Design Strategies in Fitting a Nonlinear Model

Michael Van Supranes
School of Statistics
University of the Philippines Diliman

Estimation of parameters in a nonlinear model depends on the distribution of data points along various levels of curvature in the function to be estimated. Using Monte Carlo simulation, an optimal allocation procedure for building stratified designs was derived. The optimal allocation procedure conforms well to a proportionality property, directly relating the number of observations with the total curvature and measure or length of the domain. The proportionality property can be used to easily construct an allocation procedure that is near optimal. Simulation studies show that strategic stratification can improve the prediction accuracy of uniform designs.

Keywords: *stratification, experimental designs, spline regression, Monte Carlo simulation*

1. Introduction

1.1 Preliminaries

Only two points have to be joined in order to define a straight line, but several points have to be interpolated to define or merely approximate a curve. These geometric properties served as the motivation for formulating curvature-based strategies in building experimental designs for estimating a nonlinear model.

Suppose that we are interested in estimating a nonlinear function from the results of an experiment, response variable can be predicted within the range of the observed values. The idea is to provide an optimal solution to the problem by minimizing the error of the estimated function through an experimental design that allocates more observed points along the more curved part of the function.

The idea is very intuitive and would be helpful in optimizing a model's prediction accuracy within a fixed number of sample points. Simulation results were consistent to this expectation but other factors in allocation were also observed. In addition, the idea may open the possibility of design hybridization, which may be explored in future studies. Design hybridization is possible in a stratified experimental design, as how different sampling procedures can be used in different strata in the context of sampling designs. For instance, a complicated model may be handled by approximating some parts using a polynomial function. A polynomial function would be simpler and is a much easier function to differentiate or integrate. Under a stratified framework, an optimal design can be constructed from the approximated parts, while a uniform design of points can be applied in complex or unknown parts. Hence, there is a hybrid of optimal and uniform designs.

The purpose of this study is to find an optimal allocation procedure proportional to the total curvature of a function within a subset of its domain or stratum. In addition, we are interested in determining the effect of stratification in uniform designs. A series of scenarios was generated via Monte Carlo Simulation, and the estimation performance of the stratified designs was also calculated for each scenario. Then, the approach of this research was to fit a curve based on the results of the Monte Carlo simulation in order to find a general pattern in defining a stratified optimal allocation procedure. Finding optimal designs through curve fitting of Monte Carlo experiments has been used in several researches (Muller and Parmigiani, 1995; Hertel and Kohler, 2012). To facilitate ease in visualization, this study was limited to bivariate scenarios. There is only one response variable and one explanatory variable. In addition, the study focused on the use of splines as smoothers for the estimation of the nonlinear function. The class of designs considered was forced to have observed values at the boundaries of the experiment domain.

1.2 Statistical designs of experiment

In experiments, construction of the design is as important as the implementation of the experiment. As early as 1950s, there were already researches on methods of finding efficient designs statistically. Nowadays, statistically designed experiments are becoming a trend in certain fields such as biochemistry and engineering. For instance, statistically designed experiments, specifically uniform designs, are becoming more popular in industrial management (Fang and Lin, 2003).

Designing an experiment may be viewed as choosing a subset of points in an experimental domain for which the response variable will be observed. The experiment domain is defined as a subset or subspace of all the possible values of the explanatory variables (O'Hagan, 1977). In general, experimental designs have varied from completely randomized designs or blocked designs—in which the response variable is observed a number of times for a set of representative points in an experiment domain (Montgomery 1997; Bailey, 2012)—to optimal or space-filling designs, in which the response variable is observed only once for a strategic set of points in the experiment domain (Kiefer and Wolfowitz, 1958). The latter, optimal designs for regression, was developed for the case when factorial designs involve too many treatment combinations that replication of observation has to be sacrificed in order to have a good coverage of scenarios. Hence, designing an experiment varies from the problem of finding the optimal number of replicates and treatment levels, to the problem of finding a set of observable points in the experiment domain for which an optimality criterion is attained. The design that meets the desired optimality criterion is called an optimal design (Atkinson and Donev, 1992). There are many types of optimal design depending on the desired optimality criterion.

In finding an optimal design, a mathematical model is commonly assumed to be known but has unknown parameters (Fang and Lin, 2003). The common approach is to use the information matrix of the assumed model as an objective function for optimization, which is analytically or computationally tractable. In addition, optimization is constrained within a predefined experiment domain and a predefined subset of the parameter space (Atkinson and Donev, 1992; Muller and Parmigiani, 1995). To facilitate optimization, the objective function is usually expressed as a function of a design vector.

