

## Analysis Methods for Supersaturated Design: Some Comparisons

Runze Li<sup>1</sup> and Dennis K. J. Lin<sup>2</sup>

*The Pennsylvania State University*

*Abstract:* Supersaturated designs are very cost-effective with respect to the number of runs and as such are highly desirable in many preliminary studies in industrial experimentation. Variable selection plays an important role in analyzing data from the supersaturated designs. Traditional approaches, such as the best subset variable selection and stepwise regression, may not be appropriate in this situation. In this paper, we introduce a variable selection procedure to screen active effects in the SSDs via nonconvex penalized least squares approach. Empirical comparison with Bayesian variable selection approaches is conducted. Our simulation shows that the nonconvex penalized least squares method compares very favorably with the Bayesian variable selection approach proposed in Beattie, Fong and Lin (2001).

*Key words:* Bayesian variable selection, penalized least squares, SCAD.

### 1. Introduction

Supersaturated designs (SSD) are useful in many preliminary studies in industry in the presence of a large number of potentially relevant factors, of which actual active effects are believed to be sparse. The SSD was introduced by Satterthwaite (1959) and has again received increasing attention since the appearance of Lin (1993), evidenced by the growth of the literature. Lin (1993) provided a new class of supersaturated designs based on half-fractions of Hadamard matrices. Wu (1993) augmented Hadamard

matrices by adding interaction columns. Algorithms for constructing SSD have been studied by many authors, for instance, Lin (1991, 1995), Nguyen (1996), Li and Wu (1997), Tang and Wu (1997), Yamada and Lin (1997), Deng, Lin and Wang (1999) and Fang, Lin and Ma (2000). Since the number of experiments is less than the number of candidate factors in SSD, the analysis is very challenging.

To find the sparse active effects, variable selection becomes fundamental in the analysis stage of such screening experiments. While stepwise variable selection may not be appropriate (see, for example, Westfall, Young and Lin, 1998), some traditional approaches, such as the best subset variable selection, are known to be infeasible. Beattie, Fong and Lin (2001) discuss how to implement Bayesian variable selection for data analysis of SSDs. They first employed the stochastic search variable selection (SSVS) approach, proposed by George and McCulloch (1993) then applied the intrinsic Bayesian factor (IBF), proposed by Berger and Pericchi (1996), to select active effects from the remaining factors in the first stage. In this paper, we introduce an alternative approach via nonconvex penalized least squares. The theory behind our approach is different from the Bayesian variable selection approaches.

Fan and Li (2001) proposed a class of variable selection procedures by nonconcave penalized likelihood approaches. Extending the idea of the nonconcave penalized likelihood, Li and Lin (2002) suggested the use of nonconvex penalized least squares to screen active effects for supersaturated design. From their comparison, the penalized least squares performs much better than than the stepwise procedures using various variable selection criteria. In this paper, the performance of the penalized least squares is compared with that of Bayesian variable selection procedures. Our simulation shows that the procedure, proposed by Li and Lin (2002), compares very favorably with Bayesian variable selection approaches.

This paper is organized as follows. In Section 2, we introduce nonconvex penalized least squares, and discuss how to implement the nonconvex penalized least squares approach in practice. Section 3 proposes a new approach to screening active effects via penalized least squares. A real example and empirical comparison are given in Section 4. Final conclusions are given in Section 5.

## 2. Nonconvex penalized least squares

Suppose that observations  $y_1, \dots, y_n$  are an independent sample from the linear regression model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \quad (2.1)$$

with  $E(\varepsilon_i) = 0$ , and  $\text{var}(\varepsilon_i) = \sigma^2$ , where  $\mathbf{x}_i$  corresponds to a fixed experiment design. Denote  $\mathbf{y} = (y_1, \dots, y_n)^T$ , and  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ . Following Fan and Li (2001), define penalized least squares as

$$Q(\boldsymbol{\beta}) \equiv \frac{1}{2n} \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \sum_{j=1}^d p_\lambda(|\beta_j|), \quad (2.2)$$

where  $\|\cdot\|$  is the  $L_2$ -norm,  $d$  is the dimension of  $\boldsymbol{\beta}$ ,  $p_\lambda(\cdot)$  is a penalty function, and  $\lambda$  is a tuning parameter which controls model complexity. Note that  $\lambda$  may depend on the sample size  $n$  and can be selected by some data-driven methods, such as generalized cross validation (GCV, Craven and Wahba, 1979).

