Biometrika (2002), **89,** 4, *pp.* 893–904

Printed in Great Britain

Uniform designs limit aliasing

By FRED J. HICKERNELL

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China

fred@hkbu.edu.hk

and MIN-QIAN LIU

School of Management, Tianjin University, Tianjin 300072, China

mqliu@263.net

Summary

When fitting a linear regression model to data, aliasing can adversely affect the estimates of the model coefficients and the decision of whether or not a term is significant. Optimal experimental designs give efficient estimators assuming that the true form of the model is known, while robust experimental designs guard against inaccurate estimates caused by model misspecification. Although it is rare for a single design to be both maximally efficient and robust, it is shown here that uniform designs limit the effects of aliasing to yield reasonable efficiency and robustness together. Aberration and resolution measure how well fractional factorial designs guard against the effects of aliasing. Here it is shown that the definitions of aberration and resolution may be generalised to other types of design using the discrepancy.

Some key words: Discrepancy; Effects aliasing; Fractional factorial design; Minimum aberration; Orthogonal array; Reproducing kernel Hilbert space; Resolution.

1. Introduction

Efficiency and robustness are important concepts in the design of experiments. If one knows the form of the model relating the response to the factors, then optimal designs lead to efficient estimators of the unknown parameters. However, often one must infer the form of the model from the data using regression diagnostics. In such cases aliasing can adversely affect model selection and estimation if the experiment is not well designed. Thus, the robustness of a design may be important as well.

Uniform designs (Wang & Fang, 1981; Fang & Wang, 1994, Ch. 5; Bates et al., 1996) are model independent and spread experimental points evenly over the domain. Their goal is to minimise the discrepancy, introduced by Weyl (1916), which measures the difference between the empirical distribution of the design and the uniform distribution. Theorem 1 shows that uniform designs limit the adverse effects of aliasing on efficiency and robustness under general assumptions.

Regular fractional factorial designs are often constructed to be of maximum resolution (Box et al., 1978, Ch. 12) and minimum aberration (Fries & Hunter, 1980), since these criteria limit the adverse effects of aliasing. Aberration has been generalised to nonregular

fractional factorial designs (Deng & Tang, 1999; Tang & Deng, 1999; Ma & Fang, 2001; Tang, 2001; Xu & Wu, 2001). Definition 2 and Theorem 2 show that aberration may be further generalised to cover designs that are not fractional factorial by using the discrepancy. It is also shown in Theorem 3 that minimum aberration designs and minimum discrepancy designs are equivalent in a certain limit.

The remainder of this section introduces the model and the design problem to be considered. Section 2 describes how to define a Hilbert space of possible response functions, and the discrepancy arising from that Hilbert space is defined in § 3. The ways in which low discrepancy designs limit the effects of aliasing are discussed in §§ 4 and 5.

Suppose that an experiment has s factors, and let $\mathscr{X}_j$ be a measurable set of all possible levels of the jth factor. Common examples of $\mathscr{X}_j$ are $\{0, \ldots, q_j - 1\}$ and $[0, 1]$. The experimental region, $\mathscr{X}$, is some measurable subset of $\mathscr{X}_1 \times \ldots \times \mathscr{X}_s$. An experimental design with n points, $P = \{z_i = (z_{i1}, \ldots, z_{is}) : i = 1, \ldots, n\}$, is a subset of $\mathscr{X}$ with multiple copies of the same point allowed.

Let y_i denote the observed response when the factors take on the value z_i. A linear regression model with misspecification may be written as

$$y_i = f(z_i) + \varepsilon_i = \sum_{j=1}^{p} g_j(z_i)\theta_j + h(z_i) + \varepsilon_i \quad (i = 1, \ldots, n),$$

where the specified, possibly complex-valued, functions g_j are linearly independent, $h(x)$ is an unknown deviation or misspecification, $f(x) = g^{\mathrm{T}}(x)\theta + h(x)$, the ε_i are independent and identically distributed random errors with mean 0 and variance σ^2, and $^{\mathrm{T}}$ denotes the transpose. Moreover, the true parameter vector is $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$, and $g(x) = (g_1(x), \ldots, g_p(x))^{\mathrm{T}}$, where $x = (x_1, \ldots, x_s)$. To ensure that the model is identifiable, θ is assumed to satisfy $G\theta = \mathscr{I}(\bar{g}f)$, where $G = \mathscr{I}(\bar{g}g^{\mathrm{T}})$, and so the misspecification satisfies $\mathscr{I}(\bar{g}h) = (0, \ldots, 0)^{\mathrm{T}}$. Here $^{-}$ denotes the complex conjugate, $\mathscr{I}(f) := \int_{\mathscr{X}} f(x)\, dF_{\mathscr{X}}(x)$, and $F_{\mathscr{X}}$ denotes the uniform distribution over the experimental domain $\mathscr{X}$. The matrix G is Hermitian, that is $G^H = G$, where ^{H} denotes the complex-conjugate of the transpose.

The ordinary least squares estimator of the parameter θ is $\hat{\theta} = M^{-1}X^H y$, where $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ is the vector of response data, the design matrix is $X = (g(z_1), \ldots, g(z_n))^{\mathrm{T}}$, and the information matrix, $M = X^H X$, is assumed to be nonsingular. The mean squared difference between the fitted response function, $\hat{f}(x) = g^{\mathrm{T}}(x)\hat{\theta}$, and the true one is

$$E\{\mathscr{I}(|f - \tilde{f}|^2)\} = \mathscr{I}(\{E|g^H(\theta - \hat{\theta})|^2 + |h|^2\}) = \mathrm{var}(P, g) + \mathrm{bias}(P, g, h) + \mathscr{I}(|h|^2) \quad (1)$$

(Yue & Hickernell, 1999), where

$$\mathrm{var}(P, g) = \sigma^2 \,\mathrm{tr}(M^{-1}G), \quad \mathrm{bias}(P, g, h) = \eta^H X M^{-1} G M^{-1} X^H \eta$$

and $\eta = (h(z_1), \ldots, h(z_n))^{\mathrm{T}}$. The last term in (1) is independent of the design and can only be made smaller by a better choice of model.

