

Working Paper M09/19

Methodology

Optimal Designs For Generalised

Nonlinear Models With

Application

To Second Harmonic Generation

Experiments

Stefanie Biedermann, David C. Woods

Abstract

The design of experiments for generalised nonlinear models is investigated and applied to an optical process for characterising interfaces which is widely used in the physical and natural sciences. Design strategies for overcoming the dependence of a D-optimal design on the values of the model parameters are explored, including the use of Bayesian designs. Designs for the accurate estimation of model parameters are presented and compared, as are designs for the estimation of a set of ratios of parameters, which is of particular importance in the motivating example. The effectiveness of various design methods are studied, and the benefits of well designed experiments are demonstrated.

Optimal designs for generalised nonlinear models with application to second harmonic generation experiments

Stefanie Biedermann and David C. Woods¹
Southampton Statistical Sciences Research Institute,
University of Southampton,
Southampton, UK

The design of experiments for generalised nonlinear models is investigated and applied to an optical process for characterising interfaces which is widely used in the physical and natural sciences. Design strategies for overcoming the dependence of a D -optimal design on the values of the model parameters are explored, including the use of Bayesian designs. Designs for the accurate estimation of model parameters are presented and compared, as are designs for the estimation of a set of ratios of parameters, which is of particular importance in the motivating example. The effectiveness of various design methods are studied, and the benefits of well designed experiments are demonstrated.

Keywords: Bayesian design; Cluster design; D -optimality; Laser-surface chemistry; Poisson regression; Nonlinear regression

1. Introduction

Nonlinear parametric regression models are widely used in the physical and natural sciences to describe the influence of one or more explanatory variables on a response. Typically, these models are phenomenological (i.e. derived from scientific theory), have parameters whose values are directly interpretable by the experimenters, and have additive independent random errors. When designing an experiment to estimate such a model, it is common practice to assume Normal errors with constant variance. If these assumptions are believed unlikely to hold, then a wider class of models should be entertained and methods for obtaining efficient experimental designs for this class employed.

Generalised nonlinear models extend nonlinear regression models to allow non-normally distributed error structures; see Wei (1997); Batchelor et al. (2007); Kosmidis and Firth (2008). Their relationship to nonlinear regression is analogous to that between generalised linear models (McCullagh and Nelder, 1989) and linear regression, in that the mean is related to the explanatory variables through a differentiable and strictly monotonic link function. Unlike generalised linear models, the relationship between the transformed mean and the explanatory variables is nonlinear in the unknown parameters (see Section 1.2).

The purpose of this paper is to develop and apply methods for finding optimal and efficient experimental designs for generalised nonlinear models. The methods are motivated and demonstrated through an application from laser-surface chemistry.

1.1. Motivating application

The behaviour of the interface between two phases of matter, such as a liquid and gas or a solid and a liquid, is an important field of research in modern chemistry with applications

¹*Address for correspondence:* David Woods, School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. E-mail: D.Woods@southampton.ac.uk

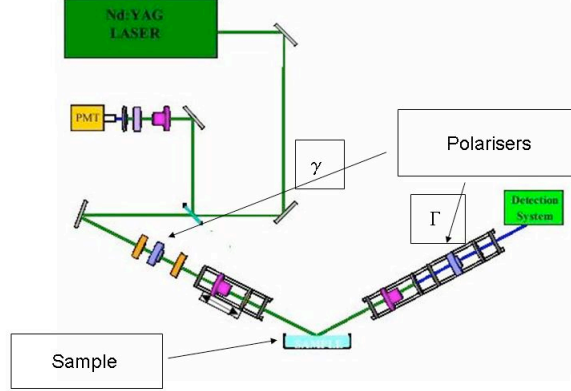


Figure 1: Schematic of a second harmonic generation experiment: γ and Γ denote input and output polarisation angles

in surfactants, catalysis, membranes and electrochemistry. Second harmonic generation (SHG; Shen, 1989) is an optical process for characterising surfaces and interfaces through counting the number of molecules at the interface of the two phases and determining molecule orientation. It has many advantages over other methods, such as spectroscopy and other optical techniques: it is simpler to apply, is directional (allowing in-situ measurements of surfaces), and is high frequency, and hence suitable for dynamic systems that change quickly over time. SHG has applications in science and engineering, including physics, electronics and biology (Campagnola and Loew, 2003; Salafsky, 2007).

Our application concerns an air/liquid interface, specifically the characterisation of an L-phenylalanine surface. In an experiment, a laser beam is fired through a polariser and strikes the liquid surface. The reflected beam of light passes through another polariser and its intensity is measured by a count of photons which is used to investigate the number and orientation of surface molecules (Fig. 1). See also Fordyce et al. (2001) for further details of the experimental procedure. The experimenter can control the size of the input and output polarisation angles γ and Γ , respectively.

The observed intensity Y_{ij} of the laser beam at the j th observation at polarisation angles γ_i and Γ_i is modelled as a Poisson random variable where the expectation is the theoretical intensity $|E_i|^2$, derived from a phenomenological model (Mizrahi and Sipe, 1988; Fordyce et al., 2001). That is,

$$\begin{aligned} Y_{ij} &\sim \text{Po}(|E_j|^2), \quad i = 1, \dots, n, \quad j = 1, \dots, n_i, \\ E_j &= E_{s,j} \sin(\Gamma_j) + E_{p,j} \cos(\Gamma_j), \\ E_{s,j} &= C \sin(2\gamma_j), \quad E_{p,j} = A \cos^2(\gamma_j) + B \sin^2(\gamma_j), \end{aligned} \tag{1}$$

where n is the number of distinct combinations of input and output angles in the experiment, known as support points of the design, and n_i is the number of replicates of the i th support point, with $n_1 + \dots + n_n = N$. The parameters A , B and C may be complex numbers but to

ensure identifiability, A is constrained to be real. These parameters can be expressed in either Cartesian or polar coordinates (see Section 2); in the Euler parameterisation, they have the form $A = r_a$, $B = r_b \exp(i\phi_b)$, and $C = r_c \exp(i\phi_c)$. Hence the parameters that require estimation in this model are $\boldsymbol{\theta} = (r_a, r_b, r_c, \phi_b, \phi_c)$. Welsh et al. (2005) applied this phenomenological model to data from the air/liquid interface for L-phenylalanine. The data were obtained using a full 3×14 factorial design, ξ_0 , with levels $\Gamma = 0, \pi/4, \pi/2$ and $\gamma = 0, \pi/26, \dots, \pi/2$. Each of the 42 combinations was equally replicated ($n_1 = \dots = n_n$).

As model (1) is a generalised nonlinear model (Section 1.2), this application demonstrates the need to find optimal and highly efficient designs for the estimation of the model parameters for this class of models (see Section 2).

When chemists wish to compare results from several experiments, then important quantities of interest are the ratios of the coefficients A , B and C . In the Euler parametrisation, these ratios have the form

$$\frac{S}{T} = \frac{r_s e^{i\phi_s}}{r_t e^{i\phi_t}} = \frac{r_s}{r_t} e^{i(\phi_s - \phi_t)}, \quad S, T = A, B, C; \quad s, t = a, b, c; \quad \phi_a = 0. \quad (2)$$

Hence ratios of the radii r_s , $s = a, b, c$, and differences between the angles ϕ_s , $s = a, b, c$, require efficient estimation. The problem of finding optimal designs for a set of ratios is addressed in Section 3.

