Functionally Induced Priors for the Analysis of Experiments

V. Roshan Joseph and James D. Delaney

H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology

Atlanta, GA 30332-0205

roshan@isye.gatech.edu jdelaney@isye.gatech.edu

June 29, 2006

Abstract

This work extends and develops the idea of using functional priors for the design and analysis of three and higher level experiments. Developing a prior distribution for model parameters is challenging because a factor can be qualitative or quantitative. We propose appropriate correlation functions and coding schemes so that the prior distribution is simple and the results interpretable. The prior incorporates well known principles such as effect hierarchy and effect heredity, which helps to resolve the aliasing problems in fractional designs almost automatically. The usefulness of the new approach is illustrated through the analysis of some real experiments.

KEY WORDS: Bayesian Analysis; Experimental Design; Gaussian process; Helmert coding.

1. INTRODUCTION

In the analysis of a typical experiment with any number of three and higher level factors, the amount and nature of the calculations in the traditional analysis do not facilitate the adoption of analysis strategies that can be easily automated. Historically, frequentist analysis strategies have had to rely upon tedious calculations to establish the aliasing relationships that comprise the design's degrees of freedom. Tools like half-normal plots and interaction plots would be used to perform variable selection and determine optimal factor settings, respectively, while computing was used to perform calculations necessary to estimate effects. While sometimes adequate, the traditional approach can be quite time consuming and does not lend itself well to utilization of the computing power that is now available.

Designs of three-level and four-level factors figure prominently in physical experiments. For example, all of the case studies reported in Taguchi, Chowdhury, and Taguchi (2000) use mixed two, three, and higher level designs. See Taguchi (1987) and Wu and Hamada (2000) for several other examples. Beginning with three-level factors, much more information about the shape of the response surface can be extracted from a good design. Unfortunately, the run size of full factorial 3^p and 4^p designs can be prohibitively large. Fractional factorial designs are used for reducing the run size, but they lead to aliasing of the effects. Several Bayesian approaches for estimating the effects from fractional designs have been suggested in the literature. The technique used to incorporate prior information plays an important role in both optimal design choice and the subsequent estimation of effects and variable selection.

Some previous approaches to Bayesian methods to the design and analysis of experiments have focused on Bayesian hierarchical models that require eliciting or estimating many hyper-parameters in order to specify priors for a linear model's parameters. See for instance the review of the literature by Chaloner and Verdinelli (1995). Problems arise in both the proper specification of the numerous priors as well as potential complications with calculation. An interesting Bayesian approach, with model priors that facilitate the incorporation of principles like *effect heredity* is suggested by Chipman, Hamada, and Wu (1997). In that paper, estimation is through a Gibbs sampling procedure.

The specification of a prior for the model parameters is not a trivial matter. The sheer quantity of the parameters is a major problem, but there are several other issues. For example, consider a 3^2 design. Suppose u_1 and u_2 represent the two coded variables of the first factor and u_3 and u_4 those of the second factor. Then the linear model that we would like to fit is

$$Y = \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 + \beta_4 u_4 + \beta_5 u_1 u_3 + \beta_6 u_1 u_4 + \beta_7 u_2 u_3 + \beta_8 u_2 u_4 + \epsilon.$$

What should be the prior distribution for the β's? The usual approach is to take them as $\mathcal{N}(0, \tau_0^2)$, see for example Chipman et al. (1997). Although this choice looks reasonable, several questions remain unanswered. For example, by the effect hierarchy principle (see Hamada and Wu 1992), we know that a two-factor interaction (2fi) is less likely to be significant than a main effect. Therefore is it ideal to use the same distribution for a main effect and 2fi? Moreover, we can use different coding schemes to represent the two degrees of freedom for each factor. How should we change the prior specification depending on the coding scheme? Are the two effects of the same factor, say β_1 and β_2, equally important? It is known that if we use a linear-quadratic system, then the linear effect is more important than the quadratic effect. How do we incorporate such a difference in the prior? Is it reasonable to take all the parameters to be independent? How should the prior be modified depending on the type of factor viz. qualitative or quantitative? It is imperative to develop a coherent and systematic approach to prior specification, so that we can answer all these questions.

In this work we propose the use of functionally induced priors for prior specification (Joseph 2006). Here a prior using a Gaussian process is postulated for the underlying transfer function and then the prior distribution for all the model parameters is induced from it. The work in Joseph (2006) focuses on two-level experiments. Mitchell, Morris, and Ylvisaker (1995) and Kerr (2001) have also studied the use of stochastic processes for the design of two-level experiments. Here we extend the approach for the case of three and higher level experiments. The extension is not trivial as there are many issues involved in higher level experiments that are not present in two-level experiments. For example, the type of factor, the type of correlation function, the type of coding scheme, the mixed-level nature

of the experiments, etc. become important when dealing with higher level experiments, but are irrelevant for two-level experiments.

A very nice property of the functionally induced prior is that it agrees with many widely accepted principles in the design and analysis of experiments such as effect sparsity, effect hierarchy, and effect heredity (Wu and Hamada 2000). The introduction of these priors has provided for a very nice setting that enables the automation of many analytical tasks, that in previous approaches would have required a great deal of time consuming manual work.

The technical report is organized as follows. We begin by reviewing the general functionally induced prior Bayesian framework. We present a decomposition result that is extremely useful for studying three-level, four-level, $\cdots$, and mixed-level designs. The results are different for the case of qualitative and quantitative factors. In Section 3, we present the results for qualitative factors. The results of this section are very simple and general, so that they can be used with any number of levels. In Section 4, the building blocks for applying the Bayesian methodology to three-level and four-level quantitative factors are presented. Here, there is a brief discussion of complications that can arise due to the choice of coding-scheme for the model matrix. We also demonstrate that a direct consequence of the functionally induced prior is a systematic methodology for ordering the effects. The utility of this Bayesian setting is illustrated through examples where the forward variable selection procedure is adapted to designs with three-level and four-level factors. This appears in Section 5. Finally, concluding remarks and suggestions for future research are given in Section 6.

2. GENERAL METHODOLOGY

Suppose that there are p factors $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$, where the factor x_i is experimented with at m_i levels. Assume the model

$$Y = f(\boldsymbol{x}) + e, \quad e \sim \mathcal{N}(0, \sigma^2),$$

where e represents the random error in the response due to the uncontrollable variables in the system. The transfer function f could be nonlinear and highly complex, but we would

like to approximate it by a linear model containing the main effects and interactions of the factors. The factor x_i can be represented by $m_i - 1$ coded variables and the interactions can be defined through the products of these coded variables. Thus we would like to approximate $f(\boldsymbol{x})$ by

$$f(\boldsymbol{x}) \approx \sum_{i=0}^{q-1} \beta_i u_i,$$

where $q = \prod_{i=1}^{p} m_i$. For example, in the 3^2 design discussed in Section 1, we let $u_5 = u_1 u_3$, $u_6 = u_1 u_4$, $u_7 = u_2 u_3$, and $u_8 = u_2 u_4$.

As the number of factors and/or the number of levels increase, the total number of parameters (q) can become very large. Therefore, postulating a prior distribution for $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_{q-1})'$ is a difficult task. Joseph (2006) used a simple idea to overcome this problem. The idea is to postulate a functional prior for the transfer function and use that to induce a prior for all of the parameters in the linear model. Therefore, let

$$f(\boldsymbol{x}) \sim GP(\mu_0, \sigma_0^2 \psi),$$

where μ_0 is the mean and $\sigma_0^2\psi$ is the covariance function of the Gaussian process (GP). The covariance function is defined as $cov(Y(\boldsymbol{x}), Y(\boldsymbol{x} + \boldsymbol{h})) = \sigma_0^2 \psi(\boldsymbol{h})$. Because there are q parameters in the linear model, they can be chosen to exactly match the function values at q points. A simple choice for the q points is the full factorial design. Let $\boldsymbol{U}$ be the $q \times q$ model matrix for the parameter $\boldsymbol{\beta}$ and let $\boldsymbol{\Psi}$ be the corresponding correlation matrix. To simplify the results, consider instead $f(\boldsymbol{x}) = \mu_0 + \sum_{i=0}^{q-1} \beta_i u_i$ at the q points in the full factorial design. Then,

$$\boldsymbol{\beta} \sim \mathcal{N}\left(0, \sigma_0^2 \boldsymbol{U}^{-1} \boldsymbol{\Psi} (\boldsymbol{U}^{-1})'\right).$$

For obvious reasons, we call this a *functionally induced prior distribution*. For large q, the variance-covariance matrix is huge, which can be difficult to construct and handle. Therefore it is important to simplify the representation of the above matrix so that the results can be easily used in practice. We achieve this under some assumptions.

Assume that the correlation function ψ has a product correlation structure of the form:

$$\psi(\boldsymbol{h}) = \prod_{j=1}^{p} \psi_j(|h_j|). \tag{1}$$

Let $\boldsymbol{U}_j$ be the model matrix for factor x_j and let $\boldsymbol{\Psi}_j$ be the corresponding correlation matrix. For example, for a 3-level factor with possible levels 1, 2, and 3, the model matrix using orthogonal polynomial coding (with common column lengths of $\sqrt{3}$) is

$$\boldsymbol{U}_j = \begin{pmatrix} 1 & -\sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \\ 1 & 0 & -\sqrt{2} \\ 1 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \end{pmatrix} \tag{2}$$

and the correlation matrix is

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \psi_j(1) & \psi_j(2) \\ \psi_j(1) & 1 & \psi_j(1) \\ \psi_j(2) & \psi_j(1) & 1 \end{pmatrix}. \tag{3}$$

Now we have the following result. All of the proofs are given in the Appendix 6.

THEOREM 1. Under the product correlation structure in (1):

$$var(\boldsymbol{\beta}) = \sigma_0^2 \bigotimes_{j=1}^{p} \boldsymbol{U}_j^{-1} \boldsymbol{\Psi}_j (\boldsymbol{U}_j^{-1})'.$$

The impact of this theorem is that we can focus on each factor, one at a time, in choosing whatever coding scheme and correlation function suits our modeling needs. For example, to construct the variance-covariance matrix in a $2^{p_2} \times 3^{p_3} \times 4^{p_4}$ design, we only need to establish the structure of $\boldsymbol{U}_j^{-1} \boldsymbol{\Psi}_j (\boldsymbol{U}_j^{-1})'$ for a two-level, three-level, and four-level factor. The results can then be combined by taking Kronecker products to get the desired variance-covariance matrix for any values of p_2, p_3, and p_4.

