An Application of the Lasso in Experimental Design

Kentaro Tanaka, Masami Miyakawa

Abstract

We present a mathematical formulation to construct fractional factorial designs automatically. In experiments whose designs have many factors, it is sometimes difficult to try all possible combinations of the levels of the factors because of time and cost restriction. Therefore, for each experiment, we often use a fractional factorial design i.e. a design which consist of carefully chosen fraction of experimental runs of full factorial design. The difficulty here is that the criterion of the choice varies according to the situation. In this paper, we present a method to choose experimental runs from a full factorial desgin using the group lasso. The method allows us to obtain the optimal design matrix with respect to a given criterion by setting the tuning parameters adequately. We give some numerical examples to show that we can obtain the orthogonal arrays as the solutions of the group lasso problems.

1 Introduction

Design of experiments is widely used in a variety of fields such as agriculture, quality control, and simulation. One of the purpose of design of experiments is to construct the optimal experimental design with respect to a criterion under some constraints reflecting real problem. However, it is sometimes hard to obtain the optimal designs theoretically. Recently, several computer-based approaches have been developed for this problem. Especially, mixed integer programming is used to construct balanced incomplete block designs (Yokoya and Yamada (2011)), orthogonal designs (Vieira et al (2011a)) and nearly orthogonal nearly balanced mixed designs (Vieira et al (2011b)). Additionally, Jones and Nachtsheim (2011) gives randomized algorithm to obtain preferable designs for screening with factors having three levels. Interestingly, in Xiao et al (2012), they theoretically give the optimal designs for the same problem of Jones and Nachtsheim (2011), using conference matrices (Belevitch (1950)) with respect to the criterion introduced in Jones and Nachtsheim (2011).

In this paper, we propose a new machine learning approach to obtain the optimal design matrix with respect to a given criterion. As the first step to generate an optimal design matrix automatically, we consider a full factorial design, i.e. a design which consists of all possible

combinations of the levels of the factors. However, it is impractical to obtain the responses for all experimental runs of a full factorial design because of time and cost restriction. In many cases, it is necessary to reduce the number of experimental runs. Therefore, in the next step, we generate a fractional factorial design, i.e. a design which consists of carefully chosen subset of experimental runs of full factorial design. In Section 2.2, we explain that this procedure is represented as the group lasso problem. The good point of our approach is that it is able to treat a variety of criteria to choose experimental runs by changing the tuning parameters in the group lasso. Furthermore, our method benefit from the development of the algorithm to solve the group lasso problem.

The organization of this paper is as follows. In Section 2, we review the formulation of the group lasso (Section 2.1) and apply it to the problem of constructing an optimal design (Section 2.2). Section 3 is devoted to the numerical examples of our approach. In Section 4, we summarize the features of our approach.

2 An application of the group lasso to design of experiments

2.1 The group lasso

The method of the group lasso (Yuan and Lin (2007)) which is a kind of generalization of the lasso (Tibshirani (1996)) has become a popular method of variable selection for linear regression. Let us consider the usual linear regression: we have continuous outputs $\boldsymbol{y} \in \mathbb{R}^N$ and a $N \times D$ design matrix X, where N is the sample size and D is the number of input variables. Let $I_1, \ldots, I_G$ be disjoint subsets of $\{1, \ldots, D\}$, i.e. $I_g \cap I_{g'} = \emptyset$ for $g \neq g' \in \{1, \ldots, G\}$, $|I_g|$ be the number of elements in I_g, and denote the elements of I_g for $g \in \{1, \ldots, G\}$ by $i_{1g}, \ldots, i_{|I_g|g}$. The estimator of the group lasso $\hat{\boldsymbol{\beta}} \in \mathbb{R}^D$ is defined as

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^D}{\operatorname{argmin}} \left(||\boldsymbol{y} - X\boldsymbol{\beta}||_2^2 + \sum_{g=1}^{G} \lambda_g ||\boldsymbol{\beta}_{I_g}||_2 \right), \tag{1}$$

where $\lambda_1, \ldots, \lambda_G \in \mathbb{R}$ are tuning parameters, $||\cdot||_2$ stands for the Euclidean norm (not squared), and $\boldsymbol{\beta}_{I_g} = (\beta_{i_{1g}}, \ldots, \beta_{i_{|I_g|g}})$ for $g \in \{1, \ldots, G\}$. If $I_1, \ldots, I_G$ are all singletons, then the group lasso of (1) coincides with the lasso. The group lasso has the property that it does variable selection at the group level, i.e., an entire group of input variables may drop out of the model. Therefore, if we choose the tuning parameters adequately, then we obtain the sparse solution

