An Efficient Variable Selection Approach for Analyzing Designed Experiments

Ming Yuan[a], V. Roshan Joseph[a], and Yi Lin[b]

[a] School of Industrial and Systems Engineering, Georgia Institute of Technology,
755 Ferst Drive NW, Atlanta, GA 30332-0205

[b] Department of Statistics, University of Wisconsin,
1300 University Avenue, Madison, WI 53706

(November 23, 2005)

Abstract

The analysis of experiments where a large number of potential variables are examined is driven by the principles of effect sparsity, effect hierarchy, and effect heredity. We propose an efficient variable selection strategy to specifically address the unique challenges faced by such analysis. The proposed methods are natural extensions of a general-purpose variable selection algorithm, LARS (Efron et al., 2004). They are very fast to compute and can find sparse models that better satisfy the goals of experiments. Simulations and real examples are used to illustrate the wide applicability of the proposed methods.

1 Introduction

We consider the analysis of experiments where a large number of potential variables are examined. In most practical situations, however, only a relatively small number of observations are affordable. Because of their run size economy and flexibility, fractional factorial designs are widely used in such experiments. But the analysis of such designs is complicated due to the aliasing of effects. The analysis is driven by the principles of effect sparsity, effect hierarchy, and effect heredity (Wu and Hamada 2000). The effect sparsity principle states that only a small number of effects are significant. The effect hierarchy principle states that lower order effects are more important than higher order effects. Using this principle we can focus on lower order effects say, main effects and two-factor interactions, assuming the

higher order interactions to be negligible. The effect heredity principle indicates that an interaction can be active only if one or both of its parent effects are also active. For example, a two-factor interaction can be active only if one or both of the corresponding main effects are active. These principles have proven to be effective tools in resolving the aliasing patterns.

The analysis of experiments can be formulated in the form of the general linear regression where we have n observations on a dependent variable Y and p predictors $(X_1, X_2, \ldots, X_p)$, and

$$Y = X\beta + \epsilon. \tag{1}$$

where $\epsilon \sim N_n(0, \sigma^2 I)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$. Throughout this paper, we center each input variable so that the observed mean is zero, and scale each predictor so that the sample standard deviation is one.

The principle of effect sparsity can be achieved by the variable selection whose goal is to search for the model that best describes the data generating mechanism among the 2^p candidate models. However, as pointed out by Chipman, Hamada and Wu (1997), the analysis of designed experiment poses several new challenges for variable selection. Firstly, the number of predictors greatly exceeds the number of runs. For example, in the 12-run Plackett-Burman design as described in Table 1, 11 main effects and 55 interactions are to be considered. Secondly, due to the large number of potential predictors the number of candidate models most often is rather huge, which calls for computationally efficient methods. Thirdly, the predictors are always related due to the presence of interactions or polynomial terms of factors. Because of this, the principle of effect heredity is required in order to achieve a reasonable model. For example, a general purpose variable selection method may select two-factor interactions without the corresponding main effects. Such models are hard to interpret in practice.

Classical variable selection methods, such as C_p, AIC, and BIC, choose among possible models using penalized sum of squares criteria, with the penalty being an increasing function of the model dimension. These methods, however, are computationally infeasible, and their stepwise implementation is shown to be inappropriate for analyzing the designed experiments (Westfall, Young and Lin, 1998). A number of other variable selection methods have also been introduced in recent years (George and McCulloch, 1993; Foster and George, 1994; Breiman, 1995; Tibshirani, 1996; George and Foster, 2000; Fan and Li, 2001; Shen and Ye,

2002; Efron, Johnston, Hastie and Tibshirani, 2004; and Yuan and Lin, 2004a). In particular, the stochastic search variable selection method developed by George and McCulloch (1993) has been adopted by Chipman, Hamada and Wu (1997) to analyze experiments with complex aliasing patterns. As noted by Chipman, Hamada and Wu (1997), their proposal is still computationally demanding. More recently, Li and Lin (2002) applied the variable selection procedure of Fan and Li (2001) to analyze supersaturated designs. Despite its nice theoretical properties, their approach does not impose heredity principle.

The lack of a fully satisfactory variable selection strategy for analyzing experiments motivates our work here. We consider the extension of an effective variable selection algorithm LARS (least angle regression) proposed by Efron et al. (2004). The LARS is extremely fast and is closely connected with boosting and another popular variable selection method, the LASSO (Tibshirani, 1996). While the LARS enjoys great computational advantage and excellent predictive performance, it is devised for general-purpose variable selection and often produces models that are hard to interpret in practice. In this paper we propose modified LARS algorithms so that the heredity principle can be taken into account in the variable selection. It is demonstrated that incorporating such constraints in variable selection often leads to a model better satisfying the goals of experiment.

The rest of the paper is organized as follows. The LARS methodology is first reviewed in the next section. In Section 3, we present different modifications to the LARS algorithm so that the heredity principles can be taken into account. Although our main focus in this paper is on main effects and two-factor interactions, we have also explained how the methods can be extended to the case of more complicated situations in Section 3.3. The wide applicability of the proposed methods is demonstrated by three examples in Section 4. We conclude with some discussions in Section 5.

