2009-08

Designing Experiments for Nonlinear Models - An Introduction

Johnson, Rachel T.; Montgomery, Douglas C.

Quality and Reliability Engineering International, Volume 26, pp. 431-441, 2009
https://hdl.handle.net/10945/44163

Research Article

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1063

Published online 17 August 2009 in Wiley InterScience

Designing Experiments for Nonlinear Models—An Introduction‡

Rachel T. Johnson[a]*† and Douglas C. Montgomery[b]

We illustrate the construction of Bayesian *D*-optimal designs for nonlinear models and compare the relative efficiency of standard designs with these designs for several models and prior distributions on the parameters. Through a relative efficiency analysis, we show that standard designs can perform well in situations where the nonlinear model is intrinsically linear. However, if the model is nonlinear and its expectation function cannot be linearized by simple transformations, the nonlinear optimal design is considerably more efficient than the standard design. Published in 2009 by John Wiley & Sons, Ltd.

Keywords: optimal design; factorial design; Bayesian *D*-optimal

1. Introduction

Experimenters often have a model in mind when planning an experiment. For example, if the experiment involves three quantitative factors logical model choices are the main effects plus interactions model or a second-order (quadratic) response surface model. The type of model one anticipates often helps to determine the choice of experimental design. In the present example, the main effects plus interactions model would lead to a factorial experiment such as a 2^3, and the second-order model could lead to a central composite design or a Box–Behnken design. It is even possible that after running the factorial experiment, the experimenter could decide that the anticipated model is wrong, and then augment the factorial with additional runs to form a central composite design and fit the quadratic.

Experience and process knowledge often dictate the type of model that one anticipates. The same experience and process knowledge often suggest that the response variable will be best analyzed and modeled in a transformed scale. For example, suppose that we are planning an experiment to study the viscosity of a product (y) as a function of time x_1 and temperature x_2. Previously, success has been found by modeling viscosity on a log scale. A first-order model with interaction is a likely candidate for the final model, so we expect to end up with a model such as

$$\ln\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2 \tag{1}$$

Now in the original viscosity scale, the fitted model is

$$\hat{y} = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2)$$

This implies that the true relationship between viscosity, time, and temperature is

$$y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)\varepsilon \tag{2}$$

where ε is the random error in the system. Notice that the error term is multiplicative. Because the error term multiplies the expectation function part of the model, taking logarithms of both sides produces

$$\ln y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

and this leads to the fitted model in Equation (1).

It is the multiplicative error term in Equation (2) that makes this work out so nicely. Equation (2) is a nonlinear model, but it is a special case; one that can be made linear by a simple transformation. This is called an intrinsically linear model. However, suppose that the true system model is

$$y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2) + \varepsilon \tag{3}$$

[a] *Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, U.S.A.*
[b] *Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287, U.S.A.*
* *Correspondence to: Rachel T. Johnson, Naval Postgraduate School, Operations Research Department, 1411 Cunningham Rd., Monterey, CA 93943, U.S.A.*
† *E-mail: rtjohnso@nps.edu*

Published in 2009 by John Wiley & Sons, Ltd.

Qual. Reliab. Engng. Int. ***2010***, 26 431–441

This is a nonlinear model because it can not be linearized by taking logarithms. However, some experimenters try to do this; they consider just the expectation function

$$E(y) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2) \tag{4}$$

and take logarithms, then add the error term back to the transformed expectation function. If the model really is nonlinear, this linearization may not be completely satisfactory, and the resulting fitted model may perform very poorly, not fitting the experimental data well and providing inaccurate predictions of future viscosity values.

This has significant implications for experimental design. If the model really is intrinsically linear and taking logarithms 'works', then a factorial design is an optimal choice of design. On the other hand, if the model really is nonlinear, the factorial may be a poor choice of design. We illustrate this in the following section, and then make some recommendations about how to choose a design for the nonlinear case.

2. The viscosity experiment

Reconsider the viscosity experiment described in the previous section. The experimenter believes that a simple log transformation on the response will be adequate and plans to fit a first-order model plus interaction to the log viscosity. He decides to use a replicated 2^2 factorial design, with eight runs. This is an excellent choice of design, if the experimenter is right about the form of the model. For a first-order model with interaction, the 2^k factorial is *D*-optimal (the standard errors of the model regression coefficients are minimized), *G*-optimal (the maximum variance of the predicted response over the design space is minimized), and *I*-optimal (the average variance of the predicted response over the design space is minimized). In other words, if the model is correct, you can not put the eight runs in better places. The design and the observed response are shown in Table I.

The experimenter fits the main effects plus interaction to the natural logarithm of the viscosity response. The results are

$$\ln(\hat{y}) = 7.65 + 0.14x_1 + 0.42x_2 + 0.34x_1x_2$$

The overall model is significant at $\alpha = 0.001$ and the standard errors of the model coefficients are all equal, with $se(\hat{\beta}) = 0.00844$. Clearly both main effects and the interaction are significant and have important effects on the viscosity. In the original scale the fitted model is

$$\hat{y} = e^{7.6484 + 0.1412x_1 + 0.4239x_2 + 0.3369x_1x_2} \tag{5}$$

The plot of the viscosity response surface from the log model but with viscosity in the original scale is shown in Figure 1.

