Lattices and Dual Lattices in Optimal Experimental Design for Fourier Models

R.A. Bates, University of Warwick
E. Riccomagno, University of Warwick
R. Schwabe, Free University, Berlin
H.P. Wynn, University of Warwick

Abstract

Number theoretic lattices, used in integration theory, are studied from the viewpoint of the design and analysis of experiments. For certain Fourier regression models lattices are optimal as experimental designs because they produce orthogonal information matrices. This generalises early results by Hoel (1965). When the Fourier model is restricted, that is a special subset of the full factorial (cross–spectral) model is used, there is a difficult inversion problem to find generators for an optimal design for the given model. Asymptotic results are derived for certain models as the dimension of the space goes to infinity. These can be thought of as a complexity theory connecting designs and models or as special type of Nyquist sampling theory.

Keywords: experimental design, integer lattices, Nyquist sampling, complexity

1 Introduction

Integer lattices have been used extensively in numerical integration. The books by Niederreiter (1992), Conway and Sloane (1992), Sloan and Joe (1994) and Fang and Wang (1994) are basic references. Periodic, and indeed Fourier models are the natural models to be associated with the methods because of the periodic nature of the lattices themselves. In this paper we investigate lattices as experimental designs for fitting Fourier models. It will be seen that as with classical polynomial models orthogonality plays an important role and yields D–optimum designs in the sense of Kiefer and Wolfowitz. Orthogonality will be related to the so–called "dual lattice" and there is a strong connection with classical designs based on defining subgroups and aliasing.

The challenge of the theory, as we see it, is to solve "inversion" problems. That is, given a particular model find the lattice which as an experimental design allows orthogonal estimation of all terms in the model, without aliasing. The paper follows a very brief mentioning of the main idea in Bates, Buck, Riccomagno and Wynn (1996) and emphasis on the experimnetal design background in Riccomagno, Schwabe and Wynn (1995).

1.1 Fourier model

Let A^+ be any set of d–dimensional integer vectors not containing 0 and such that $h \in A^+$ implies $-h \notin A^+$. Then let $A = A^+ \cup \{-A^+\} \cup \{0\}$. Let $\#(A^+) = m$ so that $\#(A) = 2m+1$. Although the results have general applicability we shall refer to the vectors h as *frequencies* to help motivation. The set A defines the following trigonometric regression model

$$\begin{aligned} E(Y(x)) = \mu(x) &= \theta_0 + \sqrt{2} \sum_{h \in A^+} \left[\theta_h \sin(2\pi h^t x) + \phi_h \cos(2\pi h^t x)\right] \\ &= \alpha_0 + \sum_{h \in A^+} \left[\beta_h e^{2\pi i h^t x} + \delta_h e^{-2\pi i h^t x}\right] \end{aligned}$$

where θ_h, ϕ_h and β_h, δ_h are related by

$$\begin{aligned} \alpha_0 &= \theta_0 \\ \beta_h &= \frac{\phi_h - i\theta_h}{\sqrt{2}} \\ \delta_h &= \frac{\phi_h + i\theta_h}{\sqrt{2}} \end{aligned}$$

The two forms of the model, real and complex, can be written in matrix form

$$\mu(x) = X(x)\ \theta = Z(x)\ \alpha$$

where

$$\begin{aligned} \theta &= \{\theta_0 : \{\theta_h\}_{h\in A^+} : \{\phi_h\}_{h \in A^+}\}^t \\ X(x) &= \left\{1 : \{\sqrt{2}\sin(2\pi h^t x)\}_{h\in A^+} : \{\sqrt{2}\cos(2\pi h^t x)\}_{h\in A^+}\right\} \\ \alpha &= \{\alpha_0 : \{\beta_h\}_{h\in A^+} : \{\delta_h\}_{h\in A^+}\}^t \\ Z(x) &= \left\{1 : \{e^{2\pi i h^t x}\}_{h\in A^+} : \{e^{-2\pi i h^t x}\}_{h\in A^+}\right\}. \end{aligned}$$

The equation relating α and θ is

$$\alpha = H\theta$$

where

$$H = \left(\begin{array}{c|c|c} 1 & 0 & 0 \\ \hline 0 & -\frac{i}{\sqrt{2}} I & \frac{1}{\sqrt{2}} I \\ \hline 0 & \frac{i}{\sqrt{2}} I & \frac{1}{\sqrt{2}} I \end{array}\right)$$

where I is the $m \times m$ identity matrix and H is an unitary matrix: $H^*H = I$ equivalently $H^{-1} = H^*$, where H^* is the transpose and conjugate matrix of H.