A design vector is a function that maps a subset of the experiment domain to a design space, which is a vector space of elements describing a particular design. Examples of these elements are the number of replicates, distances between points, cost of design and what not (Muller and Parmigiani, 1995). Mapping the experiment domain to a design domain reduces the dimension of the pre-image of the objective function. Thus, an optimal design can also be viewed as a set of designs that satisfy a design vector for which the objective function is optimal within the predefined parameter space (O'Hagan, 1977).

However, a mathematical model is usually unknown in practice. In addition, it is common in computer experiments to have a known model that is too complicated, in such cases the model can be viewed as a black box or unknown. For both cases, the needed strategy is a space filling design. Fang and Wang proposed uniform design (UD) for the case of black box or unknown models (Fang, 1980).

The uniform design is a space filling design that is found to be robust to model specification provided that the variation does not change (Xie and Fang, 2000). Uniform design is a set of points in the experiment domain that are uniformly spread on the projection of the unknown or complicated model (Fang and Lin, 2003).

1.3 Regression and designs

Experiments are conducted to establish cause-and-effect. Cause-and-effect is just a specific form of relationship, which can be expressed in terms of a regression model. For this case, design becomes a factor in building a regression model.

Concepts on optimum design theory for regression problems were discussed by Kiefer and Wolfowitz (1958) and O'Hagan (1977).

Design optimization has been applied to both nonlinear and linear regression, parametric or nonparametric. Design optimization is especially important for costly and time-consuming experiments. For instance, fatigue behaviour of materials is commonly described in terms of a nonlinear function relating different factors such as strain and stress. Experiments in these scenarios usually involve running a strain amplitude over a material until it fails and then the stress amplitudes are observed. The procedure is time consuming that the levels of strain amplitude must be strategically selected such that the nonlinear function will still be estimated with a desired optimality (Hertel and Kohler, 2012). With such constraints, researches are commonly limited to few observations.

In kernel regression, additional factors such as bandwidth allocation are considered in finding the optimal design and asymptotically optimal designs (Muller, 1984).

1.4 Curvature and splines

A curvature of a line is the degree of deviation from being straight. It can also be viewed as the tendency of the rate of change to shift from point to point in an arc length. The curvature of a function fx (x) at $X = x$ is measured as the value of the second derivative $\frac{d^2 f(x)}{dx^2}$ at $X = x$. The first derivative of a function describes the slope of a tangent line, while the second derivative describes the change in slope of a tangent line as it moves along a curve.

A straight line would have the same slope for any two points within the line. The slope of the tangent line being constant at any point would translate to a curvature of 0. There is no tendency for the tangent line to change its slope. On the other hand, variation in slope is expected as a line deviates from being straight (Leithold, 1996).

Splines are smooth curve functions that can be used to interpolate or approximate a set of control points (Gillies, 2014). In addition, a spline is a smoother. It summarizes the trend of a response measurement of a predictor. Experiments usually may lead to nonlinear relationships. Hence, splines can be used as a smooth function in estimating a general additive model. In this study, the estimation was limited to the usage of splines. Since splines can only summarize a trend within the range of the observed values, the design space considered in the study is a collection of designs such that boundary points are always observed. Smoothers, such as splines, has a single parameter. It is a smoothing parameter. A higher smoothing parameter value results to a smoother curve. This smoothing parameter can either be estimated using cross validation method or generalized cross validation method (Wahba, 1990). For this study, estimation of the smoothing parameter was limited to the cross validation method.

2. Methodology

2.1 Stratified Optimal Allocation Procedure

The first goal is to determine an optimal allocation procedure that meets a certain optimality criterion. The optimal allocation of points in the class of stratified designs is the Stratified Optimal Allocation Procedure (SOAP). For this study, the optimality criterion is chosen to be the minimum value within the set of maximum Mean Absolute Percentage Error (MAPE). It is desired to come up with a stratified design of optimal prediction accuracy on any point within the observed range. The optimal allocation procedure can be interpreted as a design with the most desirable worst estimate.

2.2 Optimality criterion: Strategies in simulation

Computer simulation was used to generate several modeling scenarios and the function of the data generating process was predetermined. Hence, MAPE can be defined as the prediction error of the estimated function with respect to the true function, instead of the prediction error of the estimated value from the observed value. Since it is desired to compute for the MAPE of an estimated function $\hat{f}(x)$ over a continuum Ω, the MAPE must be computed in an integrated form as shown in (1).

$$\text{MAPE of } \hat{f}(x) = \frac{1}{m(\Omega)} \int_{\Omega} \left| \frac{f(x) - \hat{f}(x)}{f(x)} \right| \tag{1}$$

where m(Ω) is the measure of Ω.

