R.A. Bates, E. Riccomagno, R. Schwabe and H.P. Wynn

**Fast Fourier Emulators based on space–filling Lattices**

*Integer lattices are well–known to have optimal space–filling characteristics from number theory and numerical integration. Single generator lattices are studied as an alternative to a full grid. They are used to fit emulators to high dimensional functions, for example arising from C.A.E. (Computer Aided Engineering) or other modelling situations. Fourier models are fitted using Fast Fourier Transform and the power spectrum used for fast detection of important factors.*

# 1 Introduction

Complex systems can be very expensive to model, either because of their high dimensionality or a heavy computational or experimental cost. Nowadays a common strategy is to model a complex system with an emulator or surrogate performing experiments on the system directly. In the statistical literature this has been referred to as DACE: the Design and Analysis of Computer Experiments, for a reference see Sacks, Welch, Mitchell and Wynn (1989). In the electrical and C.A.E. literature it is referred to as the method of "surrogates", see Yeşilyurt and Patera (1995).

In this paper we force the multi–dimensionality of the input into one-dimension by using a suitably chosen one generator integer lattice, see Fang and Wang (1994), Sloan and Joe (1994) and Niederreiter (1992). The mathematical model we then fit is a Fourier regression. It has been shown [7] that there always exist lattice experimental designs which are $D$–optimum in the sense of Kiefer and Wolfowitz (1959) for such a model. The $D$–optimum criterion involves maximising the determinant of the information matrix from a standard least squares regression fit.

We wish to investigate which input variables or process parameters, and up to what order, can be identified to build an emulator consisting of a submodel of some larger Fourier model. To identify these parameters (including higher order terms) we work in a one-dimensional frequency domain using the lattice structure and apply the Fast Fourier Transform (FFT [10]). We then return to the original parameters using the inverse of the mapping which brought us into the one-dimensional frequency domain.

# 2 Procedure

The full regression model in exponential notation is

$$E(Y(x_1,...,x_d)) = \mu(x_1,...,x_d) = d_0 + \sqrt{2}\sum_{\ell=1}^{M}\sum_{k_1<\ldots<k_\ell}\sum_{r_{k_1}=1}^{m}\cdots\sum_{r_{k_\ell}=1}^{m}\sum_{\alpha\in A_\ell}$$
$$\left(d^{\alpha}_{+(k_1\ldots k_\ell,r_{k_1}\ldots r_{k_\ell})}e^{2\pi i(\alpha_1 r_{k_1}x_{k_1}+\ldots+\alpha_\ell r_{k_\ell}x_{k_\ell})} + d^{\alpha}_{-(k_1\ldots k_\ell,r_{k_1}\ldots r_{k_\ell})}e^{-2\pi i(\alpha_1 r_{k_1}x_{k_1}+\ldots+\alpha_\ell r_{k_\ell}x_{k_\ell})}\right)$$

where $x_i \in [0,1[$ for all $i \in \{1,\ldots,d\}$ and $A_\ell = \left\{\alpha \in \{-1,1\}^\ell; \alpha_1 = 1\right\}$ is the set of all $\ell$–dimensional multi–indices from $\{-1,1\}$ with unit first entries. In the notation introduced in [7] this is a $F(d;m,\ldots,m;M)$ where $d$ is the dimension of the input space, the order of the marginals is $m$ and we are including only interactions up to the $M$th order. The number of coefficients in the model is

$$1 + 2m\sum_{k=1}^{d} 1 + (2m)^2\frac{d(d-1)}{2} + \cdots + (2m)^M\left(\sum_{j_1=1}^{d-M+1}\cdots\sum_{j_M=j_{M-1}+1}^{d} 1\right)$$

For Fourier models of the above kind the product of $d$ univariate $D$–optimal designs is $D$–optimum (see Hoel, 1965). Also for other criteria such as $A$–, $E$– and $IMSE$–(integrated mean square error) the product design is optimum. However the size of simple grid–based designs may be very high compared with the number of parameters required to be estimated. To overcome this difficulty in [7] a theory of optimal orthogonal fractions was developed using integer lattice designs whose sizes increase polynomially with the dimension, using special algorithms based both on algebraic generation and more direct search. Here we consider a particular class of lattice designs algebraically generated.

