ROBUST OPTIMAL DESIGNS WITH REDUCED CURVATURE

Timothy E. O'Brien
Dept. of Mathematics & Statistics, Loyola University Chicago,
6525 N. Sheridan Rd., Chicago, IL 60626.   Email: tobrie1@luc.edu

Abstract

Researchers often find that nonlinear regression models are more relevant for their studies than are linear ones, and are thus often in a position of requiring efficient designs for a given nonlinear model. A common shortcoming of most optimal designs for nonlinear models used in practical settings, however, is that these designs typically focus only on first-order parameter variance or predicted variance, and thus ignore the inherent nonlinear of the assumed model function. Another shortcoming is that optimal designs often have only p (the number of model parameters) support points, thereby providing no ability to test for model misspecification. Measures of marginal curvature, first introduced in Clarke (1987) and further explored in Haines *et al* (2004), provide a useful means of assessing model nonlinearity. This paper discusses the reliability of Clarke's marginal curvature measures (viz-a-viz other curvature and nonlinearity measures) in practical settings, and introduces a design criterion that combines variance minimization with nonlinearity minimization.

Key Words: Generalized Nonlinear Models; Geometric Designs; Lack of Fit; Marginal Curvature; Model Robustness; Nonlinearity.

1.   Introduction

Researchers often find that nonlinear regression models are more applicable for modelling various biological, physical and chemical processes than are linear ones since they tend to fit the data well and since these models (and model parameters) are more scientifically meaningful. These researchers are thus often in a position of requiring optimal or near-optimal designs for a given nonlinear model. A common shortcoming of most optimal designs for nonlinear models used in practical settings, however, is that these designs typically focus only on first-order parameter variance or predicted variance, and thus ignore the inherent nonlinear of the assumed model function. Another shortcoming of optimal designs is that

they often have only p support points, where p is the number of model parameters. In this paper, we set forth and illustrate two new robust design procedures for nonlinear models – the first underscoring the use of geometric designs, and the second focusing on efficient parameter estimation as well as reduced curvature. We also provide recommendations as to a means to choosing one of these procedures over the other. These two design procedures are given and discussed in Sections 5 (geometric method) and 6 (reduced curvature); we first provide some background in optimal design, curvature, and illustrations in Sections 2, 3 and 4, respectively.

2.   Background in Design

For a given process and/or set of data, researchers typically have in mind a statistical model. This model includes (1) an assumed distribution for the response variable, (2) a link function connecting the expected response with the independent variable(s), (3) a (mean) model function $\eta(\mathbf{x},\mathbf{q})$, and (4) a variance function (perhaps depending on $\mathbf{q}$ and/or additional parameters). The goal of the study is often to choose a design to efficiently estimate the p parameters in $\mathbf{q}$ - although sometimes designs are chosen to best discriminate between model functions in a class, or both.

An n-point design, which is a probability measure, is denoted ξ here and written

$$\mathbf{x} = \left\{ \begin{array}{l} x_1, x_2, ..., x_n \\ w_1, w_2, ..., w_n \end{array} \right\}$$

Here the ω_k are non-negative weights that sum to one, and the x_k (may be vectors) belong to the design space and are ***not necessarily distinct***.

For the model function $\eta(x,\mathbf{q})$, the *Jacobian matrix* is $\mathbf{V} = \partial \mathbf{h}/\partial \mathbf{q}$ (of dimension nxp) and the pxp Fisher *information matrix* is

$$\mathbf{M}(\xi,\mathbf{q}) = \mathbf{V}^{\mathbf{T}}\mathbf{W}\mathbf{V}$$

where $\mathbf{W} = \text{diag}\{\omega_1, \omega_2, ..., \omega_n\}$. The first-order

(and asymptotic) variance of the least-squares estimator of $\mathbf{q}$ is proportional to $\mathbf{M}^{-1}$, so designs are often chosen to minimize some convex function of $\mathbf{M}^{-1}$. For example, designs which minimize it's determinant are called *D-optimal*, and those that minimize it's trace are called *A-optimal*. Since for nonlinear models, $\mathbf{V}$ depends upon $\mathbf{q}$, a Bayesian strategy is sometimes used.

The (first-order) variance of the predicted response at X = x is given by

$$d(x,\xi,\mathbf{q}) = [\partial\mathbf{h}(x)/\partial\mathbf{q}]^T\mathbf{M}^{-1}[\partial\mathbf{h}(x)/\partial\mathbf{q}]$$

and designs that minimize (over ξ) the maximum (over x) of $d(x,\xi,\mathbf{q})$ are called *G-optimal*.

