_{-1/2}^{1/2} \langle \theta - \mu_{\text{unwrap}} \rangle^2 f(\theta) d\theta.$$

Intuitively the unwrapped mean is such that the pdf of the rotated random variable $\langle \Theta - \mu_{\text{unwrap}} \rangle$ is mostly centered around zero.

For some circular distributions the minimisation that defines the unwrapped mean might not be unique. For example, consider the bimodal probability density function depicted in Figure 5.4 for which there would by two minima, one at $1/4$ and one at $-1/4$. In this case we say that the distribution *has no unwrapped mean.*

Definition 5.2. (**Unwrapped mean**) *Let Θ be a circular random variable with pdf f. Then Θ has unwrapped mean equal to μ_{unwrap} if and only if the minimisation from* (5.2.2) *is unique. Otherwise Θ has no unwrapped mean.*

A circular random variable, Θ, with zero unwrapped mean, i.e. $\mu_{\text{unwrap}} = 0$, has the special property that the unwrapped mean is equal to the mean, that is

$$\mu_{\text{unwrap}} = E[\Theta] = 0 \tag{5.2.3}$$

and the unwrapped variance is equal to the variance, that is

$$\sigma^2 = \text{var}[\Theta]. \tag{5.2.4}$$

The above properties will be quite useful in later chapters where will model noise processes as circular random variables with zero unwrapped mean.

5.2.3 Relationships between the circular and unwrapped mean

In general the unwrapped mean and the circular mean are different. In fact it is reasonably easy to construct distributions which *have* a circular mean but do *not*

have an unwrapped mean and other distributions which *have* an unwrapped mean but do *not* have a circular mean. However, for some distributions both means are defined and they are also equal. We call such distributions **unimean**. It is potentially feasible to classify all unimean distributions, but such a task is beyond the scope of this thesis. We do, however, find a particular class of useful circular distributions that are unimean. These are described in the next theorem.

Theorem 5.1. *Let Θ be a circular random variable with piecewise continuous pdf $f(\theta)$ that attains its maximum at 0 with $f(0) > 1$ and f even and nondecreasing on $[-1/2, 0]$. Then:*

1. *Θ is unimean with circular and unwrapped means equal to zero.*

2. *For every $\phi \in \mathbb{R}$ the rotated random variable $\langle \Theta + \phi \rangle$ is unimean with circular and unwrapped mean equal to ϕ.*

Before we begin the proof, it is worth noting that this result is quite intuitively obvious as the requirements placed on f force it to be *bunched* in a symmetric manner about zero. Notice that because f is even and nondecreasing on $[-1/2, 0]$ then it is also non increasing on $[0, 1/2)$. Combining this with the fact that $f(0) > 1$ immediately implies that $f(-1/2) < 1$ otherwise the integral $\int_{-1/2}^{1/2} f(\theta)d\theta$ would not equal one.

Notice that the requirements automatically include distributions that are even and unimodal and symmetric with mode at zero, but they also include distributions with *flat* pieces. We have deliberately included flat pieces in order to allow for a particular circular distribution called the **wrapped uniform distribution** that we will describe in Section 5.5. Finally, notice that the requirements of the theorem *do not* include the circular uniform distribution (Figure 5.3) because the circular uniform distribution has $f(0)$ equal to one but not greater than one.

In later chapters we will want to use circular random variables to model noise processes and we find that the class of distributions described by this theorem covers all of the types of distributions we need. We will now prove the theorem.

Proof. Statement (2) follows directly from statement (1) because if Θ has circular mean μ_{circ} and unwrapped mean μ_{unwrap} then the rotated circular random variable $\langle \Theta + \phi \rangle$ has circular mean $\langle \mu_{\text{circ}} + \phi \rangle$ and unwrapped mean $\langle \mu_{\text{unwrap}} + \phi \rangle$. It remains to prove statement (1).

We will first prove that the circular mean $\mu_{\text{circ}} = \frac{\angle c}{2\pi}$ is zero. It suffices to show that the integral

$$c = \int_{-1/2}^{1/2} e^{2\pi i\theta} f(\theta)d\theta$$

is a positive real number. Breaking the integral into real and imaginary parts we obtain

$$c = \int_{-1/2}^{1/2} \cos(2\pi\theta) f(\theta)d\theta + j \int_{-1/2}^{1/2} \sin(2\pi\theta) f(\theta)d\theta.$$

Now, because f is even and $\sin(\cdot)$ is odd then the imaginary part of c is zero. Also because f is nondecreasing of $[-1/2, 0]$ and $f(-1/2) < 1$ the integral

$$\int_{-1/2}^{0} \cos(2\pi\theta) f(\theta) d\theta > 0$$

because $\cos(\cdot)$ is increasing and odd about $-1/4$ on the interval $[-1/2, 0]$. Now as $\cos(\cdot)$ and f are even c is equal to twice the integral above and therefore $c > 0$ and $\angle c = 0$ and the circular mean is equal to zero.

We will now prove that the unwrapped mean is also zero. The approach we take is similar to Lemma 1 from Quinn [2007] and also Lemma 3 from McKilliam et al. [2010a]. We need to show that the integral

$$g(\mu) = \int_{-1/2}^{1/2} \langle \theta - \mu \rangle^2 f(\theta) d\theta$$

is uniquely minimised at $\mu = 0$. Firstly note that because f is even and $\langle -x \rangle = -\langle x \rangle$, i.e. the fractional part function is odd, we have

$$g(-\mu) = \int_{-1/2}^{1/2} \langle \theta + \mu \rangle^2 f(\theta) d\theta = \int_{-1/2}^{1/2} \langle -\theta + \mu \rangle^2 f(\theta) d\theta = g(\mu)$$

and therefore g is even. Now, if $\mu \geq 0$ then

$$\langle \theta - \mu \rangle = \begin{cases} \theta - \mu, & \text{if } \theta > -1/2 + \mu \\ \theta - \mu + 1, & \text{if } \theta < -1/2 + \mu \end{cases}$$

and g is given by

$$\begin{aligned} g(\mu) = g(-\mu) &= \int_{-1/2+\mu}^{1/2} (\theta - \mu)^2 f(\theta) d\theta + \int_{-1/2}^{-1/2+\mu} (\theta - \mu + 1)^2 f(\theta) d\theta \\ &= g(0) + \mu^2 - \mu \int_{-1/2}^{1/2} \theta f(\theta) d\theta + \int_{-1/2}^{-1/2+\mu} (1 - 2\mu + 2\theta) f(\theta) d\theta \\ &= g(0) + \mu^2 + \int_{-1/2}^{-1/2+\mu} (1 - 2\mu + 2\theta) f(\theta) d\theta \end{aligned}$$

where the integral $\int_{-1/2}^{1/2} \theta f(\theta) d\theta$ is equal to zero because f is even. Now, when $\mu > 0$, by differentiating g with respect to μ we obtain

$$\begin{aligned} g'(\mu) &= 2\mu - 2 \int_{-1/2}^{-1/2+\mu} f(\theta) d\theta \\ &= 2\mu - 2F(\mu - 1/2) \end{aligned} \tag{5.2.5}$$

where F is the cumulative distribution function (cdf) of Θ. Since f is even and non-decreasing on $[-1/2, 0)$, $F(-1/2) = 0$, $F(0) = 1/2$ and $F(\mu - 1/2)$ is convex on $[0, 1/2)$. Also because $f(-1/2) < 1$ it follows that F is *strictly* convex on $[0, 1/2)$. So g is monotonically increasing on $[0, 1/2)$ and, being even, is monotonically decreasing on $[-1/2, 0)$. Therefore g is uniquely minimised at zero and the unwrapped mean is equal to zero. □

In the remaining sections we describe a number of circular distributions that are common in the literature. We will find that under the right conditions, these distributions are unimean. We start with perhaps the best known circular distribution, the von Mises distribution.

5.3 The von Mises distribution

The von Mises distribution is probably the most commonly used distribution in the circular statistics literature [Mardia and Jupp, 2000, page 36]. The distribution is always unimodal and symmetric and therefore, by Theorem 5.1, it is also unimean. We denote the von Mises distribution by VonMises(μ, κ) where the circular mean and unwrapped mean is given by μ and κ is a concentration parameter that decreases with increasing circular and unwrapped variance. The probability density function is given by

$$f(\theta) = \frac{e^{\kappa \cos(2\pi(\theta-\mu))}}{I_0(\kappa)}$$

were $I_0(\cdot)$ is the zeroth order modified Bessel function and the circular variance is given by

$$1 - \frac{I_1(\kappa)}{I_0(\kappa)}.$$

There does not appear to be a closed-form for the unwrapped variance, but it can be computed numerically. The distribution approaches the normal distribution as $\kappa \to \infty$ and the circular uniform distribution as $\kappa \to 0$. This effect is shown in Figures 5.6 to 5.8.

5.4 The wrapped normal distribution

The wrapped normal distribution, denoted WrappedNormal(m, σ_g^2), is constructed by taking the normal distribution with mean m and variance σ_g^2 on the real line and *wrapping* it around the circle [Mardia and Jupp, 2000, page 50]. The pdf is then

$$f(\theta) = \sum_{k \in \mathbb{Z}} f_g(x + k)$$

where f_g is the pdf of the normal distribution on the real line with mean m and variance σ_g^2. The wrapped normal is symmetric and unimodal for all m and σ_g^2, so it follows from Theorem 5.1 that it is also unimean for all m and σ^2 [Stadje, 1984; Wintner, 1947; Lévy, 1939]. The circular and unwrapped mean are given by $\langle m \rangle$ and the circular variance is given by $1 - e^{\sigma_g^2/2}$ [Fisher, 1993, page 47]. The unwrapped variance is given by the infinite sum

$$\frac{1}{12} + \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} e^{-2\pi^2 \sigma_g^2 k^2}.$$

The coefficients in the sum converge quite quickly, so a close approximation can be obtained by summing only the first few terms.

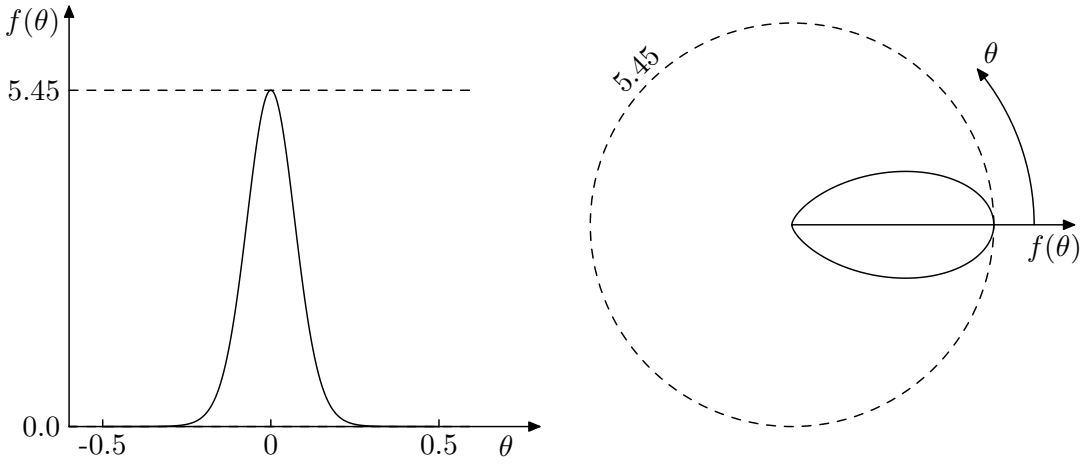


FIGURE 5.6: von Mises pdf where $\mu = 0$ and $\kappa = 5$

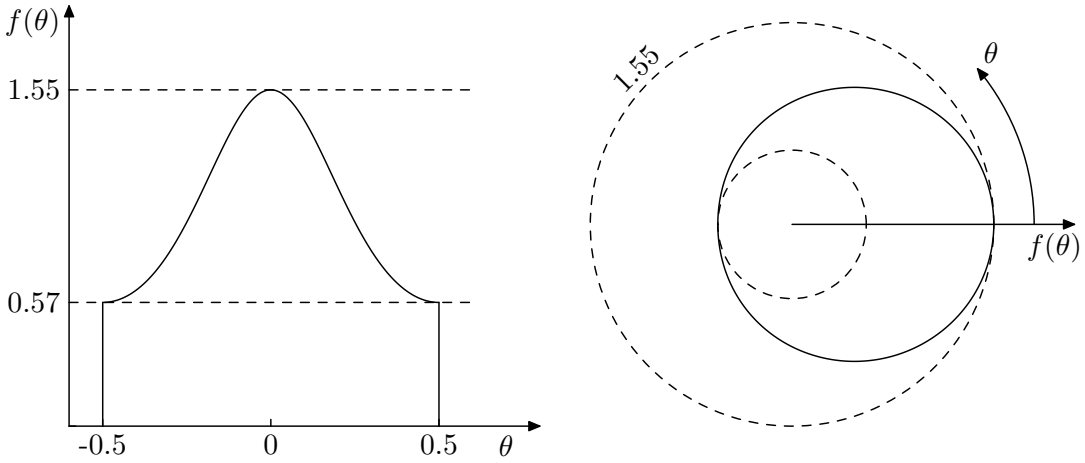


FIGURE 5.7: von Mises pdf where $\mu = 0$ and $\kappa = 0.5$

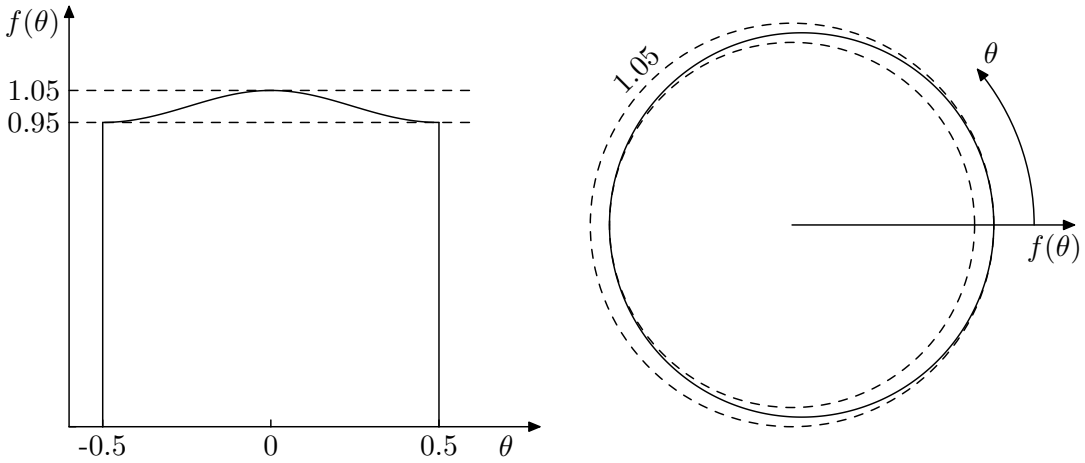


FIGURE 5.8: von Mises pdf where $\mu = 0$ and $\kappa = 0.05$

The distribution converges to the normal with mean $\langle m \rangle$ as $\sigma_g^2 \to 0$ and to the circular uniform distribution as $\sigma_g^2 \to \infty$. This effect is shown in Figures 5.9 to 5.11.

5.5 The wrapped uniform distribution

The wrapped uniform distribution, denoted WrappedUniform(m, σ_u^2), is constructed by taking the uniform distribution with mean m and variance σ_u^2 on the real line and *wrapping* it around the circle. The pdf is then

$$f(\theta) = \sum_{k \in \mathbb{Z}} f_U(x + k)$$

where f_U is the pdf of the uniform distribution on the real line with mean m and variance σ_u^2. In this thesis we will only have use of the case when $m = 0$ and $\sigma_u^2 < 1/12$. In this case no wrapping occurs and the distribution looks like the uniform distribution on the real line (Figure 5.13). Also, in this case, the distribution is unimean with circular and unwrapped mean 0, unwrapped variance σ_u^2 and circular variance given by

$$1 - \frac{\sin(2r)}{2r}$$

where $r = \sqrt{3}\pi\sigma_u$.

5.6 Projected circular distributions

A common way to construct a circular random variable is to take a complex random variable and *project* it onto the unit circle. We call these **projected circular distributions**. If X is a complex random variable with pdf $f_{\mathbb{C}}$ then the corresponding projected circular random variable Θ is given by the complex argument of X divided by 2π, that is

$$\Theta = \tfrac{1}{2\pi} \angle X.$$

The pdf of Θ is given by

$$f(\theta) = \int_0^\infty r f_{\mathbb{C}}\left(re^{2\pi i\theta}\right) dr.$$

Projected circular distributions will be particularly useful in Chapter 9 and 10 when we consider frequency estimation and polynomial phase signals. In these chapters we will be particularly interested in distributions that are unimean. The next theorem describes a large class of unimean projected circular distributions. The methodology behind this proof and the statement of the theorem is due to Barry Quinn, but any mistake or omission is the author's alone.

Theorem 5.2. *Let X be the complex random variable given by*

$$X = 1 + Ze^{2\pi j\Phi}$$

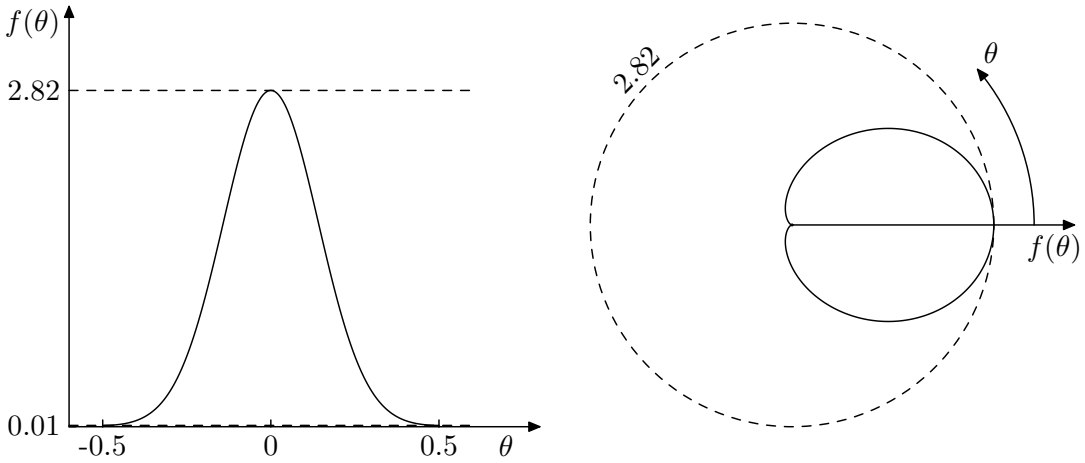


FIGURE 5.9: Wrapped normal pdf where $m = 0$ and $\sigma_g^2 = 0.02$

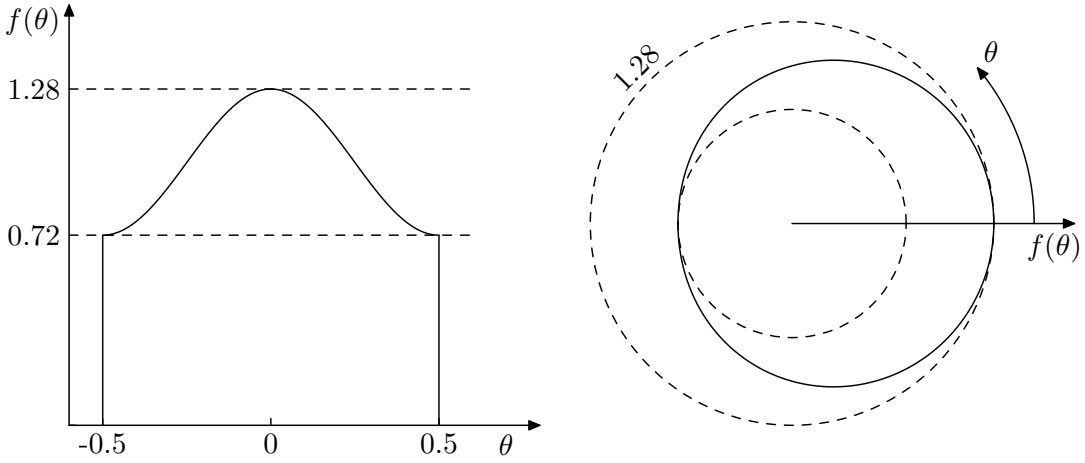


FIGURE 5.10: Wrapped normal pdf where $m = 0$ and $\sigma_g^2 = 0.1$

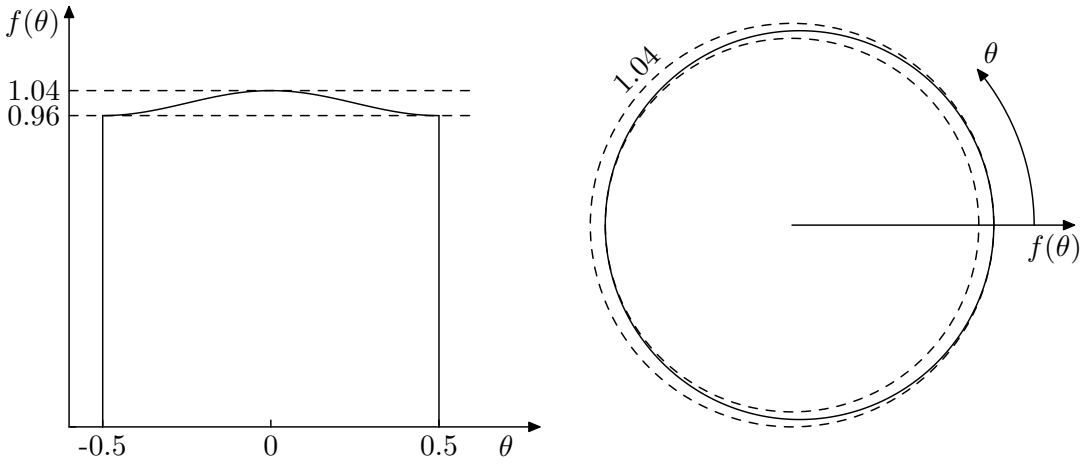


FIGURE 5.11: Wrapped normal pdf where $m = 0$ and $\sigma_g^2 = 0.2$

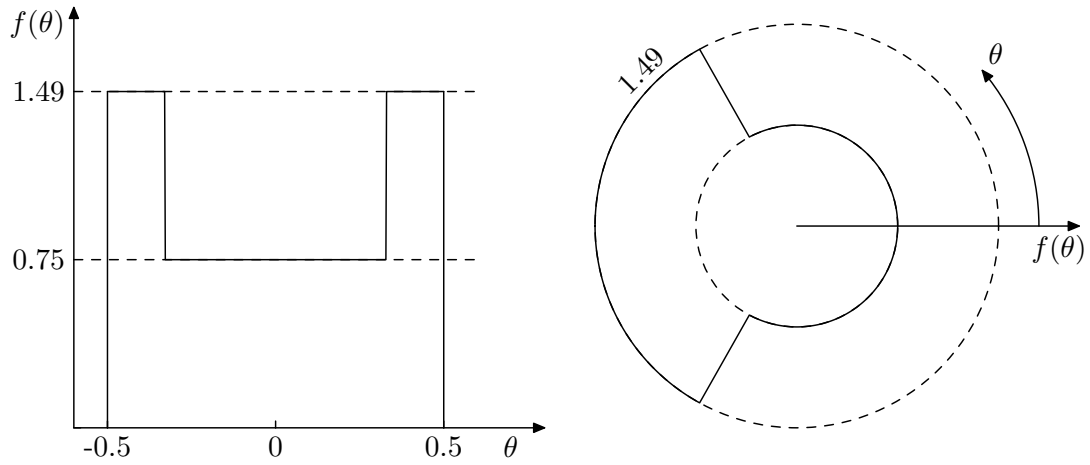


FIGURE 5.12: Wrapped uniform pdf with $m = 0$ and $\sigma_u^2 = 0.15$

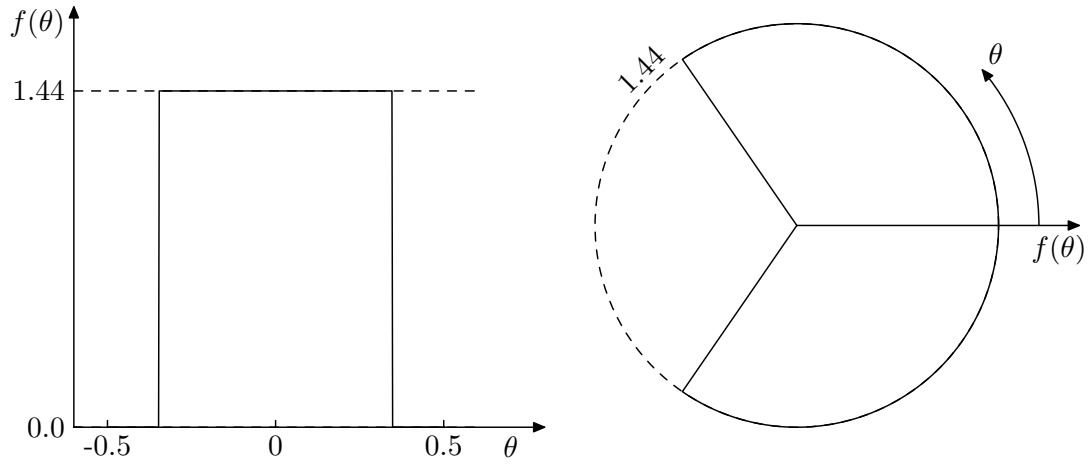


FIGURE 5.13: Wrapped uniform pdf with $m = 0$ and $\sigma_u^2 = 0.04$. No wrapping occurs because $\sigma_u^2 \leq 1/12$.

where Z and Φ are independent random variables. Let Z have pdf $f_Z(z)$ with support on the positive reals such that $z^{-1}f_Z(z)$ is non increasing and continuous and differentiable in z. Let Φ be uniformly distributed on $[-1/2, 1/2)$, i.e. Φ has the circular uniform distribution and let Θ be the projected circular random variable

$$\Theta = \frac{1}{2\pi}\angle X.$$

Then Θ is symmetrically distributed about 0, and unimodal with mode at 0, and Θ is unimean with circular and unwrapped means equal to 0.

Before we begin the proof note that the requirement for $z^{-1}f_Z(z)$ to be non increasing implies that the probability density function of $Ze^{2\pi j\Phi}$ decreases as we move away from the origin. That is, the pdf of $Ze^{2\pi j\Phi}$ in rectangular coordinates is given by $z^{-1}f_Z(z)$, where $z = \sqrt{x^2 + y^2}$ and x and y denotes the real and imaginary parts of $Ze^{2\pi j\Phi}$, and this pdf is non increasing with z. For example, the zero mean complex Gaussian distribution with independent real and imaginary parts satisfies this requirement.

Proof. The fact that Θ is unimean will follow from Theorem 5.1 after we show that Θ is symmetric and unimodal. The probability density function of Θ can be shown to be

$$f(\theta) = \int_0^\infty \frac{r}{\sqrt{r^2 - 2r\cos 2\pi\theta + 1}} f_Z\left(\sqrt{r^2 - 2r\cos 2\pi\theta + 1}\right) dr.$$

Note that f is continuous because f_Z is continuous. Also f is even because $\cos(\cdot)$ is even and therefore f is symmetric about zero, so it only remains to show that f is unimodal with mode at zero.

Let $a = \cos 2\pi\theta$, so $a \in [-1, 1]$ and $\theta = 0$ when $a = 1$ and as a decreases from 1 to -1 the magnitude $|\theta|$ increases. Let $z = \sqrt{r^2 - 2ra + 1}$ and note that $z \geq 0$ with equality only when $r = 1$ and $a = \pm 1$ or $a = 0$. The term inside the integral asymptotes when z is equal zero so is not differentiable at these points. For now assume that a is not 0 or ± 1 to avoid these asymptotes. Differentiating f with respect to a we obtain

$$\frac{d}{da} f = \int_0^\infty -\frac{r^2}{z} \frac{d}{dz}\left(\frac{f_Z(z)}{z}\right) dr.$$

Now because $z^{-1} f_Z(z)$ is non increasing in z and because z and r are positive the term inside the integral above is always positive. Therefore the integral is positive and f is increasing with a. The magnitude $|\theta|$ increases as a decreases so z is decreasing with $|\theta|$. It remains to show that no jump discontinuities occur in z when $a = \pm 1$ or $a = 0$, i.e. when $\theta = \pm 1/2$ for $\theta = 0$, but this is trivially the case due to the continuity of f. Therefore, as $f(\theta)$ is decreasing with $|\theta|$ and $f(\theta)$ is continuous we see that $f(\theta)$ is unimodal with mode at $\theta = 0$. □

It is easy to see that this theorem extends to the case where $X = c + Ze^{2\pi j\Phi}$ and c is any complex number. The circular random variable $\Theta = \frac{1}{2\pi}\angle X$ will then be unimodal and symmetric with mode $\frac{1}{2\pi}\angle c$ and from Theorem 5.1 we see that Θ will also be unimean with circular and unwrapped means equal to $\frac{1}{2\pi}\angle c$. We will now consider a particularly important circular distribution that results from projecting the complex normal distribution.

5.6.1 The projected normal distribution

The projected normal distribution, denoted ProjectedNormal(s, Σ), is the distribution of the complex argument (divided by 2π) of a complex normal random variable with mean $s \in \mathbb{C}$ and covariance between real and imaginary parts given by the 2×2 matrix Σ. This distribution has been extensively studied by [Mardia and Jupp, 2000, p. 46].

Consider the special case ProjectedNormal$(1, \sigma_N^2 \mathbf{I})$ where $\mathbf{I}$ is the 2×2 identity matrix. It follows immediately from Theorem 5.2 that this distribution is symmetric and unimodal and unimean with circular and unwrapped mean 0. The pdf is given by Quinn [2007] as

$$f(\theta) = e^{-\frac{1}{2\sigma_N^2}} + \cos(2\pi\theta) e^{-\frac{\sin^2(2\pi\theta)}{2\sigma_N^2}} \sqrt{\frac{\pi}{2\sigma_N^2}} \left(1 + \operatorname{erf}\left(\sqrt{\frac{\sigma_N^2}{2}} \cos(2\pi\theta)\right)\right)$$

and the circular variance is given by

$$1 - \sqrt{\frac{\pi\nu}{2}} e^{2\nu} \big(I_0(\nu) + I_1(\nu) \big)$$

where $\nu = (4\sigma^2)^{-1}$. The unwrapped variance is required to be numerically evaluated. It is straightforward to see that

$$\text{ProjectedNormal}(p, \Sigma) \qquad \text{and} \qquad \text{ProjectedNormal}(1, p^{-2}\Sigma)$$

are equivalent for any $p > 0$. We will have use of this distribution in Chapters 6, 9 and 10 where it will be used for modeling noise processes.

5.7 Summary

In this chapter we have introduced the field of circular statistics that aims to describe the nature of data that is measured in angles, or unit vectors. In Section 5.1 we defined circular random variables as random variables that have a probability density function with support on $[-1/2, 1/2)$. We considered the *usual* mean of a circular random variable and showed that it does not map well to our intuitive sense of **mean direction**. To solve this problem we described the **circular mean**, that is common in the literature, and the **unwrapped mean** that is less common. Both the circular and unwrapped means have intuitively appealing definitions and both map well to our intuitive sense of mean direction.

In Section 5.2.3 we considered some relationships between the unwrapped mean and the circular mean. The two means are not in general equal and for some distributions they are not even defined. In Theorem 5.1 we described a large class of circular distributions that have equal unwrapped and circular means and we called such distributions **unimean**. We will have particular interest in unimean distributions in the following chapters where they will be used to model noise processes.

In Sections 5.3 and 5.4 we considered two popular unimean circular distributions, the **von Mises distribution** and the **wrapped normal distribution**. In Section 5.5 we consider the **wrapped uniform distribution** that is unimean under certain assumptions about its unwrapped mean and variance.

In Section 5.6 we considered projected circular random variables. These are the result of taking the complex argument (divided by 2π) of a random variable in the complex plane. Projected circular distributions will be of particular interest in Chapters 9 and 10 when we consider the problems of frequency estimation and polynomial phase estimation. In Theorem 5.2 we described a large class of unimean projected circular distributions and in Section 5.6.1 we described the **projected normal distribution** that results from taking the complex argument of a complex normal random variable. This distribution is probably the most common and useful of all projected circular distributions. We find some conditions under which the projected normal is unimean with zero unwrapped and circular means.

In the next chapter we consider methods for estimating the circular and unwrapped means of a circular random variable from a set of observations. We describe accurate and computationally efficient methods for estimating both means.

The technique we describe for estimating the unwrapped mean is based on computing a nearest lattice point in the lattice A_n^*.

—Tell me, what is the practical use of all of this?
Kimberley Nunes

6 Estimating direction

One of the most fundamental tasks in statistics and estimation theory is estimating the mean of a random variable given a set of observations. Accordingly two of the most important results in statistics are the **law of large numbers**, that asserts we *can* obtain an accurate estimate of the mean from a sufficiently large number of observations, and the **central limit theorem** that describes how accurate this estimate will be.

It is not surprising that one of the fundamental tasks in circular statistics is estimating the **mean direction** of a circular random variable from a set of observations. This is the topic we consider in this chapter. Estimators of the mean direction have a wide variety of applications in science and engineering. For example, if you listen to the weather report you are often told (an estimate of) the direction of the wind. Obtaining an accurate estimate requires a method for accurately estimating the mean wind direction from a number of observations of the wind direction. This task is not as straightforward as it might initially seem.

The notorious wrapping problem

Consider making N observations $\Theta_1, \Theta_2, \ldots, \Theta_N$ of a circular random variable. We would like to estimate the mean direction of the observations. A naïve approach is to take the *usual* estimate of the mean, that is by taking the average over the samples

$$\frac{1}{N}\sum_{n=1}^{N}\Theta_n.$$

Obvious problems arise. Consider when $N = 2$ and $\Theta_1 = 0.4$ and $\Theta_2 = -0.4$. The naïve approach would yield the estimate 0. Intuitively this is wrong and a more reasonable estimate would be -0.5 (see Figure 6.1).

This is the *notorious wrapping problem* from meteorology [Fisher, 1993]. The

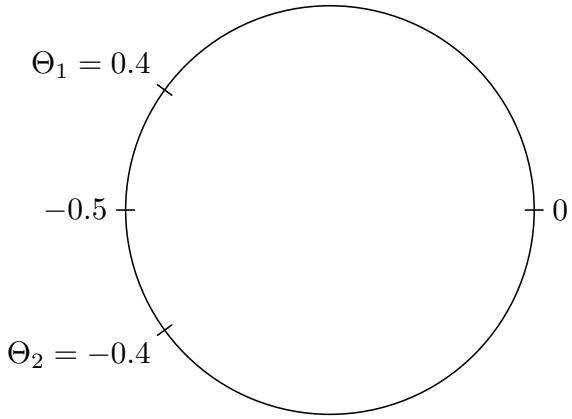


FIGURE 6.1: Care must be taken when calculating the mean direction. In the figure there are two observations Θ_1 and Θ_2. Taking the naïve approach yields $(\Theta_1 + \Theta_2)/2 = 0$, but clearly a better estimate of the mean direction is -0.5.

problem is that the average produces an estimate of the expected value, or mean $E[\Theta_n]$, and, as discussed in the previous chapter, this mean does not conform well with our intuitive notion of mean direction. What we want are estimators for the **circular mean** or the **unwrapped mean** as these conform well with our intuitions. We consider approaches to estimating the circular mean and unwrapped mean in this chapter. We will show that for both circular and unwrapped means there are estimators that satisfy an appropriate strong law of large numbers and a central limit theorem. Moreover, the estimators are computationally very simple, requiring a number of operations that is linear in the number of observations, N.

A simple estimator of the circular mean is the **sample circular mean** and we describe this in Section 6.1. This is the most common estimator in the circular statistics literature. If the circular random variable that describes the observations does not have a circular mean, then this estimator cannot be applied. However, if the circular mean does exist, then this estimator is strongly consistent (it converges almost surely to the true circular mean as the number of observations increases) and we can derive a central limit theorem that describes its accuracy. The estimator involves summing N complex numbers and therefore requires a number of operations that is linear in the number of observations, N.

In Section 5.2 we describe an alternative estimator that we call the **angular least squares** estimator. This estimates the unwrapped mean of a circular random variable and appears to have received little attention in the literature, with the notable exception of Quinn [2007]. We show that the angular least squares estimator is strongly consistent (it converges almost surely to the true unwrapped mean whenever it exists) and derive a central limit theorem that describes its accuracy.

It is possible that the angular least squares estimator has been mostly ignored in the literature due to the (mis)conception that it is difficult to compute. It turns out that the angular least squares estimator can be computed very efficiently by finding a nearest lattice point in the lattice A^*_{N-1}. We discovered a linear-time algorithm to find a nearest point in A^*_{N-1} in Chapter 3 and it follows that, like the sample circular mean, computing the angular least squares estimator requires a number of

operations that is linear in the number of observations, N. One advantage of this algorithm is that it largely avoids trigonometric operations. Particularly on small computing devices this is likely to lead to significant computational savings. We consider some computational differences between the angular least squares estimator and the sample circular mean in Section 6.7.

In Sections 6.4, 6.5 and 6.6 we consider a number of problems from signal processing and communications engineering that can be solved very efficiently and accurately by estimating the circular mean or the unwrapped mean. The first problem we consider is **phase estimation**, for which it is rather intuitively obvious that the sample circular mean or the angular least squares estimators can be applied. Phase estimation is equivalent to the problem of estimating the single coefficient of a polynomial phase signal of order zero. Therefore this section serves as an introduction to the more difficult problems of frequency estimation and polynomial phase estimation that we consider in Chapters 9 and 10.

In Sections 6.5 and 6.6 we consider two problems, the **noncoherent detection** of **phase-shift-keyed** digital signals and **delay estimation** from incomplete data. For these problems it is not immediately obvious that circular statistics are required. However, we show how these problems can be solved very accurately and efficiently using either the sample circular mean or the angular least square estimators.

6.1 The sample circular mean

To compute the sample circular mean of the observations $\Theta_1, \ldots, \Theta_N$ we first convert the observations to the complex numbers $e^{2\pi j \Theta_n}$ and take the average

$$\bar{C} = \frac{1}{N} \sum_{n=1}^{N} e^{2\pi j \Theta_n}. \tag{6.1.1}$$

The sample circular mean estimator is given by

$$\hat{\mu} = \frac{\angle \bar{C}}{2\pi}. \tag{6.1.2}$$

This estimator has received significant attention in the circular statistics literature. It is known that the sample circular mean is the maximum likelihood estimator of the circular mean when the Θ_n are independent and identically distributed with the von Mises distribution [Mardia and Jupp, 2000; Fisher, 1993]. The next theorem, due to Quinn [2010], precisely describes the asymptotic behaviour of this estimator under some assumptions about the distribution of the Θ_n.