The variance of the estimated model, $\mathrm{var}(P, g)$, is independent of the misspecification. Here L-optimal designs maximise efficiency by minimising $\mathrm{tr}(M^{-1}G)$, which could be arbitrarily large because of aliasing for a poorly chosen design.

The bias of the estimated model, $\mathrm{bias}(P, g, h)$, is independent of the noise and arises from the difference between θ and $\hat{\theta}$ because of the aliasing of h and g. Aliasing may cause a term $\theta_j g_j(x)$ to appear statistically significant and to be retained in the model, when it does not belong. Variable selection techniques will not solve this problem. One must use a more robust experimental design.

Yue & Hickernell (1999) investigated the relative importance of the terms $\mathrm{var}(P, g)$ and

bias(P, g, h). They showed that variance-minimising designs can yield substantial bias, whereas bias-minimising designs are rather efficient. Moreover, bias-minimising designs tend to spread the points evenly over the domain. A similar conclusion was reached by Box & Draper (1959) for polynomial models.

To obtain a design minimising $\mathrm{var}(P, g) + \mathrm{bias}(P, g, h)$ one must know the model in advance, including the relative size of $|h|$ with respect to σ^2. As a result of the practical difficulties of this approach, the designs considered here are based on the model-independent criterion of discrepancy. Uniform designs are shown to limit the effects of aliasing in the terms $\mathrm{var}(P, g)$ and $\mathrm{bias}(P, g, h)$ for a large class of models and misspecifications: 'model-independent' does not mean 'assumption-independent'. The definition of discrepancy depends on the definition of the space of possible response functions and the definition of the norm defined on $\mathscr{F}$. Thus, one can put any a priori knowledge about the model into the definition of the discrepancy.

2. Reproducing kernel Hilbert spaces of response functions

To make precise statements about the effects of the design on aliasing it is necessary to define the space, $\mathscr{F}$, containing the response, f, and its square. This is assumed to be a separable Hilbert space of complex-valued functions with a reproducing kernel, K (Aronszajn, 1950; Saitoh, 1988; Wahba, 1990, p. 1). This means that the evaluation functional is bounded,

$$K(.,w) \in \mathscr{F}, \quad f(w) = \langle f, K(.,w)\rangle_{\mathscr{F}} \quad (\text{for all } w \in \mathscr{X}, f \in \mathscr{F}).$$

If $\{\phi_\nu\}$ is any orthonormal basis for $\mathscr{F}$, then the reproducing kernel may be written as $K(x, w) = \sum_\nu \phi_\nu(x)\bar{\phi}_\nu(w)$.

Example 1. Suppose that there is one factor with q levels, $\mathscr{X} = \{0, \ldots, q-1\}$, and $\mathscr{F}$ is the set of all functions on $\mathscr{X}$. Then the quantity $\mathscr{I}(f) = q^{-1}\sum_{x\in\mathscr{X}} f(x)$ is the average value of f over the q levels. Define the norm on this space as $\|f\|_{\mathscr{F}} = \langle f, f\rangle_{\mathscr{F}}^{1/2}$, in terms of the following weighted $\mathscr{L}_2$-inner product:

$$\langle f, g\rangle_{\mathscr{F}} = \mathscr{I}(f)\mathscr{I}(\bar{g}) + \frac{1}{\gamma}\mathscr{I}[\{f - \mathscr{I}(f)\}\{\bar{g} - \mathscr{I}(\bar{g})\}] \quad (\text{for all } f, g \in \mathscr{F}).$$

The choice of the arbitrary weight $\gamma > 0$ is discussed in § 5. An orthonormal basis for $\mathscr{F}$ is $\{1, \gamma^{\frac{1}{2}}e^{2\pi\imath x/q}, \gamma^{\frac{1}{2}}e^{4\pi\imath x/q}, \ldots, \gamma^{\frac{1}{2}}e^{2(q-1)\pi\imath x/q}\}$, where $\imath = (-1)^{\frac{1}{2}}$, and the reproducing kernel is

$$K(x, w) = 1 + \gamma\sum_{\nu=1}^{q-1} e^{2\pi\imath\nu(w-x)/q} = 1 + \gamma(-1 + q\delta_{xw}),$$

where δ_{xw} denotes the Kronecker delta function.

Example 2. Suppose that there is one factor with a continuous range, $\mathscr{X} = [0, 1]$. The space of all square-integrable functions does not admit a reproducing kernel because the evaluation functional is not bounded. However, one may choose $\mathscr{F}$ to be the space of functions whose first derivatives are square integrable. The inner product and reproducing kernel for this space are

$$\langle f, g\rangle_{\mathscr{F}} = f(1)\bar{g}(1) + \frac{1}{\gamma}\mathscr{I}(f'\bar{g}') \quad (\text{for all } f, g \in \mathscr{F}),$$

$$K(x, w) = 1 + \gamma\{1 - \max(x, w)\} \quad (\text{for all } x, w \in \mathscr{X} = [0, 1])$$

(Hickernell, 1998a), where f' denotes the derivative of f.

Example 3. In Example 2 $\mathscr{F}$ is infinite-dimensional. It is also possible for $\mathscr{F}$ to be finite-dimensional when $\mathscr{X} = [0, 1]$. Suppose that the response function is known to be linear in x. Since $\mathscr{F}$ must contain the square of the response function, one may choose $\mathscr{F}$ to have $\{1, x, x^2/2\}$ as its orthonormal basis. The inner product and reproducing kernel for this space are

$$\langle f, g \rangle_{\mathscr{F}} = f(0)\bar{g}(0) + f'(0)\bar{g}'(0) + f''(0)g''(0) \quad (\text{for all } f, g \in \mathscr{F}),$$

$$K(x, w) = 1 + xw + x^2w^2/4 \quad (\text{for all } x, w \in \mathscr{X} = [0, 1]),$$

where f'' denotes the second derivative of f.