The *General Equivalence Theorem* (GET) of Kiefer and Wolfowitz (1960) proves that D- and G-optimal designs are equivalent, and that the variance function evaluated using the D-/G-optimal design does not exceed the line (or hyper-plane) y = p (i.e., the number of model parameters) – but that it will exceed this line for *all* other designs. A corollary establishes that the maximum of the variance function (p) is achieved for the D-/G-optimal design at the support points of this design. The GET was extended to homoskedastic nonlinear models in White (1973), and to heteroskedastic nonlinear models in Downing *et al* (2001).

A-optimality is generally used in "block design" situations (including cyclic-, row-, column- and alpha-designs) since it focuses on minimizing the average variance of the parameters (or contrasts of parameters). Thus, Kerr and Churchill (2001) recommend this criterion for gene expression microarray data. In contrast, D- (or G-) optimality is recommended for regression models since this criterion (but not A- or E-optimality) is invariant to a linear or nonlinear change in scale; see Silvey (1980). Uniqueness and cardinality associated with optimal design (though somewhat convoluted) are discussed in Gaffke (1987), Vila (1991), and references provided therein.

In most practical situations, optimal designs for p parameters have only p support points, thereby providing no ability to test for lack of fit of the assumed model. We therefore desire near-optimal "robust" designs which have extra support points.

3. Background in Confidence Intervals and Curvature

Popular confidence regions for nonlinear models include the Wald and likelihood-based regions. These are connected in that Wald confidence regions (WCR) use the following linear approximation. The corresponding $(1-\alpha)*100\%$ WCR is given by

$$\{\mathbf{q}\,\epsilon\,\mathbf{Q}: (\mathbf{q} - \mathbf{q}_*)^T\mathbf{V}_*{}^T\mathbf{V}_*\,(\mathbf{q} - \mathbf{q}_*) \leqslant ps^2F_\alpha\,\}$$

whereas the $(1-\alpha)*100\%$ likelihood-based region (LBCR) is given by

$$\{\mathbf{q}\,\epsilon\,\mathbf{Q}: S(\mathbf{q}) - S(\mathbf{q}_*) \leqslant ps^2F_\alpha\,\}$$

Here $S(\theta) = \varepsilon^T\varepsilon = \|y - \eta(\theta)\|^2$ is the sum of squares function. So, if we have

$$\eta(\mathbf{q}) \approx \eta(\mathbf{q}_*) + \mathbf{V}_*(\mathbf{q} - \mathbf{q}_*),$$

then $\varepsilon = \varepsilon_* - \mathbf{V}_*(\mathbf{q} - \mathbf{q}_*)$, and $S(\mathbf{q})-S(\mathbf{q}_*) \approx (\mathbf{q} - \mathbf{q}_*)^T\mathbf{V}_*{}^T\mathbf{V}_*\,(\mathbf{q} - \mathbf{q}_*)$. Thus, the extent to which $\eta(\mathbf{q}) \not\approx \eta(\mathbf{q}_*) + \mathbf{V}_*(\mathbf{q} - \mathbf{q}_*)$ is captured in *curvature/nonlinearity,* and is discussed below; first, we provide an important illustration.

**Example 1.** The two-parameter log-logistic (LL2) model function, given by

$$h(x,\mathbf{q}) = \frac{1}{1 + (x/\mathbf{q}_2)^{\mathbf{q}_3}},$$

is fitted to (and plotted with) the dataset portrayed in the following Figure.


LL2 fit to 6-point uniform design data
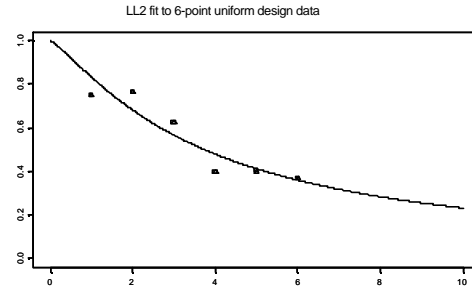
Note that for this model function, θ_2 is the LD_{50} or ED_{50} parameter (the value of the independent variable such that the expected response is ½) and that θ_3 is the slope parameter.; for the given dataset, the parameter estimates are 3.74 and 1.22 respectively. The 99% (outer region), 95% and 90% (inner region) likelihood-based confidence regions (LBCRs) for these two model

parameters and this dataset are given in the following Figure. (The central point in the plot corresponds to the least-squares estimate of the model parameters.)
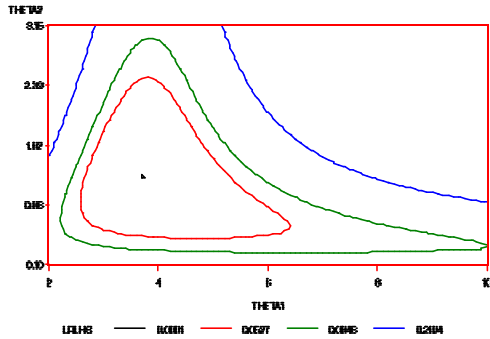