1.2 Orthogonality conditions

The set of x_i values $D = \{x_1, \ldots, x_N\}$ is called here the *experimental design* and we are interested in the design–model pair (D, A). Reparametrizing the vector of the values of A in D we have for the vector of observations Y

$$E(Y) = X\theta = XH^{-1}H\theta = XH^{-1}\alpha = Z\alpha$$

where the ith row of X is $X(x_i)$ and

$$Z = XH^{-1}$$

thus Z is the design matrix from the model in complex form. It is easily verified that the ith row of Z is $Z(x_i)$. A statistical version of the model which until now has only been expressed in expectation form is

$$Y(x) = \mu(x) + \epsilon$$

where ϵ are uncorrelated errors with zero mean and equal variance, $\sigma^2 < \infty$. Then the best linear unbiased estimator for θ, the usual least squares estimator, assuming X has full rank, is

$$\hat{\theta} = \left(X^tX\right)^{-1} X^tY$$

Since $Z^* = (XH^{-1})^* = HX^t$, the corresponding estimator for $\hat{\alpha}$ is given by

$$\begin{aligned}
\hat{\alpha} &= (Z^*Z)^{-1} Z^*Y \\
&= \left(HX^tXH^{-1}\right)^{-1} HX^tY \\
&= H\left(X^tX\right)^{-1} X^tY \\
&= H\hat{\theta},
\end{aligned}$$

as expected from standard theory.

This paper is critically concerned with the choice of the experimental design D for a given model A. The theory of optimal experimental design can be used to show that D–optimality namely achieving $\max|X^tX|$ over all designs with a given sample size N is achieved for the models described here when $X^tX = NI$, namely for *orthogonal* designs. This is an extension of early work by Hoel (1965). See also Kiefer and Wolfowitz (1959). The D–optimality property of lattices was pointed out by Bates, Buck, Riccomagno and Wynn (1996). Riccomagno, Schwabe and Wynn (1995) present the full D–optimality theory. For the models of type A, above, namely which contain h and $-h$, orthogonality is equivalent to D-optimality and several other types of optimality. The full details appear in Riccomagno, Schwabe and Wynn (1995). Here we concentrate on the orthogonality.

Definition 1 *A design with N points is called completely orthogonal if $X^tX = NI$ where X is the design matrix.*

Lemma 1 *For a design model pair (D, A)*

$$X^tX = NI \text{ if and only if } Z^*Z = NI$$

Proof. From the definitions $Z^*Z = HX^tXH^{-1}$ and $X^tX = H^{-1}Z^*ZH$. $\square$

Now Z^*Z has a block structure

$$Z^*Z = \left(\begin{array}{c|c|c} N & a_1 & b_1 \\ \hline a_2 & A & C \\ \hline b_2 & D & B \end{array}\right)$$

where for all $h_1.h_2 \in A^+$

$$\begin{aligned} a_{1_{h_1}} = b_{2_{h_1}} &= \sum_{x\in D} e^{2\pi j h_1^t x} \\ a_{2_{h_1}} = b_{1_{h_1}} &= \sum_{x\in D} e^{-2\pi j h_1^t x} \\ A_{h_1,h_2} &= \sum_{x\in D} e^{2\pi j(-h_1+h_2)^t x} \\ C_{h_1,h_2} &= \sum_{x\in D} e^{2\pi j(-h_1-h_2)^t x} \\ D_{h_1,h_2} &= \sum_{x\in D} e^{2\pi j(+h_1+h_2)^t x} \\ B_{h_1,h_2} &= \sum_{x\in D} e^{2\pi j(+h_1-h_2)^t x} \end{aligned} \tag{1}$$

and $a_1 = a_2^* = b_2^t = \bar{b}_1$, $A = B^*$ and $C = D^*$.

The conditions for complete orthogonality are, then, $a_1^t = a_2^t = b_1 = b_2 = 0$ and $C = D = 0$ and $A = B = NI$. On the diagonal, that is when $h_1 = h_2$, the conditions are met since $A_{h_1,h_1} = B_{h_1,h_1} = 1$. Outside the diagonal they have to be derived from the conditions on the design. These conditions are discussed in the next section.

2 Experimental design

The designs we investigate are those given by so–called integration lattices. These are used extensively in algebraic number theory and numerical integration. One can consider this paper as an investigation of the role of lattices in functional estimatio n or approximation rather than integration.

Definition 2 *An infinite lattice L_∞ in $\mathbb{R}^d$ is a discrete subset of $\mathbb{R}^d$ closed under summation and subtraction. An infinite lattice is called an integration infinite lattice if it includes $\mathbb{Z}^d$. We define a (finite) lattice $L := L_\infty \cap [0,1[^d$.*

In group theory an infinite lattice is an infinite Abelian group. A finite Abelian group, L, is also formed under the operation $x_1 \circ x_2 = \{x_1 + x_2\}$, where $\{x\}$ is the point in $[0,1[^d$ whose coordinates are the fractional part of the coordinates of x. The order (number of elements) of a lattice will be denoted by N. We work with a particular class of integration infinite lattices formed by the so–called *good lattice points* or *one–generator lattices*, defined as follows: let $g = (g_1, \ldots, g_d) \in \mathbb{Z}^d$ and $N \in \mathbb{Z}_+$, with the constraint that each component of g is prime with respect to N, then

$$L_\infty = \left\{ \left\{ \frac{kg}{N} \right\} \oplus \mathbb{Z}^d : k \in \{0, \ldots, N-1\} \right\}$$