In the following sections, we investigate the structure of the variance-covariance matrix, so that the result can be easily interpreted and applied in the design and analysis of experiments. The choice of correlation functions and coding schemes depend on the type of factors. Therefore we study the case of qualitative and quantitative factors separately.

3. QUALITATIVE FACTORS

By *qualitative factor* we mean a factor whose levels are nominal. That is, a qualitative factor might be the name of: the vendor for a part, a machine, a method, etc.

3.1 Correlation Function

For a qualitative factor, we should assign equal correlation between any two levels. This is because among our prior assumptions, there is no information as to how to either order these factor levels or to determine the relative distances between any of the levels. As mentioned previously, we assume that the prior Gaussian process is stationary. So for the j^{th} factor, we need only be concerned with $h_j = |x_{ij} - x_{kj}|$, for two runs i and k. That is,

$$\psi_j(h_j) = \begin{cases} 1 & \text{if } h_j = 0 \\ \rho_j & \text{if } h_j \neq 0 \end{cases},$$

where $0 < \rho_j < 1$. Then the $m_j \times m_j$ correlation matrix has the *compound symmetry* form:

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \rho_j & \dots & \rho_j \\ \rho_j & 1 & \dots & \rho_j \\ \vdots & \dots & \ddots & \vdots \\ \rho_j & \rho_j & \dots & 1 \end{pmatrix}. \tag{4}$$

3.2 Prior Distribution

Suppose that for whatever coding schemes are selected for each of the single factor model matrices $\boldsymbol{U}_j$ for $j = 1, \dots, p$, we impose only the restrictions that the first column of each $\boldsymbol{U}_j$ is $\mathbf{1}_{m_j}$ to correspond to the "y-intercept" effect, and that the remaining $m_j - 1$ columns of each $\boldsymbol{U}_j$ are a set of mutually orthogonal contrasts normalized to the length $\sqrt{m_j}$. Thus $\boldsymbol{U}_j'\boldsymbol{U}_j = m_j I_{m_j}$, where I_{m_j} is the identity matrix of dimension m_j. Then for $\boldsymbol{\Psi}_j$ as in (4),

$$\boldsymbol{U}_j'\boldsymbol{\Psi}_j\boldsymbol{U}_j = m_j \begin{pmatrix} 1 + (m_j - 1)\rho_j & 0 & \dots & 0 \\ 0 & 1 - \rho_j & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 - \rho_j \end{pmatrix}. \tag{5}$$

Of course, since the columns of U_j are mutually orthogonal vectors, each with squared length m_j, we have:

$$\boldsymbol{U}_j^{-1}\boldsymbol{\Psi}_j(\boldsymbol{U}_j^{-1})' = \frac{1}{m_j^2}\boldsymbol{U}_j'\boldsymbol{\Psi}_j\boldsymbol{U}_j. \tag{6}$$

Now we can propose the following very general result for the joint prior distribution of the effects for a design incorporating some number of qualitative factors with any mixture of levels. This follows directly from Theorem 1, Equation (5) and Equation (6). Let

$$\tau_0^2 = \sigma_0^2 \prod_{j=1}^{p} \frac{1 + (m_j - 1)\rho_j}{m_j} \qquad \text{and} \qquad r_j = \frac{1 - \rho_j}{1 + (m_j - 1)\rho_j}.$$

Let $\delta_{ij} = 1$ if β_i includes the factor j and 0 otherwise.

PROPOSITION 1. For factorial experiments on p qualitative factors, if we use an orthogonal coding for each factor and correlation matrix as in (4), then

$$\beta_i \sim \mathcal{N}(0, \tau_0^2 \prod_{j=1}^{p} r_j^{\delta_{ij}}), \quad i = 0, 1, \cdots, \prod_{j=1}^{p} m_j - 1$$

and the effects are all mutually independent.

Note that because of the independence, the variance-covariance matrix is diagonal, which makes it very easy to construct. While the expression in the above proposition may seem a bit complicated, it is easy to summarize what is happening. The variance of any effect depends not on what the interpretation of the effect is, which contrasts are involved, but only on which factors are involved in that effect. For further clarity, consider the following example.

Example: Suppose we have two factors: A and B each experimented at three levels. Let a_1 and a_2 represent the two coded variables of factor A and b_1 and b_2 that of factor B. With the correlation matrix for factor A (and similarly for factor B):

$$\mathbf{\Psi}_A = \begin{pmatrix} 1 & \rho_A & \rho_A \\ \rho_A & 1 & \rho_A \\ \rho_A & \rho_A & 1 \end{pmatrix},$$

we have $\tau_0^2 = \frac{\sigma_0^2}{9}(1+2\rho_A)(1+2\rho_B), r_A = (1-\rho_A)/(1+2\rho_A)$, and $r_B = (1-\rho_B)/(1+2\rho_B)$. So that from Proposition 1: $\beta_0 \sim \mathcal{N}(0, \tau_0^2)$, β_{a_1} and $\beta_{a_2} \sim \mathcal{N}(0, \tau_0^2 r_A)$, β_{b_1} and $\beta_{b_2} \sim \mathcal{N}(0, \tau_0^2 r_B)$, and $\beta_{a_1 b_1}$, $\beta_{a_1 b_2}$, $\beta_{a_2 b_1}$, and $\beta_{a_2 b_2} \sim \mathcal{N}(0, \tau_0^2 r_A r_B)$.

Since each $0 < r_j < 1$ can be specified or estimated, both concepts of effect hierarchy and effect heredity are appropriately integrated into the prior. Generally, as the number of factors involved in an interaction increases, the *a priori* variance around the effect's mean, which is zero, decreases, justifying effect hierarchy. If a particular ρ_j is small, then the corresponding r_j is large, which would suggest a comparatively larger variance for effects that include that factor than those interactions of the same order that do not, justifying effect heredity.

There is a very simple case of Proposition 1 which arises when all of the correlation matrices for the factors are the same. When this occurs, the marginal prior of the effect depends on whether that effect is a "main effect" (me), "two-factor interaction" (2fi), ..., "p-factor interaction" (pfi):

COROLLARY 1. . For $r_1 = r_2 = \ldots = r_p = r$,

$$\begin{aligned}
\beta_0 &\sim \mathcal{N}(0, \tau_0^2) \\
\beta_{me} &\sim \mathcal{N}(0, \tau_0^2 r) \\
\beta_{2fi} &\sim \mathcal{N}(0, \tau_0^2 r^2) \\
&\vdots \\
\beta_{pfi} &\sim \mathcal{N}(0, \tau_0^2 r^p),
\end{aligned}$$

and the effects are all mutually independent.

3.3 Coding Schemes

The restrictions imposed on our model matrix to obtain the convenient result of Proposition 1 actually admit many reasonable choices for coding schemes. We would naturally find it desirable to estimate an overall mean effect, so the leading column of 1's is not really an imposing constraint. That the other columns need be orthogonal contrasts is also quite natural. We are still free to choose contrasts that have a sensible interpretation for the type of factor we are considering in this section, a qualitative factor. Below we discuss two such

coding schemes that satisfy the assumptions of Proposition 1, but have been suggested in the frequentist design setting, indicating their value in interpretation.

For ease of implementation and interpretation, the orthogonal contrast coding scheme we recommend for a qualitative factor is Helmert coding (see Harville 1997). Other commonly used alternative coding schemes present problems. For instance, the effects from orthogonal polynomial coding do not have a natural interpretation for a qualitative factor. Although we do note that for two-level and three-level factors, Helmert coding and orthogonal polynomial coding are the same. Wu and Hamada (2000) offers some other alternatives. For example, for a three-level factor, the choice of using two of the following coding vectors: $\boldsymbol{D}_{01} = (-1, 1, 0)$, $\boldsymbol{D}_{02} = (-1, 0, 1)$, or $\boldsymbol{D}_{12} = (0, -1, 1)$ yield estimates for interpretable effects, however they are not mutually orthogonal. The problem more generally with treatment coding or zero sum coding is that the columns of $\boldsymbol{U}_j$ would not be mutually orthogonal. This would violate the assumptions that led to Proposition 1. Helmert coding, on the other hand, along with providing for the calculation of effects that may be interesting for the analysis of a qualitative factor, is quite easy to implement for any number of levels. In Helmert coding, the first effect is the difference between the second level and the first level. The second effect is the difference between the third level and the average of the first two, etc. Below is the model matrix that makes the interpretation of effects more obvious:

$$\begin{pmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & -1 & -1 & \cdots & -1 \\ 1 & 0 & 2 & -1 & \cdots & -1 \\ 1 & 0 & 0 & 3 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & (m_j - 1) \end{pmatrix}.$$

We need to "normalize" each column to have the same squared length, m_j. To accomplish this, the factors to multiply each column by are:

$$\left(1, \sqrt{\frac{m_j}{2}}, \sqrt{\frac{m_j}{6}}, \sqrt{\frac{m_j}{12}}, \ldots, \sqrt{\frac{1}{m_j - 1}}\right).$$

One should not feel restricted to using the above recommended coding scheme. Any set of mutually orthogonal contrasts will do. So if there is a set of such effects that is more interesting to the experimenter, they should be used. For a four-level qualitative factor, Wu and Hamada (2000) offers a convenient coding scheme. Their recommendation provides effects that can be interpreted as differences between pairs of levels. The model matrix is below:

$$\mathbf{U}_j = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

The above coding scheme can be extended to factors with 8, 12, 16, ... levels using Hadamard matrices. Such a coding scheme using only $\{-1, 1\}$ would not be naturally applicable to a three-level factor or a five-level factor. Whereas model matrices based on Helmert coding are easy to construct and provide interpretable effects for any number of factor levels.

4. QUANTITATIVE FACTORS

When a factor is continuous or discrete, but ordinal, where there exists some way to quantify the differences between the factor's level, we may treat it as a *quantitative* factor. For a quantitative factor, we may wish to run the experiment at evenly spaced levels, but this might not always be possible. Below we make recommendations for each situation.