On the other hand, the corresponding Wald confidence regions (WCRs) are graphed below; note the substantial departure of the Wald approximate regions from the more-reliable likelihood-based regions.
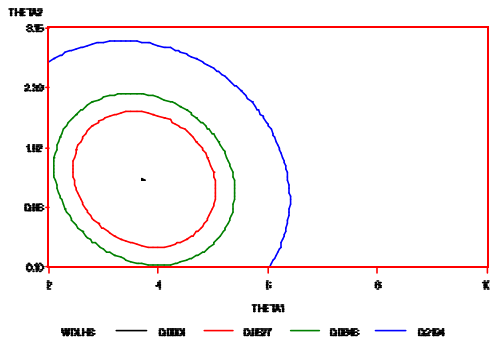
Since our focus is usually on single parameters, confidence intervals are obtained by "profiling out" the nuisance parameters (see Seber and Wild, 1989, or Pawitan, 2001). Graphed below is the profile likelihood curve for θ_2, as are cut-lines corresponding to 99% (top line), 95%, and 90% (bottom line).
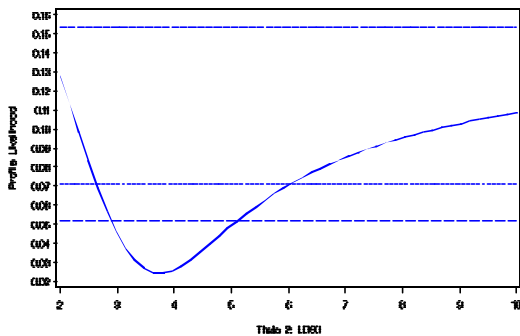
Profile likelihood confidence intervals (PLCIs) are then obtained from the intersection of the profile curve and the corresponding cut-line. In contrast, the Wald methodology replaces the actual profile likelihood curve with a parabola, thereby always yielding symmetric Wald confidence intervals (WCIs). In the graph above, note the degree to which the profile curve departs from a parabola. In contrast, the profile curve for parameter θ_3, graphed below, is closer to looking like a parabola.
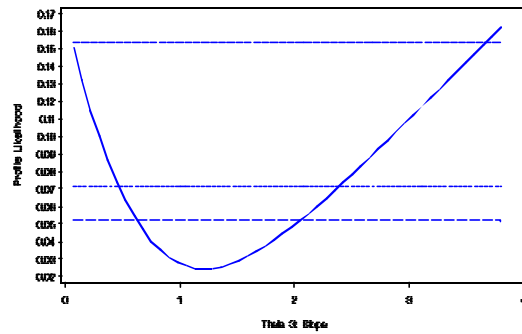
Since the latter profile likelihood curve is closer to a parabola, one would expect that the degree of curvature or nonlinearity associated with θ_2 should be higher than that associated with θ_3. Indeed this is the case here, as demonstrated by the following Wald and profile likelihood 95% confidence intervals for the two model parameters; one can claim a low degree of nonlinearity if the WCI and PLCI are nearly coincidental.

	WCI		PLCI	
$\theta_2 - LD_{50}$	2.512	4.968	2.631	6.046
$\theta_3 - Slope$	0.401	2.033	0.471	2.390

Before leaving this example, we point out that the (unique) "local" D-optimal design for this model function places half the weight at $x_1 = 1.59$ and the other half weight at $x_2 = 8.82$. This two-point design is the one which minimizes the volume (area in this instance) of the Wald CR, therefore focusing equally on both model parameters. Again, since this design has only two support points, the model function will fit the data perfectly and provide no test for model mis-specification (or lack of fit). ∎

Turning our attention to measures of curvature or nonlinearity, the following notation and definitions are useful.

Θ = the p-dimensional parameter space.
E = expectation surface; p-dimensional in the n-dimensional sample space.
IN (intrinsic) curvature = degree of flatness of E
PE (parameter effects) curvature = degree to which $SPEL$ in Θ are mapped onto $SPEL$ on E
$SPEL$ = straight, parallel, equi-spaced lines (a rectangular grid)

Over the past several years, numerous measures of nonlinearity and the like have been proposed including

- root-mean-square (RMS) IN and PE measures in Beale (1960)
- maximal IN and PE in Bates and Watts (1988)
- maximal curvature measures for parameter subsets in Cook and Goldberg (1986) and Hamilton (1986),
- bias measures in Box (1971) and Cook *et al* (1986)
- a second-order variance approximation in Clarke (1980)
- a second-order information approximation in Pázman (1986)
- asymmetry measures in Ratkowsky (1983,1990) and Hougaard (1985)
- reparameterization methods in Hougaard (1986)
- graphical procedures such as confidence curves in Cook and Weisberg (1990).