If $\mathscr{F}$ contains constant functions and the norm of the unit function is one, then one may write $\mathscr{F} = \{c\} \oplus \hat{\mathscr{F}}$ and $K = 1 + \hat{K}$, where $\hat{K}$ is the reproducing kernel for $\hat{\mathscr{F}}$. Hickernell & Wang (2002) show how to decompose an arbitrary kernel for a space of univariate functions into this sum. Examples 1–3 are all of this form.

Based on this decomposition there is a straightforward construction of reproducing kernel Hilbert spaces of functions for multifactor experiments when the experimental domain is a Cartesian product of one-dimensional domains, that is $\mathscr{X} = \mathscr{X}_1 \times \ldots \times \mathscr{X}_s$. Suppose that $\mathscr{F}_j = \{c\} \oplus \hat{\mathscr{F}}_j$ is a reproducing kernel Hilbert space of functions defined on $\mathscr{X}_j$, with reproducing kernel $K_j = 1 + \hat{K}_j$. Let $1:s$ denote the set $\{1, \ldots, s\}$, and let u be any subset of $1:s$. Let x_u denote the elements of the vector x indexed by the elements of u. Let $|u|$ denote the cardinality of u, and let $\mathscr{X}_u$ denote the Cartesian product of $\mathscr{X}_j$ with $j \in u$. The tensor product space $\mathscr{F}$ of functions on $\mathscr{X}$ with reproducing kernel K may be defined as

$$\mathscr{F} = \bigotimes_{j=1}^{s} \mathscr{F}_j = \bigoplus_{\varnothing \subseteq u \subseteq 1:s} \hat{\mathscr{F}}_u, \quad \hat{\mathscr{F}}_u = \bigotimes_{j \in u} \hat{\mathscr{F}}_j,$$

$$K(x, w) = \prod_{j=1}^{s} K_j(x_j, w_j) = \sum_{\varnothing \subseteq u \subseteq 1:s} \hat{K}_u(x_u, w_u), \quad \hat{K}_u(x_u, w_u) = \prod_{j \in u} \hat{K}_j(x_j, w_j). \tag{2}$$

Note that $\hat{K}_\varnothing = 1$ and $\mathscr{F}_\varnothing$ is the space of constant functions by convention.

The Hilbert space $\hat{\mathscr{F}}_u$ comprises functions depending only on the variables x_j for $j \in u$, and $\hat{\mathscr{F}}_u$ contains no nonzero function that is constant with respect to any of these x_j. In fact, one may express any $f \in \mathscr{F}$ as a unique decomposition into effects, f_u, lying in the Hilbert space $\hat{\mathscr{F}}_u$ with reproducing kernel $\hat{K}_u$:

$$f(x) = \sum_{\varnothing \subseteq u \subseteq 1:s} f_u(x_u) \quad (f_u \in \hat{\mathscr{F}}_u). \tag{3}$$

The $\mathscr{F}$ defined above allows for multi-factor interactions of all orders. If prior knowledge allows one to ignore higher-order interactions, then the definition of $\mathscr{F}$ may be modified accordingly. The definition

$$\mathscr{F} = \bigoplus_{\substack{\varnothing \subseteq u \subseteq 1:s \\ |u| \leqslant d}} \hat{\mathscr{F}}_u, \quad K(x, w) = \sum_{\substack{\varnothing \subseteq u \subseteq 1:s \\ |u| \leqslant d}} \hat{K}_u(x_u, w_u) \tag{4}$$

includes functions with interactions of up to order d.

Example 4. Suppose that the jth factor has q_j levels, so that $\mathscr{X}_j = \{0, \ldots, q_j - 1\}$. Using Example 1 one can identify

$$\hat{K}_u(x_u, w_u) = \gamma^{|u|} \prod_{j \in u} (-1 + q_j \delta_{x_j w_j}), \quad K(x, w) = \sum_{j=1}^{s} \{1 + \gamma(-1 + q_j \delta_{x_j w_j})\}.$$

The decomposition of $f \in \mathscr{F}$ given by (3) is an analysis of variance decomposition, which may be defined recursively as follows:

$$f_{\varnothing} = \mathscr{I}_{1:s}(f), \quad f_u(x_u) = \mathscr{I}_{1:s-u}(f) - \sum_{\varnothing \subseteq v \subset u} f_v(x_v).$$

Here $\mathscr{I}_u$ denotes the integral with respect to $F_{\mathscr{X}_u}$ over $\mathscr{X}_u$. The inner product on $\mathscr{F}$ induced by this reproducing kernel is a weighted $\mathscr{L}_2$-inner product of the effects f_u:

$$\langle f, g \rangle_{\mathscr{F}}^2 = \sum_{\varnothing \subseteq u \subseteq 1:s} \gamma^{|u|} \mathscr{I}_u(f_u \bar{g}_u) \quad \text{(for all } f, g \in \mathscr{F}\text{)}.$$

Example 5. Suppose that the experimental domain is $\mathscr{X} = [0, 1]^s$, an s-factor generalisation of Example 2. One can identify

$$\hat{K}_u(x_u, w_u) = \gamma^{|u|} \prod_{j \in u} \{1 - \max(x_j, w_j)\}, \quad K(x, w) = \prod_{j=1}^{s} [1 + \gamma\{1 - \max(x_j, w_j)\}].$$

The decomposition of $f \in \mathscr{F}$ given by (3) may be defined recursively as follows:

$$f_{\varnothing} = f(1_{1:s}), \quad f_u = f(x_u, 1_{1:s-u}) - \sum_{\varnothing \subseteq v \subset u} f_v(x_v),$$

where 1_u denotes the $|u|$-dimensional vector of ones. The inner product on $\mathscr{F}$ induced by this reproducing kernel is

$$\langle f, g \rangle_{\mathscr{F}}^2 = \sum_{\varnothing \subseteq u \subseteq 1:s} \gamma^{|u|} \mathscr{I}_u \left(\frac{\partial f_u}{\partial x_u} \frac{\partial \bar{g}_u}{\partial x_u} \right) \quad \text{(for all } f, g \in \mathscr{F}\text{)}.$$

3. The discrepancy

Having defined a Hilbert space of possible response functions, $\mathscr{F}$, one can now define the discrepancy. For any design, $P = \{z_1, \ldots, z_n\}$, one may write the associated empirical distribution as

$$F_P(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{[z_i, +\infty)}(x),$$

where $1_{\{.\}}(x)$ is the indicator function.