1.2. Generalised nonlinear models

Let $\mathbf{x}_j = (x_{j1}, \dots, x_{jk})$ hold the values for the k quantitative explanatory variables in the j th run of the experiment ($j = 1, \dots, N$) and Y_j denote the corresponding response. A generalised nonlinear model (GNM) is defined by the following components:

1. a *distribution* for the response, Y , from the exponential family, which may have variance depending on the mean:

$$f_Y(y_j; \alpha_j, \delta) = \exp \{ [y_j \alpha_j - b(\alpha_j)] / a(\delta) + c(y, \delta) \}, \quad j = 1, \dots, N,$$

which has canonical parameters α_j , a scale parameter δ and functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ specific to the member of the exponential family,

2. a *systematic predictor*, $\eta(\mathbf{x}; \boldsymbol{\theta})$, involving functions of the explanatory variables which may be nonlinear in the p unknown model parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$, and
3. a *link function*, $g(\cdot)$, that maps $E(Y) = \mu$ to the predictor; $g(\mu) = \eta(\mathbf{x}; \boldsymbol{\theta})$.

Wei (1997) gave technical details for the existence of this class of models; we note only that $\eta(\mathbf{x}; \boldsymbol{\theta})$ must be at least twice differentiable with respect to $\boldsymbol{\theta}$ and that $E(Y_j) = \partial b(\alpha_j) / \partial \alpha_j = \mu_j$ and $\text{Var}(Y_j) = a(\delta) \partial^2 b(\alpha_j) / \partial \alpha_j^2 = a(\delta) h(\mu_j)$. A more general definition is also possible via the quasi-likelihood approach of Wedderburn (1974) which requires only specification of a mean and variance relationship.

GNMs can be fitted via maximum likelihood using an iteratively reweighted least squares algorithm (see, for example, Turner and Firth, 2007). An implementation is available in the `gnm` package (Turner and Firth, 2006), available for R (R Development Core Team, 2009).

Asymptotic variances for $\hat{\boldsymbol{\theta}}$, the maximum likelihood estimators of $\boldsymbol{\theta}$, can be obtained from the inverse of the Fisher information matrix

$$-E \left[\frac{\partial^2 f_Y}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] = \sum_{j=1}^N v(\mathbf{x}_j) \mathbf{f}(\mathbf{x}_j; \boldsymbol{\theta}) \mathbf{f}(\mathbf{x}_j; \boldsymbol{\theta})^T, \quad (3)$$

where $v(\mathbf{x}_j) = h(\mu_j)^{-1}(\partial \mu_j / \partial g(\mu_j))^2$ and $\mathbf{f}(\mathbf{x}_j; \boldsymbol{\theta})$ is the vector of derivatives of $\eta(\mathbf{x}_j; \boldsymbol{\theta})$ with respect to the elements of $\boldsymbol{\theta}$.

It is clear that model (1) is a GNM in which the response follows a Poisson distribution, the predictor is a nonlinear combination of the parameters $\boldsymbol{\theta}$, and the link is the identity.

In the remainder of this paper, the problem of designing experiments for GNMs is investigated. In common with other nonlinear models, the information matrix (3) depends on $\boldsymbol{\theta}$ and hence the performance of a design, as measured by functionals of (3), may vary with $\boldsymbol{\theta}$. Misspecification of $\boldsymbol{\theta}$ when planning an experiment may result in an inefficient design. In Section 2 the Bayesian D -optimality design criterion is described, together with supporting results on the reparameterisation of the model and methods for finding approximations to Bayesian designs to reduce the computational burden. In Section 3, optimal designs are found and assessed for the estimation of ratios of the model parameters. Section 4 gives a brief discussion.

2. Optimal designs for parameter estimation

A continuous design ξ is a probability measure defining a finite set of distinct support points in a design space $\mathcal{X} \subset \mathbb{R}^k$ and the proportions, ω_i , of observations to be taken at these points:

$$\xi = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ \omega_1 & \omega_2 & \cdots & \omega_n \end{array} \right\}, \quad 0 < \omega_i \leq 1, \quad \sum_{i=1}^n \omega_i = 1.$$

To implement a design in practice, a rounding procedure usually has to be applied to obtain exact designs such that the replication, n_i , of each support point is integer (for example, Pukelsheim and Rieder, 1992). The vector of polarisation angles (γ_i, Γ_i) defines a support point \mathbf{x}_i in an SHG experiment.

Most popular design selection criteria optimise a functional of the information matrix of ξ for parameter $\boldsymbol{\theta}$ which, from (3), is proportional to

$$M(\xi, \boldsymbol{\theta}) = \sum_{i=1}^n \omega_i v(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta}) \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta})^T.$$

When interest is in accurate parameter estimation, a D -optimal design, ξ^* , may be sought which minimises the asymptotic joint confidence ellipsoid for the parameter estimators $\hat{\boldsymbol{\theta}}$ or, equivalently, maximises

$$\phi(\xi; \boldsymbol{\theta}) = \frac{1}{p} \log |M(\xi, \boldsymbol{\theta})|. \quad (4)$$

For nonlinear models, objective function (4) is dependent on the values of the parameters $\boldsymbol{\theta}$, and a design that maximises (4) is said to be *locally D-optimal*. To compare the performance

of an arbitrary design ξ to a locally D -optimal design ξ^* , we use the efficiency of ξ , calculated as

$$\text{Eff}(\xi) = \left(\frac{|M(\xi, \boldsymbol{\theta})|}{|M(\xi^*, \boldsymbol{\theta})|} \right)^{1/p}. \quad (5)$$

As $\boldsymbol{\theta}$ is unknown prior to experimentation, a Bayesian criterion may be employed to find designs *robust* to the values of the model parameters. A Bayesian D -optimal design ξ_b^* maximises the average of (4) across the parameter space Θ with respect to a prior distribution $\pi(\boldsymbol{\theta})$

$$\Phi(\xi) = \int_{\Theta} \phi(\xi; \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}), \quad (6)$$

where $\Theta \subset \mathbb{R}^p$ is the parameter space. This objective function can be viewed as the pre-posterior loss from an asymptotic Normal approximation to the posterior distribution (Chaloner and Larntz, 1989). It was used by Woods et al. (2006) to obtain robust D -optimal designs for multi-variable GLMs.

As (6) is a concave and differentiable function of the design, an extension of the arguments of Chaloner and Larntz (1989, p.194) can be used to establish an equivalence theorem. In particular, it can be shown that, for a compact design space \mathcal{X} , a necessary and sufficient condition for a design ξ_b^* to be Bayesian D -optimal is

$$\int_{\Theta} v(\mathbf{x}) f(\mathbf{x}, \boldsymbol{\theta})^T M^{-1}(\xi_b^*, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}) \leq p, \quad (7)$$

for all $\mathbf{x} \in \mathcal{X}$, with equality at the support points of ξ_b^* . The corresponding result for a locally D -optimal design is the special case in which the prior distribution is a point mass located at the value of $\boldsymbol{\theta}$.

For many nonlinear models, there may be several possible parameterisations with different properties; for example, a particular parameterisation may improve the conditioning of a model (Batchelor et al., 2007). In the SHG example, the phenomenological formulation leads to more than one natural parameterisation (using Cartesian or polar coordinates). The following theorem, proved in the Appendix, gives necessary and sufficient conditions for the Bayesian D -optimality of a design for a GNM to be invariant to the choice of parameterisation. The invariance of local D -optimality follows easily as a special case of the theorem.

Theorem 1. Let $\psi(\cdot)$ be a differentiable function from $\Theta_1 \mapsto \Theta_2$, where $\Theta_i \subset \Theta$ ($i = 1, 2$) is an open subset of \mathbb{R}^p and let the derivative of $\psi(\cdot)$ be invertible. Then a design which is Bayesian D -optimal for estimating $\boldsymbol{\theta} \in \Theta_1$ is also Bayesian D -optimal for estimating $\psi(\boldsymbol{\theta}) \in \Theta_2$.

Two methods have been recently proposed for finding approximate solutions to the robust design problem which are less computationally demanding than the Bayesian approach. Both methods use a set of m locally D -optimal designs obtained for a sample of parameter values $\tilde{\boldsymbol{\theta}}_1, \dots, \tilde{\boldsymbol{\theta}}_m$ drawn from Θ .

