Uniform Design: Design, Analysis and Applications

Runze Li[1], Dennis K. J. Lin[2] and Yan Chen[3]

[1] Department of Statistics, The Pennsylvania State University, University Park, PA 16802-2111
[2] Department of Management Science and Information Systems, The Pennsylvania State University, University Park, PA 16802
[3] College of Science, Shenyang University of Technology, Shenyang, Laoning 110023, P. R. China

Abstract

A brief introduction is given to explain the fundamental idea of uniform design and measure of uniformity. A new variable selection procedure is proposed for the analysis of data obtained from uniform designs. This new procedure is distinguished from the traditional ones in the way that it simultaneously deletes insignificant variables and estimates the coefficients of significant variables. This procedure possesses an oracle property, which means that it performs as well as if the true model were known in advance. An example is given to illustrate the application of the uniform design and this variable selection procedure.

Key Words: Uniform design; Discrepancy; Penalized least squares; SCAD.

1 Introduction

Design of experiments is a statistical tool that has been widely used in the industry in product design, process design, quality improvement, and productivity improvement. This method is proved to be a powerful means for identifying active effects and screening out unimportant effects. Many challenging questions are arising from practice and various designs are to be proposed to meet practical needs, which makes design of experiments an active area of research in statistics. In this article, we will introduce the **uniform design** proposed by Fang (1980), and discuss a data analysis method for such designs.

A uniform design (UD) is a design in which the design points scatter uniformly on the experimental domain. Such a design has the advantages that: (1) within a small number of experimental

runs, a significant amount of information can be obtained for exploring the relationships between the response and the contributing factors; and (2) it is robust to the underlying model assumption, which means that it performs well even if the form of the regression model is not known. Because of these advantages, the UD has been applied successfully in many industrial and scientific experiments. See Fang (1994) and references therein. Fang, Lin, Winker and Zhang (2000) is excellent review on the theory and application of the UD. Fang and Lin (2001) presents a comprehensive account on the potential applications of UD in industry.

Unlike the fractional factorial design, a UD is usually not orthogonal. This poses a challenge in data analysis. Via a nonconvex penalized least squares approach (Li and Lin, 2002), in this paper we propose an effective screening procedure for identification of active effects from an experiment performed using a UD. This screening procedure possesses an oracle property, a property desirable for model selection, and has a theoretical backup from a statistical consideration. See Fan and Li (2001) and Li and Lin (2002) for a systematic study of variable selection via penalized least squares and penalized likelihood. In this paper, we will use a real example to illustrate in detail how to implement the proposed screening procedure.

This paper is organized as follows. In Section 2, we discuss the background and the fundamental idea of the UD, and discuss issues related to the construction of UD's. Section 3 presents the data obtained in an industrial experiment performed using a uniform design for manufacture of liquid crystal displays (LCD's). A new screening procedure is proposed and discussed in Section 4. Section 5 shows application of the proposed data analysis procedure on the LCD example. Conclusions are drawn in Section 6.

2 Uniform Designs

2.1 Factorial Design and Optimal Design

An objective for designing an experiment is to effectively explore the relationship between a response variable and the relevant contributing factors. Such a relationship may be formulated as the regression model

$$\text{output} = f(\text{inputs}) + \text{random error}, \tag{2.1}$$

where $f(\cdot)$ is the response function. Different functional forms of $f(\cdot)$ may require different designs. A classical approach in design of experiments is based on ANanlysis Of VAriance (ANOVA). A two-way full ANOVA model with Factors A and B, which contains the main effects, the interaction

and the random error, can be represented by

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where Y_{ijk} is the response variable, μ is the grand mean, α_i and β_j are main effects for the two factors, and γ_{ij} is the interaction between the two factors, and ε_{ijk} is the random error. The purpose of an experiment is to estimate all the parameters involved with sufficient accuracy. Design of experiments aims at obtaining sufficient information for this estimation in a reasonable number of runs. When the number of factors increases, the ANOVA model with all possible interactions will become very complex, and the number of parameters to estimate increases exponentially. This requires the number of experimental runs to increase exponentially. A commonly used method to get over this problem is to ignore the high-order interactions. A *fractional factorial design* allows estimation of all main effects and low-order interactions, which can significantly reduce the number of experimental runs.