At this point, the experimenter considered fitting the nonlinear model in Equation (3). Using the data in Table 1, the nonlinear fit obtained from the JMP software package is

$$\hat{y} = e^{7.6485 + 0.1412x_1 + 0.4239x_2 + 0.3368x_1x_2}$$

which is very similar (identical to three decimal places) to the fit obtained from the linear model. The standard errors of the model parameters in the nonlinear model are all equal to $se(\hat{\beta}) = 0.01125$, which are very similar to the standard errors of the parameters in the linear model.

This is an example where the linearized model works nicely, and the 2^2 factorial design in Table I is an excellent choice. The experimenters suspected that this was the case because they had experienced previous success modeling viscosity on a log scale as a function of time and temperature. However, there are some advantages in fitting the nonlinear model as we now describe.

3. Advantages of a nonlinear model

Suppose that the experimenters were interested in predicting the mean viscosity for a variety of settings of the process variables time and temperature. If they used the linear model, then they have a model in the natural logarithm of viscosity, and to make

Table I. The 2^2 design for the viscosity experiment

Time (x_1)	Temperature (x_2)	Viscosity (y)
−1	−1	1654
1	−1	1155
−1	1	1947
1	1	5225
−1	−1	1685
1	−1	1103
−1	1	2027
1	1	5113

Published in 2009 by John Wiley & Sons, Ltd.
Qual. Reliab. Engng. Int. **2010**, 26 431–441

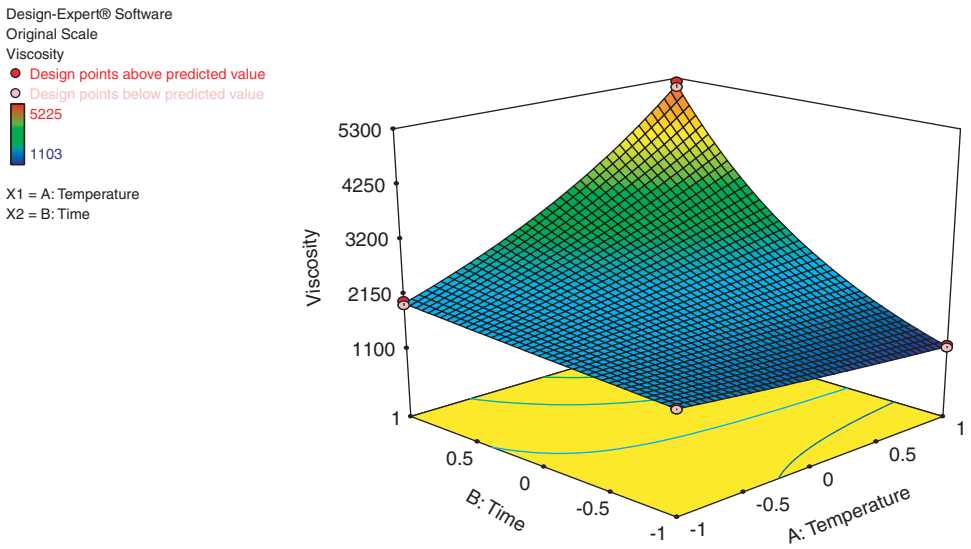

Figure 1. The contour plot of viscosity from the inverse model (from Design-Expert). This figure is available in colour online at www.interscience.wiley.com/journal/qre

Table II. Prediction interval ranges for the fitted linear and nonlinear models

	Prediction interval range	
Design point	Linear model	Nonlinear model
1	8.60	14.79
2	13.43	14.79
3	30.18	14.79
4	99.83	14.79
5	8.60	14.79
6	13.43	14.79
7	30.18	14.79
8	99.83	14.79
Average	38.01	14.79

predictions they have to apply the inverse transformation, which leads to Equation (5). However, it turns out that if the model in the logarithm of viscosity produces an estimate of the mean, the inverse transformed prediction from Equation (5) is biased. Specifically, it is an estimate of the median viscosity and not the mean. If viscosity is skewed positively, this means that the linear model will consistently underpredict the mean viscosity. See Montgomery *et al.*[1] for a discussion of this, including references to techniques for correcting the bias. Now in our specific example this may not be a serious problem, because the untransformed linear model and the nonlinear model are very similar.

Another advantage of the nonlinear model is that it may produce more accurate predictions under some conditions. This is most likely if the true system relating the response and the predictor variables is the nonlinear model in Equation (3) and not the intrinsically linear model of Equation (2). Table II provides the prediction interval ranges for the intrinsically linear model that is fit using the linear model versus the nonlinear model. The prediction interval ranges given are the asymptotic limits of prediction for the individual future-observed response. From the table, it is seen that six out of eight prediction interval ranges for the nonlinear fits as well as the average prediction interval range are smaller. Notice that the nonlinear model has equal prediction interval ranges, while the linear model has prediction interval ranges that vary. This is because a transformation is required to obtain the linear model prediction lower and upper limits, which does not preserve the equality of ranges across the design space, but does for each design point. Note that the first and the fifth design points for the linear model have the same prediction interval range, as does each replicated point.