where $\oplus$ denotes the direct sum. The integer vector g is called the generator of the lattice and N is the order of $L_\infty \cap [0,1[^d$. We write the corresponding lattice in a more useful way as

$$\begin{aligned} L &= \left\{ \left\{ \frac{kg}{N} \right\} : k = 0, \ldots, N-1 \right\} \\ &= \left\{ \left(\frac{kg_1 \bmod N}{N}, \ldots, \frac{kg_d \bmod N}{N} \right) ; \; k = 0, \ldots, N-1 \right\} \end{aligned}$$

which can be visualised as points on a line wrapped around $[0,1[^d$ interpreted as the d–dimensional torus. We shall write $L_{g,N} = L$ for a one-generator lattice with generator g and order N.

A main observation of this paper is that the orthogonality conditions can be expressed in terms of the so–called *dual lattice* of the lattice L.

Definition 3 *The dual lattice of a finite or infinite lattice L is defined as*

$$L^\perp = \left\{ h \in \mathbb{Z}^d \text{ such that } h^t x \in \mathbb{Z} \quad \text{for all } x \in L \right\}$$

Lemma 2 *If $L = L_{g,N}$ is a one–generator lattice, then*

$$L^\perp_{g,N} = \left\{ h \in \mathbb{Z}^d \text{ such that } h^t g \equiv 0 \pmod{N} \right\}$$

Proof. For a one–generator lattice $h \in L^\perp_{g,N}$ if (and only if) $h^t kg \equiv 0 \pmod{N}$ for all $k \in \mathbb{Z}$. Obviously, the latter is equivalent to $h^t g \equiv 0 \pmod{N}$. □

Notice that $L^\perp$ is itself a lattice and in particular if $h \in L^\perp$ then $sh \in L^\perp$ for any integer s. In engineering terms $L^\perp$ contains all higher order frequencies (harmonics) of any member.

The key to the theory is the next lemma whose proof can be found in Sloan and Joe (1994, Lemma 2.1) and in Niederreiter (1992, Lemma 5.2.1) based on elementary properties of group characters.

Lemma 3 *Let L be an integration lattice with N points in the unit hypercube $[0,1)^d$, then*

$$\frac{1}{N}\sum_{x\in L} e^{2\pi i h^t\cdot x} = \begin{cases} 1 & \text{if } h \in L^\perp \\ 0 & \text{otherwise} \end{cases}$$

This lemma immediately gives conditions for complete orthogonality which we present in the following theorem. The proof is immediate from (1). Note that the complete orthogonality is more easily expressed via Z following Lemma 1.

Theorem 1 *For $D \equiv L$, a one–generator lattice in $[0,1[^d$ with generator g, the Z–matrix satisfies $Z^*Z = NI$, complete orthogonality, if and only if*
(i) $\pm h_1 \notin L^\perp$
(ii) $\pm(h_1 - h_2) \notin L^\perp$ with $h_1 \neq h_2$
(iii) $\pm(h_1 + h_2) \notin L^\perp$
for all h_1, h_2 frequencies in the Fourier regression model A^+; if and only if (i) $(h_1 - h_2) \notin L^\perp$ for all $h_1, h_2 \in A$ with $h_1 \neq h_2$.*

Note that (iii) requires in particular $\pm 2h_1 \notin L^\perp$, which is in turn implied by $h_1 \notin L^\perp$ and $h_1 - h_2 \notin L^\perp$ if N is odd. To avoid additional technicalities we will assume N odd. Then (iii) simplifies to $h_1 + h_2 \notin L^\perp$ for $h_1 \neq h_2$. Or equivalently, $Z^*Z = NI$ if and only if for all $h, h' \in A$ the relation $h + h' \in L^\perp$ or $h - h' \in L^\perp$ implies $h = h'$ or $h = -h'$.

Since if L contains an element then it also contains its additive inverse. In practice, for all $h_1, h_2 \in A^+$, one need only check

$$\begin{aligned} & h_1 \notin L^\perp \\ & (h_1 - h_2) \notin L^\perp \ (h_1 \neq h_2) \\ & (h_1 + h_2) \notin L^\perp \end{aligned}$$

and the last relation only for $h_1 \neq h_2$ in case N is odd.

It is convenient to express the orthogonality as a pairwise property between different frequencies h_1, h_2.

Definition 4 *A pair of frequencies (h_1, h_2) with $h_1, h_2 \notin L^\perp$ is called mutually orthogonal if $h_1, h_2 \notin L^\perp$ and $\pm h_1 \pm h_2 \notin L^\perp$.*

Note that the above definition implies $h_1 \neq h_2$ for orthogonal h_1 and h_2. The following lemma explains the importance of the above definition. Its proof is straightforward by Lemma 3.

Lemma 4 *If every pair $h_1, h_2 \in A^+$ is mutually orthogonal with respect to $L^\perp$ then L is completely orthogonal for the model $A = A^+ \cup (-A^+) \cup \{0\}$.*

2.1 Cosets

There is a necessary and sufficient condition for a pair (h_1, h_2) to be mutually orthogonal which is similar to the condition of not being aliased in factorial experimental design.