Theorem 6.1. [Quinn, 2010] *Let $\Theta_1, \Theta_2, \cdots, \Theta_N$ be N consequitive observations of the form*

$$\Theta_n = \langle \Phi_n + \tilde{\mu} \rangle$$

where $\Phi_1, \ldots, \Phi_N$ are independent and identical circular random variables with zero circular mean, circular variance ν and pdf f. Let $\hat{\mu}$ be the sample circular mean estimator of the Θ_n. Then:

1. *(Strong consistency) The fractional part* $\langle \hat{\mu} - \tilde{\mu} \rangle \to 0$ *almost surely as the number of observations* $N \to \infty$.

2. *(Central limit theorem) The distribution of* $\sqrt{N}\,\langle \hat{\mu} - \tilde{\mu} \rangle$ *approaches the normal with zero mean and variance*

$$\frac{\sigma_s^2}{4\pi^2(1-\nu)^2}$$

where

$$\sigma_s^2 = \int_{-1/2}^{1/2} \sin^2(2\pi\theta) f(\theta) d\theta.$$

Before we give the proof note that the theorem places conditions on the *fractional part* of the difference, i.e. $\langle \hat{\mu} - \tilde{\mu} \rangle$, between the *true* circular mean $\tilde{\mu}$ and the estimated circular mean $\hat{\mu}$ rather than directly on the difference $\hat{\mu} - \tilde{\mu}$. This makes intuitive sense because the angles $\tilde{\mu}$ and $\tilde{\mu} + k$ are equivalent for any integer k (see Figure 6.1). We will find that a similar condition must be imposed for the purpose of polynomial phase estimation in Chapter 8. The theorem guarantees that $\langle \hat{\mu} - \tilde{\mu} \rangle$ converges whenever the circular mean is defined.

Computing the asymptotic variance given by the theorem requires calculating σ_s^2. In general this can be numerically evaluated, but for some circular distributions reasonably simple expressions can be found. It has been shown by Quinn [2010] that when Φ_n has the VonMises$(0, \kappa)$ distribution

$$\frac{\sigma_s^2}{(1-\nu)^2} = \frac{I_0(\kappa)}{\kappa I_1(\kappa)}$$

and when Φ_n has the ProjectedNormal$(1, \sigma^2\mathbf{I})$ distribution

$$\frac{\sigma_s^2}{(1-\nu)^2} = \frac{8\sigma^4\left(e^{2\nu}-1\right)}{\pi\left(I_0\left(\nu\right)+I_1\left(\nu\right)\right)^2}$$

where $\nu = (4\sigma^2)^{-1}$. It is also straightforward to show that when Φ_n has the WrappedUniform$(0, \sigma^2)$ distribution, where $\sigma^2 < 1/12$, then

$$\frac{\sigma_s^2}{(1-\nu)^2} = \frac{4r^2 - r\sin(4r)}{2\sin^2(2r)},$$

where $r = \sqrt{3}\pi\sigma$. A closed-form expression for the wrapped normal distribution does not appear to exist and in this case we resort to numerical evaluation. We will now prove the theorem.

Proof. From (6.1.1) and (6.1.2) we obtain

$$\begin{aligned}\hat{\mu} &= \frac{1}{2\pi}\angle\left(N^{-1}\sum_{n=1}^{N} e^{2\pi j\langle\tilde{\mu}+\Phi_n\rangle}\right) \\ &= \frac{1}{2\pi}\angle\left(e^{2\pi j\tilde{\mu}}N^{-1}\sum_{n=1}^{N} e^{2\pi j\Phi_n}\right).\end{aligned}$$

Subtracting $\tilde{\mu}$ from both sides and taking fractional parts

$$\langle \hat{\mu} - \tilde{\mu} \rangle = \frac{1}{2\pi} \angle \left(N^{-1} \sum_{n=1}^{N} e^{2\pi j \Phi_n} \right).$$

The Φ_n have zero circular mean so the expectation $E[e^{2\pi j \Phi_n}] = 1 - \nu$ is a positive real. By the strong law of large numbers

$$N^{-1} \sum_{n=1}^{N} e^{2\pi j \Phi_n} \to E[e^{2\pi j \Phi_n}] = 1 - \nu$$

almost surely as N goes to infinity, and because $\angle(1-\nu) = 0$ then $\langle \hat{\mu} - \tilde{\mu} \rangle$ converges almost surely to zero as N goes to infinity. This completes the proof of strong consistency.

To prove the central limit theorem let

$$1 - \nu + X = N^{-1} \sum_{n=1}^{N} \cos(2\pi \Phi_n) \qquad \text{and} \qquad Y = N^{-1} \sum_{n=1}^{N} \sin(2\pi \Phi_n)$$

denote the real and imaginary parts of $N^{-1} \sum_{n=1}^{N} e^{2\pi j \Phi_n}$. We will use O_P and o_P to denote variables that converge in probability as $N \to \infty$. Then both X and Y are $O_P(N^{-1/2})$ and $\sqrt{N}Y$ converges to the normal with zero mean and variance σ_s^2. Now,

$$\begin{aligned} \sqrt{N} \langle \hat{\mu} - \tilde{\mu} \rangle &= \frac{\sqrt{N}}{2\pi} \angle \left(1 + \frac{X + jY}{1 - \nu} \right) \\ &= \frac{\sqrt{N}}{2\pi} \left(\frac{Y}{1 - \nu} + O_P(N^{-1}) \right) \\ &= \frac{\sqrt{N}Y}{2\pi(1 - \nu)} + O_P(N^{-1/2}) \end{aligned}$$

follows by taking a first order approximation of the arctangent function. So, $\sqrt{N} \langle \hat{\mu} - \tilde{\mu} \rangle$ converges in probability to $\frac{\sqrt{N}}{2\pi} \frac{Y}{1-\nu}$ and therefore converges in distribution to the normal with zero mean and variance $\frac{\sigma_s^2}{4\pi^2(1-\nu)^2}$. □

6.2 Angular least squares

We now describe the angular least squares estimator of the unwrapped mean. We define the sum of squares function

$$SS(\mu) = \sum_{n=1}^{N} \langle \Theta_n - \mu \rangle^2.$$

The angular least squares estimator is defined as the minimiser of $SS(\mu)$. Intuitively this estimator chooses the angle $\hat{\mu}$ such that the rotated random variables $\langle \Theta_n - \hat{\mu} \rangle$ are closest to zero.

The angular least squares estimator can be computed very quickly by finding a nearest point in the lattice A^*_{N-1}. To see this write the sum of squares function SS as

$$SS(\mu) = \sum_{n=1}^{N} (\Theta_n - \mu - W_n)^2$$

where W_n are integers given by $\lceil \Theta_n - \mu \rfloor$. The W_n are called **wrapping variables** because they describe how the $\Theta_n - \mu$ *wrap* around the circle. If we consider the W_n as nuisance parameters to be estimated then we may write SS as a function of both μ and the W_n. The angular least squares estimator can then be found by minimising over both μ and the W_n. It may at first appear that we have just made this problem very hard for ourselves. There was only one variable, μ, to minimise over and now there are $N+1$ variables, μ and all of the $W_1, \ldots, W_N$. However, we will show how this joint minimisation problem can be solved very efficiently by computing a nearest point in the lattice A^*_{N-1}. Write SS as a function of μ and the W_n using vectors as

$$SS(\mu, \mathbf{w}) = \|\boldsymbol{\theta} - \mu\mathbf{1} - \mathbf{w}\|^2$$

where $\boldsymbol{\theta} = [\Theta_1, \ldots, \Theta_N]^\dagger$ and $\mathbf{w} = [W_1, \ldots, W_N]^\dagger$ and where $\mathbf{1}$ is the all ones vector[1]. Fixing $\mathbf{w}$ and minimising with respect to μ gives

$$\hat{\mu} = \frac{(\boldsymbol{\theta} - \mathbf{w}) \cdot \mathbf{1}}{\mathbf{1} \cdot \mathbf{1}}. \tag{6.2.1}$$

Substituting this into $SS(\mu, \mathbf{w})$ gives $SS(\mu, \mathbf{w})$ conditioned on minimisation with respect to μ as

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

where $\mathbf{Q}$ is the orthogonal projection matrix into the space orthogonal to the all ones vector $\mathbf{1}$. We saw in Section 3.3.1 how this projection matrix is related to a generator for A^*_{N-1}. It follows that $\mathbf{Q}\mathbf{w}$ is a lattice point in A^*_{N-1} and that minimising $SS(\mathbf{w})$ is equivalent to finding the nearest lattice point in A^*_{N-1} to $\mathbf{Q}\boldsymbol{\theta}$. We may use any of the algorithms described in Section 3.5.2 to compute the nearest point, which we denote $\mathbf{Q}\hat{\mathbf{w}}$. Once $\hat{\mathbf{w}}$ has been found, the estimate $\hat{\mu}$ is given by substituting $\hat{\mathbf{w}}$ for $\mathbf{w}$ in (6.2.1). Noting that the nearest point can be found in $O(N)$ operations this estimator can thus be computed in linear-time. The next theorem describes the asymptotic behaviour of this estimator.

Theorem 6.2. *Let $\Theta_1, \Theta_2, \cdots, \Theta_N$ be N consequitive observations of the form*

$$\Theta_n = \langle \Phi_n + \tilde{\mu} \rangle$$

where $\Phi_1, \ldots, \Phi_N$ are independent and identical circular random variables with zero unwrapped mean, unwrapped variance σ^2 and pdf f. Let $\hat{\mu}$ denote angular least squares estimator of the Θ_n. Then:

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

1. *(Strong consistency) The fractional part* $\langle \hat{\mu} - \tilde{\mu} \rangle \to 0$ *almost surely as the number of observations* $N \to \infty$.

2. *(Central limit theorem) If the periodic function* $f(\langle x \rangle)$ *is continuous at* $x = -1/2$ *and* $f(-1/2) \neq 1$ *then the distribution of* $\sqrt{N}\langle \hat{\mu} - \tilde{\mu} \rangle$ *approaches the normal with zero mean and variance*

$$\frac{\sigma^2}{\big(1 - f(-1/2)\big)^2}. \tag{6.2.2}$$

Proof. This theorem follows as a special case of Theorem 8.1 that we will prove in Chapter 8. An alternative proof is also given by Quinn [2007]. □

This theorem asserts that the fractional part $\langle \hat{\mu} - \tilde{\mu} \rangle$ converges whenever the unwrapped mean is defined. For the central limit theorem we require some extra assumptions about the pdf f of the Φ_n. The formula given for the asymptotic variance only holds when the periodic function $f(\langle x \rangle)$ is continuous at $x = -1/2$ and if $f(-1/2) \neq 1$. If these conditions are not met then other expressions for the asymptotic variance can be found, but this comes at a rather substantial increase in complexity, so we have opted to omit them. Circular distributions that do not satisfy these requirements are somewhat pathological and it is highly unlikely that they would be needed in practice. For all of the distributions considered in this thesis these requirements hold.

6.3 Comparing the two estimators

In this section we compare the angular least squares and the sample circular mean estimators. It is important to realise that, in general, these are estimators of different quantities. Angular least squares estimates the unwrapped mean, while the sample circular mean estimates the circular mean. For circular distributions that have different circular and unwrapped means, this would make a comparison somewhat meaningless. However, comparisons can be made for circular distributions with equal circular and unwrapped means, i.e. unimean distributions.

The unimean distributions that we will consider are the von Mises, projected normal, wrapped Gaussian and wrapped uniform distributions that were described in Chapter 5. As a general rule of thumb we find that the sample circular mean is slightly more accurate when the distribution is 'von Mises-like' and the angular least squares estimator is more accurate when the distribution is 'uniform-like'.

Figures 6.2 and 6.3 display the sample mean square error (MSE) of the estimators when the number of observations is $N = 1, 4, 16, 64, 256, 1024$ and 4096. The quantity displayed on the horizontal axis is the unwrapped variance of the random variable being estimated. For each value of unwrapped variance $T = 4000$ trials were run to obtain T separate estimates $\hat{\mu}_1, \ldots, \hat{\mu}_T$. The sample MSE is then computed by averaging the squared *fractional parts* according to

$$\frac{1}{T}\sum_{t=1}^{T} \langle \hat{\mu}_t - \tilde{\mu} \rangle^2 .$$

Figure 6.2 displays the MSE when the observations are sampled from the von Mises distribution. The sample circular mean performs slightly better in this case. This is perhaps not surprising as it is known that the sample circular mean is the maximum likelihood estimator of the circular mean (and hence the unwrapped mean) of the von Mises distribution Mardia and Jupp [2000]. Figure 6.3 displays the MSE when observations are sampled from the wrapped uniform distribution. For this distribution, angular least squares is more accurate. Both figures display the asymptotic variances predicted in Theorems 6.1 and 6.2. The predictions are remarkably accurate for both the von Mises and the uniform distribution, particularly when N is large.

In the next sections we will consider some practical signal processing problems to which these estimators can be applied. These are **phase estimation**, detection of **phase-shift-keyed** digital signals and **delay estimation**. These applications will motivate the use of the projected normal and wrapped normal distributions.

6.4 Phase estimation

A problem closely related to estimating the circular mean is that of **phase estimation**. Here, the model will motivate the use of the **projected circular distributions** that we introduced in Section 5.6. This section also serves as a introduction to the more complicated problems of frequency estimation and polynomial phase estimation that we consider in Chapters 9 and 10. Consider observing N complex random variables $Y_1, Y_2, \ldots, Y_N$ of the form

$$Y_n = \rho e^{2\pi j\tilde{\mu}} + X_n$$

where $\rho > 0$ is an unknown amplitude and $X_1, \ldots, X_N$ are zero mean, independent and identically distributed complex random variables. We wish to estimate the phase parameter $\tilde{\mu}$. An obvious approach is the least squares estimator of $\tilde{\mu}$ that is given by $\angle\bar{Y}/(2\pi)$ where $\bar{Y} = \sum_{n=1}^{N} Y_n$ is the mean of the Y_n.

An alternative approach is to discard the magnitude of the Y_n and consider only the phases $\angle Y_1, \ldots, \angle Y_N$. Dividing the phases by 2π we obtain the circular random variables

$$\Theta_n = \frac{\angle Y_n}{2\pi} = \frac{\angle\left(\rho e^{2\pi j\tilde{\mu}} + X_n\right)}{2\pi} = \langle\tilde{\mu} + \Phi_n\rangle \tag{6.4.1}$$

where the $\Phi_1, \cdots \Phi_N$ are circular random variables related to the X_n by

$$\Phi_n = \frac{\angle(\rho + X_n)}{2\pi}.$$

The Φ_n are projected circular random variables and if the $\rho + X_n$ satisfy the requirements of Theorem 5.2 then the Φ_n will be unimean with zero circular and unwrapped means. That is, if the X_n are zero mean complex random variables with pdf that is non increasing with the magnitude $|X_n|$ and is independent of the argument $\angle X_n$ then the Φ_n are unimean with zero circular and unwrapped means. Under these assumptions we see that $\tilde{\mu}$ can be estimated as either the circular or the unwrapped

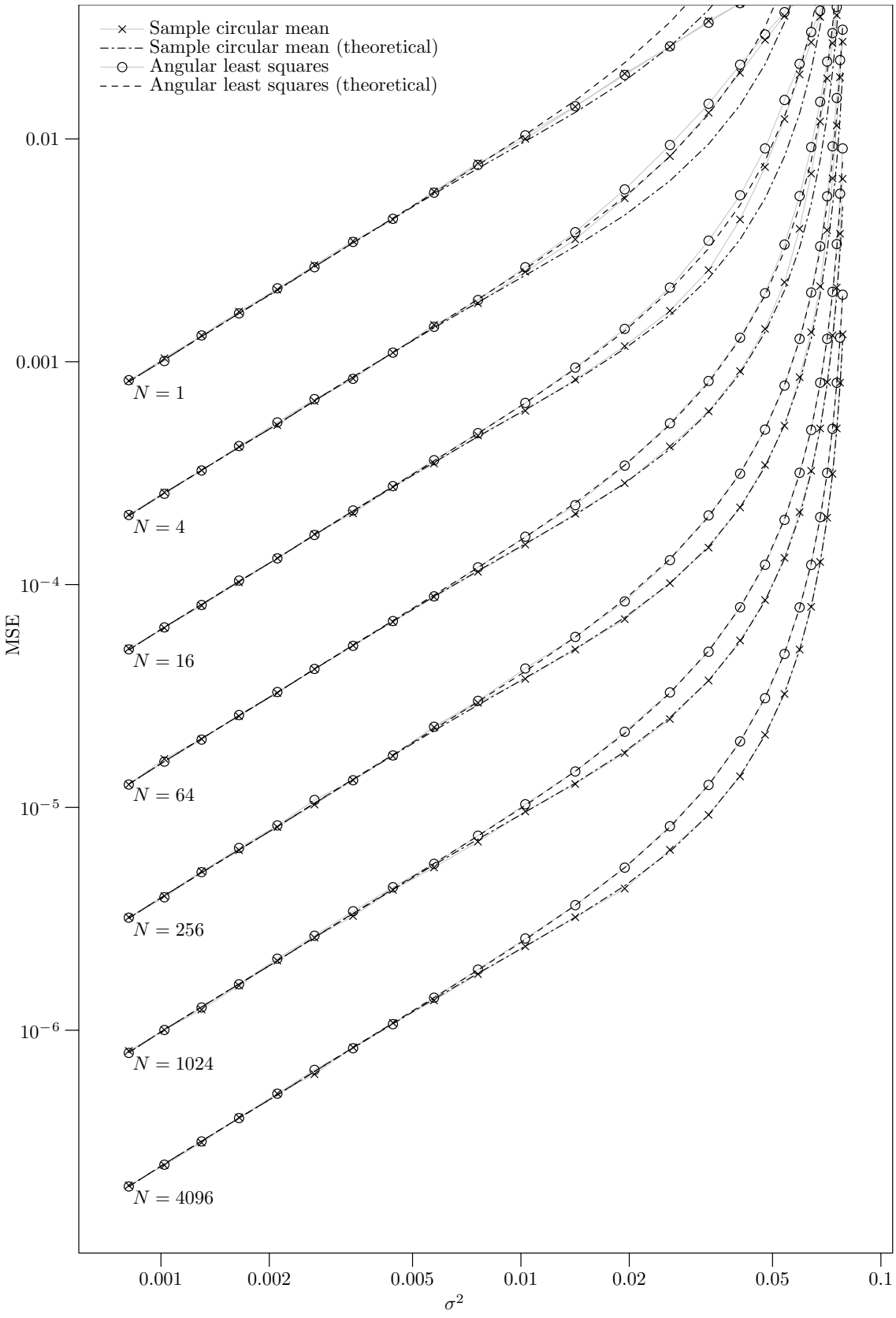


Figure 6.2: MSE versus unwrapped variance when $f(x)$ is the von Mises distribution.

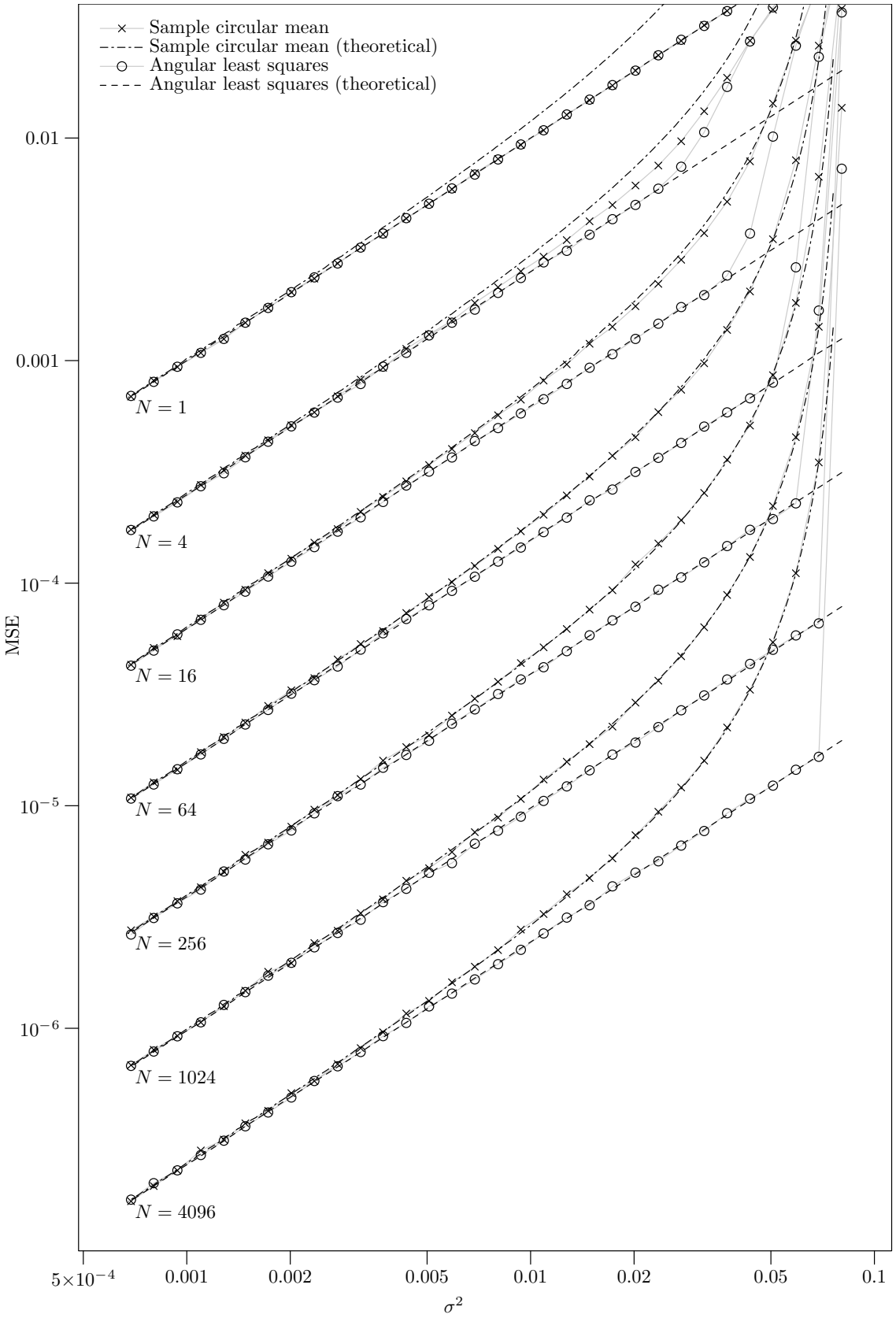


FIGURE 6.3: MSE versus unwrapped variance when $f(x)$ is the wrapped uniform distribution.

mean of the Θ_n. We can use either the sample circular mean or the angular least squares estimator.

In Figure 6.4 we plot the MSE of both the least squares estimator, the sample circular mean estimator and the angular least squares estimator for $N = 1, 4, 16, 64, 256, 1024$ and 4096. For this simulation we have set the amplitude $\rho = 1$ and the X_n are independent and identical with zero mean complex Gaussian random variable with independent real and imaginary parts having variance σ_c^2. Under this assumption the least squares estimator, given by $\angle\bar{Y}/(2\pi)$, is also the maximum likelihood (ML) estimator and the Cramer Rao lower bound (CRB) is given by

$$\frac{\sigma_c^2}{4\rho^2\pi^2 N}. \tag{6.4.2}$$

Also, the Φ_n take the ProjectedNormal$(1, \sigma_c^2\Sigma)$ distribution. It is clear from Figure 6.4 that the maximum likelihood estimator gives the best performance, closely followed by the sample circular mean and then the angular least squares estimator. We have plotted the asymptotic variance as predicted by Theorems 6.1 and 6.2 which can be seen to closely model the behaviour of the sample circular mean and the angular least squares estimators, particularly when N is large. The CRB (6.4.2) is also plotted and it can be seen that the maximum likelihood estimator gives a mean square error very close to the CRB.

It may seem that discarding the amplitudes and estimating the circular (or unwrapped) mean is somewhat contrived in this case, particularly in view of the fact that the least squares estimator performs so well and is simple to compute. In practice we recommend using the least squares estimator for phase estimation. However, in Chapter 10 when we consider polynomial phase signals, the equivalent least squares estimator will be hopelessly computationally expensive. In this case, discarding the amplitudes can lead to estimators that are significantly less computationally expensive. In Section 10.4 we will consider using an analogue of the angular least squares estimator for polynomial phase signals and we find that it produces remarkably accurate results in only a fraction of the time it takes to compute the least squares estimator.

6.5 Noncoherent detection of PSK

In this section we will apply the angular least square estimator and the sample circular mean estimator to the problem of **noncoherent detection** of data signals that arise in communications engineering. We will find that the angular least squares estimator and the sample circular mean are computationally less expensive than existing techniques and produce virtually identical detection performance.

Consider the transmission of digital signals in an unknown transmission channel that varies over time. In order to correctly decode a transmitted signal the receiver must obtain an estimate of the channel. A standard technique for channel estimation is for the transmitter to send a known signal, called a **pilot signal**, that can be used to estimate the channel at the receiver. This is called **pilot assisted transmission** and is in regular use in many communications systems. In situations where the

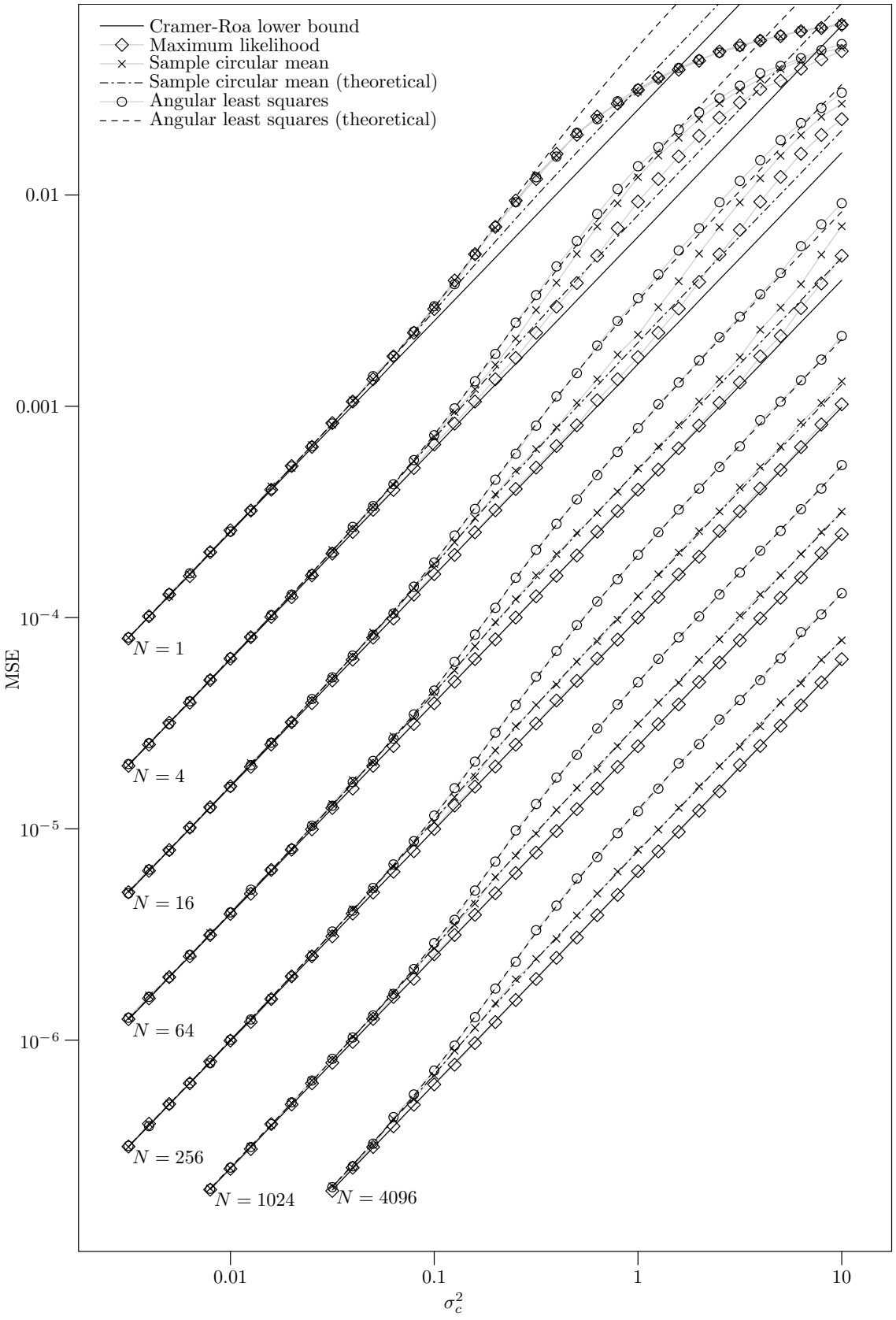


FIGURE 6.4: MSE versus σ_c^2 for phase estimation in complex Gaussian noise.

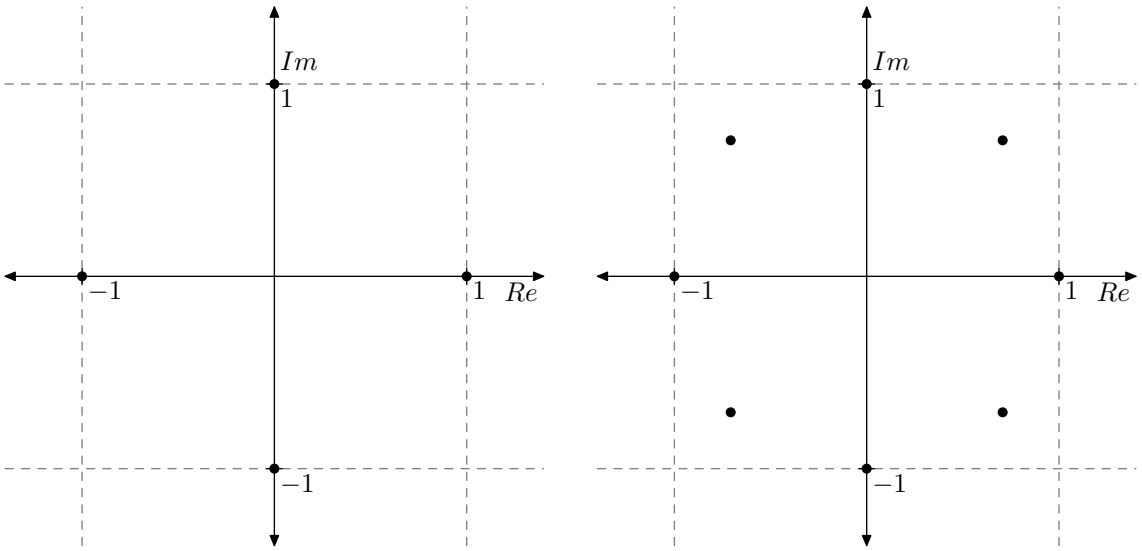


Figure 6.5: The 4 PSK constellation (left) and the 8 PSK constellation (right)

channel varies over time, pilot signals must be transmitted regularly, at a suitable interval, so that the estimate of the channel remains valid. This suitable interval is called the **coherence interval**.

Alternatively, noncoherent detection attempts to estimate the channel and the data simultaneously, without the need for pilot signals. Noncoherent detection is particularly applicable to systems with small coherence intervals where the frequent transmission of pilot signals is wasteful. Such situations typically occur in mobile communications. Moreover, pilot assisted transmission is, in a sense, inherently suboptimal because it only uses the energy of a small number of pilot signals for channel estimation, rather than also exploiting the (typically larger) energy in the unknown data signals [Chen et al., 2003]. In this section we consider noncoherent transmission schemes for a particular symbol constellation called M**-ary phase-shift-keying** (M-PSK) and when the channel adheres to a model called the **block fading** channel model.

The M-PSK constellation contains **symbols** of the form $e^{2\pi ju/M}$ where $u \in \{0, 1, \cdots, M-1\}$. Figure 6.5 is a picture of the 4-PSK constellation and the 8-PSK constellation in the complex plane. The 4-PSK constellation is often also called the **quadrature-phase-shift-keying** (QPSK) or 4-QAM constellation.

The block fading channel model

We consider receiving N symbols from the M-PSK constellation transmitted though a noisy block fading channel. The received symbols, $Y_1, Y_2, \ldots, Y_N$, take the form

$$Y_n = \tilde{h} e^{2\pi j \tilde{w}_n / M} + X_n \tag{6.5.1}$$

where $\tilde{h}$ is a complex scalar representing the channel, the $X_1, \ldots, X_N$ are zero mean complex random variables and the $\tilde{w}_n \in \{0, 1, \ldots, M-1\}$ represent the data transmitted. The aim is to estimate the data $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_N$.

A number of algorithms have been devised for this purpose. Wilson et al. [1989] and Makrakis and Feher [1990] both propose algorithms of complexity $O(e^N)$. Liu et al. [1991] describe a suboptimal algorithm that requires $O(N^2)$ arithmetic operations. Warrier and Madhow [2002] describe an approximate least squares algorithm that they claim to require $O(N)$ operations. It was shown by Sweldens [2001] that the algorithm actually requires $O(N^2)$ operations in order for the approximation to remain valid as N increases[2]. Mackenthun [1994] proposed a least squares algorithm that required $O(N \log N)$ arithmetic operations. Later, Sweldens [2001] rediscovered the same algorithm. Here, we will describe two related algorithms that require only $O(N)$ arithmetic operations. We will firstly describe the least squares estimator found by Mackenthun [1994] and Sweldens [2001].

We write the sum of squares function

$$SS_c(h, w_1, w_2, \cdots) = \sum_{n=1}^{N} \left| Y_n - he^{2\pi i w_n/M} \right|^2$$

and consider the estimator that returns the minimisers of SS_c. Mackenthun [1994] and Sweldens [2001] showed how the minimisers could be found efficiently. The computational complexity of this algorithm is dominated by a sorting procedure and therefore requires $O(N \log N)$ operations. It is interesting to note that the algorithm bears a close resemblance to the log-linear time algorithm for the lattice A_n^* described in Section 3.5.2.

We will consider a different approach that is based on estimating the circular mean. Let $\tilde{h} = \tilde{\rho}e^{2\pi i\tilde{\mu}}$ and compute the complex arguments of the Y_n to obtain

$$\Theta_n = \frac{\angle Y_n}{2\pi} = \left\langle \tilde{\mu} + \frac{\tilde{w}_n}{M} + \Phi_n \right\rangle$$

where the $\Phi_n = \frac{1}{2\pi}\angle\tilde{\rho} + X_n$ are projected circular random variables. Multiplying both sides by M and taking fractional parts we obtain

$$\langle M\Theta_n \rangle = \langle M\tilde{\mu} + M\Phi_n \rangle .$$

Now the $M\tilde{\mu}$ can be estimated as the circular or unwrapped mean of the $\langle M\Theta_n \rangle$. A word of caution here is that there is no guarantee that the $\langle M\Theta_n \rangle$ is unimean or even that the true circular or unwrapped mean is equal to $M\tilde{\mu}$. Worse still, we are not even sure that $\langle M\Theta_n \rangle$ *has* circular and unwrapped means! The problem is that we have *rewrapped* the Θ_n by multiplying by M and taking the fractional part. It is likely possible to devise conditions on the X_n such that the $\langle M\Theta_n \rangle$ are unimean with circular and unwrapped means equal to $M\tilde{\mu}$ but we will not consider this in this thesis.

These somewhat theoretical considerations aside, we will find that, for the noise distributions of interest for PSK, it appears that applying either the sample circular

[2]Sweldens [2001] was actually published before Warrier and Madhow [2002]. It appears that Sweldens was working from a 1999 preprint of Warrier and Madhow [2002] that was available online.

mean or the angular least squares estimator to $\langle M\Theta_n \rangle$ produces excellent results in practice. Let $M\hat{\mu}$ be the estimate obtained. Then the estimate of the data is

$$\hat{w}_n = \lceil M\Theta_n - M\hat{\mu} \rfloor \operatorname{rem} M$$

where we recall that $a \operatorname{rem} b$ is the remainder of a/b.

An important property of noncoherent detection of M-PSK is the ambiguity between the transmitted data $w_1, w_2, \ldots, w_N$ and the transmitted data

$$(w_1 + k) \operatorname{rem} M, (w_2 + k) \operatorname{rem} M, \ldots, (w_n + k) \operatorname{rem} M$$

for some integer k. This is easily observed from SS_c because

$$SS_c(h, w_1, w_2, \ldots) = SS_c(e^{-2\pi k/M} h, (w_1 + k) \operatorname{rem} M, (w_2 + k) \operatorname{rem} M, \ldots)$$

These ambiguities can be resolved by differential encoding and this is the approach taken here [Weber, 1978; Mackenthun, 1994].

Simulations

Simulations were run to compare the bit error rate (BER) of the least squares estimator, the sample circular mean estimator and the angular least squares estimator for QPSK (i.e. $M = 4$) as the signal to noise ratio per bit (E_b/N_0) was varied from 2 dB to 10 dB and the number of symbols N was set to 4 and 40. The channel, $\tilde{h}$, was generated such that $\tilde{h} = e^{2\pi i \mu}$ where $\tilde{\mu}$ is uniformly distributed in the range $[0, 1)$. The noise, X_n, is independent and identically distributed Gaussian complex noise with independent real and imaginary parts having variance $N_0/2$.

The results are plotted in Figure 6.6. It is evident that there is negligible difference in performance between the various estimators. All perform better than the conventional 2-symbol differential detector. As N increases all estimators approach the performance of differentially encoded 4-PSK when perfect channel knowledge is available. These results are not surprising. Notice in Figure 6.4 that all of the estimators produce equally good estimates of phase when σ_c^2 is small. For the signal to noise ratios of interest in 4-PSK σ_c^2 *is* small and therefore the estimators obtain very similar channel estimates, and therefore have similar BERs. The primary advantage of the angular least squares and the sample circular mean estimators is that they require only $O(N)$ rather than $O(N \log N)$ operations.