Some existing variable selection criteria are closely related to the penalized least squares criterion (2.2). Take the penalty function to be the entropy penalty, namely,  $p_\lambda(|\theta|) = \frac{1}{2}\lambda^2 I(|\theta| \neq 0)$ , which is also referred to as  $L_0$ -penalty in the literature, where  $I(\cdot)$  is an indicator function. Note that the dimension or the size of a model equals to the number of nonzero regression coefficients in the model. This actually equals  $\sum_j I(|\beta_j| \neq 0)$ . In other words, the PLS (2.2) with the entropy penalty can be rewritten as

$$\frac{1}{2n} \sum_{i=1}^n (y_i - \mathbf{x}_i \boldsymbol{\beta})^2 + \frac{1}{2} \lambda_n^2 |M|, \quad (2.3)$$

where  $|M| = \sum_j I(|\beta_j| \neq 0)$  is the size of the underlying candidate model. Hence, many popular variable selection criteria can be derived from the PLS (2.3) by choosing different values of  $\lambda_n$ . For instance, the AIC (Akaike, 1974) and BIC (Schwarz, 1978) correspond to  $\lambda_n = 2(\sigma/\sqrt{n})$  and  $\sqrt{2 \log n}(\sigma/\sqrt{n})$ , respectively, although these criteria were motivated from different principles. The  $L_1$  penalty  $p_\lambda(|\beta|) = \lambda|\beta|$  results in LASSO (Tibshirani, 1996),

and the  $L_q$  penalty in general leads to a bridge regression (see Frank and Friedman, 1993). Furthermore, the  $L_q$  penalty can provide the resulting solution a Bayesian interpretation.

To achieve the purpose of variable selection, some conditions on  $p_\lambda(|\cdot|)$  are needed. Fan and Li (2001) suggested the use of the smoothly clipped absolute deviation (SCAD), whose first order derivative is defined by

$$p'_\lambda(\beta) = \lambda \left\{ I(\beta \leq \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)\lambda} I(\beta > \lambda) \right\}, \text{ for some } a > 2 \text{ and } \beta > 0,$$

with  $p_\lambda(0) = 0$ . For simplicity of presentation, we will use the name of SCAD for all procedures using the SCAD penalty. The SCAD involves two unknown parameters  $\lambda$  and  $a$ . Fan and Li (2001) suggested using  $a = 3.7$  from a Bayesian point of view. They found via Monte Carlo simulation that the performance of SCAD is not sensitive to the choice of  $a$  provided that  $a \in (2, 10)$ . They also found that the empirical performances of the SCAD with  $a = 3.7$  is as good as those of the SCAD with the value of  $a$  chosen by the GCV. Hence, this value will be used throughout the entire paper.

It is challenging to find solutions of the nonconvex penalized least squares with the SCAD penalty, because the target function is nonconvex and could be high-dimensional. Furthermore, the SCAD penalty function may not have the second order derivative at some points. In order to apply the Newton Raphson algorithm to the penalized least squares, Fan and Li (2001) suggested to locally approximate the SCAD penalty function by a quadratic function as follows.

Given an initial value  $\beta^{(0)}$  that is close to the true value of  $\beta$ , when  $\beta_j^{(0)}$  is not very close to 0, the penalty  $p_\lambda(|\beta_j|)$  can be locally approximated by the quadratic function as

$$[p_\lambda(|\beta_j|)]' = p'_\lambda(|\beta_j|)\text{sgn}(\beta_j) \approx \{p'_\lambda(|\beta_j^{(0)}|)/|\beta_j^{(0)}|\}\beta_j, \quad (2.4)$$

otherwise, set  $\hat{\beta}_j = 0$ . With the local quadratic approximation, the solution for the penalized least squares can be found by iteratively computing the following ridge regression with an initial value  $\beta^{(0)}$ :

$$\beta^{(1)} = \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta^{(0)})\}^{-1} \mathbf{X}^T \mathbf{y}, \quad (2.5)$$

where  $\Sigma_\lambda(\boldsymbol{\beta}^{(0)}) = \text{diag}\{p'_\lambda(|\beta_1^{(0)}|)/|\beta_1^{(0)}|, \dots, p'_\lambda(|\beta_d^{(0)}|)/|\beta_d^{(0)}|\}$ . A standard error formula can be derived from the iterative ridge regression algorithm (2.5); specifically,

$$\widehat{\text{cov}}(\hat{\boldsymbol{\beta}}) = \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\hat{\boldsymbol{\beta}})\}^{-1} \mathbf{X}^T \mathbf{X} \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\hat{\boldsymbol{\beta}})\}^{-1}.$$