The lattice design generated by $(g_1, \ldots, g_d)$ and of size $n$ is defined as the uniform design which support points lie on the set

$$\left\{\left(\frac{kg_1 \text{mod } n}{n}, ..., \frac{kg_d \text{mod } n}{n}\right); \, k = 0, ..., n-1\right\}$$

Due to the periodic structure of these designs we were able to infer their optimality for periodic models.

It is shown in [7] that for the Fourier regression model with 2 factor interactions ($M = 2$) the design generated by the following recursively–built generator is $D$-optimum for making inference on the main parameters. If the dimension of the model is $d$ the iterated sequence for the generator vector is

$$\begin{aligned} g_1 &= 1, \\ g_d &= g_{d-1} + 2m\alpha_{d-1}, \quad d \geq 2; \end{aligned}$$

where

$$\begin{aligned} \alpha_1 &= 1, \\ \alpha_{2^j} &= (m+1)\alpha_{2^{j-1}} - 1, \quad j \geq 2 \\ \alpha_d &= \alpha_{2^j}, \qquad\qquad \text{for } d = 2^j q, \quad q \text{ odd} \end{aligned}$$

When the order of the marginals is 2, the $\alpha$'s or difference sequence starts as

$$\begin{aligned} &1, 2, 1, 5, 1, 2, 1, 14, 1, 2, 1, 5, 1, 2, 1, 41, \\ &1, 2, 1, 5, 1, 2, 1, 14, 1, 2, 1, 5, 1, 2, 1, 122, \\ &1, 2, 1, 5, 1, 2, 1, 14, 1, 2, 1, 5, 1, 2, 1, 41, \\ &1, 2, 1, 5, 1, 2, 1, 14, 1, 2, 1, 5, 1, 2, 1, 365, ... \end{aligned}$$

which has a *self similar* (*fractal*) structure. The solution is $\alpha_{2^j} = \frac{(m-1)(m+1)^j+1}{m}$ and $g_{2^j} = 2m(m+1)^j + 1$. This implies that $g_d$ increases like $d^\gamma$ where $\gamma = \frac{\log(m+1)}{\log 2}$. If $m = 2$ then $\gamma = \frac{\log 3}{\log 2} \approx 1.5850$.

The corresponding design sample size is $2mg_d + mg_{d-1} + 1$ which increases polynomially with $d^\gamma$ as the dimension $d$ increases. For $m = 2$ the generator sequence is

$$(1, 5, 13, 17, 37, 41, 49, 53, 109, 113, 121, 125, 145, 149, 157, 161, 325, 329, ...)$$

and the size sequence is

$$(*, 23, 63, 95, 183, 239, 279, 311, 543, 671, 711, 743, 831, 887, 927, 959, 1623, 1967, ...)$$

It can be easily shown that because of the diagonal structure of the information matrix the lattice designs are also optimum with respect to any other convex criterion based on the eigenvalues of the information matrix itself such as $A$–, $E$– and integrated mean squared error criteria.

# 3 FFT for analysis

We clarify the use of the FFT tool with an example. Suppose we want to fit the additive model $F(2; 1, 1; 1)$ to a real process evaluated at design points which form a $D$–optimum set for the fuller complete model $F(2; 1, 1; 2)$. We write the full model in complex form

$$
\begin{aligned}
E(Y(x_1,x_2)) = d_0 \quad &+ \quad \sqrt{2}d_1 e^{i2\pi x_1} + \sqrt{2}d_{-1}e^{-i2\pi x_1} + \sqrt{2}d_2 e^{i2\pi x_2} + \sqrt{2}d_{-2}e^{-i2\pi x_2} \\
&+ \quad \sqrt{2}d_3 e^{i2\pi(x_1+x_2)} + \sqrt{2}d_{-3}e^{-i2\pi(x_1+x_2)} + \sqrt{2}d_4 e^{i2\pi(x_1-x_2)} + \sqrt{2}d_{-4}e^{-i2\pi(x_1-x_2)}
\end{aligned}
$$

Consider the design generated by $(1,5)$ and with 13–points, which, as shown in [7], is $D$–optimum for making inference on the main coefficients $(d_{-2}, d_{-1}, d_0, d_1, d_2)$. The lattice points for $t = \{0,\ldots,12\}$ are

$$
\begin{array}{lcccccccccccccc}
x_{1_t} & : & \frac{0}{13} & \frac{1}{13} & \frac{2}{13} & \frac{3}{13} & \frac{4}{13} & \frac{5}{13} & \frac{6}{13} & \frac{7}{13} & \frac{8}{13} & \frac{9}{13} & \frac{10}{13} & \frac{11}{13} & \frac{12}{13} \\
x_{2_t} & : & \frac{0}{13} & \frac{5}{13} & \frac{10}{13} & \frac{2}{13} & \frac{7}{13} & \frac{12}{13} & \frac{4}{13} & \frac{9}{13} & \frac{1}{13} & \frac{6}{13} & \frac{11}{13} & \frac{3}{13} & \frac{8}{13}
\end{array}
$$