4. Designing experiments for a nonlinear model

Constructing designs for linear models is relatively straightforward. There are many standard designs such as the factorial design in Table I and response surface designs such as the central composite design, and many of these designs are either optimal or

Published in 2009 by John Wiley & Sons, Ltd.

near-optimal designs with respect to the *D*, *G*, or *I* criteria. It is also relatively straightforward to construct optimal designs for either the *D* or *I* criteria. For example, a *D*-optimal design for a linear model is found by selecting the design points to maximize the determinant of $\mathbf{X}'\mathbf{X}$, where $\mathbf{X}$ is the model matrix constructed by expanding the design matrix to model form. In linear models, the model matrix and consequently the $\mathbf{X}'\mathbf{X}$ matrix contain only functions of the design points.

To find a *D*-optimal design for a nonlinear model we must find design points that maximize the determinant of $\mathbf{D}'\mathbf{D}$, where $\mathbf{D}$ is a matrix of partial derivatives of the nonlinear model expectation function with respect to each model parameter evaluated at each design point. This matrix will contain the unknown model parameters. For example, in the model of Equation (3) the $\mathbf{D}$ matrix is

$$\mathbf{D}=\begin{bmatrix} e^{\beta_0+\beta_1 x_{11}+\beta_2 x_{12}+\beta_3 x_{11}x_{12}} & x_{11}e^{\beta_0+\beta_1 x_{11}+\beta_2 x_{12}+\beta_3 x_{11}x_{12}} & x_{12}e^{\beta_0+\beta_1 x_{11}+\beta_2 x_{12}+\beta_3 x_{11}x_{12}} & x_{11}x_{12}e^{\beta_0+\beta_1 x_{11}+\beta_2 x_{12}+\beta_3 x_{11}x_{12}} \\ e^{\beta_0+\beta_1 x_{21}+\beta_2 x_{22}+\beta_3 x_{21}x_{22}} & x_{21}e^{\beta_0+\beta_1 x_{21}+\beta_2 x_{22}+\beta_3 x_{21}x_{22}} & x_{22}e^{\beta_0+\beta_1 x_{21}+\beta_2 x_{22}+\beta_3 x_{21}x_{22}} & x_{21}x_{22}e^{\beta_0+\beta_1 x_{21}+\beta_2 x_{22}+\beta_3 x_{21}x_{22}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{\beta_0+\beta_1 x_{n1}+\beta_2 x_{n2}+\beta_3 x_{n1}x_{n2}} & x_{n1}e^{\beta_0+\beta_1 x_{n1}+\beta_2 x_{n2}+\beta_3 x_{n1}x_{n2}} & x_{n2}e^{\beta_0+\beta_1 x_{n1}+\beta_2 x_{n2}+\beta_3 x_{n1}x_{n2}} & x_{n1}x_{n2}e^{\beta_0+\beta_1 x_{n1}+\beta_2 x_{n2}+\beta_3 x_{n1}x_{n2}} \end{bmatrix}$$

where x_{ij} is the ith observation for design factor j. The $\mathbf{D}'\mathbf{D}$ matrix is

$$\mathbf{D}'\mathbf{D}=\begin{bmatrix} \sum_{i=1}^{n} e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} & \sum_{i=1}^{n} x_{i1}e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} & \sum_{i=1}^{n} x_{i2}e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} & \sum_{i=1}^{n} x_{i1}x_{i2}e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} \\ & \sum_{i=1}^{n} x_{i1}^2 e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} & \sum_{i=1}^{n} x_{i1}x_{i2}e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} & \sum_{i=1}^{n} x_{i1}^2 x_{i2}e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} \\ & & \sum_{i=1}^{n} x_{i2}^2 e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} & \sum_{i=1}^{n} x_{i1}x_{i2}^2 e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} \\ & & & \sum_{i=1}^{n} x_{i1}^2 x_{i2}^2 e^{2(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}x_{i2})} \end{bmatrix}$$

Clearly it is impossible to choose a *D*-optimal design for this model without knowing the model parameters β_0, β_1, β_2, and β_3.

Box and Lucas[2] have applied the *D*-criterion to find designs for regression models. There are reviews of this subject in Ford *et al.*[3] and Atkinson *et al.*[4]. Chernoff[5] proposed choosing values for the unknown model parameters and finding the design that maximized the determinant of $\mathbf{D}'\mathbf{D}$ for the set of parameters. This leads to the idea of a local *D*-optimal design. This can work well if the estimates of the unknown parameters are close to the actual values. Another approach is to use a sequential design strategy; begin with a design that is smaller than the size of the final design, run this experiment and obtain preliminary estimates of the model parameters, then use these parameter estimates as if they were the true values of the parameters, and augment the original design with additional runs to produce the final design.

