Robust Design Strategies
for Nonlinear Regression Models

T. E. O'Brien
University of Georgia

Abstract

In the context of nonlinear regression models, this paper outlines recent developments in design strategies when the assumed model function, initial parameter guesses, and/or error structure are not known with complete certainty. Designs obtained using these strategies are termed robust designs as they are intended to be robust to specified departures. Robust designs are clearly advantageous in many practical settings since these designs can be used to test for, say, lack of fit of the assumed model function or error heteroskedasticity, whereas so-called optimal designs often cannot.

1 Introduction

The validity and practicality of research in drug studies, agricultural studies and engineering studies depends upon the reliability and efficiency of the experimental designs used by researchers in these fields. Optimal design theory provides researchers with the means to select the "optimal" design for their experiment, in the sense of yielding a design that will require the fewest repetitions to yield accurate results. Optimal designs depend upon the chosen model function, on the error structure, and, in the case of nonlinear models, on the initial parameter choice.

Our focus here is to outline recent developments in design strategies for situations where the assumed (nonlinear) model function, initial parameter guesses, and/or error structure may not be known with complete certainty. Since a methodology can by termed "robust" if it is efficient even if some of the underlying assumptions are not met, designs obtained using the strategies discussed here are termed robust designs. Since an integral part of data analysis is to perform diagnostics such as tests for lack of fit and for variance homogeneity, robust designs are clearly advantageous in many practical settings as these designs can be used to perform these tests whereas optimal designs often cannot.
Matters of Notation

Classic univariate nonlinear models are typically expressed as

\[ y_k = \eta(x_k, \theta) + \varepsilon_k, \quad k = 1, \ldots, n, \]  

though in this paper we extend our focus to include compartmental models, in which the model is defined by a linear or nonlinear set of ordinary differential equations (see, e.g., Chap. 5 of [4] or Chap. 8 of [26]). The design problem in this situation is to obtain an n-point design, \( \xi \), and to estimate some function of the p-dimensional parameter vector, \( \theta \), with high efficiency. Our focus here is on approximate designs, or designs of the form

\[ \xi = \{ x_1, x_2, \ldots, x_n \}, \]

where the design points (or vectors) \( x_1, x_2, \ldots, x_n \) are elements of the design space \( X \) (and are not necessarily distinct) and the associated weights \( \omega_1, \omega_2, \ldots, \omega_n \) are non-negative real numbers which sum to unity. Algorithms to convert optimal approximate designs to near-optimal exact ones, or designs where each \( \omega_k \) is of the form \( \frac{n_k}{n} \), for \( n_k \) an integer, are given in [21] and [24].

When the errors associated with the assumed model (1) are uncorrelated normal random variables with zero mean and constant variance (taken without loss of generality to equal one), the (Fisher) information per observation is given by

\[ \mathbf{M}(\xi, \theta) = \sum_{i=1}^{n} \omega_i \left( \frac{\partial \eta_i(x_i)}{\partial \theta} \right) \left( \frac{\partial \eta_i(x_i)}{\partial \theta} \right)^T = \mathbf{V} \mathbf{\Omega} \mathbf{V}. \]  

Here \( \mathbf{V} \) is the \( n \times p \) Jacobian of \( \eta \) with \( i^{th} \) row equal to \( \frac{\partial \eta_i(x_i)}{\partial \theta} \), and \( \mathbf{\Omega} \) is the diagonal matrix with diagonal elements \( \omega_1, \omega_2, \ldots, \omega_n \). Also, the variance function ([3], p. 95) of \( \eta \) is given by

\[ d(x, \xi, \theta) = \left( \frac{\partial \eta(x)}{\partial \theta} \right)^T \mathbf{M}^{-1}(\xi, \theta) \frac{\partial \eta(x)}{\partial \theta}, \]

where \( \frac{\partial \eta(x)}{\partial \theta} \) is of dimension \( p \times 1 \), and a generalized inverse is used whenever \( \mathbf{M} \) is singular.