An *optimal design* is a design which is constructed for optimizing the estimates of a certain parameters in a prespecified statistical model. An example of such a model is the multiple linear regression model

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_s x_s + \varepsilon,$$

where x_i's are input variable, β_i's are unknown parameters of interest, and ε is the random error. Different criteria of optimality may yield different designs even for the same model. Readers may refer to Pukelsheim (1993) for more discussions on optimal designs.

2.2 Space Filling Design

Fractional factorial design and optimal design can be used as a screening mechanism for identifying the potentially active factors in a statistical model. When these designs are used, the design points usually depends on the functional form of $f(\cdot)$. In reality, however, the regression function $f(\cdot)$ is unknown, and for such case the *uniform design* which has a good spread of design points over the entire design space has been proved to be a very suitable candidate for exploring the relationship between the output variable and input variables.

Theoretic Background

The theoretical backup of uniform design rests on the theory of numbers and quasi-Monte Carlo method. Related methods for studying uniform designs are therefore called number theoretic

methods (Fang and Wang, 1994). Suppose that we need to evaluating the integral

$$\int_D h(\mathbf{x})\, d\mathbf{x}, \tag{2.2}$$

where $\mathbf{x}$ is an s-dimensional vector, $h(\mathbf{x})$ is a known function, and D is the domain of integration. When the integral cannot be expressed in a closed form, numerical methods have to be used for evaluation of the integral. The integral in (2.2) is in fact the expected value of $h(\mathbf{x})$ times the volume $|D|$ of D, when $\mathbf{x}$ has a uniform distribution over D. That is,

$$E\{h(\mathbf{x})\} = \int_D h(\mathbf{x})\, d\mathbf{x}/|D|.$$

A numerical method that can be used to evaluate the integral is the Monte Carlo Method to generate n points $\mathcal{P} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ that are independently and identically distributed (i.i.d.) over D with uniform distribution.

$$\bar{h} = \frac{1}{n} \sum_{\mathbf{x} \in \mathcal{P}} h(\mathbf{x})$$

as an estimate of $E\{h(\mathbf{x})\}$.

The Central Limit Theorem (CLT) in statistics shows that the rate of convergence of Monte Carlo method is $n^{-1/2}$, more precisely,

$$|\bar{h} - E\{h(\mathbf{x})\}| = O_P(n^{-1/2}).$$

The rate of convergence $n^{-1/2}$ can be improved by choosing $\mathcal{P}$ in other ways. In this paper, our attention will focus on the case in which D is the unit cube C^s. McKay, Beckman and Conover (1979) proposed a method of generating a set of points $\mathcal{P}$ over C^s, called *Latin hypercube sampling* (LHS), which provides a more accurate estimate of $E\{h(\mathbf{x})\}$ than the average on a set of i.i.d. points with uniform distribution on C^s. Replacing the random sample by a set of deterministic points, quasi-Monte Carlo method provides a set of points $\mathcal{P}$ that estimates $E\{h(\mathbf{x})\}$ in a much more efficient way. Based on the principal of quasi-Monte Carlo method, Fang (1980) and Wang and Fang (1981) proposed the *uniform design*, which allocates experimental points uniformly on the experimental domain. The first successful application of the UD was provided a solution to a challenging experimental design problem with 31 runs and 5 factors each having 31 levels (Fang, 1980).

Both the LHS and the UD are "spacing filling" experimental designs. The fundamental difference between them is that when the experimental domain is continuous, points in the LHS designs are selected at random from the cells, whereas points in the UD are selected from the center of the cells. Thus, the LHS can be viewed as a space filling design in **randomly** uniform manner, while the

UD in **deterministically** uniform manner. Furthermore, an LHS design requires one-dimensional balance of all levels for each factor, but a UD requires one-dimensional balance and s-dimensional uniformity. Thus, these designs are similar in one dimension but can be very different in higher dimensions.