A Bayesian approach uses a prior distribution $f(\boldsymbol{\beta})$ to specify the uncertainty in the parameter values. This leads to a design criterion

$$\phi(\mathbf{D})=\int \log|\mathbf{D}'\mathbf{D}|f(\boldsymbol{\beta})\,d\boldsymbol{\beta} \tag{6}$$

This is the expectation of the logarithm of the information matrix. This criterion was proposed by Chaloner and Larntz[6] for single-factor logistic regression. The difficulty in using Equation (6) as a design criterion is that the multidimensional integral must be evaluated a large number of times. Gotwalt *et al.*[7] have recently developed a clever quadrature scheme that greatly improves the computing time to evaluate the integral in Equation (6) and which exhibits excellent accuracy. This procedure is implemented in the nonlinear design platform of JMP, and uses a coordinate exchange algorithm as the basis of design construction.

We will use the nonlinear design platform in JMP to construct several *D*-optimal designs for the model in Equation (3). Suppose that we want to find an 8-run *D*-optimal design for this model and the prior knowledge about each model parameter can be summarized by a normal distribution with limits:

$$\begin{aligned} 1 &\le \beta_0 \le 10 \\ 0 &\le \beta_1 \le 2 \\ 0 &\le \beta_2 \le 2 \\ 0 &\le \beta_{12} \le 2 \end{aligned} \tag{7}$$

These limits are $\pm 2\sigma$ limits in the normal distribution. The *D*-optimal design for this situation obtained from JMP is shown in Table III and Figure 2. We used 1000 random starts to construct this design. This design is similar to the 2^2 factorial; it has seven distinct design points including all four vertices of the square, but only one vertex is replicated [(1, 1)], and there are three other design points that are reasonably close to the (1, 1) vertex.

Published in 2009 by John Wiley & Sons, Ltd.
Qual. Reliab. Engng. Int. **2010**, 26 431–441

Table III. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (7)

x_1	x_2
1	1
1	0.63077069
1	−1
1	1
−1	1
−1	−1
0.62779336	1
0.54002902	0.53503102

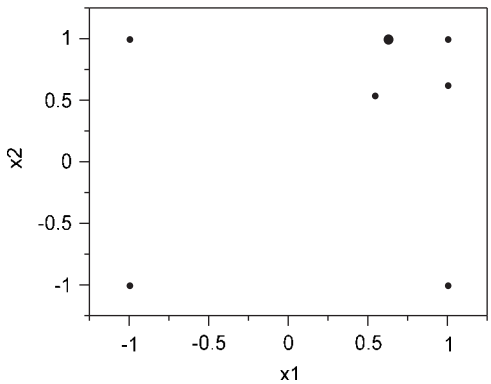

Figure 2. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (7)

Table IV. A 12-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (7)

x_1	x_2
−1	1
1	1
0.53328413	0.52914731
0.61439972	1
0.61439972	1
1	−1
−1	−1
1	1
1	1
−1	−1
1	0.61624804
1	0.61624804

If the experimenter had used 12 runs instead of 8, the factorial choice would have been simple; just use three replicates of the 2^2. The 12-run *D*-optimal design for the nonlinear model in Equation (3) and the prior distribution on the parameters in Equation (7) are shown in Table IV and Figure 3. It also has seven distinct design points like the 8-run design in Table III and Figure 2, and it also bears some resemblance to the 12-run 2^2 factorial; in that it has all four vertices of the square with two replicated vertices [$(-1,-1)$ and $(+1,+1)$], but it has three other distinct design points, two of which are replicated.

The *D*-optimal design for a nonlinear model also depends on the prior distribution, on the form of the prior distribution, and on the parameters. The *D*-optimal design for the model in Equation (3) and the ranges on the parameters in Equation (7) with a uniform prior distribution is shown in Table V and Figure 4. This design does not even closely resemble the 2^2 factorial. It has five distinct design points, only two of which are vertices of the square. One of the vertices is replicated twice, and there are two other non-vertex design points that are also replicated twice.

Published in 2009 by John Wiley & Sons, Ltd.
Qual. Reliab. Engng. Int. ***2010***, 26 431–441

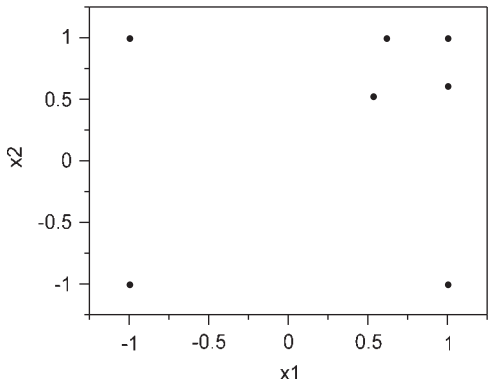

Figure 3. A 12-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (7)