6.6 Delay estimation from incomplete data

Consider sampling the time of arrival of N periodic events with known period T and some unknown delay $T\tilde{\mu}$ and assume that the sampling process is both noisy, so that the recorded event times do not exactly match the 'true' event times, and sparse, so that some of the events are missed. We may represent the N sample times, $Y_1, Y_2, \ldots, Y_N$, according to the model

$$Y_n = T\tilde{w}_n + T\tilde{\mu} + X_n \tag{6.6.1}$$

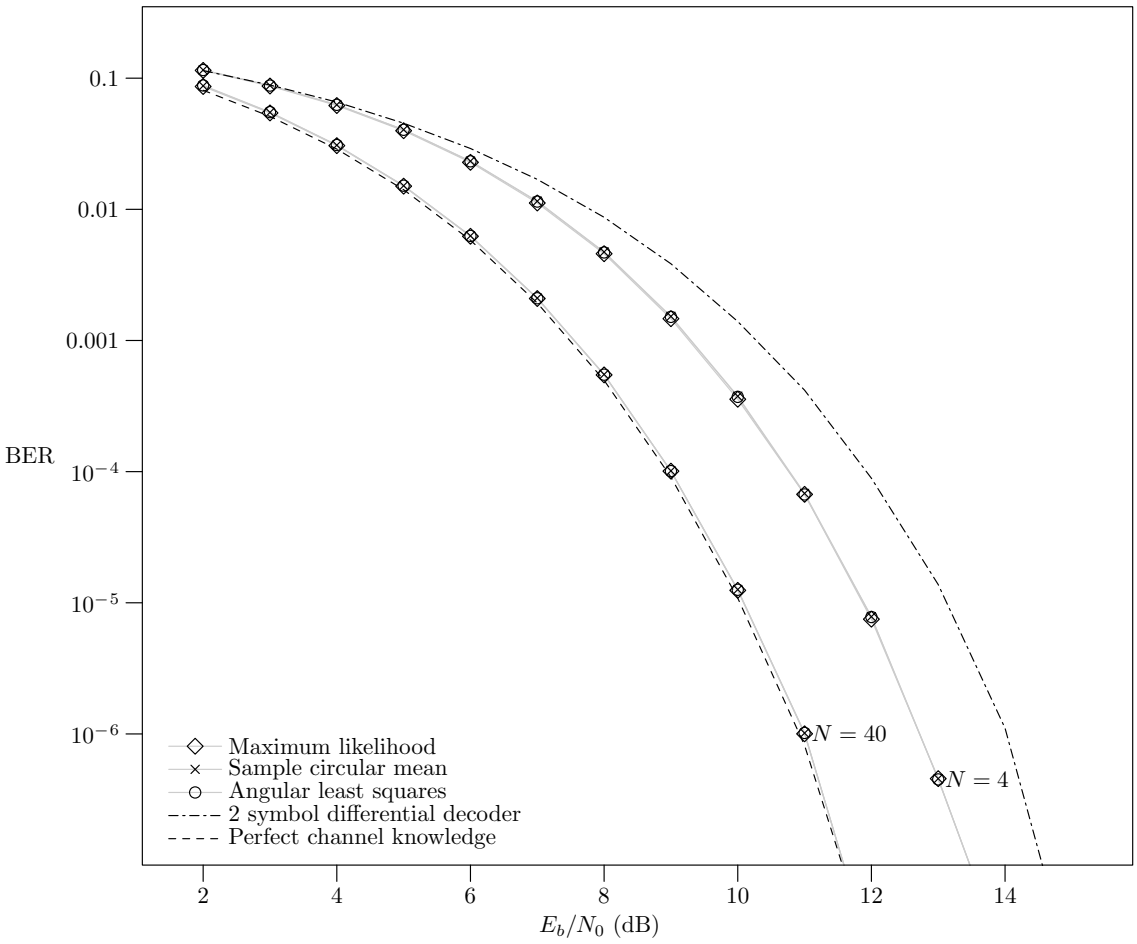


FIGURE 6.6: Bit Error Rate (BER) versus E_b/N_0.

where T is the period, the $\tilde{w}_n$ are *unknown* integers that represent which events were received and $X_1, \ldots, X_N$ are zero mean random variables. The aim is to estimate the delay parameter $\tilde{\mu}$. This model is depicted in Figure 6.7.

In the case that the $\tilde{w}_n$ are known a priori the least squares estimator is given by

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N}\frac{Y_n}{T} - \tilde{w}_n = \tilde{\mu} + \frac{1}{NT}\sum_{n=1}^{N} X_n. \tag{6.6.2}$$

If the X_n are zero mean independent and identically distributed with variance σ_g^2 then the estimator has variance

$$\operatorname{var}(\hat{\mu} - \tilde{\mu}) = \frac{\sigma_g^2}{NT^2}. \tag{6.6.3}$$

We are more interested in the case where the $\tilde{w}_n$ are *unknown*. This model occurs in practice in bit-synchronisation in telecommunications devices [Fogel and Gavish, 1989] and also pulse-train estimation in electronic support [Wiley, 1982]. The model is a simpler version of the more general problem of estimating both $\tilde{\mu}$ and T where T is assumed to be unknown [Fogel and Gavish, 1988; Clarkson et al., 1996; Clarkson,

2008; McKilliam and Clarkson, 2008; Sidiropoulos et al., 2005]. This generalisation is substantially more difficult to analyse and will not be addressed in this thesis.

Dividing (6.6.1) by T and taking fractional parts we obtain the circular random variables

$$\Theta_n = \left\langle \frac{Y_n}{T} \right\rangle = \left\langle \tilde{w}_n + \tilde{\mu} + \frac{X_n}{T} \right\rangle = \left\langle \tilde{\mu} + \left\langle \frac{X_n}{T} \right\rangle \right\rangle = \langle \tilde{\mu} + \Phi_n \rangle$$

where under appropriate assumptions concerning the X_n, the Φ_n are circular random variables with zero circular and unwrapped means. For example if the X_n are zero mean Gaussian random variables then the Φ_n are unimean wrapped normal random variables with zero circular and unwrapped means. If the X_n are uniformly distributed with zero mean and variance less than $\frac{T^2}{12}$ then the Φ_n are wrapped uniform random variables with zero circular and unwrapped means. In either case $\tilde{\mu}$ can be estimated as the circular or unwrapped mean of the Θ_n. We can again use either the angular least squares estimator or the sample circular mean estimator.

Figure 6.8 displays the performance of the estimators when the X_n are assumed to be independent and identically distributed Gaussian random variables with variance $T\sigma_g^2$. In this case the circular noise terms Φ_n have the WrappedNormal$(0, \sigma_g^2)$ distribution. The integer variables $\tilde{w}_n$ are chosen so that the differences $\tilde{w}_n - \tilde{w}_{n-1}$ are independent and Poisson distributed with mean 2 and the first integer $\tilde{w}_1$ is also Poisson distribution with mean 2. We see that the sample circular mean performs slightly better when $T\sigma_g^2$ is approximately greater than 0.05 and the angular least squares estimator performs slightly better when $T\sigma_c^2$ is approximately less than 0.05. Theorems 6.1 and 6.2 are again excellent predictors for the performance of the estimators. Also displayed is the asymptotic variance of the estimator in the case that the $\tilde{u}_n$ are known (6.6.3). When $T\sigma_g^2$ is sufficiently small the estimators perform very close to the performance attainable when the $\tilde{w}_n$ are known a priori and there is little lost by having incomplete data. However, when $T\sigma_g^2$ is large a significant performance penalty is paid by having incomplete data.

It is informative to view the pdf of the wrapped normal variables Φ_n. Figures 5.9, 5.10 and 5.11 show the pdf when $\sigma_g^2 = 0.02, 0.1$ and 0.2. It is clear that f makes a rather abrupt transition from 'Gaussian looking' to 'uniform looking' at around $\sigma_g^2 \approx 0.1$. It is obviously unreasonable to expect the estimators to perform well when f is close to the circular uniform distribution and this gives an intuitive explanation for the results observed in Figure 6.8.

6.7 Computational considerations

Both the sample circular mean and the angular least squares estimator require $O(N)$ operations to compute. However, we find that in practice the estimators differ somewhat in their computational complexity. The varying complexity depends highly on the problem at hand and arises largely from the need (or lack of need) to perform trigonometric operations. The disparity is likely to be more evident on small computing devices where trigonometric operations are expensive.

Firstly, consider the problems of phase estimation and noncoherent detection discussed in Sections 6.4 and 6.5. Here the N observations obtained are complex

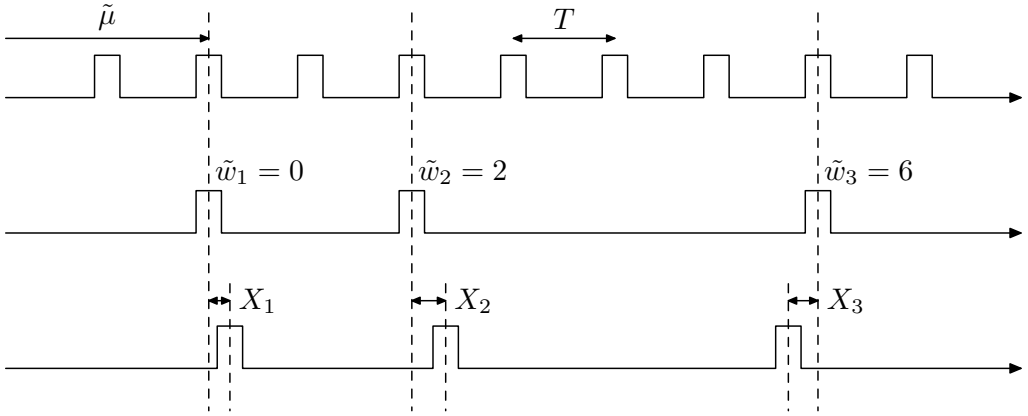


FIGURE 6.7: Delay estimation from incomplete data.

numbers. In order to use the angular least squares estimator we must first compute the complex argument of the observations and this requires computing the arctangent N times. For the sample circular mean we need only convert the complex numbers to unit magnitude and this is typically less computationally expensive than computing the arctangent. The sample circular mean does require a single arctangent operation once the average of the unit complex numbers is computed.

On the other hand, consider applications where the angles are measured directly, for example, when wind direction is measured N times over the course of the day. In this case the sample circular mean requires converting these N angles into complex numbers. This requires N cosine operations and N sine operations and also a single arctangent operation to convert the complex mean back to an angle. This is $2N + 1$ trigonometric operations in total. By contrast, the angular least squares estimator requires no trigonometric operations at all. We need only compute a nearest lattice point in the lattice A^*_{N-1}. If Algorithm 3.6 is used then the vast majority of operations are either memory operations or floating point addition, subtraction, comparison. These operations are typically very fast. For this reason, the angular least squares estimator is likely to be a better choice from a computational point of view when the angles are measured directly.

6.8 Summary

In Chapter 5 we saw two notions of **mean direction** that both conform well with intuition, these are the **circular mean** and the **unwrapped mean**. In this chapter we have considered methods for estimating these means. Applications of these estimators are widespread throughout science and engineering.

The first is the **sample circular mean** estimator of the circular mean. This estimator has received significant attention in the literature. Theorem 6.1 showed that the sample circular mean is strongly consistent and derived its central limit theorem. The second estimator is the **angular least squares estimator** of the unwrapped mean. We showed how this estimator can be rapidly computed by finding a nearest point in the lattice A^*_n. Theorem 6.2 showed that the angular least squares

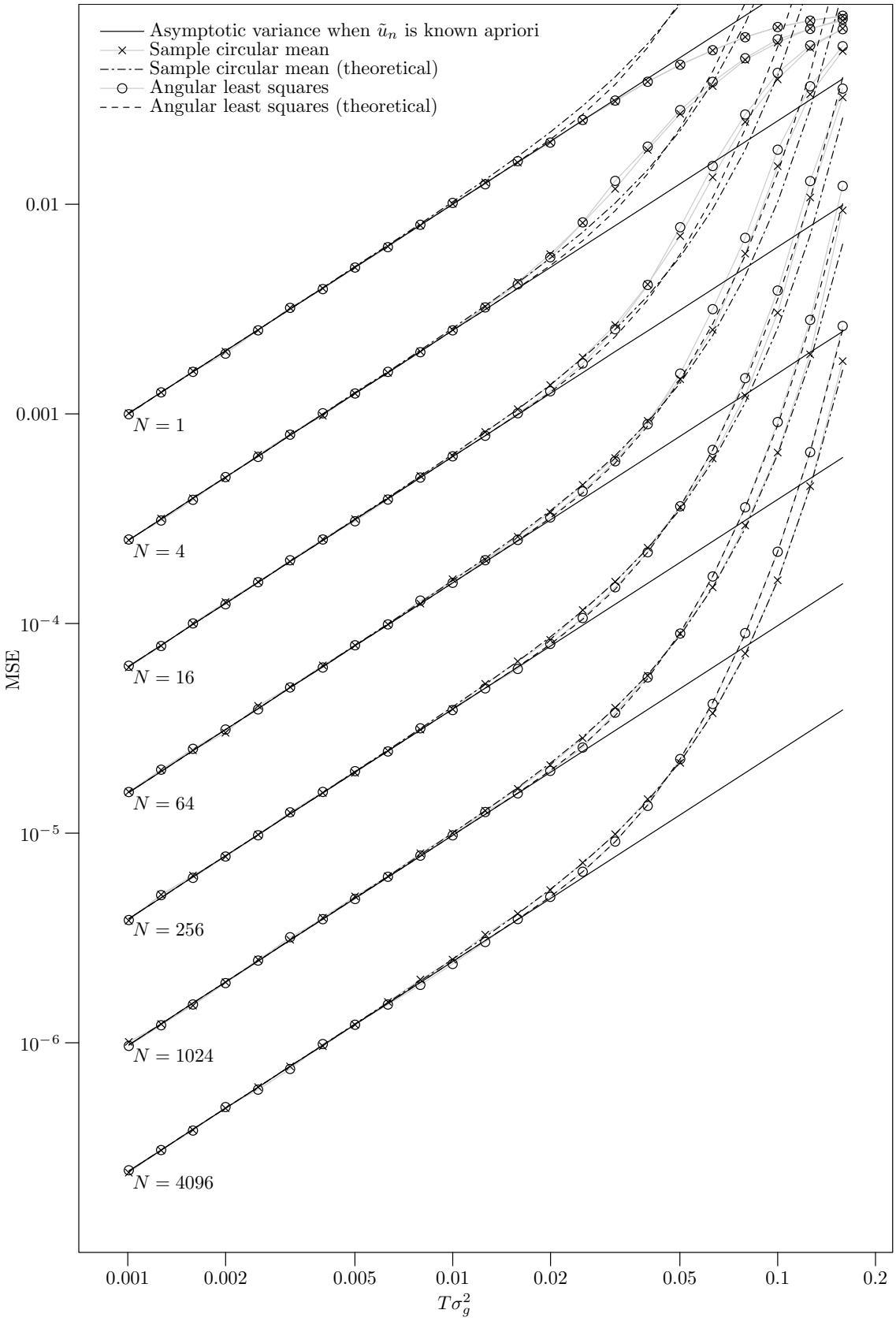


Figure 6.8: MSE versus $T\sigma_g^2$ where X_n is normally distributed.

estimator is strongly consistent and satisfied a central limit theorem.

In Section 6.3 we considered the performance of these estimators for some unimean distributions. We found that the angular least squares estimator tends to perform better when the distribution is 'uniform-like' whereas the sample circular mean tends to perform better when the distribution is 'von Mises-like'. We also found that the performance of the estimators is very accurately modeled by the central limit theorems derived in Theorems 6.1 and 6.2.

In Sections 6.4, 6.5 and 6.6 we applied the estimators to the problems of **phase estimation**, **noncoherent detection**, and **delay estimation**. For phase estimation we found that it is probably better to simply use the standard least squares estimator, but we hinted at how the angular least squares squares estimator will be computationally a more feasible choice in Chapter 10 where, rather than *constant* phase estimation, we consider polynomial phase estimation.

For noncoherent detection of PSK signals we found that highly accurate detection could be performed in practice using the angular least squares or the sample circular mean estimators. This approach is computationally attractive because it requires only a linear number of operations in the **block length**, whereas existing least squares approaches require a log-linear number of operations.

We also considered the problem of delay estimation from noisy and incomplete data. This problem has applications to bit-synchronisation in telecommunications devices [Fogel and Gavish, 1989] and also pulse-train estimation in electronic support [Wiley, 1982]. It was observed that the angular least square estimator or the sample circular mean could be used to produce very accurate estimates of the delay regardless of the amount of data that is missing. However, if the noise level is very high, then a significant accuracy penalty is paid for having incomplete data.

Finally, in Section 6.7 we discussed some of the computational properties of the two estimators. In particular we focused on the number of trigonometric operations that are required. We found that if the N observations are complex numbers, such as in the problem of phase estimation, then the sample circular mean requires only a single arctangent operation, but the angular least squares estimator requires N arctangent operations. On the other hand, if the angles are observed directly, as is likely to be the case in meteorology and other applications, then the sample circular mean requires $2N + 1$ trigonometric operations, but the angular least squares estimator does not require any. If trigonometric operations are particularly expensive, as is typically the case on small computing devices, then consideration of these properties will likely lead to computational savings.

Part III

Polynomial phase signals

—When you are a Bear of Very Little Brain, and you Think of Things, you find sometimes that a Thing which seemed very Thingish inside you is quite different when it gets out into the open and has other people looking at it.

A. A. Milne

7

The aliasing of polynomial phase signals

In Part III of this thesis we study **polynomial phase signals**. These signals have substantial application in science, in particular in astronomy, optics and biology and also in engineering, particularly in communications and radar. Of significant practical importance is the estimation of a polynomial phase signal from a number of noisy observations, or *samples* and this estimation problem is the focus of Chapters 8, 9 and 10. However, before we consider this estimation problem we must understand the phenomenon of **aliasing** that occurs when polynomial phase signals are sampled. This aliasing is the subject of this chapter.

In Chapter 5 we introduced circular statistics and gave two different definitions for the **mean direction**, the **unwrapped mean** and the **circular mean**. In Chapter 6 we considered methods for estimating these means from a number of observations of a circular random variable. In these chapters we implicitly assumed that *angles* were restricted to take values in the interval $[-1/2, 1/2)$. This restriction naturally occurred because angles are equivalent modulo one.

A similar phenomenon occurs for sampled polynomial phase signals. It turns out that two (or more) distinct polynomial phase signals can sometimes *take exactly the same values* when they are sampled. We call such signals **aliases** and in Section 7.1 we completely describe how the aliasing occurs using some ideas from lattice theory and also the integer valued polynomials that we introduced in Chapter 4. These aliasing results are also given by McKilliam and Clarkson [2009], but the presentation here is more thorough. The results have also been independently discovered by Abatzoglou [1986] and Ängeby [2000a] for polynomial phase signals of order 2, but we generalise this to polynomial phase signals of any order.

In Chapters 8, 9 and 10 we will want to estimate the coefficients of a polynomial phase signal from a number of noisy samples and understanding the nature of this aliasing is of paramount importance. In Section 7.2 we show that in order to ensure the **identifiability** of any estimator we must restrict the polynomial coefficients to a particular region. We call this the **identifiable region** and we show how

it can be represented as the **tessellating region** of a particular lattice. We also describe how to resolve aliased coefficients, compute square error between coefficients unambiguously and generate coefficients uniformly in the identifiable region.

7.1 Sampling polynomial phase signals

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

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

where t is a real number, typically time, and

$$y(t) = \mu_0 + \mu_1 t + \mu_2 t^2 + \dots \mu_m t^m$$

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. In practice the signal obtained is typically *sampled* at discrete points in time, t. In this thesis we only consider **uniform sampling**, that is, where the gap between consecutive samples is a 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

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

where n is an integer. The phase of $s(n)$ is described by the sampled polynomial

$$y(n) = \mu_0 + \mu_1 n + \mu_2 n^2 + \dots + \mu_m n^m.$$

Recall from Section 4.2 (page 55) that we defined $\mathcal{Z}$ to 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* polynomial 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 we get

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

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 and corollary.

Theorem 7.1. *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, $y - z \in \mathcal{Z}$.*

Corollary 7.1. *The polynomial phase signals $s(t)$ and $r(t)$ are aliases if and only if $s(t) = r(t)e^{2\pi jp(t)}$ where p is a polynomial from $\mathcal{Z}$.*

It may be helpful to observe Figures 7.1, 7.2, 7.3 and 7.4. 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 is plotted (given by taking the fractional part of the polynomials on the left). The circles display the *samples* at the integers. Note that the samples of the principal components intersect and the corresponding polynomial phase signals will be 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 Chapters 8, 9 and 10. We will make use of the coef$(\cdot)$ notation for converting polynomials to vectors that was introduced on page 54, that is, for y a polynomial of order m, coef(y) is the column vector of length $m+1$ containing the coefficients of y. 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 coef$(y) =$ coef$(z) +$ coef(p)[1]. Consider the set of vectors coef(p) for all polynomials $p \in \mathcal{Z}$, that is

$$L_{m+1} = \{\text{coef}(p) \mid p \in \mathcal{Z}\}.$$

Recall from Chapter 4 that we defined the **integer valued polynomials**, p_k, and showed in Lemma 4.1 (page 56) that the p_k form an integer basis for $\mathcal{Z}$. So

$$\begin{aligned} L_{m+1} &= \{\text{coef}(c_0p_0 + c_1p_1 + \cdots + c_mp_m) \mid c_i \in \mathbb{Z}\} \\ &= \{c_0\,\text{coef}(p_0) + c_1\,\text{coef}(p_1) + \cdots + c_m\,\text{coef}(p_m) \mid c_i \in \mathbb{Z}\}. \end{aligned}$$

Also, recall that we defined $\boldsymbol{\mathcal{P}}$ (Section 4.2 page 55) as the $m+1$ by $m+1$ matrix with columns given by the coefficients of the integer valued polynomials, that is, $\boldsymbol{\mathcal{P}}$ is the matrix

$$\boldsymbol{\mathcal{P}} = \left[\begin{array}{cccc} \text{coef}(p_0) & \text{coef}(p_1) & \ldots & \text{coef}(p_m) \end{array}\right].$$

Then,

$$L_{m+1} = \{\boldsymbol{\mathcal{P}}\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}\} = \boldsymbol{\mathcal{P}}\mathbb{Z}^{m+1}$$

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 $\boldsymbol{\mathcal{P}}$. We can restate Theorem 7.1 as:

Corollary 7.2. *Two polynomial phase signals $s(t) = e^{2\pi jy(t)}$ and $r(t) = e^{2\pi jz(t)}$ are aliases if and only if* coef(y) *and* coef(z) *differ by a lattice point in L_{m+1}.*

7.2 Estimation and identifiability

In Chapters 8, 9 and 10 we will want to estimate the coefficients $\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_m$ of a polynomial phase signal of order m, from N noisy observations $Y_1, Y_2, \ldots, Y_N$ of the form

$$Y_n = e^{2\pi(\tilde{\mu}_0 + \tilde{\mu}_1 n + \cdots + \tilde{\mu}_m n^m)} + X_n$$

[1] In group theory terminology coef$(\cdot)$ coupled with vector addition is called a **group homomorphism**.

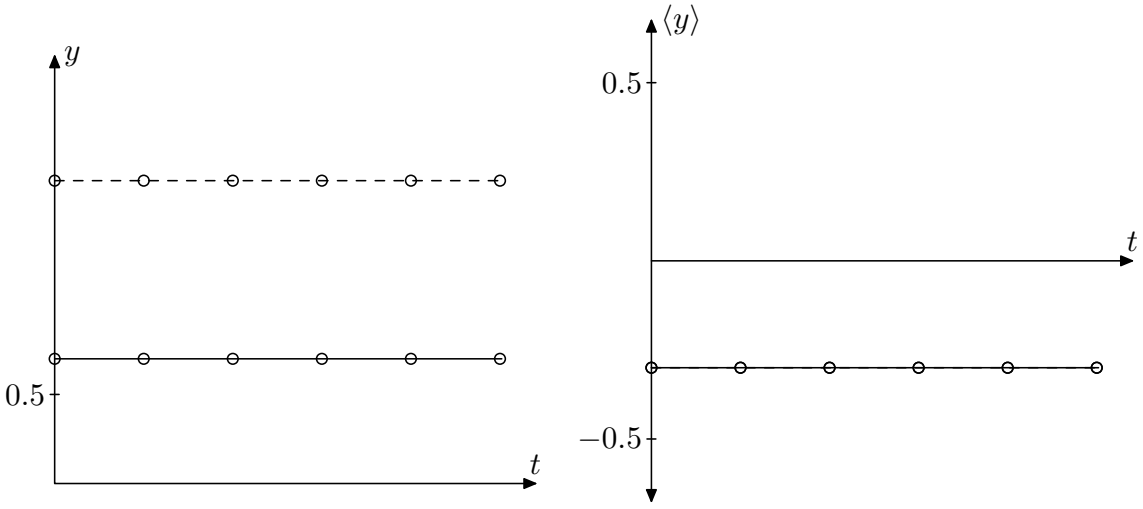


FIGURE 7.1: The zeroth order polynomials $\frac{7}{10}$ (solid line) and $\frac{17}{10}$ (dashed line).

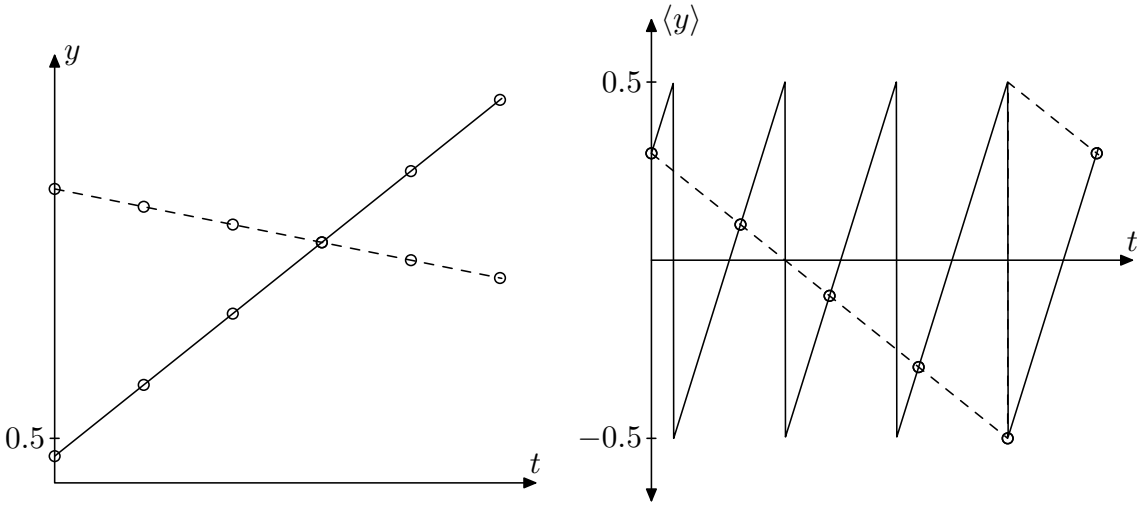


FIGURE 7.2: The first order polynomials $\frac{1}{10}(3+8t)$ (solid) and $\frac{1}{10}(33-2t)$ (dashed line).

where $n \in \mathbb{Z}$ and the X_n are random variables representing the noise. In order to ensure the **identifiability** of any estimator of the coefficients we must take account of aliasing. That is, we must restrict the set of allowable coefficients so that no two polynomial phase signals are aliases. In consideration of Corollary 7.2 we require that the coefficients, written in vector form $\boldsymbol{\mu}$, are contained in a set of coset representatives for the quotient $\mathbb{R}^{m+1}/L_{m+1}$, or in other words they are contained in a tessellating region of the lattice L_{m+1} (see Section 2.2). We call the chosen tessellating region the **identifiable region**.

To give an example consider a polynomial phase signal of order zero, then $e^{2\pi j\mu_0} = e^{2\pi j(\mu_0+k)}$ for any integer k and in order to ensure identifiability, we must restrict the μ_0 to some interval of length 1. A natural choice is the interval $[-1/2, 1/2)$. When $m = 0$ 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

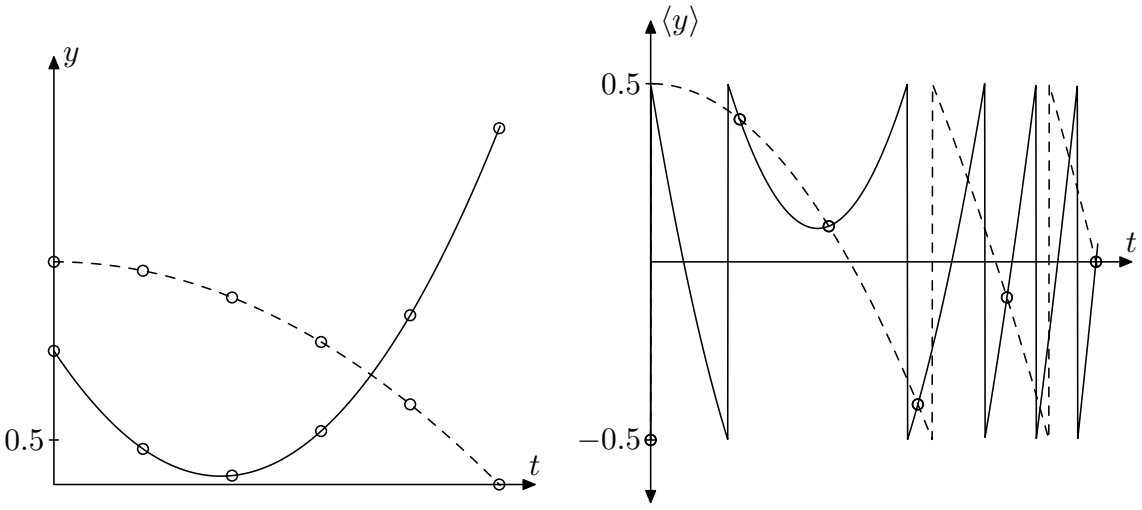


Figure 7.3: The quadratic polynomials $\frac{1}{10}(15 - 15t + 4t^2)$ (solid line) and $\frac{1}{10}(25 - t^2)$ (dashed line).

a natural choice of identifiable region is the square box $[-1/2, 1/2)^2$. This result is also quite intuitive as it 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} is not given by the integer lattice. However, we can always construct an identifiable region by taking a tessellating region of lattice L_{m+1}.

There are an infinity of choices for the identifiable region. A natural choice is the Voronoi cell of L_{m+1} and this was used by McKilliam and Clarkson [2009]. Another potential choice is a fundamental parallelepiped of L_{m+1}. In this thesis we shall use the rectangular tessellating region constructed using Proposition 2.1 (page 17). As the generator $\boldsymbol{\mathcal{P}}$ is upper triangular with kth diagonal given by $\frac{1}{k!}$ this rectangular tessellating region is

$$B = \prod_{k=0}^{m} \left[-\frac{0.5}{k!}, \frac{0.5}{k!} \right). \tag{7.2.1}$$

We will make use of this region when deriving the statistical properties of the angular least squares estimator in Section 8.2.

There are a number of useful things that we can do now that we have a definition of the identifiable region B. We can resolve aliased coefficients, that is, given some polynomial coefficients not necessarily in the identifiable region, we can find the equivalent coefficients in the identifiable region. For evaluating the performance of estimators it is convenient to calculate the square error between the true and estimated polynomial coefficients. Some problems arise due to aliasing, but, we will show how to compute square error in an unambiguous way. Finally we show how to generate coefficients that are uniformly distributed in the identifiable region. This is useful if we wish to evaluate the performance of an estimator over the entire identifiable region. These procedures make use of the function dealias($\boldsymbol{\mu}$) that maps the vector of coefficients $\boldsymbol{\mu}$ to its equivalent point inside the identifiable region B.

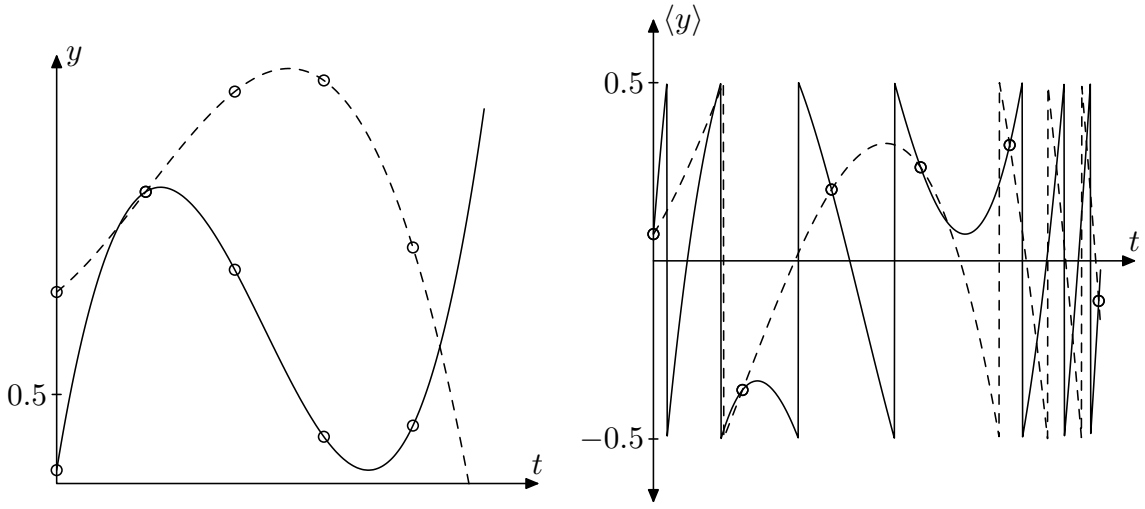


FIGURE 7.4: The cubic polynomials $\frac{1}{160}(174 + 85t - 118t^2 + 40t^3)$ (solid line) and $\frac{1}{48}(84 + 19t + 12t^2 - 4t^3)$ (dashed line).

This can be implemented as

$$\text{dealias}(\boldsymbol{\mu}) = \boldsymbol{\mu} - \mathbf{p}$$

where $\mathbf{p}$ is the lattice point in L_{m+1} computed by applying Algorithm 2.2 (page 31) to $\boldsymbol{\mu}$.

7.2.1 Resolving aliasing

Given the polynomial coefficients $\mu_0, \ldots, \mu_m$ the equivalent coefficients within the identifiable region B are given, in vector form, by $\text{dealias}(\boldsymbol{\mu})$.

7.2.2 Computing square error

When evaluating the performance of an estimator by simulation we usually have some true coefficients $\tilde{\mu}_0, \ldots, \tilde{\mu}_m$ and obtain some estimated coefficients $\hat{\mu}_0, \ldots, \hat{\mu}_m$. In order to gauge the accuracy of the estimate we typically wish to compute the square error between the *true* and estimated coefficients, that is, the values $(\tilde{\mu}_k - \hat{\mu}_k)^2$. Some difficulties arise due to aliasing. For example, consider when $m = 0$. It may be that the true coefficient is $\tilde{\mu}_0 = 0.4$ and the estimated coefficient is $\hat{\mu}_0 = -0.4$. Naïvely we might compute the square error as $(\tilde{\mu}_0 - \hat{\mu}_0)^2 = 0.8^2$. Intuitively this is wrong because $\tilde{\mu}_0$ and $\hat{\mu}_0$ are angles that are close together on the circle (see Figure 7.5). We can correctly compute the square error as

$$(\hat{\mu}_0 - \tilde{\mu}_0 - \lceil \hat{\mu}_0 - \tilde{\mu}_0 \rfloor)^2 = \langle \hat{\mu}_0 - \tilde{\mu}_0 \rangle^2 = 0.2^2.$$

Analogously, to compute the square error for any $m \geq 0$ we first compute the vector

$$\boldsymbol{\lambda} = \text{dealias}(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}), \tag{7.2.2}$$

then square error of the kth coefficient is given by λ_k^2.

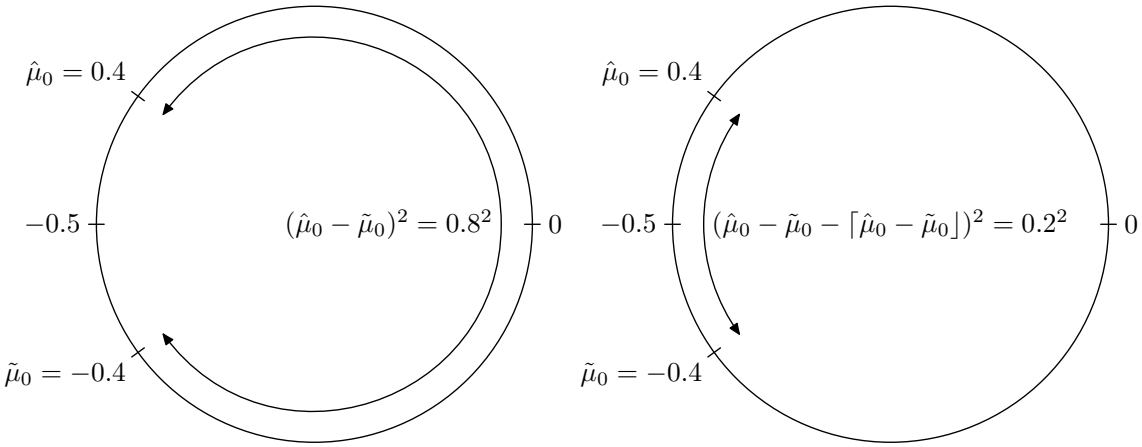


FIGURE 7.5: The figure on the left incorrectly computes the square error between the *true* coefficient $\tilde{\mu}_0 = -0.4$ and the estimated coefficient $\hat{\mu}_0 = 0.4$. The figure on the right correctly computes the error.

7.2.3 Generating coefficients

A coefficient uniformly distributed in the identifiable region B can be generated as

$$\boldsymbol{\mu} = \text{dealias}(\boldsymbol{\mathcal{P}}\mathbf{w}) \tag{7.2.3}$$

where $\mathbf{w}$ is a vector with elements that are independent and uniformly distributed on $[0, 1)$.

7.3 Summary

In this chapter we have investigated the phenomenon of aliasing that occurs when polynomial phase signals are sampled uniformly. We found that in order to ensure the identifiability of any estimator for the polynomial phase coefficients we must restrict the set of allowable coefficient to an **identifiable region**. We showed how a suitable identifiable region is described as a tessellating region of the lattice L_{m+1} with generator matrix described using the coefficients of the integer valued polynomials. There are potentially many choices for the identifiable region, one for every distinct tessellating region. However, for this thesis we will always use the rectangular region given by B (7.2.1). Using this lattice we show how to resolve aliased parameters, compute square error and generate parameters uniformly in an identifiable region. These procedures will be useful in the next chapters where we consider the estimation of polynomial phase signals.

—The computer can't tell you the emotional story. It can give you the exact mathematical design, but what's missing is the eyebrows.