Measure of Uniformity

The rationale for construction of a good UD is based on the following Koksma-Hlawka inequality

$$|\bar{h} - E\{h(\mathbf{x})\}| \leq D(\mathcal{P})V(h),$$

where $V(h)$ is a measure of the variation of h, and $D(\mathcal{P})$ is the discrepancy of $\mathcal{P}$ which is a measure of the uniformity of $\mathcal{P}$. See below for definition of $D(\mathcal{P})$. The quantity $V(h)$ is defined in the sense of Hardy and Krause (Niederreiter, 1992, p.19), which is independent of the design points. The Koksma-Hlawka inequality show that given a bounded $V(h)$, the more uniform a set $\mathcal{P}$ of points distributes over the experimental region C^s, the more accurate $\bar{h}$ is as an estimate of $E\{h(\mathbf{x})\}$. Therefore, given the number of factors and the number of experimental runs, one should choose a set of experimental points with the smallest discrepancy. This is the motivation for the quest of UD's. Since $V(h)$ does not depend on the set $\mathcal{P}$, a UD is robust against change of the function h. Thus, given the number of points and the class of functions with bounded $V(h)$, a UD can be used for obtaining the "most accurate" estimate of $E\{h(\mathbf{x})\}$.

Measures of uniformity are closely related to goodness-of-fit statistics. Let $F_n(\mathbf{x})$ be the empirical distribution function of $\mathcal{P}_n = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$:

$$F_n(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} I(\mathbf{x}_i \leq \mathbf{x}),$$

where $I(\cdot)$ is the indicator function and the last inequality is defined according to componentwise order in R^s. The L_p *discrepancy* $(0 < p < \infty)$ on C^s is defined as

$$D_p(\mathcal{P}_n) = \{\int_{C^s} |F_n(\mathbf{x}) - F(x)|^p \, dx\}^{1/p},$$

where $F(\cdot)$ is the uniform distribution function on C^s. When $p = \infty$, the corresponding L_p discrepancy is

$$D(\mathcal{P}_n) = \sup_{\mathbf{x} \in C^s} |F_n(\mathbf{x}) - F(\mathbf{x})|,$$

which is a commonly used measure of uniformity in the literature of quasi-Monte Carlo method. From the definition of $D_p(\mathcal{P}_n)$, $D_2(\mathcal{P}_n)$ is the Cramer-von Moses statistic for goodness-of-fit test,

and the $D(\mathcal{P}_n)$ is the Kolmogorov-Smirnov goodness-of-fit statistic. Hickernell (1998) proposed other discrepancies that have some desirable properties, and less computational load is required for construction of UD's using those discrepancies. Section 5.1 of Fang and Lin (2001) has a detailed discussion on measures of uniformity.

For a given measure of uniformity, to construct a UD on C^s is to find a set of n points in C^s that has the smallest discrepancy – which is a formidable combinatorial optimization problem even for a moderate value of n. Algorithms such as Threshold Accepting (TA) can be used for solving the problem numerically, but no algorithm will guarantee that a global optimal solution can be obtained. Using the TA algorithm, UD's with different s and n have been constructed and tabulated on the website www.math.hkbu.edu.hk/UniformDesign for convenient use.

Results that reveal the relationship between uniformity and orthogonality have been obtained. Fang and Mukerjee (2000) first established a connection between uniformity and aberration in regular fractions of two-level factorials. This result has been extended to multi-level factorials in Fang and Ma (1999). Such results allow one to construct orthogonal designs from uniform designs.

Uniform designs on constraint experimental domains have also been investigated. An example is UD's constructed in the simplex

$$T_s = \{(x_1, \cdots, x_s) : x_1 + \cdots + x_s = 1, x_j \geq 0, j = 1, \cdots, s\}$$

for experiments with mixtures (Fang and Wang (1990), Wang and Fang (1996)). Readers may refer to Chan (2000) and Cornell (2002) for expositions on experiments with mixtures. Wang and Fang (1996) and Fang and Yang (1999) proposed algorithms to generate sets of points uniformly scattered over the domain

$$T_s\{(a_1, \cdots, a_s), (b_1, \cdots, b_s)\} = \{(x_1, \cdots, x_s) \in T_s, 0 \leq a_j \leq x_j \leq b_j \leq 1, j = 1 \cdots, s\}$$

which is a generalization of T_s, and such a generation is useful in practice and poses challenge questions in constructing a UD over constraint domains.