4.1 Correlation Function

When the levels are equally spaced, the correlation matrix $\mathbf{\Psi}_j$ has the *symmetric Toeplitz* form given in (16), with $\psi_j(h_j) \to 0$ as $|h_j| \to \infty$. There are many parametric forms for $\psi_j(h_j)$. For example, generally we could make use of the two parameter exponential correlation function:

$$\psi_j(h_j) = \exp\left(-\theta_j |h_j|^{\alpha_j}\right) \qquad 0 < \alpha_j \leq 2 \qquad 0 < \theta_j < \infty. \tag{7}$$

This is the most popular correlation function used in computer experiments, but other correlation functions such as the Matérn correlation function, cubic correlation function, etc. could also be used (see Santner et al. 2003). A convenient special case of the exponential correlation function is when the parameter $\alpha_j = 2$. This case is useful for modeling functions that are infinitely differentiable. Let $\rho_j = \exp(-\theta_j)$. Then

$$\mathbf{\Psi}_j = \begin{pmatrix} 1 & \rho_j & \dots & \rho_j^{(m_j-1)^2} \\ \rho_j & 1 & \dots & \rho_j^{(m_j-2)^2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_j^{(m_j-1)^2} & \rho_j^{(m_j-2)^2} & \dots & 1 \end{pmatrix}, \tag{8}$$

which will be used in most of the examples presented here.

We suggest that when the levels are not evenly spaced for factor j, that instead of using the values $x_j \in \{1, 2, \dots, m_j\}$ for the arguments of the correlation function, that the end points: 1 and m_j be used for the first and last levels, and that the other levels be represented by interpolating between these points. For example, if the unevenly spaced levels for a quantitative three-level factor are: $25, 30, 37$, then use the quantities $1, 11/6, 3$, instead of $1, 2, 3$ in the correlation function. One might anticipate that the factor levels would have to be grossly unevenly spaced for it to result in any noticeable changes in practice. However, this is one part of our suggested methodology where some caution should be exercised.

The most common coding scheme for quantitative factors is *orthogonal polynomial coding* (see Wu and Hamada 2000). Unfortunately, a general result like Proposition 1 does not exist for quantitative factors under this coding scheme. Therefore, we examine the most important cases of three-level and four-level designs in detail.

4.2 Prior Distribution for Three-Level Experiments

Let us consider an experiment with p quantitative, evenly spaced three-level factors. We assume a correlation matrix for each factor $j = 1, 2, \ldots, p$ of the form:

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \psi_j(1) & \psi_j(2) \\ \psi_j(1) & 1 & \psi_j(1) \\ \psi_j(2) & \psi_j(1) & 1 \end{pmatrix}.$$

The model matrix using orthogonal polynomial coding (with common column lengths of $\sqrt{3}$) is:

$$\boldsymbol{U}_j = \begin{pmatrix} 1 & -\sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \\ 1 & 0 & -\sqrt{2} \\ 1 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \end{pmatrix}.$$

So that by matrix multiplication, we have:

$$\boldsymbol{U}_j'\boldsymbol{\Psi}_j\boldsymbol{U}_j = \begin{pmatrix} 3 + 4\psi_j(1) + 2\psi_j(2) & 0 & -\sqrt{2}(\psi_j(1) - \psi_j(2)) \\ 0 & 3(1 - \psi_j(2)) & 0 \\ -\sqrt{2}(\psi_j(1) - \psi_j(2)) & 0 & 3 - 4\psi_j(1) + \psi_j(2) \end{pmatrix}. \tag{9}$$

Notice that the "quadratic" and "y-intercept" effects are going to be (negatively) correlated. This is an important difference from qualitative factors, where this matrix was diagonal. We can now propose expressions for the model parameters' marginal prior distributions. For the following proposition, let:

$$\tau_0^2 = \frac{\sigma_0^2}{3^{2p}} \prod_{j=1}^{p} (3 + 4\psi_j(1) + 2\psi_j(2)),$$

$$r_{jl} = \frac{3 - 3\psi_j(2)}{3 + 4\psi_j(1) + 2\psi_j(2)}, \qquad r_{jq} = \frac{3 - 4\psi_j(1) + \psi_j(2)}{3 + 4\psi_j(1) + 2\psi_j(2)}.$$

Let $l_{ij} = 1$ if β_i includes the linear effect of factor j and 0 otherwise. Similarly, $q_{ij} = 1$ if β_i includes the quadratic effect of factor j and 0 otherwise. Then the following expressions follow directly from (9) and Theorem 1:

PROPOSITION 2. For p quantitative three-level factors, with a model matrix coded according to orthogonal polynomial contrasts we have:

$$\beta_i \sim \mathcal{N}\left(0, \tau_0^2 \prod_{j=1}^{p} r_{jl}^{l_{ij}} r_{jq}^{q_{ij}}\right), \quad i = 0, 1, \cdots, 3^p - 1.$$

Note that unlike the result in Proposition 1, the β_i's are not independent. We will consider some properties of a special case of this result. Suppose for each factor we assume a 3×3 correlation matrix in the form of (8), with $\rho_1 = \rho_2 = \ldots = \rho_p = \rho$. Let,

$$\tau_0^2 = \frac{\sigma_0^2}{3^{2p}}(3 + 4\rho + 2\rho^4)^p, \qquad r_l = \frac{3 - 3\rho^4}{3 + 4\rho + 2\rho^4}, \qquad r_q = \frac{3 - 4\rho + \rho^4}{3 + 4\rho + 2\rho^4}.$$

Then the following expressions illustrate a useful special case of Proposition 2:

COROLLARY 2. . For p quantitative three-level factors, with a model matrix coded according to orthogonal polynomial contrasts, if we further assume $\rho_1 = \rho_2 = \cdots = \rho_p = \rho$ in (8), then:

$$\begin{aligned}
\beta_0 &\sim \mathcal{N}\left(0, \tau_0^2\right) \\
\beta_l &\sim \mathcal{N}\left(0, \tau_0^2 r_l\right) \\
\beta_q &\sim \mathcal{N}\left(0, \tau_0^2 r_q\right) \\
\beta_{ll} &\sim \mathcal{N}\left(0, \tau_0^2 r_l^2\right) \\
\beta_{lq} &\sim \mathcal{N}\left(0, \tau_0^2 r_l r_q\right) \\
&\vdots \\
\beta_{q\cdots q} &\sim \mathcal{N}\left(0, \tau_0^2 r_q^p\right).
\end{aligned}$$

To clarify the above notation, the subscript lq, for example, indicates that β_{lq} is the interaction effect between the linear effect of any one of the p factors and the quadratic effect of any of the other $p - 1$ remaining factors.

We have for all $\rho \in (0, 1)$

$$0 < r_l^3 < r_q < r_l^2 < r_l < 1. \tag{10}$$

It is quite common to say that a quadratic effect is less important than a linear effect (notationally $q \prec l$) . Because $r_q < r_l$, we now have a mathematical justification of the above statement. Similarly, $r_q < r_l^2$ shows that $q \prec ll$, which is an interesting result. The property (10) can be used to order many higher order effects. For example, $qq \prec llq \prec lq \prec ll$.

The ordering of effects is important for properly defining a design criterion similar to minimum aberration. Cheng and Ye (2005) proposes two rules:

$$(a) \quad : \quad l \succ q \succ ll \succ lq \succ lll \succ qq \succ llq \succ llll \succ lqq \succ lllq \succ qqq \succ llqq \succ lqqq \succ qqqq,$$

$$(b) \quad : \quad l \succ q \succ ll \succ lq \succ qq \succ lll \succ llq \succ lqq \succ qqq \succ llll \succ lllq \succ llqq \succ lqqq \succ qqqq.$$

Rule (a) is obtained by ordering effects first by the degree of the polynomial, and then within that by the number of factors involved in the interaction, whereas rule (b) is by ordering effects first by the number of terms in the interaction, and then by the degree of that polynomial. Alternatively, the rule implied by (10) is:

$$l \succ ll \succ q \succ lll \succ lq \succ llll \succ llq \succ qq \succ lllq \succ lqq \succ llqq \succ qqq \succ lqqq \succ qqqq.$$

As the number of factors increases, the ordering rule can get more complicated. For example, when there are five three-level factors, and $\alpha = 2$, we would need a result like (10) to include a comparison of the two terms r_l^5 and r_q^2. Numerically, it can be shown that: for $\rho \in (0, 0.357)$ or $(0.847, 1)$,

$$0 < r_l^5 < r_q^2 < r_l^3 < r_q < r_l^2 < r_l < 1,$$

whereas for $\rho \in (0.357, 0.847)$,

$$0 < r_q^2 < r_l^5 < r_l^3 < r_q < r_l^2 < r_l < 1.$$

This should be enough to order the orthogonal polynomial factorial effects for five three-level factors when the value of ρ is known. When $\alpha = 1$ the ordering of effects additionally depends on ρ with as few as four factors. The nice thing about the Bayesian approach is that we do not need to worry about these complicated ordering of effects, it will be automatically built-in in the design and analysis of experiments.