It is of historic interest that the term alias occurs both in the time series/signal processing literature and in the experimental design literature. That this is true in the one–dimensional case is well–known. We believe that the present extension to lattice designs is new. For readers acquainted with the classical development of the alias table in experimental design the following development in terms of cosets will be familiar. Note however how the Fourier model expressed as $A^+ \cup \{-A^+\}$ gives a double coset interpretation of aliasing. We need to recall some elementary group theory.

Let $(G, +)$ be an Abelian group, U a subgroup of G and a be in G. The set $K_a = K - a := \{-a + u : u \in U\} K_a = K - a := \{-a + u : u \in U\}$ is called the ath-coset of G with respect to U. More technically the equivalence relation on G defined by $a \sim b$ if $-a + b \in U$ generates the quotient G/U whose elements, the residue classes, are the cosets of G with respect to U. The cosets form a partition of G. It turns out that the orthogonality conditions in Theorem 1 are related to pairs of cosets $K_a \cup K_{-a}$. We denote by $L_h^\perp$ the h–coset of $L^\perp$.

Theorem 2 *The pair of frequencies h_1 and h_2 is mutually orthogonal if and only if there exist $h \notin L^\perp$ and $h' \notin L^\perp$ such that $h_1 \in L_h^\perp \cup L_{-h}^\perp$ and $h_2 \in L_{h'}^\perp \cup L_{-h'}^\perp$ and the two double cosets are disjoint.*

Proof. Let h_1 and h_2 be mutually orthogonal. Assume that h_1 and h_2 are in the same double coset $L_h^\perp \cup L_{-h}^\perp$. Without loss of generality $h_1 \in L_h^\perp$ and, hence, $h_1 - h \in L^\perp$. Then either $h_2 - h \in L^\perp$ or $h_2 + h \in L^\perp$ and, consequently, $h_1 - h_2 \in L^\perp$ or $h_1 + h_2 \in L^\perp$ in contradiction to the asumption which proves the direct part.

Conversely, if h_1 and h_2 are not mutually orthogonal, then either $h_1 - h_2 \in L^\perp$ or $h_1 + h_2 \in L^\perp$. Hence, $h_1 \in L_{h_2}^\perp \cup L_{-h_2}^\perp$ which shows that h_1 and h_2 are in the same double coset. $\square$

Note that since our models always contain the constant term θ_0 (α_0) we are only considering $h_1, h_2 \notin L^\perp$ since any $h \in L^\perp$ will be aliased with the constant term. The theory can clearly be adjusted to the case with a constant term missing.

We exhibit the partition on $\mathbb{Z}^d$ defined by the cosets of the dual lattice of a one–generator lattice design L with N points, N odd. Since we choose N mutually prime with all the g_i for all $i = 1, \ldots, d$, then without loss of generality we may assume $g_1 = 1$. Thus the lattice is

$$L = \left\{ \frac{k(1, g_2, \ldots, g_d) \pmod N}{N} : \quad k = 0, \ldots, N-1 \right\}$$

and its dual is

$$L^\perp = \left\{ (a_1, \ldots, a_d) \in \mathbb{Z}^d : \quad (a_1, \ldots, a_d)^t (1, \ldots, g_d) \equiv 0 \pmod N \right\}$$

The $(h_1,\ldots,h_d)$–coset is

$$L^{\perp}_{(h_1,\ldots,h_d)} = \left\{(a_1-h_1,\ldots,a_d-h_d) \in \mathbb{Z}^d : \quad (a_1,\ldots,a_d) \in L^{\perp}\right\}$$

It turns out that there are N cosets forming a partition of $\mathbb{Z}^d$. They are enumerated as

$$L^{\perp}_{(0,\ldots,0)}, L^{\perp}_{(1,\ldots,0)}, \ldots, L^{\perp}_{(N-1,\ldots,0)}$$

For notational convenience we write $L^{\perp}_k = L^{\perp}_{(k,\ldots,0)}$ for $k = 0,\ldots,N-1$ and $L^{\perp}_0 = L^{\perp}$.

Theorem 3 *For the lattice L, $L^{\perp}_{(h_1,\ldots,h_d)} = L^{\perp}_k$ if and only if $h_1+g_2h_2+\ldots+g_dh_d \equiv k \pmod N$*

Proof. Suppose $h \in L^{\perp}_k$. We have to prove that if $h \in L^{\perp}_{(h_1,\ldots,h_d)}$ then $h_1+g_2h_2+\ldots+g_dh_d \equiv k \pmod N$. Then h may be written as

$$\begin{aligned} h &= (a_1-k, a_2, \ldots, a_d) \qquad (a_1,\ldots,a_d) \in L^{\perp} \\ &= (b_1-h_1,\ldots,b_d-h_d) \quad (b_1,\ldots,b_d) \in L^{\perp} \end{aligned}$$