2 LARS

The LARS uses a variable selection strategy similar to the forward selection. Starting with all coefficients equal to zero, the algorithm finds the predictor that is most correlated with the response variable and proceeds in this direction. Instead of taking a full step towards the projection of Y on the variable, as would be done in a forward selection, the LARS only takes the largest step possible in this direction until some other variable has as much

correlation with the current residual. Then this new predictor is entered and the process is continued. The great computational advantage of the LARS comes from the fact that the LARS path is piecewise linear and all we need to do is to locate the change points. More specifically, the LARS algorithm can be described as follows.

Algorithm – LARS

1. Start from $\beta^{[0]} = 0$, $k = 1$ and $r^{[0]} = Y$

2. Find a predictor X_j that is most correlated with $r^{[0]}$ and set $\mathcal{B}_k = \{j\}$.

3. Compute the current direction γ which is a p dimensional vector with $\gamma_{\mathcal{B}_k^c} = 0$ and

$$\gamma_{\mathcal{B}_k} = \left(X'_{\mathcal{B}_k} X_{\mathcal{B}_k}\right)^{-} X'_{\mathcal{B}_k} r^{[k-1]}. \tag{2}$$

4. For every $i \notin \mathcal{B}_k$ compute how far the algorithm will march in direction γ before X_i has the same amount of correlation with the residual as the variables in $\mathcal{B}_k$. This can be measured by the smallest $\alpha_i \in [0, 1]$ such that

$$|X'_i(r^{[k-1]} - \alpha_i X\gamma)| = |X'_{\mathcal{B}_1}(r^{[k-1]} - \alpha_i X\gamma)|. \tag{3}$$

5. If $\mathcal{B}_k \neq \{1, \ldots, p\}$, let $\alpha = \min_{i \notin \mathcal{B}_k} \alpha_i \equiv \alpha_{i^*}$ and update $\mathcal{B}_{k+1} = \mathcal{B}_k \cup \{i^*\}$. Otherwise, set $\alpha = 1$.

6. Update $\beta^{[k]} = \beta^{[k-1]} + \alpha\gamma$, $r^{[k]} = Y - X\beta^{[k]}$ and $k = k + 1$. Go back to step 3 until $\alpha = 1$.

Here $\mathcal{B}_k$ keeps track of the variables that are included in the model at the kth stage, γ determines the direction in which the coefficient estimate will move along, and α measures how far the algorithm will march along such direction. Note that (3) is equivalent to

$$X'_i(r^{[k-1]} - \alpha_i X\gamma) = \pm X'_{\mathcal{B}_1}(r^{[k-1]} - \alpha_i X\gamma), \tag{4}$$

which can be easily solved for α. The interested readers are referred to Efron et al. (2004) for more details.

3 LARS under Heredity Principles

A LARS type algorithm is driven by the measurement of "predictability". In its original form, "predictability" is measured by the correlation with the residual. At any point on the solution path of the LARS, the variables selected are the ones that are most correlated with the current residual. Define $\theta(r, X_i)$ as the angle between the two n-vectors, r and X_i. It is clear that the squared correlation between X_i and r can be written as $\cos^2(\theta(r, X_i)) = ||X_i'r||^2/||r||^2$. This is also the proportion of the total variation in r that is explained by the regression on X_i, i.e. the R^2 when r is regressed on X_i. In other words, a variable enters the LARS path if it has the highest "predictability" on its own. Now that the heredity principles are in place, some adjustment to the LARS algorithm is needed.

We consider two versions of the heredity principle (Chipman 1996). Under *strong heredity*, for a two-factor interaction effect to be active both its parent factors should be active, whereas under *weak heredity* only one of its parent factors need to be active. We will now propose modifications to the LARS algorithm so that the selected models will obey either the strong or the weak heredity principles. This will lead to better models, provided the true model which is unknown to the experimenter, obeys the heredity principles. Exceptions are possible, but many empirical studies have confirmed the use of these principles. A Bayesian justification of effect heredity can be found in Joseph (2005).

3.1 Strong Heredity Principle

We begin with the strong heredity principle. To develop the idea, consider only the main effects and two-factor interactions for the moment. The extensions to higher order interactions will be given in a later section. In this case, if an interaction is selected, so do its corresponding main effects. To account for such dependence, in determining whether an interaction should be entered, it is natural to measure the average "predictability" of all effects that must be included. Adopting the idea of the original LARS algorithm, the predictability of a set of variables can be measured by the squared cosine of the angle between the residual and the linear space spanned by the set of variables. This idea can be illustrated by the following diagram. To measure the predictability of X_1, which is bivariate, we look at the squared cosine of α_1, the angle between Y^* and the two-dimensional linear space spanned by

the two components of X_1. This is a natural extension of the idea behind LARS. When the predictor is one-dimensional such as X_2 in the diagram, LARS looks at the squared cosine of α_2, which is the angle between two vectors, Y^* and X_2.