4.3 Prior Distribution for Four-Level Experiments

Let us now consider an experiment with p quantitative, evenly spaced four-level factors. We assume a correlation matrix for each evenly spaced factor $j = 1, 2, \ldots, p$ of the form:

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \psi_j(1) & \psi_j(2) & \psi_j(3) \\ \psi_j(1) & 1 & \psi_j(1) & \psi_j(2) \\ \psi_j(2) & \psi_j(1) & 1 & \psi_j(1) \\ \psi_j(3) & \psi_j(2) & \psi_j(1) & 1 \end{pmatrix}. \tag{11}$$

As we did before we may attempt to use orthogonal polynomial coding, albeit with some reservations, in anticipation that some of the off diagonal terms in the prior parameter covariance matrix will be nonzero. The "normalized" model matrix for the j^{th} four-level factor using orthogonal polynomial coding is:

$$\boldsymbol{U}_j = \begin{pmatrix} 1 & -\frac{3}{\sqrt{5}} & 1 & -\frac{1}{\sqrt{5}} \\ 1 & -\frac{1}{\sqrt{5}} & -1 & \frac{3}{\sqrt{5}} \\ 1 & \frac{1}{\sqrt{5}} & -1 & -\frac{3}{\sqrt{5}} \\ 1 & \frac{3}{\sqrt{5}} & 1 & \frac{1}{\sqrt{5}} \end{pmatrix}.$$

Now this model matrix would enable us to calculate "y-intercept", "linear", "quadratic", and "cubic" effects. However, as suggested by the matrix calculation below, the prior covariance matrix is in fact not diagonal. There are some nonzero covariances between the "y-intercept" and the "quadratic" effect as well as between the "linear" and "cubic" effects. For the following equation, let us denote: $\psi_j(1) = \psi_{j1}$, $\psi_j(2) = \psi_{j2}$, and $\psi_j(3) = \psi_{j3}$. Then we obtain:

$$\boldsymbol{U}_j'\boldsymbol{\Psi}_j\boldsymbol{U}_j =$$

$$\begin{pmatrix} 4+6\psi_{j1}+4\psi_{j2}+2\psi_{j3} & 0 & -2(\psi_{j1}-\psi_{j3}) & 0 \\ 0 & 4+2\psi_{j1}-\frac{12}{5}\psi_{j2}-\frac{18}{5}\psi_{j3} & 0 & -2\psi_{j1}+\frac{16}{5}\psi_{j2}-\frac{6}{5}\psi_{j3} \\ -2(\psi_{j1}-\psi_{j3}) & 0 & 4-2\psi_{j1}-4\psi_{j2}+2\psi_{j3} & 0 \\ 0 & -2\psi_{j1}+\frac{16}{5}\psi_{j2}-\frac{6}{5}\psi_{j3} & 0 & 4-6\psi_{j1}+\frac{12}{5}\psi_{j2}-\frac{2}{5}\psi_{j3} \end{pmatrix}. \tag{12}$$

Now using the above result and Theorem 1, we can obtain a result similar to Proposition 2. For notational simplicity, we will only provide a special case where the correlation matrix is

as in (8) with $\rho_1 = \rho_2 = \ldots = \rho_p = \rho$. Let

$$\tau_0^2 = \frac{\sigma_0^2}{4^{2p}}(4 + 6\rho + 4\rho^4 + 2\rho^9)^p, \qquad r_l = \frac{4 + 2\rho - \frac{12}{5}\rho^4 - \frac{18}{5}\rho^9}{4 + 6\rho + 4\rho^4 + 2\rho^9},$$

$$r_q = \frac{4 - 2\rho - 4\rho^4 + 2\rho^9}{4 + 6\rho + 4\rho^4 + 2\rho^9}, \qquad r_c = \frac{4 - 6\rho + \frac{12}{5}\rho^4 - \frac{2}{5}\rho^9}{4 + 6\rho + 4\rho^4 + 2\rho^9}.$$

PROPOSITION 3. For p quantitative four-level factors, with a model matrix coded according to orthogonal polynomial contrasts, if we further assume $\rho_1 = \rho_2 = \cdots = \rho_p = \rho$ in (8), then:

$$\beta_i \sim \mathcal{N}\left(0, \tau_0^2 r_l^{l_i} r_q^{q_i} r_c^{c_i}\right),$$

where l_i, q_i, and c_i are the number of linear, quadratic, and cubic terms in β_i.

Thus, using the notations used in Corollary 2, $\beta_c \sim \mathcal{N}(0, \tau_0^2 r_c), \beta_{lc} \sim \mathcal{N}(0, \tau_0^2 r_l r_c)$, etc. Note that as in Corollary 2, these effects are not independent.

At this point, it should be abundantly clear that it is a trivial matter to construct Proposition 1, Proposition 2, and Proposition 3-like results for designs of any mixture of factors with possibly different numbers of levels, possibly different types: qualitative or quantitative, and different parametric forms for the correlation functions. The user need only construct a model matrix U_j and correlation matrix Ψ_j appropriate for each factor and then deduce the full factorial prior covariance results using Theorem 1.

4.4 Coding Schemes

In this section on quantitative factors, we presented results for three-level and four-level factors assuming that orthogonal polynomial coding was the most desirable coding scheme. This coding scheme does indeed have some nice properties. The model matrix is easy to generate for a factor of any number of levels. The effects generated from such a model matrix are also easy to interpret for a quantitative factor.

One drawback to the orthogonal polynomial coding scheme for a design containing a quantitative factor is that the resulting prior covariance matrix is not diagonal. In fact constructing the matrix $\boldsymbol{R} = \tau_0^{-2}\, var(\boldsymbol{\beta})$, which will be used in the estimation, is not a

trivial matter. If the matrix $\boldsymbol{R}$ is not calculated directly, which in itself could be prohibitively computationally intensive, it is quite a difficult matter of accounting to calculate and position these off-diagonal elements correctly in the matrix. In addition, the matrix $\boldsymbol{R}$ represented in its full form, may be quite large, requiring sparse matrix techniques.

So suppose instead that our motivation was to find an orthogonal coding scheme for a quantitative factor's model matrix $\boldsymbol{U}_j$ that produces a diagonal $\boldsymbol{U}_j^{-1}\boldsymbol{\Psi}_j(\boldsymbol{U}_j^{-1})'$. Let $\boldsymbol{\Lambda}_j = diag(\lambda_{j,1}, \lambda_{j,2}, \ldots, \lambda_{j,m_j})$, with each $\lambda_{j,k}$, $k = 1, \ldots, m_j$ being the eigenvalues of $\boldsymbol{\Psi}_j$ and $\boldsymbol{E}_j$ is a $m_j \times m_j$ matrix whose columns are orthonormal eigenvectors corresponding to those eigenvalues. Now by letting $\boldsymbol{U}_j = \sqrt{m_j}\boldsymbol{E}_j$, we obtain

$$\boldsymbol{U}_j^{-1}\boldsymbol{\Psi}_j(\boldsymbol{U}_j^{-1})' = \frac{1}{m_j}\boldsymbol{\Lambda}_j, \tag{13}$$

which is a diagonal matrix. Now by Theorem 1, the variance-covariance matrix is also diagonal. Therefore, the matrix R could be easily constructed. A related idea exists in Steinberg and Bursztyn (2004), which contains a procedure for data analysis that involves relating regression coefficients to those produced from the eigenvectors of the correlation matrix of the corresponding random field model. However, using a model matrix whose columns are proportional to the eigenvectors of $\boldsymbol{\Psi}_j$ presents its own problems in the context here. The coding scheme will vary with the correlation matrix. That is, the model matrix $\boldsymbol{U}_j$ will depend on $\psi_j(1), \psi_j(2), \ldots, \psi_j(m_j - 1)$. Also, that the leading column of this coding scheme will not precisely be a column of ones, failing to yield a true "y-intercept" effect, complicates construction of the full design model matrix with Kronecker products less predictable. So rather than suggesting the use of this"eigen-Coding" scheme here, we merely use the observation that orthogonal polynomial coding is very nearly the coding scheme obtained from the eigenvectors of $\boldsymbol{\Psi}_j$'s as evidence in support of the belief that dismissing the off-diagonal elements of the matrix $\boldsymbol{R}$ may be acceptable in practice.

In Figure 1 we see a comparison of the orthogonal polynomial coding scheme and the eigen-coding scheme. From the figure we see the curvature in the y-intercept and linear effects demonstrating their dependence on the quadratic and cubic effects respectively. In addition, the plots demonstrate the relative "closeness" of the orthogonal polynomial effects

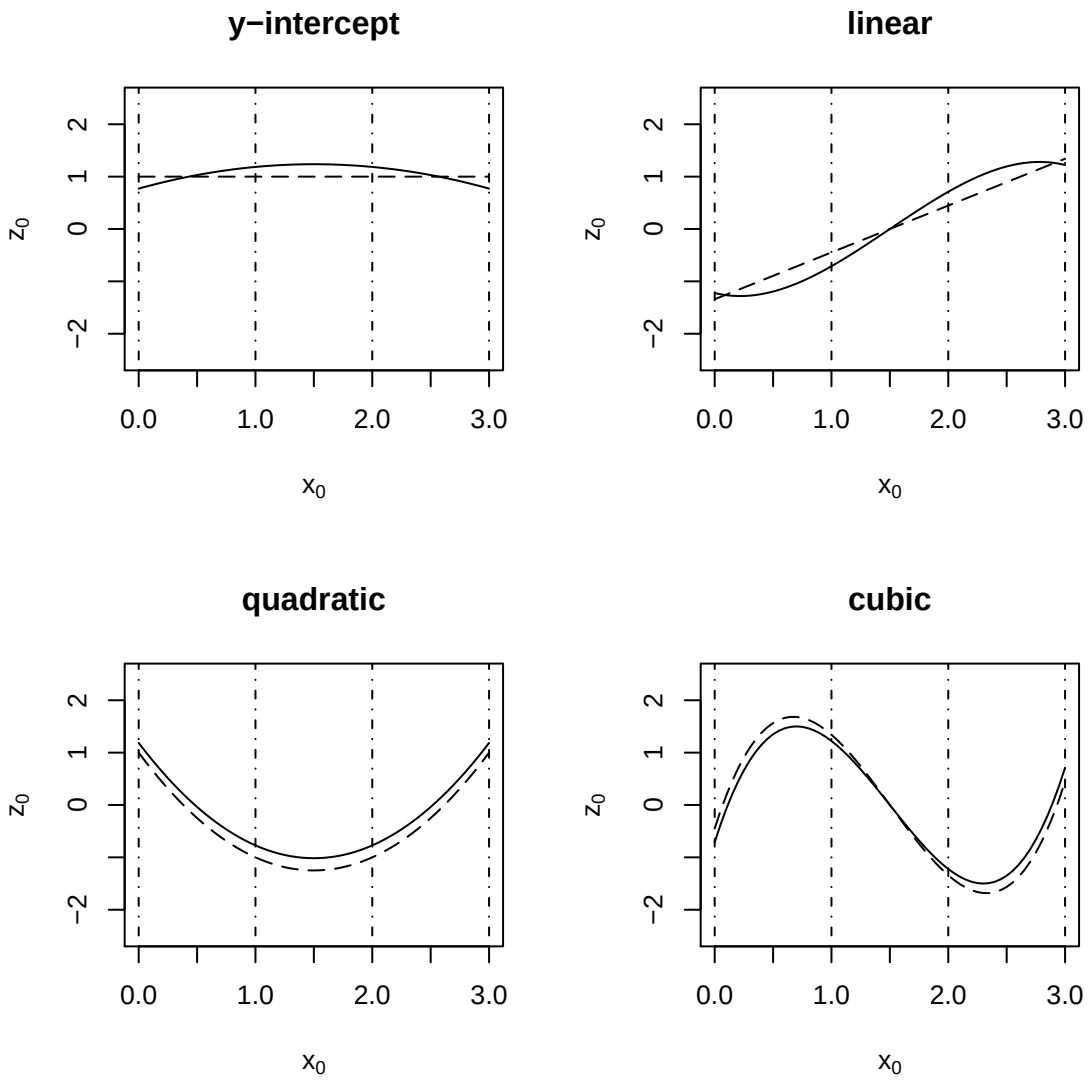


Figure 1: Comparison of Eigen-Coding (solid) and Orthogonal Polynomial Coding (dashed) for $\rho = 0.5$.

and each of their corresponding eigen-coding effect. For a single factor it is also easy to verify numerically that the correlation between a polynomial effect and its corresponding eigen-coding effect is very high. For reference, through numerical studies it can be shown that a single evenly-spaced four-level factor with a correlation matrix like (8), the correlation between a polynomial effect and its corresponding eigen-coded effect is greater than 0.97 for all $\rho \in (0, 1)$.