There are other examples of uniform designs constructed on constraint domains. One is uniform designs constructed for directional data. The experimental domain for directional data is the unit sphere

$$S^d = \{(x_1, \cdots, x_d) : x_1^2 + \cdots + x_d^2 = 1\}.$$

An efficient algorithm to generate UD's over S^d is available in Fang and Wang (1994). Fang and Li (1997) propose an algorithm to generate a UD over a Stiefel manifold

$$\mathcal{O}(d, k) = \{\Gamma \in R^{d \times k} : \Gamma'\Gamma = I_k\},$$

the collection of all possible $d \times k$ column-orthogonal matrices. It is clear that S^d coincides with $\mathcal{O}(d, k)$ with $k = 1$.

3 The LCD Example

An industrial case study on process design for quality improvement in the manufacturer of liquid crystal displays (LCD's) is extracted from Chan and Huang (2002) to illustrate application of the uniform design and analysis of the data obtained.

The experiment has 5 contributing factor, V, F, T, L, S, and the response variable Y is the number of acceptable LCD's produced in a tray of 66 units. Each of the factors V, F, T, S has 5 settings, while the factor L has only 3 settings. The values of the settings of these factors are shown in Table 1.

Table 1: Settings of Factors

	Levels				
Factor	1	2	3	4	5
V	16	18	20	22	24
F	9	11	13	15	17
T	4	5	6	7	8
L	2	11	23	–	–
S	1	1.5	2	2.5	3

It was decided to perform an experiment of 15 runs. The procedure for the design of this experiment started with a $U_{15}(15^5)$ UD extracted from Fang (1994), which has 15 runs, 5 factors each having 15 levels, and is shown in the first 5 columns from the left in Table 2. (In all tables, column numbers are counted from the left.) Using quasi-level technique, adjacent levels in each columns of the $U_{15}(15^5)$ UD were combined to give a $U_{15}(5^4 \times 3^1)$ UD. In the 1st, 2nd, 3rd and 5th columns of Table 2, levels 1,2,3 were combined to give a new level 1, levels 4,5,6 were combined to give a new level 2, ... In the 4th column of Table 2, levels 1,2,3,4,5 were combined to give a new level 1, levels 6,7,8,9,10 were combined to give a new level 2, etc. The $U_{15}(5^4 \times 3^1)$ UD constructed is shown in columns 6 to 10

For each of the 15 runs of the experiment, 5 replicates were performed. Columns 2 to 6 of Table 3 show the actual values of the settings of the factors, and columns 7 to 11 show the numbers of acceptable LCD's from trays of 66. Table 3 shows that the output is strongly related to the input factors. For instance, Run 7 produces very low numbers of accepted LCD's, while Run 1 yields

the best result among the 15 runs. Runs 8 and 12 also produce reasonably good results. Browsing through Table 3, one may guess that the best choice for Factor L is 23, but interactions among the factors may exist. Regression analysis will be performed in Section 5.

Table 2: Uniform Design $U_{15}(15^5)$ and $U_{15}(5^4 \times 3^1)$

V	F	T	L	S	V	F	T	L	S
1	4	7	11	13	1	2	3	3	5
2	8	14	7	11	1	3	5	2	4
3	12	6	3	9	1	4	2	1	3
4	1	13	14	7	2	1	5	3	3
5	5	5	10	5	2	2	2	2	2
6	9	12	6	3	2	3	4	2	1
7	13	4	2	1	3	5	2	1	1
8	2	11	13	14	3	1	4	3	5
9	6	3	9	12	3	2	1	2	4
10	10	10	5	10	4	4	4	1	4
11	14	2	1	8	4	5	1	1	3
12	3	9	12	6	4	1	3	3	2
13	7	1	8	4	5	3	1	2	2
14	11	8	4	2	5	4	3	1	1
15	15	15	15	15	5	5	5	3	5

4 Data Analysis of Uniform Design

When an orthogonal design (such as factorial design) is applied in an experiment for estimation of coefficients in a linear regression model, it is well-known that the coefficients can be independently estimated and tested (Walepole, Myers, Myers and Yi (2002)), which makes screening of variables a relatively straightforward task. When a UD or LHS is used, the screening task will not be so straightforward as these designs are not orthogonal. In this section, we shall use the LCD data to explain how screening can be carried out when UD's are used.