On the other hand, it is hard to analytically solve for the integrated form of MAPE for an estimated function, most especially that the method of estimation involves splines. Splines are smooth interpolations, which make it integrable, but splines are too complex for analytical computations. Thus, the integrated MAPE was rather approximated via Monte Carlo simulation as shown in (2).

$$\text{MAPE of } \hat{f}(x) \approx \frac{1}{n} \sum_{x \in \omega} \left| \frac{f(x) - \hat{f}(x)}{f(x)} \right| \tag{2}$$

where ω is a set of n points in Ω.

The approximation of the integrated MAPE becomes more accurate as the number of points simulated approaches infinity. The approximation is basically an application of summation approximation for any multidimensional integral. The error of approximation has an error bound, which is defined the Koksma-Hlawka inequality (Niederreiter as cited by Xie and Fang, 2000). It states that the upper bound for the error in approximation is a direct function of uniformity of points in $\boldsymbol{\omega}$ and the variability of the function in the integrand. Uniformity is being used in the same context as in uniform designs. Thus, a good approximation can be attained as more equidistant points over Ω are simulated for the computation of MAPE. For this research, exactly 100,000 uniformly spaced points within every interval of length 1 was simulated to have good approximations for the true function, estimated function and integrated MAPE.

2.3 Stratified uniform designs

After identifying the SOAP, the stratified optimal allocation procedure was applied to a more specific class of designs called uniform designs. Uniform designs are advantageous when the assumed model is hard to integrate or is too complex; hence, the case when the model is called a black box. The idea is to improve the design based on additional information, which is curvature. Such additional information is not maximized when the function is too complex, since the information is considered as good as unknown. Nevertheless, simulation and computational tools can be used to approximate such functions.

2.4 Computation of curvature: Strategies in simulation

Computing the total curvature is necessary in the process of finding the optimal allocation procedure. The total curvature is important in implementing the allocation procedure, since the procedure was contextualized according to the total curvature of the function within each stratum. In addition, the allocation method being formulated is only concerned with the magnitude of the curvature; hence, the total curvature is computed from the absolute value of the curvature function. The formula for the total magnitude of curvature over a domain is shown in (3).

$$\tau(\mathfrak{X}) = \int_{\mathfrak{X}} \left| f''(x) \right| dx \tag{3}$$

However, analytical computation of the total curvature is a hurdle for most nonlinear functions. Integration is too difficult for complicated functions or “black box” functions. Computational approximations were used to solve this dilemma.

For the case of computer experiments, these “black box” functions usually exist internally in the system and values can be computed from the function computationally. Hence, it is still possible to compute for the total curvature of the function within a stratum through Monte Carlo simulation. The total curvature can be approximated by simulating several points on the curve and use the fact that curvature is simply the change in slope. The approximation of the total curvature of a function is shown in (4).

$$\tau(\mathfrak{X}) \approx \frac{m(\mathfrak{X})}{N-2} \sum_{i=1}^{N-2} \left| \left(\frac{f(x_{(i+2)}) - f(x_{(i+1)})}{x_{(i+2)} - x_{(i+1)}} - \frac{f(x_{(i+1)}) - f(x_{(i)})}{x_{(i+1)} - x_{(i)}} \right) \left(\frac{1}{x_{(i+1)} - x_{(i)}} \right) \right| \quad (4)$$

where $m(\mathfrak{X})$ is the measure of $\mathfrak{X}$

$x_{(i)} \in \mathfrak{X}, \forall i = 1, 2, ..., N,$ and $x_{(1)} < x_{(2)} < ... < x_{(N-1)} < x_{(N)}$.

The approximation is asymptotically accurate. Similar to approximating the integrated MAPE, the points simulated must be uniformly spread in order to attain the asymptotic behaviour of the approximation. The error minimizes as the number of points simulated increases. An illustration is shown in Table 1, wherein the total curvature along the exponential function over the interval $\mathfrak{X}$:[1,4] was approximated on an increasing number of simulated points.

Table 1. Asymptotic Property of the Approximation

Number of Simulated Points (N)	Approximated Total Curvature $\tau^*(\mathfrak{X})$	Exact Total Curvature $\tau(\mathfrak{X})$
10	130.947	145.695
50	142.595	145.695
100	144.137	145.695
500	145.382	145.695
1000	145.538	145.695
5000	145.664	145.695

2.5 Monte Carlo simulation: Procedures and strategies

Monte Carlo simulation was use to identify the optimal allocation procedure. A set of predetermined nonlinear functions and stratified designs were simulated. After observing the maximum MAPE for each simulated design, a curve was fitted to best approximate the relationship between the design and maximum MAPE. The optimal allocation procedure was then determined by locating the lowest point of the curve.