Frank Zappa

8

Angular least squares and its asymptotic properties

In Chapter 6 we introduced the **angular least squares** estimator of the unwrapped mean of a circular random variable and in Section 6.4 we showed how this estimator could be used for **phase estimation**. In this chapter we generalise the angular least squares estimator for the estimation of polynomial phase signals. We will apply the theory developed in this chapter to **frequency estimation** and **polynomial phase estimation** in Chapters 9 and 10.

We will show in Section 8.1 how the angular least squares estimator for a polynomial phase signals of order m can be computed by finding a nearest lattice point in the lattice $V^*_{n/m}$. This is analogous to how the unwrapped mean was estimated by angular least squares in Section 6.2 by computing a nearest lattice point in A^*_n. It is no coincidence that $V^*_{n/0} = A^*_n$ (see Section 4.1). We derived a polynomial time algorithm to compute a nearest point in $V^*_{n/m}$ in Section 4.3 and it will follow, by using this algorithm, that the angular least squares estimator for polynomial phase signals can be computed in a number of operations that is polynomial in the number of observations N.

In Section 8.2 we derive the statistical properties of the angular least squares estimator. We show that the estimator is strongly consistent and we derive its central limit theorem. For the case of polynomial phase signals of order greater than one, the statistical results we derive in Section 8.2 are the first of their kind.

8.1 Angular least squares and the lattice $V^*_{n/m}$

In this section we describe the angular least squares estimator for the coefficients of a polynomial phase signal. Consider observing N complex numbers $Y_1, Y_2, \ldots, Y_N$

of the form

$$Y_n = \rho e^{2\pi(\tilde{\mu}_0+\tilde{\mu}_1 n+\cdots+\tilde{\mu}_m n^m)} + X_n$$

where the X_n are zero mean complex random variables and ρ is an unknown amplitude. The aim is to estimate the polynomial coefficients $\tilde{\mu}_0, \ldots, \tilde{\mu}_m$. For the angular least squares estimator we take the complex argument of the Y_n and divide by 2π to obtain the circular random variables

$$\Theta_n = \left\langle \Phi_n + \tilde{\mu}_0 + \tilde{\mu}_1 n + \tilde{\mu}_2 n^2 + \cdots + \tilde{\mu}_m n^m \right\rangle \tag{8.1.1}$$

where the $\Phi_n = \frac{1}{2\pi}\angle(\rho + X_n)$ are projected circular random variables. We define the sum of squares function

$$SS(\mu_0, \mu_1, \ldots, \mu_m) = \sum_{n=1}^{N} \left\langle \Theta_n - \sum_{k=0}^{m} \mu_k n^k \right\rangle^2. \tag{8.1.2}$$

The angular least squares estimator is defined as the coefficients $\hat{\mu}_0, \ldots, \hat{\mu}_m$ that minimise SS over the identifiable region B (7.2.1). That is, in vector form,

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}\in B} SS(\boldsymbol{\mu}). \tag{8.1.3}$$

We will show how this estimator is connected with the lattice $V^*_{N-m-1/m}$. Our presentation mimics that of Section 6.2. We can write the sum of squares function SS as

$$SS(\mu_0, \mu_1, \ldots, \mu_m) = \sum_{n=1}^{N} \left(\Theta_n - W_n - \sum_{k=0}^{m} \mu_k n^k \right)^2$$

where $W_n = \left\lceil \Theta_n - \sum_{k=0}^{m} \mu_k n^k \right\rfloor$ are the integer wrapping variables. If we consider the W_n as nuisance parameters to be estimated then SS can be written as a function of both the coefficients $\mu_0, \ldots, \mu_m$ and the W_n. The angular least squares estimator is then found by minimising over the $\mu_0, \ldots, \mu_m$ and the W_n. This joint minimisation problem can be solved by computing a nearest point in the lattice $V^*_{N-m-1/m}$. To see this write SS as a function of both the $\mu_0, \ldots, \mu_m$ and the W_n using vectors as

$$SS(\boldsymbol{\mu}, \mathbf{w}) = \|\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\mu} - \mathbf{w}\|^2 \tag{8.1.4}$$

where $\mathbf{X}$ is the rectangular Vandermonde matrix defined in (4.2.2) and where we define the column vectors $\mathbf{w} = [W_1, \ldots, W_N]^\dagger$ and $\boldsymbol{\theta} = [\Theta_1, \ldots, \Theta_N]^\dagger$ and the column vector of coefficients $\boldsymbol{\mu} = [\mu_0, \ldots, \mu_m]^\dagger$. Fixing $\mathbf{w}$ and minimising with respect to $\boldsymbol{\mu}$ gives

$$\hat{\boldsymbol{\mu}} = (\mathbf{X}^\dagger\mathbf{X})^{-1}\mathbf{X}^\dagger(\boldsymbol{\theta} - \mathbf{w}) = \mathbf{X}^+(\boldsymbol{\theta} - \mathbf{w}) \tag{8.1.5}$$

where $\mathbf{X}^+ = (\mathbf{X}^\dagger\mathbf{X})^{-1}\mathbf{X}^\dagger$ is the pseudoinverse of $\mathbf{X}$.

Substituting this into $SS(\boldsymbol{\mu}, \mathbf{w})$ we obtain the sum of squares function conditioned on minimisation with respect to $\boldsymbol{\mu}$ as

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

where $\mathbf{Q} = \mathbf{I} - \mathbf{X}\mathbf{X}^+$ is the orthogonal projection matrix defined in (4.2.12) on page 57. It follows that $\mathbf{Q}\mathbf{w}$ is a lattice point in $V^*_{N-m-1/m}$ and that minimising $SS(\mathbf{w})$ is equivalent to finding the nearest lattice point in $V^*_{N-m-1/m}$ to $\mathbf{Q}\boldsymbol{\theta}$. We can use, for example, the algorithm described in Section 4.3 to compute the nearest point in a number of operations that is polynomial in N. Denote this point by $\mathbf{Q}\hat{\mathbf{w}}$. Now the estimate $\hat{\boldsymbol{\mu}}$ is given by substituting $\hat{\mathbf{w}}$ for $\mathbf{w}$ in (8.1.5). After this procedure it is possible that the $\hat{\boldsymbol{\mu}}$ obtained is not in the identifiable region but instead is an *aliased version* of the desired estimate. This can be resolved using the dealiasing procedure described in Section 7.2.1, i.e. by computing dealias($\boldsymbol{\mu}$).

8.2 Asymptotic properties of angular least squares

In this section the angular least squares estimator is shown to be strongly consistent and its central limit theorem is derived. The main result is Theorem 8.1, the proof of which is given in two parts within this section. The reader not interested in the technical details of the proof could read the statement of Theorem 8.1 and then move onto the remaining chapters that involve the application of the angular least squares estimator to the problems of frequency estimation and polynomial phase estimation.

Theorem 8.1. *Let $\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_m$ be the minimisers of the sum of squares function SS from* (8.1.2) *over the identifiable region B. That is, in vector form, $\hat{\boldsymbol{\mu}}$ is given by* (8.1.3)*. Let $\hat{\boldsymbol{\lambda}} = \text{dealias}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})$ be the dealiased difference between the true and estimated coefficients. If the circular random variables $\Phi_1, \ldots, \Phi_N$ are independent and identically distributed with zero unwrapped mean and pdf f then:*

1. *(Strong consistency) The normalised elements $N^k\hat{\lambda}_k$ converge almost surely to zero as $N \to \infty$ for all $k = 0, 1, \ldots, m$.*

2. *(Central limit theorem) If the periodic function $f(\langle x \rangle)$ is continuous at $x = -1/2$ and $f(-1/2) \neq 1$ then the distribution of the vector*

$$\left[\begin{array}{cccc} N^{1/2}\hat{\lambda}_0 & N^{3/2}\hat{\lambda}_1 & \ldots & N^{(2m+1)/2}\hat{\lambda}_m \end{array}\right]^\dagger$$

converges to the normal with zero mean and covariance matrix

$$\frac{\sigma^2}{\left(1 - f(-1/2)\right)^2}\mathbf{C}^{-1} \tag{8.2.1}$$

where σ^2 is the unwrapped variance of the Φ_n and $\mathbf{C}$ is the $(m+1) \times (m+1)$ matrix with elements $C_{i,j} = 1/(i+j-1)$.

The proof of this theorem is broken over the next two sections. Section 8.2.1 proves the strong consistency and Section 8.2.2 proves the central limit theorem. The proofs for the simpler case of $m = 1$ where given by McKilliam et al. [2010a] and the proofs here take a similar approach.

The theorem describes conditions on the dealiased difference $\hat{\boldsymbol{\lambda}} = \text{dealias}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})$ between the true coefficients $\tilde{\boldsymbol{\mu}}$ and the estimated coefficients $\hat{\boldsymbol{\mu}}$ rather than directly

on the difference $\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}$. This makes intuitive sense in view of the discussion in Section 7.2.2 where we described the *correct* way to compute error between polynomial phase coefficients.

For strong consistency we require that the Φ_n are independent and identically distributed with zero unwrapped mean and pdf f. For the central limit theorem we include the requirement that the periodic function $f(\langle x \rangle)$ is continuous at $x = -1/2$ and that $f(-1/2) \neq 1$. If these conditions are not met then other expressions for the asymptotic covariance can potentially be found, but we will not consider this. Circular distributions that do not satisfy these requirements are highly unlikely to be needed in practice. For all of the distributions considered in this thesis these requirements hold. We will make some discussion about these assumptions after the proof.

Before we begin it is worthwhile stating that $\mathbf{C}$ is known as a **Hilbert matrix** and that the elements of the inverse $\mathbf{C}^{-1}$ are given by

$$C_{i,j}^{-1} = (-1)^{i+j}(i+j-1)\binom{n+i-1}{n-j}\binom{n+j-1}{n-i}\binom{i+j-2}{i-1}^2.$$

Hilbert matrices are particularly ill conditioned and difficult to numerically invert. This problem can be avoided using the above formula.

8.2.1 Strong consistency

Substituting (8.1.1) into SS we obtain

$$\begin{aligned} SS\left(\mu_0, \mu_1, \ldots, \mu_m\right) &= \sum_{n=1}^{N} \left\langle \left\langle \Phi_n + \sum_{k=0}^{m} \tilde{\mu}_k n^k \right\rangle - \sum_{k=0}^{m} \mu_k n^k \right\rangle^2 \\ &= \sum_{n=1}^{N} \left\langle \Phi_n + \sum_{k=0}^{m} (\tilde{\mu}_k - \mu_k) n^k \right\rangle^2. \end{aligned}$$

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 the lattice L_{m+1} (see Chapter 7). So, from the definition of L_{m+1} in terms of integer valued polynomials we have that $p_0 + p_1 n + \cdots + p_m n^m$ is an integer whenever n is an integer and therefore

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

and the sum of squares function can be written using $\boldsymbol{\lambda}$ as

$$SS\left(\mu_0, \mu_1, \ldots, \mu_m\right) = \sum_{n=1}^{N} \left\langle \Phi_n + \sum_{k=0}^{m} \lambda_k n^k \right\rangle^2 = N S_N\left(\lambda_0, \lambda_1, \ldots, \lambda_m\right),$$

where $S_N(\boldsymbol{\lambda}) = \frac{1}{N} SS(\boldsymbol{\mu})$. From the definition of the dealias$(\cdot)$ function we have that $\boldsymbol{\lambda}$ is inside the identifiable region B so the elements of $\boldsymbol{\lambda}$ satisfy

$$-\frac{0.5}{k!} \leq \lambda_k < \frac{0.5}{k!}. \tag{8.2.2}$$

Now the dealiased difference between the true and estimated coefficients $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda} = \text{dealias}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})$ is the minimiser of S_N over B. We shall show that the minimiser of S_N is such that the elements $N^k \hat{\lambda}_k \to 0$ almost surely as $N \to \infty$ for all $k = 0, 1, \ldots, m$ and from this the proof of strong consistency will follow.

We take the following approach. We first show, in Lemma 8.1, that $S_N(\boldsymbol{\lambda})$ converges in a strong sense (almost surely and uniformly in $\boldsymbol{\lambda} \in B$) to its expectation $ES_N(\boldsymbol{\lambda})$. We can then reason about the minimisers of S_N using the minimisers of its expectation ES_N and this leads, via Lemma 8.10, to the proof. The basis of this approach is described in a general fashion by Amemiya [1985, Theorem 4.1.1]. We start by proving that S_N converges almost surely and uniformly to its expectation.

Lemma 8.1. *The function $S_N(\boldsymbol{\lambda})$ converges almost surely and uniformly in $\boldsymbol{\lambda} \in B$ to its expectation $ES_N(\boldsymbol{\lambda})$. That is*

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

almost surely as $N \to \infty$.

Before we begin this proof it is worth noting that this type of result is common to a body of literature that, in recent times, has been called the **uniform law of large numbers**. Early results were given by, for example, Jenrich [1969] and Hoadley [1971]. These results typically make rather strong assumptions about the function S_N (or equivalently SS) and also about the random variables Φ_n that model the noise. For example, Jenrich [1969] assumes that the noise variables Φ_n are independent and identically distributed (as we have here). More modern results are given by Andrews [1987], Pötscher and Prucha [1989] and Newey and McFadden [1994] and these make weaker assumptions about S_N and the Φ_n. An overview of some of these techniques is given by Amemiya [Amemiya, 1985, Chapter 4], however, it is fair to say that this literature is still, somewhat, scattered. It is worth pointing out that the majority of results in the literature only prove convergence in *probability*, whereas here, we will prove (and prefer) the stronger mode of convergence *almost surely*.

It is tempting to try to map our problem to one of the techniques in the literature, and leverage some existing work in order to prove Lemma 8.1. However, we have found that the mapping process is more complicated than proving the lemma directly using some well known results in probability theory, namely **Markov's inequality** and the **Borel-Cantelli lemma** [Billingsley, 1979, page 46]. However, we note that the more high powered techniques in the literature might allow this proof to be made under weaker assumptions about the Φ_n.

Proof. The idea is to consider a rectangular grid of points spaced over the identifiable region B. Lemma 8.3 will show that S_N converges almost surely to its expectation on all of the grid points. Lemma 8.5 will show that the grid points are spaced such that S_N cannot change much between consecutive grid points and from this it will follow that S_N converges to its expectation uniformly in B. To specify a grid point we use the notation $\boldsymbol{\lambda}[\mathbf{r}]$, where $\mathbf{r} \in \mathbb{Z}^{m+1}$, to denote the point

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

for some $b > 0$. The variable b defines how closely spaced the grid points are. Note that the points are spaced so that adjacent points are separated by $\frac{1}{N^b}$ in the first coordinate, $\frac{1}{N^{b+1}}$ in the second coordinate and $\frac{1}{N^{b+k}}$ in the kth coordinate. We use $B[\mathbf{r}]$ to denote the small dense space about the grid point $\boldsymbol{\lambda}[\mathbf{r}]$, that is

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

We only want to consider a finite number of grid points in B and these are given by the set of points $\boldsymbol{\lambda}[\mathbf{r}]$ where $\mathbf{r}$ is in the set

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

So the total number of grid points is $|G| = N^{(m+1)(2b+m)/2}$. Note that the union of the $B[\mathbf{r}]$ over all $\mathbf{r} \in G$ contains the identifiable region B, and in fact, the $B[\mathbf{r}]$ partition B.

It will be convenient to define the difference between S_N and its expectation as

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

Now we wish to show that $\sup_{\boldsymbol{\lambda} \in B} |D_N(\boldsymbol{\lambda})|$ converges to zero almost surely as N converges to ∞. We may write

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

So, the proof is complete if we can show, firstly that the difference D_N converges to zero on all of the grid points $\boldsymbol{\lambda}[\mathbf{r}]$, i.e. that $\sup_{\mathbf{r} \in G} |D_N(\boldsymbol{\lambda}[\mathbf{r}])| \to 0$ almost surely as $N \to \infty$, and, secondly that the difference D_N cannot change much within any of the small dense regions $B[\mathbf{r}]$ surrounding a grid point, that is,

$$\sup_{\mathbf{r} \in G} \sup_{\boldsymbol{\lambda} \in B[\mathbf{r}]} |D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])| \to 0$$

almost surely as $N \to \infty$. These will be proved in lemmas 8.3 and 8.5 to follow. □

We first need the following result about sums of independent random variables.

Lemma 8.2. *Let $Z_1, Z_2, \ldots, Z_N$ be independent, zero-mean random variables with all magnitudes $|Z_j|$ bounded by some constant. Then, for any integer $\beta > 0$ we have*

$$S = E\left[(Z_1 + \cdots + Z_N)^{2\beta} \right] = O(N^\beta)$$

as $N \to \infty$.

Proof. A proof can be found in Lemma 9 of [McKilliam et al., 2010a] and another proof is given by Brillinger [1962]. This is actually a rather weak result and stronger versions are known that, for example, do not require independence. See, for example, Yokoyama [1980].

□

Lemma 8.3. *The difference D_N converges to zero almost surely on the all of the grid points. That is,*

$$\sup_{\mathbf{r}\in G} |D_N(\boldsymbol{\lambda}[\mathbf{r}])| \to 0$$

almost surely as $N \to \infty$.

Proof. The approach we take it to first bound the probability that $\sup_{\mathbf{r}} |D_N(\boldsymbol{\lambda}[\mathbf{r}])|$ is larger than some small positive constant using Markov's inequality on a positive even power of the D_N. This proves the convergence in *probability*. We then show that the bounds are strong enough to imply that the convergence is *almost sure* by the Borel-Cantelli lemma.

For some $\boldsymbol{\lambda}$, consider the random variable that results from taking $D_N(\boldsymbol{\lambda})$ to the power of 2β with β a positive integer. By applying Markov's inequality we obtain, for any arbitrarily small constant $\varepsilon > 0$, that

$$\text{Prob}\left(\left|D_N^{2\beta}(\boldsymbol{\lambda})\right| \leq \epsilon^{2\beta}\right) \leq \frac{E\left[\left|D_N^{2\beta}(\boldsymbol{\lambda})\right|\right]}{\varepsilon^{2\beta}}$$

and because β is a positive integer we have $\left|D_N^{2\beta}(\boldsymbol{\lambda})\right| = D_N^{2\beta}(\boldsymbol{\lambda})$ and therefore

$$\text{Prob}\left(|D_N(\boldsymbol{\lambda})| \leq \epsilon\right) \leq \frac{E\left[D_N^{2\beta}(\boldsymbol{\lambda})\right]}{\varepsilon^{2\beta}}.$$

Let

$$E\left[D_N^{2\beta}(\boldsymbol{\lambda})\right] = \frac{1}{N^{2\beta}} E\left[\left(\sum_{n=1}^{N} Z_n\right)^{2\beta}\right]$$

where each of

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

are independent with zero mean and are bounded in $[-1, 1]$ due to the fractional parts. From Lemma 8.2 we see that

$$E\left[D_N^{2\beta}(\boldsymbol{\lambda})\right] = \frac{1}{N^{2\beta}} E\left[\left(\sum_{n=1}^{N} Z_n\right)^{2\beta}\right] = O\left(N^{-\beta}\right). \tag{8.2.3}$$

Now the probability

$$\begin{aligned}\text{Prob}\left(\sup_{\mathbf{r}\in G}|D_N(\boldsymbol{\lambda}[\mathbf{r}])|>\varepsilon\right) &\le \sum_{\mathbf{r}\in G}\text{Prob}\left(|D_N(\boldsymbol{\lambda}[\mathbf{r}])|>\varepsilon\right)\\ &\le \sum_{\mathbf{r}\in G}\frac{E\left[D_N^{2\beta}(\boldsymbol{\lambda}[\mathbf{r}])\right]}{\varepsilon^{2\beta}}\\ &= |G|O\left(N^{-\beta}\right)=O\left(N^{(m+1)(2b+m)/2-\beta}\right).\end{aligned}$$

For any $b>0$ we can choose β so that the exponent

$$(m+1)(2b+m)/2-\beta<0$$

and this proves that $\sup_{\mathbf{r}\in G}|D_N(\boldsymbol{\lambda}[\mathbf{r}])|$ converges in *probability* to zero as $N\to 0$. To extend the convergence to almost surely simply choose β so that the exponent

$$(m+1)(2b+m)/2-\beta<-1.$$

Now the sum

$$\sum_{N=1}^{\infty}\text{Prob}\left(\sup_{\mathbf{r}\in G}|D_N(\boldsymbol{\lambda}[\mathbf{r}])|>\varepsilon\right)=\sum_{N=1}^{\infty}O\left(N^{(m+1)(2b+m)/2-\beta}\right)$$

converges and consequently from the Borel-Cantelli lemma $\sup_{\mathbf{r}\in G}|D_N(\boldsymbol{\lambda}[\mathbf{r}])|$ converges almost surely to zero as $N\to\infty$. □

Before proving Lemma 8.5 we need the following result about fractional parts.

Lemma 8.4. *Let x and δ be real numbers. Then*

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

Proof. We will prove the lemma for $\delta\ge 0$, the proof for $\delta\le 0$ is similar. Note that $0\le\langle x\rangle^2\le\frac{1}{4}$ for all real numbers x. So if $\delta\ge\frac{1}{4}$ then the lemma is obviously true because

$$\langle x\rangle^2-|\delta|\le 0\le\langle x+\delta\rangle^2\le\tfrac{1}{4}\le\langle x\rangle^2+|\delta|.$$

So we may assume that $\delta<\frac{1}{4}$. We consider two cases, firstly when $\langle x\rangle+\delta\in[-1/2,1/2)$, and secondly when $\langle x\rangle+\delta\ge 1/2$. Assume that $\langle x\rangle+\delta\in[-1/2,1/2]$ and therefore

$$-\tfrac{1}{2}\le\langle x\rangle\le\tfrac{1}{2}-\delta$$

and also

$$\langle x+\delta\rangle^2=\langle\langle x\rangle+\delta\rangle^2=(\langle x\rangle+\delta)^2=\langle x\rangle^2+2\langle x\rangle\delta+\delta^2.$$

Substituting the bounds on $\langle x\rangle$ above into the right hand side of this equation we obtain

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

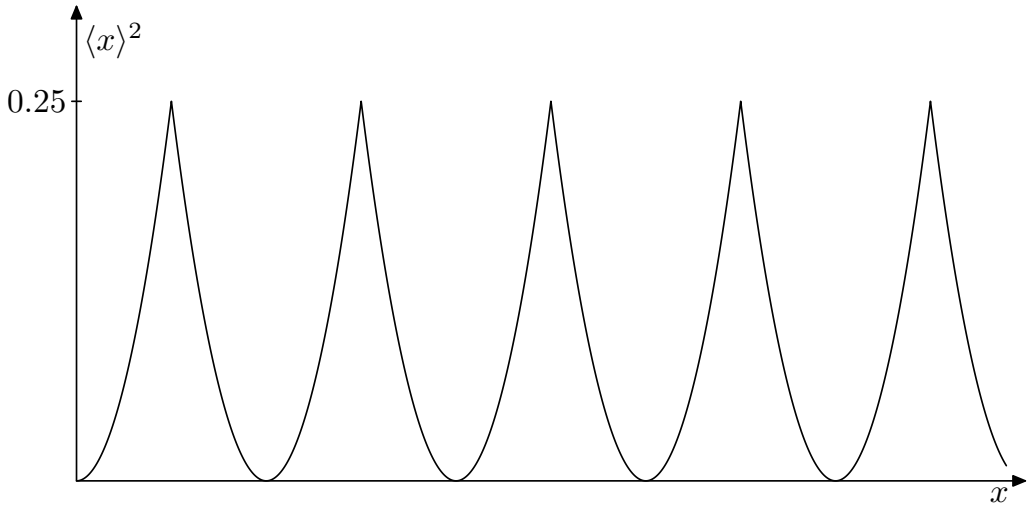


Figure 8.1: The function $\langle x \rangle^2$. Note that the function is continuous and piecewise differentiable. The derivative has magnitude less than one whenever it exists.

The proof now follows because $\delta(1-\delta) \leq |\delta|$ because $0 \leq \delta < \frac{1}{4}$. Now assume that $\langle x \rangle + \delta \geq 1/2$. We have

$$\langle x + \delta \rangle = \langle x \rangle + \delta - 1$$

and therefore

$$\tfrac{1}{2} - \delta \leq \langle x \rangle \leq \tfrac{1}{2}$$

and also

$$\langle x + \delta \rangle^2 = (\langle x \rangle + \delta - 1)^2 = \langle x \rangle^2 + 2 \langle x \rangle (\delta - 1) + (\delta - 1)^2.$$

Substituting the bounds on $\langle x \rangle$ above into the right hand side of this equation we again obtain

$$\langle x \rangle^2 - \delta(1-\delta) \leq \langle x + \delta \rangle^2 \leq \langle x \rangle^2 + \delta(1-\delta)$$

and the proof follows because $\delta(1-\delta) \leq |\delta|$ because $0 \leq \delta < \frac{1}{4}$. □

Consideration of Figure 8.1 makes Lemma 8.4 *obvious* because the function $\langle x \rangle^2$ is clearly continuous and piecewise differentiable and the derivative has magnitude less than one whenever it exists. The function is similar to the 'row of glasses' function described by [Huxley, 1996, page 95]. It is a shame that the proof mostly loses this intuitive picture, however, we have not found a more intuitive proof with the same amount of rigour. In any case, we are now in a position to prove Lemma 8.5.

Lemma 8.5. *The difference between D_N evaluated at the grid point $\lambda[\mathbf{r}]$ and D_N evaluated anywhere in the small dense region $B[\mathbf{r}]$ containing $\lambda[\mathbf{r}]$ converges to zero almost surely as $N \to \infty$. That is,*

$$\sup_{\mathbf{r} \in G} \sup_{\boldsymbol{\lambda} \in B[\mathbf{r}]} |D_N(\boldsymbol{\lambda}) - D_N(\boldsymbol{\lambda}[\mathbf{r}])| \to 0$$

almost surely as $N \to \infty$.

Proof. The proof we give is not just *almost surely* but *surely* as the convergence will be shown to occur irrespective of the values of the Φ_n. For convenience let

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

where $\lambda[\mathbf{r}]_k$ denotes the kth element of the grid point $\boldsymbol{\lambda}[\mathbf{r}]$. For $\boldsymbol{\lambda}$ contained in the small region about the grid point $\boldsymbol{\lambda}[\mathbf{r}]$, i.e. $\boldsymbol{\lambda} \in B[\mathbf{r}]$, we have

$$b_n = a_n + \delta_n$$

where all of the δ_n have magnitude $|\delta_n| \leq \frac{m+1}{N^b}$. Taking factional parts, squaring, and applying Lemma 8.4 we see that, for all real numbers x, the difference between $\langle x + b_n \rangle^2$ and $\langle x - a_n \rangle^2$ is bounded above and below like

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

or equivalently $|\langle x + b_n \rangle^2 - \langle x - a_n \rangle^2| \leq \frac{m+1}{N^b}$. Now note that

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

and then

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

and therefore, for all $\boldsymbol{\lambda} \in B[\mathbf{r}]$, we have

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

Seeing as this bound is independent of the Φ_n we also immediately have the same result for the expectation, so, using Jensen's inequality,

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

Therefore, for all $\boldsymbol{\lambda} \in B[\mathbf{r}]$, we have

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

and as this bound does not depend on the grid point chosen, i.e. the bound is independent of $\mathbf{r}$, we have

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

and the proof follows. □

We have now given all the results required for the proof of Lemma 8.1, that $S_N(\boldsymbol{\lambda})$ converges to its expectation $ES_N(\boldsymbol{\lambda})$ almost surely and uniformly in $\boldsymbol{\lambda} \in B$ as $N \to \infty$. We now require some information about the minimisers of the expectation ES_N. Because the distribution of the Φ_n has zero unwrapped mean and has unwrapped variance σ^2 then the expected value of $\langle \Phi_n + z \rangle^2$ is

$$E \langle \Phi_n + z \rangle^2 = \int_{-1/2}^{1/2} \langle \theta + z \rangle^2 f(\theta) d\theta \tag{8.2.4}$$

and over $z \in [-1/2, 1/2)$ this is minimised uniquely at $z = 0$ (see Section 5.2.2). Also, when $z = 0$ we see that

$$E \langle \Phi_n \rangle^2 = \operatorname{var} \Phi_n = \sigma^2 = \int_{-1/2}^{1/2} \theta^2 f(\theta) d\theta \tag{8.2.5}$$

is the unwrapped variance of the Φ_n.

Lemma 8.6. *For $\boldsymbol{\lambda}$ in the identifiable region B the expectation $ES_N(\boldsymbol{\lambda})$ is minimised uniquely at $\mathbf{0}$ and at this minimum $ES_N(\mathbf{0}) = \sigma^2$.*

Proof. Let z be the polynomial

$$z(n) = \lambda_0 + \lambda_1 n + \lambda_2 n^2 + \cdots + \lambda_m n^m$$

then we can write

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

We know that $E \langle \Phi_n + \langle z(n) \rangle \rangle^2$ 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 $\operatorname{coef}(z)$ is a lattice point in L_{m+1}. From the definition of the identifiable region B (7.2.1) contains precisely one lattice point from L_{m+1}, this being the origin $\mathbf{0}$. Therefore ES_N is minimised uniquely at $\mathbf{0}$ at which point it takes the value σ^2. □

Lemma 8.7. *The value of the expectation $ES_N(\hat{\boldsymbol{\lambda}})$ converges almost surely to $ES_N(\mathbf{0}) = \sigma^2$ as $N \to \infty$. That is*

$$ES_N(\hat{\boldsymbol{\lambda}}) - \sigma^2 \to 0 \tag{8.2.6}$$

almost surely as $N \to \infty$.

Proof. By definition

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda} \in B} S_N(\boldsymbol{\lambda}) \tag{8.2.7}$$

and therefore

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

Also, because ES_N is minimised at $\mathbf{0}$, then

$$0 \leq ES_N(\hat{\boldsymbol{\lambda}}) - ES_N(\mathbf{0})$$

and adding these two inequalities gives

$$\begin{aligned} 0 \leq ES_N(\hat{\boldsymbol{\lambda}}) - ES_N(\mathbf{0}) &\leq ES_N(\hat{\boldsymbol{\lambda}}) - ES_N(\mathbf{0}) + S_N(\mathbf{0}) - S_N(\hat{\boldsymbol{\lambda}}) \\ &\leq |ES_N(\hat{\boldsymbol{\lambda}}) - S_N(\hat{\boldsymbol{\lambda}})| + |S_N(\mathbf{0}) - ES_N(\mathbf{0})| \end{aligned}$$

and from Lemma 8.1, the right hand side of the above equation converges almost surely to 0 as $N \to \infty$. So, $ES_N(\hat{\boldsymbol{\lambda}}) - ES_N(\mathbf{0})$ converges to zero almost surely as $N \to \infty$ and because $ES_N(\mathbf{0}) = \sigma^2$ we have

$$ES_N(\hat{\boldsymbol{\lambda}}) - \sigma^2 \to 0$$

almost surely as $N \to \infty$. □

The proof of strong consistency is now partly complete because $ES_N(\hat{\boldsymbol{\lambda}})$ is *uniquely* minimised at $\mathbf{0}$ and therefore $\hat{\boldsymbol{\lambda}}$ must converge to zero. However, this tells us nothing about the order of convergence of the elements in $\hat{\boldsymbol{\lambda}}$ as required by the theorem. We will show in Lemma 8.10 that all sequences $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_N$ that satisfy (8.2.6) have the property that $N^k \lambda_{k,N} \to 0$ for all $k = 0, 1, \ldots, m$ where $\lambda_{k,N}$ is the kth element in the Nth vector in the sequence. As $\hat{\boldsymbol{\lambda}}$ satisfies (8.2.6) almost surely it will immediately follow that $N^k \hat{\lambda}_k \to 0$ almost surely as $N \to \infty$ and this will complete the proof of strong consistency. To prove Lemma 8.10 we need some preliminary results about arithmetic progressions and from the calculus of finite differences.

Let $W = \{1, 2, \ldots, N\}$ be the set of integers from 1 to N and let K be a subset of W. For any integer h we define the set $A(h, K)$ to contain 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. That is

$$A(h, K) = \big\{ n \mid n + ih \in K \ \forall\ i \in \{0, 1, \ldots, m\} \big\}. \tag{8.2.8}$$

If K is a small subset of W then $A(h, K)$ might contain no elements at all, i.e. there may be no arithmetic progressions $n, n+h, \ldots$ of length $m+1$ in K. 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 will write $K \backslash r$ to denote the set K with the element r removed.

Lemma 8.8. *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 $n + ih = r$ where $i \in \{0, 1, \ldots, m\}$. That is, there are at most $m+1$ arithmetic progressions of type $n, n+h, \ldots n+mh$ that contain r. □

Lemma 8.9. *Let $K \subseteq W$. Then $|A(h,K)| \geq N - mh - (N - |K|)(m+1)$*

Proof. Note that $|A(h,W)| = N - mh$. The proof follows by starting with $A(h,W)$ and applying Lemma 8.8 $|W| - |K| = N - |K|$ 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. □

Corollary 8.1. *Let $K \subseteq W$ such that $|K| > \frac{2m+1}{2m+2}N$. Then for all $1 \leq h \leq \frac{N}{2m}$ 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 $h \leq \frac{N}{2m}$ directly into the inequality from Lemma 8.9 we immediately obtain $|A(h,K)| > 0$. □

The final piece of machinery we require comes from the calculus of finite differences. For any function $d(n)$, let

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

denote the first difference with interval h. Let

$$\Delta_h^2 d(n) = \Delta_h d(n+h) - \Delta_h d(n) = d(n+2h) - 2d(n+h) + d(n)$$

denote the second difference with interval h and similarly let

$$\Delta_h^r d(n) = \Delta_h^{r-1} d(n+h) - \Delta_h^{r-1} d(n) = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} d(n+kh) \tag{8.2.9}$$

denote the rth difference with interval h. It is not hard to show (and is well known) that the sum of the magnitude of the coefficients (i.e the binomial coefficients) inside the summand above is equal to 2^r, that is

$$\sum_{k=0}^{r} \binom{r}{k} = 2^r.$$

In other words, $\Delta_h^r d(n)$, can be represented by adding and subtracting the

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

precisely 2^r times.

The differencing operator Δ_h has some special properties when applied to polynomials. If $d(n) = a_r n^r + \cdots + a_0$ is a polynomial of order r then the rth difference of order h of the polynomial d is given by

$$\Delta_h^r d(n) = h^r r! a_r. \tag{8.2.10}$$

So, the rth difference of the polynomial is a constant depending on h, r and the rth polynomial coefficient a_r. A derivation of this well known property is given by Jordan [1965, page 51]. We can now prove Lemma 8.10 from which the proof of strong consistency will follow.

Lemma 8.10. *Suppose* $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots$ *is a sequence with* $\boldsymbol{\lambda}_n \in B$ *for all* $n = 1, 2, \ldots$ *and with*

$$ES_N(\boldsymbol{\lambda}_N) - \sigma^2 \to 0$$

as $N \to \infty$. *Then* $N^k \lambda_{k,N} \to 0$ *for all* $k = 0, 1, \ldots, m$.

Proof. For the sake of notational simplicity we will abbreviate $\lambda_{k,N}$ to λ_k. Define the function

$$g(z) = E\left\langle \Phi_n + z \right\rangle^2 - \sigma^2 = \int_{-1/2}^{1/2} \left\langle \theta + z \right\rangle^2 f(\theta) d\theta - \sigma^2$$

which is continuous in z. Also, because of (8.2.5) and (8.2.4), we have $g(z) \geq 0$ with equality only at $z = 0$ for $z \in [-1/2, 1/2)$. Now

$$ES_N(\lambda_0, \ldots, \lambda_m) - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} g\left(\left\langle \sum_{k=0}^{m} n^k \lambda_k \right\rangle \right) \to 0$$

as $N \to \infty$. It will be convenient to define the polynomial

$$z(n) = \lambda_0 + \lambda_1 n + \lambda_2 n^2 + \cdots + \lambda_m n^m.$$

Now we may write

$$ES_N(\lambda_0, \ldots, \lambda_m) - \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} g\left(\left\langle z(n) \right\rangle \right) \to 0.$$

We need the following lemma.

Lemma 8.11. *For any constants* $0 \leq c < 1$ *and* $\delta > 0$ *there exists an* N_0 *such that for all* $N > N_0$ *the proportion of* $\langle z(n) \rangle$ *with magnitude less than* δ *is greater than* c. *That is, the set*

$$K_N = \{ n \leq N \mid | \left\langle z(n) \right\rangle | < \delta \}$$

has more than cN *elements for all* $N > N_0$.

Proof. Assume not. Then for every N_0 there exists an $N > N_0$ such that there are more than $(1 - c)N$ integers from 1 to N with $| \left\langle z(n) \right\rangle | > \delta$. Let γ be the minimum value of g over the interval given by the union $[-1/2, -\delta] \cup [\delta, 1/2)$. Because g is minimised uniquely at 0 then γ is strictly greater than 0 and the sum

$$\frac{1}{N} \sum_{n=1}^{N} g\left(\left\langle z(n) \right\rangle \right) \geq (1 - c)\gamma$$

with $(1 - c)\gamma$ a positive constant. This violates the fact that g converges to zero as $N \to \infty$ and the lemma is true by contradiction. □

It is particularly useful to choose the constants

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

So, the set K_N has $|K_N| > \frac{2m+1}{2m+2}N$ elements and from Corollary 8.1 it follows that for all $1 \leq h \leq \frac{N}{2m}$ the set $A(h, K_N)$ contains at least one element, that is, there exist an $n' \in A(h, K_N)$ such that all the elements from the arithmetic progression $n', n'+h, \ldots, n'+mh$ are in K_N and therefore

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

are all less than δ.