Statistical Formulation

Consider the regression model in (2.1). Suppose that s factors $u_1, ..., u_s$ contribute to the output, but functional form of $f(\cdot)$ is known and is approximated by the regression model

$$y = \beta_1 x_1 + \cdots + \beta_d x_d + \varepsilon,$$

where $x_1, ..., x_d$ are power functions of $u_1, ..., u_s$, $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = \sigma^2$. In order to attenuate modeling bias at the initial stage of modeling, it is desirable to include the linear terms in $u_1, ..., u_s$

Table 3: Design and Responses for the LCD experiment

No.	V	F	T	L	S	$Y^{(1)}$	$Y^{(2)}$	$Y^{(3)}$	$Y^{(4)}$	$Y^{(5)}$
1	16	11	6	23	3	66	66	66	66	66
2	16	13	8	11	2.5	33	36	36	36	51
3	16	15	5	2	2	36	36	45	39	45
4	18	9	8	23	2	45	54	51	60	54
5	18	11	5	11	1.5	45	54	54	51	54
6	18	13	7	11	1	33	36	33	36	36
7	20	17	5	2	1	0	6	6	12	9
8	20	9	7	23	3	54	66	63	66	66
9	20	11	4	11	2.5	30	30	54	45	45
10	22	15	7	2	2.5	33	45	57	54	57
11	22	17	4	2	2	21	30	33	36	33
12	22	9	6	23	1.5	60	60	66	66	66
13	24	13	4	11	1.5	21	45	45	42	45
14	24	15	6	2	1	18	21	27	27	30
15	24	17	8	23	3	30	36	33	42	48

in the model at the initial stage, and incorporate other terms such as quadratic and interaction terms as appropriate, in order to give a full potential model. For example, one may set $x_1 \equiv 1$ to give the constant term, and set $x_2 = u_1$, $x_3 = u_1^2$, $x_4 = u_1 u_2$, etc. It is therefore possible that the number of regression coefficients d in the model will be much larger than s. The insignificant x_j's $(j = 1, ..., d)$ will be finally excluded from the model.

Variable Selection via Penalized Least Squares

Some matrix notation will be defined. Denote by y_i the response from the run of the experiment at which x_j assumes the value x_{ij} $(j = 1, ..., d)$, and let $\mathbf{x}_i = (x_{i1}, \cdots, x_{id})^T$. Define $\mathbf{y} = (y_1, \cdots, y_n)^T$ and $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)^T$. Many variable selection criteria for the linear regression model are based on the penalized least squares

$$Q(\boldsymbol{\beta}) = \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \sum_{j=1}^{d} p_{\lambda_n}(|\beta_j|), \tag{4.1}$$

where $\|\mathbf{a}\| = (\mathbf{a}^T\mathbf{a})^{1/2}$, $p_{\lambda_n}(\cdot)$ is a penalty function, and λ_n is a tuning parameter which controls the complexity of the model and is chosen by a data-driven method such as generalized cross validation (GCV, Craven and Wahba, 1979).

The family of L_p penalty $p_{\lambda_n}(|\theta|) = \lambda_n p^{-1}|\theta|^p$ has been used for the penalized least squares. The L_2 penalty results in a ridge regression estimator. The L_1 penalty yields the LASSO (Tibshirani,

1996), The L_p $(0 < p < 1)$ penalty yields the bridge regression (Frank and Friedman, 1993). The entropy penalty, namely, $p_{\lambda_n}(|\theta|) = \frac{1}{2}\lambda_n^2 I(|\theta| \neq 0)$, where $I(\cdot)$ is an indicator function, corresponds to the L_0 penalty. Note that

$$\sum_{j=1}^{d} I(|\beta_j| \neq 0) = \text{the number of nonzero components of } \boldsymbol{\beta}.$$

Many variable selection criteria can be derived from the penalized least squares (4.1) by setting different values for λ_n. For instance, the AIC (Akaike, 1974) and BIC (Schwarz, 1978) correspond to $\lambda_n = \sqrt{2}(\sigma/\sqrt{n})$ and $\sqrt{\log n}(\sigma/\sqrt{n})$, respectively, although these two criteria were motivated from different principles. The entropy penalty function is discontinuous, and requires searching over all possible subsets for finding the solution of this penalized least squares which makes very expensive in terms of computational cost. Furthermore, it produces unstable models (Breiman, 1996).