Creating an optimal design via simulation involves a mathematical model $y = f(x|\theta)$ with an unknown parameter θ. The usual approach is to simulate observations from several possible designs, typically by randomly generating a subset of points from a predefined experiment domain $\mathfrak{X}$, at several randomly selected parameter values θ within a predefined parameter space Θ, so that the optimal solution would be a design that attains optimality for any value in the predefined parameter space.

A similar approach was conducted in this study, several random subsets of the experiment domain were simulated, but instead of exploring a single function with varying parameter values, several preselected functions $f_i\ (x|\ \theta+1)$ of different shapes were considered. That is to be able to explore the optimal stratified solution for different forms of a function. Equations (5), (6) and (7) show the functions used in the simulation.

$$y_1 = f_1\ (x) = \exp(x) + 10 \tag{5}$$

$$y_2 = f_2\ (x) = x^3 + 10 \tag{6}$$

$$y_3 = f_3\ (x) = x^6 + x^3 + 10 \tag{7}$$

The first data generating process was an exponential function and the next two were polynomials with different degrees. The error term ε for all equations have a standard normal distribution. Note that the chosen functions are integrable and have a nonzero second derivative to ensure that the total curvature is analytically tractable. The number of runs/observations was kept fixed such that the only factor that varies for each design was the definition of strata and the procedure of allocation. For convenience in illustration, only two strata were considered. Since there were only two strata, stratification can be easily defined by a single cutoff point. Hence, a design vector d(x) for a stratified design would contain information on both stratification and allocation as shown in (8).

$$\text{d}:(c, n_1, n_2 \mid \text{c} \in \mathfrak{X}, n_1 + n_2 = n) \tag{8}$$

Stratification is defined by a cutoff point c, and allocation is defined by the stratum sample sizes n_1, n_2 for a fixed total sample size n. The total sample size was fixed at 30, since most optimization problems are constrained under a small sample size. It is important to fix the sample size to isolate the effect of allocation that is not due to the overall size of the sample.

The predetermined experimental domain for the exponential function was $\mathfrak{X}_1$:{x ∈ [-5,5]}, while the predetermined experimental domain for the polynomial functions was $\mathfrak{X}_2 = \mathfrak{X}_3$:{x ∈ [0,8]}. Factors and corresponding levels considered in simulation are summarized in Table 2.

Table 2: Summary of Factors and Levels in Monte Carlo Simulation

Factors	Levels
Allocation Procedure (n1 , n2) - n_1 is the sample size for $(-\infty,c] \cap \mathfrak{X}$ - n_2 is the sample size for $(c,\infty) \cap \mathfrak{X}$	(5,25)(10,20)(15,15) (20,10)(25,5)
Cutoff Points for Exponential case	-3, -2, -1, 0, 1, 2, 3
Cutoff Points for Polynomial cases	0, 1, 2, 3, 4, 5, 6, 7
Data Generating Process	Exponential, Cubic, Sixth Degree Polynomial

There is a total of 35 cases for each data generating process. Since this is an exploratory study, each case was only iterated 20 times. The simulation design is like a nested factorial design, since the set of levels for the cutoff points is different for the polynomial cases. The cutoffs are different because the predetermined experiment domain is different for the exponential and polynomial cases. The experiment domain is strategically selected for each function, such that the range of the curvature function over the domain is nonnegative. The curvature function is not required to be nonnegative, but it makes the computation of total curvature easier.

For each scenario, the maximum MAPE was computed, as well as the corresponding total curvature for each stratum. Then, a surface was fitted over the results of simulation for each data generating process. The response surface relates the response of the maximum MAPE to the different strata definitions and allocation procedures for each function.

The final step was locating the lowest point of the surface for each data generating process. Thus, the product of the simulation was a stratified optimal allocation procedure for each data generating process.

To test the application of stratified optimal allocation procedure, the maximum MAPE of the usual uniform design is compared against the maximum MAPE of stratified uniform design via Monte Carlo simulation. That is to determine if stratification and optimal allocation procedure have an advantageous effect in the prediction accuracy of nonlinear models estimated from a predefined design, specifically the uniform design.