Definition 1 (*Hickernell*, 2000). *Given a design, P, and a kernel, K, the discrepancy is defined as*

$$\begin{aligned} D(P; K) &= \left[\int_{\mathscr{X}^2} K(x, w)\, d\{F_{\mathscr{X}}(x) - F_P(x)\}\, d\{F_{\mathscr{X}}(w) - F_P(w)\} \right]^{\frac{1}{2}} \\ &= \left[\int_{\mathscr{X}^2} K(x, w)\, dF_{\mathscr{X}}(x)\, dF_{\mathscr{X}}(w) - \frac{1}{n} \sum_{i=1}^{n} \int_{\mathscr{X}} \{K(z_i, w) + K(w, z_i)\}\, dF_{\mathscr{X}}(w) \right. \\ &\qquad \left. + \frac{1}{n^2} \sum_{i,k=1}^{n} K(z_i, z_k) \right]^{\frac{1}{2}}. \end{aligned}$$

The discrepancy arises in the theory of quadrature error for multivariate integrals, and the analysis here relies on these results. For a given design P the integral $\mathscr{I}(f)$ may be approximated by the mean of the values of the integrand sampled at the design points.

The error of this approximation is

$$|\mathrm{err}(f;P)| = \left|\mathscr{I}(f) - \frac{1}{n}\sum_{i=1}^{n} f(z_i)\right| = \left|\mathscr{I}(f) - \int_{\mathscr{X}} f(x)\,dF_P(x)\right|$$
$$\leqslant D(P;K)V(f;K) \quad (\text{for all } f \in \mathscr{F}) \tag{5}$$

(Hickernell, 2000), where the variation of the integrand is defined as $V(f;K) = \inf_c \|f - c\|_{\mathscr{F}}$. Error bound (5) is tight. For every P there exists a worst-case integrand, $\xi_P \in \mathscr{F}$, for which (5) becomes an equality. This worst-case integrand is, in fact, the representer of the linear functional err(.; P) (Hickernell, 2000). Note that $D(P;K) = \|\xi_P\|_{\mathscr{F}}$.

The discrepancy depends only on the design and the reproducing kernel, and it measures how well the empirical distribution of the design, F_P, approximates the uniform distribution, $F_{\mathscr{X}}$. One may also define the discrepancy without reference to quadrature error as $D(P;K) = \|F_{\mathscr{X}} - F_P\|_{\mathscr{M}}$, where the definition of $\|.\|_{\mathscr{M}}$ depends on the kernel K. For the details of this approach see Hickernell (1999).

The squared discrepancy defined by the kernel in Example 4 may be written as

$$D^2(P;K) = -1 + \frac{1}{n^2}\sum_{i,k=1}^{n}\prod_{j=1}^{s}\{1+\gamma(-1+q_j\delta_{z_{ij}z_{kj}})\}.$$

The squared discrepancy defined by the kernel in Example 5 reduces to

$$D^2(P;K) = \left(1+\frac{\gamma}{3}\right)^s - \frac{2}{n}\sum_{i=1}^{n}\prod_{j=1}^{s}\left(1+\gamma\frac{1-z_{ij}^2}{2}\right)$$
$$+\frac{1}{n^2}\sum_{i,k=1}^{n}\prod_{j=1}^{s}[1+\gamma\{1-\max(z_{ij},z_{kj})\}]$$

(Hickernell, 1998b, 2000). For a review of other discrepancy measures see Hickernell (1998b, 2000).

It is typically the case that $D(P;K) = 0$ if and only if $F_P = F_{\mathscr{X}}$. However, it is possible to have $D(P;K) = 0$ with $F_P \neq F_{\mathscr{X}}$ if F_P mimics $F_{\mathscr{X}}$ perfectly when sampling the functions in the reproducing kernel Hilbert space defined by K. Consider the K defined in Example 3, and the design $P = \{(3 \pm 3^{\frac{1}{2}})/6\}$. It is straightforward to check that

$$\int_{\mathscr{X}} K(x,w)\,dF_{\mathscr{X}}(w) = 1 + \frac{x}{2} + \frac{x^2}{12} = \int_{\mathscr{X}} K(x,w)\,dF_P(w).$$

Since $\int_{\mathscr{X}} K(x,w)\,d\{F_{\mathscr{X}}(w) - F_P(w)\} = 0$, for all x, it follows that $D(P;K) = 0$ even though $F_{\mathscr{X}} \neq F_P$.

4. THE EFFICIENCY AND ROBUSTNESS OF LOW DISCREPANCY DESIGNS

The quadrature error bound in (5) implies bounds on $\mathrm{var}(P,g)$ and $\mathrm{bias}(P,g,h)$ in (1) in terms of the discrepancy. Suppose that the response, $f(x)$, lies in some reproducing kernel Hilbert space, $\mathscr{F}$, with kernel K. The misspecification, $h(x)$, lies in a subspace, $\mathscr{H}$, with reproducing kernel $K_\perp$. The Hilbert space $\mathscr{H}$ consists of those functions that are orthogonal, in an $\mathscr{L}_2$-sense, to $g_1,\ldots,g_p$. It is known that

$$K_\perp(x,w) = K(x,w) - \Upsilon^H(x)\Psi^{-1}\Upsilon(w),$$

where

$$\Upsilon(w) = \mathscr{I}\{gK(.,w)\}, \quad \Psi = \int_{\mathscr{X}^2} \bar{g}(x)K(x,w)g^{\mathrm{T}}(w)\,dF_{\mathscr{X}}(x)\,dF_{\mathscr{X}}(w)$$

(Yue & Hickernell, 1999).