5. EXAMPLES

We need the following notation. Let $\boldsymbol{D}$ be the design matrix, which has n rows and p columns corresponding to the p factors and $\boldsymbol{y} = (y_1, \cdots, y_n)'$ be the response values obtained from the experiment. Let $\boldsymbol{U}_D$ be the model matrix generated from D and Ψ_D the corresponding correlation matrix. Let $var(\boldsymbol{\beta}) = \tau_0^2 \boldsymbol{R}$, where the construction of the matrix $\boldsymbol{R}$ was discussed in detail in the previous sections. The examples presented in this section do not have replicates. Because we do not have any information about σ^2, we set $\sigma^2 = 0$. We obtain

$$\widehat{\boldsymbol{\beta}} = \mathbb{E}(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{\tau_0^2}{\sigma_0^2} \boldsymbol{R}\boldsymbol{U}'_D \boldsymbol{\Psi}_D^{-1} (\boldsymbol{y} - \mu_0 \mathbf{1}_n)$$

and

$$var(\boldsymbol{\beta}|\boldsymbol{y}) = \tau_0^2 \left(\boldsymbol{R} - \frac{\tau_0^2}{\sigma_0^2} \boldsymbol{R}\boldsymbol{U}'_D \boldsymbol{\Psi}_D^{-1} \boldsymbol{U}_D \boldsymbol{R} \right).$$

A general expression for τ_0^2/σ_0^2 is given by

$$\frac{\tau_0^2}{\sigma_0^2} = \frac{\prod_{j=1}^{p} sum(\boldsymbol{\Psi}_j)}{q^2}, \tag{14}$$

where $q = \prod_{i=1}^{p} m_i$ and $sum(\boldsymbol{\Psi}_j)$ denotes the sum of all the elements of the matrix $\boldsymbol{\Psi}_j$. We can calculate the ratios

$$t_i = \frac{|\widehat{\beta}_i|}{\widehat{\sigma}_{\beta_i}},$$

to identify the important effects, where $\widehat{\sigma}^2_{\beta_i}$ is the diagonal element in $var(\boldsymbol{\beta}|\boldsymbol{y})$ corresponding to β_i. The most important effect is the one with the largest t_i. The other important effects can be similarly identified one-by-one using a forward selection strategy as explained in Joseph (2006).

The hyper-parameters can be estimated using empirical Bayes methods. Let $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_p)'$. Then

$$\widehat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} n \log \widehat{\sigma}_0^2 + \log\det(\boldsymbol{\Psi}_D),$$
$$\widehat{\mu}_0 = (\mathbf{1}_n' \boldsymbol{\Psi}_D^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n' \boldsymbol{\Psi}_D^{-1} \boldsymbol{y},$$

and

$$\widehat{\sigma}_0^2 = \frac{1}{n}(\boldsymbol{y} - \widehat{\mu}_0 \mathbf{1}_n)' \boldsymbol{\Psi}_D^{-1} (\boldsymbol{y} - \widehat{\mu}_0 \mathbf{1}_n).$$

For numerical stability, we must put some mild constraints on the feasible region of $\boldsymbol{\rho}$ in the above optimization, such as $\rho_i \in [0, 0.99]$. We could have instead implemented the penalized likelihood recommendations from Li and Sudjianto (2005). There are some additional considerations in the empirical Bayes step for estimating $\boldsymbol{\rho}$. It is important to obtain the constrained *global* optimum. Most software will converge on some local optima. We employ a naive approach to global optimization and implement a sequence of local optimizations over randomly generated initial values, choosing the best local optimum as the global optimum. We caution that it is possible to begin the algorithm with a value for $\boldsymbol{\rho}$ that is not a true global optimum due to either the precautions taken to prevent inverting an ill-conditioned $\boldsymbol{\Psi}$ matrix, or by not being able to pragmatically do an exhaustive search of the feasible region for all of the local optima. The $\boldsymbol{\rho}$ is estimated only at step 0 of the forward selection procedure. We use this estimate for each subsequent step. From this estimate of $\boldsymbol{\rho}$, we are able to calculate the factor τ_0^2/σ_0^2, as well as the matrices $\boldsymbol{R}$ and $\boldsymbol{\Psi}_D$ used in the calculations at all later steps. The first example illustrates a situation where the matrix $\boldsymbol{R}$ has nonzero off-diagonal elements. This matrix can be constructed through the explicit matrix calculations suggested by Theorem 1. However, we found that the diagonal approximation to $\boldsymbol{R}$ in this example is adequate for discovering the first few important effects. In the second example, $\boldsymbol{R}$ is a diagonal matrix. So for this example constructing $\boldsymbol{R}$ is a simple matter, where the diagonal elements of $\boldsymbol{R}$ correspond to the appropriate factor calculation preceding each of the propositions and entered into the matrix $\boldsymbol{R}$ in the order the effects appear as columns of $\boldsymbol{U}_D$.

Here we emphasize the ease with which the methodology of Joseph (2006) is extended beyond two-level experiments. In addition, we stress that very often the procedure is entirely automatic, yielding no ambiguity in situations where the traditional frequentist approach would require deeper attention. Finally, we note that there may exist situations where the iterative procedure is unnecessary. That is, a quick proxy to the forward selection may be to use a half-normal plot to select effects in Step 0. This technique seems to produce results equivalent to the forward selection when the significant effects after k steps form a projection of the factor space onto a lower dimensional, but orthogonal factor space.

5.1 Blood Glucose Experiment

Hamada and Wu (1992) analyzed an experiment designed to study blood glucose reading levels from a testing device. In this experiment, there was one two-level factor and seven three-level factors (Table 1). The three-level factors were all considered to be quantitative factors. These factors did not all have evenly spaced levels, but they were approximately evenly spaced. The design was a nonregular fraction of a $2^1 \times 3^7$ design, the 18-run design popularized by Taguchi (1987). The design and the data are given in Table 2.

Table 1: Factors and Levels, the Blood Glucose Experiment

Factor		Level 1	2	3
A.	wash	no	yes	
B.	microvial volume (ml)	2.0	2.5	3.0
C.	caras H_2O level (ml)	20	28	35
D.	centrifuge RPM	2100	2300	2500
E.	centrifuge time (min)	1.75	3	4.5
F.	(sensitivity, absorption)	(0.10,2.5)	(0.25,2)	(0.50,1.5)
G.	temperature (^{0}C)	25	30	37
H.	dilution ratio	1:51	1:101	1:151

In the frequentist analysis, it is computationally cumbersome to entertain all of the 4,374 possible factorial effects. Therefore, we consider only the main effects and two-factor

Table 2: Design Matrix and Data, the Blood Glucose Experiment

	Factor								Mean
Run	A	G	B	C	D	E	F	H	Reading
1	1	1	1	1	1	1	1	1	97.94
2	1	1	2	2	2	2	2	2	83.40
3	1	1	3	3	3	3	3	3	95.88
4	1	2	1	1	2	2	3	3	88.86
5	1	2	2	2	3	3	1	1	106.58
6	1	2	3	3	1	1	2	2	89.57
7	1	3	1	2	1	3	2	3	91.98
8	1	3	2	3	2	1	3	1	98.41
9	1	3	3	1	3	2	1	2	87.56
10	2	1	1	3	3	2	2	1	88.11
11	2	1	2	1	1	3	3	2	83.81
12	2	1	3	2	2	1	1	3	98.27
13	2	2	1	2	3	1	3	2	115.52
14	2	2	2	3	1	2	1	3	94.89
15	2	2	3	1	2	3	2	1	94.70
16	2	3	1	3	2	3	1	2	121.62
17	2	3	2	1	3	1	2	3	93.86
18	2	3	3	2	1	2	3	1	96.10

interactions. This analysis identifies the effects B_lH_q, B_qH_q, E_lG_l, AH_q, $\cdots$, as having high explanatory power, which is shown in Figure 2(b). Thus, the frequentist approach does not lead to a model satisfying effect hierarchy or effect heredity.

By contrast, the proposed methodology does respect effect hierarchy and effect heredity, and is able to entertain all of the factorial effects. In step 0 of the Bayesian forward selection, the empirical Bayes estimate of the correlation matrix parameters is given by the vector,

$$\hat{\boldsymbol{\rho}} = (0.93, 0.00, 0.99, 0.99, 0.98, 0.98, 0.99, 0.00)'.$$

Figure 3(a) shows the half-normal plot of the t_i ratios at this step. We can see that B_lH_q is the most significant effect. After selecting this effect and continuing with the forward selection, we identify the effects B_qH_q, B_l, B_q, $\cdots$ as having high explanatory power. This is shown in the R^2-plot in Figure 3(b).