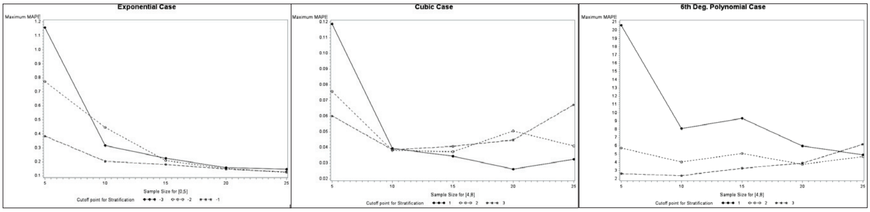
3. Results and Discussion

3.1 Stratified Optimal Allocation Procedure

Before we proceed to discussing the Stratified Optimal Allocation Procedure for each of the function, let us first observe the properties and general patterns in finding a better allocation procedure. Note that in all discussions the overall total sample size is fixed. Hence, allocation and the definition of strata are the varying factors in the design.

First, we consider the case in which there is a fixed stratification rule. These are the cases when there exist a natural cutoff in defining strata. For instance, if the variable of interest is income, there are already existing natural cutoff points in forming income brackets. For such scenarios, the problem is confined on the search for the best allocation of points.

Based on simulation results, the maximum MAPE tends to decline as more points are allocated to the stratum with a higher total curvature among the designs with the same cutoff point. Figure 1 shows a plot of the maximum MAPE against the sample size allocated in the stratum with the much higher total curvature. In figure 1, each line represents a fixed cutoff point and the cutoff points chosen are those that would divide the total curvature unevenly, where the second stratum dominates in terms of total curvature.



(Left) Exponential Case, (Middle) Cubic Case, (Right) Polynomial Case

Figure 1. Plot of Maximum MAPE over Increasing Allocation in the Second Stratum for a fixed set of Cutoff points

Given a fixed stratification rule, increasing the density of observations in the segment of experiment domain with a higher curvature would result to better designs. The observation is common to all of the three functions simulated. Note that for polynomials there seem to be a 'breaking point,' in the sense that Maximum MAPE is not always decreasing. There seem to be a point when the number of points allocated in the more curvaceous stratum is too many that the error on the less curvaceous stratum pulls up the MAPE. This result may be an indication that curvature within a stratum may not be the only governing factor in allocating the points.

The next case would be the scenario in which there are no natural strata definitions. Hence, the optimization problem becomes a balance between the right combination of cutoff point (definition of strata) and allocation procedure (stratum sample sizes). In exploring this scenario, there is a need to search for the scenarios that have a 'good' match of stratum definition and allocation procedure.

Let us define a design as 'good' match, if for a given cutoff point, it is the simulated design with the minimum maximum MAPE, or if for a given allocation procedure, it is the simulated design with minimum maximum MAPE. All designs that are 'good' match were observed and it was found out that there are two governing factors that could lead to an optimal allocation and stratification combination. The first factor is curvature, which was shown to have a direct proportionality to allocation for a given stratification rule. The second factor is the measure of the experiment domain corresponding to the stratum, which is the length of interval of interest for a one-factor scenario. As seen in figure 1, the maximum MAPE is not always decreasing with respect to increasing allocation to a more curvaceous stratum and this should not be the case if curvature is the only factor in finding a good allocation. Apparently the length of interval also affects allocation, a longer stratum with a lower total curvature may still require additional observations to attain optimality. This is more likely due to coverage problems. A long curve with a moderate curvature would be difficult to estimate with very few points.

$$n_i \propto \ell\left(\mathfrak{X}_i\right)\tau\left(\mathfrak{X}_i\right) \tag{9}$$

where n_i is the number of observations allocated to the i^{th} stratum
$\ell(\mathfrak{X}_i)$ is the length of the i^{th} stratum
$\tau(\mathfrak{X}_i)$ is the total curvature in the i^{th} stratum

To show that indeed allocation of observation is proportional to the interaction of length and curvature, the proportionality showed in statement (9) is explored among the ‘good’ match scenarios. Note that ‘good’ match designs are relatively better designs. Hence, the proportionality observed in this set of scenarios might be a property of an optimal allocation procedure. It is actually observed that the interaction of length and curvature of a stratum is directly proportional to the allocation of observations in a stratum for ‘good’ designs.