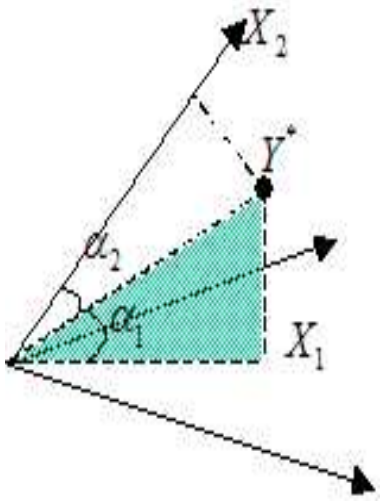

Figure 1: Predictability Measure for a Set of Variables

Once the measure of predictability for a set of variables is obtained, we have to adjust for the fact that different sets of variables have different numbers of degrees of freedom. It is clear that the more variables a set has, the better it can explain the residual for the given data. One way of adjusting for this is to measure the predictability per degree of freedom, which can be defined as $\cos^2(\theta(r, X_{\mathcal{A}}))/n_{\mathcal{A}}$ where r is the current residual, $X_{\mathcal{A}}$ is the set of effects to be entertained, and $n_{\mathcal{A}}$ is the cardinality of $\mathcal{A}$. Recall that $\cos^2(\theta(r, X_{\mathcal{A}}))$ is the R^2 when r is regressed on $X_{\mathcal{A}}$ and can be computed by fitting one linear regression. Similar idea has also been used by Yuan and Lin (2004b) in a different context. They argued that such measure of averaged predictability can also be motivated from the classical ANOVA analysis.

With such notion of average "predictability", the proposed LARS type algorithm proceeds in the following way.

Algorithm – Strong Heredity

0. Initialize $\mathcal{D}_i = \phi$ if the ith effect is a main effect. Otherwise, let $\mathcal{D}_i$ be the set of the main effects corresponding to the ith effect.

1. Start from $\beta^{[0]} = 0$, $k = 1$ and $r^{[0]} = Y$

2. Compute the "prediction score" for each candidate variable i:

$$s_i = \cos^2(\theta(r^{k-1}, X_{\{i\}\cup\mathcal{D}_i}))/(1 + n_{\mathcal{D}_i}). \tag{5}$$

   Denote $i^* = \arg\max_i s_i$. Define the current "most predictive variable" as $\mathcal{A}_1 = \{i^*\}$ and the "active set" as $\mathcal{B}_1 = \mathcal{A}_1 \cup \mathcal{D}_{i^*}$.

3. Compute the current direction γ which is a p dimensional vector with $\gamma_{\mathcal{B}_k^c} = 0$ and

$$\gamma_{\mathcal{B}_k} = \left(X'_{\mathcal{B}_k} X_{\mathcal{B}_k}\right)^- X'_{\mathcal{B}_k} r^{[k-1]}. \tag{6}$$

4. For every $i \notin \mathcal{B}_k$, update $\mathcal{D}_i = \mathcal{D}_i \cap \mathcal{B}_k^c$ and compute how far the algorithm will march in direction γ before X_i enters the most predictive set. This can be measured by the smallest $\alpha_i \in [0, 1]$ such that

$$||X'_{\{i\}\cup\mathcal{D}_i}(r^{[k-1]} - \alpha_i X\gamma)||^2/(1 + n_{\mathcal{D}_i}) \geq ||X'_{\mathcal{B}_1}(r^{[k-1]} - \alpha_i X\gamma)||^2/n_{\mathcal{B}_1}. \tag{7}$$

5. If $\mathcal{B}_k \neq \{1, \ldots, p\}$, let $\alpha = \min_{i\notin\mathcal{B}_k} \alpha_i \equiv \alpha_{i^*}$ and update $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{i^*\}$ and $\mathcal{B}_{k+1} = \mathcal{B}_k \cup \{i^*\} \cup \mathcal{D}_{i^*}$. Otherwise, set $\alpha = 1$.

6. Update $\beta^{[k]} = \beta^{[k-1]} + \alpha\gamma$, $r^{[k]} = Y - X\beta^{[k]}$ and $k = k + 1$. Go back to step 3 until $\alpha = 1$.

As in the LARS, we start with all coefficients being zero, then we compare the candidate effects in terms of the average predictability (5). For main effects, the average predictability is defined as the magnitude of the correlation between the factor and the residual. For two-factor interactions, it is defined as the average predictability of all variables from $\{i\} \cup \mathcal{D}_i$ since all of them have to enter the model as a group if the ith variable is selected. After the most predictive variable is identified, we form two different sets to keep track of (a) the most predictive variables and (b) variables that enter the model (i.e., $\mathcal{A}_k$ and $\mathcal{B}_k$). In the case of LARS, the two sets coincide. But in our case, the two may differ because some variables enter the model only because its child is highly predictive. The algorithm continues along the least square estimate with only the variables from the active set, a direction that reduces the residual sum of squares the most. We march in this direction until another variable has at least the same amount of predictability as the variables from the current most predictive set.

By the definition of γ, (7) holds when $\alpha_i = 1$ because its right hand side equals 0 in this case. Therefore, α_i in Step 4 is always well defined. Different from (3), finding α_i amounts to solving a quadratic equation, which can also be obtained in explicit form.

Another difference from the LARS algorithm is that the amount of progression measured by α is now defined through an inequality (7), rather than an equality. Such modification is necessary because $\mathcal{D}_i$ may change in the process. The averaged predictability for the ith effect can increase as a result of the inclusion of an element of $\mathcal{D}_i$. For example, consider the case of two main effects: A, B and a two-factor interaction AB. In the beginning, the predictability of AB is measured by $s_{AB,1} = \cos^2(\theta(Y, X_{\{A,B,AB\}}))/3$. Suppose that $s_{AB,1}$ is dominated by the predictive score of A, $s_A \equiv \cos^2(\theta(Y, X_{\{A\}}))$. Since A enters the model, the predictability of AB should now be measured by $s_{AB,2} = \cos^2(\theta(Y, X_{\{B,AB\}}))/2$, which might be even greater than s_A. In this case, the solution to (7) is $\alpha = 0$ and $\{AB, B\}$ enter the model immediately after A enters the model.