To achieve the purpose of variable selection, the penalty function must satisfy certain conditions. Fan and Li (2001) advocate that a good penalty function should have the following three properties. (a) **Sparsity**: The resulting estimator may automatically set the small estimated coefficients to be zero in order to reduce model complexity. In other words, the resulting estimator should have a thresholding rule. (b) **Unbiasedness**: The resulting estimator is nearly unbiased when the true unknown coefficient is large. This will avoid unnecessary modeling bias. (c) **Continuity**: The resulting estimator should be continuous in some sense in order to avoid instability in model prediction.

Unfortunately, from a mathematical consideration, the L_p penalty $(0 < p \leq \infty)$ does not posses the above three properties. Fan and Li (2001) suggest to use the smoothly clipped absolute deviation (SCAD) penalty, which is given by

$$p_\lambda(\theta) = \lambda|\theta| I(|\theta| \leq \lambda) + \frac{2a\lambda|\theta| - \theta^2}{2(a-1)} I(\lambda \leq |\theta| < a\lambda) + \frac{a^2\lambda^2}{2(a-1)} I(|\theta| \geq a\lambda) \text{ for some } a > 2, \quad (4.2)$$

which satisfies all the above three conditions mathematically. In what follows, SCAD will mean any procedure derived using the SCAD penalty. The SCAD involves two unknown parameters λ and a, where λ is a regularization parameter chosen by the generalized cross validation (GCV), a data-driven method. Fan and Li (2001) suggested using $a = 3.7$ based on a Bayesian argument, and this value will be used throughout in this paper.

Oracle Property

Suppose that in the following model, the errors $\boldsymbol{\varepsilon}$ are i.i.d.:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

Suppose that $\mathbf{X}$ is partitioned as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. Without loss of generality, assume that all components in $\mathbf{X}_1$ are active, while those in $\mathbf{X}_2$ are not active. An ideal estimator is the oracle estimator:

$$\widehat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{y}, \qquad \text{and} \qquad \widehat{\boldsymbol{\beta}}_2 = \mathbf{0},$$

which correctly specifies the true model and provides estimates of the regression coefficient associated with $\mathbf{X}_1$. This is a desired property in variable selection. With a proper selection of λ_n, Li and Lin (2002) showed the SCAD satisfies this oracle property asymptotically, which makes it an ideal variable selector in many cases.

An Algorithm

Solving a nonconvex penalized least squares problem is a challenging task, simply because the target function is nonconvex and the problem could be high-dimensional. The penalized least squares function of the SCAD do not have the second derivative at some points, due to the form of the penalty function. In order to apply the Newton-Raphson algorithm to solve the penalized least squares problem for the SCAD, the following local approximation method will be used.

When $\beta_j^{(0)}$ is not very close to 0, given an initial value $\boldsymbol{\beta}^{(0)}$ that is close to the true value of $\boldsymbol{\beta}$, approximate the penalty $p_\lambda(|\beta_j|)$ locally by the quadratic function as follows:

$$p_{\lambda_n}(|\beta_j|) \approx p_{\lambda_n}(|\beta_j^{(0)}|) + \frac{1}{2}\{p'_{\lambda_n}(|\beta_j^{(0)}|)/|\beta_j^{(0)}|\}(\beta_j^2 - \beta_j^{(0)2}), \tag{4.3}$$

otherwise, delete the corresponding term from the model. With this local quadratic approximation, the solution for the penalized least squares problem can be found by iteratively computing the following ridge regression with an initial value $\boldsymbol{\beta}^{(0)}$:

$$\boldsymbol{\beta}^{(1)} = \{\mathbf{X}^T\mathbf{X} + n\Sigma_\lambda(\boldsymbol{\beta}^{(0)})\}^{-1}\mathbf{X}^T\mathbf{y}, \tag{4.4}$$

where

$$\Sigma_\lambda(\boldsymbol{\beta}^{(0)}) = \text{diag}\{p'_{\lambda_1}(|\beta_1^{(0)}|)/|\beta_1^{(0)}|, \cdots, p'_{\lambda_d}(|\beta_d^{(0)}|)/|\beta_d^{(0)}|\}.$$

Choice of tuning parameter λ_n

From the iterative ridge regression (4.4), the fitted value of $\mathbf{y}$ is

$$\widehat{\mathbf{y}} = \mathbf{X}\{\mathbf{X}^T\mathbf{X} + n\Sigma_{\lambda_n}(\widehat{\boldsymbol{\beta}})\}^{-1}\mathbf{X}^T\mathbf{y}.$$