This implies

$$(a_1,a_2,\ldots,a_d) = (b_1-h_1+k, b_2-h_2,\ldots,b_d-h_d) \in L^{\perp}$$

This condition expands to

$$\begin{aligned} (b_1-h_1+k)+g_2(b_2-h_2)+\ldots+g_d(b_d-h_d) &\equiv 0 \pmod N \\ -(h_1+h_2g_2+\ldots+g_dh_d)+k &\equiv 0 \pmod N \end{aligned}$$

that is $(h_1+h_2g_2+\ldots+g_dh_d) \equiv k \pmod N$, which proves the assertion. Similar arguments prove that $h \in L^{\perp}_k$. $\square$

Since N is odd there are $\frac{N-1}{2}$ double cosets plus $L^{\perp}$ itself, specifically

$$L^{\perp}_0, L^{\perp}_1 \cup L^{\perp}_{N-1}, \ldots, L^{\perp}_k \cup L^{\perp}_{N-k}, \ldots, L^{\perp}_{\frac{N-1}{2}} \cup L^{\perp}_{\frac{N+1}{2}} \qquad (k=2,\ldots,\tfrac{N-3}{2})$$

This fact is of course clear from the underlying group theory.

Theorem 2 can be reformulated as h and h' are mutually orthogonal if and only if

$$h_1+g_2h_2+\ldots+g_dh_d \not\equiv \pm(h'_1+g_2h'_2+\ldots+g_dh'_d) \pmod N$$

Because of Lemma 1 if there are any two pairs of frequencies which are *not* mutually orthogonal Z (and X) becomes singular that is the model as a whole is not identifiable (a single off diagonal 1 makes Z^tZ singular). Thus models can have at most $\frac{N-1}{2}$ terms and for every chosen $k=0,\ldots,\frac{N-1}{2}$ at most one model–frequency satisfies the equation

$$h_{1k}+g_2h_{2k}+\ldots+g_dh_{dk} \equiv k \pmod N$$

It is not surprising that there are at most $\frac{N-1}{2}$ terms in A^+ because there are at most N parameters.

As an example consider a two dimensional lattice, L, generated by $(1,5)$ and with $N = 13$ points. The design $D = L$ is completely orthogonal for the model defined by $A = (1,0),(0,1),(1,1),(1,-1)$ because its frequencies generate respectively

$$\begin{array}{ll} L^{\perp}_{(1,0)} \cup L^{\perp}_{(12,0)} & (\ 1 + 5 \cdot 0 \equiv 1 \pmod{13}) \\ L^{\perp}_{(5,0)} \cup L^{\perp}_{(8,0)} & (\ 0 + 5 \equiv 5 \pmod{13}) \\ L^{\perp}_{(6,0)} \cup L^{\perp}_{(7,0)} & (\ 1 + 5 \equiv 6 \pmod{13}) \\ L^{\perp}_{(4,0)} \cup L^{\perp}_{(9,0)} & (\ 1 - 5 \equiv -4 = 9 \pmod{13}) \end{array}$$

It is a useful exercise to find other identifiable models for this design.

The lattice design generated by $(1,5)$ and $N = 23$ is suitable to identify the main effects for the model $\{(1,0),(2,0),(0,1),(0,2),(1,\pm1),(1,\pm2),(2,\pm1),(2,\pm2)\}$, but not the interactions since the frequencies $(1,2)$ and $(2,2)$ are aliased, being both in $L^{\perp}_{11} \cup L^{\perp}_{-11}$.

3 Finding good generators: inversion

The previous section can be considered as a foundation for a more difficult problem: given a model, specified by a set of frequencies A, find the generator $g = (g_1, \ldots, g_d)$ and N for which the orthogonality conditions of Theorem 1 hold. In a moment we shall classify particular models of interest. It will then work out that certain models, for example without high–order cross frequencies can lead to designs (generators) for which N only grows polynomially or even linearly in the dimension d. This points to a complexity theory for the design model pair (A, D). Let us put this more precisely: what is the smallest sample size N^* such that there exists a generator g^* with $D = L_{g^*,N^*}$ satisfying the orthogonality conditions of Theorem 1? In many cases it appears that N^* is hard to find explicitly and we shall find $L_{g,N}$ with a suitable upper bound $N \geq N^*$.

Let A be in a particular class of models and let $K(A)$ be some measure of the size of A such as the total number of frequencies in A, or the order of the model. Considering N^* as a function of A and namely: $N^*(A)$, so that $N(A) \geq N^*(A)$. A plot of $N^*(A)$ against $K(A)$ gives a frontier below which no orthogonal design exists for A. Typically as $K(A)$ increases $N^*(A)$ will increase. Thus the broad objective in terms of complexity is to find this frontier. For example in the case of no cross spectra we shall find a linear boundary. Finally as the dimension tends to infinity for a fixed class of models we can express $K(A)$ in terms of d or present (N^*, d) plots or (N, d) plots for an upper bound $N \geq N^*$.

Since N^* represents a broad measure of the sampling density and $K(A)$ will typically measure quantities like the largest fitted frequency or the number of distinct frequencies, the result can be thought of as a kind of Nyquist–Shannon sampling theory.