3.2 Weak Heredity Principle

Unlike the strong heredity principle, under the weak heredity principle, which main effect is to be included so that an interaction can be entered is not pre-determined. Therefore, any element from $\mathcal{D}_i$ can enter the model together with the ith variable. We pick the one that yields the highest predictive score. More specifically, the predictive score for the ith variable is now defined as

$$\max_{j \in \mathcal{D}_i} \cos^2(\theta(r^{k-1}, X_{\{i,j\}}))/2. \tag{8}$$

Thus we have the following algorithm:

Algorithm – Weak Heredity

0. Initialize $\mathcal{D}_i = \phi$ if the ith effect is a main effect. Otherwise, let $\mathcal{D}_i$ be the set of the main effects corresponding to the ith effect.
1. Start from $\beta^{[0]} = 0$, $k = 1$ and $r^{[0]} = Y$
2. If $\mathcal{D}_i = \phi$, define the "predictive score" of a candidate effect as $s_i = \cos^2(\theta(r^{k-1}, X_i))$. If $\mathcal{D}_i \neq \phi$, compute the "prediction scores" for each candidate effect i and each effect

in $\mathcal{D}_i$:

$$s_{ij} = \cos^2(\theta(r^{k-1}, X_{\{i,j\}}))/2. \tag{9}$$

and define $s_i = \max_j s_{ij}$. Denote $i^* = \arg\max_i s_i$. Define the current "most predictive set" as $\mathcal{A}_1 = \{i^*\}$. If $\mathcal{D}_{i^*} = \phi$, define the "active set" as $\mathcal{B}_1 = \mathcal{A}_1$. Otherwise, denote $j^* = \arg\max_j s_{i^*j}$ and define $\mathcal{B}_1 = \{i^*, j^*\}$.

3. Compute the current direction γ which is a p dimensional vector with $\gamma_{\mathcal{B}_k^c} = 0$ and

$$\gamma_{\mathcal{B}_k} = \left(X'_{\mathcal{B}_k} X_{\mathcal{B}_k}\right)^- X'_{\mathcal{B}_k} r^{[k-1]}. \tag{10}$$

4. For every $i \notin \mathcal{B}_k$, update $\mathcal{D}_i = \mathcal{D}_i \cap \mathcal{B}_k^c$. Compute how far the algorithm will march in direction γ before X_i enters the most predictive set. This can be measured by $\alpha_i \in [0, 1]$ defined as follows:

   (i) If $\mathcal{D}_i = \phi$, α_i is the smallest value such that

$$||X_i'(r^{[k-1]} - \alpha_i X\gamma)||^2 \geq ||X'_{\mathcal{B}_1}(r^{[k-1]} - \alpha_i X\gamma)||^2/n_{\mathcal{B}_1}. \tag{11}$$

   (ii) If $\mathcal{D}_i \neq \phi$, for each $j \in \mathcal{D}_i$, define α_{ij} as the smallest value in $[0, 1]$ such that

$$||X'_{\{i,j\}}(r^{[k-1]} - \alpha_{ij} X\gamma)||^2/2 \geq ||X'_{\mathcal{B}_1}(r^{[k-1]} - \alpha_{ij} X\gamma)||^2/n_{\mathcal{B}_1}. \tag{12}$$

   and $\alpha_i = \min_j \alpha_{ij}$.

5. Denote $i^* = \arg\min_i \alpha_i$ and update $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{i^*\}$. If $\mathcal{D}_{i^*} = \phi$, set $\mathcal{B}_{k+1} = \mathcal{B}_k \cup \{i^*\}$. Otherwise, define $j^* = \arg\min_j \alpha_{i^*j}$ and update $\mathcal{B}_{k+1} = \mathcal{B}_k \cup \{i^*, j^*\}$.

6. Denote $\alpha = \min_{i \notin \mathcal{B}_k} \alpha_i$ and update $\beta^{[k]} = \beta^{[k-1]} + \alpha\gamma$, $r^{[k]} = Y - X\beta^{[k]}$ and $k = k + 1$. Go back to step 3 until $\alpha = 1$.

Not every variable in $\mathcal{D}_i$ enters the model together with the ith variable under the weak heredity principle. Only the variable that yields the high predictability score together the ith variable enter the active set $\mathcal{B}_k$. The algorithm proceeds in the same fashion as the one under the strong heredity principle.

3.3 Beyond Two-factor Interactions

In an abstract level, we can represent the heredity principles by sets $\{\mathcal{D}_i : i = 1, \ldots, p\}$ where $\mathcal{D}_i$ contains a set of variables. In order that the ith variable can be considered for

inclusion, all elements of $\mathcal{D}_i$ must be included under strong heredity principle, and at least one element of $\mathcal{D}_i$ should be included under weak heredity principle. It is worth pointing out that our definition of strong and weak heredity principles is more general than their traditional version. For example, Nelder (1998) mentioned a heredity principle that requires certain main effect to be included so that an interaction is to be considered. Such a *partial heredity* principle can be induced by the strong heredity principle with the choices of $\mathcal{D}_{AB} = \{A\}$ or $\mathcal{D}_{AB} = \{B\}$. In our previous discussion, we have focused on dealing with two-factor interactions. More generally, both algorithms work for the case where $\mathcal{D}_j = \phi$ if $j \in \mathcal{D}_i$ for any i. If this is not the case, for example in the case when the three-factor interaction is also entertained, modifications to the above algorithms are necessary.