Optimal designs typically minimize some convex function of \( \mathbf{M}^{-1} \) (see [24]; c.f. [29]). For example, designs which minimize the determinant \( |\mathbf{M}^{-1}(\xi, \theta^*)| \) are called locally D-optimal, those which minimize the maximum (over all \( x \in X \)) of \( d(x, \xi, \theta^*) \) are called locally G-optimal, and if \( \lambda_1, \ldots, \lambda_p \) are the p eigenvalues
Robust designs strategies for nonlinear regression models

of $M^{-1}(\xi, \theta^*), \text{then designs which minimize } \left\{ \frac{1}{k}(\lambda_1^k + \ldots + \lambda_p^k) \right\}^{1/k} \text{for } k \in (0, \infty) \text{ are called locally } \Phi_k-\text{optimal. The term "locally" is used here to emphasize that the corresponding design is based on an initial parameter choice, } \theta^*. \text{ Further, the General Equivalence Theorem of [13] and [28] establishes that D-optimal designs are equivalently G-optimal.}

D-optimal designs also minimize the first-order approximation to the volume of the confidence ellipsoid for the parameter estimates. In contrast, Hamilton and Watts [11] use a quadratic approximation to show that the volume of a $100(1-\alpha)%$ confidence region for the true parameter vector is approximately

$$\nu = c |V^TV|^{-1/2} |D|^{-1/2} \left( 1 + k^2 \text{ tr } \left[ D^{-1}C \right] \right), \quad (4)$$

where $c$ and $k$ are constants relative to the design, $C$ is a function of the parameter-effects curvature, and $D$ measures the intrinsic curvature in the direction of the residual vector (see [4] for a discussion of curvature). Here $D = I_p - B$, and $B = L^T [c^T|W] L$. Claiming that (4) could not be used as a design criterion as it requires knowledge of the unknown residuals, the authors replaced the residual vector in (4) with a vector of zeros (and so $D = I_p$), and obtained a criterion that seeks designs which minimize the volume

$$\nu' = c |V^TV0|^{-1/2} \left( 1 + k^2 \text{ tr } [C] \right). \quad (5)$$

Such designs ignore the intrinsic nonlinearity of the expectation surface. We call designs which minimize the volume in (4) Q-optimal and those which minimize the volume in (5) Q'-optimal.

3 Design and Detecting Lack of Fit of the Assumed Model Function

Although so-called optimal designs may yield efficient parameter estimates when the assumed model function is known with complete certainty, researchers who find themselves in less-certain circumstances often desire designs with design points which can be used to test for lack of fit of the hypothesized model function. Further, since empirical evidence ([11], [27]; c.f., [10]) indicates that optimal designs for models with $p$ model parameters often have only $p$ support points (and hence cannot be used to test for model mis-specification), we often require near-optimal designs with "extra" support points. In this section, we discuss four such robust design strategies.
3.1 Q-optimality as a Robust Design Strategy

O'Brien [15] replaces the residual vector in (4) with a reasonable (non-zero) approximation and introduces (p+1)-point Q-optimality, a design strategy which yields (p+1)-point exact designs that take account of all of the curvature of the expectation surface. This design procedure is extended in [20] to yield (p+s)-point approximate designs by using the relation $\epsilon = N\alpha$, where the $n \times s$ matrix $N$ (whose columns form an orthonormal basis for the space orthogonal to the tangent plane) is obtained via the QR-decomposition of $V$, and where

$$
\alpha = \sqrt{s\sigma} \begin{pmatrix}
\sin\varphi_1 & \sin\varphi_2 & \ldots & \sin\varphi_{s-1} \\
\cos\varphi_1 & \sin\varphi_2 & \ldots & \sin\varphi_{s-1} \\
\ldots & & & \\
\cos\varphi_{s-1}
\end{pmatrix}
$$

A (p+s)-point design is then said to be Q-optimal if it minimizes the expected volume

$$
\int_0^{\pi/2} \cdots \int_0^{\pi/2} \nu(\varphi_1, \ldots, \varphi_{s-1}) \, d\varphi_1 \cdots d\varphi_{s-1},
$$

for $\nu = \nu(\varphi_1, \ldots, \varphi_{s-1})$ given in (4).

This procedure is applied to the two-parameter intermediate product model function (6) in [15] and [20] to obtain a four-point exact design and a three-point approximate design. Interestingly, all exact and approximate designs with more support points “collapsed” to these two designs, indicating that this procedure may be used to obtain only a few additional “check” points.