1. Dror and Steinberg (2006) applied a K -means clustering algorithm to the union of the sets of design points of exact locally D -optimal designs. A robust design is then formed by taking the centroids of the resulting clusters as equally-weighted support points of the design (see also Russell et al., 2009).

2. Melas (2005) took a set of m locally optimal continuous designs and, for each design, calculated the D -efficiency (5) for each of the m values of the parameter vector. The design with the maximum value of the minimum efficiency is then considered a *most-robust locally optimal design*. This method offers considerable computational advantage over the application of a maximin criterion across all designs and parameter values.

These authors applied and assessed their methods on designs for GLMs, and for logistic, hyperexponential and rational regression models respectively.

2.1. Application to second harmonic generation experiments

We apply and compare the above methods using the SHG example. We start by establishing two results, both proved in the Appendix, which aid in finding Bayesian D -optimal designs. The first result shows that the same design is Bayesian D -optimal under both Cartesian and polar (Euler) representations of the parameters B and C in model (1) (Corollary 1 below). This allows attention to be restricted to the Euler parameterisation in the remainder of the paper. The second result (Lemma 1) enables a reduction in the design space that needs to be searched.

Corollary 1. Let $r_b \neq 0$ and $r_c \neq 0$. Then it follows from Theorem 1 that a design which is Bayesian D -optimal for estimating the parameters θ in the Euler representation is also Bayesian D -optimal for estimating their Cartesian counterparts.

Lemma 1. The design space $\mathcal{X} = [0, 2\pi] \times [0, 2\pi]$ can, without loss of generality, be reduced to either $\mathcal{X} = [0, \pi/2] \times [0, \pi]$ or $\mathcal{X} = [0, \pi] \times [0, \pi/2]$.

For the SHG experiment, both locally and Bayesian D -optimal designs were found numerically using a quasi-Newton method with numerical derivatives (the BFGS algorithm; Dennis and Schnabel, 1983). To aid convergence, an unconstrained optimisation problem was obtained through use of an arctan transformation for $\mathbf{x} = (\gamma, \Gamma)$ and a standard transformation for the weights ω , see Atkinson et al. (2007, ch. 9).

2.1.1. Locally D -optimal designs We first consider locally D -optimal designs for a vector of parameter values θ . The existence of such designs with $5 \leq n \leq 15$ support points is ensured through the compactness of the induced design space $\Xi = \{\sqrt{\nu(\mathbf{x})}f(\mathbf{x}; \theta) \mid \mathbf{x} = (\gamma, \Gamma) \in \mathcal{X}\}$ and Caratheodory's Theorem. Optimal designs were calculated for different values of θ , including the maximum likelihood estimates from Welsh et al. (2005), $\theta_0 = (0.889, 0.317, 0.588, 1.142, 0.301)$. Optimality of these designs was established using (7). Most designs had $n = 5, 6$ or 7 , and only designs with 5 support points had equal weights ω_i . Examples of these designs are given in Table 1.

The support of the locally D -optimal designs considered includes the point $(0, 0)$, or the equivalent points $(0, \pi)$, $(\pi, 0)$, (π, π) for this model, and also the point $(\pi/2, 0)$. Most designs also had support points close to $(\pi/4, \pi/2)$. The points $(0, 0)$ and $(\pi/2, 0)$ are stationary points of the regression function irrespective of the values of the parameters. Therefore considerable changes in the behaviour of the function may occur at these points, and this can be best detected by taking observations at these points. The designs in Table 1 are typical in that they have similar support points.

Table 2 gives the D -efficiencies of the locally D -optimal designs of Table 1 when the parameter values are misspecified. For example, if the true value of θ is θ_2 and we use the locally

Table 1: Support points (Γ, γ) and weights (ω) of the locally D -optimal designs for values $\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$

$\xi_{\boldsymbol{\theta}_0}: \boldsymbol{\theta}_0 = (0.889, 0.317, 0.588, 1.142, 0.301)$						
γ	$\pi/2$	1.068	0	0.788	1.952	2.593
Γ	0	0.332	0	1.446	0.788	0.934
ω	0.181	0.061	0.199	0.188	0.175	0.195
$\xi_{\boldsymbol{\theta}_1}: \boldsymbol{\theta}_1 = (1, 0.25, 0.5, 1, 0.5)$						
γ	1.928	0.784	0	2.555	$\pi/2$	
Γ	0.782	1.310	0	1.037	0	
ω	0.2	0.2	0.2	0.2	0.2	
$\xi_{\boldsymbol{\theta}_2}: \boldsymbol{\theta}_2 = (1, 0.5, 0.8, 1, 0.6)$						
γ	0.778	0	2.628	$\pi/2$	1.967	
Γ	1.326	0	0.922	0	0.766	
ω	0.2	0.2	0.2	0.2	0.2	

Table 2: D -efficiencies for the locally D -optimal designs $\xi_{\boldsymbol{\theta}_i}$ ($i = 1, 2, 3$) of Table 1 for $\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$

Design	Parameter vector		
	$\boldsymbol{\theta}_0$	$\boldsymbol{\theta}_1$	$\boldsymbol{\theta}_2$
$\xi_{\boldsymbol{\theta}_0}$	1	0.904	0.974
$\xi_{\boldsymbol{\theta}_1}$	0.814	1	0.905
$\xi_{\boldsymbol{\theta}_2}$	0.942	0.862	1

optimal design, $\xi_{\boldsymbol{\theta}_0}$, for $\boldsymbol{\theta}_0$, then the efficiency for estimating the model parameters is 0.974. The locally D -optimal designs have generally high D -efficiencies for the three sets of parameter values. The lowest efficiency, 0.814, is obtained if the true value is $\boldsymbol{\theta}_0$, and the experimenter has misspecified this value as $\boldsymbol{\theta}_1$. Design $\xi_{\boldsymbol{\theta}_0}$ is slightly more robust with a lowest efficiency of 0.904 among the three parameter vectors; this may be a result of the design having an extra support point.

2.1.2. Robust designs We investigate Bayesian D -optimal designs for the SHG experiment using two joint prior distributions for $\boldsymbol{\theta}$ which reflect different degrees of prior uncertainty. A beta distribution on $[0, 1]$ was used to describe each radius, r_a , r_b and r_c , and a von Mises distribution on $[0, 2\pi]$ was used for each angle, ϕ_b and ϕ_c . For the first prior distribution, $\pi_1^c(\boldsymbol{\theta})$, the respective modes of the beta distributions were chosen to be 0.889, 0.317, and 0.588, to match the first three entries of $\boldsymbol{\theta}_0$ (Section 2.1.1). Similarly, the von Mises distributions for ϕ_b and ϕ_c were centred on 1.142 and 0.301, respectively, both with dispersion parameter $\kappa = 4$, corresponding to moderate prior knowledge. The second prior distribution, $\pi_2^c(\boldsymbol{\theta})$, represents greater prior uncertainty through using flat beta(1, 1) distributions (i.e. uniform) for the radii, and more diffuse von Mises distributions with $\kappa = 2$ and centred as above.