It is helpful to think of the dependence structure described by the $\mathcal{D}'$s as a directed graph where all p variables are the nodes and an edge from i to j is present if and only if $j \in \mathcal{D}_i$. To handle the strong heredity principle, we first re-evaluate the dependence set $\mathcal{D}'$s so that $\mathcal{D}_i$ contains all nodes that can be reached from the ith node. This is can be done efficiently using for example the breadth first algorithm (Cormen, Leiserson, Rivest and Rivest, 1990). After this step, the LARS algorithm with the strong heredity principle presented before can be applied.

The situation for the weak heredity principle is more complicated since we need to first figure out which variables are to be included together with the ith variable. We first determine which nodes are terminal nodes, i.e. the nodes whose dependence set is empty. Then the candidate variable sets to be considered for the ith factor to be included can be described by all the possible paths from it to any of the terminal nodes. This can be efficiently done using the depth first algorithm (Cormen et al., 1990). Denote $\{P_1, \ldots, P_k\}$ the collection of such paths. We compare these paths again using the averaged predictive scores when all nodes on the path are included. The factors on the path whose averaged predictive score is the highest will enter the model and these factors will be eliminated from the dependence sets of the remaining factors. The process then continues as the weak heredity principle algorithm presented in Section 3.2.

4 Examples

In this section, we demonstrate the proposed variable selection strategy using three examples. The first one uses a 12-run two-level nonregular design, the second one uses a 16-run two-level regular design, and the third one uses an 18-run nonregular mixed-level design. These examples are selected to show the wide applicability of our procedure.

Example 1 This is a simulated example proposed by Hamada and Wu (1992). Eleven two-level factors and their second order interactions are considered. The design is given in Table 1. The response is simulated according to the following linear model:

$$Y = A + 2AB + 2AC + \epsilon \tag{13}$$

where $\epsilon \sim \mathcal{N}(0, 0.25^2)$.

Run	A	B	C	D	E	F	G	H	I	J	K
1	+	+	−	+	+	+	−	−	−	+	−
2	+	−	+	+	+	−	−	−	+	−	+
3	−	+	+	+	−	−	−	+	−	+	+
4	+	+	+	−	−	−	+	−	+	+	−
5	+	+	−	−	−	+	−	+	+	−	+
6	+	−	−	−	+	−	+	+	−	+	+
7	−	−	−	+	−	+	+	−	+	+	+
8	−	−	+	−	+	+	−	+	+	+	−
9	−	+	−	+	+	−	+	+	+	−	−
10	+	−	+	+	−	+	+	+	−	−	−
11	−	+	+	−	+	+	+	−	−	−	+
12	−	−	−	−	−	−	−	−	−	−	−

Table 1: 12-run Plackett-Burman Design for Example 1

There are a total of 66 candidate variables (11 main effects and 55 two-factor interactions). Figure 2 compares the solution paths obtained by the new methods and the LARS algorithm. The trace of the estimated regression coefficients for the first few steps are plotted. Here the weak heredity version of the new method is able to pick up the right effects in only

two steps, while the LARS could not identify the main effect of A until the eighth step. In this example, the strong heredity did not work, which should be expected because the true model does not contain the main effects of B and C. In practice, we will not know which version of the heredity principle to be used. Therefore we should run both of them. It will be easy to select the right one by looking at the solution paths. In this example, comparing the solution paths generated by the strong and weak heredity versions of the algorithm, we can immediately understand that we should be using the weak heredity version. We also note that one of the ordinary forward selection methods proposed in Hamada and Wu (1992) could not identify any of the important effects. This clearly shows the advantages of the proposed procedure.