Table V. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the parameter ranges in Equation (7) and a uniform prior

x_1	x_2
1	0.50707574
0.50700296	1
1	1
0.50701284	1
0.45589894	0.45176904
1	1
−1	−1
1	0.50707574

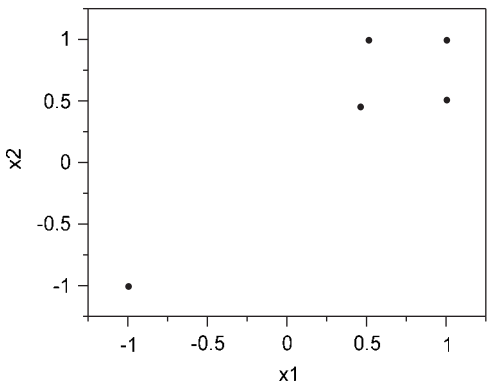

Figure 4. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the parameter ranges in Equation (7) and a uniform prior

Changing the ranges of the parameters in the prior distribution also has an impact on the *D*-optimal design. Suppose that the experimenter wants to use a normal prior with $\pm 2\sigma$ limits on the parameters as follows:

$$\begin{aligned} 1 \le \beta_0 \le 15 \\ -1 \le \beta_1 \le 2 \\ -1 \le \beta_2 \le 2 \\ -1 \le \beta_{12} \le 2 \end{aligned} \tag{8}$$

These limits are wider than those specified in Equation (7). The 8-run *D*-optimal design from JMP that results from 1000 random starts of the nonlinear design algorithm is shown in Table VI and Figure 5. This design has eight distinct runs (no replication) and is somewhat different from the 8-run *D*-optimal design in Table-replicated 2^2 factorial. Four of the points are vertices of the square and there are four other design points that are reasonably close to three of the four vertices.

In addition to changing the ranges of the parameters in the prior distribution, shifting the range away from including zero has an impact on the *D*-optimal design. Suppose that the experimenter wants to use a normal prior with $\pm 2\sigma$ limits on the

Published in 2009 by John Wiley & Sons, Ltd.

Qual. Reliab. Engng. Int. ***2010***, 26 431–441

Table VI. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (8)

x_1	x_2
1	1
−0.7323753	0.57396031
−1	1
1	0.61735857
−1	−1
1	−1
0.51893018	−0.9927921
0.66327665	1

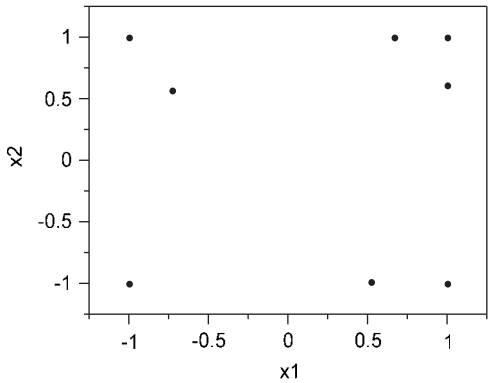

Figure 5. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (8)

Table VII. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (9)

x_1	x_2
1	1
−1	−1
1	1
1	0.75241
0.72952	0.7308
1	0.75241
0.75236	1
0.75236	1

parameters as follows:

$$\begin{aligned} 1 \le \beta_0 \le 10 \\ 1 \le \beta_1 \le 3 \\ 1 \le \beta_2 \le 3 \\ 1 \le \beta_{12} \le 3 \end{aligned} \tag{9}$$

These limits have the same range as those specified in Equation (7). The 8-run *D*-optimal design from JMP that results from 1000 random starts of the nonlinear design algorithm is shown in Table VII and Figure 6. This design does not even closely resemble the 2^2 factorial, but is very similar to the design shown in Figure 4. It has five distinct design points, only two of which are vertices of the square. One of the vertices is replicated twice, and there are two other non-vertex design points that are also replicated twice. The difference between this design and the one displayed in Figure 4 is that the cluster of points in the upper right-hand corner of Figure 6 is a bit tighter and higher than the cluster of points in Figure 4.

Now suppose that the experimenter suspects that a first-order model with interaction is unlikely to be an adequate way to represent the response. He decides to use a second-order model in the log viscosity scale. The corresponding nonlinear model is

$$y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2) + \varepsilon \tag{10}$$

Published in 2009 by John Wiley & Sons, Ltd.