3.2 Estimation and Discrimination as a Robust Design Strategy

For situations in which at least two alternative model functions can be given to describe a given process, O'Brien and Rawlings [22] introduced a design strategy which provides for efficient estimation of all model parameters and for the ability to discriminate between the rival model functions. This strategy can also be used to obtain robust designs when the researcher has only one model function in mind by first obtaining a second rival model function (see below) and then using the estimation-discrimination design procedure developed in [22]. Such designs will typically have $2p$ support points (where $p$ is the number of parameters in
the original model function), and hence may be used to test for adequacy of the original model function.

To illustrate, suppose that a chemist feels that the two-parameter intermediate product model function (IP2)

$$\eta_1 = \frac{\theta_1}{\theta_1 - \theta_2} \left( e^{-\theta_2 x} - e^{-\theta_1 x} \right)$$

(6)

adequately describes the movement of a chemical through a medium with initial parameter choices \((\theta_1^*, \theta_2^*) = (0.70, 0.20)\). In this case, the locally D-optimal design associates the weight \(w = 1/2\) with each of the support points \(x = 1.229\) and 6.858. A rival to this IP2 function is the two-parameter inverse quadratic model function (IQ2)

$$\eta_2 = \frac{\theta_3 x}{(1 + \theta_3 x)(1 + \theta_4 x)}$$

(obtained from \(\eta_1\) by using the approximation \(e^{-z} \approx \frac{1}{1 + z}\)) with initial parameter guesses \((\theta_3^*, \theta_4^*) = (1.80, 0.20)\). Our estimation-discrimination design procedure produces the design which places the weight \(w = 1/4\) at each of the points \(x = 0.241, 1.104, 3.036,\) and 7.087, a design which indeed can be used to test for the adequacy of the original (IP2) model function.

This estimation-discrimination design strategy can also be amended to incorporate the degree of belief that the researcher has in the respective model functions. For example, if the chemist in the above example has a great deal of faith in the IP2 model function (but still desires a robust design), then the corresponding estimation-discrimination design could be obtained as above but subject to the constraint that the estimation efficiency (see [3], p. 116) of the IP2 model parameters be at least, say, 80%.

3.3 Obtaining Robust Designs by Model Nesting

A third strategy to find designs with “extra” design points is to embed or “nest” the assumed (“original”) model function in a larger (“super”) model function which reduces to the original model function when some parameters of the super model function are equal to specified extended real numbers.

Specifically, let \(\eta_1(\theta_1)\) denote the original model function (with \(p_1\) model parameters), let \(\eta_2(\theta_1, \theta_2)\) denote the super model function (with \(p_2\) additional model parameters), and let \(\eta_1\) generalize \(\eta_1\) in the sense that \(\eta_1\) and \(\eta_2\) are identical for all \(x \in X\) when \(\theta_2 = c\) (for \(c\) some extended real \(p_2 \times 1\) vector). In this
setting, a measure of the information that the design $\xi$ contains regarding $\theta_1$ in the original function $\eta_1$ is $|M_{11}|$, and a measure of the information $\xi$ contains regarding departures from $\eta_1$ in the direction of $\eta_2$ is $|M_{22} - M_{21}M_{11}^{-1}M_{12}|$, where $M_{jk} = V^T_j \Omega V_k$ for $j,k = 1,2$. O'Brien [17] combines these measures into the single information measure

$$\psi(\xi, \theta, \lambda) = \frac{\lambda}{p_1} \log |M_{11}| + \frac{1-\lambda}{p_2} \log |M_{22} - M_{21}M_{11}^{-1}M_{12}|,$$

and seeks designs to maximize $\psi$ for given choices of $\theta^T = (\theta_1^T, \theta_2^T)$ and $\lambda$, designs which we call (locally) D$_\lambda$-optimal. Here $\lambda$ controls the amount of information obtained regarding estimation of $\theta_1$ relative to detecting departures from $\eta_1$, and is usually chosen so that the final design has an efficiency of a prescribed level (as described in the previous section).