Now we make 13 observations $Y_t$'s at the lattice design points, where

$$
\begin{array}{lcccccccccccccc}
Y_t & : & 1 & 2 & 3 & 4 & 4 & 1 & 1 & 1 & 2 & 3 & 3 & 3 & 2
\end{array}
$$

These values are plotted in Figure 1.

Then applying ordinary Fourier transform method we may compute the numerical values of $c_i$'s in the orthogonal trigonometric representation (Fourier series) of the $Y_t$, $t = 0,\ldots,12$

$$
\begin{aligned}
Y_t = c_0 \quad &+ \quad c_1 e^{i\frac{2\pi}{13}(1)t} + c_{-1}e^{-i\frac{2\pi}{13}t} + c_2 e^{i\frac{2\pi}{13}(2)t} + c_{-2}e^{-i\frac{2\pi}{13}2t} \\
&+ \quad c_3 e^{i\frac{2\pi}{13}(3)t} + c_{-3}e^{-i\frac{2\pi}{13}3t} + c_4 e^{i\frac{2\pi}{13}(4)t} + c_{-4}e^{-i\frac{2\pi}{13}4t} \\
&+ \quad c_5 e^{i\frac{2\pi}{13}(5)t} + c_{-5}e^{-i\frac{2\pi}{13}5t} + c_6 e^{i\frac{2\pi}{13}(6)t} + c_{-6}e^{-i\frac{2\pi}{13}6t} \qquad (1)
\end{aligned}
$$

For completeness we repeat the following well–known facts from Fourier and Spectral Analysis. The power of a finite sequence $\{Y_t\}$ is defined as

$$\sum Y_t^2 = n \sum |c_t|^2$$

When the Fourier series of the observed values is written in a real form, let $\theta_k$ and $\phi_k$ be the coefficients of the cosine and sine terms respectively. Then the value of the power spectrum $f_k$ at the $k$–frequency is defined as

$$
\begin{cases}
f_0 = |c_0|^2 = \theta_0^2 \\
f_k = 2|c_k|^2 = \frac{1}{2}\left(\theta_k^2 + \phi_k^2\right) \quad \text{for} \quad k = 1,\ldots,\left[\frac{n}{2}\right]
\end{cases}
$$

Hence the plot of $f_k$ against $k$ describes how the total power is distributed over the various frequency components of the observed sequence. This corresponds to an analysis of variance in statistics.

For the example,we restrict the input parameters of the model $F(2;1,1;2)$ to the "line" $\{(s, 5s(\text{mod } 1)) : s \in \mathbb{R}\}$ on which the lattices lies. Visualised on the plane $[0,1[^2$ this "line" is a step function but on the torus defined by the standard isomorphism from $[0,1[^2$ to $S_1 \times S_1$ it is a spiral. We take parametric coordinates on this spiral normalised to 13 and denote it by $s$. Thus for $s = [0,13[$ the $F(2;1,1;2)$ model becomes

$$
\begin{aligned}
E(Y(x_1(s),x_2(s))) = d_0 \quad &+ \quad \sqrt{2}d_1 e^{i\frac{2\pi}{13}s} + \sqrt{2}d_{-1}e^{-i\frac{2\pi}{13}s} + \sqrt{2}d_2 e^{i\frac{2\pi}{13}5s} + \sqrt{2}d_{-2}e^{-i\frac{2\pi}{13}5s} \\
&+ \quad \sqrt{2}d_3 e^{i\frac{2\pi}{13}6s} + \sqrt{2}d_{-3}e^{-i\frac{2\pi}{13}6s} + \sqrt{2}d_4 e^{i\frac{2\pi}{13}9s} + \sqrt{2}d_{-4}e^{-i\frac{2\pi}{13}9s} \qquad (2)
\end{aligned}
$$