Lemma 8.12. *Let $a_1, a_2, \ldots, a_r$ be r real numbers such $|\langle a_n\rangle| < \delta$ 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 $|\langle \sum_{n=1}^{r} a_n\rangle| \leq \frac{1}{2}$ for all $a_n \in \mathbb{R}$. If $\delta \leq \frac{1}{2r}$ then $\langle \sum_{n=1}^{r} a_n\rangle = \sum_{n=1}^{r} \langle a_n\rangle$ and from the triangle inequality

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

□

Consider the mth difference of the $\langle z(n)\rangle$ evaluated at n'. Because the mth difference is a linear combination of 2^m elements (see (8.2.9)) from

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

all with magnitude less than δ we obtain, from Lemma 8.12, that

$$|\langle \Delta_h^m z(n')\rangle| \leq |\Delta_h^m \langle z(n')\rangle| \leq \Delta_h^m |\langle z(n')\rangle| < 2^m\delta. \tag{8.2.11}$$

Using the properties of the mth difference of a polynomial from (8.2.10) it follows that the left hand side is equal to a constant involving h, m and λ_m giving the bound

$$|\langle h^m m!\lambda_m\rangle| = |\langle \Delta_h^m z(n')\rangle| < 2^m\delta \tag{8.2.12}$$

for all $1 \leq h \leq \frac{N}{2m}$. Setting $h = 1$ and recalling from the definition of the identifiable region (8.2.2) that $\lambda_m \in [-\frac{0.5}{m!}, \frac{0.5}{m!})$, we have

$$|\langle m!\lambda_m\rangle| = |m!\lambda_m| < 2^m\delta$$

because λ_m is small enough for the fractional part to disappear. Now, because we chose $\delta < \frac{1}{2^{2m}}$ it follows that

$$|\lambda_m| < \tfrac{2^m}{m!}\delta < \tfrac{1}{m!2^{m+1}}.$$

So, when $h = 2$,

$$|\langle 2^m m!\lambda_m\rangle| = |2^m m!\lambda_m| < 2^m\delta$$

because $2^m m!\lambda_m \in [-0.5, 0.5)$ so the fractional parts disappear again. Therefore

$$|\lambda_m| < \frac{1}{m!}\delta < \frac{1}{m!2^{2m+1}}.$$

Now, with $h = 4$, we similarly obtain $|\,\langle 4^m m!\lambda_m\rangle\,| = |4^m m!\lambda_m| < 2^m\delta$ and iterating this process we eventually obtain

$$|\lambda_m| < \frac{2^m}{2^{rm}m!}\delta < \frac{1}{m!2^{(r+1)m}+1}$$

where 2^r is the largest power of 2 less than or equal to $\frac{N}{2m}$. Then $2^{r+1} > \frac{N}{2m}$ and substituting this into the inequality above gives

$$|\lambda_m| < \frac{1}{2m!\left(\frac{N}{2m}\right)^m}\delta$$

from which it follows that

$$N^m|\lambda_m| < \frac{2^{m-1}m^m}{m!}\delta. \tag{8.2.13}$$

As δ can be made arbitrarily small and m is constant, it follows that $N^m\lambda_m \to 0$ as $N \to \infty$.

We have now shown that the highest order polynomial coefficient λ_m converges as required by Theorem 8.1. The remaining coefficients will be shown to converge by induction. Assume that $N^k\lambda_k \to 0$ for all $k = r+1, r+2, \ldots, m$, that is, assume that the $m-r$ highest order coefficients all converge. Define the polynomial

$$z_r(n) = \lambda_0 + \lambda_1 n + \lambda_2 n^2 + \cdots + \lambda_r n^r.$$

Because the $m-r$ highest order coefficients converge we can write $z(n) = z_r(n)+\gamma(n)$ where $\gamma(n)$ can be made simultaneously arbitrarily small for all n. Now the bound from (8.2.11), but applied using the rth difference, gives

$$|\langle\Delta_h^r z(n')\rangle| = |\langle\Delta_h^r\gamma(n') + \Delta_h^r z_r(n')\rangle| = |\,\langle\epsilon + h^r r!\lambda_r\rangle\,| < 2^r\delta \tag{8.2.14}$$

where $\epsilon = \Delta_h^r\gamma(n')$ and therefore $|\epsilon| \geq 0$ can be chosen arbitrarily small. We need the following lemma.

Lemma 8.13. *Let $|\langle a+\epsilon\rangle| < \delta$ where $|\epsilon| < 1/4$ and $0 < \delta < 1/4$. Then $|\langle a\rangle| < \delta + |\epsilon|$.*

Proof. The idea behind this proof is to show that under the conditions given $\langle\langle a\rangle + \epsilon\rangle$ *does not wrap*, i.e. that $\langle a\rangle + \epsilon \in [-1/2, 1/2)$. The proof will then follow easily. To start, assume that $\langle a\rangle + \epsilon \geq 1/2$. Then $1/2 \leq \langle a\rangle + \epsilon < 3/4$ and

$$-1/2 \leq \langle\langle a\rangle + \epsilon\rangle = \langle a+\epsilon\rangle < -1/4$$

and therefore $|\,\langle a+\epsilon\rangle\,| > 1/4 > \delta$, a contradiction. Similarly, assume that $\langle a\rangle + \epsilon < -1/2$. Then $-1/2 > \langle a\rangle + \epsilon > -3/4$ and $1/2 > \langle a+\epsilon\rangle > 1/4$ and therefore $|\langle a+\epsilon\rangle| > 1/4 > \delta$, a contradiction. So, $\langle a\rangle + \epsilon \in [-1/2, 1/2)$ and

$$|\,\langle a+\epsilon\rangle\,| = |\langle a\rangle + \epsilon| < \delta$$

from which it follows that $|\langle a\rangle| < \delta + |\epsilon|$. □

We can choose δ and ϵ such that $2^r\delta < \frac{1}{4}$ and $|\epsilon| < \frac{1}{4}$. Then from the left hand side of (8.2.14) and the above lemma we have

$$|\,\langle h^r r!\lambda_r\rangle\,| < 2^r\delta + |\epsilon|$$

and by choosing $2^r\delta + |\epsilon| < \frac{1}{2^{2r+1}}$ and using the same iterative process we did for the highest order coefficient λ_m (see (8.2.12) to (8.2.13)) we find that $N^r\lambda_r \to 0$ as $N \to \infty$. The proof now follows by induction. □

8.2.2 The central limit theorem

The derivation of the central limit theorem is fortunately more straightforward than the proof of strong consistency. Most of this proof is more easily derived using vector notation. From (8.1.5) we obtain

$$\hat{\boldsymbol{\mu}} = \mathbf{X}^+(\boldsymbol{\theta} - \hat{\mathbf{w}})$$

where the elements of $\hat{\mathbf{w}} = \lceil \boldsymbol{\theta} - \mathbf{X}\hat{\boldsymbol{\mu}} \rfloor$ are the *estimated* wrapping variables $\hat{W}_n$ and $\mathbf{X}^+ = (\mathbf{X}^\dagger\mathbf{X})^{-1}\mathbf{X}^\dagger$ is the pseudoinverse of the N by $m+1$ Vandermonde matrix $\mathbf{X}$ from (4.2.2). Writing the observed random variables $\Theta_1, \Theta_2, \ldots, \Theta_N$ from (8.1.1) in vector form we obtain

$$\boldsymbol{\theta} = \langle \boldsymbol{\phi} + \mathbf{X}\tilde{\boldsymbol{\mu}}\rangle$$

where the fractional part function $\langle\cdot\rangle$ (and also the round function $\lceil\cdot\rfloor$) works elementwise on vectors. Substituting this into the equation for $\hat{\boldsymbol{\mu}}$ above we obtain

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \mathbf{X}^+\big(\,\langle \boldsymbol{\phi} + \mathbf{X}\tilde{\boldsymbol{\mu}}\rangle - \mathbf{w}\big) \\
&= \mathbf{X}^+\big(\,\langle \boldsymbol{\phi} + \mathbf{X}\tilde{\boldsymbol{\mu}}\rangle - \lceil \boldsymbol{\theta} - \mathbf{X}\hat{\boldsymbol{\mu}}\rfloor\,\big) \\
&= \mathbf{X}^+\big(\,\langle \boldsymbol{\phi} + \mathbf{X}\tilde{\boldsymbol{\mu}}\rangle - \lceil \langle\boldsymbol{\phi} + \mathbf{X}\tilde{\boldsymbol{\mu}}\rangle - \mathbf{X}\hat{\boldsymbol{\mu}}\rfloor\,\big) \\
&= \mathbf{X}^+\big(\boldsymbol{\phi} + \mathbf{X}\tilde{\boldsymbol{\mu}} - \lceil \boldsymbol{\phi} + \mathbf{X}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})\rfloor\,\big)
\end{aligned}$$

where, in the last line, we have used the fact that $\lceil\langle x\rangle + y\rfloor = \lceil x + y\rfloor - \lceil x\rfloor$ for any real numbers x and y. Now because $\mathbf{X}^+\mathbf{X}$ is the $m+1$ by $m+1$ identity matrix we have

$$\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}} = \mathbf{X}^+\big(\boldsymbol{\phi} - \lceil \boldsymbol{\phi} + \mathbf{X}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})\rfloor\,\big). \tag{8.2.15}$$

Recall that $\hat{\boldsymbol{\lambda}} = \text{dealias}(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}) = \tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}} - \mathbf{p}$ where $\mathbf{p}$ is a lattice point from L_{m+1} (see Chapter 7). From the definition L_{m+1} in terms of integer valued polynomials we have that all the elements in the vector $\mathbf{Xp}$ are integers so

$$\left\lceil \boldsymbol{\phi} + \mathbf{X}\hat{\boldsymbol{\lambda}}\right\rfloor = \lceil \boldsymbol{\phi} + \mathbf{X}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}} - \mathbf{p})\rfloor = \lceil \boldsymbol{\phi} + \mathbf{X}(\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})\rfloor - \mathbf{Xp}.$$

Negating and subtracting $\mathbf{p}$ from both sides of (8.2.15) gives

$$\hat{\boldsymbol{\lambda}} = \mathbf{X}^+\left\lceil \boldsymbol{\phi} + \mathbf{X}\hat{\boldsymbol{\lambda}}\right\rfloor - \mathbf{X}^+\boldsymbol{\phi}. \tag{8.2.16}$$

So, to derive the central limit theorem we need to find the distribution of

$$\mathbf{D}\hat{\boldsymbol{\lambda}} = \mathbf{D}\mathbf{X}^+\left\lceil \boldsymbol{\phi} + \mathbf{X}\hat{\boldsymbol{\lambda}}\right\rfloor - \mathbf{D}\mathbf{X}^+\boldsymbol{\phi} \tag{8.2.17}$$

where $\mathbf{D}$ is the $m+1$ by $m+1$ diagonal matrix with diagonal elements $d_{i,i} = N^{i/2}$. For notational convenience we let $\mathbf{z}$ be the vector

$$\mathbf{z} = \mathbf{D}\mathbf{X}^{+}\left[\boldsymbol{\phi} + \mathbf{X}\hat{\boldsymbol{\lambda}}\right]. \tag{8.2.18}$$

It is proved in Lemma 8.15 below that the expected value of $\mathbf{z}$ satisfies

$$E[\mathbf{z}] = \mathbf{D}\hat{\boldsymbol{\lambda}}\big(f(-1/2) + o(1)\big)$$

where $o(1)$ converges almost surely to zero as $N \to \infty$, and it is proved in Lemma 8.16 below that the vector $\mathbf{z} - E[\mathbf{z}]$ converges in probability to zero, i.e. the elements of $\mathbf{z} - E[\mathbf{z}]$ are all in $o_P(1)$.

Subtracting $E[\mathbf{z}]$ from both sides of (8.2.17) and using Lemmas 8.15 and 8.16 gives

$$\mathbf{D}\hat{\boldsymbol{\lambda}}(1 - f(-1/2) + o(1)) = o_P(1) - \mathbf{D}\mathbf{X}^{+}\boldsymbol{\phi}$$

where, by a mild abuse of notation, $o_P(1)$ here denotes a *column vector* of length $m+1$ with all elements converging in probability to zero as $N \to \infty$. Because $f(-1/2) < 1$ from the statement of Theorem 8.1 we may rearrange this to

$$\mathbf{D}\hat{\boldsymbol{\lambda}} = o_P(1) - \frac{\mathbf{D}\mathbf{X}^{+}\boldsymbol{\phi}}{1 - f(-1/2) + o(1)}.$$

Because the Φ_n are identical and independent with zero mean and variance σ^2 it follows from the standard central limit theorem [Billingsley, 1979, page 308] that the distribution of $\mathbf{D}\mathbf{X}^{+}\boldsymbol{\phi}$ converges to the normal with zero mean and covariance $\sigma^2(\mathbf{X}^{+})^{\dagger}\mathbf{D}^2\mathbf{X}^{+}$. It is not hard to show that the matrix $(\mathbf{X}^{+})^{\dagger}\mathbf{D}^2\mathbf{X}^{+}$ converges to the inverse of the $m+1$ by $m+1$ Hilbert matrix $\mathbf{C}^{-1}$. It follows immediately that $\mathbf{D}\hat{\boldsymbol{\lambda}}$ converges to the normal with zero mean and covariance $\sigma^2(1 - f(-1/2))^{-2}\mathbf{C}^{-1}$ as required. It remains to prove Lemmas 8.15 and 8.16 that we have used. We first need the following lemma.

Lemma 8.14. *Let ϵ be a positive constant less than one and let*

$$f_\epsilon = \sup_{-\epsilon \le x \le \epsilon} |f(-1/2) - f(\langle -1/2 + x \rangle)|.$$

Then, for all x such that $|x| \le \epsilon$ the expected value of $\lceil \Phi_1 + x \rfloor$ satisfies the bound

$$x(f(-1/2) - f_\epsilon) \le E\lceil \Phi_1 + x \rfloor \le x(f(-1/2) + f_\epsilon)$$

Proof. First note that, because $\epsilon < 1$ then,

$$\lceil \Phi_1 + x \rfloor = \begin{cases} 1 & \Phi_1 + x \ge 1/2 \\ -1 & \Phi_1 + x < -1/2 \\ 0 & \text{otherwise.} \end{cases}$$

Consider when x is positive so that $0 \le x \le \epsilon$ and the expectation

$$E\lceil \Phi_1 + x \rfloor = \text{Prob}(\Phi_1 + x \ge 1/2) = \int_{1/2 - x}^{1/2} f(\theta)d\theta.$$

From the definition of f_ϵ the integral above can be lower and upper bounded giving

$$x(f(-1/2) - f_\epsilon) \;\leq\; E\lceil \Phi_1 + x \rfloor = \int_{1/2-x}^{1/2} f(\theta)d\theta \;\leq\; x(f(-1/2) + f_\epsilon)$$

for all $0 \leq x \leq \epsilon$. Similarly for all $-\epsilon \leq x < 0$ we have

$$E\lceil \Phi_1 + x \rfloor = -\operatorname{Prob}(\Phi_1 + x < 1/2) = \int_{-1/2}^{-1/2-x} f(\theta)d\theta$$

and again the integral can be lower and upper bounded giving

$$x(f(-1/2) - f_\epsilon) \;\leq\; E\lceil \Phi_1 + x \rfloor \;\leq\; x(f(-1/2) + f_\epsilon).$$

□

Lemma 8.15. *The expected value of the vector* $\mathbf{z}$ *from* (8.2.18) *satisfies*

$$E[\mathbf{z}] = \mathbf{D}\hat{\boldsymbol{\lambda}}\big(f(-1/2) + o(1)\big)$$

Proof. We have

$$E[\mathbf{z}] = \mathbf{D}\mathbf{X}^+ E\left[\boldsymbol{\phi} + \mathbf{X}\hat{\boldsymbol{\lambda}}\right] = \mathbf{D}\mathbf{X}^+ E\lceil \boldsymbol{\phi} + \mathbf{a} \rfloor$$

where $\mathbf{a} = \mathbf{X}\hat{\boldsymbol{\lambda}}$. In view of strong consistency all the elements in $\mathbf{a}$ converge almost surely to zero as $N \to \infty$. Let

$$E\lceil \Phi_n + a_n \rfloor = E\lceil \Phi_1 + a_n \rfloor$$

because the Φ_n are identical. Now consider when $0 \leq a_n < 1$. For any positive constant ϵ less than one we have $|a_n| < \epsilon$ almost surely as $N \to \infty$. So from Lemma 8.14,

$$a_n(f(-1/2) - f_\epsilon) \;\leq\; E\lceil \Phi_n + a_n \rfloor \;\leq\; a_n(f(-1/2) + f_\epsilon)$$

almost surely for all $n = 1, \ldots, N$. From the statement of Theorem 8.1 the periodic function $f(\langle x \rangle)$ is continuous at $x = -1/2$ so, f_ϵ can be made arbitrarily close to zero by choosing ϵ small. Using this the inequality above can be written in vector form as

$$E\lceil \Phi_n + \mathbf{a} \rfloor = \mathbf{a}(f(-1/2) + o(1))$$

where $o(1)$ denotes a number going to zero as $N \to \infty$. Substituting $\mathbf{X}\boldsymbol{\lambda} = \mathbf{a}$ gives

$$E\lceil \Phi_n + \mathbf{X}\boldsymbol{\lambda} \rfloor = \mathbf{X}\boldsymbol{\lambda}(f(-1/2) + o(1))$$

and multiplying both sides by $\mathbf{D}\mathbf{X}^+$ and using that $\mathbf{X}^+\mathbf{X}$ is the $m+1$ by $m+1$ identity matrix completes the proof. □

Lemma 8.16. *The elements in the vector of random variables* $\mathbf{z} - E[\mathbf{z}]$ *converge in probability to zero as* $N \to \infty$.

Proof. The variance of the nth element of the vector $\mathbf{z} - E[\mathbf{z}]$ is

$$\operatorname{var}[Z_n - EZ_n] = E[Z_n^2] - E[Z_N]^2 \leq E[Z_n^2].$$

By the same argument used in Lemma 8.15 it can be shown that

$$E[Z_n^2] \leq a_n(f(-1/2) + o(1)).$$

The proof follows because all the a_n converge almost surely to zero. □

Discussion about the assumptions on f and Φ_n

For the proof of the central limit theorem we have made the assumption that the periodic function $f(\langle x \rangle)$ is continuous at $x = -1/2$ and that $f(-1/2) \neq 1$. Intuitively the assumption that $f(\langle x \rangle)$ is continuous ties the *ends* of f together, that is

$$f(-1/2) = \lim_{\theta \to 1/2} f(\theta)$$

where the limit approaches from below. We used this fact in Lemmas 8.15 and 8.16 to compute the expected value of the vector $\mathbf{z}$ (8.2.18). It might be possible to relax this continuity assumption but we will not consider this here.

A result we have omitted is that a circular random variable, say Φ, with zero unwrapped mean and pdf f must have the property that $f(-1/2) \leq 1$. A proof of this follows by contradiction and using the integral (5.2.5) from the proof of Theorem 5.1. So, the requirement that $f(-1/2) \neq 1$ could equivalently be stated as $f(-1/2) \leq 1$. Dealing with the case that $f(-1/2) = 1$ probably requires taking a higher order approximation of f at $-1/2$ in order to obtain an expression for $E[\mathbf{z}]$ that avoids the asymptote in (8.2.1) that occurs when $f(-1/2) = 1$. We will not consider this.

To a large extent circular distributions for which $f(\langle x \rangle)$ is discontinuous at $-1/2$ and $f(-1/2) = 1$ are pathological and unlikely to occur in practice. For all of the circular distributions considered in this thesis these assumptions hold.

8.3 Summary

In this chapter we have considered estimating the coefficients of a **polynomial phase signal**. In Section 8.1 we derived the angular least squares estimator for the polynomial coefficients and showed how the estimator could be computed by finding a nearest lattice point in the lattice $V^*_{n/m}$. We derived the asymptotic properties of this estimator showing that it is strongly consistent and obtaining its central limit theorem under some assumptions about the circular noise terms Φ_n. For the case of polynomials of order greater than one, the statistical results derived in this chapter are the first of their kind.

In the remaining chapters we will employ the angular least squares estimator in frequency estimation and polynomial phase estimation. We will find that the angular least squares estimator is in practice statistically much more accurate than many of the estimators that exist in the literature.

—We encounter periodic phenomena every day of our lives. Those of us who still use analogue clocks are acutely aware of the 60 second, 60 minute and 12 hour periods associated with the sweeps of the second, minute and hour hands. We are conscious of the fact that the Earth rotates on its axis roughly every 365 days. These periodicities are reasonably accurate. The quantities we are interested in measuring are not precisely periodic and there will also be error associated with their measurement.

from *The Estimation and Tracking of Frequency*, B. G. Quinn and E. J. Hannan.

9 Frequency estimation

In this chapter we consider estimating the two coefficients of a polynomial phase signal of order one. This is equivalent to a well studied problem called **frequency estimation** and has application to, for example, radar, sonar, telecommunications, astronomy and medicine [Quinn and Hannan, 2001].

In Sections 9.1, 9.2 and 9.3 we describe some of the estimators that exist in the literature. These are the **periodogram estimator**, the **Quinn-Fernandes estimator** and **Kay's unwrapping estimator**. In Section 9.5 we discuss the **angular least squares estimator** that can be computed by finding a nearest point in the lattice $V_{n/1}^*$. We could use the nearest point algorithm described in Section 4.3, but we find that it is quite slow. Instead we describe a simple approximate nearest point algorithm that is much faster, and for frequency estimation, has almost identical statistical performance to the exact nearest point algorithm. Section 9.5 compares the estimators by Monte-Carlo simulation. It is found that that the angular least squares estimator is very statistically accurate, but is computationally more expensive than the other estimators.

Signal model

Typically the signal model for frequency estimation is given in complex exponential form, that is, we observe N complex numbers of the form

$$Y_n = \tilde{\rho} e^{2\pi j(\tilde{f} n + \tilde{\mu})} + X_n \tag{9.0.1}$$

where $\tilde{\rho} > 0$ is an unknown amplitude and $X_1, X_2, \ldots, X_N$ are zero mean complex random variables. The parameter $\tilde{f}$ is typically called the **frequency** and $\tilde{\mu}$ is called the **phase**. Note that to ensure identifiability we must restrict $(\tilde{\mu}, \tilde{f})$ to the square identifiable region $[-1/2, 1/2)^2$. This corresponds to the fact that the lattice $L_2 = \mathbb{Z}^2$ when $m = 1$ (see Section 7.2). It also corresponds with the Nyquist criterion.

Our focus will be on estimating the frequency, $\tilde{f}$, as this is typically the parameter of interest in the literature and also in practice. Once an estimate of the frequency, say $\hat{f}$, is found the phase can be estimated from $Y_n e^{-2\pi j n \hat{f}}$ using any of the approaches to **phase estimation** described in Section 6.4. We now briefly describe a number of popular estimators from the literature.

9.1 Least squares and the periodogram estimator

An obvious approach is the least squares estimator. This is given by the minimisers of the sum of squares function

$$S(\rho, \mu, f) = \sum_{n=1}^{N} \left| Y_n - \rho e^{2\pi i(\mu + fn)} \right|^2.$$

The minimisation is over the frequency, f, the phase, μ, and also the amplitude, ρ. It may at first seem a search over all of three parameters is needed, but the problem can be simplified by conditioning the sum of squares function with respect to the amplitude and the phase. To see this let $r = \rho e^{2\pi i \mu}$ and write the sum of squares function as

$$S(r, f) = \sum_{n=1}^{N} \left| Y_n - r e^{2\pi i f n} \right|^2.$$

Fixing f and differentiating with respect to the conjugate of r gives

$$\frac{d}{dr^*} S(r, f) = \sum_{n=1}^{N} Y_n e^{-2\pi i f n} - r,$$

and setting this to zero gives

$$\hat{r} = N^{-1} \sum_{n=1}^{N} Y_n e^{-2\pi i f n},$$

and it is not hard to check that this stationary point corresponds to a minimum. Substituting this into $S(r, f)$ gives the sum of squares function conditioned upon minimisation with respect to r as[1]

$$S(f) = \sum_{n=1}^{N} |Y_n|^2 - N^{-1} \left| \sum_{n=1}^{N} Y_n e^{-2\pi i f n} \right|^2.$$

Now the frequency can be estimated by finding the minimiser of $S(f)$ and this only involves a minimisation over the single parameter, f. The sum $\sum_{n=1}^{N} |Y_n|^2$ is

[1] We have slightly abused notation by reusing S, as in $S(\rho, \mu, f)$, $S(r, f)$ and $S(f)$, but this should not cause any confusion as the different objective functions can easily be told apart by their different inputs.

constant with respect to f and therefore, for the purpose of minimisation, can be ignored. By negating the remaining term we obtain

$$P_s(f) = N^{-1} \left| \sum_{n=1}^{N} Y_n e^{-2\pi i f n} \right|^2$$

which is referred to as the **periodogram** in the frequency estimation literature. The **periodogram estimator** results from selecting the frequency that maximises $P_s(f)$. The periodogram is the squared magnitude of the **Fourier transform** of the Y_n and so maximisation of $P_s(f)$ amounts to selecting the frequency component of the signal Y_n that has largest energy. In this context the periodogram estimator is particularly intuitively appealing. The statistical asymptotic properties of the periodogram estimator have been known for some time [Quinn and Hannan, 2001; Hannan, 1973; Walker, 1971].

There still remains the problem of computing the maximiser of the periodogram in practice. Rife and Boorstyn [1974] have suggested a practical method, by using the **fast Fourier transform** to obtain the value of the periodogram at the **Fourier frequencies**

$$f = -\frac{1}{2}, -\frac{1}{2} + \frac{1}{N}, -\frac{1}{2} + \frac{2}{N}, \ldots, \frac{1}{2} - \frac{1}{N}. \tag{9.1.1}$$

The Fourier frequency that maximises the periodogram is found and this estimate is then further refined by a numerical procedure such as Newton's method. A problem is that the numerical procedure can fail to locate the correct maximiser [Rice and Rosenblatt, 1988]. To avoid the problem Rife and Boorstyn [1974] suggested zero padding the signal to length $4N$ before performing the fast Fourier transform to obtain samples of the periodogram on the finer intervals

$$f = -\frac{1}{2}, -\frac{1}{2} + \frac{1}{4N}, -\frac{1}{2} + \frac{2}{4N}, \ldots, \frac{1}{2} - \frac{1}{4N}.$$

This zero padding has recently been shown to work by Quinn et al. [2008], who also show that applying Newton's method to the derivative of certain monotonic functions of the periodogram, rather than to the periodogram itself, ensures that Newton's method will succeed even without any zero padding.

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

9.2 The Quinn-Fernandes estimator

In order to avoid the numerical difficulties associated with maximising the periodogram Quinn and Fernandes [1991] suggested a different frequency estimator. The estimator begins with an initial guess of the frequency. A specialised iterative procedure then converges from this guess to the estimated frequency. In our

implementation the guess is given by choosing the Fourier frequency (9.1.1) that maximises the periodogram. These are calculated using the fast Fourier transform and therefore our implementation requires $O(N \log N)$ operations. The advantage of the Quinn-Fernandes estimator is that the specialised iterative procedure is substantially more robust than Newton's method. The asymptotic properties of the estimator are also well studied [Quinn and Fernandes, 1991; Quinn, 1999, 2008].

9.3 Kay's unwrapping estimator

It appears to have been Tretter [1985] who first suggested estimating the frequency using only the complex argument of the observations $\angle Y_n$. Tretter [1985] noticed that the $\angle Y_n$ looked like a straight line that is *wrapped* modulo 2π and suggested that if the line could be *unwrapped* then the frequency could be estimated using standard linear regression. This is precisely the mentality behind the angular least squares estimator. Tretter [1985] did not suggest how the unwrapping should be performed. For the angular least squares estimator we have taken a least squares approach, but many authors have considered different unwrapping methods, for example Kay [1989] and Lovell et al. [1991]. We shall briefly describe one of the more popular approaches that is due to Kay [1989].

Kay [1989] considered the first differences of the phase signal $\angle Y_n - \angle Y_{n+1}$ and noted that the resulting signal resembles a moving average process, whose parameters can be estimated by standard linear techniques. A significant advantage of this approach is that the moving average process has enough structure for the estimates to be computed with only $O(N)$ arithmetic operations. The estimator also appears to perform well when the signal-to-noise ratio is large. A major drawback of Kay's estimator is that it is not as statistically accurate as the other estimators. Another problem is that Kay's estimator only works well when the true frequency $\tilde{f}$ is bounded away from ± 0.5. This phenomenon has been studied by Quinn [2000]. In Chapter 10 we shall see that many popular estimators for polynomial phase signals suffer from a similar, but more severe, problem.

9.4 Approximating angular least squares

In order to use the angular least squares estimator we take the complex argument of the Y_n and divide by 2π to obtain the circular random variables

$$\Theta_n = \tfrac{\angle Y_n}{2\pi} = \left\langle \Phi_n + \tilde{f} n + \tilde{\mu} \right\rangle \tag{9.4.1}$$

where the $\Phi_n = \frac{1}{2\pi}\angle(\rho + X_n)$ are projected circular random variables (see Section 5.6). The Θ_n are now in the form of (8.1.1) and we can use the angular least square estimator in the way described in Section 8.1.

Computing the angular least squares estimator requires computing a nearest lattice point in the lattice $V^*_{(N-2)/1}$. Using the algorithm developed in Section 4.3 this requires $O(N^5)$ operations. Although polynomial time this is far too slow for practical use, particularly when the other estimators require only $O(N)$ or $O(N \log N)$

operations. For this reason we will describe an approximate angular least squares estimator that requires $O(N^2 \log N)$ operations and performs in practice very similarly to the exact angular least squares estimator.

The geometric intuition behind the approximate algorithm is that of computing the nearest point in the lattice A^*_{N-1} to a *line* of points. When $m = 1$ we can rewrite the sum of squares function from (8.1.4) in vector form as

$$SS(\mu, f, \mathbf{w}) = \|\boldsymbol{\theta} - f\mathbf{n} - \mu\mathbf{1} - \mathbf{w}\|^2. \tag{9.4.2}$$

where $\boldsymbol{\theta} = [\Theta_1, \ldots, \Theta_N]^\dagger$ and $\mathbf{1}$ is the all ones vector, $\mathbf{n}$ is the vector $[1, 2, \ldots, N]$ and $\mathbf{w}$ is a vector of integer wrapping variables that are considered as nuisance parameters. Fixing both f and $\mathbf{w}$ and minimising with respect to μ we obtain

$$\mu = \frac{\mathbf{1} \cdot (\mathbf{y} - f\mathbf{n} - \mathbf{w})}{N}. \tag{9.4.3}$$

Substituting this into (9.4.2) we find that the angular least square estimator of the frequency, conditioned on minimisation with respect to μ, is given by

$$\hat{f} = \arg \min_{f \in [-1/2, 1/2)} \min_{\mathbf{w} \in \mathbb{Z}^N} \|\mathbf{Q}\mathbf{y} - f\mathbf{Q}\mathbf{n} - \mathbf{Q}\mathbf{w}\|^2, \tag{9.4.4}$$

where $\mathbf{Q}$ is the projection matrix into the space orthogonal to the all ones vector $\mathbf{1}$ (3.3.1). By definition the $\mathbf{Q}\mathbf{w}$ are lattice points in A^*_{N-1} (see Section 3.3) and we can rewrite the minimisation problem above as

$$\hat{f} = \arg \min_{f \in [-1/2, 1/2)} \min_{\mathbf{x} \in A^*_{N-1}} \|\mathbf{z} - f\mathbf{g} - \mathbf{x}\|^2$$

where $\mathbf{z} = \mathbf{Q}\mathbf{y}$ and $\mathbf{g} = \mathbf{Q}\mathbf{n}$. Geometrically this minimisation problem is that of finding the nearest lattice point in A^*_{N-1} to the line segment given by $\mathbf{z} + f\mathbf{g}$ for $f \in [-1/2, 1/2)$. Seeing as very fast algorithms exist to compute a nearest lattice point in A^*_{N-1} we suggest the following obvious approach to computing $\hat{f}$. Take a set V of uniformly spaced samples from $[-1/2, 1/2)$ so that the elements in V are

$$-\frac{1}{2}, -\frac{1}{2} + \frac{1}{|V|}, -\frac{1}{2} + \frac{2}{|V|}, \ldots, \frac{1}{2} - \frac{1}{|V|}$$

and consider the minimisation problem

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in A^*_{N-1}} \min_{f \in V} \|\mathbf{z} - f\mathbf{g} - \mathbf{w}\|^2.$$

That is $\hat{\mathbf{w}}$ is the nearest point in A^*_{N-1} to the sampled line segment $\mathbf{z} + f\mathbf{g}$ for $f \in V$. An estimate of frequency is then given by

$$\hat{f} = \frac{(\mathbf{z} - \hat{\mathbf{w}}) \cdot \mathbf{g}}{\mathbf{g} \cdot \mathbf{g}}.$$

The nearest point in A^*_{N-1} can be computed in $O(N)$ operations using Algorithm 3.4 so computing $\hat{\mathbf{w}}$ and $\hat{f}$ requires $O(|V|N)$ operations. A problem is that we

do not know how many samples of the line segment are needed to guarantee that the true minimiser is found, that is, we do not know how large V should be. Therefore this approach can only be considered an approximation to the exact angular least squares estimator. Experimentation seems to suggest that choosing $|V| = 2N \log_2 N$ gives very good results and in this case the algorithm requires $O(N^2 \log N)$ arithmetic operations. Using similar ideas the *exact* angular least squares estimator can be computed in $O(N^3 \log N)$ arithmetic operations. We will not describe this algorithm here, but details can be found in McKilliam et al. [2010a].

9.5 Simulations

In this section we use Monte Carlo simulation to compare the performance of the angular least squares estimator, the periodogram estimator, Kay's estimator and the Quinn-Fernandes estimator for three different noise distributions, complex Gaussian, von Mises and wrapped uniform. We consider five different values of $N = 4, 16, 64, 256, 1024$ over a range of noise variance. One thousand trials were run for each noise variance and the true frequency $\tilde{f}$ and the true phase $\tilde{\mu}$ are varied uniformly over $[-1/2, 1/2)$ and the true amplitude $\tilde{\rho} = 1$. The simulated mean square error (MSE) of the frequency is plotted in Figures 9.1, 9.2 and 9.3. The MSE is computed using the dealised error as described in Section 7.2.2. The frequency is the more interesting parameter so we do not provide plots for the MSE of the phase. The plots for either frequency or phase lead to similar conclusions about the performance of the estimators.

In Figure 9.1 the complex noise term X_n from (9.0.1) is zero mean complex Gaussian with independent real and imaginary parts each with variance σ_c^2. In this case the Cramér-Rao lower bound (CRB) has been derived by Rife and Boorstyn [1974] and is given by

$$\operatorname{covar}\left[N^{1/2}(\tilde{\mu} - \hat{\mu}), N^{3/2}(\tilde{f} - \hat{f})\right]^{\dagger} \geq \frac{\sigma_c^2}{4\pi^2} \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix}.$$

Under these conditions the periodogram estimator is also the maximum likelihood estimator. It can be seen that the periodogram estimator and Quinn-Fernandes estimator perform very close to the CRB when the noise variance is below a particular threshold value. The threshold grows as N increases. The angular least squares estimator performs very closely to the variance predicted by Theorem 8.1. We have plotted the exact angular least square estimator (diamonds) for $N = 4, 16, 64$ and it can be seen that the approximate angular least square estimator (crosses) gives a close approximation, but with substantially less computationally complexity.

Kay's estimator performs very poorly. This is due to the estimator failing when the true frequency $\tilde{f}$ is near $\pm 1/2$. In practice it might be possible to bound the frequency away from $\pm 1/2$ by, for example, increasing the rate (in Hz) at which observations are acquired. For this reason we have also plotted the MSE of Kay's estimator where $\tilde{f}$ is chosen uniformly in $[-0.3, 0.3)$. Kay's estimator works better under these conditions but is still not comparable with the other estimators when the noise variance is large.

Figures 9.2 and 9.3 display the performance of the estimators when the circular noise term Φ_n from (9.4.1) takes respectively the von Mises and the wrapped uniform distribution. In this case the complex samples Y_n are given by $e^{2\pi j\Theta_n}$ where Θ_n is calculated according to (9.4.1). For the von Mises distribution the periodogram and Quinn-Fernandes estimators give the best performance with the angular least squares estimator being marginally less accurate. For the wrapped uniform distribution we see that the angular least squares estimator gives the best performance. This is similar to the results found in Chapter 5 where it was noticed that for mean direction estimation the angular least squares estimator performs well when the noise distribution is 'uniform-like'. The variance given by Theorem 8.1 again accurately models the behaviour of the angular least squares estimator. Kay's estimator again only works well when the frequency and phase are bounded away from $\pm 1/2$ and even then Kay's estimator is not as accurate as the others.

9.6 Summary

This chapter considered the problem of signal frequency estimation. Three estimators that exist in the literature were described, the **periodogram estimator**, the **Quinn-Fernandes estimator** and **Kay's unwrapping estimator**. We also considered the **angular least squares estimator**.

We showed by Monte-Carlo simulation that the periodogram estimator, the Quinn-Fernandes estimator and the angular least squares estimator are all very accurate. The performance of the angular least squares estimator is well modelled by the central limit theorem derived in Theorem 8.1. Kay's unwrapping estimator only works well when the true value of the frequency parameter is bounded away from ± 0.5 and even then the performance is not as good as the other estimators.

The periodogram and Quinn-Fernandes estimators both require $O(N \log N)$ arithmetic operations and Kay's unwrapping estimator requires $O(N)$ operations where N is the number of observations. The angular least square estimator requires a nearest point in the lattice $V^*_{N-2/1}$ to be computed. If we use the algorithm from Chapter 4 then $O(N^5)$ operations are required. This is very slow, so we described a simple method to approximate the nearest point in $O(N^2 \log N)$ arithmetic operations. Although much faster, the complexity is still high when compared with other frequency estimators. However, it may be that much faster nearest point algorithms exist for $V^*_{N-2/1}$. As we have shown, even fast *approximate* nearest point algorithms would prove very useful for frequency estimation.