This extensive list often leaves the practitioner feeling overwhelmed and confused. Further, some of the conventional curvature measures are spurious in the sense of indicating a problem when there is none and vice versa; examples are given in Cook and Witmer (1985). Clarke (1987) points out that the IN and PE measures "suffer from the practical defect, however, of attempting to measure a multidimensional phenomenon by a single quantity" (p.844).

By expanding the estimate of θ_K in powers of σ^2 so as to adjust the endpoints of the WCI to bring them more in line with the PLCI, Clarke (1987) introduces so-called marginal curvatures. These marginal curvatures are then used to adjust Wald confidence intervals (WCI) for the parameters, one at a time. If we let the WCI for θ_K be written

$$[W_L^K, W_U^K], \text{ for}$$
$$W_L^K = \theta_K - t*SE_K$$
$$W_U^K = \theta_K + t*SE_K$$

then we can write the corresponding marginal curvature adjusted confidence interval (MCCI) as

$$[M_L^K, M_U^K], \text{ for}$$
$$M_L^K = \theta_K - t*(1 - \Gamma_a t + \beta_a t^2)*SE_K$$
$$M_U^K = \theta_K + t*(1 + \Gamma_a t + \beta_a t^2)*SE_K$$

Note that Γ_a and β_a are functions of the second and third derivatives of η with respect to θ; these calculations can be somewhat involved since the first derivative (Jacobian matrix) is of dimension nxp, the second derivative is an nxpxp array, and so on. With regard to the roles of these two measures, we underscore the length and skewness of MCCIs given in the following table.

	Wald	MC (Clarke)
Length	$2*t*SE_K$	$2*t*SE_K*(1 + \beta_a t^2)$
Skewness	0	$2*\Gamma_a*t^2*SE_K$

So, like likelihood intervals, MCCI's can be skewed (captured by Γ_a) and/or widened/ narrowed (captured by β_a). Skewed intervals are usually more sensible since information can often be asymmetric. Finally, note that the Hougaard skewness measure given in SAS (PROC NLIN) is directly related to Γ_a but ignores β_a; that this is not always a good idea is discussed in Haines *et al* 2004.

Example 1 continued. Here for θ_2, $\Gamma_a = 0.0903$, $\beta_a = 0.0281$. For $\alpha = 5\%$, we obtain for θ_2:

Type	Confidence interval	Overlap to PLCI
Wald	(2.512 , 4.968)	66.2%
MC	(2.554 , 5.542)	83.4%
PL	(2.631 , 6.046)	----

Empirical and theoretical evidence shows that MCCI's perform better in approximating PLCI's than do WCI's. Note that since calculating PLCI's is cumbersome, after finding a WCI and a MCCI, and distinguishing one of four cases (in the order MWWM, MWMW, WMWM, or WMMW), Haines *et al* (2004) uses a function (denoted f_1) *which assesses the overlap of the WCI to the MCCI* as an indicator of when a problem exists. Practitioners are provided an

indication: (1) use the WCI from package (e.g., SAS/PROC NLIN), (2) use the MCCI found in SAS/IML, or (3) go through the endeavor of finding the corresponding PLCI. We return to this function by including it in the design procedure given in Section 6, but first provide some additional examples that exemplify commonly chosen designs.

4. Some Additional Illustrations

To set the stage for the geometric design procedure given and discussed in the next section, we give several examples that illustrate the types of designs used by practitioners.

Example 2. Collett (2003) provides data related to tobacco budworms in which sets of 20 larvae of each gender were exposed to various doses of an insecticide, and the number killed or impaired were recorded. The results are given in the following table.

| Sex	Dose					
	1	2	4	8	16	32
Male	1	4	9	13	18	20
Female	0	2	6	10	12	16

Of course, these data might best be analyzed using binary logistic regression methods; our point here is to highlight the use of the design $x = 2^K$ for $K = 0, 1, \ldots 5$. ∎

Example 3. ELISA (enzyme-linked immunosorbent assay) data are provided in Huet *et al* (1996) wherein the response is optical density and the independent variable (x) is of the form $x = \log_{10}(1/d)$ and $1/d$ is reciprocal dilution. The chosen values of $1/d$ for this study (i.e., the experimental design) were 30, 90, 270, 810, 2430, 7290, 21869, and 65609, or essentially, $x = 30*3^K$, for $K = 0, 1, \ldots 7$. ∎