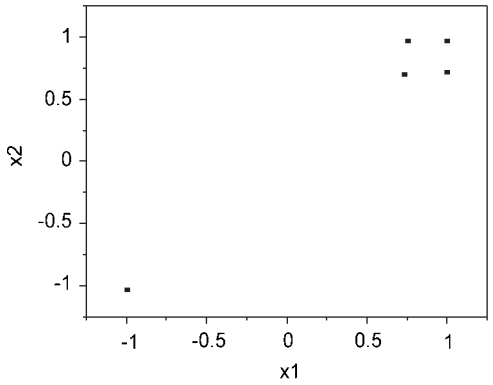

Figure 6. An 8-run *D*-optimal design for the nonlinear model in Equation (3) using the normal prior distribution in Equation (8)

Table VIII. A 12-run *D*-optimal design for the nonlinear model in Equation (10) using the normal prior distribution in Equation (11)

x_1	x_2
1	1
−1	−1
−1	−0.009748
−1	−1
1	0.01824062
1	0.5589844
−1	1
1	−1
−1	1
0.48635058	1
0.59216208	1
1	1

If the experimenter plans to fit a linear model to the log of viscosity, a reasonable design for the linear model would be a variant of a central composite design such as the face-centered cube. This design would have nine runs, not counting replicated runs at the center. A reasonable choice of design would be the face-centered cube with four center runs, yielding a 12-run design. This design could be used to fit either the linearized model or the nonlinear model in Equation (9). We will construct a 12-run *D*-optimal design for the nonlinear model in Equation (9) using a normal prior on the parameters with $\pm 2\sigma$ ranges as follows:

$$\begin{aligned} 1 \le \beta_0 \le 10 \\ 0 \le \beta_1 \le 1 \\ 0 \le \beta_2 \le 1 \\ 0 \le \beta_{12} \le 1 \\ 0 \le \beta_{11} \le 1 \\ 0 \le \beta_{22} \le 1 \end{aligned} \tag{11}$$

The 12-run *D*-optimal design from JMP found using 1000 random starts is shown in Table VIII and Figure 7. This design, like the face-centered cube, has nine distinct design points, including all four vertices. It has two runs that are very similar to the edge centers in the face-centered cube at $x_1 = \pm 1$. The remaining three runs are quite different from the design points in the face-centered cube. Three of the vertices are replicated. There are no center runs.

As a final example, consider the three-parameter model

$$y = \frac{\beta_1 x_1}{\beta_2 \left(1 + \frac{x_2}{\beta_3} + x_1\right)} + \varepsilon \tag{12}$$

This model is used in modeling enzyme kinetics. It is not intrinsically linear; it cannot be linearized by simple transformations, either on the response or the design variables. There is no obvious choice of a standard design to fit this model. However, since the model has three parameters, a 2^2 factorial would be a logical choice. With two replicates, this results in an 8-run design.

Published in 2009 by John Wiley & Sons, Ltd.
Qual. Reliab. Engng. Int. ***2010***, 26 431–441

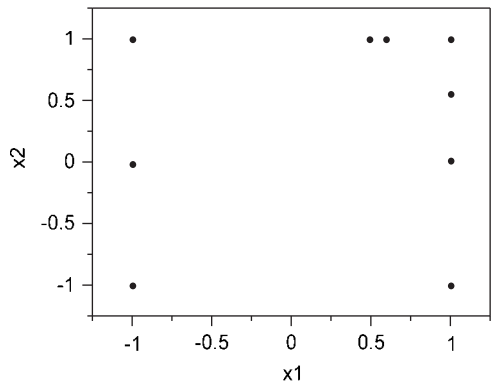

Figure 7. A 12-run *D*-optimal design for the nonlinear model in Equation (10) using the normal prior distribution in Equation (11)

Table IX. An 8-run *D*-optimal design for the nonlinear model in Equation (12) using the normal prior distribution in Equation (14)

x_1	x_2
−0.2256351	0.06409333
−0.050464	0.14435988
−0.3534687	−0.1549625
−0.6003572	0.40773179
−0.4050571	0.04333717
0.25323739	−0.62942
−0.5710757	0.92768144
−0.3538035	−0.004651

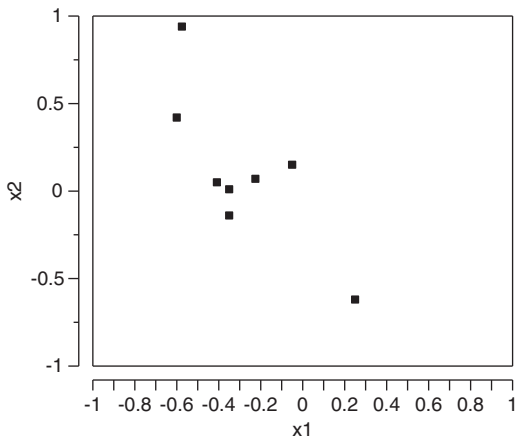

Figure 8. An 8-run *D*-optimal design for the nonlinear model in Equation (12) using the normal prior distribution in Equation (14)

We constructed an 8-run *D*-optimal design for this model from JMP, using a normal prior with the following $\pm 2\sigma$ ranges on the parameters:

$$\begin{aligned} 1 \le \beta_1 \le 3 \\ 0.1 \le \beta_2 \le 0.3 \\ 0.25 \le \beta_3 \le 0.75 \end{aligned} \tag{13}$$

The design found from JMP after 1000 random starts of the coordinate exchange procedure is shown in Table IX and Figure 8. This design is very different from the 2^2 factorial. It has eight distinct design points, and none of the points are vertices. It does not share any point in common with the factorial.