This formula only applies for non-vanished components of  $\hat{\boldsymbol{\beta}}$ .

### 3. The proposed procedure

We now propose a screening active effect procedure for the SSD. For a given value of  $\lambda$  and an initial value of  $\boldsymbol{\beta}$ , we iteratively compute the ridge regression (2.5) with updating local quadratic approximation (2.4) at each iteration. This can be easily implemented in many statistical packages. Some components of the resulting estimate will exactly be 0. These components correspond to the coefficient of inactive effects. In other words, nonzero components of the resulting estimate correspond to the active effects in the SSD. Following Li and Lin (2002), we choose  $\lambda$  by minimizing the corresponding GCV scores.

Compared with stepwise regression, the proposed procedure is actually an estimation procedure rather than a variable selection procedure. Stochastic errors may be ignored during the course of selecting variables. Compared with the best subset variable selection, our procedure significantly reduces the computational burden. Our procedure simultaneously excludes inactive factors and estimates the regression coefficients of active effects. Assume observations are taken from the model  $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$ . Without loss of generality, further assume that all components of the first portion ( $\mathbf{X}_1$ ) are active, while the second portion ( $\mathbf{X}_2$ ) is not active. An ideal estimator is the oracle estimator:

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

The oracle estimator correctly specifies the true model and efficiently estimates the regression coefficients of the significant part. This is a desired property in variable selection. It has been shown in Li and Lin (2002) that the SCAD yields an estimate possessing the oracle property in asymptotic sense.

#### 4. Numerical comparison and example

The goal of this section is to compare the performance of the penalized least squares approach with the SCAD penalty and the two Bayesian variable selection procedures, SSVS, proposed by George and McCulloch (1993) and extended for SSD by Chipman, Hamada and Wu (1993), denoted by CHW for short, and the two stage procedure (SSVS/IBF) by Beattie, Fong and Lin (2001), denoted by BFL. We assess the performance of these variable selection procedures in terms of their ability to identify the true model, i.e., their ability of identifying the smallest active effect and the size of selected model. In the following examples, we set the level of significance to be 0.1 for both F-enter and F-remove in the stepwise variable selection procedure for choosing an initial value for the penalized least squares approaches. All simulations are conducted using MATLAB codes.

**Example 1.** In this example, we apply SCAD to the supersaturated design demonstrated first in Lin (1993) and listed in Table 1. Table 2 depicts the five models with the highest posterior probability obtained in CHW, using the SSVS approach. From Table 2, the first two models have very close posterior probability. A question arising here is whether Factor 10 is active. In fact, this is a very challenging example for various approaches. Most of the existing approaches have difficulty in determining whether or not Factor 10 is active. See BFL for more detailed discussions.

We now apply SCAD for this data set. The tuning parameter  $\lambda$  is selected by GCV, resulting  $\hat{\lambda}=6.5673$ . The final selected models, including estimated coefficients and their standard errors, are shown in Table 3. The SCAD identifies  $X_{15}$ ,  $X_{12}$ ,  $X_{20}$  and  $X_4$  as active effects. This is consistent with the model concluded in Williams (1968), in which 28 rather than 14 experiments are included in the analysis.