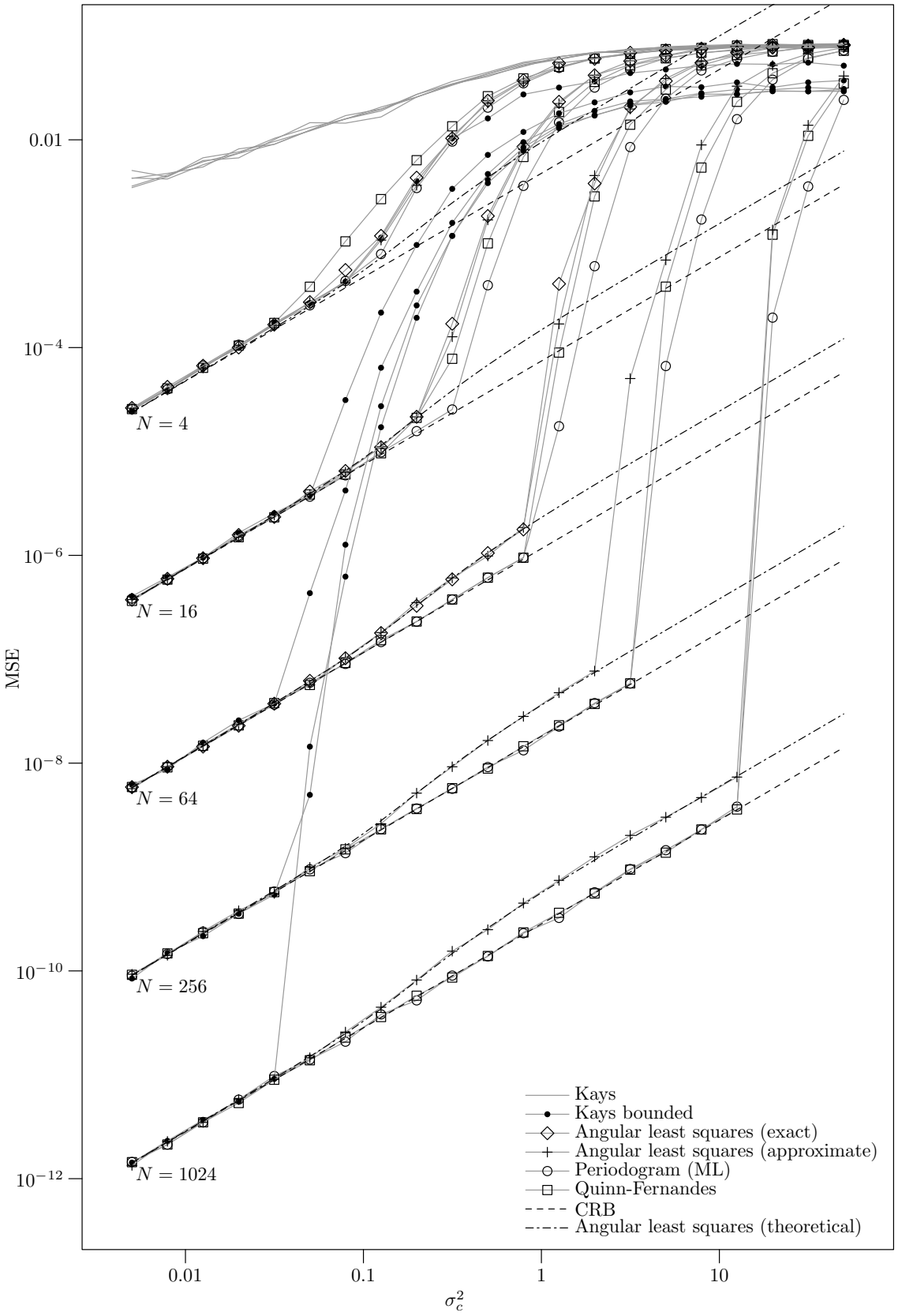


FIGURE 9.1: Mean square error in frequency with zero mean complex Gaussian noise having independent real and imaginary parts with variance σ_c^2.

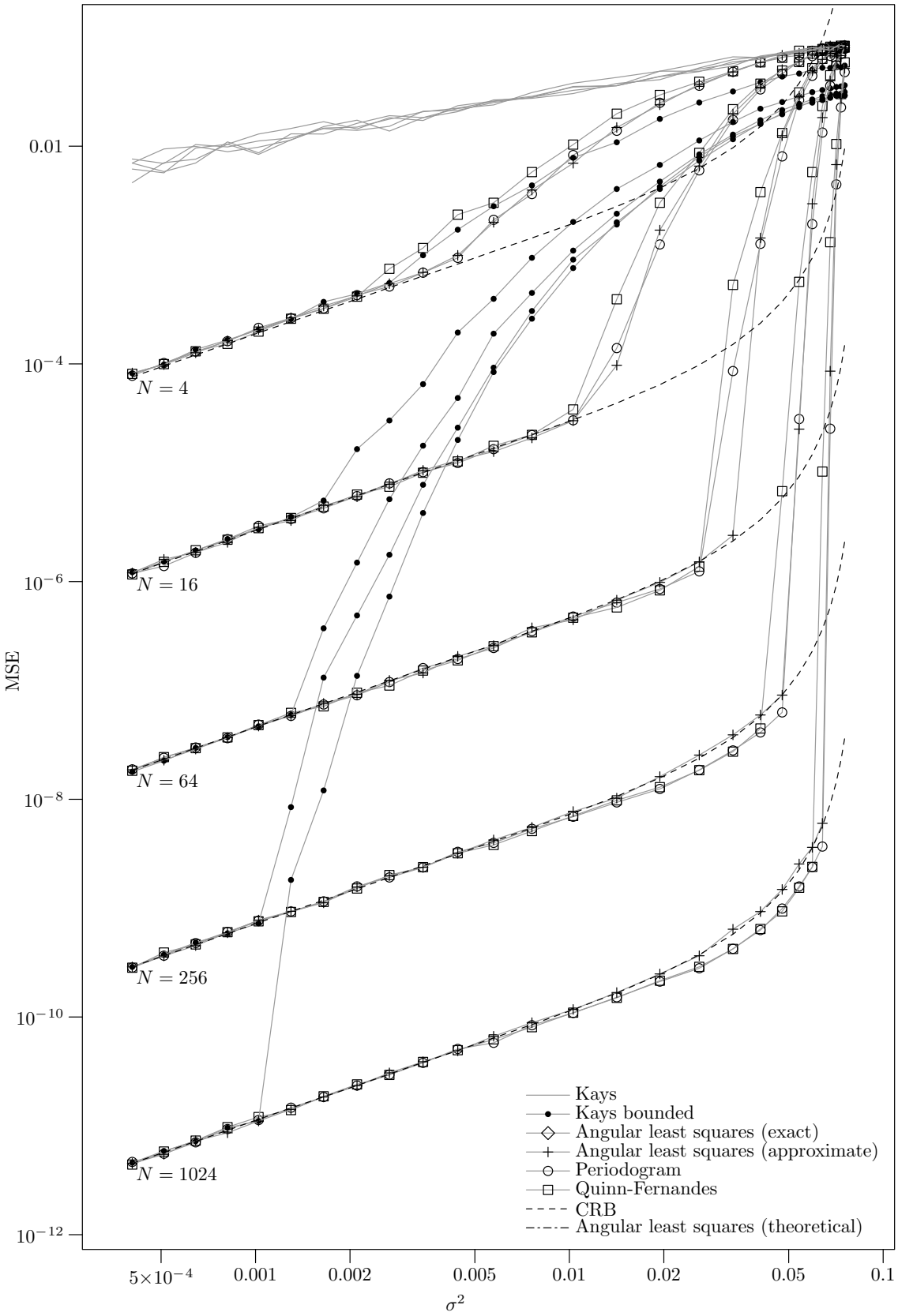


Figure 9.2: Mean square error in frequency with von Mises noise with zero unwrapped mean and unwrapped variance equal to σ^2.

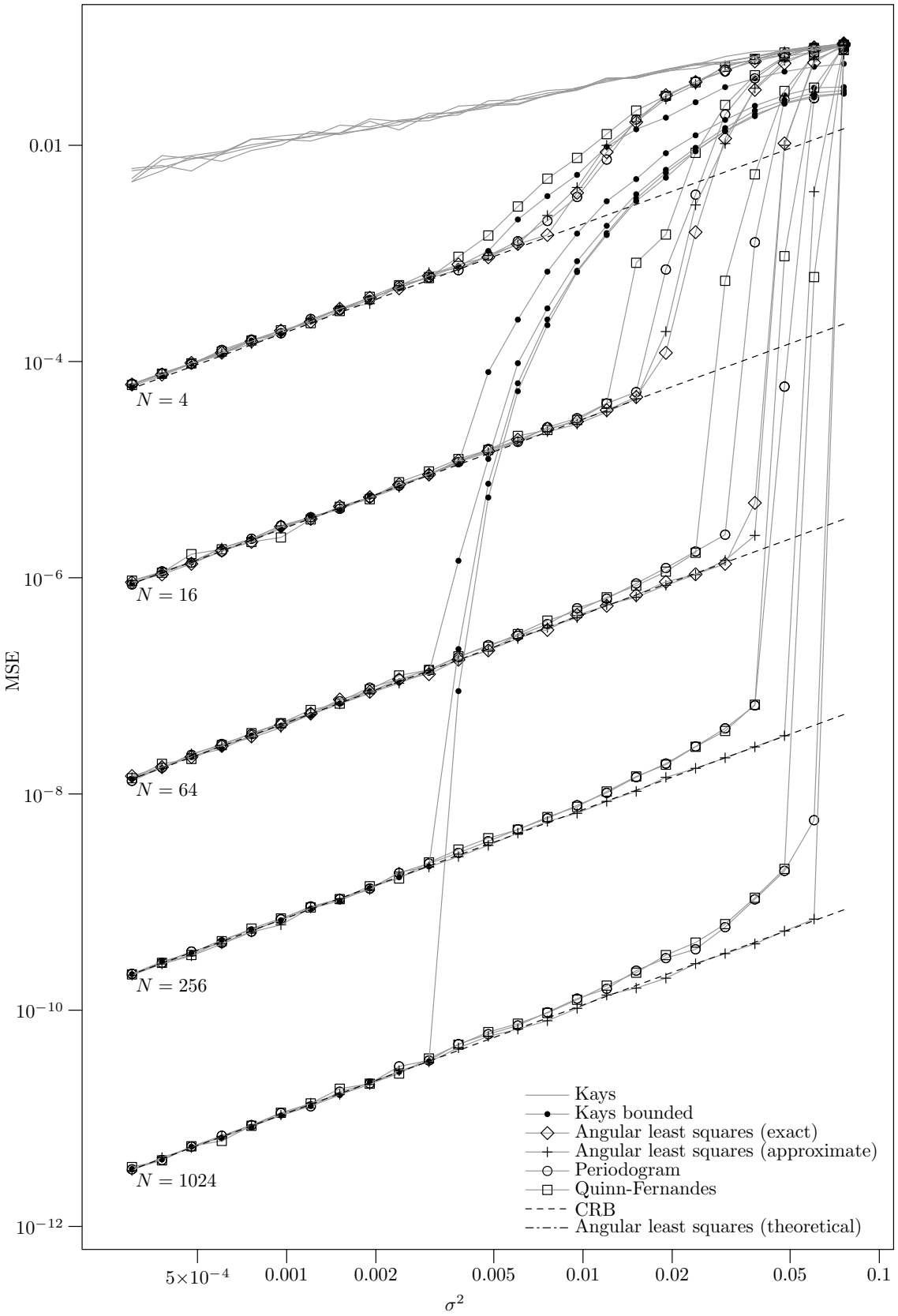


FIGURE 9.3: Mean square error in frequency with wrapped uniform noise with zero unwrapped mean and unwrapped variance equal to σ^2.

—Out of this haphazard excursion into number theory, it is my fervent hope that something coherent, interesting and useful has emerged.

I. Vaughan L. Clarkson

10 Polynomial phase estimation

In this chapter we consider estimating the $m+1$ coefficients of a polynomial phase signal of order m. The problem has applications in electrical engineering, particularly in radar, sonar and telecommunications and also in science, particularly in optics, geophysics and biology. In radar and sonar applications polynomial phase signals occur when acquiring radial velocity and acceleration (and higher order motion descriptors) of a target from a reflected signal. In biology, polynomial phase signals can be used to describe the sounds emitted by bats and dolphins for echo location [Suga et al., 1975; Thomas et al., 2005; Peleg and Friedlander, 1995].

In Section 10.1 we describe the **least squares estimator** for polynomial phase signals and show how it can be computed by iterating the **periodogram estimator** from frequency estimation (see Section 9.1). The least squares estimator is very computationally intensive and for this reason many authors have considered methods to reduce the computational complexity [Peleg and Friedlander, 1995; Kitchen, 1994; Morelande, 2009; Morelande and Zoubir, 2002; Djuric and Kay, 1990; O'Shea, 1996; Golden and Friedlander, 1998b, 1999, 1998a; Farquharson, 2006; Farquharson et al., 2005]. Loosely these techniques can be grouped into two classes, those based on **polynomial phase transforms** and those based on **phase unwrapping**. It is not the intention of this thesis to give a exhaustive review of the numerous estimators that exist in the literature so we will focus on two estimators that we feel are the most representative of these two classes. These are the **discrete polynomial phase transform** [Peleg and Friedlander, 1995] and **Kitchen's unwrapping estimator** [Kitchen, 1994].

Section 10.2 describes the **discrete polynomial phase transform** (DPT) that works by transforming the observed data so that each polynomial coefficient can be estimated individually using the periodogram estimator. The DPT is computationally efficient requiring only $O(N \log N)$ arithmetic operations where N is the number of observations. The estimator can also be statistically quite accurate. However, we find that the DPT does not work very well for some coefficients in the **identifiable**

region. This is, in a way, similar to how Kay's frequency estimator performs poorly when the true frequency is near ± 0.5. The problem is actually *far* more severe for the DPT estimator and we find that the range of coefficients for which the DPT estimator works correctly can be *extremely* small when compared to the identifiable region. Moreover, this range shrinks rapidly as the number of observations, N, increases. We consider how this problem might be overcome by increasing the rate at which observation are acquired (the **sample rate**), but we show that this comes with inevitable statistical penalties.

Section 10.3 describes **Kitchen's unwrapping estimator**. This estimator is a natural generalisation of Kay's unwrapping estimator for frequency estimation (Section 9.3). Kitchen's estimator shares many of the properties of Kay's unwrapping estimator, in that it is very computationally efficient, requiring only $O(N)$ operations, but it only works well when the noise variance and the number of observations is small. Also, Kitchen's estimator only works when the range of parameters is restricted (bounded away from the edge of the identifiable region), but this restriction is much less severe than it is for the discrete polynomial phase transform.

Section 10.4 considers the angular least squares estimator that was studied in Chapter 8. Computing the angular least squares estimator requires finding a nearest lattice point in the lattice $V^*_{n/m}$. The nearest point algorithm described in Section 4.3 can be to compute the estimator in a number of operations that is polynomial in N, but this turns out to be very computationally expensive in practice. So we consider alternative algorithms, the **sphere decoder**, the **K-best algorithm** and **Babai's nearest plane algorithm**, to compute or approximate the nearest point. These approaches are feasible, but they are still computationally more expensive than the discrete polynomial phase transform and Kitchen's unwrapping estimator.

In Section 10.5 we use Monte-Carlo simulation to compare the performance of the estimators in practice. We find that the angular least squares estimators and the least squares estimators are both very accurate and work well for coefficients anywhere in the identifiable region. Kitchen's unwrapping estimator and the DPT are less accurate. The DPT suffers from the fact that it operates very poorly on a large range of coefficients inside the identifiable region. We also discuss some computational properties of the various estimators.

Signal Model

Like frequency estimation the signal model for polynomial phase estimation is typically given in complex exponential form, that is, by N complex observations of the form

$$Y_n = \tilde{\rho} e^{2\pi i(\tilde{\mu}_0 + \tilde{\mu}_1 n + \tilde{\mu}_2 n^2 + \cdots + \tilde{\mu}_m n^m)} + X_n \tag{10.0.1}$$

where $\tilde{\rho} > 0$ is an unknown amplitude and $X_1, X_2, \ldots, X_N$ are zero mean complex random variables. The aim is to estimate the polynomial coefficients $\tilde{\mu}_0, \ldots, \tilde{\mu}_m$. As discussed in Section 7.2 in order to ensure identifiability it is necessary to restrict the coefficients such that $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_0, \ldots, \tilde{\mu}_m]$ is inside the identifiable region from (7.2.1),

$$B = \prod_{k=0}^{m} \left[-\frac{0.5}{k!}, \frac{0.5}{k!} \right)^2$$

that tessellates on the lattice L_{m+1}.

10.1 The least squares estimator

An obvious approach to estimating the polynomial coefficients is the least squares estimator. This is given by the minimisers of the sum of squares function

$$S(\rho, \mu_0, \mu_1, \ldots, \mu_m) = \sum_{n=1}^{N} \left| Y_n - \rho e^{2\pi i(\tilde{\mu}_0 + \tilde{\mu}_1 n + + \tilde{\mu}_2 n^2 + \cdots + \tilde{\mu}_m n^m)} \right|^2.$$

The minimisation is over the $m+1$ polynomial coefficients and also the amplitude ρ. It may at first seem that we need to perform a search over all of these parameters to find the minimum, but, similar to the approach taken for the periodogram frequency estimator, S can be conditioned with respect to ρ and μ_0 to obtain the simpler sum of squares function

$$S(\mu_1, \ldots, \mu_m) = \sum_{n=1}^{N} |Y_n|^2 - N^{-1} \left| \sum_{n=1}^{N} Y_n e^{-2\pi i(\tilde{\mu}_1 n + \tilde{\mu}_2 n^2 + \cdots + \tilde{\mu}_m n^m)} \right|^2.$$

Now the $\mu_1, \ldots, \mu_m$ can be estimated by searching over only these m coefficients.

We recommend the following approach to minimising this function. First perform a discrete search over the $m-1$ coefficients $\mu_2, \ldots, \mu_m$ using suitably many samples spaced over the identifiable region. We recommend using $4N^2$ samples for μ_2 and $4N^3$ samples for μ_3 and in general $4N^m$ for μ_m. We wish to spread these samples over the subset of the identifiable region associated with $\mu_2, \ldots, \mu_m$. To do this we take a generator for L_{m+1}, i.e. the matrix $\boldsymbol{\mathcal{P}}$, then the set of discrete samples is given (in vector form) as

$$V = \left\{ \boldsymbol{\mathcal{P}}\mathbf{u} \mid u_0 = 0, u_1 = 0, u_k = \left\{ 0, \tfrac{1}{4N^k}, \tfrac{2}{4N^k}, \ldots, 1 - \tfrac{1}{4N^k} \right\} \right\}.$$

Note that we have decided to index the vector $\mathbf{u}$ by starting at zero because this matches with the way we have indexed the polynomial coefficients. Each vector in $\mathbf{v} \in V$ corresponds to a discrete sample of $\mu_2, \ldots, \mu_m$ by setting

$$\mu_2 = v_2, \;\; \mu_3 = v_3, \;\; \ldots, \;\; \mu_m = v_m.$$

For each discrete sample we compute the signal

$$Y_n e^{-2\pi i(\mu_2 n^2 + \cdots + \mu_m n^m)}$$

and then obtain an estimate of the frequency parameter μ_1 by applying the periodogram estimator (see Section 9.1). Once the minimiser of $S(\mu_1, \ldots, \mu_m)$ over the discrete samples has been found the estimate is further refined using Newton's method. After this procedure it is possible that the estimate obtained is not in the identifiable region but instead is an *aliased version* of the desired estimate. This can be resolved using the dealiasing procedure described in Section 7.2.1, i.e. by applying the dealias$(\cdot)$ function. It is important to realise that this proceedure does

not guarantee that the global minimiser is found because Newton's method might fail. By increasing the number of discrete samples (i.e. the size of the set V) we can increase the chance of Newton's method succeeding. Using the number of discrete samples we have suggested does give good results in practice, but it may fail sometimes, we don't know. Abatzoglou [1986] considers this problem for the specific case when $m = 2$ and arrives at a similar conclusion as to the number of discrete samples required that we have here.

The overall complexity of this algorithm is proportional to the number of discrete samples $|V|$ multiplied by the computational complexity of the periodogram estimator which is $O(N \log N)$. Using the number of samples we have recommended results in an algorithm that requires $O(N^{m(m+1)/2} \log N)$ operations. For large N and moderate m this is computationally very expensive and for this reason many authors have considered methods to reduce the computational complexity. Loosely these techniques can be grouped into two classes, those based on **polynomial phase transforms** and those based on **phase unwrapping**.

10.2 The polynomial phase transform

The class of estimators based on polynomial phase transforms typically attempt to transform the received signal so that each coefficient can be estimated independently. This reduces the $m-1$ dimensional search used for the least squares estimator to $m-1$ one-dimensional searches. Perhaps the primary example of this approach is the **discrete polynomial phase transform** (DPT) suggested by Peleg and Friedlander [1995]. Modifications of the DPT have been suggested by O'Shea [1996] and Golden and Friedlander [1998b]. Another example of an estimator in this class is the so called **high-order phase function** considered by Farquharson et al. [2005] and Wang et al. [2008].

The basic idea behind the DPT is that of *convolution with a delayed complex conjugate*. The effect of this is to *difference* the phase of the signal and, by the calculus of finite differences, obtain a signal with phase described by a polynomial of a smaller order [Jordan, 1965]. The convolution is applied $m-1$ times so that the resulting signal resembles a single frequency signal and the coefficients can then be estimated using any of the techniques for frequency estimation discussed in the Chapter 9. Peleg and Friedlander [1995] suggest using the periodogram estimator and that is also what we recommend. This process of $m-1$ delayed convolutions followed by a discrete Fourier transform (as apart of the periodogram estimator) is what Peleg and Friedlander [1995] call the **discrete polynomial phase transform of order** m.

An adjustable quantity is the *delay* (sometimes also called the *lag*) used in the convolution operation which Peleg and Friedlander [1995] denote by τ. The statistical performance of the estimator is dependent on τ and, in order to maximise the performance Peleg and Friedlander [1995] suggest choosing $\tau = \frac{N}{m}$ or $\tau = \frac{N}{m+2}$. Here we will assume that $\tau = \frac{N}{m}$. The DPT estimator works correctly when the

coefficients satisfy

$$2|\mu_k| \leq \frac{1}{k!\tau^{k-1}} = \frac{1}{k!}\left(\frac{m}{N}\right)^{k-1} \tag{10.2.1}$$

for all $k = 0, \ldots m$ [Peleg and Friedlander, 1995, eq. (15)].

This means that the DPT works correctly only when the coefficients lie within a $m+1$ dimensional rectangular prism of volume

$$V_{DPT} = \prod_{k=0}^{m} \frac{1}{k!}\left(\frac{m}{N}\right)^{k-1} = \left(\frac{m}{N}\right)^{m(m-1)/2} \prod_{k=0}^{m} \frac{1}{k!}.$$

The volume of the identifiable region is given by the square root of the determinant of the lattice L_{m+1}, that is

$$V_m = \mathrm{vol}(\mathrm{Vor}(L_{m+1})) = \sqrt{\det L_{m+1}} = \prod_{k=0}^{m} \frac{1}{k!}.$$

Observing the ratio

$$\frac{V_{DPT}}{V_m} = \left(\frac{m}{N}\right)^{m(m-1)/2} \tag{10.2.2}$$

it is clear that, when $m > 1$, the range of coefficients for which the DPT works is small when compared to the size of the identifiable region. Moreover, the ratio $\frac{V_{DPT}}{V_m}$ shrinks quite rapidly as N increases. For large N this may make the DPT unusable as the true coefficients might lie outside the rectangular prism defined by (10.2.1). Note that the identifiable region is independent of N. As we shall see in Section 10.5 see the least squares estimator and the angular least square estimator work correctly for coefficients anywhere in the identifiable region.

In practice the observations are acquired at a particular **sample rate**, say δ, in Hz. If we desire the volume of acceptable coefficients V_{DPT} to remain constant as N increases it is necessary to increase the sampling rate. Including the rate parameter δ we find that the DPT works correctly if the coefficients satisfy

$$\frac{2|\mu_k|}{\delta^k} \leq \frac{1}{k!\tau^{k-1}} = \frac{1}{k!}\left(\frac{m}{N}\right)^{k-1}. \tag{10.2.3}$$

These inequalities correspond to coefficients lying inside a rectangular prism of volume

$$V_{DPT}(\delta) = \delta^{m(m+1)/2} V_{DPT} = \delta^{m(m+1)/2}\left(\frac{m}{N}\right)^{m(m-1)/2} V_m.$$

It makes sense to pick δ so that $V_{DPT}(\delta) = V_m$ so that the volume of acceptable coefficients for the DPT is equal to the volume of the identifiable region. In this case

$$\delta = \left(\frac{N}{m}\right)^{(m-1)/(m+1)}. \tag{10.2.4}$$

So the sampling rate must increase like $O(N^{1/3})$ when $m = 2$, like $O(N^{1/2})$ when $m = 3$, like $O(N^{3/5})$ when $m = 4$ and for large m the sample rate must increases close to linearly with N. We have plotted the required increase in the sampling rate for $m = 2, 3, 4$ in Figure 10.1.

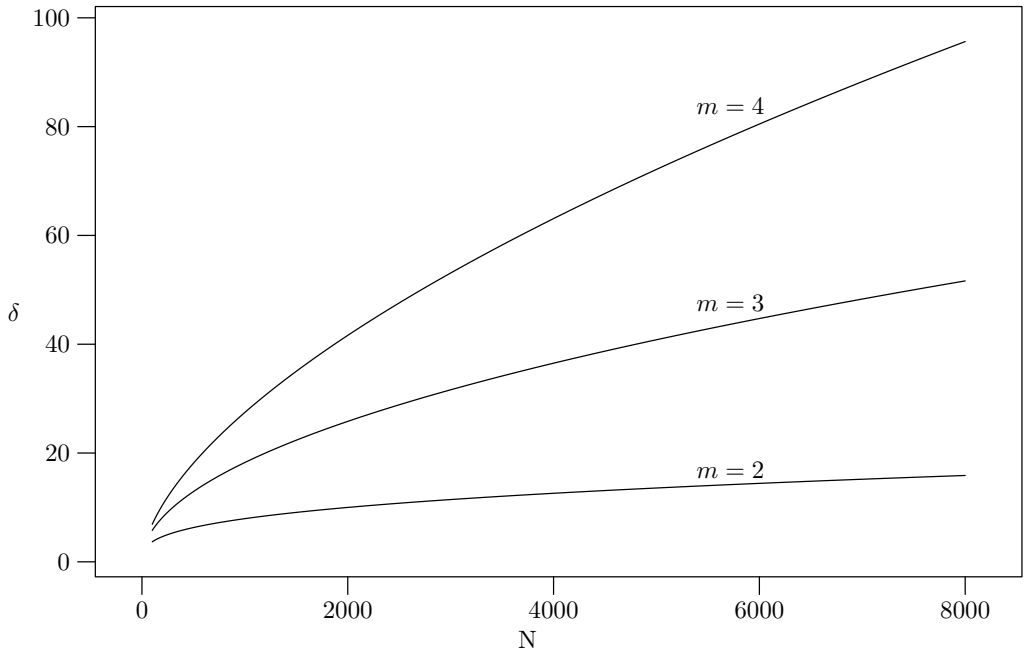


FIGURE 10.1: Required increase in sample rate for the DPT as the number of observations N increases for $m = 2, 3, 4$.

A word of caution is that although the volumes $V_{DPT}(\delta)$ and V_m are equal this *does not* mean that the $m + 1$ dimensional rectangular prism corresponding to $V_{DPT}(\delta)$ is equal to the identifiable region B. The inequalities from (10.2.3) do not describe a tessellating region for the lattice L_{m+1}. However, for the purpose of comparing the DPT with the other estimators, selecting δ so that the volumes $V_{DPT}(\delta)$ and V_m are equal appears to be the fairest approach. At least, under this condition, the *volume* of acceptable coefficients is the same. Whether it is possible to increase the sampling rate in practice is likely to be highly application dependent. However, as we shall see in Section 10.5, increasing the sampling rate comes with inevitable statistical penalties.

The **high-order phase function** estimator of Farquharson et al. [2005] also only works on a very restricted set of coefficients. Here the estimator, in the most general case, only works correctly if the coefficients lie inside a rectangular prism of volume

$$V_{HPF} = \left(\frac{2}{N}\right)^{m(m-1)/2} \prod_{k=0}^{m} \frac{1}{k!}$$

which is even smaller than V_{DPT}. In order to make the volume V_{HPF} equal to the volume of the identifiable region V_m we would need to select the sample rate

$$\delta = \left(\frac{N}{2}\right)^{(m-1)/(m+1)}$$

which is even faster than that required by the DPT.

10.3 Kitchen's unwrapping estimator

Motivated by the phase unwrapping approach to frequency estimation suggested by Tretter [1985] and Kay [1989], both Djuric and Kay [1990] and Kitchen [1994] suggested approaches to polynomial phase estimation based on phase unwrapping ideas. We will consider Kitchen's estimator here.

Kitchen considered the mth difference of the complex argument of the Y_n. Like Kay's frequency estimator the resulting signal resembles a moving average process and can be estimated by standard linear techniques. A significant advantage is that the moving average process has enough structure for the estimates to be computed with only $O(N)$ arithmetic computations. Kitchen's estimator shares many of the properties of Kay's frequency estimator in that it appears to be statistically efficient only when both the noise variance and N are small. Also, Kitchen's estimator does not appear to work well over the entire identifiable region. This is similar to how Kay's estimator fails when the true frequency and phase of the signal is near ± 0.5. The situation is far less severe than for the DPT. Experimentation reveals that the estimator works correctly when the kth coefficient is bounded away from $\pm\frac{0.5}{k!}$. For the simulations in Section 10.5 we find that if the coefficients lie in the prism

$$\prod_{k=1}^{m}\left[-\frac{0.3}{k!},\frac{0.3}{k!}\right)$$

then Kitchen's estimator appears to work reasonably well. Note that the volume of this prism is smaller than the identifiable region by the factor $(\frac{3}{5})^{m+1}$ but, in contrast to the discrete polynomial phase transform, the volume does not shrink as the number of observations N increases.

A problem with Kitchen's estimator is that we need to bound the phase coefficient, $\tilde{\mu}_0$ away from ± 0.5. In practice the higher order coefficients $\tilde{\mu}_1,\ldots,\tilde{\mu}_m$ can all be bounded away from $\pm\frac{0.5}{k!}$ by slightly increasing the sampling rate, but we do not have similar control over the phase coefficient. Fortunately, an easy fix is to use Kitchen's estimator to obtain the estimates $\hat{\mu}_1,\ldots,\hat{\mu}_m$ of the higher order coefficients, then compute the estimate $\hat{\mu}_0$ of the phase coefficients from the signal $Y_n e^{-2\pi(\hat{\mu}_1+\cdots+\hat{\mu}_m)}$ by applying any of the techniques for **phase estimation** that we discussed in Section 6.4. In our implementation we have used the least squares estimator. The complexity of the estimator is still $O(N)$ with this modification, but the estimator now works well when the coefficients lie in the prism

$$\left[-\frac{1}{2},\frac{1}{2}\right)\times\prod_{k=1}^{m}\left[-\frac{0.3}{k!},\frac{0.3}{k!}\right) \tag{10.3.1}$$

which has volume $(\frac{3}{5})^m$. In the simulations in Section 10.5 we will use this modification of Kitchen's estimator.

10.4 Approximating angular least squares

In order to use the angular least squares estimator we take the complex argument of the Y_n and divide by 2π to obtain the circular random variables

$$\Theta_n = \tfrac{\angle Y_n}{2\pi} = \left\langle \Phi_n + \tilde{\mu}_0 + \tilde{\mu}_1 n + + \tilde{\mu}_2 n^2 + \cdots + \tilde{\mu}_m n^m \right\rangle \tag{10.4.1}$$

where the $\Phi_n = \frac{1}{2\pi}\angle(\tilde{\rho} + X_n)$ are projected circular random variables. The Θ_n are now in the form of (8.1.1).

Computing the angular least squares estimator requires finding a nearest point in the lattice $V^*_{n/m}$. We could use the algorithm suggested in Section 4.3 to compute a nearest point in $O(N^{(m+1)^2+1})$ operations. However, when $m = 2$ this is $O(N^{10})$ and when $m = 3$ this is $O(N^{17})$ so although these are polynomial-time in N they are far too slow for practical use. We suggest the following approximate approaches in practice.

When N is small (approximately less than 50) the **sphere decoder** can be used to compute the nearest point exactly and quite rapidly. For any N **Babai's nearest plane algorithm** can be used to approximate the nearest point in only $O(N^2)$ operations. However, we will find that a reasonably significant performance penalty is suffered by using Babai's nearest plane algorithm if the noise variance is large. Finally the **K-best algorithm** can be used to approximate the nearest point in $O(K^2N^2 \log K)$ operations [Guo and Nilsson, 2006]. We have found that setting $K = 4N$ seems to work well and in this case the algorithm requires $O(N^3 \log N)$ operations.

Even these approximate approaches are quite computationally expensive compared to the DPT that requires only $O(N \log N)$ operations and Kitchen's unwrapping estimator that requires on $O(N)$ operations. For now at least this might restrict the use of the angular least squares estimator to cases where N is small. However, there may exist significantly faster methods for computing or approximating a nearest point in the lattice $V^*_{n/m}$ that we have not discovered yet. In the next section we will display the excellent statistical performance of the angular least squares estimator. Due to this performance, studying the properties of $V^*_{n/m}$ more thoroughly, particularly with a view to finding faster nearest point algorithms is a worthy direction for future research.

10.5 Simulations

In this section Monte-Carlo simulation is used to compare the performance of the least squares estimator, the DPT, Kitchen's estimator and the angular least squares estimator computed using Babai's nearest plane algorithm, the sphere decoder, and the K-best algorithm.

Figures 10.4 to 10.15 display the performance of the estimators for a polynomial phase signal of order 3. The number of observations is $N = 16, 64$ and 256. It is only possible to compute the exact angular least squares estimator using the sphere decoder when $N = 16$ and 64. The noise term X_n is complex Gaussian with

independent real and imaginary parts having variance σ_c^2. In this case the **Cramer-Rao lower bound** (CRB) has been derived by Peleg and Porat [1991] who note that computing the exact CRB for small N is numerically difficult, but for large N a close approximation is given by

$$\text{covar}\left[\begin{array}{cccc} N^{1/2}(\tilde{\mu}_0-\hat{\mu}_0) & \ldots & N^{(2m+1)/2}(\tilde{\mu}_m-\hat{\mu}_m) \end{array}\right] \geq \frac{\sigma_c^2}{4\pi^2}\mathbf{C}^{-1}$$

where $\mathbf{C}$ is the $m+1$ by $m+1$ **Hilbert matrix** (see Section 8.2 page 117).

For Figures 10.4 to 10.7 (pages 158 and 159) the true coefficients $\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3$ are distributed uniformly in the identifiable region. Both the DPT and Kitchen's unwrapping estimator display very poor performance. The sphere decoder and K-best methods perform similarly and a rather significant performance penalty is suffered by using Babai's nearest plane algorithm. Also plotted is the CRB and the asymptotic variance predicted by Theorem 8.1 which can be seen to accurately model the behaviour of the angular least squares estimator provided that the noise variance is small enough to avoid the threshold effect.

For Figures 10.8 to 10.11 (pages 160 and 161) the true coefficients $\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3$ are restricted to suit the particular estimators. For the DPT the coefficients all satisfy (10.2.1) and for Kitchen's estimator the coefficients are all inside the prism defined by (10.3.1). For the angular least squares estimator the coefficients are again generated uniformly within the identifiable region. Under these conditions both Kitchen's estimator and the DPT work, although they are still not as accurate as the angular least squares estimator.

The results displayed in Figures 10.8 to 10.11 are, in a sense, *unfair*. The reason is that the DPT and Kitchen's estimator have been given *extra information* about the true coefficients. In the case of the Kitchen's estimator the coefficients have only been restricted slightly by a factor (in volume) of $\left(\frac{3}{5}\right)^m$ and this is probably negligible. However, for the DPT the coefficients have been *severely* restricted. From (10.2.2), when $N = 256$, the volume of coefficients searched by the DPT is $\left(\frac{3}{256}\right)^3 \approx 1.6 \times 10^{-6}$ times the size of the identifiable region. The DPT is searching a space almost *one million times smaller* than the angular least squares estimator! In this context we would probably expect the DPT to work much better than the angular least square estimator. It is somewhat surprising that it does not.

In order to display more *fair* results we can increase the sample rate of the DPT so that the volume of coefficients to which it applies is equal to the volume of the identifiable region. This comes with some inevitable statistical penalties. If we rederive the Cramèr-Rao lower bound taking account of the fact that the sample rate is now δ given by (10.2.4) then we obtain

$$\text{covar}\left[\begin{array}{cccc} N^{1/2}(\tilde{\mu}_0-\hat{\mu}_0) & \ldots & N^{(2m+1)/2}(\tilde{\mu}_m-\hat{\mu}_m) \end{array}\right] \geq \frac{\sigma_c^2}{4\pi^2}\mathbf{S}\mathbf{C}^{-1}$$

where $\mathbf{S}$ is the $m+1$ by $m+1$ diagonal matrix with diagonal entries given by $1, \delta^2, \delta^4, \ldots, \delta^{2m}$. The DPT, applied at the sample rate δ, will not achieve a mean square error less than this new CRB. The results are displayed in Figures 10.12 to 10.15 and it can be seen that in this scenario the performance of the DPT is, as predicted, substantially worse than before.

In Figures 10.2 and 10.3 we compare the angular least squares estimator with the least squares estimator. Here, the polynomial phase signal has order 2 and we have only plotted results for the highest order parameter μ_2. It is only possible to run the least squares estimator for small N and we have used $N = 10$ and 50. In Figure 10.2 the noise terms X_n are complex Gaussian with independent real and imaginary parts having variance σ_c^2. Under these conditions the least squares estimator is also the maximum likelihood estimator. The performance of the angular least squares estimator computed exactly using the sphere decoder and approximately using Babai's nearest plane algorithm is displayed. It can be seen that the least squares estimator has better performance. When $N = 50$ the threshold effect occurs for larger variance with the least squares estimator. It is possible that this gap will close as N increases as this type of behaviour is observed in the results displayed for frequency estimation in Section 9.5 (see Figure 9.1). We unfortunately are unable to test this due to the computational complexity of the least squares estimator.

Figure 10.3 displays the performance when the circular noise terms Φ_n from (10.4.1) take the wrapped uniform distribution. In this case the complex samples Y_n are given by $e^{2\pi j\Theta_n}$ where Θ_n is calculated according to (10.4.1). Here, the angular least squares estimator computed using the sphere decoder performs slightly better than the least squares estimator.

Computational considerations

The least squares estimator can only be computed in a reasonable amount of time for N approximately less than 50 and $m = 2$, but is very computationally intensive for any N when $m > 2$. By comparison the sphere decoder estimator can be computed quite quickly for any m with N approximately less than 50. In situations where N is small, but high statistical accuracy is required, the sphere decoder is computationally a better choice than the least squares estimator, especially for polynomial phase signals of order greater than 2.

Babai's nearest plane algorithm only requires $O(N^2)$ operations and can be run in a reasonable amount of time for any m and even very large N. It should be noted that Babai's nearest plane algorithm requires a **Lovàs reduced basis** [Lenstra et al., 1982] for the lattice $V_{n/m}^*$. This requires $O(N^4)$ operations to compute, but, the Lovàs reduced basis can be computed once offline. It does not need to be computed each time the estimator is run. The K-best algorithm runs in a reasonable amount of time for any m and when N is less than about 300. If the noise variance is small, then both the sphere decoder and the K-best algorithm are actually very fast. This is because both of these estimators begin with the lattice point found by Babai's nearest plane algorithm. If this point is close to (or *is*) the nearest point then both the sphere decoder and the K-best algorithm terminate quite quickly. This is in contrast to the least squares estimator that is slow regardless of the noise variance.