Example 4. Seefeldt et al (1995) reports data in which the response is dry weight of weeds and x is the dose of a herbicide, and a graph of the data follows. A reasonable model function for these data is the four-parameter log-logistic model, given by

$$h(x, \boldsymbol{q}) = \boldsymbol{q}_4 + \frac{\boldsymbol{q}_1 - \boldsymbol{q}_4}{1 + (x / \boldsymbol{q}_2)^{\boldsymbol{q}_3}}$$

so that, as in Example 1, θ_2 is the LD_{50} and θ_3 is the slope; here, θ_1 and θ_4 are the upper and lower asymptotes respectively.
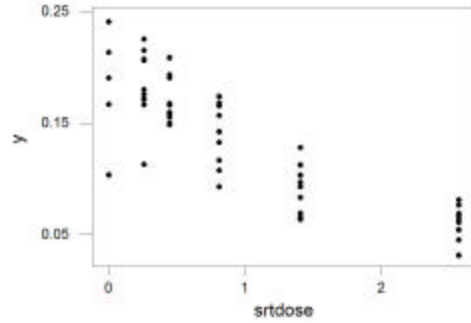
Noting the non-constant variances, we would be wise to model the variance as well, viz,

$$\mathrm{Var}(Y) = \sigma^2 \eta^\rho(x),$$

where ρ is an additional model parameter to be estimated. ∎

Example 5. In pharmacological (PK) modelling, the response variable is drug concentration and the independent variable is time, and repeated measurements (over time) are recorded for each subject. Such is the case in Atkinson et al (1993), where the three-parameter intermediate product (IP3) model function,
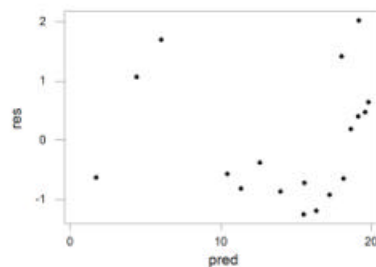
$$h = \boldsymbol{q}_3 \left\{ e^{-\boldsymbol{q}_2 t} - e^{-\boldsymbol{q}_1 t} \right\}$$

is used. We choose to write this model function as

$$h = \frac{f\boldsymbol{q}_1\boldsymbol{q}_2}{\boldsymbol{q}_1 - \boldsymbol{q}_2} \left\{ e^{-\boldsymbol{q}_2 t} - e^{-\boldsymbol{q}_1 t} \right\}$$

so that ϕ is the so-called area under the curve (AUC), or sum total of the drug delivered to the subject. The following residual plot highlights

Residual Plot for ACHJ data with heterosk. IP3 fit

that the residuals may follow a structure such as an AR1 error structure, and should be modeled as such. ∎

The above illustrations highlight the importance of the correct choice for the (mean) model function, the variance function, and the correlation structure. These examples also underscore some of the designs used in practice. Examples 2 and 3 follow the "x = a*b^K" pattern, with a = 1 and b = 2 for Example 2, and a = 30 and b = 3 for Example 3. Interestingly, the doses chosen for Example 4 are 0, 0.066, 0.198, 0.660, 1.98, and 6.60, or 0, c, 3c, 10c, 30c, 100c for c = 0.066. Finally, the time points chosen for Example 5 are 1/6, 1/3, 1/2, 2/3, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 8, 10, 12, 24, 30 and 48 hours; optimal design strategies for PK models is discussed in Retout *et al* (2002). In the next section, we explore the geometric design approach used in Examples 2 and 3.

5. Robust Geometric Design Strategy

Returning to the LL2 model highlighted in Example 1, and for t = $(x/q_2)^{q_3}$, the local D-optimal design places the weight ω = ½ at each of t_1 = 0.352175 and t_2 = 1 / t_1 = 2.839498; D-optimality of this design is confirmed by noting that the corresponding variance function, graphed below, does not exceed the line y = 2.
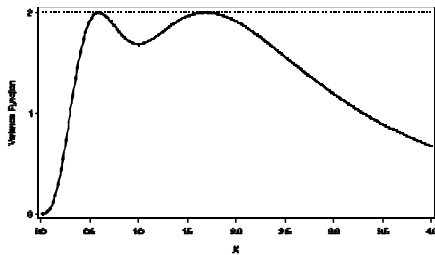

It follows that one can solve for the values of x by solving the equations $x_1 = \theta_2 (0.352175)^{1/\theta_3}$ and $x_2 = \theta_2 (2.839498)^{1/\theta_3}$ for an array of choices of θ_2 and θ_3. Barring such a (pseudo-Bayesian) strategy, optimal designs for this model suffer from only having two support points.