**Example 2.** The design matrix  $\mathbf{X}$  is displayed in Table 1. We generate data from the linear model

$$Y = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon,$$

where the random error  $\varepsilon$  is  $N(0, 1)$ . We consider three cases for  $\boldsymbol{\beta}$ :

Table 1: Supersaturated Design: Half Fraction of Willaims (1968) Data

	Run													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	1	1	1	-1	-1	-1	-1	-1	1	-1	1	1	-1
2	1	-1	1	1	-1	-1	-1	1	-1	1	1	-1	1	-1
3	1	-1	-1	-1	1	1	-1	1	-1	1	-1	-1	1	1
4	-1	-1	1	1	1	1	-1	-1	-1	1	1	-1	1	-1
5	-1	-1	1	-1	1	1	1	-1	-1	-1	1	1	1	-1
6	-1	-1	-1	1	1	1	-1	1	1	1	-1	1	-1	-1
7	1	1	-1	-1	-1	1	-1	-1	1	1	-1	1	1	-1
8	1	1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	-1
9	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	1	-1
10	1	-1	1	1	-1	1	1	1	-1	-1	-1	1	-1	-1
11	1	-1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	1
12	-1	-1	1	-1	-1	-1	1	-1	1	1	-1	1	1	1
13	1	1	1	1	1	-1	1	-1	-1	-1	-1	1	-1	-1
14	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1
15	-1	1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1
17	1	-1	1	-1	1	1	1	-1	1	1	-1	-1	-1	-1
18	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1
19	-1	-1	-1	1	-1	1	1	1	-1	1	1	1	-1	-1
20	1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	1	1
21	-1	1	-1	-1	-1	1	1	1	-1	1	-1	1	1	-1
22	-1	1	1	-1	1	1	-1	1	-1	-1	-1	1	-1	1
23	-1	-1	1	-1	1	-1	-1	1	1	-1	1	1	1	-1
24	1	-1	-1	1	-1	1	-1	1	1	1	1	1	-1	-1
	133	62	45	52	56	47	88	193	32	53	276	145	130	127

The first column is for the factors, the last row is the responses

Table 2: Posterior Model Probabilities in Example 1

Model	Prob.	$R^2$
4 12 15 20	0.0266	0.955
4 10 12 15 20	0.0259	0.973
4 10 11 12 15 20	0.0158	0.987
4 12 15 20 21	0.1200	0.969
4 11 12 15 20	0.0082	0.966

Table 3: The Final model selected by SCAD in Example 1

Factor	Intercept	X4	X12	X15	X20
$\hat{\beta}$	102.7857	20.1084	-25.3946	-69.5738	-28.7967
$SE(\hat{\beta})$	4.5377	4.6965	4.6557	5.1075	4.6965

Case I: One active factor,  $\beta_1 = 10$  and other components of  $\beta$  are equal to zero;

Case II: Three active factors,  $\beta_1 = -15$ ,  $\beta_5 = 8$ ,  $\beta_9 = -2$ , and other components of  $\beta$  equal 0;

Case III: Five active factors,  $\beta_1 = -15$ ,  $\beta_5 = 12$ ,  $\beta_9 = -8$ ,  $\beta_{13} = 6$  and  $\beta_{17} = -2$ , and other components of  $\beta$  equal 0.

Cases I and II were used in BFL, while Case III was studied by Beattie (1999). Simulation results for SCAD based on 1000 replicates are summarized in Table 4. For convenience, we also present simulation results of Bayesian variable selection approaches obtained in BFL and Beattie (1999). In Table 4, SSVS stands for the Bayesian variable selection approach, proposed by George and McCulloch (1993) and extended to the context of screening variables by CHW. See BFL for the meaning of the hyper-parameters, shown in parentheses. SSVS/IBF refers to the two stages of Bayesian model selection strategy proposed by BFL, retaining good features of the SSVS approach and IBF approach.

There is only a single strong effect among the factors in Case I. All three



approaches perform well in terms of the rate of identifying the single active effect. SSVS and SSVS/IBF tend to incorrectly identify inactive effects as active effects. This leads to a low rate of identifying the true model. In this case, SCAD performs the best.

The factors in Case II contain a very strong effect, a medium effect and a relatively small effect. This is challenging for various existing variable selection procedures. The two Bayesian variable selection approaches frequently fail to identify the smallest effect  $X_9$  as active effect and suffer very low identification rate of the true model. The performance of the SCAD is similar to that of Case I, and outperforms the other two approaches.

In Case III, there are 5 active effects, two factors have very strong effects, two have moderate effects, and one has a small effect. In this case, the SSVS/IBF cannot take its advantage in identifying the smallest active effect, compared with SSVS. Again, the SCAD performs the best.