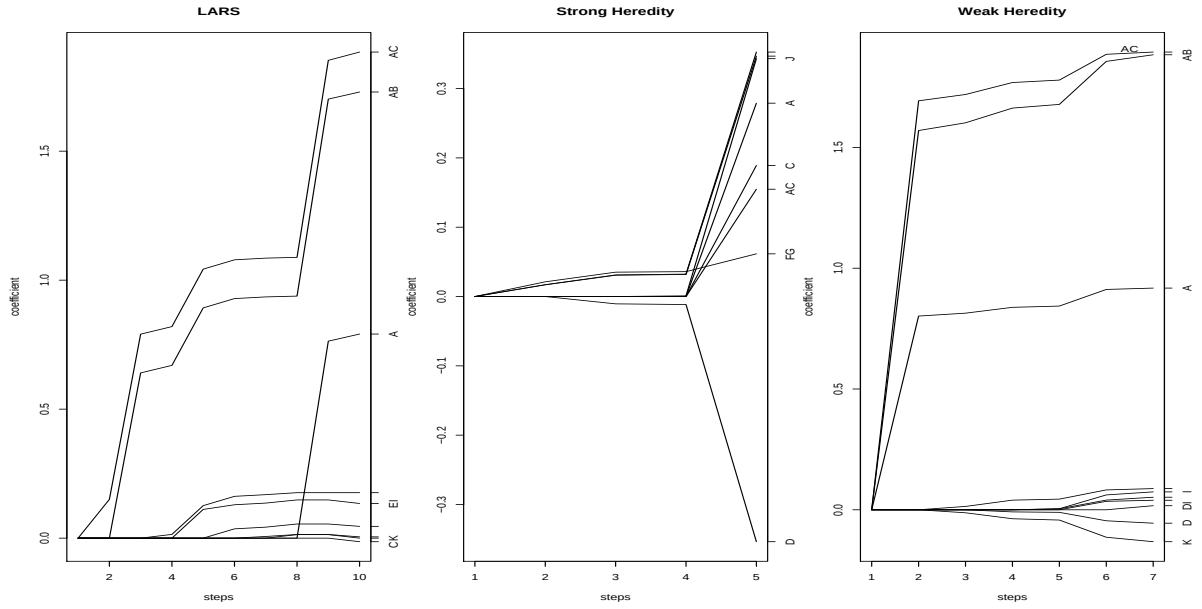

Figure 2: Solution Paths of the New Method and the LARS for the Simulated Experiment

Example 2 Consider a 2^{9-5} experiment reported by Raghavarao and Altan (2003). The design and data are given in Table 2.

The results of the analysis are plotted in Figure 3. The variables selected in the first five steps are given in Table 4. We see that LARS identifies AH, J, E, G, and CH as significant. In a 2^{9-5} design, the variables are either orthogonal or completely aliased with others. Ignoring three and higher order interactions, we can obtain the following aliasing relationships for the

above five effects:

$$\begin{aligned} AH &= EJ = DG = BG \\ J &= -CF \\ E &= -BC \\ G &= -FH = -AB \\ CH &= GJ = DE \end{aligned}$$

Thus the effect AH could actually be EJ or DG or BG. Anyone of these effects can produce the same fit. Therefore it is unclear which one to choose. LARS selected AH because it appears before EJ, DG, and BG in the list of variables (they are listed in alphabetical ordering). Hence the above selection is inconclusive. Same is the case with the selection of J, E, G, and CH. In the literature, follow-up experiments are usually recommended for dealiasing the effects (see Meyer, Steinberg, and Box 1996). On the other hand Wu and Hamada (2000) suggested that by applying effect hierarchy and effect heredity principles some of the effects can be dealiased. Our procedures incorporate both these principles and therefore the confusion due to the aliasing will be less. As given in Table 4, the first five effects identified by strong heredity are J, E, EJ, G, and GJ. Note that this is the only set of effects from the five aliasing relationships that satisfy strong heredity. Thus our procedure has no ambiguity in selecting the effects. The final model from our procedure seems to be more meaningful and interpretable. By applying the weak heredity algorithm we obtained the same effects except the last one. Instead of GJ it identifies DE. But note that these two effects are completely aliased. Weak heredity cannot break it, because one of the parent effects from both of these interactions are significant. In such a situation we recommend using strong heredity. Although DE can be significant under weak heredity, GJ is more likely to be significant because both of its parent effects G and J are significant. Interestingly, the heuristic analysis in Raghavarao and Altan (2003) also identified the same five effects J, E, EJ, G, and GJ as significant.

Example 3 The blood glucose experiment is studied by Hamada and Wu (1992) among many others. It has one two-level factor and seven three-level factors. The experimental design is the mixed-level orthogonal array $OA(18, 2^1 3^7)$. The design and the response are given in Table 3.

Run	A	B	C	D	E	F	G	H	J	Y
1	−	−	−	−	−	−	−	−	−	136.475
2	+	+	−	+	+	−	−	−	−	147.775
3	+	−	−	+	−	−	+	+	−	142.425
4	+	−	+	+	+	−	+	+	+	141.800
5	+	+	+	+	−	−	−	−	+	136.675
6	−	+	−	−	+	−	+	+	−	150.725
7	−	+	+	−	−	−	+	+	+	142.800
8	−	−	+	−	+	−	−	−	+	135.825
9	+	+	+	−	−	+	−	+	−	143.476
10	−	+	−	+	+	+	+	−	+	145.150
11	+	−	+	−	+	+	+	−	−	142.600
12	−	−	−	+	−	+	−	+	+	139.375
13	+	+	−	−	+	+	−	+	+	139.650
14	+	−	−	−	−	+	+	−	+	144.775
15	−	−	+	+	+	+	−	+	−	148.275
16	−	+	+	+	−	+	+	−	−	141.075

Table 2: The 2^{9-5} Design and Data for Example 3

Each three level factor is divided into linear and quadratic effects using the orthogonal polynomial coding (Wu and Hamada 2000). Thus there are a total of 15 main effects and 96 two-factor interactions. We treat both the linear and quadratic effects as independent main effects. This will make the application of heredity principle very simple. For example, the two-factor interaction B^2C^2 is entertained if either B^2 or C^2 is in the model (under weak heredity). For a more advanced treatment of heredity principle with polynomial and interaction terms see Chipman (1996).

Figure 4 gives the solution paths of the LARS and the two proposed methods. The plot indicates that the weak heredity principle is more likely to be true and the corresponding result is also in accordance with the previous analysis (Hamada and Wu, 1992; Chipman, Hamada, and Wu, 1997), whereas LARS identifies a model that does not satisfy any of the heredity principle (see Table 4).