3.1 Linear models

In this section we consider models with no interactions that is cross spectral terms but including all the marginal frequencies up to a certain order. The set A for the so–referred additive or linear models is formed by points on the coordinate axis, and

$$A^+ = \left\{ \begin{array}{ll} (h_1, 0, 0, \ldots, 0): & h_1 = 1, \ldots, m_1 \\ (0, h_2, 0, \ldots, 0): & h_2 = 1, \ldots, m_2 \\ \ldots & \\ (0, \ldots, 0, 0, h_d): & h_d = 1, \ldots, m_d \end{array} \right\}$$

We use the notation $F(d; m_1, \ldots, m_d; 1)$ to refer to this model.

According to Theorem 3 we look for $g = (g_1, \ldots, g_d)$ and N such that

$$0, \quad g_k h_k \quad : \quad h_k = \pm 1, \ldots, \pm m_k \quad k = 1, \ldots, d$$

are all different $\pmod N$.

The recursive scheme

$$\begin{aligned} g_1 &= 1 \\ g_d &= m_{d-1} g_{d-1} + 1 \end{aligned}$$

and $N = 2m_d g_d + 1$ gives an orthogonal design but the size N increases exponentially with the dimension d. In fact for $m_1 = \ldots = m_d = m$ we have

$$g_d = \frac{m^d - 1}{m - 1} \quad \text{and} \quad N = 2m\frac{m^d - 1}{m - 1} + 1$$

To improve upon this arrange the model factors in such a way that $m_1 \geq m_2 \geq \ldots \geq m_d$. Define the recursive procedure

$$\begin{aligned} g_1 &= 1 \\ g_d &= (d-1)m_1 + 1 \end{aligned}$$

and $N = 2m_1^2(d-1) + 1 \geq 2\max_{k=1,\ldots,d}\{g_k m_k\} + 1$ gives a design whose size N increases linearly with the dimension when

the orders m_k remain bounded. Note that a new rearrangement of the factors has to be made each time a new factor is added.

Let us comment again on the choice of N and its link to the Nyquist frequency as in Fourier Analysis. The Nyquist frequency in discrete Fourier Analysis is the highest frequency f^* used in Fourier transform. If Δt is the sampling interval in time, then $2\pi f^* = \frac{\pi}{\Delta t}$. For all the frequencies $f < f^*$ the components with frequencies $f^* + k$ ($k \in \mathbb{Z}$) are all aliased with each other and the sum of their contributions will appear collected, at frequency f. This is sometimes refered to as "Nyquist folding". To avoid this the frequencies under study must be less than the Nyquist frequency, equivalently the sampling interval will be chosen to make this happen. In other terms we require $2\pi f < \frac{\pi}{\Delta t}$. In the case of N equally spaced points Δt is $\frac{1}{N}$ we have $2\pi f < \pi N$ i.e. $2f + 1 \leq N$.

In the additive case presented here the Nyquist frequency of the one dimensional case is replaced by $\max_{k=1,\ldots,d}\{g_k m_k\}$. This is also a suitable $K(A)$ in the discussion in Section 3. We shall return to the general question of how certain aliased frequency contributions collect at other model terms in Section 4.

3.2 Two–factor interaction models

In this section we study the model $F(d; 2, \ldots, 2; 2)$ i.e. a two–factor interaction model with marginal frequencies equal to $0, 1$ or 2. The frequency set is

$$A^+ = \left\{ \begin{array}{lllll} (k,0,\ldots,0) & (0,k,\ldots,0) & \ldots & (0,\ldots,k,0) & (0,\ldots,k) \\ (k,h,\ldots,0) & (0,k,h,\ldots,0) & \ldots & (0,\ldots,k,h) & \\ \ldots & & & & \\ (k,0,\ldots,h) & (0,k,\ldots,h) & \ldots & & : k,h = 1,2 \end{array} \right\}$$

The lattice design generated by the sequence

$$\begin{aligned} g_1 &= 1 \\ g_d &= g_{d-1} + 4\alpha_d \\ \alpha_1 &= 1 \\ \alpha_{2^j} &= 3\alpha_{2^{j-1}} - 1 \\ \alpha_k &= \alpha_{2^j} \quad \text{for } k = 2^j q,\ q \text{ odd} \end{aligned}$$

with $N = 2g_d + 2g_{d-1} + 1$ allows a model with all main effects for each factor (dimension) orthogonally to all second order interaction: a resolution IV design in experimental design. Note however that the interaction can be aliased. For example $h_1 = (1, 2, 0, \ldots, 0)$ and $h_2 = (-2, 0, 1, 0, \ldots, 0)$ are in the same double coset $L_{11}^{\perp}$ since $g_1 + 2g_2 = g_3 - 2g_1 = 11$.