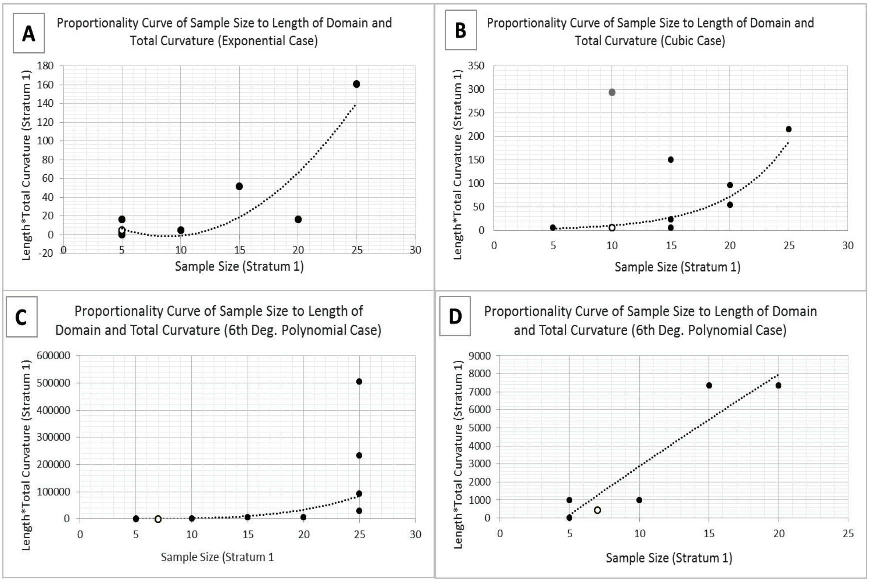
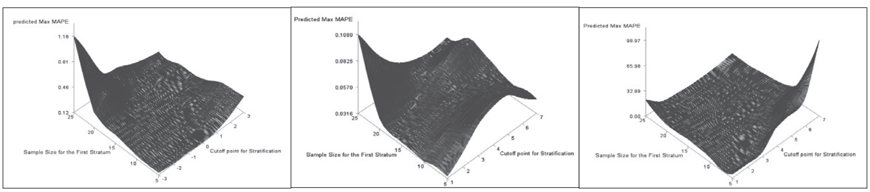


Figure 2. Proportionality Curve between Sample Size, Length and Total Curvature

Figure 2 shows the result of plotting the number of observations allocated to stratum 1 against the product of length and total curvature in stratum 1. Note that discussing patterns in one of the strata would already suffice in describing the general pattern because the experiment domain is fixed, the total number of observations is fixed and there are only two strata. Figure 2-A is for the exponential case. Figure 2-B is for the cubic cases. Notice that in 2-B there is a gray dot, which actually signifies the only point among the 'good' matches that does not conform to the general trend observed. Figure 2-C is for the sixth degree polynomial. Note that the changes in curvature in a sixth degree polynomial would be much steeper than the exponential case after some point. Hence, very large values of total curvature are computed for the high degree polynomial case. The increasing pattern may not be clear in figure 2-C due to the long range of values of curvature, but focusing on a comparable set of points would lead to a clearer increasing pattern, which is showed in figure 2-D. Generally, the proportionality showed in statement (9) was generally exhibited by the 'good' match designs.

Then, the SOAP was determined for each simulated function. If the prior mentioned proportionality is an optimal property, then the resulting SOAP should conform to the patterns showed in figure 2. This makes sense, since the SOAP is the 'best' match design. The strategy implemented in finding the SOAP for each function is through fitting a response surface over the results of the Monte Carlo experiments. Figure 3 shows the estimated response surface of the maximum MAPE over a design space defined by the sample size allocated in the first strata and the cutoff point for stratification.



(Left) Exponential Case, (Middle) Cubic Case, (Right) Polynomial Case

Figure 3. Estimated Response Surfaces for Each Function

To find the Stratified Optimal Allocation Procedure, the pre-image over the X-Y plain of the lowest point of the surface was located. For the exponential case, the lowest point is located at the design using an allocation of 5 observations in the first stratum with a cutoff point of 0.02. For the cubic case, the lowest point is located at the design using an allocation of 10 observations in the first stratum with a cutoff point of 1.00. Finally, the SOAP for the high degree polynomial case is the design allocating 7 observations in the first stratum under a stratum boundary of 1.70. Note that all designs have fixed total observations of 30.

The corresponding allocation and total-length product are computed for the SOAP of each scenario. The computed values were plotted in figure 2 as the hollow dots. Note that the hollow dots, representing the SOAP, conforms very well with the general trend in the proportionality of the sample size allocated to the length and curvature of the stratum. Hence, this is an indication that indeed such proportionality showed in statement (9) is an optimal property for allocation.

Stratified uniform designs

The goal then is to test if optimal stratification can help in improving the optimality of specific designs. For this research, uniform designs are considered. It would be interesting, if such robust designs can still be improved, when additional information on curvature is present. The Stratified Optimal Allocation Procedure was used in implementing Stratified Uniform Designs in simulation. The maximum MAPE of the Uniform Design is compared against the maximum MAPE of the implemented Stratified Uniform Design for each function simulated. The results are showed in Table 3.