Table 4: Summary of Simulation Results in Example 2

Method	True Model Identified Rate	Smallest Effect Identified Rate	Avg. Size	
			Median	Mean
Case I: One Active Effects				
SSVS(1/10,500)	40.5%	99%	2	3.1
SSVS(1/10,500)/IBF	61%	98%	1	2.5
SCAD	75.6%	100%	1	1.7
Case II: Three Active Effects				
SSVS(1/10, 500)	8.6%	30%	3	4.7
SSVS(1/10,500)/IBF	8.0%	28%	3	4.2
SCAD	74.7%	98.5%	3	3.3
Case III: Five Active Effects				
SSVS(1/10, 500)	36.4%	84%	6	8.0
SSVS(1/10,500)/IBF	40.7%	75%	5	5.6
SCAD	69.7%	99.4%	5	5.4

## 5. Conclusion

In this paper, we have introduced the SCAD for analyzing data in SSD. The popular William's data were used for illustration. Although we don't know the true answer from William's data, the results from the proposed method is rather consistent with other approaches. Empirical performance of the nonconvex penalized least squares based on heavy simulations are also compared with that of Bayesian variable selection procedures for the SSD. The comparison shows that the SCAD performs better than the proposals in CHW and BFL.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, **19**, 716-723.
- Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations (with discussions). *Journal of American Statistical Association*, **96**, 939-967.
- Beattie, S. D. Fong, D. K. H. and Lin, D. K. J. (2002). A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics*, **44**, 55-63.
- Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction, *Journal of American Statistical Association*, **91**, 109-122.
- Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, **37**, 373-384.
- Chipman, H., Hamada, M. and Wu, C. F. J. (1997). A Bayesian variable selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, **39**, 372-381.
- Craven, P. and Wahba, G. (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.
- Deng, L. Y., Lin, D. K. J. and Wang, J. N. (1999). On resolution rank criterion for supersaturated design. *Statistica Sinica*, **9**, 605-610.

- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association*, **96**, 1348-1360.
- Fang, K. T., Lin, D. J. K. and Ma, C. X. (2000). On the construction of multi-level supersaturated design. *Journal of Statistical Planning and Inference*, **86**, 239-252.
- Frank, I.E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109-148.
- George, E. I. and McMulloch, R. E. (1993). Variable selection via Gibbs sampling, *Journal of American Statistical Association*, **91**, 883-904.
- Li, R. and Lin, D. K. J. (2002). Data Analysis of Supersaturated Design, *Statistics and Probability Letters*, **59**, 135-144.
- Li, W. W. and Wu, C. F.J. (1997). Columnwise-pairwise algorithms with applications to construction of supersaturated designs. *Technometrics*, **39**, 171-179.
- Lin, D. K. J. (1991). Systematic supersaturated designs. Working Paper, No. 264, Department of Statistics, University of Tennessee.
- Lin, D. K. J. (1993). A new class of supersaturated designs, *Technometrics*, **35**, 28-31.
- Lin, D. K. J. (1995). Generating system supersaturated designs, *Technometrics*, **37**, 213-225
- Nguyen, N. K. (1996). An algorithmic approach to constructing supersaturated designs. *Technometrics*, **38**, 69-73.
- Satterthwaite, F. (1959). Random balance experimentation (with discussion), *Technometrics*, **1**, 111-137.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461-464.
- Tang, B. and Wu, C. F. J. (1997). A method for constructing supersaturated designs and its  $Es^2$  optimality. *Canadian J. Statist.*, **25**, 191-201.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, B*, **58**, 267-288.
- Westfall, P. H., Young, S. S. and Lin, D.K.J. (1998). Forward selection error control in analysis of supersaturated designs. *Statistica Sinica*, **8**, 101-117.
- Wu, C. F. J. (1993). Construction of supersaturated designs through partially aliased interactions. *Biometrika*, **80**, 661-669.
- Yamada, S. and Lin, D. K. J. (1997). Supersaturated designs including an orthogonal base, *Canadian J. Statist.*, **25**, 203-213.

Received August 30, 2002; accepted October 20, 2002

Runze Li  
Department of Statistics  
The Pennsylvania State University  
University Park, PA 16802-2111

Dennis K. J. Lin  
Department of Management Science & Information Systems  
The Pennsylvania State University  
University Park, PA 16802