The given sequence of generators starts as

$$(1, 5, 13, 17, 37, 41, 53, 109, 113, 121, 125, 145, 157, 161, 325, 329)$$

and the α–sequence has a self–similar structure as shown by

$$\begin{array}{l} 1,2,1,5,1,2,1,14,1,2,1,5,1,2,1,41, \\ 1,2,1,5,1,2,1,14,1,2,1,5,1,2,1,122 \\ 1,2,1,5,1,2,1,14,1,2,1,5,1,2,1,41, \\ 1,2,1,5,1,2,1,14,1,2,1,5,1,2,1,365... \end{array}$$

Standard methods for solving recurrence relations yield the solutions $\alpha_{2^j} = \frac{1}{2}(3^j + 1)$ and $g_{2^j+1} = 4\ 3^j + 1$. This implies that g_d, and N, increases like d^γ where $\gamma = \frac{\log 3}{\log 2} \approx 1.5858$.

To study the whole parameter set of $F(d; 2, \ldots, 2; 2)$ we suggest the recursive sequence of generators

$$\begin{aligned} g_1 &= 1 \\ g_2 &= 5 \\ g_d &= 4g_{d-1} + 2g_{d-2} + 1 \end{aligned}$$

i. e. $g_d = \frac{1}{10}(\sqrt{6}(2+\sqrt{6})^d - \sqrt{6}(2-\sqrt{6})^d - 2) \approx \frac{\sqrt{6}}{10}(2+\sqrt{6})^d$ with $N = 4(g_d + g_{d-1}) + 1 \approx \frac{6}{5}(2+\sqrt{6})^d$.

Another computer–generated sequence up to dimension 14 is

$$(1, 5, 23, 60, 77, 173, 222, 409, 535, 634, 935, 1182, 1361, 1497)$$

and the corresponding sample sizes for $d \geq 2$

$$25, 113, 333, 549, 1001, 1581, 2525, 4677, 6277, 8469, 10173, 11533.$$

These sample sizes are obtained by the formula $4(g_d + g_{d-1}) + 1$ where $2(g_d + g_{d-1})$ is the highest frequency in the model. Again we see the important Nyquist–like role played by the highest frequencies in driving up the sample size.

4 The fitted model

In the usual regression setup, assuming the original model is correct, we can estimate $\mu(x)$ at a particular x by the unbiased estimator

$$\hat{\mu}(x) = X(x)\hat{\theta} = Z(x)\hat{\alpha}$$

Under the orthogonality conditions, the variance of estimation under the model A is easily seen to be

$$Var\,(\hat{\mu}(x)) = E\,(\hat{\mu}(x) - \mu(x))^2 = \frac{\sigma^2}{n}\#(A)$$

where σ^2 is the error variance. Under the Kiefer–Wolfowitz theory max $Var(\hat{\mu}(x))$ over $[0,1]^d$ is also minimised: G-optimality. The integrated variance is

$$I_1 = \int_{[0,1)^d} \#(A)\frac{\sigma^2}{n}\, dx = \#(A)\frac{\sigma^2}{n}$$

It is instructive to consider the consequences of taking the observations as arising as if the underlying model was given by $\mu(x) = f(x)$ for a wide choice of $f(x)$. Thus suppose that $f(x)$ is expressible in a standard Fourier representation

$$f(x) = \sum_{h \in \mathbb{Z}^d} \alpha(h) e^{2\pi i h^t x} \qquad (2)$$

where $\alpha(h) = \int_{[0,1]^d} f(x) e^{-2\pi i h^t x}\, dx$. Sufficient conditions for uniform convergence are well known and omitted.

Under suitable conditions the set $\{e^{2\pi i h^t \cdot x} : h \in \mathbb{Z}^d\}$ forms an orthogonal basis of $L^2(\mathbb{R}^d)$ and through the Fourier transform there is a one–to–one mapping from $l^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$. We may thus fix an order on $\mathbb{Z}^d$ (the same as in Section 1) and write the Fourier decomposition of $f(x)$ as the scalar product of the infinite complex vectors

$$\begin{aligned} Z_\infty(x) &= \left(1, \ldots, e^{2\pi i h^t x}, \ldots, e^{-2\pi i h^t x}, \ldots\right)^t \\ \alpha_\infty &= (\alpha(0), \ldots, \alpha(h), \ldots, \alpha(-h), \ldots)^t \end{aligned}$$

and

$$f(x) = Z_\infty(x)^t \alpha_\infty$$

Suppose we have a Fourier model A and a one–generator lattice design $D = L$ with N points is chosen such that (D, A) is completely orthogonal. If we observe without error the observation vector is

$$Y = (f(x_1), \ldots, f(x_N))^t = Z_\infty \alpha_\infty$$

where the kth row of Z_∞ is $Z_\infty(x_k)^t$. Then, using the orthogonality, the least squared estimator is

$$\hat{\alpha} = (Z^* Z)^{-1} Z^* Y = \frac{1}{N} Z^* Y = \frac{1}{N} Z^* Z_\infty \alpha_\infty$$