Table 3. Maximum MAPE by Design and Data Generating Process

Data Generating Process	Uniform Design	Stratified Uniform Design
$f_1(x) = \exp(x) + 10$	0.12096	0.11576
$f_2(x) = x^3 + 10$	0.040068	0.043998
$f_3(x) = x^6 + x^3 + 10$	2.75452	2.75147

**Note that the allocation procedure applied in StUD is the respective SOAP.*

Based on the simulation results in Table 3, there seemed to be not much of an improvement in the maximum MAPE of the Uniform Design, when it was stratified using the stratified optimal allocation procedure. Nevertheless, the maximum MAPE under the StUD for exponential and high-degree polynomial is slightly better than the maximum MAPE of the uniform design.

This is more likely due to the fact that the curvature is more unbalanced in the exponential and high-degree polynomial case. Hence, the two cases might be more sensitive to stratification. Moreover, two strata may not be sufficient in giving a much more significant improvement in optimality. The variation of the curvature function is highly diminished, when the design is stratified into few intervals and points are allocated based on the total curvature of each stratum. Hence, the number of strata might be too few. Note that the above 100% maximum MAPE is observed for the high-degree polynomial case, which is more likely due to the inadequacy of a small overall sample size for a highly curvaceous function.

In addition, it would be of interest to know if stratification based on the proportionality property observed in the previous section can serve as a more convenient way of designing an optimal stratification. Moreover, SOAP may not be located in complicated or black box scenarios, in which a Uniform Design usually finds its niche. Consider the case wherein a SOAP is not available, stratification may be done by simply making uniformly sized strata and used the proportionality in (9) in solving the number of points to allocate. Table 4 summarizes the resulting designs of the simplified stratification procedure.

The resulting designs in table 4 are not really the exact designs that would result from the proportionality property. Nevertheless, it still follows the proportionality to a certain degree. Note that splines will be used to estimate the nonlinear functions; hence, there is a need to include the boundary points in all the designs. That is to ensure that prediction can be made at any values within the experiment domain. Hence, the allocation procedure are restricted to at least allocating 2 boundary points for each stratum. But for this case, the partitioning of the total curvature is highly unbalanced. Hence, if the proportionality is strictly implemented, 0 points will be allocated in the less curvaceous stratum. Thus, the stratum with a very low proportion of total curvature was forced to have an allocation of 3 points, wherein 2 points are points near/at the boundaries of the stratum.

Table 4. Designs Based on a Simplified Stratification Procedure

Data Generating Process	Cutoff Point	Total Curvature		Sample Size	
		1st Stratum	2nd Stratum	1st Stratum	2nd Stratum
$f_1(x) = \exp(x) + 10$	0	1	147	3	27
$f_2(x) = x^3 + 10$	4	24	24	15	15
$f_3(x) = x^6 + x^3 + 10$	4	7704	115224	3	27

Table 5 compares the simplified allocation procedure to the SOAP and as well to the ordinary uniform design. As expected, the maximum MAPE of the Stratified Uniform Design using a simplified allocation procedure is greater than the SOAP's, since the minimum maximum MAPE is expected in a stratified design using the SOAP in a set of stratified designs with the same total number of observations and number of strata. Nevertheless, the results of the simplified procedure do not deviate that much from the results of the SOAP. Hence, such simplified approach to allocation may be conducted if an optimal allocation procedure is computationally and analytically inconvenient. On the other hand, it is observed that the simplified version would not actually improve the Uniform Design. But similarly, the results from the simplified procedure are of close resemblance of the uniform design. In addition, stratification for this case may have not worked effectively, since there are only two strata being considered.

Table 5. Maximum MAPE by Design and Data Generating Process

Data Generating Process	Uniform Design	Stratified Uniform Design (SOAP)	Stratified Uniform Design (Simplified)
$f_1(x) = \exp(x) + 10$	0.12096	0.11576	0.12547
$f_2(x) = x^3 + 10$	0.040068	0.043998	0.049982
$f_3(x) = x^6 + x^3 + 10$	2.75452	2.75147	2.75402

Hence, 3-strata design with a conveniently chosen allocation procedure is considered in Table 6. The convenient procedure allocates 5 observations in [-5,-1], 10 observations in (-5,2], and 15 observations in (2,5] for the exponential case, while it allocates 5 observations in [0,3], 10 observations in (3,6] and 15 observations in (6,8] for the polynomial cases. There is an increasing allocation of points, since the curvature function is increasing for all functions, while the length is almost evenly distributed. Thus, the allocation is proportional to the total curvature to a certain degree. Based on the results in Table 6, notice that the design with 3 strata has a better maximum MAPE in the exponential case but has the relatively worst maximum MAPE for the polynomial cases. Hence, it can be observed that a number of strata are also important factors in optimizing stratified designs. Strategies involving the number of strata shall be investigated further in future studies. For now, the properties of a good allocation procedure were illustrated using a two-strata design.