To facilitate numerical computation of Bayesian D -optimal designs using (6), each $\pi_i^c(\boldsymbol{\theta})$ ($i = 1, 2$) was approximated by a discrete distribution, $\pi_i(\boldsymbol{\theta})$, with support consisting of a random sample $\tilde{\Theta}_i \subset \Theta$ of 1000 parameter vectors from $\pi_i^c(\boldsymbol{\theta})$ which were assigned equal probabilities. The designs found, ξ_1 and ξ_2 , have 16 and 23 support points, respectively. The

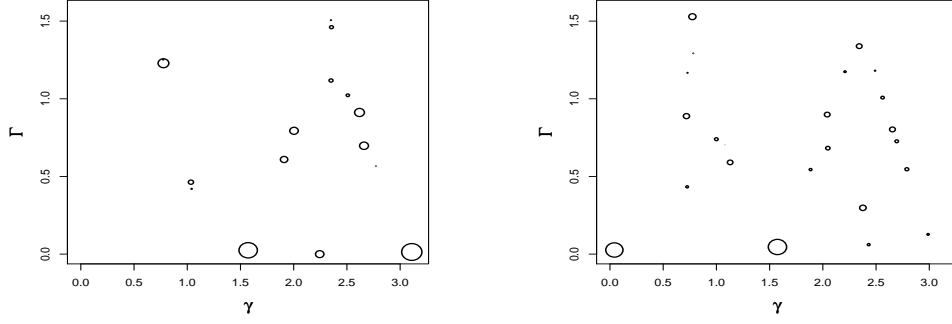


Figure 2: Bayesian D -optimal designs with respect to $\pi_1(\boldsymbol{\theta})$ (ξ_1 , left) and $\pi_2(\boldsymbol{\theta})$ (ξ_2 , right)

support points are shown in Fig. 2, using a plotting symbol whose size is proportional to the corresponding weight ω_i , in common with all plots of designs in this paper. Both designs have support points which include $(0, 0)$ and $(\pi/2, 0)$. These points have the largest weights and also are included in the support of the locally D -optimal designs in Section 2.1.1.

The performance of ξ_i ($i = 1, 2$) was evaluated by calculating D -efficiency (5) with respect to the locally D -optimal design for each element of $\tilde{\Theta}_1$ and $\tilde{\Theta}_2$. For design ξ_1 the median and minimum efficiencies are respectively 0.82 and 0.14 across $\tilde{\Theta}_1$, with corresponding values for $\tilde{\Theta}_2$ of 0.75 and 0.09. For design ξ_2 , the respective efficiencies are 0.78 and 0.16 for $\tilde{\Theta}_1$, and 0.79 and 0.21 for $\tilde{\Theta}_2$. Thus each design is fairly robust to the choice of prior distribution, with higher efficiencies achieved under the less diffuse discrete distribution $\pi_1(\boldsymbol{\theta})$, and ξ_2 having higher minimum efficiency than ξ_1 across the support of both discrete distributions.

2.1.3. Cluster designs Two possible designs for the SHG experiment were found by a modification of the method of Dror and Steinberg (2006). Each design was obtained by clustering the points in the union U_i of the design points from the $m = 1000$ exact locally D -optimal designs $\tilde{\xi}_{ij}$ ($j = 1, \dots, m; i = 1, 2$). The designs $\tilde{\xi}_{ij}$ were obtained for $N = 50$ by applying rounding to a continuous locally D -optimal design for each parameter vector in $\tilde{\Theta}_i$. The application of clustering to design points, rather than support points, has the advantage of overcoming any large differences in the design weights, ω . A preliminary visual inspection of the points in each U_i showed evidence of clustering which could be exploited by this method.

To construct a continuous cluster design, we generalise the method of Dror and Steinberg (2006) by allowing unequally-weighted support points, where each weight is proportional to the corresponding cluster size to reflect the relative importance of each cluster. The number of clusters, i.e. support points, was chosen by investigating a range of possible values. For each value, the D -efficiency of a cluster design was calculated via (5) relative to each of the m locally D -optimal designs found for the parameter values in $\tilde{\Theta}_i$ ($i = 1, 2$), summarised in Fig. 3. The median efficiencies first increase sharply with the number of clusters and then level-off to about 0.79 for $\tilde{\Theta}_1$ and 0.75 for $\tilde{\Theta}_2$; the minimum efficiencies show greater local variation. Hence we chose 15 clusters for both cases, as a value where the median has reached a plateau and satisfactory minimum efficiencies are also achieved. The resulting cluster designs, ξ_3 and ξ_4 , shown in Fig. 4, have minimum efficiencies of 0.16 and 0.15, respectively, for $\tilde{\Theta}_1$ and $\tilde{\Theta}_2$. In common with the Bayesian designs (Fig. 2), each design gives high weight to $(0, 0)$ and $(\pi/2, 0)$, and moderately high weight to $(\pi/4, \pi/2)$ and $(3\pi/4, \pi/2)$ or points nearby.

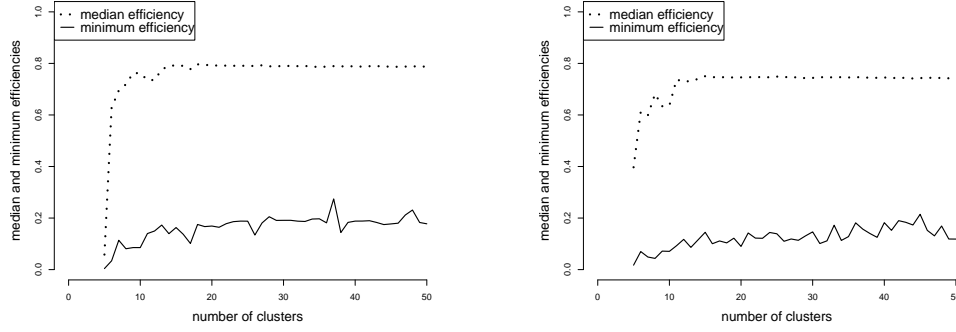


Figure 3: The median and the minimum D -efficiencies of the cluster designs with 5 to 50 support points for $\tilde{\Theta}_1$ (left) and $\tilde{\Theta}_2$ (right)

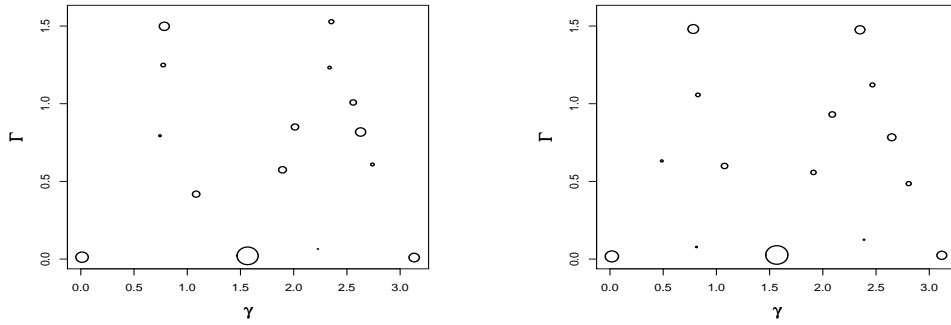


Figure 4: The 15-point cluster designs for $\tilde{\Theta}_1$ (ξ_3 , left) and $\tilde{\Theta}_2$ (ξ_4 , right)

2.1.4. Most-robust locally D -optimal designs We found the most-robust locally D -optimal designs, ξ_5 and ξ_6 , through comparison with locally D -optimal designs for $\tilde{\Theta}_1$ and $\tilde{\Theta}_2$. Both designs have 6 support points, see Fig. 5; for ξ_6 the point $(0.70, 0.45)$ with small weight was found to be necessary for optimality. Again, the points $(\pi, 0)$, equivalent to $(0, 0)$, and $(\pi/2, 0)$ features with high weight. Designs ξ_5 and ξ_6 are locally D -optimal for $\boldsymbol{\theta} = (0.67, 0.55, 0.57, 1.26, 6.20)$ and $\boldsymbol{\theta} = (0.79, 0.84, 0.94, 1.34, 0.69)$, respectively. The median (minimum) efficiencies are 0.76 (0.16) for ξ_5 and 0.63 (0.09) for ξ_6 . The minimum efficiencies are achieved at $\boldsymbol{\theta} = (0.33, 0.26, 0.69, 0.15, 6.26)$ and $\boldsymbol{\theta} = (0.87, 0.10, 0.48, 0.16, 0.07)$.