Then

$$\begin{aligned} \frac{1}{N} Z^* Z_\infty &= \frac{1}{N} \left\{ \sum_{x \in L} e^{2\pi i(-h+k)^t x} \right\}_{h \in A, k \in \mathbb{Z}^d} \\ &= \{\delta_{h,k}\}_{h \in A \cup \{0\}, k \in \mathbb{Z}^d} \end{aligned}$$

where

$$\delta_{h,k} = \begin{cases} 1 & \text{if } -h + k \in L^\perp \\ 0 & \text{otherwise} \end{cases}$$

or equivalently

$$\delta_{h,k} = \begin{cases} 1 & \text{if } k \in L_h^\perp \\ 0 & \text{otherwise} \end{cases}$$

This analysis generalises the Nyquist folding to lattices. Thus the fitted model, when there is no error, is

$$\hat{\mu}(x) = \sum_{h \in A} \left(\sum_{k \in L_h^\perp} \alpha(k) \right) e^{2\pi i x^t h}$$

It is instructive also to give the integrated squared error. Employing the L^2 orthogonality of the Fourier terms over $[0,1]^d$ we have

$$I_2 = \int_{[0,1]^d} \left(\hat{Y}(x) - f(x)\right)^2 dx = \sum_{h \in A} \left(\sum_{k \in L_h^{\perp};\ k \neq h} \alpha(k) \right)^2 + \sum_{k \in Z^d - A} \alpha(k)^2$$

If there is observation error then I_2 represents the bias term and the full integrated mean square error is

$$IMSE = I_1 + I_2$$

It is clear from (4) that I_2 is unbounded unless further conditions are added to the Fourier coefficients $\alpha(k)$ over and above L^2–conditions: $\sum_{k \in \mathbb{Z}} \alpha(k)^2 < \infty$. A natural condition arising is

$$\sum_{k \in \mathbb{Z}} |\alpha(k)| < \infty$$

which is equivalent to absolute convergence of $f(x)$. Stronger conditions still are the $E_\alpha(c)$ conditions used in the integration theory of lattices (see Niederreiter, 1992, Definition 5.1).

Optimum design theory can be used to show that in addition to D–optimality for the Fourier model the orthogonal designs described here also minimize the integrated variance under the regression model. Further research is needed to find designs which guard against growth of the bias term I_2. In such work the relationship between the design $D = L$ and the source of biases as given by (4) should play a major part and there is a trade off between I_1 and I_2.

5 Discussion

This paper can be considered as the theoretical backgrounds for building Fourier models on integer lattices with special reference to the design model pair (D,A). Models built on space filling designs are becoming popular in building emulators, or surrogates, of large and expensive–to–run simulation codes in com puter aided design and similar areas. This body of research started with the work by Sacks et al. (1989). A recent paper on this is Bates, Riccomagno, Schwabe and Wynn (1995), having first been suggested in Bates, Buck, Riccomagno and Wynn (1996). A restriction of the Fourier models in the case where the vector x is some multidimensional set of inputs is the assumption of periodicity. For example $\mu(0,\ldots,0) = \mu(1,\ldots,1)$ so that both x–values are clearly not included. One can try to avoid this restriction by enlarging, shifting, or randomising the start of the design. Also there are methods for periodising non–periodic function, see Sloan and Joe (1994). The authors' preliminary conclusions from the practical use of these methods is that they are valuable, at least, for *screening*, that is detection of significant factors very much in the style of screening designs based on orthogonal arrays which have become popular in "Taguchi" methodology in Robust Engineering Design.

6 Acknowledgements

The first, second and fourth authors acknowledge the support of the UK Engineering and Physical Science Research Council. The research of the third author was supported by grants Ku719/2-1 and Ku719/2-2 of the Deutsche Forschungsgemeinschaft. The authors thank the British Council and the Deutscher Akademischer Austauschdienst for funding their joint ARC project which supports this collaboration.

References

[1] Bates, R.A., Buck, R.J., Riccomagno, E. and Wynn, H.P. Experimental design and observation for large systems. J. Royal Statistical Society, Series B, **58** (1996) 76–94.

[2] Bates, R.A., Riccomagno, E., Schwabe, R. and Wynn, H.P. Fast Fourier emulators based on space–filling lattices, Improceedings of ICIAM95.

[3] Conway, J.H. and Sloane, N.J.A. *Sphere Packings, Lattices and Groups.* Springer–Verlag, New York 1992.

[4] Fang, W.T. and Wang, Y. *Number–theoretic Methods in Statistics.* Chapman & Hall (London), 1994.

[5] Kiefer, J.C. and Wolfowitz, J. Optimum design in regression problem. Ann. Math. Statist., **30** (1959), 271–294.

[6] Hoel, P.G. Minimax distance designs in two dimensional regression. Ann. Math. Stat., **36** (1965), 1097–1106.

[7] Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-NFS, SIAM (Philadelphia), 1992.

[8] Riccomagno, E., Schwabe, R. and Wynn, H.P. Lattice–based optimum design for Fourier regression. Annals of Statistics, 1995 (submitted).

[9] Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. Design and analysis of computer experiments. Statistical Science, **4** (1989), 409–435.

[10] Sloan, I.H. and Joe, S. *Lattice Methods for Multiple Integration.* Oxford Science Publications, Oxford, 1994.