Similar results are observed for the three-parameter log-logistic (LL3) model function given by the expression

$$h = \frac{q_1}{1+t} = \frac{q_1}{1+(x/q_2)^{q_3}}$$

Not surprisingly, the D-optimal design for this LL3 model includes the two given for the LL2 model along with the point x = t = 0, and D-optimality is indeed established from the following graph of the corresponding variance function.
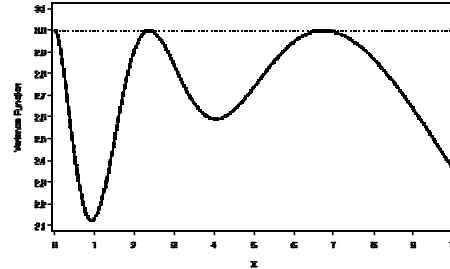
A rival strategy to choosing an optimal design – provided the loss in information is not too great – is to seek a design with a geometric pattern akin to the designs used in Examples 2 and 3 in the last section. Thus, for the LL3 model function, we choose

$x_1 = 0$, $x_2 = a$, $x_3 = a*b$, $x_4 = a*b^2$, and $x_5 = a*b^3$,

with the weight ω placed at x_1 = 0 and the remaining weight 1 - ω divided evenly across the remaining four points. Subject to these constraints and choosing the D-optimality criterion, we thus seek designs to maximize the determinant of the information matrix over the triple (a, b, ω). Regardless of the values of θ_2 and θ_3, this approach produces designs for which the expected response are approximately 100%, 80%, 60%, 40% and 20% of the maximum value of E(Y) (i.e., θ_1) , and are thus easily obtained given a reasonable sketch of the anticipated model function. This situation is graphed below for θ_2 = 4 and θ_3 = 2, for the optimal a = 1.945, b = 1.597, and ω = 0.326. Note that the x-values are then approximately 0, 2, 3, 5 and 8.
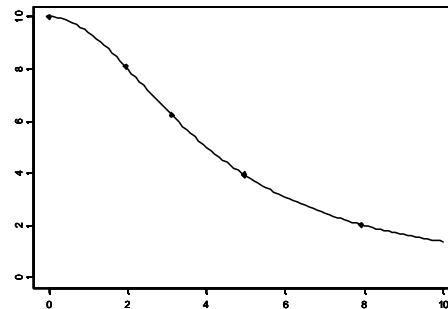
Since the weight at x = 0 is approximately 1/3, a

final sample size of $N = 6n$ is indicated.

We can assess the loss of information of the use of the design ξ relative to the three-point D-optimal design (ξ_D) by using the D-efficiency,

$$D_{EFF} = [\det(\xi)/\det(\xi_D)]^{1/p}$$

In this case, it turns out that $D_{EFF} = 94.8\%$, meaning that the use of this five-point geometric design results in just a 5.2% information loss, yet it provides us with a means to test for lack-of-fit, and may thus be more practical. We point out that the rival geometric design which takes measurements at $x = 0, 1, 2, 4, 8$, results in an information loss of approximately 12.5%.

6. Robust MC-Compromise Design Strategy

As indicated at the end of Section 3, a reasonable and reliable measure of nonlinearity associated with the kth model parameter is the function f_1, which itself depends on the corresponding marginal curvature. We can thus combine variance minimization and curvature reduction into one design criterion function by considering

$$CF(\xi) = \lambda\, \frac{\det\left(M^{-1}(\boldsymbol{x})\right)}{\det\left(M^{-1}(\boldsymbol{x}_D)\right)} + (1-\lambda)\, \frac{f_1\left(\boldsymbol{x}_{MAX};\boldsymbol{q}_K\right)}{f_1\left(\boldsymbol{x};\boldsymbol{q}_K\right)}$$

Here λ is chosen between 0 and 1; note that $\lambda = 1$ corresponds to D-optimality, and $\lambda = 0$ corresponds to the reduced curvature (RC) design, or maximizing the overlap between the WCI and the MCCI for θ_k.