2.2. Comparison of designs

The performance of the following 10 designs were compared using D -efficiency: the factorial design ξ_0 used by Welsh et al. (2005); locally D -optimal designs ξ_{θ_0} and ξ_{θ_2} from Section 2.1.1; robust designs ξ_1, \dots, ξ_6 discussed in Section 2.1.2; and a second factorial design ξ_7 , with 8 equally spaced input angles $\Gamma = 0, \pi/8, \dots, 7\pi/8$ and 5 equally spaced output angles $\gamma = 0, \pi/8, \pi/4, 3\pi/8, \pi/2$. This last design has greater range of values for Γ than is present in ξ_0 .

Designs were assessed via simulation from $\pi_1^c(\boldsymbol{\theta})$ and $\pi_2^c(\boldsymbol{\theta})$, following Woods et al. (2006). Independent samples of parameter vectors of size 1000 are drawn from each distribution and locally D -optimal designs are found for each parameter vector. The D -efficiency (5) of each of

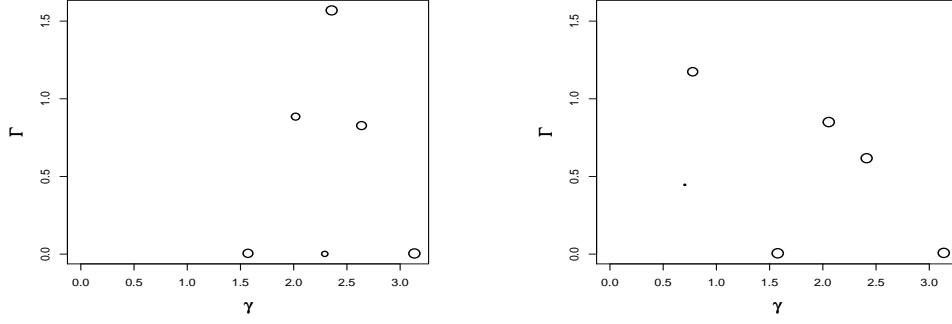


Figure 5: Most-robust locally optimal designs for $\tilde{\Theta}_1$ (ξ_5 , left) and $\tilde{\Theta}_2$ (ξ_6 , right)

$\xi_0, \dots, \xi_7, \xi_{\theta_0}, \xi_{\theta_2}$ was then calculated with respect to each locally optimal design. The resulting approximations to the distributions of the efficiencies can be used to compare the designs using, for example, median and minimum efficiencies, and are summarised in Fig. 6 and 7.

For both $\pi_1^c(\boldsymbol{\theta})$ and $\pi_2^c(\boldsymbol{\theta})$, the Bayesian D -optimal designs, ξ_1 and ξ_2 , for the correctly specified prior distribution have the best performance (median efficiencies 0.81 and 0.77 respectively). Mis-specification of the prior distribution has only a small impact on design performance, with ξ_1 having median efficiency 0.74 under $\pi_2^c(\boldsymbol{\theta})$ and ξ_2 having median efficiency 0.78 under $\pi_1^c(\boldsymbol{\theta})$.

The cluster designs, ξ_3 and ξ_4 , perform similarly to the corresponding Bayesian D -optimal designs, ξ_1 and ξ_2 , with respective median efficiencies 0.79 and 0.75 under $\pi_1^c(\boldsymbol{\theta})$ and $\pi_2^c(\boldsymbol{\theta})$. Hence both cluster designs are good alternatives to the Bayesian designs with their construction requiring less computation. The most-robust locally D -optimal design, ξ_5 , for $\tilde{\Theta}_1$ performs well under both $\pi_1^c(\boldsymbol{\theta})$ and $\pi_2^c(\boldsymbol{\theta})$ (median efficiencies 0.77 and 0.71). However, the same is not true for ξ_6 , the corresponding design for $\tilde{\Theta}_2$. The wider spread of parameter values in $\tilde{\Theta}_2$ leads to a design which is locally optimal for a more extreme parameter vector, and hence has lower efficiency across the majority of the parameter space.

All the remaining designs perform poorly overall. The 3×14 factorial, ξ_0 , has the poorest performance with median efficiency 0.35 under $\pi_1^c(\boldsymbol{\theta})$ and 0.41 under $\pi_2^c(\boldsymbol{\theta})$. A comparison of Figs. 6 and 7 shows that the differences between the performance of the designs are less for the more diffuse distribution $\pi_2^c(\boldsymbol{\theta})$.

3. Optimal designs for estimating ratios of parameters

In several application areas, experimenters require accurate estimates of a set of ratios of model parameters. Examples occur in science, as in the SHG example, and in discrete choice and ranking experiments in marketing, where the marginal rate of substitution is of particular interest. This rate measures the relative importance of different product attributes or features. An important marginal rate of substitution is the willingness-to-pay, i.e. the ratio of model parameters relative to the parameter for price.

A key step in finding optimal designs is the determination of a first-order approximation to the variance-covariance matrix of the estimators. Suppose that $\tau(\cdot)$ is a function which maps the p -vector $\boldsymbol{\theta}$ to ν ratios of parameters ($1 \leq \nu \leq p - 1$). As $\tau(\cdot)$ is a differentiable function of the original parameters, the delta-method (see, for example, Serfling, 1980) can be applied. We consider any set of ν *admissible* ratios such that the $p \times \nu$ Jacobian matrix J_τ of $\tau(\cdot)$ is of

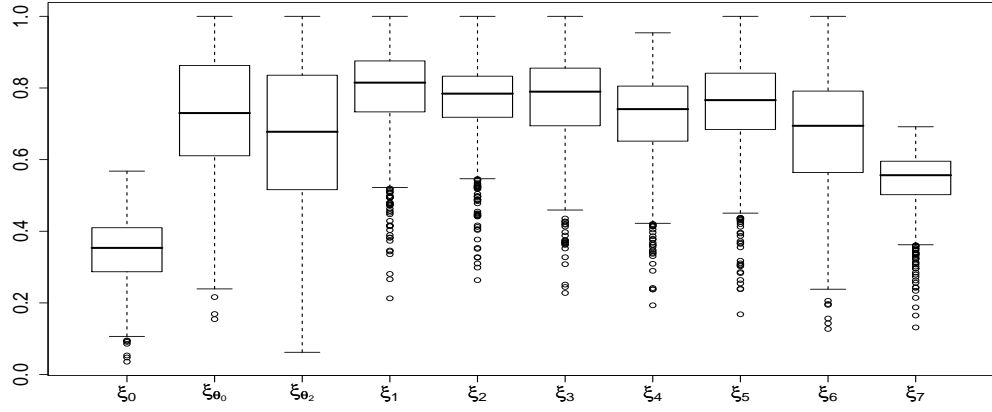


Figure 6: Boxplots of the D -efficiencies for 10 different designs with respect to 1000 parameter vectors from $\pi_1^c(\boldsymbol{\theta})$

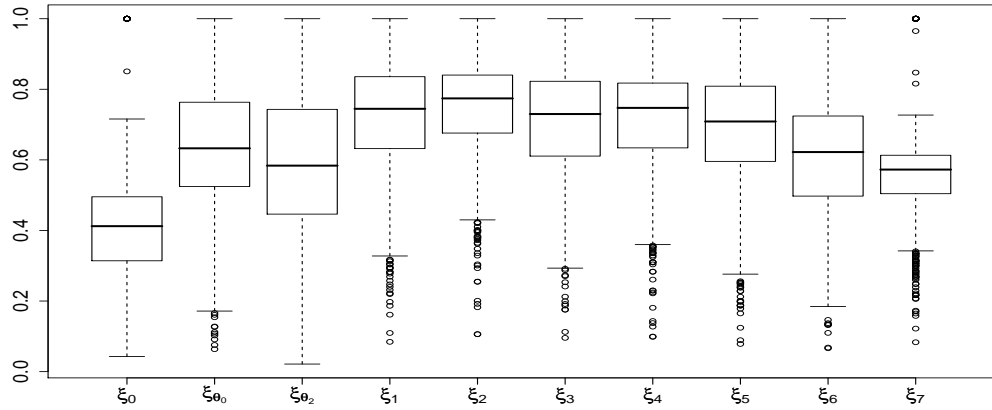


Figure 7: Boxplots of the D -efficiencies for 10 different designs with respect to 1000 parameter vectors from $\pi_2^c(\boldsymbol{\theta})$

rank ν and hence $C(\xi, \boldsymbol{\theta}) = J_\tau^T M^-(\xi, \boldsymbol{\theta}) J_\tau$ is the approximate asymptotic variance-covariance matrix of $\hat{\tau}(\boldsymbol{\theta})$, with $M^-(\xi, \boldsymbol{\theta})$ being a generalised inverse of the information matrix. Therefore, analogous to (6), a Bayesian D -optimal design for estimating the ratios $\tau(\boldsymbol{\theta})$ maximises the objective function

$$\Phi_r(\xi; \boldsymbol{\theta}) = \int_{\Theta} \log |C^{-1}(\xi, \boldsymbol{\theta})| d\pi(\boldsymbol{\theta}).$$

Lemma 2, proved in the Appendix, shows that the optimal design for estimating ν ratios, involving $2 \leq p_1 \leq p$ parameters, does not depend on the particular choice of ν ratios of these parameters.