O'Brien [17] points out that most (if not all) growth models in current use are members of either the Weibull, the Log-Logistic, or the Richards families (see Chap. 4 of [25] and Chap. 7 of [26]), and provides a (six-parameter) super model which (simultaneously) generalizes each of these families. Thus, whenever a researcher seeks a robust design for a given growth function, this design can be obtained by using the above nesting design strategy and this six-parameter super model; such a design would then be useful in detecting departures from the assumed model function in the direction of all other growth curves. Finally, note that this design strategy is easily generalized to nonlinear models other than growth functions.

### 3.4 A Robust Design Strategy Useful to Detect General Departures

Departures from a given model function can also be detected by using the design algorithm introduced in O'Brien [19], which provides designs which are near-optimal (typically with a D-efficiency around 90%), yet which have extra design points. This design algorithm adds to the locally D-optimal design, $\xi_D$, those values of $x$ for which the variance function $d(x, \xi_D, \theta^o)$ given in (3) intersects the line

$$y = p \left\{ \frac{(p+1)\delta}{p} \right\}^p - 1,$$

(7)

where $\delta$ is the desired D-efficiency and $p$ is the number of model parameters.

To illustrate, consider the two-parameter simple exponential model $\eta = \theta_1 e^{-\theta_2 x}$ over the design space $X = [0.34, \infty)$. Note that for this model function and
the initial parameter choice $\theta_2 = 0.6944$, the (locally) D-optimal design takes half its observations at $x = 0.340$ and the other half at $x = 1.761$. Further, the design which uses our general departure procedure and $\delta = 0.90$ (and for which the right-hand side in (7) is 1.645) takes three observations at each of $x = 0.340$ and 1.761, and one observation each at $x = 0.398$, 1.133, and 2.529. This latter design, which has a D-efficiency of 93.6%, can therefore be used to test for model mis-specification yet is "nearly optimal" in the sense that it represents only a 6.4% loss in efficiency relative to the (locally) D-optimal design.

3.5 Choosing the Appropriate Robust Design Strategy

The inadequacy of so-called optimal designs with only $p$ support points has been emphasized by several researchers (e.g., [5] and [9]), and so the four robust design strategies discussed above provide reasonable and important alternatives. Choice of the specific robust procedure to use in a given situation is a function of the degree of confidence that the researcher has in the assumed model function. For example, if we are reasonably confident in our model function but desire a design with extra support points, it is prudent for us to obtain a $(p+1)$-point Q-optimal design in which only one "extra" support point is obtained. If our confidence in the assumed model function were somewhat lower, and we were entertaining two or three rival model functions, then the estimation-discrimination outlined above could be used to obtain an efficient design which highlights which function best fits the data. Should we have a given model function in mind which is a member of a class of functions (such as the class of growth functions), and desire a design which is robust against the other elements of this class, the nesting strategy should be used. Finally, when we wish to guard against all departures from the assumed model functions (instead of specifying the direction of these departures), we are well-advised to use the general-departures design procedure. We therefore work down this list of robust design strategies as our prior belief in the assumed model function decreases.

4 Design and Robust Parameter Choices

By definition and in contrast with linear models, the Jacobian matrix for a non-linear model depends upon at least one of the model parameters. It follows that all admissible designs, or designs which depend upon the corresponding information matrix (2), also depend upon some model parameters; we therefore find ourselves in the troublesome situation of needing to know the values of the model parame-
ters before we can find a good design to estimate these same model parameters efficiently. Techniques to circumvent this dilemma include obtaining locally optimal designs, minimax designs, sequential designs, and Bayesian designs.

One way to obtain efficient designs for nonlinear models is to use an initial "guess" (or one-point prior), $\theta^o$, for the parameter vector in (2), and this strategy is equivalent to replacing the assumed nonlinear model function (1) by its first-order Taylor approximation at $\theta = \theta^o$. This technique, illustrated above with each of the robust design strategies discussed in the previous section, was originally introduced in [8] for classic design criteria. An obvious objection to its use, however, is that a design which may be optimal for one value of $\theta$ may be very inefficient if the true parameter value, $\theta^*$, is far from $\theta^o$.