Published in 2009 by John Wiley & Sons, Ltd.
Qual. Reliab. Engng. Int. ***2010***, 26 431–441

5. Design efficiency

In the previous section we have observed that the *D*-optimal design for a nonlinear model, even a model that is intrinsically linear, can differ somewhat from a standard design, such as a factorial. It is of interest to compare the nonlinear *D*-optimal design with the standard design. We will do this using a measure of design efficiency based on the design criterion in Equation (6). Specifically, for a particular scenario (nonlinear model, prior distribution, and design) we will use Monte Carlo methods to randomly sample 1000 times from the prior distribution for that scenario and evaluate $\log|\mathbf{D}'\mathbf{D}|$ for the standard design and the *D*-optimal design. The sum of these $\log|\mathbf{D}'\mathbf{D}|$ values for each design is an approximation of the integral in Equation (6). We take as the efficiency of the factorial design relative to the *D*-optimal design the ratio

$$E=\left(\frac{\sum_{i=1}^{1000}\log|\mathbf{D}_i'\mathbf{D}_i|_{\text{Factorial}}}{\sum_{i=1}^{1000}\log|\mathbf{D}_i'\mathbf{D}_i|_{D\text{-optimal}}}\right)^{1/p} \tag{14}$$

where p is the number of parameters in the model. Values of this ratio that are less than unity indicate that the factorial is less efficient than the *D*-optimal design for that particular scenario. We then repeat this 1000 times so that we can evaluate the distribution of the efficiency values.

Figure 9 summarizes these efficiency calculations for the designs in Tables III–IX. The first four sections of this figure compare the *D*-optimal designs for the model in Equation (3) with the 2^2 factorial design. Note that the average efficiencies range from 0.9125 to 0.9936. The only design for which the average efficiency is below 0.9704 is the case in Table VII, where the ranges on the parameters in the prior distribution did not include zero. This illustrates the sensitivity of the design to the prior. However, in this set of examples, we conclude that the 2^2 factorial design is highly *D*-efficient and almost *D*-optimal for the assumed model. So for this particular model, the experimenter does not lose much from using the 2^2 factorial design. Examining the designs in Tables III–VII reveals that the optimal designs share several points in common with the factorial and considering that the model we are fitting is intrinsically linear, these results are not terribly surprising. For the model in Equation (10), which is intrinsically linear but has a full quadratic in the exponent, the face-centered cube has a mean efficiency of 0.9940 compared with the *D*-optimal design in Table VIII. Again, this is not too surprising, as the optimal design shares several points in common with the face-centered cube, and the face-centered cube is known to be reasonably *D*-efficient for the second-order model.

Now consider the designs for the nonlinear model in Equation (12). From Figure IX we see that the mean relative efficiency of the 2^2 factorial design relative to the computer-generated *D*-optimal design in Table IX is 0.7882; hence, the *D*-optimal design enjoys a significant advantage in this case. Note that the optimal design does not share any points in common with the factorial and the model that we are using is a true nonlinear model.

Finally, note that the histograms and summary statistics in Figure IX indicate a very narrow range for the efficiencies for each design comparison. The widest range is for the nonlinear design in Table IX, from 0.7588 to 0.9166. Only 10% of the observed efficiencies were above 0.8035. We conclude that 1000 repetitions of the efficiency calculations were sufficient to provide useful information about the relative efficiency of the designs in this study.