Lemma 2. Let ξ_b^* be the Bayesian D -optimal design for estimating $1 \leq \nu \leq p - 1$ admissible ratios given by $\boldsymbol{\tau} = \tau(\boldsymbol{\theta})$, involving $2 \leq p_1 \leq p$ parameters. Then ξ_b^* is also Bayesian D -optimal for estimating any other set, $\boldsymbol{\kappa} = \kappa(\boldsymbol{\theta})$, of ν admissible ratios involving the same p_1 parameters.

3.1. Optimal designs for the estimation of two ratios in the example

For model (1) the ratios of the complex values A , B and C take the form (2). From Lemma 2, without loss of generality, we can restrict ourselves to considering just B/A and C/A , that is, estimation of $\psi(\boldsymbol{\theta}) = (r_b/r_a, r_c/r_a, \phi_b, \phi_c)$.

A related problem is to estimate two ratios and, in addition, one of the radii. Here, the ratios can be used to compare different experiments and the additional radius would serve as a baseline. In this case, the optimal designs for estimating the model parameters, found in the previous section, can be shown to be D -optimal for this new problem. The proof of Corollary 2 can be found in the Appendix.

Corollary 2. Suppose r_a, r_b and r_c are non-zero, and let ξ_b^* be the Bayesian D -optimal design for estimating $\boldsymbol{\theta} = (r_a, r_b, r_c, \phi_b, \phi_c)$. Then ξ_b^* is also Bayesian D -optimal for simultaneously estimating two admissible ratios and one of the radii, either r_a , r_b or r_c .

For notational convenience, the ratios $\tau(\boldsymbol{\theta})$ are denoted by $\boldsymbol{\theta}^r$. Fig. 8 shows the locally D -optimal designs for the ratios $\boldsymbol{\theta}_0^r$ and $\boldsymbol{\theta}_2^r$ corresponding to the parameter vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_2$ from Section 2.1.1.

Both locally optimal designs have six support points. As for the locally D -optimal designs for parameter estimation, we found from our study that most locally D -optimal designs for estimating ratios have 5-7 support points, and all are supported on the points $(0, 0)$, or equivalently $(\pi, 0)$, and $(\pi/2, 0)$.

To find robust designs, we use the same prior distributions, $\pi_1^c(\boldsymbol{\theta})$ and $\pi_2^c(\boldsymbol{\theta})$, as in Section 2.1.2. The resulting marginal densities for the two differences between angles are again von Mises densities. For the two ratios of radii, the induced marginal densities are ratios of two independent beta densities (Pham-Gia, 2000) for $\pi_1^c(\boldsymbol{\theta})$, and a ratio of two independent uniform densities for $\pi_2^c(\boldsymbol{\theta})$, given by

$$f(u) = \begin{cases} 0.5 & 0 \leq u \leq 1 \\ 1/2u^2 & u > 1. \end{cases}$$

Fig. 9 shows the marginal prior densities for the ratio r_c/r_a corresponding to the prior distributions $\pi_1^c(\boldsymbol{\theta})$ (ratio of beta densities) and $\pi_2^c(\boldsymbol{\theta})$ (ratio of uniform densities) for $\boldsymbol{\theta}$.

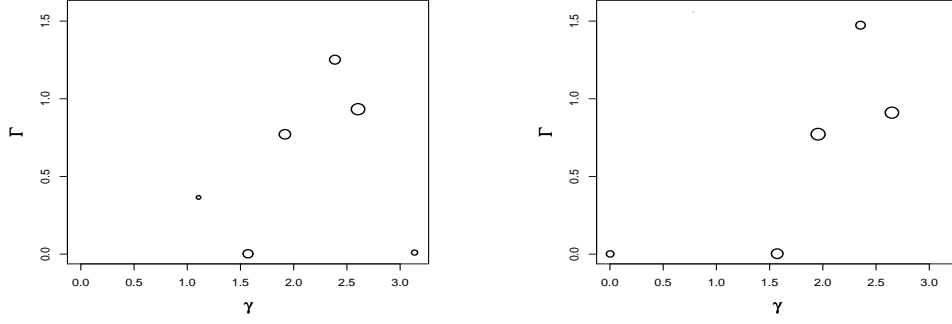


Figure 8: The locally D -optimal designs for estimating the ratios $\theta_0^r = (0.357, 0.661, 1.142, 0.301)$ ($\xi_{\theta_0}^r$, left) and $\theta_2^r = (0.5, 0.8, 1, 0.6)$ ($\xi_{\theta_2}^r$, right)

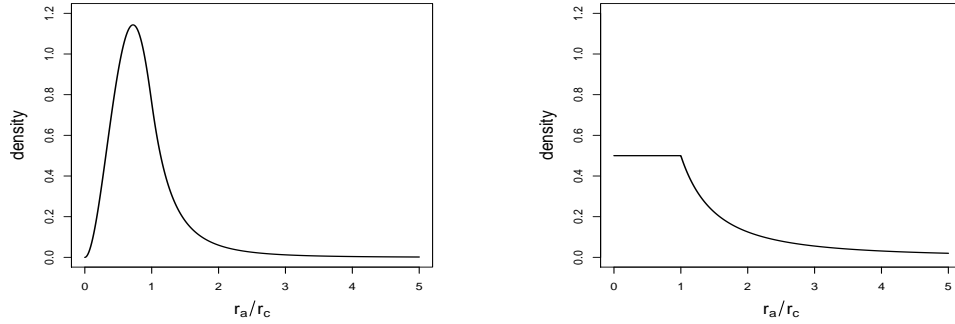


Figure 9: Densities of the prior distribution for r_c/r_a with respect to $\pi_1^c(\theta)$ (left) and $\pi_2^c(\theta)$ (right)

As in Section 2.1.2, the continuous distributions $\pi_1^c(\theta)$ and $\pi_2^c(\theta)$ are approximated by discrete distributions $\pi_1(\theta)$ and $\pi_2(\theta)$. Fig. 10 shows the Bayesian D -optimal designs for $\pi_1(\theta)$ and $\pi_2(\theta)$; cluster designs on 15 points, found using $\tilde{\Theta}_1$ and $\tilde{\Theta}_2$ as in Section 2.1.3, had similar support points and weights. The distribution of points in these designs is very similar to the corresponding designs from Section 2.1.2. This suggests that the performance of designs optimal for parameter estimation will be similar to the performance of designs for the estimation of the ratios.