in (1), i.e. the solution that contains many zero components. Note that the choice of the values of λ_g for $g \in \{1, \ldots, G\}$ is important: we obtain a sparse solution ($\boldsymbol{\beta}_{I_g} = \mathbf{0}$) for a sufficiently large λ_g, and a non-zero solution for a sufficiently small λ_g. The group lasso problem of (1) can be formulated as a second order cone programming and solved by the interior point methods. Furthermore, there are some specialized algorithms which solve the group lasso problem faster than the interior point methods.

2.2 Design of experiments

Let $\mathcal{F}$ be a finite subset of $\mathbb{Z}^F_{\geq 0}$, Assume that there are F factors $a_1, \ldots, a_F$ and the relation between the response variable R and the factors is formulated as

$$R = \sum_{\boldsymbol{f} \in \mathcal{F}} \gamma_{\boldsymbol{f}} \boldsymbol{a}^{\boldsymbol{f}} + \epsilon, \tag{2}$$

where $\boldsymbol{a}^{\boldsymbol{f}} = \prod_{i=1}^{F} a_i^{f_i}$ for $\boldsymbol{f} = (f_1, \ldots, f_F) \in \mathcal{F}$, $\gamma_{\boldsymbol{f}} \in \mathbb{R}$ is an unknown coefficient of $\boldsymbol{a}^{\boldsymbol{f}}$, and ϵ is the error with mean zero and variance σ^2. Let us denote the set of parameters which we want to estimate by $\boldsymbol{\gamma}'$. Then, $\boldsymbol{\gamma}'$ is a subset of $\{\gamma_{\boldsymbol{f}} \in \mathbb{R} \mid \boldsymbol{f} \in \mathcal{F}\}$. For simplicity, in this paper, we only consider the case where each factor has finite fixed levels and assume that, for each $f \in \{1, \ldots, F\}$, the f-th factor a_f has l_f levels $\alpha_{f1}, \ldots, \alpha_{fl_f}$. Let A be the set of all possible combinations of the levels of the factors, i.e. the set which consists of all experimental runs of a full factorial design, and each of the elements is represented as a column vector. Then A has $\prod_{f=1}^{F} l_f$ elements and A can be represented as $A = \{(\alpha_{1i_1}, \ldots, \alpha_{Fi_F})^T \mid i_f \in \{1, \ldots, l_f\},\ f \in \{1, \ldots, F\}\}$ where T stands for transpose. Let G be the number of elements of A and $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_G\}$ be the elements of A, i.e. the experimental runs of A, in an arbitrary order. The design matrix C is defined such that the g-th column is $\boldsymbol{a}_g$, i.e., $C = [\boldsymbol{a}_1 \ \ldots \ \boldsymbol{a}_G]$.

Example 1. *Let us consider the case where there are $F = 3$ factors a_1, a_2, a_3 and each factor has two levels 1 or -1. That means that values of $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{32}$ can be defined as $\alpha_{f1} = 1$ and $\alpha_{f2} = -1$ for each $f \in \{1, 2, 3\}$. In this case, there are $G = 2^3 = 8$ experimental runs in the full factorial design and A is represented as*

$$\begin{aligned} A &= \{(\alpha_{1i_1}, \alpha_{2i_2}, \alpha_{3i_3})^T \mid i_1, i_2, i_3 \in \{1, 2\}\} \\ &= \{(1, 1, 1)^T, (1, 1, -1)^T, \ldots, (-1, -1, -1)^T\} \end{aligned}$$

Furthermore, the column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_8$ *can be defined as* $\boldsymbol{a}_1 = (1,1,1)^T$, $\boldsymbol{a}_2 = (1,1,-1)^T$, $\ldots, \boldsymbol{a}_8 = (-1,-1,-1)^T$. *Then the design matrix* C *which consists of* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_8$ *is defined as follows:*

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}. \tag{3}$$

□

Let $|\mathcal{F}|$ be the number of elements in $\mathcal{F}$ and $\mathcal{F} = \{\boldsymbol{f}_0, \ldots, \boldsymbol{f}_{|\mathcal{F}|-1}\}$ sorting the elements of $\mathcal{F}$ in an arbitrary order. For each $\boldsymbol{a}_g$ for $g \in \{1, \ldots, G\}$, let $\tilde{\boldsymbol{a}}_g = (\boldsymbol{a}_g^{\boldsymbol{f}_0}, \ldots, \boldsymbol{a}_g^{\boldsymbol{f}_{|\mathcal{F}|-1}})^T$. Then the model matrix M of the experimental runs of A is defined such that the g-th column is $\tilde{\boldsymbol{a}}_g$, i.e., $M = [\tilde{\boldsymbol{a}}_1 \ \ldots \ \tilde{\boldsymbol{a}}_G]$.