Although $\widehat{\mathbf{y}}$ is not a linear in term of $\mathbf{y}$ since $\widehat{\boldsymbol{\beta}}$ may be associated with $\mathbf{y}$,

$$\mathbf{P}_{\mathbf{X}}\{\widehat{\boldsymbol{\beta}}(\lambda)\} = \mathbf{X}\{\mathbf{X}^T\mathbf{X} + n\Sigma_\lambda(\widehat{\boldsymbol{\beta}})\}^{-1}\mathbf{X}^T$$

can be regarded as a projection matrix. Thus, a given λ,

$$e(\lambda) = \mathrm{tr}[\mathbf{P}_{\mathbf{X}}\{\widehat{\boldsymbol{\beta}}(\lambda)\}]$$

can be regarded as effective number of parameters. The generalized cross-validation statistics can be defined by

$$\mathrm{GCV}(\lambda) = \frac{\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\lambda)\|^2}{\{n - e(\lambda)\}^2} \tag{4.5}$$

and λ_n is estimated by $\widehat{\lambda}_n = \mathrm{argmin}_\lambda\{\mathrm{GCV}(\lambda)\}$.

A Variable Selection Procedure

For a given value of λ_n and an initial value of $\boldsymbol{\beta}$, we iteratively compute the ridge regression (4.4) with updating local quadratic approximation (4.3) at each step during the iteration. This can be easily implemented in many statistical packages. Some components of the resulting estimate will be exactly 0, and such components correspond to the coefficient of inactive effects. In other words, nonzero components of the resulting estimate are correspondent to the active effects in the SSD.

5 Analysis of the LCD Example

We now demonstrate how to implement the SCAD screening active effects for the LCD data. The means and standard deviations of the five controllable variables are listed below.

	V	F	T	L	S
Mean	20	13	6	12	2
SD	2.8475	2.8475	1.4237	8.6603	0.7119

To avoid numerical unstability, all controllable variables are standardized. The standardized variables are denoted by V_z, F_z, T_z, L_z, S_z.

Using the GCV to estimate the tuning parameter λ, we obtain $\widehat{\lambda}$=0.6099. With this value of λ, the estimated coefficients of the selected model, their standard error and the corresponding P-value are shown in Table 4. From Table 4, all selected terms are significant at the 0.05 significance level, and the resulting model for $f(\cdot)$ is

$$\widehat{f} = 52.5231 - 9.6258F_z + 9.9433S_z - 2.1342L_z - 9.5534T_z^2 + 7.9921T_z * L_z - 2.6838V_z * L_z - 2.5289V_z * T_z.$$

The last equation shows that in order to have large $\widehat{f}$, the controllable variable F should be set at small values. This is consistent with the outputs of Runs 1, 8 and 12. Factor S should be set at high value. This is also consistent with the outputs of Runs 1 and 8. The final model shows that

there exists a strong two-order interactions among V, T and L. To find an optimal setting for the controllable variables V, T and L, we have to maximize $\widehat{f}$ with respect to V, T and L over the experimental domain, which will not be pursued here.

Note that the standard deviations of outputs among the 5 replicates for each run vary quite significantly. This indicates that the i.i.d. assumption on the random error might not be valid. Testing of the homogeneous assumption and further statistical analysis will be dealt elsewhere and will not be presented here.

Table 4: Estimates, Standard Errors and P-value

X-variable	Estimate	Standard Error	P-value
Intercept	52.5231	1.5146	0.0000
F_z	-9.6258	0.8790	0.0000
S_z	9.9433	1.0042	0.0000
L_z	-2.1342	0.8296	0.0123
T_z^2	-9.5534	1.3683	0.0000
$T_z * L_z$	7.9921	1.0267	0.0000
$V_z * L_z$	-2.6838	0.9715	0.0074
$V_z * T_z$	-2.5289	1.1491	0.0312

6 Conclusion

In this paper, the background, the fundamental idea and the construction of the uniform design are introduced. A screening procedure, SCAD, for regression analysis is explained. A real example is used to illustrate application of the uniform design in solving an industrial problem, and analysis of data using SCAD is presented. This paper serves a reference for practitioners who wish to apply the uniform design to solve real industrial problems.