This reveals the key point. The input parameter $x_1$ is associated with the frequency $\omega_1 = 1$, $x_2$ with $\omega_5 = 5$, the interaction $x_1 + x_2$ with $\omega_6 = 6$ and $x_1 - x_2$ with $\omega_9 = 9$. Indeed compare formulae (1) and (2) by letting $t = s$ and $c_i = \sqrt{2}d_i$. The terms giving the main effects are in both formulae and from the spectral analysis for the $Y_i$'s we are able to extract the important frequencies for (2). Then using the mapping $x_1 \leftrightarrow 1$, $x_2 \leftrightarrow 5$, $x_1 + x_2 \leftrightarrow 6$ we know which coefficients in $F(2;1,1;2)$ are to be selected for investigation. Note that $x_1 - x_2 \leftrightarrow 9$ is not available because only the frequencies up to order $\left[\frac{13}{2}\right]$ are recognised by the Fourier transform. Since we are interested in the main effects this does not prove restrictive.

¿From the power spectrum plot (see Figure 2), we deduce that the most influent frequency is at $k = 2$. This is confirmed by the direct evaluation of the Fourier coefficients in (1)

$$\begin{array}{ll} c_0 = 30.000000 + 0.000000i & \\ c_1 = 2.124319 - 1.264602i & c_2 = -8.759016 - 0.568902i \\ c_3 = -0.420373 + 2.222723i & c_4 = 0.355769 - 0.767301i \\ c_5 = -2.143804 - 1.048391i & c_6 = 0.343106 + 1.094877i \end{array}$$

and their squared norms show the key role of the second frequency

900.000000   6.111952   77.044014   5.117213   0.715323   5.695021   1.316477

The value at the frequency 0 corresponds to the mean of the observations. It is convenient to exclude this from plots because of its typically high value.

In conclusion, the power spectrum shows both $x_1$ and $x_2$ (corresponding to frequency points 1 and 5 respectively) as being significant with $x_1$ having the larger effect. ¿From this we can deduce the slightly greater importance of the variable $x_1$. The largest point on the spectrum is however the second frequency point and we offer the following explanation. The lattice design generated by $(1, 5)$ with 16 points is $D$–optimum for the model $F(2; 2, 2; 2)$. The analysis carried for $F(2; 2, 2; 2)$ with the same generator $(1, 5)$ but with 16 points suggests that a screening on the second order model for the first variable could be significant, due to the predominant role of the 3rd frequency corresponding to $2x_1$. This is confirmed by the plot of the power spectrum for the 16-point model comprising the 13 original points with $Y_{14} = 1, Y_{15} = 1, Y_{16} = 2$, see Figures 3 and 4.

### Acknowledgements

*The authors are supported by the UK Engineering and Physical Sciences Research Council and by the Deutsche Forschungsgemeinschaft.*

### 1. References

1 Bates, R.A., Buck, R.J., Riccomagno, E., Wynn, H.P.: Experimental design and observation for large systems; J. Royal Statistical Society, Series B, 1995 (in press).

2 Fang, W.T., Wang, Y.: Number–theoretic Methods in Statistics; Chapman & Hall (London), 1994.

3 Ferguson, R.S., Sprevak, D., Walton, A.J., Fallon, M., Newsam, M.I.: Screening of variables for the empirical modelling of semiconductor devices; Technical Report of the Queens Univerisity in Belfast.

4 Hoel, P.G.: Minimax distance designs in two dimensional regression; Ann. Math. Stat., **36** (1965),1097–1106.

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

6 Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods; CBMS-NFS, SIAM (Philadelphia), 1992.

7 Riccomagno, E., Schwabe, R., Wynn, H.P.: Lattice–based optimum design for fourier regression; Annals of Statistics, 1995 (submitted).

8 Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and ‘analysis of Computer Experiments; Statistical Science, **4** (1989), 409–435.

9 Sloan, I.H., Joe, S.: Lattice Methods for Multiple Integration; Oxford Science Publications, Oxford.

10 Wei, W.W.: Time series Analysis. Univariate and Multivariate Methods; Addison–Wesley, California, 1990.

11 Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D.: Screening, Predicting, and Computer Experiments; Technometrics, **34** (1992):15–25.

12 Yeşilyurt, S., Patera, A.T.: Surrogates for numerical simulations; optimization of eddy–promoter heat exchangers; Comp. Meth. in Appl. Mech. and Eng., (in press).

*Addresses:* R A Bates, City University, Engineering Design Centre, London EC1V 0HB. E Riccomagno, University of Warwick, Coventry. R Schwabe, The Free University of Berlin. H P Wynn, University of Warwick, Coventry.