3.2. Comparison of designs

Designs are compared using simulation, analogous to Section 2.2, with 1000 parameter vectors drawn from each of $\pi_1(\theta)$ and $\pi_2(\theta)$. As the Bayesian D -optimal and the cluster designs from Section 2.1.2 for the problem of parameter estimation outperform the other designs, we evaluate the corresponding designs, ξ_1^r , ξ_2^r , ξ_3^r and ξ_4^r , for estimating the ratios. To assess the robustness of locally optimal designs, we also compare locally D -optimal designs, $\xi_{\theta_0}^r$ and $\xi_{\theta_2}^r$, for estimating the ratios under $\theta_0^r = (0.357, 0.661, 1.142, 0.301)$ and $\theta_2^r = (0.5, 0.8, 1, 0.6)$ respectively. We also consider ξ_0 and the second factorial design ξ_7 to assess their performance

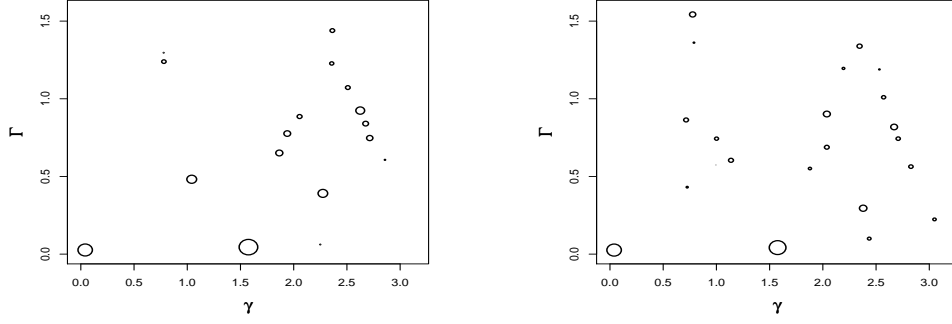


Figure 10: The Bayesian D -optimal designs for estimating ratios with respect to the discrete prior distributions $\pi_1(\boldsymbol{\theta})$ (ξ_1^r , left) and $\pi_2(\boldsymbol{\theta})$ (ξ_2^r , right)

for estimating ratios of parameters.

We also reassess the Bayesian D -optimal designs, ξ_1 and ξ_2 , and the cluster designs, ξ_3 and ξ_4 from Section 2.1. These designs require less computational effort to find, as designs for ratio estimation require a matrix inversion at each step of the optimisation. Also, if the aim of the experiment is the estimation of both the model parameters and their ratios, designs are needed which are efficient for both purposes.

Figs. 11 and 12 show the boxplots of the D -efficiencies from the comparative study for these designs. The results are similar to those in Section 2.2: the Bayesian D -optimal design for the correct prior distribution is the most robust design, closely followed by the cluster designs; mis-specification of the prior distribution does not have a substantial effect on the performance of the Bayesian and cluster designs. However, for estimation of the ratios, the locally optimal designs, $\xi_{\boldsymbol{\theta}_0}^r$ and $\xi_{\boldsymbol{\theta}_2}^r$, have much better performance for $\pi_1^c(\boldsymbol{\theta})$ than for the more diffuse $\pi_2^c(\boldsymbol{\theta})$ with median efficiency for $\xi_{\boldsymbol{\theta}_0}^r$ ($\xi_{\boldsymbol{\theta}_2}^r$) of 0.64 (0.59) for $\pi_1^c(\boldsymbol{\theta})$ and 0.47 (0.34) for $\pi_2^c(\boldsymbol{\theta})$. The optimal designs for parameter estimation, ξ_1, \dots, ξ_4 , almost match the performance of their counterparts for the estimation of the ratios, and so could be used if there was interest in both these goals.

4. Discussion

When an observed response from a scientific process is not normally distributed, a generalised nonlinear model with the systematic predictor formed from physical theory may be an appropriate description. We have investigated designs for such models, including their theoretical properties, numerical construction and efficiency, through an application of a GNM to second harmonic generation.

In common with designs for other nonlinear models, the performance of designs for GNMs may depend on the unknown model parameters, and robust designs are required to overcome this dependency. We have applied various methods of finding robust designs to an SHG example, for estimating model parameters, and also for estimating ratios of the parameters. Our findings include: (i) robust designs tend to have a greater number of support points; (ii) cluster designs built from locally optimal designs can provide effective, and more easily computed, alternatives to Bayesian designs; (iii) factorial designs with uniform spacing of levels perform poorly, as they do not incorporate any prior information or knowledge of the form of the model. In addition,

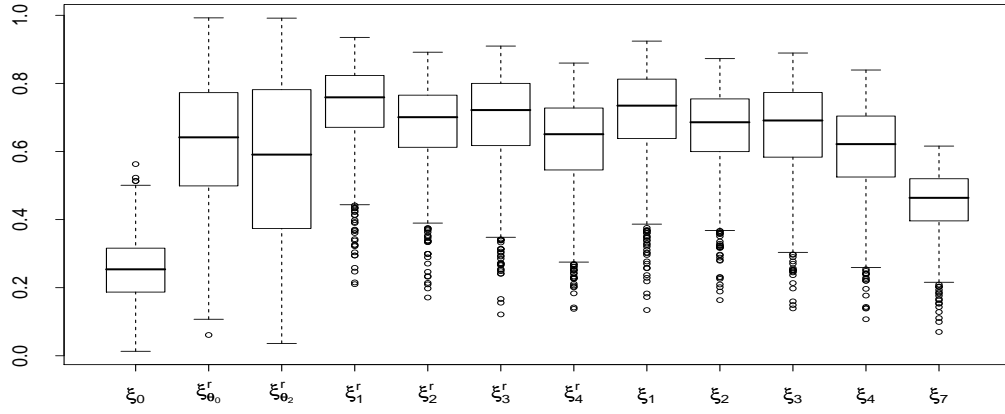


Figure 11: Boxplots of the D -efficiencies for the estimation of ratios for 12 different designs for 1000 vectors from $\pi_1^c(\boldsymbol{\theta})$

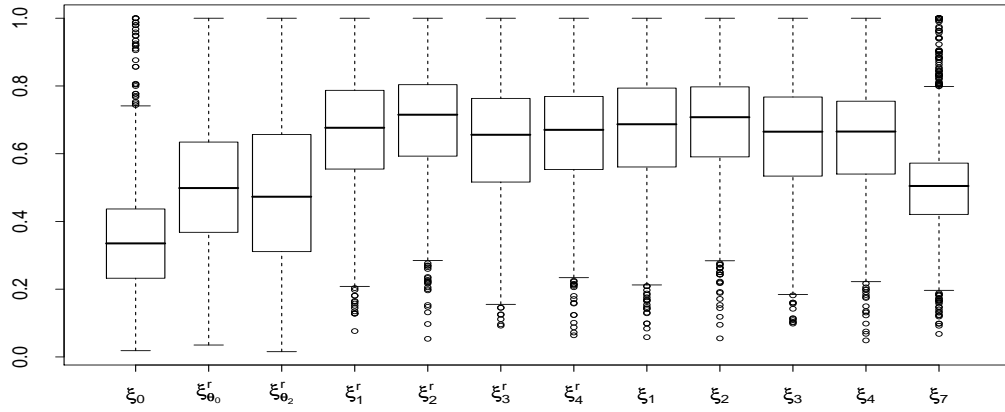


Figure 12: Boxplots of the D -efficiencies for the estimation of ratios for 12 different designs for 1000 vectors from $\pi_2^c(\boldsymbol{\theta})$

robust designs for estimating all the parameters were found to be efficient for the estimation of ratios of parameters.

The use of cluster designs has a computational advantage over Bayesian designs in that design construction can be easily implemented using distributed computing in an “embarrassingly parallel fashion”, i.e. locally optimal designs for different parameter vectors can be generated on different computers or nodes. The simulation-based design assessment, again using many locally optimal designs, can be implemented similarly.

Acknowledgments

The work of D.C. Woods was supported by EPSRC grant EP/C008863/1. We are grateful to Jeremy Frey (School of Chemistry, Southampton) for raising the problem and providing Fig. 1, and to him and Susan Lewis (Southampton Statistical Sciences Research Institute) for helpful discussions.