Both Kitchen's estimator and the DPT are very fast to compute. These estimators can feasibly be run for any m and extremely large N. However, for very large N the DPT might not be applicable because as N increases the region of coefficients for which it works correctly decreases, and for very large N, the true coefficient might lie outside this region.

10.6 Summary

In this chapter we have described a number of techniques for estimating the $m+1$ coefficients of a polynomial phase signal of order m. We first considered the least squares estimator and described how it could be computed by sampling an objective function over the $m-1$ highest order coefficients $\mu_2, \mu_3, \ldots, \mu_m$ and applying the periodogram frequency estimator. Using the number of samples that we recommend results in an estimator that requires $O(N^{m(m+1)/2} \log N)$ operations. This is very computationally intensive, and is infeasible when $m > 2$. For this reason many authors have considered computationally tractable approaches and we studied two of these: the **discrete polynomial phase transform** (DPT) [Peleg and Friedlander, 1995] and **Kitchen's unwrapping estimator** [Kitchen, 1994].

Using ideas from the calculus of finite differences the DPT operates on the observed signal so that each coefficient can be estimated iteratively using the periodogram estimator used for frequency estimation (Section 9.1). The DPT requires only $O(N \log N)$ operations. The major drawback of the DPT is that it only operates correctly for a very small portion of the identifiable region, and worse, this region shrinks rapidly as the number of observations, N, increases. This might make the DPT estimator infeasible for large N as the true coefficients might lie outside the region acceptable for the DPT. For some applications this problem may be overcome by increasing the **sampling rate**. We showed that the sampling rate must increase like $O(N^{(m-1)/(m+1)})$ in order for the volume of acceptable coefficients to remain constant. Section 10.5 showed that increasing the sample rate comes with inevitable statistical penalties.

Section 10.3 described Kitchen's unwrapping estimator that is very computationally efficient, requiring only $O(N)$ operations, but only works well when the noise variance is small. The estimator also only works when the range of coefficients is restricted to a region smaller than the identifiable region, but this restriction is far less severe than for the DPT and the volume of acceptable coefficients does not shrink as N increases.

Section 10.4 considered the **angular least squares estimator** which involves computing a nearest point in the lattice $V^*_{n/m}$. The polynomial time nearest point algorithm described in Chapter 4 is unfortunately too slow for practical use and we therefore considered some general purpose algorithms, the **sphere decoder**, **Babai's nearest plane algorithm** and the **K-best algorithm**. Both the sphere decoder and K-best algorithm are statistically very accurate. Figures 10.2 and 10.3 showed that the performance is similar to the least squares estimator. If N is small, then the sphere decoder and K-best algorithms are in practice much computationally simpler than the least squares estimator. We found that Babai's nearest plane algorithm is computationally efficient, requiring only $O(N^2)$ arithmetic operations, but its statistical performance is not as good. Unlike the DPT and Kitchen's unwrapping estimator the angular least squares estimator works uniformly well over the entire identifiable region. In cases where N is large and the noise variance is not too large, this property makes the approximate angular least squares estimator computed using Babai's nearest plane algorithm an attractive choice.

The algorithms that we have used for computing the angular least squares estimator are still slower than the DPT estimator and Kitchen's unwrapping estimator. However, there may exist fast (exact or approximate) nearest point algorithms for $V_{n/m}^*$ that we have not found yet. Considering the statistical superiority of the angular least squares estimator the search for faster nearest point algorithms for $V_{n/m}^*$ is a promising direction of future research.

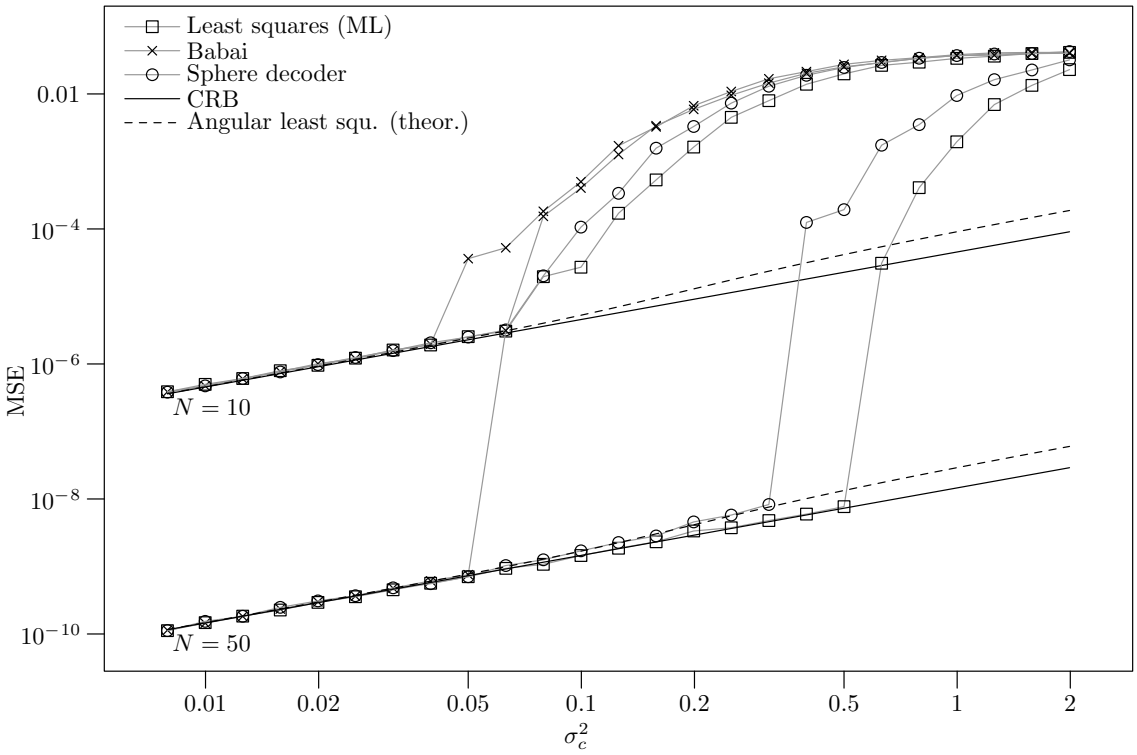


Figure 10.2: MSE in second order parameter μ_2 for $N = 10, 50$ versus the variance σ_c^2 of the X_n. The X_n are complex Gaussian random variables.

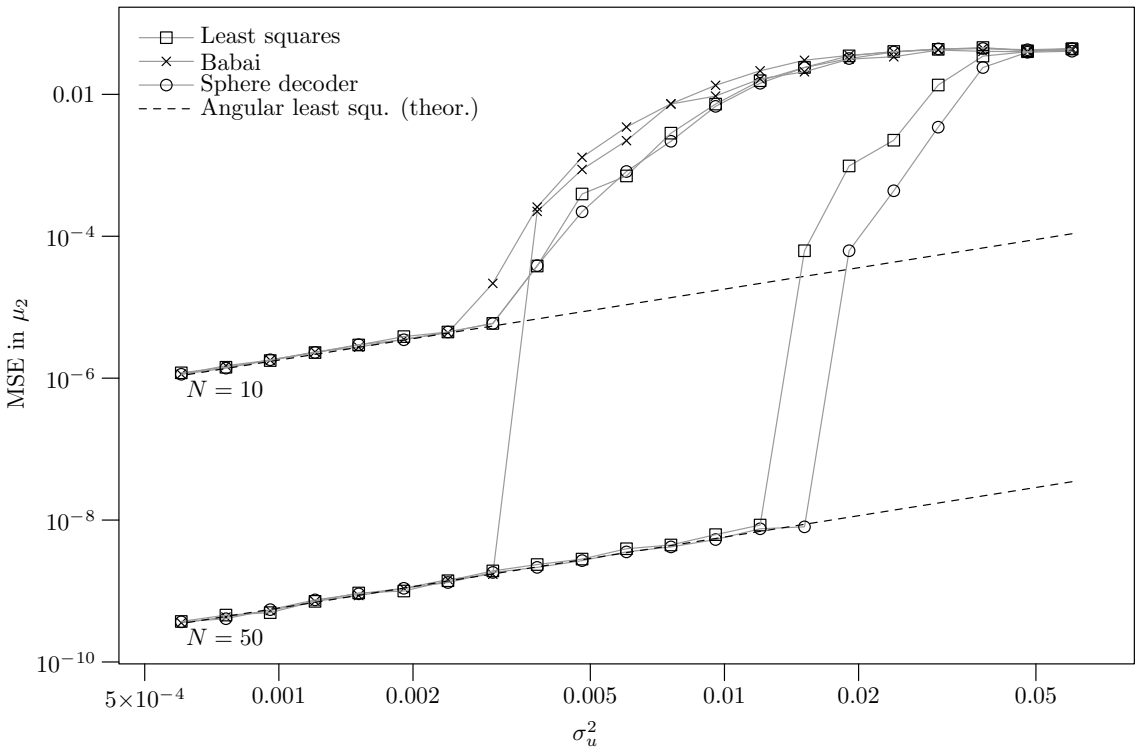


Figure 10.3: MSE in second order parameter μ_2 for $N = 10, 50$ versus the unwrapped variance σ_u^2 of the Φ_n. The Φ_n are zero mean wrapped uniform circular random variables.

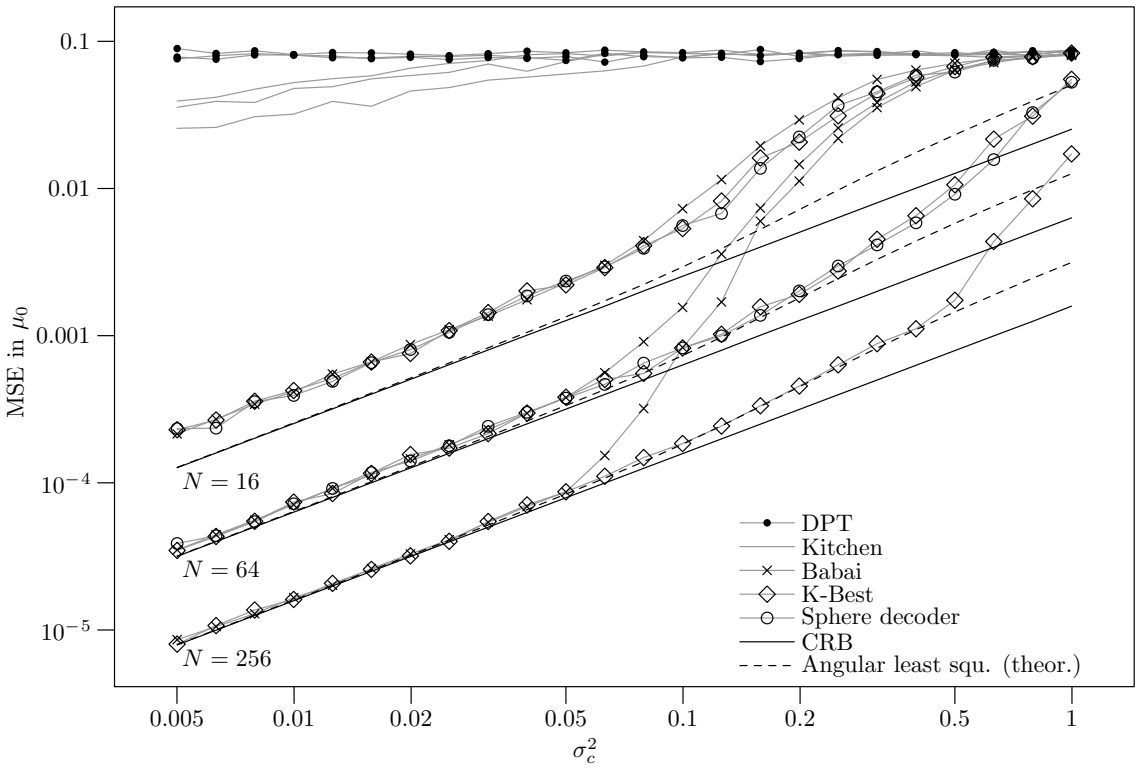


FIGURE 10.4: MSE in the phase coefficient μ_0 versus $\operatorname{var} X_n = \sigma_c^2$. The true coefficients are uniformly spread in the identifiable region.

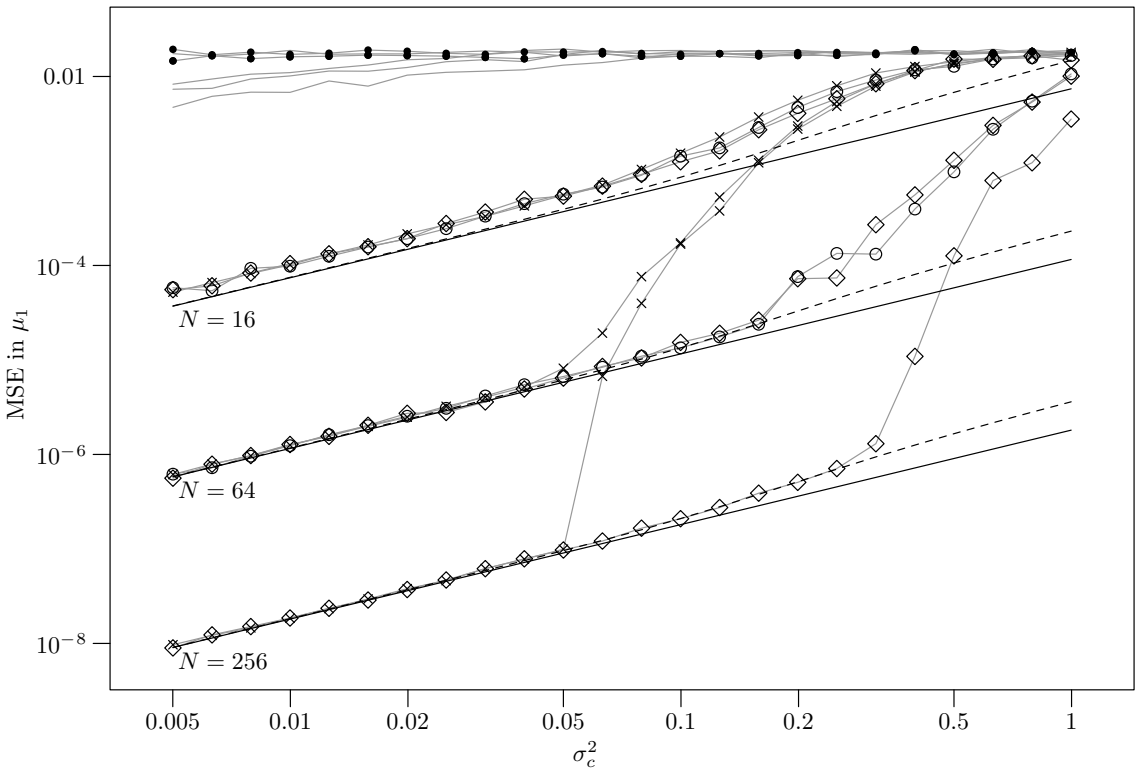


FIGURE 10.5: MSE in the frequency coefficient μ_1 versus $\operatorname{var} X_n = \sigma_c^2$. The true coefficients are uniformly spread in the identifiable region.

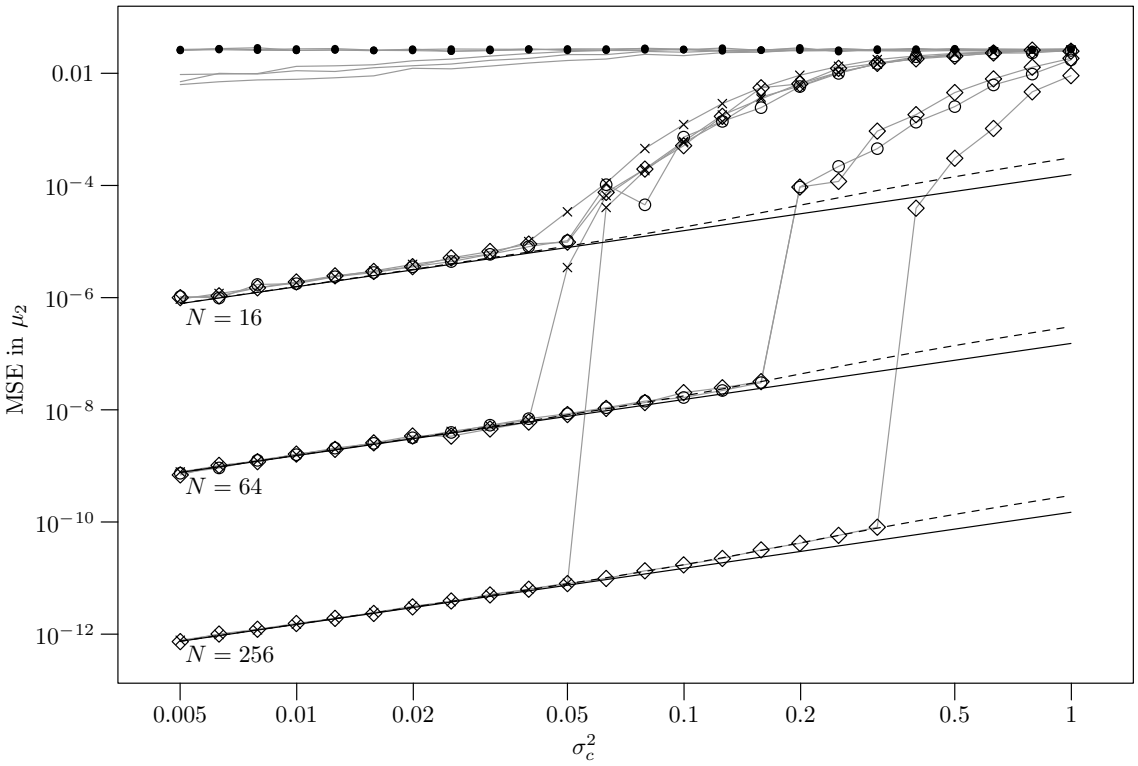


FIGURE 10.6: MSE in the second order coefficient μ_2 versus $\operatorname{var} X_n = \sigma_c^2$. The true coefficients are uniformly spread in the identifiable region.

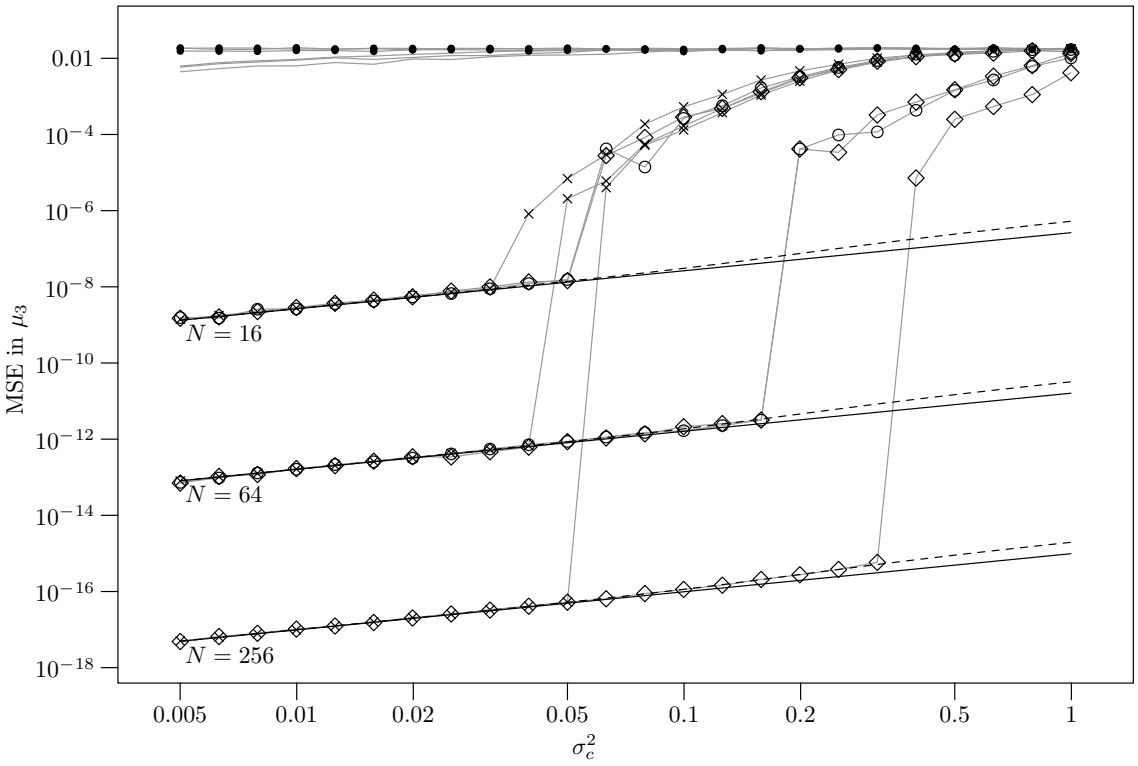


FIGURE 10.7: MSE in the third order coefficient μ_3 versus $\operatorname{var} X_n = \sigma_c^2$. The true coefficients are uniformly spread in the identifiable region.

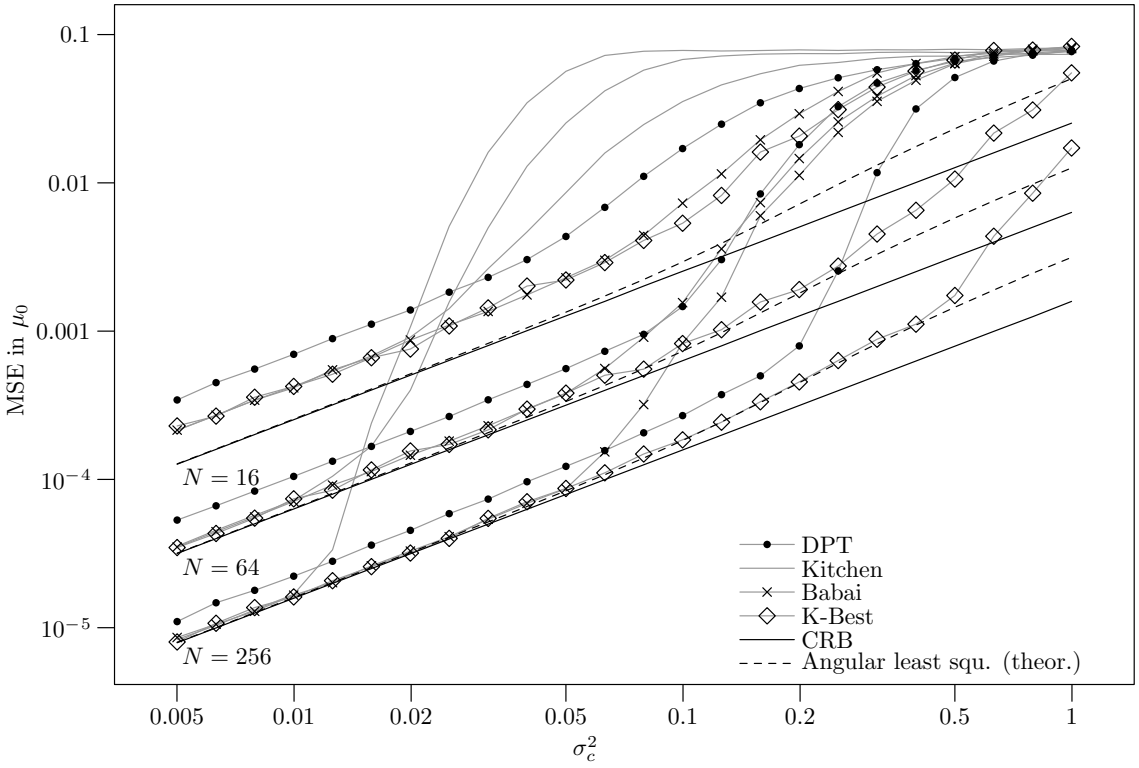


FIGURE 10.8: MSE in the phase coefficient μ_0 versus $\operatorname{var} X_n = \sigma_c^2$. The coefficient have been restricted for Kitchen's estimator and the DPT.

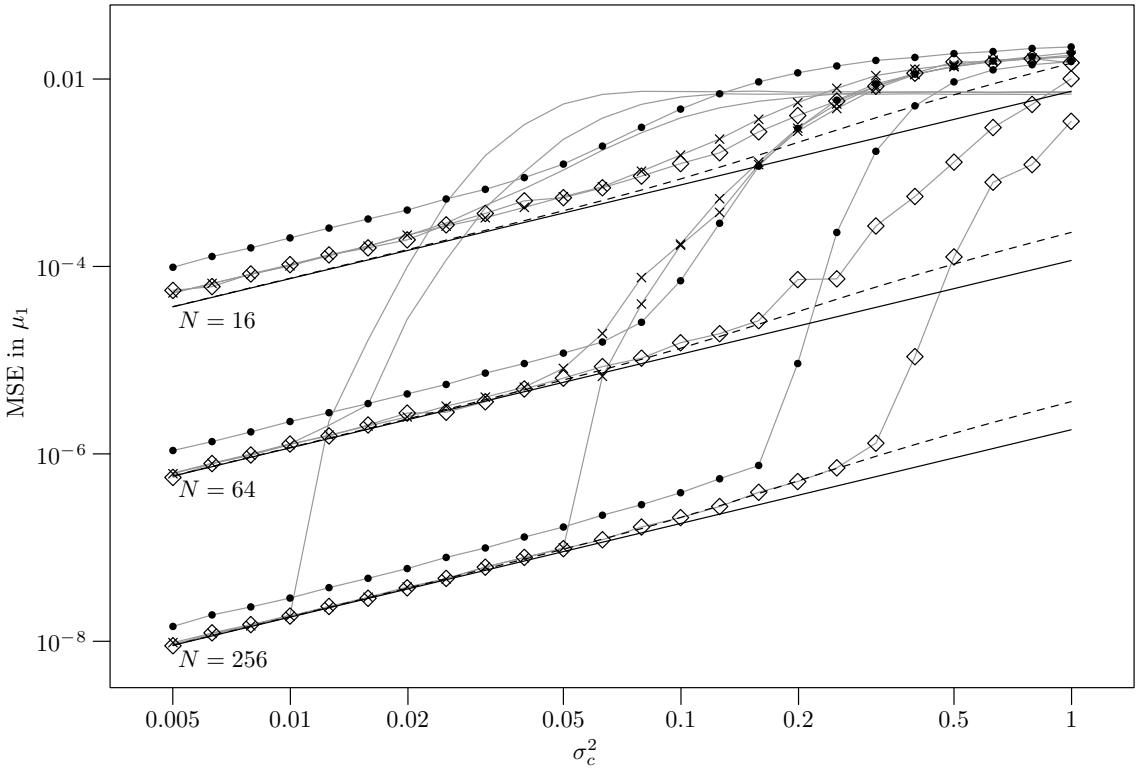


FIGURE 10.9: MSE in the frequency coefficient μ_1 versus $\operatorname{var} X_n = \sigma_c^2$. The coefficient have been restricted for Kitchen's estimator and the DPT.

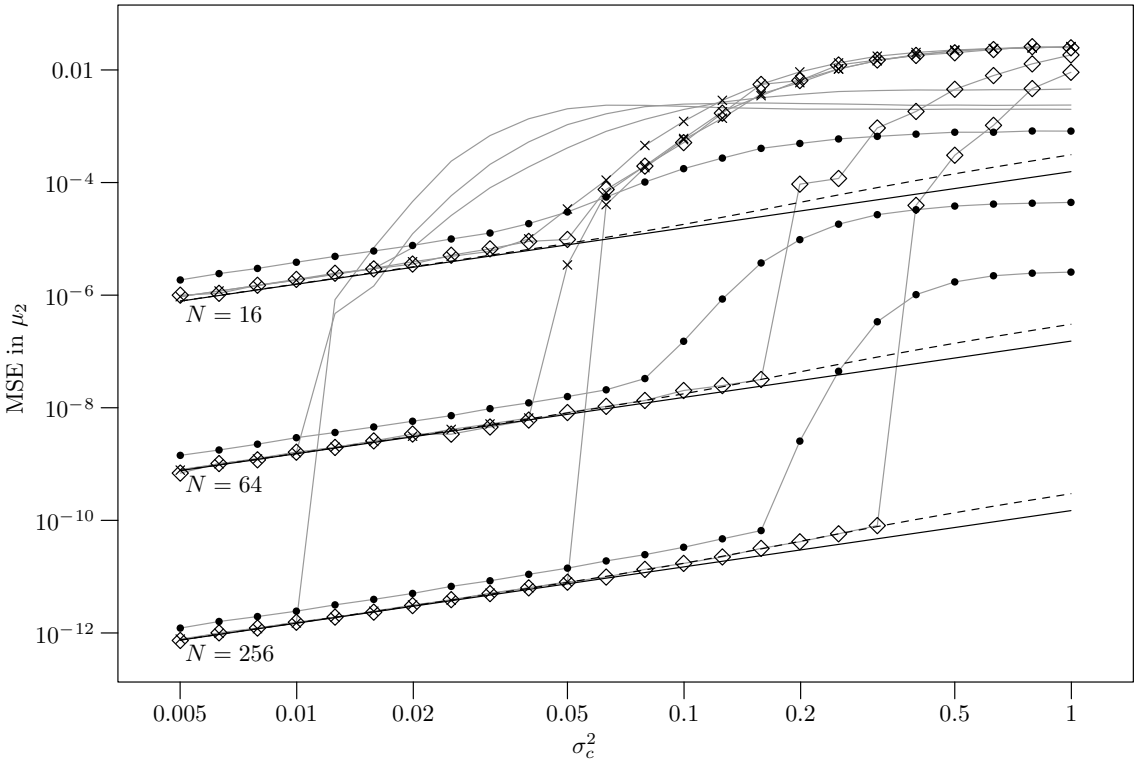


Figure 10.10: MSE in the second order coefficient μ_2 versus $\operatorname{var} X_n = \sigma_c^2$. The coefficient have been restricted for Kitchen's estimator and the DPT.

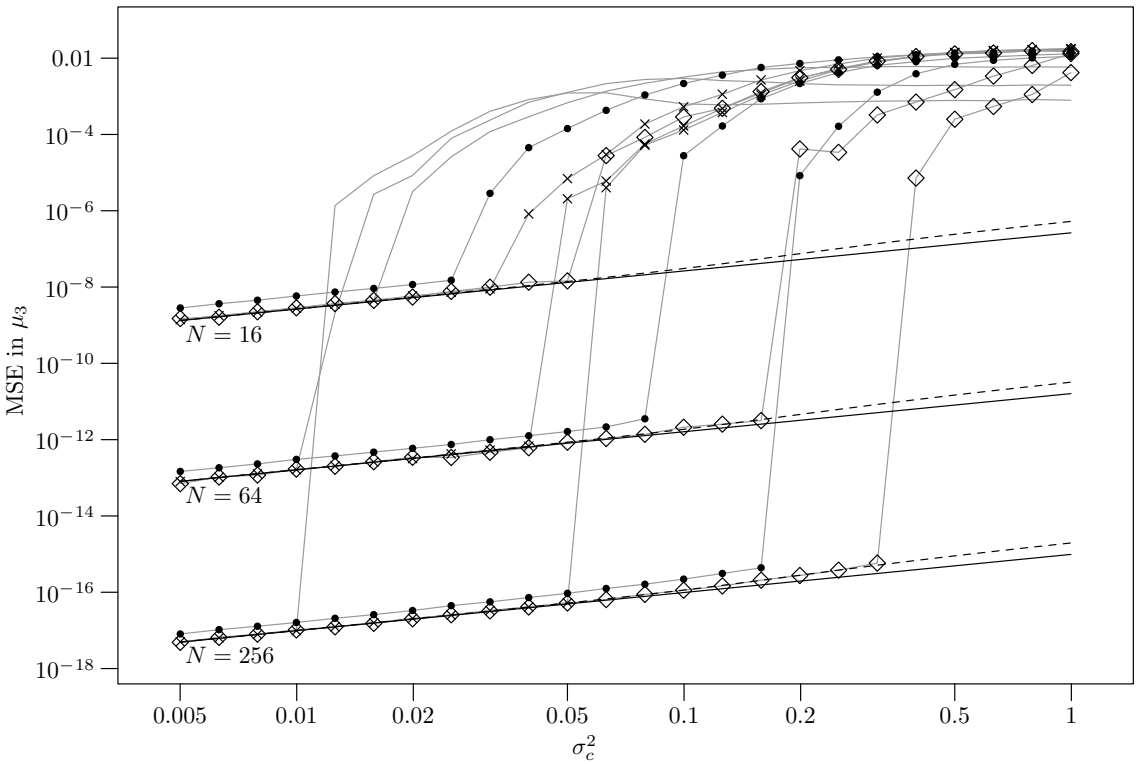


Figure 10.11: MSE in the third order coefficient μ_3 versus $\operatorname{var} X_n = \sigma_c^2$. The coefficient have been restricted for Kitchen's estimator and the DPT.

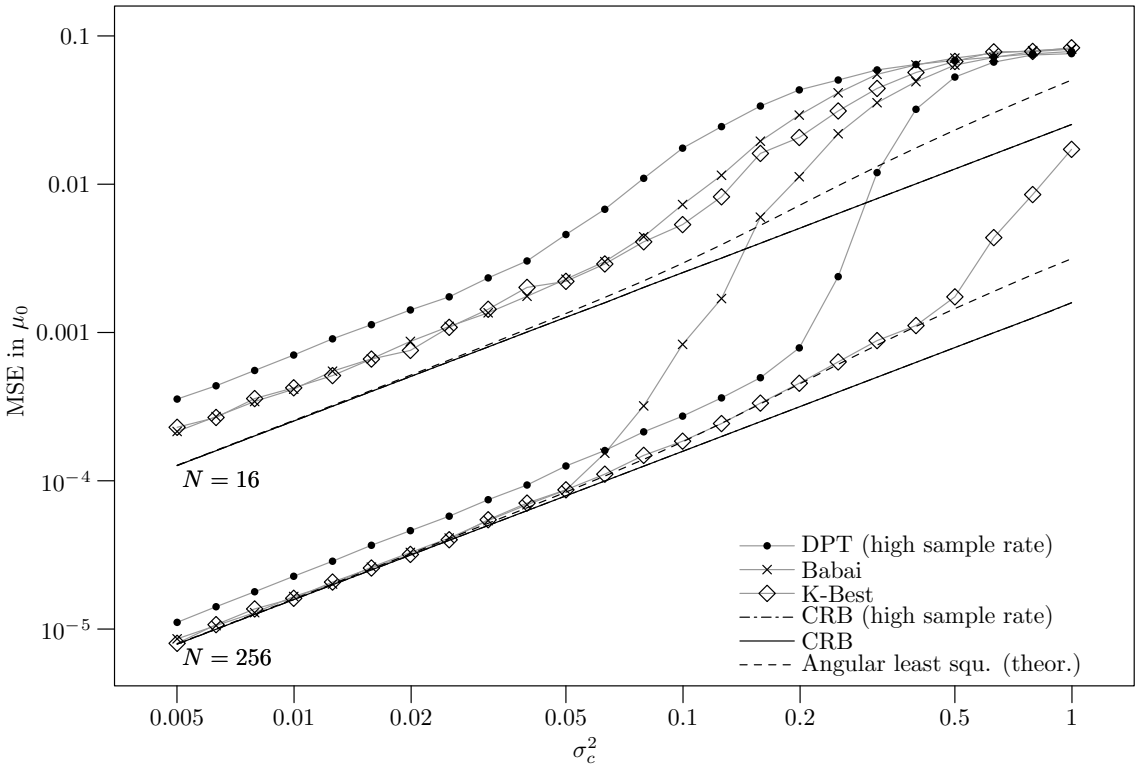


FIGURE 10.12: MSE in the phase coefficient μ_0 versus $\operatorname{var} X_n = \sigma_c^2$. The DPT estimator runs at the higher sampling rate δ so that the volumes $V_{DPT}(\delta) = V_m$.

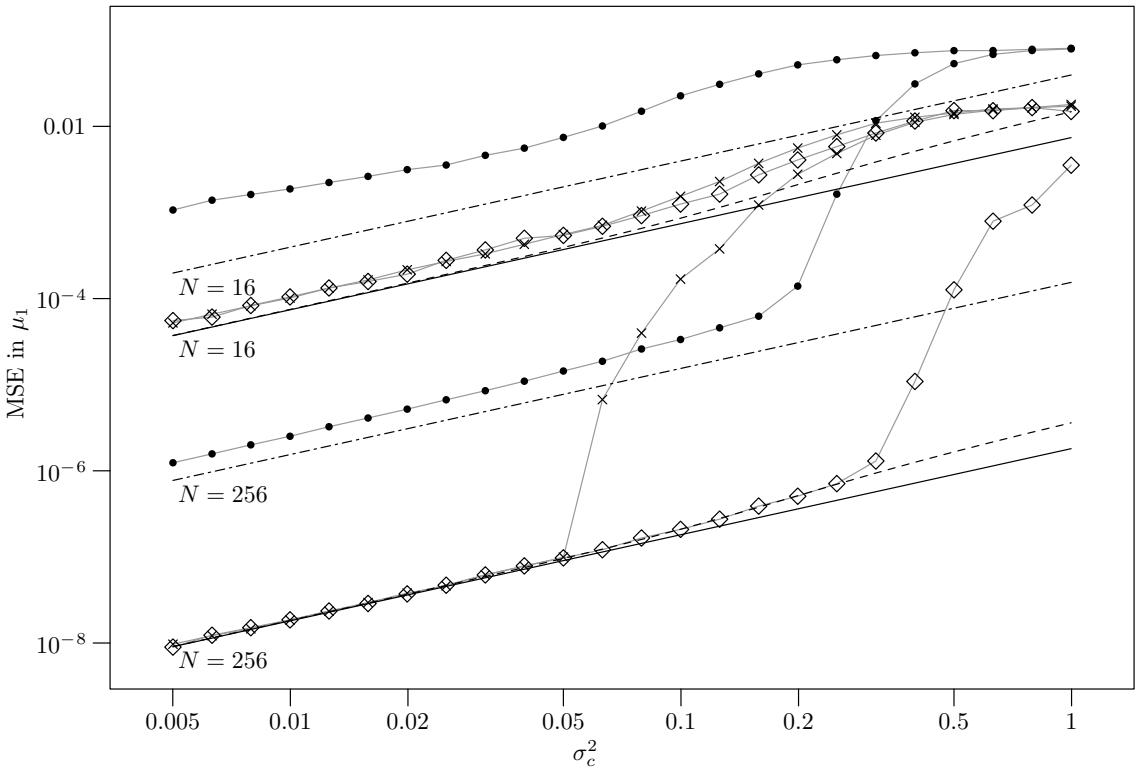


FIGURE 10.13: MSE in the frequency coefficient μ_1 versus $\operatorname{var} X_n = \sigma_c^2$. The DPT estimator runs at the higher sampling rate δ so that the volumes $V_{DPT}(\delta) = V_m$.