THEOREM 1. *Suppose that $\mathscr{F}$ is a reproducing kernel Hilbert space with reproducing kernel K, and that the set of design points is P. Assume that the functions $g_1, \ldots, g_p \in \mathscr{F}$ are chosen such that $G = I$, the $p \times p$ identity matrix. Moreover, for all complex-valued $\alpha = (\alpha_1, \ldots, \alpha_p)^{\mathrm{T}}$, all $w \in \mathscr{X}$ and all $\zeta \in \mathscr{F}$, suppose that $|g^{\mathrm{T}}\alpha|^2$, $\omega_{\alpha w} = K_{\perp}(w, x)g^{\mathrm{T}}(x)\alpha$ and $\Omega_{\alpha\zeta} = \langle \omega_{\alpha w}, \zeta \rangle_{\mathscr{F}}$ are all functions in $\mathscr{F}$. Define the following constants that depend on the form of the model and not on the design*:

$$V_{gg} = \sup_{\|\alpha\|_2 \leqslant 1} V(|g^{\mathrm{T}}\alpha|^2), \quad V_{gh} = \sup_{\substack{\|\alpha\|_2 \leqslant 1 \\ \|\zeta\|_{\mathscr{F}} \leqslant 1}} \|\Omega_{\alpha\zeta}\|_{\mathscr{F}}. \tag{6}$$

Then $\mathrm{var}(P, g)$ *and* $\mathrm{bias}(P, g, h)$ *have the following upper bounds in terms of discrepancy, provided that $D(P; K)V_{gg} < 1$*:

$$\mathrm{var}(P, g) \leqslant \frac{\sigma^2 p}{n\{1 - D(P; K)V_{gg}\}}, \tag{7}$$

$$\mathrm{bias}(P, g, h) \leqslant \left\{\frac{D(P; K)V_{gh}\|h\|_{\mathscr{F}}}{1 - D(P; K)V_{gg}}\right\}^2. \tag{8}$$

These upper bounds are monotonically decreasing as $D(P; K)$ tends to zero.

The hypotheses and conclusions of this theorem are discussed before giving the proof. Since $\mathrm{var}(P, g)$ and $\mathrm{bias}(P, g, h)$ are invariant if $g(x)$ is replaced by $Lg(x)$ for any $p \times p$ nonsingular matrix L, the assumption that $G = I$ is not essential but simplifies the form of the upper bound.

Although $D(P; K)$ does not depend on the particular model, the kernel K does reflect one's assumptions about what kinds of response are expected. Not only does K determine the Hilbert space $\mathscr{F}$ of all possible responses, it also determines the definitions of the norms and inner products that arise in the definitions of V_{gg} and V_{gh}. Thus, K may be constructed to reflect any prior knowledge about the response.

For finite-dimensional spaces, $\mathscr{F}$, the quantities V_{gg} and V_{gh} may be replaced by their upper bounds over all possible g and h. However, for infinite-dimensional $\mathscr{F}$ one can usually construct a pathological example with arbitrarily bad aliasing, so V_{gg} and V_{gh} are unbounded.

The Hilbert space $\mathscr{F}$ must also contain the square of the response function. The assumption that $\omega_{\alpha w} \in \mathscr{F}$ means that $g_j\bar{h} \in \mathscr{F}$ for any possible misspecification h. The assumptions that $|g^{\mathrm{T}}\alpha|^2 \in \mathscr{F}$ and $D(P; K)V_{gg} < 1$ are necessary to ensure that the information matrix, M, has eigenvalues bounded away from zero, thus guaranteeing that the model does not contain functions that the design cannot discriminate.

A weakness of Theorem 1 is that the upper bounds (7) and (8) are not tight. The vector g enters the definitions of $\mathrm{var}(P, g)$ and $\mathrm{bias}(P, g, h)$ in a nontrivial, nonlinear way. A tight bound over all possible models in terms of V_{gg} and V_{gh} would require numerical functional optimisation and result in a very complicated bound. Thus, we have sacrificed tightness for a bound in terms of a relatively simple measure of quality of the design, namely the

discrepancy. There is one limiting case for which Theorem 1 is tight. If $F_P = F_{\mathscr{X}}$, then $M = nI$, $D(P; K) = 0$, $\mathrm{var}(P, g) = \sigma^2 p/n$ and $\mathrm{bias}(P, g, h) = 0$.

Proof of Theorem 1. Since $G = I$ it follows that $\mathrm{var}(P, g) = \sigma^2 \,\mathrm{tr}(M^{-1})$. Let $\tilde{M} = I - M/n$. For any $\alpha \in \mathbb{C}^p$ it follows that

$$\rho(\tilde{M}) = \sup_{\|\alpha\|_2 \leqslant 1} |\alpha^H \tilde{M} \alpha| = \sup_{\|\alpha\|_2 \leqslant 1} \left| \alpha^H \left\{ \mathscr{I}(\bar{g} g^{\mathrm{T}}) - \frac{1}{n} \sum_{i=1}^{n} \bar{g}(z_i) g^{\mathrm{T}}(z_i) \right\} \alpha \right|$$

$$= \sup_{\|\alpha\|_2 \leqslant 1} |\mathrm{err}(|g^{\mathrm{T}}\alpha|^2; P)| \leqslant D(P; K) \sup_{\|\alpha\|_2 \leqslant 1} V(|g^{\mathrm{T}}\alpha|^2; K) = D(P; K) V_{gg}.$$

Here ρ denotes the spectral radius of a matrix. If $D(P; K) V_{gg} < 1$, then the smallest eigenvalue of $n^{-1}M = I - \tilde{M}$ is greater than or equal to $1 - D(P; K) V_{gg} > 0$. This guarantees that M is nonsingular. Thus, $\rho(M^{-1}) \leqslant n^{-1}\{1 - D(P; K) V_{gg}\}^{-1}$, which completes the proof of (7).