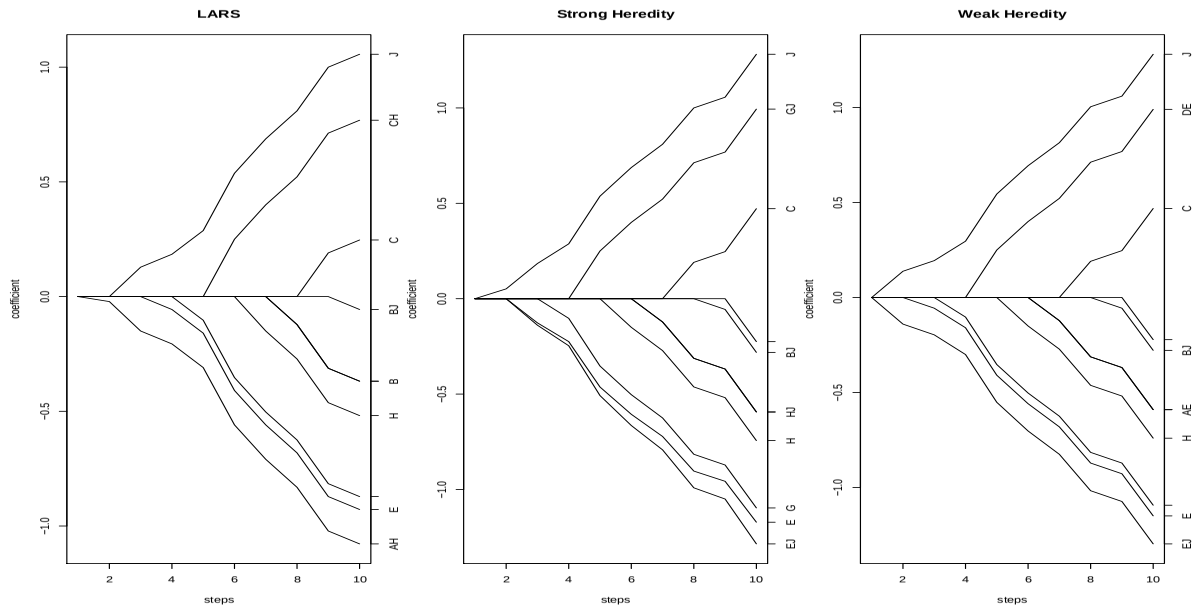

Figure 3: Solution Paths for the 2^{9-5} Factorial Experiment

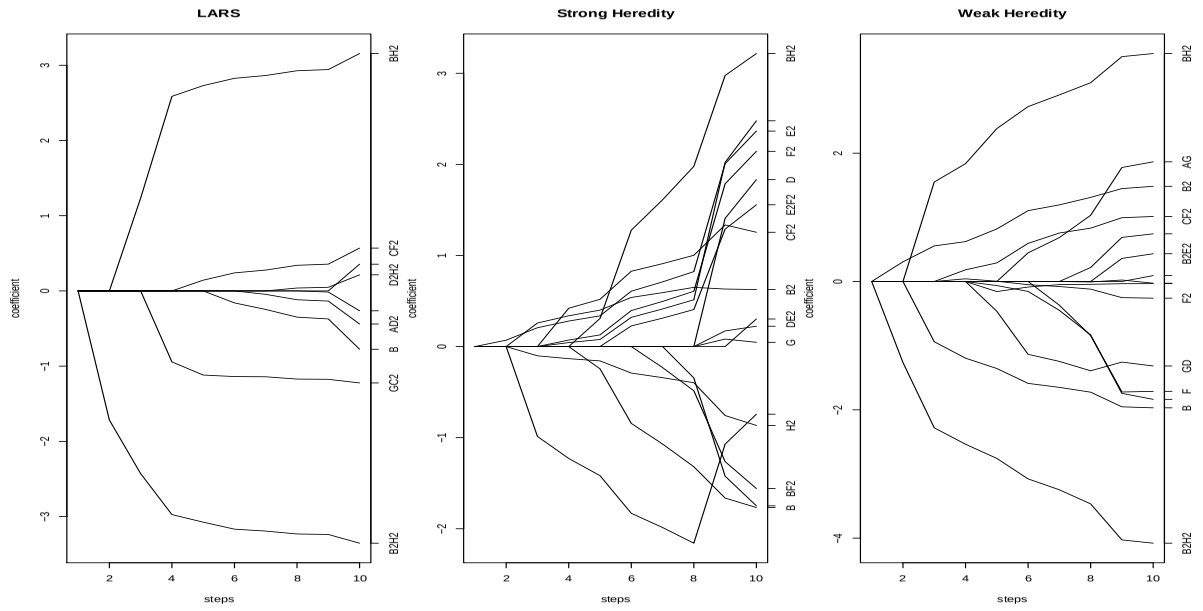

Figure 4: Solution Paths for Blood Glucose Experiment

Run	A	G	B	C	D	E	F	H	Y
1	1	1	1	1	1	1	1	1	97.94
2	1	1	2	2	2	2	2	2	83.40
3	1	1	3	3	3	3	3	3	95.88
4	1	2	1	1	2	2	3	3	88.86
5	1	2	2	2	3	3	1	1	106.58
6	1	2	3	3	1	1	2	2	89.57
7	1	3	1	2	1	3	2	3	91.98
8	1	3	2	3	2	1	3	1	98.41
9	1	3	3	1	3	2	1	2	87.56
10	2	1	1	3	3	2	2	1	88.11
11	2	1	2	1	1	3	3	2	83.81
12	2	1	3	2	2	1	1	3	98.27
13	2	2	1	2	3	1	3	2	115.52
14	2	2	2	3	1	2	1	3	94.89
15	2	2	3	1	2	3	2	1	94.70
16	2	3	1	3	2	3	1	2	121.62
17	2	3	2	1	3	1	2	3	93.86
18	2	3	3	2	1	2	3	1	96.10