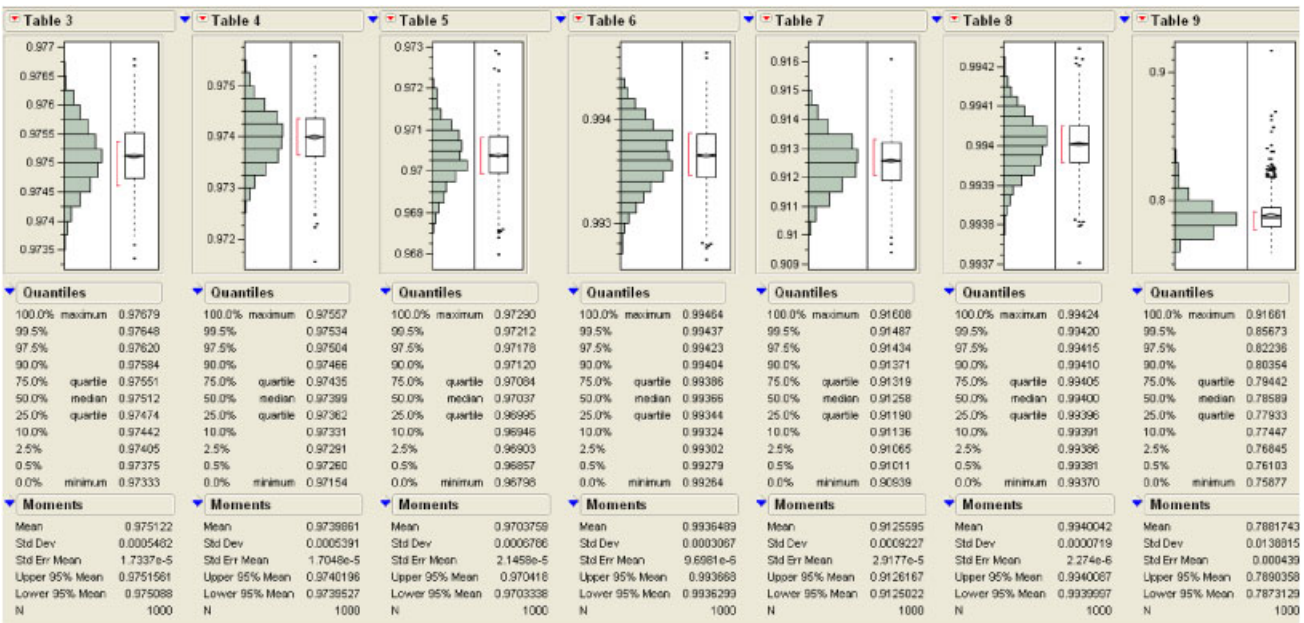

Figure 9. Summary statistics for the efficiency calculations. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Published in 2009 by John Wiley & Sons, Ltd.

6. Concluding remarks

We have shown that it is relatively straightforward, using a modern computer software implementation of the Bayesian criterion, to generate *D*-optimal designs for nonlinear models. For a specific example of an intrinsically linear model, we have shown that a 2^2 factorial design has a very high relative efficiency compared with the *D*-optimal design. We also showed that a face-centered cube compared favorably with a *D*-optimal design for an intrinsically linear model with a second-order polynomial in the exponent. While our study is limited, we suspect that in many situations standard designs will compare favorably with optimal designs for nonlinear models that are intrinsically linear, particularly in situations where the *D*-optimal design shares several points in common with the standard design. We have developed a relatively straightforward procedure to evaluate the efficiency of these comparisons.

If the model of interest is nonlinear and its expectation function cannot be linearized by simple transformations of the response and/or predictor variables, then the nonlinear *D*-optimal is considerably more efficient than a standard design. While our study is limited to two nonlinear models, we believe that the results will generalize for many nonlinear models and experimental situations.

References

1. Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis* (4th edn). Wiley: NJ, 2006.
2. Box GEP, Lucas H. Design of experiments in non-linear situations. *Biometrika* 1959; **46**:66–90.
3. Ford I, Kitsos D, Titterington DM. Recent advances in non-linear experimental design. *Technometrics* 1989; **31**:49–60.
4. Atkinson AC, Donev A, Tobias R. *Optimum Experimental Designs, with SAS*. Oxford University Press: Oxford, 2007.
5. Chernoff H. Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics* 1953; **24**:586–602.
6. Chaloner K, Larntz K. Optimal bayesian designs applied to logistic regression experiments. *Journal of Statistical Planning and Inference* 1989; **21**:191–208.
7. Gotwalt CM, Jones BA, Steinberg DM. Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics* 2009; **51**:86–95.

Authors' biographies

Rachel T. Johnson is an assistant professor in the Operations Research Department at the Naval Postgraduate School. She received her PhD and MS in Industrial Engineering from the Arizona State University and her BS in industrial engineering from the Northwestern University. Her research interests are in the design and analysis of both physical and computer experiments and simulation methodology. She is a recipient of the Mary G. and Joseph Natrella Scholarship for excellence in statistics. She is a member of the American Society for Quality and the Institute for Operations Research and Management Sciences.

Douglas C. Montgomery is Regents' professor and the ASU Foundation Professor of Engineering at the Arizona State University. He is co-director of the ASU Committee on Statistics. His research interests are in industrial statistics. He has advised 50 doctoral students and over 40 MS thesis students and statistics applied projects. He is an author of 16 books and over 190 technical papers. His research has been funded by NSF, NASA, the Office of Naval Research, the U.S. Army, and private industry. He is an ASU Outstanding Doctoral Mentor, and has received teaching awards from both his department and the Fulton School of Engineering. He is a recipient of the Shewhart Medal, the Brumbaugh Award, the Lloyd S. Nelson Award, the Hunter Award, and the Shewell Award (twice), all from the American Society for Quality, the Deming Lecture Award from the American Statistical Association, and the Ellis R. Ott Award. He is one of the Chief Editors of *Quality & Reliability Engineering International*, a former Editor of the *Journal of Quality Technology*, and a member of several other editorial boards. Professor Montgomery is a fellow of the American Statistical Association, a fellow of the American Society for Quality, a fellow of the Royal Statistical Society, a fellow of the Institute of Industrial Engineers, an elected member of the International Statistical Institute, and an academician of the International Academy for Quality.

Published in 2009 by John Wiley & Sons, Ltd.
Qual. Reliab. Engng. Int. ***2010***, 26 431–441