According to Yue & Hickernell (1999), $\mathrm{bias}(P, g, h)$ has a tight upper bound of $\rho(M^{-2} X^H K_\perp X) \|h\|_{\mathscr{F}}^2$. Since an upper bound on $\rho(n^2 M^{-2})$ is known, the remaining step is to derive an upper bound on $\rho(B)$, where $B = n^{-2} X^H K_\perp X$. Using a similar argument to the above one may write

$$\alpha^H B \alpha = \frac{1}{n^2} \sum_{i,k=1}^{n} \alpha^H \bar{g}(z_i) K_\perp(z_i, z_k) g^{\mathrm{T}}(z_k) \alpha.$$

Recall that, if $\{\phi_\nu^\perp(x)\}$ is any countable, orthonormal basis of the subspace $\mathscr{H}$, then one may write $K_\perp = \sum_\nu \phi_\nu^\perp(x) \bar{\phi}_\nu^\perp(w)$. Since $\mathscr{I}\{\bar{g} K_\perp(., w)\} = 0$ for all $w \in \mathscr{X}$ it follows that

$$\alpha^H B \alpha = \sum_\nu \left\{ \frac{1}{n^2} \sum_{i,k=1}^{n} \alpha^H \bar{g}(z_i) \phi_\nu^\perp(z_i) \bar{\phi}_\nu^\perp(z_k) g^{\mathrm{T}}(z_k) \alpha \right\}$$

$$= \sum_\nu |\mathrm{err}(\bar{\phi}_\nu^\perp g^{\mathrm{T}}\alpha; P)|^2 = \sum_\nu |\langle \bar{\phi}_\nu^\perp g^{\mathrm{T}}\alpha, \xi_P \rangle_{\mathscr{F}}|^2.$$

One the other hand, one may also write

$$\Omega_{\alpha\xi_P}(x) = \langle \omega_{\alpha x}, \xi_P \rangle_{\mathscr{F}} = \langle K_\perp(x, .) g^{\mathrm{T}}\alpha, \xi_P \rangle_{\mathscr{F}} = \left\langle \sum_\nu \phi_\nu^\perp(x) \bar{\phi}_\nu^\perp g^{\mathrm{T}}\alpha, \xi_P \right\rangle_{\mathscr{F}}$$

$$= \sum_\nu \phi_\nu^\perp(x) \langle \bar{\phi}_\nu^\perp g^{\mathrm{T}}\alpha, \xi_P \rangle_{\mathscr{F}} \in \mathscr{H}.$$

The preceding two equations imply that $\|\Omega_{\alpha\xi_P}\|_{\mathscr{H}}^2 = \sum_\nu |\langle \bar{\phi}_\nu^\perp g^{\mathrm{T}}\alpha, \xi_P \rangle_{\mathscr{F}}|^2 = \alpha^H B \alpha$. By referring to (6) one may bound the spectral radius of B by

$$\rho(B) = \sup_{\|\alpha\|_2 \leqslant 1} \|\Omega_{\alpha\xi_P}\|_{\mathscr{H}}^2 \leqslant \|\xi_P\|_{\mathscr{F}}^2 V_{gh}^2 = D^2(P; K) V_{gh}^2.$$

Together with the bound on $\rho(n^2 M^{-2})$, this completes the proof of (8). □

The construction of robust experimental designs has been studied previously by other authors. Box & Draper (1959) considered the case where the true response is a polynomial of some unknown degree. Sacks & Ylvisacker (1984) considered one-factor designs with $\mathscr{X}$ being the real line, and $\mathscr{F}$ being the set of functions that are Lipschitz continuous or have Lipschitz continuous derivatives. Wiens (1990) considered a situation similar to that considered here, but with h having bounded square integral. Since no smoothness of the misspecification is assumed, one must allow F_P to be an arbitrary absolutely continuous measure on $\mathscr{X}$, rather than the empirical distribution of a finite set of points.

In view of Theorem 1 the discrepancy is an omnibus measure of the possible effects of aliasing. However, there is a scaling problem since $D(P;\, cK) = c^{\frac{1}{2}} D(P;\, K)$ for any positive c. One may scale the discrepancy as follows:

$$\hat{D}(P;\, K) = \frac{D(P;\, K)}{\{\int_{\mathscr{X}} K(x,\, x)\, dF_{\mathscr{X}}(x) - \int_{\mathscr{X}^2} K(x,\, w)\, dF_{\mathscr{X}}(x)\, dF_{\mathscr{X}}(w)\}^{\frac{1}{2}}}.$$

For a simple random n-point or Monte Carlo design, P_{MC}, $E\{\hat{D}(P_{\mathrm{MC}};\, K)\}^2 = 1/n$, independent of the choice of kernel (Hickernell, 1998a), so a design P is as good or bad as a random design with $1/\hat{D}^2(P;\, K)$ points.

5. Aberration and resolution

Consider now the case where the experimental domain, $\mathscr{X}$, is a Cartesian product of one-dimensional domains. Let $P_u = \{z_u : z \in P\}$ denote the projection of the design into the domain $\mathscr{X}_u$. One would normally desire that, for small $|u|$, the projections P_u would be good designs on the $\mathscr{X}_u$. This is the motivation behind the definitions of resolution (Box et al., 1978, p. 385) and aberration (Fries & Hunter, 1980). The discrepancy, as defined above, does not necessarily guarantee this, but one may define an aberration in terms of the pieces of the squared discrepancy.

Since the squared discrepancy is linear in the reproducing kernel, for kernels of the form (2) one may write

$$D^2(P;\, K) = \sum_{\varnothing \subset u \subset 1:s} D^2(P_u;\, \hat{K}_u) = \sum_{j=1}^{s} D^2_{(j)}(P;\, K), \tag{9}$$

where

$$D^2_{(j)}(P;\, K) = \sum_{|u| = j} D^2(P_u;\, \hat{K}_u).$$

Since $\hat{K}_{\varnothing} = 1$ it follows that $D(P;\, \hat{K}_{\varnothing}) = 0$.