Table 3: $OA(18, 2^1 3^7)$ and data from the Blood Glucose Experiment

5 Discussion

Because of the large number of candidate variables, it is imperative to use an efficient variable selection algorithm for the analysis of experiments. The LARS algorithm is a good choice. But since the variables in experiments are related due to the presence of polynomial and interaction terms, the ordinary application of LARS may lead to models that are not interpretable. To overcome this problem we have proposed a novel extension of the LARS algorithm that incorporates the effect heredity principles. Two versions of the algorithm, viz. weak and strong heredity, are presented. The proposed algorithms are computationally efficient and are shown to be able to select models better satisfying the goals of the

	Simulated			Factorial			Blood Glucose		
Step	LARS	Strong	Weak	LARS	Strong	Weak	LARS	Strong	Weak
1	AC	F, G, FG	A, AC	AH	J	J, EJ	B^2H^2	E^2	B^2, B^2H^2
2	AB	D	AB	J	E, EJ	E	BH^2	B^2, H^2, B^2H^2	B, BH^2
3	HK	J	E	E	G	G	GC^2	C, F^2, CF^2	F^2, CF^2
4	EI	A, C, AC	H	G	GJ	DE	CF^2	B, BH^2	G, GD
5	DH	B, AB	G, GJ	CH	H	H	B	E^2F^2	GC^2

Table 4: Variables Selected at the First Five Steps

experiment.

We have demonstrated the advantages of the new algorithm by analyzing a wide range of experimental designs. In some cases the weak heredity version performed better, whereas in some other cases the strong heredity version performed better. In practice, we do not know which version to use. Therefore our recommendation is to apply both and select the best one based on the solution paths generated by them.

The analysis of the 2^{9-5} fractional factorial design reiterated the importance of using heredity principle in the analysis of experiments. The ordinary LARS algorithm produced a set of aliased effects which could not be distinguished. On the other hand, the proposed approach could identify a unique model. Ambiguities are possible with the application of our algorithm, but the chances are much less. At last, we want to point out that although we have focused on the analysis of designed experiments, the techniques developed here can also be applied to the general linear regression variable selection problems.

References

[1] Breiman, L. (1995), Better subset regression using the nonnegative garrote, *Technometrics*, **37**, 373-384.

[2] Chipman, H. (1996), Bayesian variable selection with related predictors, *Canadian Journal of Statistics*, **24**, 17-36.

[3] 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.

[4] Cormen, T., Leiserson, C., Rivest, R.L. and Rivest, R. (1990), *Introduction to Algorithms*, McGraw-Hill.

[5] Efron, B., Johnstone, I., Hastie, T. and Tibshirani, R. (2004), Least angle regression, *Ann. Statist.*, **32**, 407-499.

[6] Fan, J. and Li, R. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.*, **96**, 1348-1360.

[7] Foster, D. P. and George, E. I. (1994), The risk inflation criterion for multiple regression, *Ann. Statist.*, **22**, 1947-1975.

[8] George, E. I. and Foster, D. P. (2000), Calibration and empirical Bayes variable selection, *Biometrika*, **87**, 731-747.

[9] George, E. I. and McCulloch, R. E. (1993), Variable selection via Gibbs sampling, *J. Amer. Statist. Assoc.*, **88**, 881-889.

[10] Hamada, M. and Wu, C.F.J. (1992), Analysis of designed experiments with complex aliasing, *J. Qual. Tech.*, **24**, 130-137.

[11] Joseph, V. R. (2005), A Bayesian approach to the design and analysis of fractionated experiments, to appear in *Technometrics*. (available at `http://www.isye.gatech.edu/statistics/papers/`)

[12] Li, R. and Lin, D. (2002), Data analysis in supersaturated designs, *Statist. & Probab. Lett.*, **59**, 135-144.

[13] Meyer, R. D., Steinberg, D. M., and Box, G. (1996), Follow-up designs to resolve confounding in multifactor experiments (with discussion), *Technometrics*, 38, 303-332.

[14] Raghavarao, D. and Altan, S. (2003), A heuristic analysis of highly fractionated 2^n factorial experiments, *Metrika*, **58**, 185-191.

[15] Shen, X. and Ye, J. (2002), Adaptive model selection, *J. Amer. Statist. Assoc.*, **97**, 210-221.

[16] Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *J. Royal. Statist. Soc. B.*, **58**, 267-288.

[17] Yuan, M. and Lin, Y. (2004a), Efficient empirical Bayes variable selection and estimation in linear models, to appear in *J. Amer. Statist. Assoc.*. (available at `http://www.isye.gatech.edu/~myuan/`)

[18] Yuan, M. and Lin, Y. (2004b), Model selection and estimation in regression with grouped variables, to appear in *J. Roy. Statist. Soc. B.* (available at `http://www.isye.gatech.edu/~myuan/`)

[19] Westfall, P.H., Young, S. and Lin, D. (1998), Forward selection error control in analysis of supersaturated designs, *Statist. Sinica*, **8**, 101-117.

[20] Wu, C. and Hamada, M. (2000), Experiments, *John Wiley & Sons, Inc.*, New York.