Of course, the principles of effect hierarchy and effect heredity can be enforced in the frequentist forward selection through some modifications, such as the strategy presented in Hamada and Wu (1992). However, we believe that the Bayesian strategy is more elegant and efficient. For example, if a three-factor interaction effect is significant, the frequentist analysis will miss it, but the Bayesian analysis will identify it with high probability. Indeed, the Bayesian analysis seems to be more powerful than the frequentist analysis, as can be seen by comparing the half-normal plots of both of the analyses at step 0; that is, all of the significant effects can be identified even at step 0 of the Bayesian analysis. In the next section, we provide an example where the frequentist analysis fails, but the Bayesian analysis succeeds.

Implementation of the Bayesian methodology of Chipman et al. (1997) was also illustrated through this example. One of the most significant differences in the Bayesian methodology presented here versus that of Chipman et al. (1997) is how the prior belief in effect heredity is incorporated into the model. In the procedure presented above, specification of effect heredity is through the parameter space as a consequence of our functional prior assumption. In Chipman et al. (1997), effect heredity is reflected through prior specification in the model space. In the methodology presented here, effect heredity was a direct consequence of the

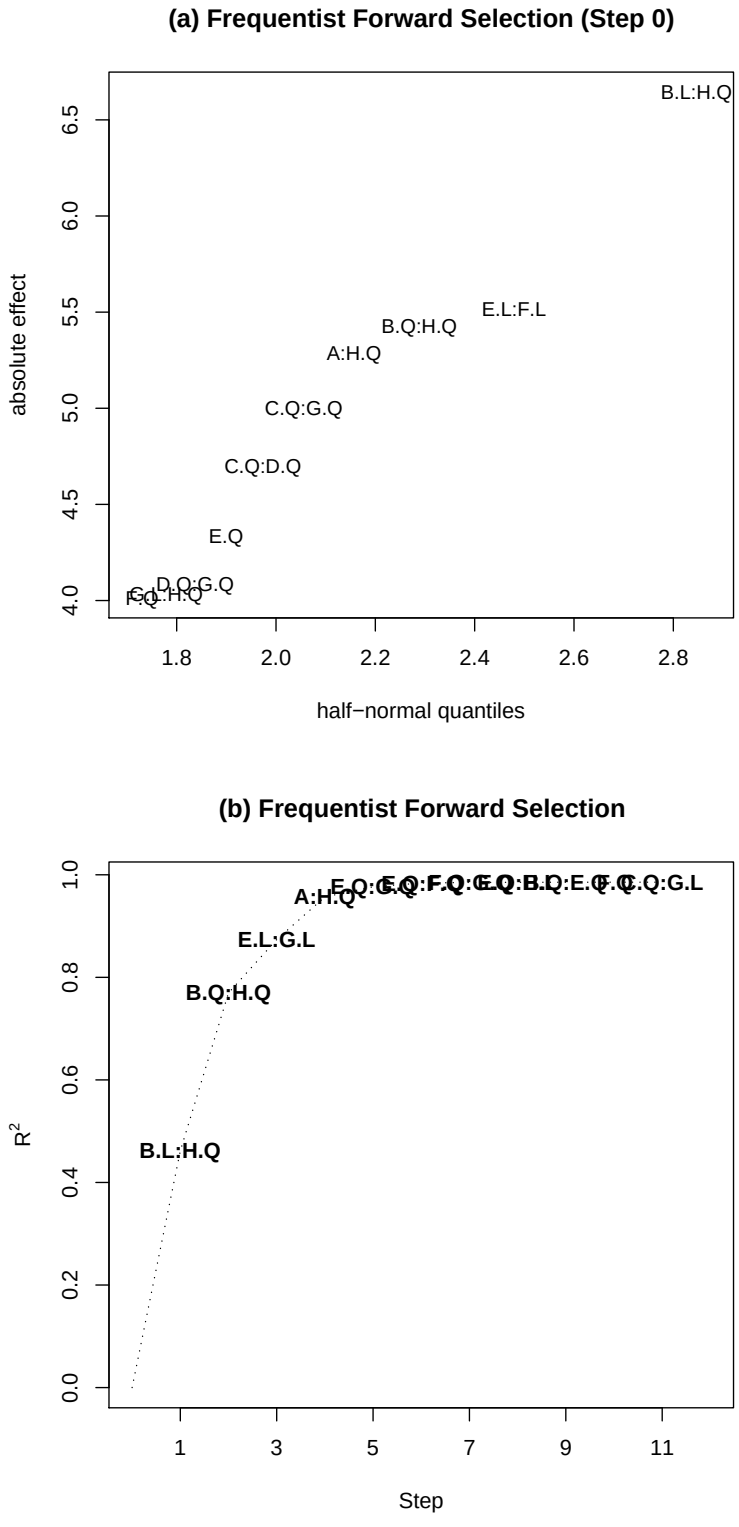
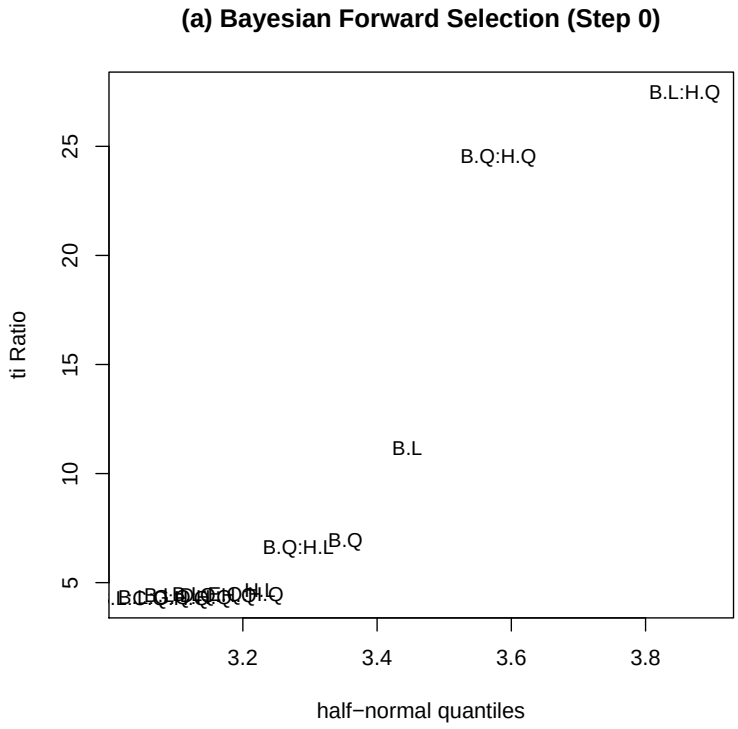


Figure 2: Frequentist Analysis of the Blood Glucose Experiment



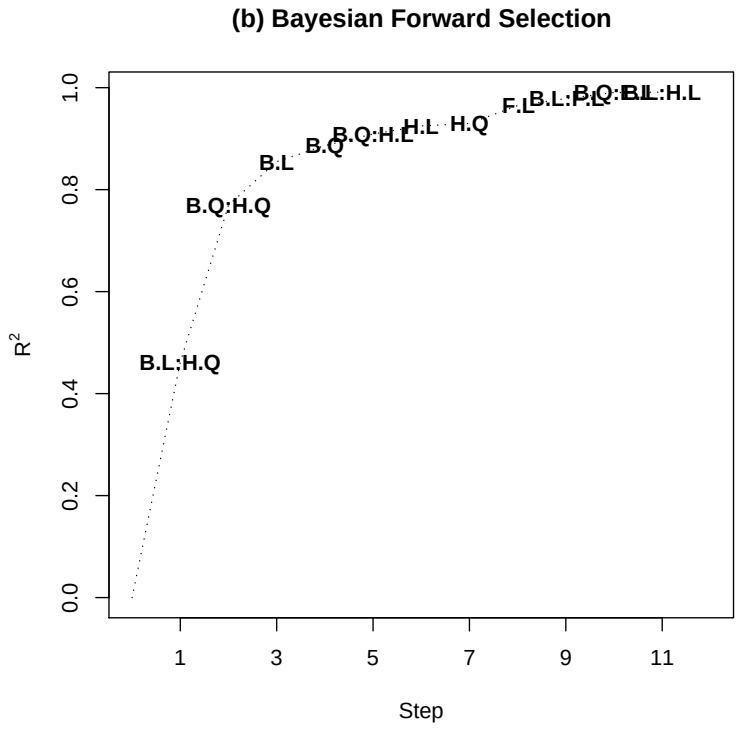


Figure 3: Bayesian Analysis of the Blood Glucose Experiment

functionally induced prior on $\boldsymbol{\beta}$, whereas in Chipman et al. (1997) hierarchical priors on all subset models had to be specified in order to incorporate prior beliefs about heredity. The technique of Chipman et al. (1997) does offer the advantage of great flexibility in enabling the incorporation of other possible *a priori* beliefs about relationships between effects through adding on to the hierarchical prior structure. However, the procedure described in this paper is a fairly automatic methodology that quite naturally imposes effect hierarchy and effect heredity. Moreover, the extension of the prior specification to include three and higher order interactions, cubic, fourth order terms, etc. is more difficult to implement with the hierarchical priors compared with the functionally induced priors.

5.2 Router Bit Experiment

Phadke (1989) reported on an experiment designed to help improve the lifetime of a router bit used to cut printed circuit boards. This experiment was also analyzed by Wu and Hamada (2000). The experiment is an unreplicated fraction of a $2^7 \times 4^2$ design. The factors and levels are shown in Table 3 and the design and data are given in Table 4. There are only 32 runs and 2,048 possible effects to consider. The two four-level factors: (D) "bit type" and (E) "spindle position" are treated as qualitative factors. So in analyzing this experiment, we have two types of factors: seven two-level factors and two qualitative four-level factors.