Example 2. *As in Example 1, we consider the case where there are* $F = 3$ *factors and each factor has two levels* 1 *or* -1. *Furthermore, we assume that the relation between the response variable and the factors is formulated as*

$$R = \gamma_{000} + \gamma_{100}a_1 + \gamma_{010}a_2 + \gamma_{001}a_3 + \epsilon. \tag{4}$$

In this case, $\mathcal{F} = \{(0,0,0), (1,0,0), (0,1,0), (0,0,1)\}$ *and* $\tilde{\boldsymbol{a}}_1, \tilde{\boldsymbol{a}}_2, \ldots, \tilde{\boldsymbol{a}}_G$ *are obtained as follows:*

$$\begin{aligned} \tilde{\boldsymbol{a}}_1 &= (1^0 1^0 1^0, 1^1 1^0 1^0, 1^0 1^1 1^0, 1^0 1^0 1^1)^T = (1,1,1,1)^T \\ \tilde{\boldsymbol{a}}_2 &= (1^0 1^0 (-1)^0, 1^1 1^0 (-1)^0, 1^0 1^1 (-1)^0, 1^0 1^0 (-1)^1)^T = (1,1,1,-1)^T \\ &\vdots \\ \tilde{\boldsymbol{a}}_8 &= ((-1)^0 (-1)^0 (-1)^0, (-1)^1 (-1)^0 (-1)^0, (-1)^0 (-1)^1 (-1)^0, (-1)^0 (-1)^0 (-1)^1)^T \\ &= (1,-1,-1,-1)^T \end{aligned}$$

Then the model matrix M *of the experimental runs of* A *is defined as follows:*

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}. \tag{5}$$

Note that, for simplicity, we may also write $\gamma_0, \gamma_1, \gamma_2$ *and* γ_3 *instead of* γ_{000}, γ_{100}, γ_{010} *and* γ_{001}.

□

For $j \in \{0, \ldots, |\mathcal{F}| - 1\}$, let $\boldsymbol{e}_j = \boldsymbol{e}_{\boldsymbol{f}_j}$ be the $|\mathcal{F}|$-dimensional column vector such that $\boldsymbol{e}_j$ has an entry 1 at the $(j+1)$-th row and 0 otherwise. Furthermore, for $j \in \{0, \ldots, |\mathcal{F}| - 1\}$, let $\gamma_j = \gamma_{\boldsymbol{f}_j}$ as in Example 2, and $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{|\mathcal{F}|-1})^T$. For $g \in \{1, \ldots, G\}$, let R_g and ϵ_g be the response variable and the error respectively when the factors $\boldsymbol{a}_g^{\boldsymbol{f}_0}, \ldots, \boldsymbol{a}_g^{\boldsymbol{f}_{|\mathcal{F}|-1}}$ are given. Let $\boldsymbol{R} = (R_1, \ldots, R_G)^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_G)^T$. We use the following proposition to formulate the problem of constructing an optimal design matrix as a problem of mathematical programming.

Proposition 3. *For a fixed $j \in \{0, \ldots, |\mathcal{F}| - 1\}$, if there exists $\boldsymbol{\beta}_{j\circ} = (\beta_{j1}, \ldots, \beta_{jG}) \in \mathbb{R}^G$ such that $\beta_{j1}\tilde{\boldsymbol{a}}_1 + \cdots + \beta_{jG}\tilde{\boldsymbol{a}}_G = \boldsymbol{e}_j$, then $\hat{\gamma_j} = \beta_{j1}R_1 + \cdots + \beta_{jG}R_G$ is an unbiased estimator of γ_j. The variance of $\hat{\gamma_j}$ is $\mathrm{Var}[\hat{\gamma_j}] = \sigma^2 ||\boldsymbol{\beta}_{j\circ}||_2^2$.*