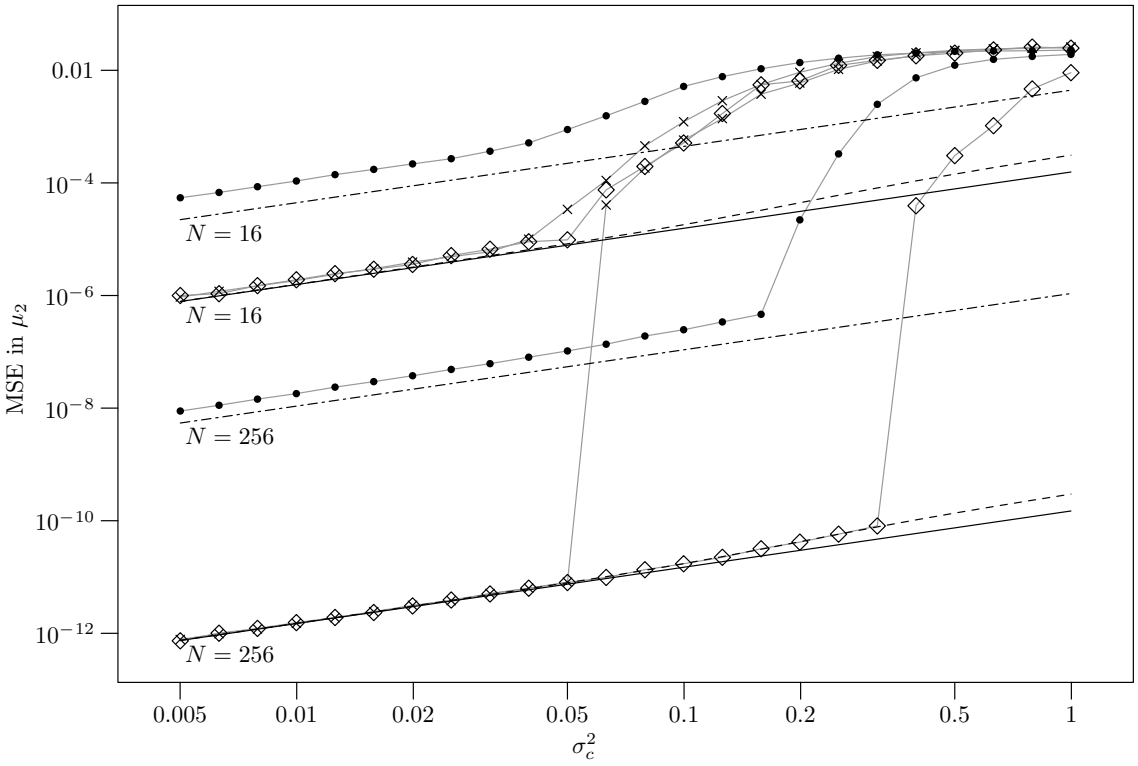


FIGURE 10.14: MSE in the second order coefficient μ_2 versus $\operatorname{var} X_n = \sigma_c^2$. The DPT estimator runs at the higher sampling rate δ so that the volumes $V_{DPT}(\delta) = V_m$.

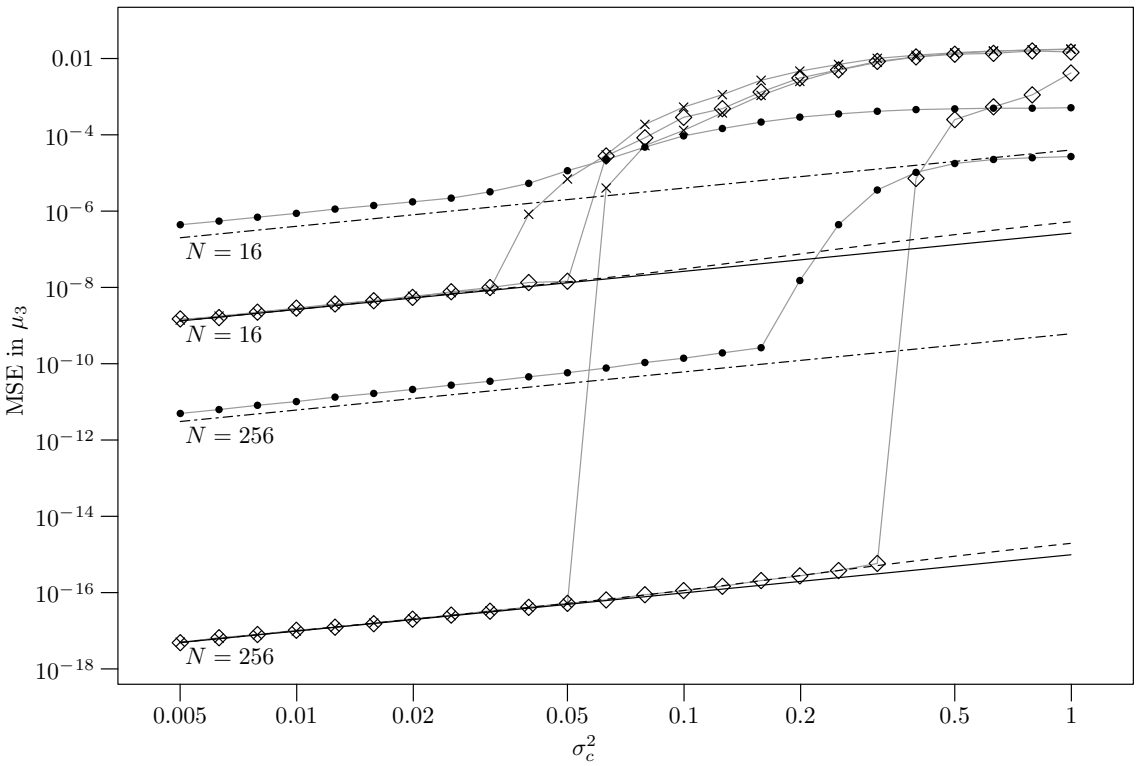


FIGURE 10.15: MSE in the third order coefficient μ_3 versus $\operatorname{var} X_n = \sigma_c^2$. The DPT estimator runs at the higher sampling rate δ so that the volumes $V_{DPT}(\delta) = V_m$.

Conclusion

In this thesis we have studied connections between two fields, **lattice theory** and **circular statistics**. We focused on the estimation and theoretical analysis of **polynomial phase signals**. These signals have a vast array of applications in science, in particular in astronomy, optics, biology, geology, geography and meteorology and also in engineering, particularly in communications and radar. Despite this, the theoretical tools for analysing these signals were found lacking. In particular the effect of **aliasing** that occurs when polynomial phase signals are sampled was not understood. We discovered some special **lattices**, called $V_{n/m}$, $V_{n/m}^*$ and $V_{n/m}^\perp$, that are particularly useful for studying polynomial phase signals. Using these lattices we completely described the aliasing, and also produced some remarkably accurate estimators. These results will be of great value to engineers, scientists and statisticians studying polynomial phase signals.

In Chapter 2 we described some introductory concepts from lattice theory. We focused on **tesselating regions**, the **Voronoi cell**, the **nearest lattice point problem**, **dual lattices**, **sublattices**, **quotient groups** and also the properties of lattices generated by intersection with or projection into a subspace.

In Chapter 3 we considered the lattices A_n, A_n^* and also a related family called the **Coxeter lattices**. We showed how A_n can be constructed as the intersection of the integer lattice $\mathbb{Z}^{n+1}$ with the subspace that is orthogonal to the all ones vector $\mathbf{1}$, and how A_n^* can be constructed as the projection of $\mathbb{Z}^{n+1}$ into this subspace. We described nearest point algorithms for these lattices that require only a linear number of operations in the dimension of the lattice n. The algorithms exploit the fact that the Voronoi cell of the lattice A_n is equivalent to the convex polytope that results from projecting the $n+1$ dimensional hypercube into the subspace that is orthogonal to $\mathbf{1}$. These new algorithms are the fastest known.

In Chapter 4 we derived a number of properties of the lattices $V_{n/m}$, $V_{n/m}^*$ and $V_{n/m}^\perp$. We showed how these lattices are generated as intersections and projections of the integer lattice $\mathbb{Z}^{n+m+1}$. We derived generator matrices for $V_{n/m}^\perp$ and $V_{n/m}^*$ and found closed-form formulas for the determinants and also the order of the dual quotient group $V_{n/m}^*/V_{n/m}$. Using these properties we found an algorithm to compute the nearest point in $V_{n/m}^*$ by computing a set of coset representatives for the dual quotient group. The nearest point algorithm requires $O(n^{(m+1)^2+1})$ operations in total which is polynomial in the dimension n but is exponential in the *projection* parameter m. This is an improvement over the fastest algorithms for *random* lattices, such as the sphere decoder, that require a number of operations that is exponential in the dimension.

In Chapter 5 we gave a brief overview of circular statistics. We described **circular random variables** and their **probability density functions** and showed how the standard definition of the mean in terms of the expected value does not map well to our intuitive notion of **mean direction**. To solve this we considered two different definitions, the **circular mean** and the **unwrapped mean**. Both of these means map well to our intuitive notion of mean direction. The two means are not in general equal and for some distributions they are not even defined. In Theorem 5.1 we described a large class of circular distribution that have equal unwrapped and circular means and we called these **unimean** distributions. We consider some popular circular distributions, the **von Mises distribution**, the **wrapped normal distribution**, the **wrapped uniform distribution** and the **projected normal distribution**. We described conditions under which these distributions are unimean.

Chapter 6 considered methods for estimating the circular and unwrapped means of a circular random variable from a number, say N, of observations. The first estimator considered is the **sample circular mean** estimator of the circular mean. Theorem 6.1 showed that the sample circular mean is strongly consistent and derived its central limit theorem. The second estimator is the **angular least squares estimator** of the unwrapped mean. We showed how this estimator can be rapidly computed by finding a nearest point in the lattice A_n^*. Theorem 6.2 showed that the angular least squares estimator is strongly consistent and that it satisfied a central limit theorem. We considered the performance of these estimators for some unimean distributions and found that the angular least squares estimator tends to perform better when the distribution is 'uniform-like' whereas the sample circular mean tends to perform better when the distribution is 'von Mises-like'. We also found that the performance of the estimators is very accurately modeled by the central limit theorems derived in Theorems 6.1 and 6.2.

In Sections 6.4, 6.5 and 6.6 we applied the estimators to the problems of **phase estimation**, **noncoherent detection**, and **delay estimation**. For phase estimation we found that it is better to simply use the standard least squares estimator. For noncoherent detection of PSK signals we found that highly accurate detection could be performed in practice using the angular least squares or the sample circular mean estimators. This approach is computationally attractive because it requires only a linear number of operations in the **block length**, whereas existing least squares approaches require a log-linear number of operations. We also considered the problem of delay estimation from noisy and incomplete data. It was observed that the angular least squares estimator or the sample circular mean could be used to produce very accurate estimates of the delay regardless of the amount of data that is missing. However, if the noise level is very high, then a significant accuracy penalty is paid for having incomplete data.

We also discussed some of the computational properties of the two estimators. We focused on the number of trigonometric operations that are required and found that if the N observations are complex numbers, such as in the problem of phase estimation, then the sample circular mean requires only a single arctangent operation, but the angular least squares estimator requires N arctangent operations. On the other hand, if the angles are observed directly, as is likely to be the case in

meteorology and other applications, then the sample circular mean requires $2N+1$ trigonometric operations, but the angular least squares estimator does not require any. If trigonometric operations are particularly expensive, as is typically the case on small computing devices, then consideration of these properties will likely lead to computational savings.

Estimating the mean direction of a circular random variable is equivalent to estimating the phase of a polynomial phase signal of order zero, otherwise called a **constant phase signal**. In Part III we generalised this concept to polynomial phase signals of arbitrary order. In Chapter 7 we considered the phenomenon of **aliasing** that occurs when polynomial phase signals are sampled uniformly and we described how the aliasing occurs using some ideas from lattice theory. For polynomial phase signals of order one, this aliasing effect is equivalent to the Nyquist criterion. In order to ensure the **identifiability** of any estimator the polynomial phase coefficients must be restricted to an **identifiable region**. We showed how an identifiable region is described as a tessellating region of the lattice L_m with generator matrix described using the **integer valued polynomials**. Using this lattice we showed how to resolve aliased parameters, compute square error and generate parameters uniformly in an identifiable region.

Chapter 8 considered the problem of estimating the coefficients of a polynomial phase signal from N observations. We derived the angular least squares estimator for the polynomial coefficients and showed how the estimator could be computed by finding a nearest lattice point in the lattice $V^*_{n/m}$. We derived the asymptotic properties of this estimator showing that it is strongly consistent and obtaining its central limit theorem. For the case of polynomials of order greater than one, the statistical results derived in this chapter are the first of their kind. We have proved these theoretical results under the assumption that the noise terms are identical and independent, but we discussed how these assumptions could potentially be relaxed.

Chapter 9 considered the special case of estimating the two coefficients of a polynomial signal of order one, otherwise known as **frequency estimation**. We described three estimators that exist in the literature, the **periodogram estimator**, the **Quinn-Fernandes estimator** and **Kay's unwrapping estimator**. We also considered the **angular least squares estimator**. We showed by Monte-Carlo simulation that the periodogram estimator, the Quinn-Fernandes estimator and the angular least squares estimator are all very accurate. The performance of the angular least squares estimator is well modeled by the central limit theorem derived in Theorem 8.1. Kay's unwrapping estimator is not as accurate as the other estimators.

The angular least square estimator requires a nearest point in the lattice $V^*_{N-2/1}$ to be computed. If we use the algorithm from Chapter 4 then $O(N^5)$ operations are required. This is very slow, so we described a simple method to approximate the nearest point in $O(N^2 \log N)$ arithmetic operations. Although much faster, the complexity is still high when compared with other frequency estimators. It may be that much faster nearest point algorithms exist for $V^*_{N-2/1}$. Considering the accuracy of this estimator, even fast *approximate* nearest point algorithms might prove useful for frequency estimation.

Chapter 10 described a number of techniques for estimating the $m+1$ coefficients of a polynomial phase signal of order m from N noisy observations. We first considered the least squares estimator and showed that it is very computationally intensive. We then described some computationally tractable approaches that exist in the literature: the **discrete polynomial phase transform** (DPT) [Peleg and Friedlander, 1995] and **Kitchen's unwrapping estimator** [Kitchen, 1994].

The DPT requires only $O(N \log N)$ operations and Kitchen's unwrapping estimator requires only $O(N)$ operations. However, both of these estimators only function correctly for a restricted range of coefficients inside the identifiable region. For the DPT estimator the problem is particularly severe and the range of coefficients is a very small fraction of the identifiable region and shrinks rapidly as the number of observations N increases. We considered how this problem might be overcome by increasing the sampling rate, but showed how this came with inevitable statistical penalties.

We then considered the **angular least squares estimator** that involves computing a nearest point in the lattice $V^*_{n/m}$. The polynomial time nearest point algorithm described in Chapter 4 is unfortunately too slow for practical use and therefore we considered some general purpose algorithms, the **sphere decoder**, **Babai's nearest plane algorithm** and the **K-best algorithm**. The sphere decoder and K-best algorithm are statistically very accurate with performance similar to the least squares estimator. We found that Babai's nearest plane algorithm is computationally efficient, requiring only $O(N^2)$ arithmetic operations, but its statistical performance is not as good. Unlike the DPT and Kitchen's unwrapping estimator the angular least squares estimator works uniformly well over the entire identifiable region. In cases where N is large and the noise variance is not too large, this property makes the approximate angular least squares estimator computed using Babai's nearest plane algorithm an attractive choice.

The algorithms that we used for computing the angular least squares estimator are slower than the DPT estimator and Kitchen's unwrapping estimator. However, there may exist fast (exact or approximate) nearest point algorithms for $V^*_{n/m}$ that we have not found yet. Considering the statistical superiority of the angular least squares estimator the search for faster nearest point algorithms for $V^*_{n/m}$ is a worthy direction for future research.

Bibliography

Abatzoglou, T. [1986]. Fast Maximum Likelihood Joint Estimation of Frequency and Frequency Rate. *IEEE Trans. Aerospace Elec. Systems*, 22(6), 708–715.

Agrell, E., Eriksson, T., Vardy, A. and Zeger, K. [2002]. Closest point search in lattices. *IEEE Trans. Inform. Theory*, 48(8), 2201–2214.

Ajtai, M. [1998]. The shortest vector problem in L_2 is NP-hard for randomized reductions. *in Proc. 30th ACM Symposium on Theory of Computing*, pp. 10–19.

Amemiya, T. [1985]. *Advanced econometrics*. Harvard University Press.

Anderson, J. and Mohan, S. [1984]. Sequential Coding Algorithms: A Survey and Cost Analysis. *IEEE Trans. Commun.*, 32(2), 169 – 176.

Andrews, D. W. K. [1987]. Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica*, 55(6), 1465–1471.

Ängeby, J. [2000a]. Aliasing of polynomial-phase signal parameters. *IEEE Trans. Sig. Process.*, 48(5), 1488–1491.

Ängeby, J. [2000b]. Estimating signal parameters using the nonlinear instantaneous least squares approach. *IEEE Trans. Sig. Process.*, 48, 2721–2732.

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

Babu, J. G. [1980]. An Inequality for Moments of Sums of Truncated ϕ-Mixing Random Variables and Its Applications. *Sankhyā*, 42(1/2), 1–8.

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

Barbero, L. G. and Thompson, J. S. [2008]. Fixing the Complexity of the Sphere Decoder for MIMO Detection. *IEEE Trans. Wireless Commun.*, 7(6), 2131 –2142.

Belov, N. V. [1965]. Theorem of the emptiness of the fundamental parallelepiped in the crystal lattice. *Journal of Structural Chemistry*, pp. 169–170.

Billingsley, P. [1979]. *Probability and measure.* Wiley.

Blum, M., Floyd, R. W., Pratt, V. R., Rivest, R. L. and Tarjan, R. E. [1973]. Time Bounds for Selection. *J. Comput. Syst. Sci.*, 7(4), 448–461.

Borwein, P. and Ingalls, C. [1994]. The Prouhet-Tarry-Escott problem revisited. *Enseign. Math.*, 40(4), 3–27.

Brillinger, D. [1962]. A note on the rate of convergence of mean. *Biometrika*, 49, 574–576.

Burger, T., Gritzmann, P. and Klee, V. [1996]. Polytope Projection and Projection Polytopes. *The American Mathematical Monthly*, 103(9), 742–755.

Cahen, P. J. and Chabert, J. L. [1997]. *Integer-valued Polynomials.* American Mathematical Society, Providence RI.

Chen, R., Koetter, R., Agrawal, D. and Madhow, U. [2003]. Joint demodulation and decoding for the noncoherent block fading channel: a practical framework for approaching Shannon capacity. *IEEE Trans. Commun.*, 51(10), 1676–1689.

Clarkson, I. V. L. [1999a]. An algorithm to compute a nearest point in the lattice A_n^*. In M. Fossorier, H. Imai, S. Lin and A. Poli, editors, *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, volume 1719 of *Lecture Notes in Computer Science*, pp. 104–120. Springer.

Clarkson, I. V. L. [1999b]. Frequency estimation, phase unwrapping and the nearest lattice point problem. *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, 3, 1609–1612.

Clarkson, I. V. L. [2008]. Approximate Maximum-Likelihood Period Estimation from Sparse, Noisy Timing Data. *IEEE Trans. Sig. Process.*, 56(5), 1779–1787.

Clarkson, I. V. L., Howard, S. D. and Mareels, I. M. Y. [1996]. Estimating the period of a pulse train from a set of sparse, noisy measurements. *Proc. Internat. Sympos. Signal Process. Appl.*, 2, 885–888.

Cohen, H. [1993]. *A course in computational algebraic number theory.* Springer Verlag.

Conway, J. H. [1997]. In *The sensual (quadratic) form*, number 26 in The Carcus mathematical monographs. The Mathematical Association of America.

Conway, J. H. and Sloane, N. J. A. [1982]. Fast quantizing and decoding and algorithms for lattice quantizers and codes. *IEEE Trans. Inform. Theory*, 28(2), 227–232.

Conway, J. H. and Sloane, N. J. A. [1986]. Soft decoding techniques for codes and lattices, including the Golay code and the Leech lattice. *IEEE Trans. Inform. Theory*, 32(1), 41–50.

Conway, J. H. and Sloane, N. J. A. [1992]. Low-Dimensional Lattices. VI. Voronoi Reduction of Three-Dimensional Lattices. *Proceedings: Mathematical and Physical Sciences*, 436(1896), 55–68.

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

Cormen, T. H., Leiserson, C. E., Rivest, R. L. and Stein, C. [2001]. *Introduction to Algorithms*. MIT Press. and McGraw-Hill, 2nd edition.

Cox, I., Miller, M. and Bloom, J. [2008]. *Digital watermarking and steganography*. Morgan Kaufmann.

Coxeter, H. S. M. [1951]. Extreme forms. *Canad. J. Math.*, 3, 391–441.

de Buda, R. [1989]. Some optimal codes have structure. *IEEE J. Sel. Areas Commun.*, 7(6), 893–899.

Dinur, I., Kindler, G. and Safras, S. [2003]. Approximating CVP to within almost-polynomial factors in NP-hard. *Combinatorica*, 23, 205–243.

Djuric, P. M. and Kay, S. M. [1990]. Parameter estimation of chirp signals. *IEEE Trans. Acoust. Speech Signal Process.*, 38(12), 2118–2126.

Eisinberg, A. and Fedele, G. [2007]. Discrete Orthogonal Polynomials on Equidistant Nodes. *International Mathematical Forum*, 2(21), 1007–1020.

Eisinberg, A., Pugliese, P. and Salerno, N. [2001]. Vandermonde matrices on integer nodes: the rectangular case. *Numerische Mathematik*, 87(4), 663–674.

El Gamal, H., Caire, G. and Damen, M. O. [2004]. Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels. *IEEE Trans. Inform. Theory*, 50(6), 968–985.

Erdös, P., Gruber, P. M. and Hammer, J. [1989]. Lattice Points. Wiley, New York.

Erez, U., Litsyn, S. and Zamir, R. [2005]. Lattices Which Are Good for (Almost) Everything. *IEEE Trans. Inform. Theory*, 51(10), 3401–3416.

Erez, U. and Zamir, R. [2004]. Achieving $1/2\log(1+SNR)$ on the AWGN channel with lattice encoding and decoding. *IEEE Trans. Inform. Theory*, 50(10), 2293–2314.

Farquharson, M. [2006]. *Estimating the parameters of polynomial phase signals*. Ph.D. thesis, Queensland University of Technology, Brisbane, Australia.

Farquharson, M., O'Shea, P. and Ledwich, G. [2005]. A computationally efficient technique for estimating the parameters of polynomial-phase signals from noisy observations. *IEEE Trans. Sig. Process.*, 53(8), 3337–3342.

Fisher, N. I. [1993]. *Statistical analysis of cirular data*. Cambridge University Press.

Floyd, R. W. and Rivest, R. L. [1975a]. The Algorithm SELECT - for Finding the ith Smallest of n Elements. *Commun. ACM*, 18(3), 173.

Floyd, R. W. and Rivest, R. L. [1975b]. Expected time bounds for selection. *Commun. ACM*, 18, 165–172.

Fogel, E. and Gavish, M. [1988]. Parameter estimation of quasi-periodic sequences. *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, 4, 2348–2351.

Fogel, E. and Gavish, M. [1989]. Performance evaluation of zero-crossing-based bit synchronizers. *IEEE Trans. Commun.*, 37(6), 663–665.

Gentry, C. [2009a]. *A fully homomorphic encryption scheme*. Ph.D. thesis, Stanford University.

Gentry, C. [2009b]. Fully homomorphic encryption using ideal lattices. *in Proc. 41th ACM Symposium on Theory of Computing*, pp. 169–178.

Golden, S. and Friedlander, B. [1998a]. Estimation and statistical analysis of exponential polynomial signals. *IEEE Trans. Sig. Process.*, 46(11), 3127–3130.

Golden, S. and Friedlander, B. [1998b]. A modification of the discrete polynomial transform. *IEEE Trans. Sig. Process.*, 46(5), 1452–1455.

Golden, S. and Friedlander, B. [1999]. Maximum likelihood estimation, analysis, and applications of exponential polynomial signals. *IEEE Trans. Sig. Process.*, 47(6), 1493–1501.

Goldreich, O., Goldwasser, S. and Halevi, S. [1997]. Public-Key Cryptosystems from Lattice Reduction Problems. *Lecture Notes in Computer Science*, 1294, 112–131.

Guo, Z. and Nilsson, P. [2006]. Algorithm and implementation of the K-best sphere decoding for MIMO detection. *IEEE J. Sel. Areas Commun.*, 24(3), 491 – 503.

Hammond, C. [2001]. *Basics of Crystallography and Diffraction*. Oxford.

Hannan, E. J. [1973]. The Estimation of Frequency. *Journal of Applied Probability*, 10(3), 510–519.

Hardy, G. H. and Wright, E. M. [2008]. *An introduction to the theory of numbers*. Oxford University Press, 6th edition.

Hassibi, A. and Boyd, S. P. [1998]. Integer parameter estimation in linear models with applications to GPS. *IEEE Trans. Sig. Process.*, 46(11), 2938–2952.

Hoadley, B. [1971]. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Stat.*, (6), 1977–1991.

Hochwald, B. M., Peel, C. B. and Swindlehurst, A. L. [2005]. A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation. *IEEE Trans. Commun.*, 53(3), 537–544.

Huxley, M. N. [1996]. *Area, lattice points, and exponential sums*. Oxford University Press, USA.

Jalden, J., Barbero, L. G., Ottersten, B. and Thompson, J. S. [2009]. The Error Probability of the Fixed-Complexity Sphere Decoder. *IEEE Trans. Sig. Process.*, 57(7), 2711 –2720.

Jalden, J. and Ottersten, B. [2005]. On the complexity of sphere decoding in digital communications. *IEEE Trans. Sig. Process.*, 53(4), 1474–1484.

Jenrich, R. I. [1969]. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Stat.*, 40(2), 633–643.

Jordan, C. [1965]. *Calculus of finite differences*. Chelsea Publishing Company, New York, N.Y.

Kannan, R. [1987]. Minkowski's Convex Body Theorem and Integer Programming. *Math. Operations Research*, 12(3), 415–440.

Kannan, R. and Bachem, A. [1979]. Polynomial Algorithms for Computing the Smith and Hermite Normal Forms of an Integer Matrix. *SIAM Journal on Computing*, 8(4), 499–507.

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

Kitchen, J. [1994]. A method for estimating the coefficients of a polynomial phase signal. *Signal Processing*, 37(1), 463–470.

Knuth, D. E. [1997]. *The Art of Computer Programming*, volume 2, Seminumerical Algorithms. Addison-Wesley, Reading, Ma., 3rd edition.

Knuth, D. E. [1998]. *The Art of Computer Programming*, volume 3, Sorting and Searching. Addison-Wesley, Reading, Ma., 3rd edition.

Lenstra, A. K., Lenstra, H. W. and Lovász, L. [1982]. Factoring polynomials with rational coefficients. *Math. Ann.*, 261, 515–534.

Levanon, N. and Mozeson, E. [2004]. *Radar Signals*. Wiley.

Lévy, R. [1939]. L'addition des variables aléatoires définies sur une circonference. *Bull. Soc. Math.*, 67, 1–41.

Liu, J., Kim, J., Kwatra, S. C. and Stevens, G. H. [1991]. An analysis of the MPSK scheme with differential recursive detection (DRD). *Proc. IEEE Conf. on Veh. Tech.*, pp. 741–746.

Loeliger, H. A. [1997]. Averaging bounds for lattices and linear codes. *IEEE Trans. Inform. Theory*, 43(6), 1767–1773.

Love, D. J., Jr., R. W. H., Santipach, W. and Honig, M. L. [2004]. What is the value of limited feedback for MIMO channels? *IEEE J. Sel. Areas Commun.*, 42(10), 54–59.

Lovell, B. C., Kootsookos, P. J. and Williamson, R. C. [1991]. The circular nature of discrete-time frequency estimates. *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, 5, 3369–3372.

Mackenthun, K. [1994]. A fast algorithm for multiple-symbol differential detection of MPSK. *IEEE Trans. Commun.*, 42, 1471–1474.

Makrakis, D. and Feher, K. [1990]. Optimal noncoherent detection of PSK signals. *Electronics Letters*, 26(6), 398–400.

Mardia, K. V. and Jupp, P. [2000]. *Directional Statistics*. Wiley, 2nd edition.

Martinet, J. [2003]. *Perfect lattices in Euclidean spaces*. Springer.

McKilliam, R. G. and Clarkson, I. V. L. [2008]. Maximum-likelihood period estimation from sparse, noisy timing data. *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, pp. 3697–3700.

McKilliam, R. G. and Clarkson, I. V. L. [2009]. Identifiability and aliasing in polynomial-phase signals. *IEEE Trans. Sig. Process.*, 57(11), 4554–4557.

McKilliam, R. G., Clarkson, I. V. L. and Quinn, B. G. [2008a]. An Algorithm to Compute the Nearest Point in the Lattice A_n^*. *IEEE Trans. Inform. Theory*, 54(9), 4378–4381.

McKilliam, R. G., Clarkson, I. V. L., Quinn, B. G. and Moran, B. [2009a]. Polynomial-phase estimation, phase unwrapping and the nearest lattice point problem. *Asilomar Conference on Signals, Systems, and Computers*, pp. 493–495.

McKilliam, R. G., Clarkson, I. V. L., Ryan, D. J. and Collings, I. B. [2009b]. Linear-time block noncoherent detection of PSK. *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, pp. 2465–2468.

McKilliam, R. G., Clarkson, I. V. L., Smith, W. D. and Quinn, B. G. [2008b]. A linear-time nearest point algorithm for the lattice A_n^*. *International Symposium on Information Theory and its Applications*.

McKilliam, R. G., Quinn, B. G., Clarkson, I. V. L. and Moran, B. [2010a]. Frequency Estimation by Phase Unwrapping. *IEEE Trans. Sig. Process.*, 58(6), 2953–2963.

McKilliam, R. G., Ryan, D. J., Clarkson, I. V. L. and Collings, I. B. [2008c]. An improved algorithm for optimal noncoherent QAM detection. *Proc. Australian Commun. Theory Workshop*, pp. 64–68.

McKilliam, R. G., Ryan, D. J., Clarkson, I. V. L. and Collings, I. B. [2010b]. Block Noncoherent Detection of Hexagonal QAM. *in Proc. Australian Commun. Theory Workshop*.

McKilliam, R. G., Smith, W. D. and Clarkson, I. V. L. [2010c]. Linear-time nearest point algorithms for Coxeter lattices. *IEEE Trans. Inform. Theory*, 56(3), 1015–1022.

Micciancio, D. [2001]. The hardness of the closest vector problem with preprocessing. *IEEE Trans. Inform. Theory*, 47(3), 1212–1215.

Micciancio, D. and Voulgaris, P. [2009]. A Deterministic Single Exponential Time Algorithm for Most Lattice Problems based on Voronoi Cell Computations (Extended Abstract). *42nd ACM Symposium on Theory of Computing*.

Micciancio, D. and Warinschi, B. [2001]. A linear space algorithm for computing the Hermite Normal Form. *Proceedings of the 2001 International Symposium on Symbolic and Algebraic Computation*, pp. 231–236.

Minkowski, H. [1910]. *Geometrie der zahlen*. BG Teubner Leipzig.

Morelande, M. [2009]. Parameter Estimation of Phase-Modulated Signals Using Bayesian Unwrapping. *IEEE Trans. Sig. Process.*, 57(11), 4209–4219.

Morelande, M. and Zoubir, A. [2002]. On the performance of cyclic moments-based parameter estimators of amplitude modulated polynomial phase signals. *Signal Processing, IEEE Transactions on*, 50(3), 590 –606.

Morelande, M. R. [2008]. Circular regression using Bayesian unwrapping. *Proc. Internat. Conf. Acoust. Spe. Sig. Process.*, pp. 3441–3444.

Newey, W. K. and McFadden, D. [1994]. Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pp. 2111–2245. Elsevier.

O'Shea, P. J. [1996]. An Iterative Algorithm For Estimating The Parameters Of Polynomial Phase Signals. *Proc. Internat. Sympos. Signal Process. Appl.*, 2, 730–731.

Peel, C. B., Hochwald, B. M. and Swindlehurst, A. L. [2005]. A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication Part I: Channel Inversion and Regularization. *IEEE Trans. Commun.*, 53(1), 195–202.

Peleg, S. and Friedlander, B. [1995]. The discrete polynomial-phase transform. *IEEE Trans. Sig. Process.*, 43(8), 1901–1914.

Peleg, S. and Porat, B. [1991]. The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase. *IEEE Trans. Sig. Process.*, 39(3), 749–752.

Pohst, M. [1981]. On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications. *SIGSAM Bull.*, 15(1), 37–44.

Poltyrev, G. [1994]. On coding without restrictions for the AWGN channel. *IEEE Trans. Inform. Theory*, 40(2), 409–417.

Pötscher, B. M. and Prucha, I. R. [1989]. A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes. *Econometrica*, 57(3), 675–683.

Quinn, B. [1999]. A fast efficient technique for the estimation of frequency: interpretation and generalisation. *Biometrika*, 86(1), 213.

Quinn, B. G. [2000]. On Kay's Frequency Estimator. *Journal of Time Series Analysis*, 21(6), 707–712.

Quinn, B. G. [2007]. Estimating the mode of a phase distribution. *Asilomar Conference on Signals, Systems and Computers*, pp. 587–591.

Quinn, B. G. [2008]. Recent advances in rapid frequency estimation. *Digital Signal Processing*.

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

Quinn, B. G. and Fernandes, J. M. [1991]. A fast efficient technique for the estimation of frequency. *Biometrika*, 78(3), 489–497.

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

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

Rice, J. A. and Rosenblatt, M. [1988]. On Frequency Estimation. *Biometrika*, 75(3), 477–484.

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

Ryan, D. J., Clarkson, I. V. L. and Collings, I. B. [2007a]. GLRT-Optimal Noncoherent Lattice Decoding. *IEEE Trans. Sig. Process.*, 55, 3773–3786.

Ryan, D. J., Clarkson, I. V. L., Collings, I. B., Guo, D. and Honig, M. L. [2007b]. QAM Codebooks for Low-Complexity Limited Feedback MIMO Beamforming. *IEEE Int. Conf. Commun.*, pp. 4162–4167.

Ryan, D. J., Clarkson, I. V. L., Collings, I. B. and Heath Jr., R. W. [2008]. Performance of vector perturbation multiuser MIMO systems with limited feedback. Accepted for *IEEE Trans. Commun.*

Ryshkov, S. S. and Baranovskii, E. P. [1979]. Classical methods in the theory of lattice packings. *Russian Mathematical Surveys*, 34(4), 1.

Sage, A. P. and Melsa, J. L. [1971]. *Estimation theory with applications to communications and control*. McGraw-Hill, New York.

Schnorr, C. P. and Euchner, M. [1993]. Lattice Basis Reduction: Improved Practical Algorithms and Solving Subset Sum Problems. In *Math. Programming*, pp. 181–191.

Secord, N. and de Buda, R. [1989]. Demodulation of a Gosset Lattice Code Having a Spectral Null at DC. *IEEE Trans. Inform. Theory*, 35(2), 472–477.

Sidiropoulos, N. D., Swami, A. and Sadler, B. M. [2005]. Quasi-ML period estimation from incomplete timing data. *IEEE Trans. Signal Process.*, 53, 733–739.

Sikirić, M. D., Schürmann, A. and Vallentin, F. [2008]. A generalistation of Voronoi's reduction theory and its application. *Duke Mathematical Journal*, 142(1), 127–164.

Stadje, W. [1984]. Wrapped Distributions and Measurement Errors. *Metrika,*, 31, 303–317.

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

Sweldens, W. [2001]. Fast block noncoherent decoding. *IEEE Comms. Letters*, 5(4), 132–134.

Szego, G. [1975]. *Orthogonal Polynomials*. Amer. Math. Soc., 4th edition.

Telatar, I. et al. [1999]. Capacity of multi-antenna Gaussian channels. *European transactions on telecommunications*, 10(6), 585–596.

Teunissen, P. J. G. [1995]. The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. *Journal of Geodesy*, 70, 65–82.

Teunissen, P. J. G. [2006]. The LAMBDA method for the GNSS compass. *Artificial Satellites*, 41(3), 89–103.

Thomas, J. A., Moss, C. F., Vater, M. and Moore, P. W. [2005]. Echolocation in bat and dolphins. *The Journal of the Acoustical Society of America*, 118, 2755.

Tretter, S. A. [1985]. Estimating the Frequency of a Noisy Sinusoid by Linear Regression. *IEEE Trans. Inform. Theory*, 31(6), 832–835.

van der Vaart, A. W. [1998]. *Asymptotic Statistics*. Cambridge University Press.

Vardy, A. and Be'ery, Y. [1993]. Maximum likelihood decoding of the Leech lattice. *IEEE Trans. Inform. Theory*, 39(4), 1435–1444.

Viterbo, E. and Boutros, J. [1999]. A universal lattice code decoder for fading channels. *IEEE Trans. Inform. Theory*, 45(5), 1639–1642.

Voronoi, G. [1908]. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Journal für die reine und angewandte Mathematik*, pp. 97–178.

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

Wang, P., Djuvorič, I. and Yang, J. [2008]. Generalized high-order phase function for parameter estimation of polynomial phase signal. *IEEE Trans. Inform. Theory*, 56(7), 3023–3028.

Warrier, D. and Madhow, U. [2002]. Spectrally efficient noncoherent communication. *IEEE Trans. Inform. Theory*, 48(3), 651–668.

Weber, W. J. [1978]. Differential encoding for multiple amplitude and phase shift keying systems. *IEEE Trans. Commun.*, 26, 385–391.

Wiley, R. G. [1982]. *Electronic Intelligence: The Analysis of Radar Signals*. Artech House, Norwood, Massachusetts.

Wilson, S. G., Freebersyser, J. and Marshall, C. [1989]. Multi-symbol detection of MPSK. *Proc. IEEE GLOBECOM*, pp. 1692–1697.

Wintner, A. [1947]. On the shape of the angular case of Cauchy's distribution curves. *Ann. Math. Statist.*, 18, 589–593.

Yokoyama, R. [1980]. Moment bounds for stationary mixing sequences. *Probability Theory and Related Fields*, 52, 45–57.