Table 3: Factors and Levels, the Router Bit Experiment

Factor		Level			
A.	suction (in of Hg)	1	2		
B.	x-y feed (in/min)	60	80		
C.	in-feed (in/min)	10	50		
D.	bit type	1	2	3	4
E.	spindle position	1	2	3	4
F.	suction foot	SR	BB		
G.	stacking height (in)	3/16	1/4		
H.	Slot depth (mils)	60	100		
J.	speed (rpm)	30000	40000		

Table 4: Design Matrix and Data, the Router Bit Experiment

	Factor									
Run	A	B	C	D	E	F	G	H	J	Lifetime
1	−	−	−	1	1	−	−	−	−	3.5
2	−	−	−	2	2	+	+	−	−	0.5
3	−	−	−	3	4	−	+	+	−	0.5
4	−	−	−	4	3	+	−	+	−	17.0
5	−	+	+	3	1	+	+	−	−	0.5
6	−	+	+	4	2	−	−	−	−	2.5
7	−	+	+	1	4	+	−	+	−	0.5
8	−	+	+	2	3	−	+	+	−	0.5
9	+	−	+	4	1	−	+	+	−	17.0
10	+	−	+	3	2	+	−	+	−	2.5
11	+	−	+	2	4	−	−	−	−	0.5
12	+	−	+	1	3	+	+	−	−	3.5
13	+	+	−	2	1	+	−	+	−	0.5
14	+	+	−	1	2	−	+	+	−	2.5
15	+	+	−	4	4	+	+	−	−	0.5
16	+	+	−	3	3	−	−	−	−	3.5
17	−	−	−	1	1	−	−	−	+	17.0
18	−	−	−	2	2	+	+	−	+	0.5
19	−	−	−	3	4	−	+	+	+	0.5
20	−	−	−	4	3	+	−	+	+	17.0
21	−	+	+	3	1	+	+	−	+	0.5
22	−	+	+	4	2	−	−	−	+	17.0
23	−	+	+	1	4	+	−	+	+	14.5
24	−	+	+	2	3	−	+	+	+	0.5
25	+	−	+	4	1	−	+	+	+	17.0
26	+	−	+	3	2	+	−	+	+	3.5
27	+	−	+	2	4	−	−	−	+	17.0
28	+	−	+	1	3	+	+	−	+	3.5
29	+	+	−	2	1	+	−	+	+	0.5
30	+	+	−	1	2	−	+	+	+	3.5
31	+	+	−	4	4	+	+	−	+	0.5
32	+	+	−	3	3	−	−	−	+	17.0

The coding scheme that we used for the four-level factors is the Wu-Hamada recommendation highlighted in a previous section. Here, those main effects are labeled D_1, D_2, D_3 and E_1, E_2, E_3. Figure 4(a) shows the half-normal plot from a traditional analysis. The effects D_2, G, J, GJ and AF appear to be significant. Note that each of them represents a set of aliased effects. Assuming three and higher order interactions are negligible, one can show that

$$AF = -D_2H = -CE_2 = BD_3 = D_1E_3 = E_1G \tag{15}$$

and

$$D_2 = AG = BE_3 = E_1F,$$

whereas the effects G, J, and GJ are clear (not aliased with any of the main effects or two-factor interactions). Follow-up experiments can be used to de-alias the above effects (see, e.g, Meyer, Steinberg, and Box 1996). An alternative to running a follow-up experiment is the approach in Wu and Hamada (2000) which appeals to the widely accepted principles of effect hierarchy and effect heredity. In that analysis, effect hierarchy was manually applied to the aliasing relationships to select the main effect D_2 as opposed to one of the two-factor interactions with which it is aliased. Similarly, effect heredity was used to justify selecting either the interaction D_2H or E_1G as opposed to the other four two-factor interaction effects. However neither of these two principles enable breaking the tie between D_2H and E_1G. Wu and Hamada (2000) argued that because the four spindles are synchronized, the effect of G should not vary substantially with the spindle position; thus ruling out the E_1G interaction, so that D_2H was the effect identified as significant.

In Step 0 of the proposed method, we obtain the empirical Bayes estimates of $\boldsymbol{\rho}$,

$$\hat{\boldsymbol{\rho}} = (0.99, 0.99, 0.99, 0.71, 0.99, 0.99, 0.60, 0.09, 0.56)'.$$

Figure 4(b) shows that the effects J, GJ, D_2, HJ, D_2H, G, and GHJ seem to be significant, which are the same as the first seven effects identified by the Bayesian forward selection strategy. Note that in the Bayesian analysis no confusion is created by the aliasing relationships. For example, at step 0, the t_i ratios for the effects in (15) are: $t_{AF} = 0.14$, $t_{D_2H} = 42.33$,

$t_{CE_2} = 0.10$, $t_{BD_3} = 0.61$, $t_{D_1E_3} = 0.43$, and $t_{E_1G} = 0.70$. Thus D_2H stands out very clearly from the others as the significant effect. This could not be achieved using the frequentist analysis. Wu and Hamada (2000) were able to choose D_2H but only after applying expert knowledge of the process. Whereas the Bayesian approach was able to identify this effect through mere data analysis. We also note that a reasonable, potentially significant three-factor interaction is identified by the Bayesian analysis, which was not even possible in the frequentist analysis.

By proposing the aforementioned Bayesian analysis, we are not trying to discourage the use of follow-up experiments. If a decision has to be made based on a one-shot experiment, the Bayesian analysis will be able to provide a unique answer. On the other hand, if resources do exist to perform follow-up experiments, then even in this situation, this type of Bayesian analysis can yield very useful information. For example, based on the t_i ratios, we can order the effects in (15): $D_2H \succ E_1G \succ BD_3 \succ D_1E_3 \succ AF \succ CE_2$. This ordering is immensely helpful for the optimal choice of follow-up runs. In frequentist analysis all six of these effects would be viewed as equally important and thus some of these additional resources will be spent on de-aliasing unimportant effects.

6. CONCLUSIONS

Typically, frequentist methods in the analysis of three and higher level experiments require significant work to resolve ambiguities. First the analyst, must identify the aliasing relationships. In the case of a regular fraction, this will enable the analyst to make variable selection decisions based on the well known principles of effect sparsity, effect hierarchy, and effect heredity. After analyzing the data, there may still be the need to run a follow-up experiment to resolve issues that arise from aliasing. In nonregular designs, such as the 18-run designs, the traditional approach would only consider estimating main effects, because of complex aliasing. Modern techniques such as those presented in Wu and Hamada (2000), while adequate in extracting more information from these designs, do not lend themselves well to being an automatic procedure. Here we have extended the use of functionally induced priors

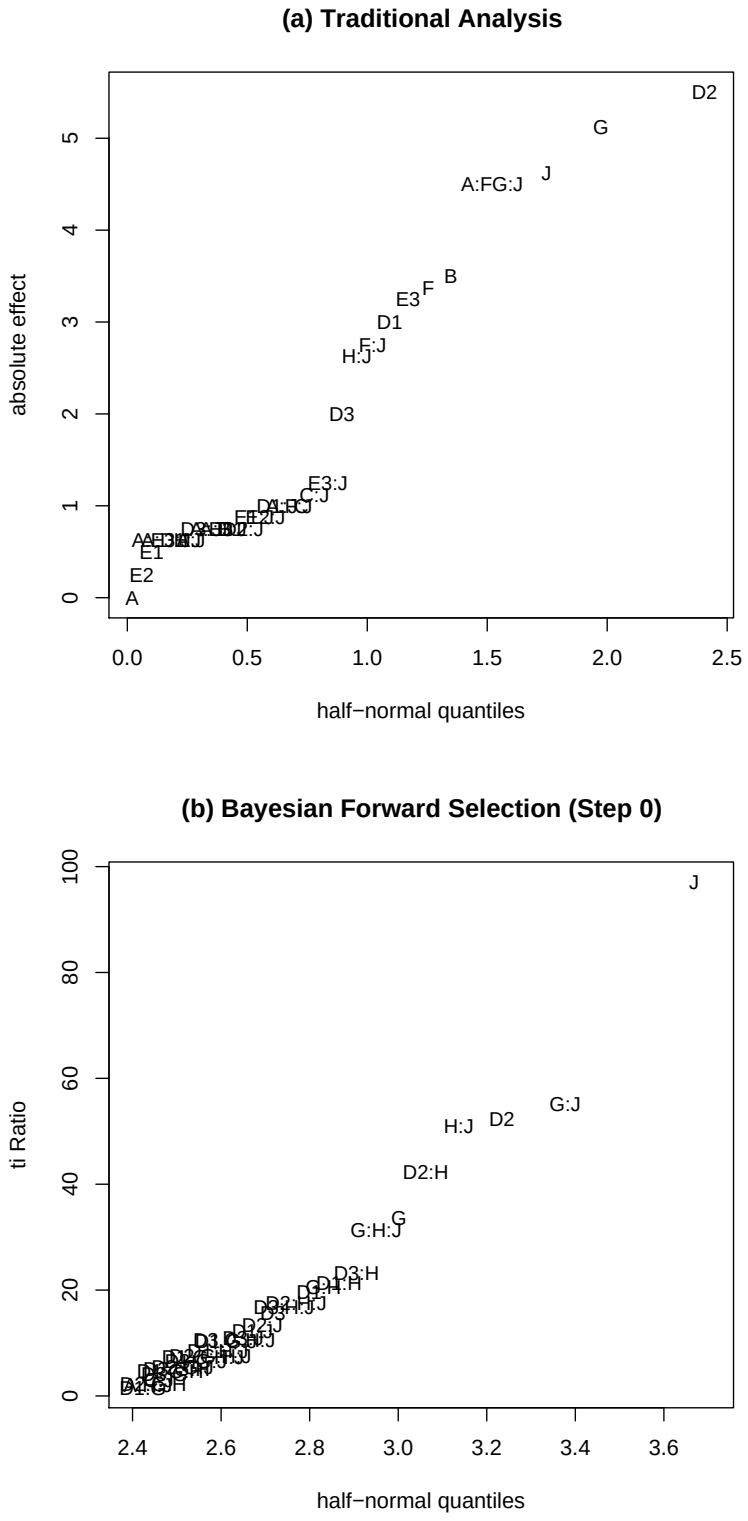


Figure 4: Analysis of the Router Bit Experiment

to designs that involve three-level and four-level factors. From this exposition, the procedure for extending the ideas for fractions of factorials not directly addressed here should be obvious. These tools provide a major step toward a reasonable fully automatic procedure for analyzing experimental data. Not only are the procedures well grounded in theory that facilitate the above mentioned principles of analysis of experiments, but the procedures are easy to implement and yield credible empirical results.

In the general framework, a Gaussian process over the design space induces a joint prior distribution for the linear model's parameters. From this, some additional assumptions about experimental design can be validated. Yet two effect ordering principles for three-level designs from Cheng and Ye (2005) could be challenged as a consequence of the theory here. We could be more specific about when the ordering assumptions are valid and explain why. Moreover new rules can be obtained when the assumptions are not valid.