Definition 2. *Suppose that $\mathscr{X} = \mathscr{X}_1 \times \ldots \times \mathscr{X}_s$, and that the reproducing kernel, K, is of the form* (2). *The projection discrepancy pattern, or generalised word-length pattern, is defined as the s-vector*

$$\mathrm{PD}(P;\, K) = (D_{(1)}(P;\, K), \ldots, D_{(s)}(P;\, K)).$$

For any two designs P, $\tilde{P} \subseteq \mathscr{X}$, one says that P has smaller aberration than $\tilde{P}$, or equivalently $\mathrm{PD}(P;\, K) < \mathrm{PD}(\tilde{P};\, K)$, *if and only if the first, from the left, nonzero component of* $\mathrm{PD}(P;\, K) - \mathrm{PD}(\tilde{P};\, K)$ *is negative. If $t = \min\{j : D_{(j)}(P;\, K) > 0\}$, then P is said to have resolution t.*

For Examples 4 and 5, the $D^2_{(j)}(P;\, K)$ are given by

$$D^2_{(j)}(P;\, K) = \gamma^j \sum_{|u| = j} \frac{1}{n^2} \sum_{i,k=1}^{n} \prod_{l \in u} (-1 + q_l \delta_{z_{il} z_{kl}}), \tag{10}$$

$$D^2_{(j)}(P;\, K) = \gamma^j \sum_{|u| = j} \left[(\tfrac{1}{3})^j - \frac{2}{n} \sum_{i=1}^{n} \prod_{l \in u} \left(\frac{1 - z_{il}^2}{2} \right) + \frac{1}{n^2} \sum_{i,k=1}^{n} \prod_{l \in u} \{1 - \max(z_{il} z_{ik})\} \right],$$

respectively. Definition 2 does not assume that the design is a regular factorial design or

even that each factor has a finite number of levels. It only assumes that the experimental domain is a Cartesian product of one-factor domains, and that the reproducing kernel is a product of one-dimensional kernels. For the special case of (10), Definition 2 reduces to the generalised aberration proposed by Xu & Wu (2001), as shown below.

THEOREM 2. *For the case of Example* 4 *where the components of projection discrepancy pattern are given by* (10), *the resulting aberration as given in Definition* 2 *is equivalent to the generalised aberration defined by Xu & Wu* (2001). *For the case of two-level designs, the aberration defined here is equivalent to the* G_2*-aberration of Tang & Deng* (1999).

Proof. Xu & Wu (2001) defined the generalised word-length pattern as the vector $(A_1(P), \ldots, A_s(P))$, where

$$A_j(P) = \frac{1}{n^2} \sum_{\mathrm{wt}(v)=j} |\chi_v(P)|^2, \quad \chi_v(P) = \sum_{i=1}^{n} \prod_{l=1}^{s} e^{2\pi \imath v_l z_{il}/q_l}.$$

Here wt(.) denotes the number of nonzero elements of the argument. From this definition it follows that

$$|\chi_v(P)|^2 = \sum_{i,k=1}^{n} \prod_{l=1}^{s} e^{2\pi \imath v_l (z_{il} - z_{kl})/q_l}, \quad A_j(P) = \frac{1}{n^2} \sum_{|u|=j} \sum_{i,k=1}^{n} \prod_{l \in u} \sum_{v_l=1}^{q_l - 1} e^{2\pi \imath v_l (z_{il} - z_{kl})/q_l}.$$

If one uses the definitions in Example 4 this expression may be further simplified to obtain

$$A_j(P) = \frac{1}{n^2} \sum_{|u|=j} \sum_{i,k=1}^{n} \prod_{l \in u} (-1 + q_l \delta_{z_{il} z_{kl}}) = \gamma^{-j} D_{(j)}^2(P; K).$$

Thus, the generalised word-length pattern of Xu & Wu (2001) is equivalent to the projection discrepancy pattern of Definition 2, and the two definitions of aberration are also equivalent. Xu & Wu (2001) showed that their aberration was equivalent to the G_2-aberration of Tang & Deng (1999). □

The connection between aberration and discrepancy has been considered for regular two-level fractional factorial designs by Fang & Mukerjee (2000) and Fang et al. (2002) for higher numbers of levels. Tang (2001) showed that the J-characteristics defining the G_2-aberration of Tang & Deng (1999) measure the uniformity of projections of the two-level design into lower dimensions. The above theorem relating discrepancy to aberration is a generalisation of these results.

Some comments are in order regarding the parameter γ that enters into the definition of the discrepancy. This parameter has no effect when comparing the projection discrepancy patterns of different designs. Thus, as far as the projection discrepancy pattern is concerned, one might as well set $\gamma = 1$. However, the value of γ does affect the comparison of the discrepancies of different designs. Recall from (9) that the squared discrepancy is a sum of the $D_{(j)}^2(P; K)$, and note that, if each $\hat{K}_j$ has a leading factor of γ, then $D_{(j)}^2(P; K)$ has a leading factor of γ^j. A larger value of γ gives a relatively heavier weight to the $D_{(j)}^2(P; K)$ with large j and implies a preference for better uniformity in the high-dimensional projections of P, whereas a small value of γ implies a preference for better uniformity in the low-dimensional projections of P. Thus, comparing the aberration of two designs is equivalent to comparing their discrepancies for vanishing γ, as is explained in the following theorem whose proof is straightfoward.

THEOREM 3. *Suppose that the reproducing kernel is of the form* (2), *and that* $\hat{K}_j$ *has a*

leading coefficient γ. *For a fixed number of experiments,* n, *let* P_a *be a minimum aberration design, and let* t *be its resolution. For any* $\gamma > 0$, *let* P_γ *denote a minimum discrepancy design. Then, if one assumes that* $D(P_a; K) > 0$, *it follows that*

$$\lim_{\gamma \downarrow 0} D(P_\gamma; K)/D(P_a; K) = 1, \quad \lim_{\gamma \downarrow 0} D_{(j)}(P_\gamma; K) = D_{(j)}(P_a; K) \quad (j = 1, \ldots, t).$$

If a factor has only a few levels, then it may be justified to choose K as in Example 1. However, in some practical cases a factor has several levels and their order is important. In such a situation, one might expect low wave-number functions to be more likely than high wave-number functions. The reproducing kernel may be constructed to reflect this assumption. A very simple example is given below to illustrate the general principle.