Example 1 continued. For the LL2 model function and the above parameter values, the D-optimal design is the equal-weight two-point design $\{x_1 = 1.587, x_2 = 8.816\}$, and the generalized variance is graphed below. In this
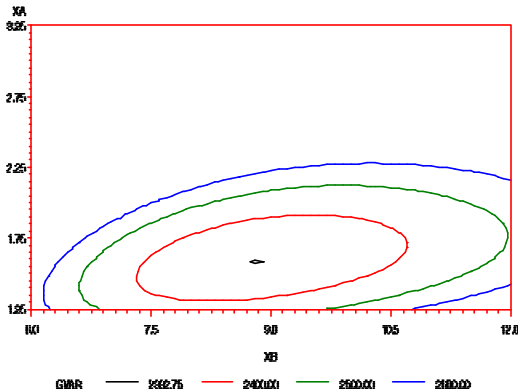
contour plot, x_1 is plotted on the vertical axis, x_2 is plotted on the horizontal axis, and the contours correspond to equal generalized variance, with the minimum at the point ($x_1 = 1.587$, $x_2 = 8.816$).

Similarly, the RC design is the equal-weight two-point design $\{x_1 = 2.458, x_2 = 9.455\}$, and contours of the function f_1 are graphed below.
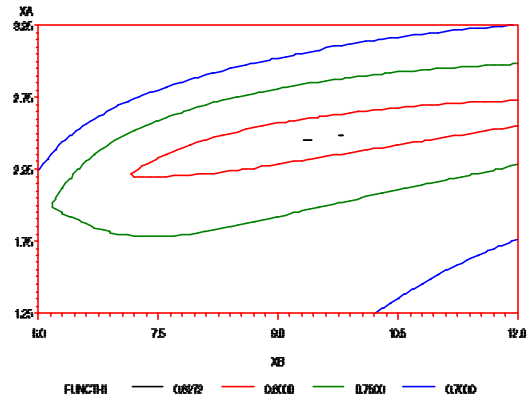
For this model, the design obtained by minimizing the above criterion function would lie somewhere between these two designs, depending on the choice of λ.

So as to compare various design strategies, we evaluate the performance of six designs here. The first design (ξ_1) is the original six-point design used in this study, $x = 1, 2, 3, 4, 5,$ and 6. The next design (ξ_2) is three replicates of the D-optimal design and the third design (ξ_3) is three replicates of the RC design. The fourth design (ξ_4) is two replicates of the D-optimal design and one replicate of the RC design, whereas the fifth design (ξ_5) is one replicate of the D-optimal design and two replicates of the RC design. In order to tie our methods here with those of the previous section, our last design considered here (ξ_6) is a six-point geometric design of the form a, $a*b$, $a*b^2$, … $a*b^5$. The optimal value of "a" here turns out to be very nearly unity and optimal $b = 1.665$, so this design is $x = 1, 1.665, 2.772, 4.615, 7.684,$ and 12.79. In order to visualize these six designs, we provide a plot in the Appendix at the end of the paper.

In order to assess each of these six designs, we record the generalized variance (denoted "G-Var") and the function f_1 (for the parameter θ_2) in the following table. Not surprisingly, designs ξ_2 - ξ_5 do quite well here in terms of both

measures, whereas the chosen uniform design

Design	G-Var	$f_1(\theta_2)$
ξ_1	444.5 (76.4%)	0.797 (89.5%)
ξ_2	259.2 (100.0%)	0.830 (93.2%)
ξ_3	308.6 (91.6%)	0.891 (100.0%)
ξ_4	269.0 (98.2%)	0.844 (94.8%)
ξ_5	284.8 (95.4%)	0.864 (97.0%)
ξ_6	314.9 (90.7%)	0.833 (93.6%)

(ξ_1) does not fair well at all. The geometric design (ξ_6), on the other hand, performs quite admirably. Similar results have been observed via simulation and for other settings. ∎

7.   **Conclusion and Recommendations**

This paper highlights several important points when designing studies involving nonlinear regression modelling. First, the complete model needs to be taken into consideration, including the assumed distribution of the response variable, the mean model function, the variance function (including covariance structure), and the link function.   Second, optimal designs with minimal support in general provide no ability to test for model mis-specification, and are therefore of only limited use. Third, in most settings, curvature needs to be considered and should be included in the design criterion; our criterion uses Clarke's marginal curvature measures. Finally, geometric designs provide a reasonable and important class of designs worth considering – especially those of the form $a*b^K$ illustrated here, and where "a" and "b" are chosen in some optimal sense. For example, in the example at the end of the last section, the choice b = 1.665 translates into dilutions of $1/b \approx$ 0.6006, or about 60%. These latter designs are thus usually practically feasible, and the additional support points provide the researcher with the ability to test for model lack of fit.

Note that the variance-curvature criterion given in the previous section is only one of several which can be suggested. For example, Clarke and Haines (1995) introduces an alternate design criterion which yields designs with reduced curvature.   Clearly more research is needed comparing these two criteria.

References

Atkinson, A.C., Chaloner, K., Herzberg, A.M. and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model, *Biometrics*, 49, 325-337.