Appendix A. Proofs

Proof of Theorem 1. Let $M^{-1}(\xi, \boldsymbol{\theta})$ denote the asymptotic covariance matrix of the estimator for $\boldsymbol{\theta} \in \Theta_1$. From the delta-method, it follows that the asymptotic covariance matrix for the estimator of $\psi(\boldsymbol{\theta})$ is given by $J_\psi^T M^{-1}(\xi, \boldsymbol{\theta}) J_\psi$ where J_ψ is the Jacobian matrix of ψ . Since the derivative of ψ is invertible, its Jacobian matrix $J_\psi \in \mathbb{R}^{p \times p}$ is of full rank. So the determinant of the covariance matrix is given by

$$|J_\psi^T M^{-1}(\xi, \boldsymbol{\theta}) J_\psi| = |J_\psi|^2 |M^{-1}(\xi, \boldsymbol{\theta})|. \quad (8)$$

Now, J_ψ does not depend on the design ξ , and so a design which minimises $M^{-1}(\xi, \boldsymbol{\theta})$ will also minimise $|J_\psi|^2 |M^{-1}(\xi, \boldsymbol{\theta})|$. For Bayesian D -optimality we take logs in (8) and integrate with respect to the given prior distribution $\pi(\boldsymbol{\theta})$ to obtain

$$\begin{aligned} \int \log(|J_\psi^T M^{-1}(\xi, \boldsymbol{\theta}) J_\psi|) d\pi(\boldsymbol{\theta}) &= \int \log(|J_\psi|^2 |M^{-1}(\xi, \boldsymbol{\theta})|) d\pi(\boldsymbol{\theta}) \\ &= \int \log(|J_\psi|^2) d\pi(\boldsymbol{\theta}) + \int \log(|M^{-1}(\xi, \boldsymbol{\theta})|) d\pi(\boldsymbol{\theta}). \end{aligned}$$

The first term does not depend on the design, which proves the assertion.

Proof of Corollary 1. Let $\psi(\cdot)$ be the function transforming the polar coordinates of a point $\boldsymbol{\theta}$ from the Euler representation to their Cartesian counterparts, i.e. $(x_a, x_b, x_c, y_b, y_c) = \psi(r_a, r_b, r_c, \phi_b, \phi_c) = (r_a, r_b \cos(\phi_b), r_c \cos(\phi_c), r_b \sin(\phi_b), r_c \sin(\phi_c))$. As the determinant of the Jacobian matrix of $\psi(\cdot)$ is given by $r_b r_c \neq 0$, the assertion follows from Theorem 1.

Proof of Lemma 1.

(1) The information matrix is shift invariant with respect to a shift of magnitude π . This follows from the equalities $f((\gamma, \Gamma); \boldsymbol{\theta}) = f((\gamma + \pi, \Gamma), \boldsymbol{\theta}) = f((\gamma, \Gamma + \pi); \boldsymbol{\theta}) = f((\gamma + \pi, \Gamma + \pi); \boldsymbol{\theta})$ and $v[(\gamma, \Gamma)] = v[(\gamma + \pi, \Gamma)] = v[(\gamma, \Gamma + \pi)] = v[(\gamma + \pi, \Gamma + \pi)]$.

(2) For each value of $\boldsymbol{\theta}$ we have $f((\gamma, \Gamma); \boldsymbol{\theta}) = f((\pi - \gamma, \pi - \Gamma); \boldsymbol{\theta})$ and $v[(\gamma, \Gamma)] = v[(\pi - \gamma, \pi - \Gamma)]$, and therefore the points (γ, Γ) and $(\pi - \gamma, \pi - \Gamma)$ provide equivalent information in the design.

The assertion now follows directly by combining (1) and (2).

Proof of Lemma 2. As $\boldsymbol{\tau}$ and $\boldsymbol{\kappa}$ involve the same set of p_1 model parameters, from the delta-method, the asymptotic variance-covariance matrix for the estimator of $\boldsymbol{\kappa}$ is given by $L^T C(\xi, \boldsymbol{\theta}) L$, where L is the $\nu \times \nu$ full rank Jacobian matrix of $\boldsymbol{\kappa}$ with respect to $\boldsymbol{\tau}$. Hence, the corollary is established via an application of Theorem 1, with $\psi(\cdot)$ mapping from $\boldsymbol{\tau}$ to $\boldsymbol{\kappa}$.

Proof of Corollary 2. If all the radii are non-zero, the Jacobian matrix J_τ of a function $\tau(\boldsymbol{\theta})$ mapping $\boldsymbol{\theta}$ to a vector consisting of the Euler parameters, which describe two admissible ratios, and of one of the radii is a 5×5 -matrix of full rank. The assertion is therefore established via an application of Theorem 1.

References

- Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007) *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press, 2nd edn.
- Batchelor, A., Turner, H. L. and Firth, D. (2007) Nonlinear discrete-time hazard models for entry into marriage. *Tech. Rep. 07-16*, Centre for Research in Statistical Methodology, University of Warwick.
- Campagnola, P. and Loew, L. M. (2003) Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nature Biotechnology*, **21**.
- Chaloner, K. and Larntz, K. (1989) Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, **21**, 191–208.
- Dennis, Jr., J. E. and Schnabel, R. B. (1983) *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice Hall Series in Computational Mathematics. Englewood Cliffs, NJ: Prentice Hall Inc.
- Dror, H. A. and Steinberg, D. M. (2006) Robust experimental design for multivariate generalized linear models. *Technometrics*, **48**, 520–529.
- Fordyce, A. J., Bullock, W. J., Timson, A. J., Haslam, S., Spencer-Smith, R. D., Alexander, A. and Frey, J. G. (2001) The temperature dependence of surface second harmonic generation from the air-water interface. *Molecular Physics*, **99**, 677–687.
- Kosmidis, I. and Firth, D. (2008) Bias reduction in exponential family nonlinear models. *Tech. Rep. 08-05*, Centre for Research in Statistical Methodology, University of Warwick.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. New York: Chapman and Hall, 2nd edn.
- Melas, V. (2005) On the functional approach to optimal designs for nonlinear models. *Journal of Statistical Planning and Inference*, **132**, 93–116.
- Mizrahi, V. and Sipe, J. E. (1988) Phenomenological treatment of surface second-harmonic generation. *Journal of the Optical Society of America B*, **5**, 660–667.
- Pham-Gia, F. (2000) Distributions of the ratios of independent beta variables and applications. *Communications in Statistics - Theory and Methods*, **12**, 2693–2715.

- Pukelsheim, F. and Rieder, S. (1992) Efficient rounding of approximate designs. *Biometrika*, **79**, 763–770.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Russell, K. G., Woods, D. C., Lewis, S. M. and Eccleston, J. A. (2009) D-optimal designs for poisson regression models. *Statistica Sinica*, **19**, 721–730.
- Salafsky, J. S. (2007) Second-harmonic generation for studying structural motion of biological molecules in real time and space. *Physical Chemistry Chemical Physics*, **9**, 5704–5711.
- Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1st edn.
- Shen, Y. R. (1989) Surface properties probed by second-harmonic and sum-frequency generation. *Nature*, **337**, 519–525.
- Turner, H. L. and Firth, D. (2006) *Generalized nonlinear models in R: An overview of the gnm package*. URL <http://go.warwick.ac.uk/heatherturner/gnm/gnmoverview.pdf>. R package version 0.9-1.
- (2007) gnm: A package for Generalized Nonlinear Models. *R News*, **7**, 8–12.
- Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–447.
- Wei, B. (1997) *Exponential Family Nonlinear Models*. New York: Springer.
- Welsh, A. H., Mansson, R. A., Frey, J. G. and Danos, L. (2005) Statistical analysis of second harmonic generation experiments: a phenomenological model. *chemometrics and Intelligent Laboratory Systems*, **75**, 45–54.
- Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006) Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, **48**, 284–292.