Example 6. Consider the same setting as in Example 1 with $q = 4$ and $\mathscr{X} = \{0, 1, 2, 3\}$, but with orthonormal basis $\{1, \gamma^{\frac{1}{2}} e^{\pi \imath x/2}, \gamma e^{\pi \imath x}, \gamma^{\frac{1}{2}} e^{-\pi \imath x/2}\}$. Choose $\gamma < 1$ to emphasise lower wave-number functions over higher wave-number functions. The reproducing kernel and squared discrepancy are

$$K(x, w) = 1 + 2\gamma \cos\{\pi(x - w)/2\} + \gamma^2(-1)^{x-w},$$

$$D(P; K) = \frac{1}{n^2} \sum_{i,k=1}^{n} [2\gamma \cos\{\pi(z_i - z_k)/2\} + \gamma^2(-1)^{z_i - z_k}].$$

For either of the two-point designs $\{0, 2\}$ or $\{1, 3\}$ the squared discrepancy is γ^2, whereas for any other designs with two distinct points the squared discrepancy is $\gamma > \gamma^2$. Thus, the design with the points more spread out has lower discrepancy. By contrast, for Example 1 all designs with two distinct points have the same discrepancy.

Suppose that the true response is known to be a sum of interaction terms involving τ or fewer coordinates. Then the product of any two terms involves at most $d \leqslant \min(2\tau, s)$ coordinates. To guarantee no aliasing in the estimation of the model one must use a design of resolution $d + 1$. This can be seen explicitly by referring back to Theorem 1. The assumptions in this theorem are that f and $|f|^2$ both lie in the space (4). Thus, $D^2(P; K)$ is only a sum of the $D^2_{(j)}(P; K)$ with $j \leqslant d$. Since the design, P, has resolution $d + 1$, this implies that $D^2_{(1)}(P; K) = \ldots = D^2_{(d)}(P; K) = 0$, and so $D^2(P; K) = 0$. By Theorem 1 and the remark preceding its proof, it follows that $\mathrm{var}(P, g) = \sigma^2 p/n$ and $\mathrm{bias}(P, g, h) = 0$.

Acknowledgement

The authors greatly appreciate helpful discussions with Kai-Tai Fang, Chang-Xing Ma, Rahul Mukerjee and Guo-Liang Tian. The suggestions of the referee, the associate editor and the editor greatly improved the paper. This work was partially supported by grants from the Hong Kong Research Grants Council, the Hong Kong Baptist University and the National Natural Science Foundation of China.

References

Aronszajn, M. (1950). Theory of reproducing kernels. *Trans. Am. Math. Soc.* **68**, 337–404.

Bates, R., Buck, R., Riccomagno, E. & Wynn, H. P. (1996). Experimental design and observation for large systems. *J. R. Statist. Soc.* B **58**, 77–94.

Box, G. E. P. & Draper, N. R. (1959). A basis for the selection of a response surface design. *J. Am. Statist. Assoc.* **54**, 622–54.

Box, G. E. P., Hunter, W. G. & Hunter, J. S. (1978). *Statistics for Experimenters.* New York: John Wiley & Sons.

DENG, L.-Y. & TANG, B. (1999). Generalized resolution and minimum aberration criteria for Plackett-Burman and other nonregular factorial designs. *Statist. Sinica* **9**, 1071–82.

FANG, K. T. & MUKERJEE, R. (2000). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika* **87**, 193–8.

FANG, K. T. & WANG, Y. (1994). *Number-Theoretic Methods in Statistics*. New York: Chapman and Hall.

FANG, K. T., MA, C. X. & MUKERJEE, R. (2002). Uniformity in fractional factorials. In *Monte Carlo and Quasi-Monte Carlo Methods 2000*, Ed. K. T. Fang, F. J. Hickernell and H. Niederreiter, pp. 232–41. Berlin: Springer-Verlag.

FRIES, A. & HUNTER, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics* **22**, 601–8.

HICKERNELL, F. J. (1998a). A generalized discrepancy and quadrature error bound. *Math. Comp.* **67**, 299–322.

HICKERNELL, F. J. (1998b). Lattice rules: How well do they measure up? In *Random and Quasi-Random Point Sets*, Lecture Notes in Statistics, **138**, Ed. P. Hellekalek and G. Larcher, pp. 109–66. New York: Springer-Verlag.

HICKERNELL, F. J. (1999). Goodness-of-fit statistics, discrepancies and robust designs. *Statist. Prob. Lett.* **44**, 73–8.

HICKERNELL, F. J. (2000). What affects the accuracy of quasi-Monte Carlo quadrature? In *Monte Carlo and Quasi-Monte Carlo Methods 1998*, Ed. H. Niederreiter and J. Spanier, pp. 16–55. Berlin: Springer-Verlag.

HICKERNELL, F. J. & WANG, X. (2002). Error bounds and tractability of quasi-Monte Carlo algorithms in infinite dimensions. *Math. Comp.* **71**. To appear.

MA, C.-X. & FANG, K.-T. (2001). A note on generalized aberration in factorial designs. *Metrika* **53**, 85–93.

SACKS, J. & YLVISACKER, D. (1984). Some model robust designs in regression. *Ann. Statist.* **12**, 1324–48.

SAITOH, S. (1988). *Theory of Reproducing Kernels and its Applications*. Harlow, Essex, England: Longman Scientific & Technical.

TANG, B. (2001). Theory of *J*-characteristics for fractional factorial designs and projection justification of minimum G_2-aberration. *Biometrika* **88**, 401–7.

TANG, B. & DENG, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Ann. Statist.* **27**, 1914–26.

WAHBA, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

WANG, Y. & FANG, K. T. (1981). A note on uniform distribution and experimental design. *Kexue Tongbao (Chin. Sci. Bull.)* **26**, 485–9.

WEYL, H. (1916). Über die Gleichverteilung der Zahlen mod eins. *Math. Ann.* **77**, 313–52.

WIENS, D. P. (1990). Robust minimax designs for multiple linear regression. *Lin. Algeb. Applic.* **127**, 327–40.

XU, H. & WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Ann. Statist.* **29**, 549–60.

YUE, R. X. & HICKERNELL, F. J. (1999). Robust designs for fitting linear models with misspecification. *Statist. Sinica* **9**, 1053–69.

[*Received January* 2001. *Revised April* 2002]