Table 6. Maximum MAPE by Design and Data Generating Process

Data Generating Process	Uniform Design	Stratified Uniform Design with 2 strata (SOAP)	Stratified Uniform Design with 2 strata (Simplified)	Stratified Uniform Design with 3 Strata (Simplified)
$f_1(x) = \exp(x) + 10$	0.12096	0.11576	0.12547	0.11815
$f_2(x) = x^3 + 10$	0.040068	0.043998	0.049982	0.10471
$f_3(x) = x^6 + x^3 + 10$	2.75452	2.75147	2.75402	7.47956

4. Conclusions

This study reviewed and proposed a lot of possible curvature related strategies in building optimal experimental designs for fitting a nonlinear model. Note that all the functions considered in this study are monotone increasing. Hence, results from this study may be applied to a fluctuating smooth function by treating the function piecewise such that each part would be a monotone nondecreasing or nonincreasing line.

In summary, an optimal allocation procedure is allocating a number of observations in a stratum that is directly proportional to the stratum's measure and total curvature. The problem simplifies if there is an existing natural stratum definitions. Hence, the problem reduces to allocating more points in more curvaceous subsets.

Uniform designs can be improved by finding the optimal number of strata and the optimal allocation procedure. If it is computationally impossible or inconvenient to solve for a stratified optimal allocation procedure, simpler allocation procedures may be applied, so long as the allocation is somewhat proportional to the measure and total curvature of each stratum. Although a number of strata are not well-highlighted in this study, it was illustrated that the number of strata may pull up or down the maximum MAPE. Thus, further study on the optimal number of strata is recommended.

The results appeal very much to intuition and existing methods. Existing methods of finding optimal designs are commonly derived from the information matrix. The information matrix is related to the hessian matrix, which is a matrix of second derivatives. The hessian matrix is a gradient of curvature functions. Hence, we can view optimization via information matrix as a method of finding a curvature-based solution in finding the optimal design.

REFERENCES

ATKINSON, A. and DONEV, A. N., 1992, *Optimum Experimental Designs*, Oxford University Press.

BAILEY, R.A., 2012, Randomization 50 years after Fisher, Royal Statistical Society Conference, Sept. 2012, Royal Statistical Society. Available at: www.maths.qmul.ac.uk/~rab/telfbeam.pdf

FANG, K.T., 1980, Experimental design by the uniform distribution, *Acta Math. Appl. Sinica* 3:363-372.

FANG, K.T. and LIN, D., 2003, Uniform experimental designs and their applications in industry, *Handbook of Statistics* 22:131-166.

GULLIES, D., 2010, B-Splines, *WIREs Computational Statistics* 2:237-242.

HERTEL, I. and KOHLER M., 2012, Estimation of the optimal design of a nonlinear parametric regression problem via Monte Carlo experiments, *Computational Statistics and Data Analysis* 59:1-12.

LEITHOLD, L., 1996, *The Calculus 7*, Harper Collins College Publishing.

KIEFER J. and WOLFOWITZ, J., 1958, Optimum designs in regression problems, *The Annals of Mathematical Statistics*, 271-293.

MONTGOMERY, D., 1997, *Design and Analysis of Experiments*, John Wiley and Sons, Inc.

MULLER, H.G., 1984, Optimal Designs for Nonparametric Regression, Statistics and Probability Letters, 2:285-290.

MULLER, P. and PARMIGIANI, G., 1995, Optimal design via curve fitting of Monte Carlo experiments, *Journal of the American Statistical Association* 90:1322-1330.

O'HAGAN, A. and KINGMAN, J.F.C., 1978, Curve fitting and optimal design for prediction, *Journal of the Royal Statistical Society*. Series B (Methodological), 40(1):1-42

WAHBA, G., 1990, *Spline Models for Observational Data: Estimating the Smoothing Parameter*, 45-66.

XIE, M.Y. and FANG, K.T., 2000, Admissibility and minimaxity of the uniform design measure in nonparametric regression model, *Journal of Statistical Planning and Inference* 83:101-111.