A second way to deal with the dependence of optimal designs on $\theta^o$ is to seek designs which maximize the information about the least informative situation, that is, to obtain a minimax design (see, e.g., [7]). To illustrate using D-optimality, a minimax D-optimal design minimizes

$$
\max_{\theta \in \Theta^o} |M^{-1}(\xi, \theta)|,
$$

where $\Theta^o$ is a subset of the parameter space. Unfortunately, this technique can also be problematic in that it depends very strongly on the chosen subset $\Theta^o$ (and so the problem of using an initial guess $\theta^o$ is replaced by the problem of using the subset $\Theta^o$), and designs which might be efficient for the least informative situation may also be very inefficient in most other situations.

A third and very sensible way to reduce dependence on initial parameter guesses is to design sequentially, in which a series of designs are obtained and each design helps give more and more information about the requisite subsequent initial parameter choice. Sequential designs, which are described in [3] and [24], are useful in settings where the time required to run and process experiments is trivial (such as in the field of chemical engineering); agronomists – who often must wait until the following growing season to repeat an experiment – often find this technique impractical. Note that each of the robust design strategies discussed in the previous section has been extended to also yield robust sequential designs.

So-called Bayesian optimal designs, as discussed in [7], have also been introduced in nonlinear settings to reduce dependence on the initial parameter choice. For example, Bayesian D-optimal designs maximize the expected log-information

$$
\int \log |M(\xi, \theta)| p(\theta) \, d\theta,
$$

where $p(\theta)$ is a hypothesized prior distribution of the model parameters and is chosen to reflect the researcher's belief about the model parameters. As may be
expected, [6] and [22] have found that the number of support points of Bayesian designs tend to increase as does the dispersion in the assumed prior \( p(\theta) \). Again, each of the above robust design strategies has been extended to yield robust Bayesian designs.

5 Design and Alternate Error Structures

As mentioned above, an important diagnostic test which researchers often wish to perform after an experiment has been run and the data analyzed is to test for constancy of the residual variance over the design space. As such, replicates should be taken at several of the support points. Further, the classic (homoskedastic Gaussian) design problem has also been extended to allow for heteroskedastic errors and for other-than-Gaussian error distributions.

Huet et al. [12] points out that the term “model” should be used to include not only the assumed model function, or the expected value of the observations, but also the assumed variance function, or the variance of the observations. In a general sense, then, “the model” includes both the \( p \) model function parameters, \( \theta \), and the \( q \) variance function parameters. For example, in many agricultural studies it is reasonable to suppose that the error variance in (1) depends upon the expected value \( \eta \) in that we can write \( E(\varepsilon^2) = \sigma^2 \eta^p(x, \theta) \); this setting obviously includes the classic setting when we take \( p = 0 \). Atkinson and Cook [2] develop D-optimal designs for heteroskedastic linear models, and this problem was considered for nonlinear models in O'Brien [16]. In general, these designs typically have \( p+q \) support points.

Most optimal design criteria discussed, say, in [3] or [24] assume that the errors follow the normal distribution, although binomial errors have been hypothesized in very specific settings (see [6] and [14]). An interesting extension to the nesting robust design strategy discussed in the previous section is to imbed the assumed normal error distribution in the exponential family error distribution; since this “super-model” includes many other error distributions, robust designs obtained using this nesting strategy can be used in many practical settings.

6 Discussion

Practitioners have had a long history of dissatisfaction with theoretical optimal designs. Two reasons for this displeasure is that software for optimal designs has
been scarce and classic optimal designs often have only $p$ support points. Yet, now that Gauss and SAS programs to obtain optimal designs are given in [16] and [18], better algorithms to obtain optimal designs are given in [1] and [30], and robust design strategies have been developed, it is hoped that near-optimal designs will be more widely utilized in practical settings.

References


D-optimal designs for heteroscedastic linear models. *Journal of the American Statistical Association*, 90, 204-212.


Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21, 191-208.


Robust designs strategies for nonlinear regression models


A further note on quadratic designs for nonlinear regression models. Submitted.

A practical design methodology for nonlinear regression models. Submitted.
A non-sequential design procedure for parameter estimation and model discrimination in nonlinear regression models. To appear in *Journal of Statistical Planning and Inference*.

A new subset design strategy for nonlinear regression models. Submitted.