Bates, D.M. and Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications.* New York: Wiley.

Beale, E.M.L. (1960). Confidence regions in non-linear estimation (with Discussion). *J. R. Stat. Soc. B*, 22, 41-88.

Box, M.J. (1971). Bias in nonlinear estimation (with Discussion). *J. R. Stat. Soc. B*, 33, 171-201.

Clarke, G.P.Y. (1980). Moments of the least squares estimators in a non-linear regression model. *J. R. Stat. Soc. B*, 42, 227-237.

Clarke, G.P.Y. (1987). Marginal curvatures and their usefulness in the analysis of nonlinear regression models. *J. Amer. Stat. Assoc.*, 82(399), 844-850.

Clarke, G.P.Y. and Haines, L.M. (1995). Optimal design for models incorporating the Richards function, in Seeber, G.U.H., Francis, B.J., Hatzinger, R. and Steckel-Berger, G., eds., *Statistical Modelling*, New York: Springer.

Collett, D. (2003). *Modelling Binary Data*, 2nd Edition. Boca Raton, FL: Chapman & Hall.

Cook, R.D. and Goldberg, M.L. (1986). Curvatures for parameter subsets in nonlinear regression. *Ann. Statist.*, 14, 1399-1418.

Cook, R.D., Tsai, C.-L., and Wei, B.-C. (1986). Bias in nonlinear regression. *Biometrika*, 73, 615-623.

Cook, R.D. and Weisberg, S. (1990). Confidence curves in nonlinear regression. *J. Amer. Stat. Assoc.*, 85(410), 544-551.

Cook, R.D. and Witmer, J.A. (1985). A note on parameter-effects curvature. *J. Amer. Stat. Assoc.*, 80, 872-878.

Donaldson, J.R and Schnabel, R.B. (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics*, 29(1), 67-82.

Downing, D., Fedorov, V.V., and Leonov, S. (2001). Extracting information from the variance function: optimal designs, in Atkinson, A.C., Hackl, P. and Mueller, W.G., eds., *MODA6 – Advances in Model-Oriented Design and Analysis*. Heidelberg: Physica-Verlag, 45-52.

Gaffke, N. (1987). On D-optimality of exact linear regression designs with minimal support.. *J. Stat. Plan. Inf.*, 15, 189-204.

Gallant, A.R. (1987). *Nonlinear Statistical Models*. New York: Wiley.

Haines, L.M., O'Brien, T.E. and Clarke, G.P.Y. (2004). Kurtosis and curvature measures for nonlinear regression models. *Statistica Sinica*, 14(2), 547-570.

Hamilton, D. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika*, 73, 57-64.

Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *J. R. Stat. Soc. B*, 47, 103-114.

Hougaard, P. (1986). Covariance stabilizing transformations in non-linear regression. *Scand. J. Stat*., 13, 207-210.

Huet, S., Bouvier, A., Gruet, M-A. and Jolivet, E. (1996). *Statistical Tools for Nonlinear Regression: A Practical Guide with S-Plus Examples*. New York: Springer.

Kerr, M.K. and Churchill, G.A. (2001). Experimental design for gene expression microarrays. *Biostatistics*, 2, 183.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Can. J. Math.*, 12, 363-366.

O'Brien, T.E. Modelling and Design to Detect Interaction of Insecticides, Herbicides and Other Similar Compounds, in Johnson, D. and

Milliken, G., eds. (2004). *Proc. 15th Conference on Applied Statistics in Agriculture*. Kansas State U. Press, 303-321.

O'Brien, T.E. and Funk, G.M. (2003). A Gentle Introduction to Optimal Design for Regression Model. *American Statistician*, 57, 4, 265-7.

Pawitan, Y. (2001). *In All Likelihood*. Oxford: Clarendon Press.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*. Dordrecht, Holland: D. Reidel.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.

Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. New York: Marcel Dekker.

Retout, S., Mentré, F. and Bruno, R. (2002). Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics, *Stat. in Med.*, 21, 2623-2639.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*. New York: Wiley.

Seefeldt, S.S., Jensen, J.E. and Fuerst, E.P. (1995). Log-logistic analysis of herbicide dose-response relationships. *Weed Technology*, 9, 218-227.

Silvey, S.D. (1980). *Optimal Design: An Introduction to the Theory for Parameter Estimation*. London: Chapman and Hall.

Vila, J.P. (1991). Local optimality of replications from a minimal D-optimal design in regression: a sufficient and quasi-necessary condition. *J. Stat. Plan. Inf.*, 29, 261-277.

White, L.V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*. 60, 345-348.

Appendix

Dotplot of x vs design no.