We make a distinction between *qualitative* factors and *quantitative* factors. This becomes important with three-level and higher designs. We also provide a consistent and logical way of addressing this distinction through the specification of the correlation function that partially characterizes the underlying Gaussian Process. This approach fits into the Gaussian process functional prior framework seamlessly.

In our discussion of the examples, we note that some designs might be more likely than others to produce ambiguities in variable selection. When these arise, they could either be resolved manually or resolved through a simple automated procedure that respects the principles of effect sparsity, effect hierarchy and effect heredity. When different components of the $\boldsymbol{\rho}$ vector are used for each factor, this issue becomes increasingly less likely to be a concern.

Proofs

Proof of Theorem 1

We first need to establish a general result concerning the construction of our correlation matrices over all p factors. Observe that for the following result to hold, that we can define the $(m_1 m_2 \cdots m_p) \times (m_1 m_2 \cdots m_p)$ full factorial model matrix over all p factors, $\boldsymbol{U}$, using whatever coding scheme we desire. Suppose we construct our full model matrix via a Kronecker product of the individual factor model matrices, taken in increasing order of the frequency the levels change. The run order for the full factorial design corresponds to one where the first factor's levels are changing the slowest and the p^{th} factor's levels are changing the quickest:

$$\boldsymbol{U} = \boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \otimes \cdots \otimes \boldsymbol{U}_p = \bigotimes_{j=1}^{p} \boldsymbol{U}_j.$$

Now, let $\boldsymbol{\Psi}$ denote the correlation matrix corresponding to the full factorial design over all p factors. The $m_j \times m_j$ correlation matrix corresponding to factor j denoted by $\boldsymbol{\Psi}_j$ will have the general structure of a symmetric Toeplitz matrix due to the stationarity assumption imposed on the Gaussian process in each factor:

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \psi_j(1) & \dots & \psi_j(m_j - 1) \\ \psi_j(1) & 1 & \ddots & \psi_j(m_j - 2) \\ \vdots & \ddots & \ddots & \vdots \\ \psi_j(m_j - 1) & \psi_j(m_j - 2) & \dots & 1 \end{pmatrix}. \tag{16}$$

Then $\boldsymbol{\Psi}$ has a convenient block symmetric structure. Let the matrix $\boldsymbol{\Psi}_{(i)}$ represent the correlation matrix for the full factorial design over the last $p-i$ factors. Then, since the first factor's levels are changing the slowest, and we have assumed a product correlation function structure, the correlation matrix has the following block form:

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{(1)} & \psi_1(1)\Psi_{(1)} & \dots & \psi_p(m_1 - 1)\boldsymbol{\Psi}_{(1)} \\ \psi_1(1)\boldsymbol{\Psi}_{(1)} & \boldsymbol{\Psi}_{(1)} & \ddots & \psi_1(m_1 - 2)\boldsymbol{\Psi}_{(1)} \\ \vdots & \ddots & \ddots & \vdots \\ \psi_1(m_1 - 1)\boldsymbol{\Psi}_{(1)} & \psi_1(m_1 - 2)\boldsymbol{\Psi}_{(1)} & \dots & \boldsymbol{\Psi}_{(1)} \end{pmatrix}, \tag{17}$$

where each of the blocks are $(m_2 m_3 \cdots m_p) \times (m_2 m_3 \cdots m_p)$. This matrix follows from the fact that the first $(m_2 m_3 \cdots m_p)$ runs in the full factorial design only differ among the last $p-1$ factors, in the same way the full factorial design with $p-1$ factors varies. Each run in the second block of $(m_2 m_3 \cdots m_p)$ runs differs from the first run in the full factorial design by one level in the first factor, and then in the same way as the full factorial design differs among the last $p-1$ factors, etc. Hence $\boldsymbol{\Psi} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_{(1)}$. Noting that $\boldsymbol{\Psi}_{(p-1)} = \boldsymbol{\Psi}_p$, we obtain

$$\boldsymbol{\Psi} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_{(1)} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_{(2)} = \cdots = \boldsymbol{\Psi}_1 \otimes \cdots \otimes \boldsymbol{\Psi}_p = \bigotimes_{j=1}^{p} \boldsymbol{\Psi}_j.$$

Now through the properties of the Kronecker product operator, we can prove Theorem 1:

$$\begin{aligned}
var(\boldsymbol{\beta}) &= \sigma_0^2 \boldsymbol{U}^{-1} \boldsymbol{\Psi} (\boldsymbol{U}^{-1})' \\
&= \sigma_0^2 (\bigotimes_{j=1}^{p} \boldsymbol{U}_j)^{-1} \bigotimes_{j=1}^{p} \boldsymbol{\Psi}_j \left((\bigotimes_{j=1}^{p} \boldsymbol{U}_j)^{-1} \right)' \\
&= \sigma_0^2 \bigotimes_{j=1}^{p} \boldsymbol{U}_j^{-1} \bigotimes_{j=1}^{p} \boldsymbol{\Psi}_j \left(\bigotimes_{j=1}^{p} \boldsymbol{U}_j^{-1} \right)' \\
&= \sigma_0^2 \bigotimes_{j=1}^{p} \boldsymbol{U}_j^{-1} \bigotimes_{j=1}^{p} \boldsymbol{\Psi}_j \bigotimes_{j=1}^{p} (\boldsymbol{U}_j^{-1})' \\
&= \sigma_f^2 \bigotimes_{j=1}^{p} \boldsymbol{U}_j^{-1} \boldsymbol{\Psi}_j (\boldsymbol{U}_j^{-1})'.
\end{aligned}$$

♢

Proof of Equation (5)

We have $\boldsymbol{\Psi}_j = (1-\rho_j)\boldsymbol{I}_{m_j} + \rho_j \boldsymbol{J}_{m_j}$, where $\boldsymbol{J}_{m_j}$ is a $m_j \times m_j$ square matrix of 1's. Then:

$$\begin{aligned}
\boldsymbol{U}_j' \boldsymbol{\Psi}_j \boldsymbol{U}_j &= (1-\rho_j) \boldsymbol{U}_j' \boldsymbol{U}_j + \rho_j \boldsymbol{U}_j' \boldsymbol{J}_{m_j} \boldsymbol{U}_j \\
&= m_j (1-\rho_j) I_{m_j} + \rho_j \boldsymbol{U}_j' \begin{pmatrix} m_j & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ m_j & 0 & \cdots & 0 \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= m_j(1-\rho_j)I_{m_j} + \rho_j \begin{pmatrix} m_j^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \\
&= m_j \begin{pmatrix} 1-\rho_j+m_j\rho_j & 0 & \cdots & 0 \\ 0 & 1-\rho_j & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1-\rho_j \end{pmatrix}.
\end{aligned}
$$

$\diamondsuit$

Proof of Equation (14)

Let $\boldsymbol{\Sigma} = \boldsymbol{U}^{-1}\boldsymbol{\Psi}(\boldsymbol{U}^{-1})'$. Then $var(\beta_0) = \tau_0^2 = \sigma_0^2\boldsymbol{\Sigma}_{1,1}$. But since $\boldsymbol{U}$ is the orthogonal full factorial model matrix with the leading column of $\boldsymbol{U}$ assumed to be $\mathbf{1}_q$, we have that $\boldsymbol{U}^{-1}$ has as its leading row: $\frac{1}{q}\mathbf{1}_q{}'$. So that:

$$
\frac{\tau_0^2}{\sigma_0^2} = \frac{\mathbf{1}_q'\boldsymbol{\Psi}\mathbf{1}_q}{q^2} = \frac{sum(\boldsymbol{\Psi})}{q^2} = \frac{sum(\bigotimes_{j=1}^p \boldsymbol{\Psi}_j)}{q^2} = \frac{\prod_{j=1}^p sum(\boldsymbol{\Psi}_j)}{q^2}.
$$

$\diamondsuit$

References

Chaloner, K. and Verdinelli, I. (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 273–304.

Cheng, S.-W. and Ye, K. (2005), "Geometric Isomorphism and Minimum Aberration," *Annals of Statistics*, 32, 2168–2185.

Chipman, H., Hamada, M., and Wu, C. F. J. (1997), "A Bayesian Variable-Selection Approach for Analyzing Designed Experiments With Complex Aliasing," *Technometrics*, 39, 372–381.

Hamada, M. and Wu, C. F. J. (1992), "Analysis of Designed Experiments With Complex Aliasing," *Journal of Quality Technology*, 24, 130–137.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York: Springer.

Joseph, V. R. (2006), "A Bayesian Approach to the Design and Analysis of Fractionated Experiments," *Technometrics*, 48, 219–229.

Kerr, M. K. (2001), "Bayesian Optimal Fractional Factorials," *Statistica Sinica*, 11, 605–630.

Li, R. and Sudjianto, A. (2005), "Analysis of Computer Experiments Using Penalized Likelihood in Gaussian Kriging Models," *Journal of the American Statistical Association*, 47, 111–120.

Meyer, R. D., Steinberg, D. M., and Box, G. (1996), "Follow-up Designs to Resolve Confounding in Multifactor Experiments, (with discussion)," *Technometrics*, 38, 303–332.

Mitchell, T. J., Morris, M. D., and Ylvisaker, D. (1995), "Two Level Fractional Factorials and Bayesian Prediction," *Statistica Sinica*, 5, 559–573.

Phadke, M. S. (1989), *Quality Engineering Using Robust Design*, Englewood Cliffs, NJ: Prentice Hall.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer.

Steinberg, D. M. and Bursztyn, D. (2004), "Data Analytic Tools for Understanding Random Field Regression Models," *Technometrics*, 46, 411–420.

Taguchi, G. (1987), *System of Experimental Design, Vol. 1 & Vol. 2*, White Plains, New York: Unipub/Kraus International.

Taguchi, G., Chowdhury, S., and Taguchi, S. (2000), *Robust Engineering*, New York:McGraw-Hill.

Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.

Wu, C. F. J. and Zhang, R. (1993), "Minimum Aberration Designs with Two-Level and Four-Level Factors," *Biometrika*, 80, 203–209.