Acknowledgement: The authors would like to thank Professor Philip L. Y. Chan and the referees for their constructive comments that substantially improve an earlier draft. Li's research was supported by a NSF grant DMS-0102505. Lin's research is partially supported by a National Security Agency Grant MDA 904-02-1-0054.

References

H. Akaike (1974). A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, **19**, 716–723.

L. Breiman (1996). Heuristics of instability and stabilization in model selection, *Ann. Statist.*, **24**, 2350-2383.

L. Y. Chan (2000). Optimal design for experiments with mixtures: a survey. *Communications in Statistics - Theory and Methods*, **29**, 2281–2312.

L. Y. Chan and G. Q. Huang (2002). Application of uniform design in quality improvement in the manufacture of liquid crystal displays, *Proceeding of the Eighth ISSAT International Conference on Reliability and Quality in Design*, edited by H. Pham and M. W. Lu, 245-249.

J. A. Cornell (2002). *Experiments with Mixtures, Designs, Models and the Analysis of Mixture Data* 3rd Edition, Wiley, New York.

P. Craven and G. Wahba (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377–403.

J. Fan and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association.* **96**, 1348-1360.

K. T. Fang (1980). The uniform design: application of number-theoretic methods in experimental design. *Acta Math. Appl. Sinica*, **3**, 363-372.

K. T. Fang (1994).*Uniform Design and Uniform Design Tables.* Science Press, Beijing.

Fang, K. T. and Li, R. (1997). Some methods for generating both NT net and the uniform distribution on a Stiefel manifold and their applications. *Computational Statistics and Data Analysis*, **24**, 29-46.

K. T. Fang and D. K. J. Lin (2001). Uniform experimental design and its application in industry. *Handbook of Statistics*, eds C. R. Rao.

K. T. Fang and C. X. Ma (1999). Relationship between uniformity and aberration, constructions of uniform designs in regrulars 3^{s-1}, *Technical Report*, **Math-243**, Hong Kong Baptist University.

K. T. Fang and R. Mukerjee (2000). A connection between uniformity and aberration in regular fractions of two-level factorials, *Biometrika*, **87**, 193-198.

K. T. Fang and Y. Wang (1994). *Applications of Number Theoretic Methods in Statistics*, Chapman and Hall, London.

K. T. Fang and Z. H. Yang (1999). On uniform design of experiments with restricted mixtures and generation of uniform distribution on some domains, *Statist. & Probab. Letters*, **46**, 113-120.

I. E. Frank and J. H. Friedman (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109-148.

F. J. Hickernell (1998). A generalized discrepancy and quadrature error bound. *Math. Comp.*, **67**, 299-322.

R. Li and D. K. J. Lin (2002). Data analysis in supersatured designs, *Statistics and Probability Letters*. **59**, 135-144.

C. L. Mallows (1973). Some comments on C_p. *Technometrics*, **15**, 661–675.

M. D. Mckay, R. J. Beckman, and W. J. Conover (1979). A comparison of three methods for selecting values of input variables in analysis of output from a computer code, *Technometrics*, **21**, 239-245.

H. Niederreiter (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia.

F. Pukelsheim (1993). *Optimal Design of Experiments*, Wiley, New York.

G. Schwarz (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.

R. Tibshirani (1996). Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society, B*, **58**, 267-288.

R.E. Walpole, R.H. Myers, S.L Myers, K. Ye (2002). *Probability and Statistics for Engineers and Scientists*, Prentice Hall, Upper Saddle River, NJ.

Y. Wang and K. T. Fang (1981). A note on uniform distribution and experiment design. *KeXue TongBao*, **26**, 485-489.

Y. Wang and K. T. Fang (1990). Number theoretic methods in applied statistics (II), *Chinese Ann. Math. Ser. B*, **11**, 264-275.

Y. Wang and K. T. Fang (1996). Uniform design of experiments with mixtures, *Science in China (Series A)*, **39**, 264-275.

About the Authors

R. Li is an assistant professor at Department of Statistics, the Pennsylvania State University at University Park.

D. K. J. Lin is a professor at Department of Management Science and Information Systems, the Pennsylvania State University at University Park.

Y. Chen is a Lecturer at College of Science, Shenyang University of Technology, Shenyang, Laoning 110023