Proof: Note that $\boldsymbol{R} = M^T\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ and $M\boldsymbol{\beta}_{j\circ} = \boldsymbol{e}_j$ from the assumption. Then we obtain

$$\begin{aligned} E[\hat{\gamma_j}] &= E[\beta_{j1}R_1 + \cdots + \beta_{jG}R_G] \;=\; E[\boldsymbol{R}^T\boldsymbol{\beta}_{j\circ}] \;=\; E[(M^T\boldsymbol{\gamma} + \boldsymbol{\epsilon})^T\boldsymbol{\beta}_{j\circ}] \\ &= E[\boldsymbol{\gamma}^T M\boldsymbol{\beta}_{j\circ}] + E[\boldsymbol{\epsilon}^T\boldsymbol{\beta}_{j\circ}] \;=\; E[\boldsymbol{\gamma}^T\boldsymbol{e}_j] \;=\; \gamma_j. \end{aligned}$$

The variance is

$$\mathrm{Var}[\hat{\gamma_j}] \;=\; \beta_{j1}^2\mathrm{Var}[R_1] + \cdots + \beta_{jG}^2\mathrm{Var}[R_G] \;=\; \sigma^2||\boldsymbol{\beta}_{j\circ}||_2^2.$$

□

In the following, we consider a method to construct a design matrix and estimators of $\boldsymbol{\gamma}'$ in the design matrix based on some criteria. As one of our design criteria, we use the A-optimality in which the sum (or the "A"verage) of the variances of the estimators is minimized. In general, the more the experimental runs, the smaller the variances of the estimates. Therefore, the full factorial design is fine if we consider only the A-optimality as a criterion. However, it is sometimes difficult to try all possible combinations of the levels of the factors because of time and cost restriction. Therefore, as another criterion, we consider the number of experimental runs. We consider that a design which has fewer experimental runs is preferable. Thus, we need to consider two tasks at the same time: i) to minimize the variances of the estimatotrs, and ii) to try to reduce the number of experimental runs. There is a trade-off between these two tasks. Hence, it is prefarable to formulate the tasks as a mathematical programming problem which enables us to control the balance between them by changing the values of tuning parameters.

Let us denote the set of the indices of the parameters which we want to estimate by $J \subseteq \{1, \ldots, |\mathcal{F}|\}$. Then, $\boldsymbol{\gamma}'$ is represented as $\{\gamma_j \mid j \in J\}$. For each $\gamma_j \in \boldsymbol{\gamma}'$, we consider a linear

estimator $\hat{\gamma_j} = \beta_{j1}R_1 + \cdots + \beta_{jG}R_G = \boldsymbol{R}^T\boldsymbol{\beta}_{j\circ}$. From Proposition 3, the sum of the variances of the estimators for $\{\gamma_j \mid j \in J\}$ is $\sigma^2 \sum_{j\in J} ||\boldsymbol{\beta}_{j\circ}||_2^2$. Therefore, based on the A-optimality criterion, we need to minimize the sum $\sum_{j\in J} ||\boldsymbol{\beta}_{j\circ}||_2^2$ under the condition of the unbiasedness: $M\boldsymbol{\beta}_{j\circ} = \boldsymbol{e}_j$ for $j \in J$. For each $g \in \{1, \ldots, G\}$, let $\boldsymbol{\beta}_{\circ g}$ be a column vector whose entries are $\{\beta_{jg} \mid j \in J\}$. Next, to try to reduce the number of experimental runs, we consider the group lasso penalty for each of $\boldsymbol{\beta}_{\circ 1}, \ldots, \boldsymbol{\beta}_{\circ G}$. If $\boldsymbol{\beta}_{\circ g} = \boldsymbol{0}$ for a $g \in \{1, \ldots, G\}$, then the response R_g at the g-th experimental run is not used for the estimators of $\boldsymbol{\gamma}'$ and hence we do not need to conduct the g-th experimental run. Thus, we use the method of the group lasso for variable selection in regression models to choose the subset of the experimental runs of full factorial design. Therefore, the following penalized least square gives the solution which minimizes the sum of the variances of the estimators under the condition of the unbiasedness with consideration of the number of experimental runs.

$$\begin{aligned} \min_{\{\boldsymbol{\beta}_{j\circ} | j\in J\}} \quad & \left(\sum_{j\in J} ||\boldsymbol{\beta}_{j\circ}||_2^2 + \sum_{g=1}^{G} \lambda_g ||\boldsymbol{\beta}_{\circ g}||_2 \right) \\ \text{s.t.} \quad & M\boldsymbol{\beta}_{j\circ} = \boldsymbol{e}_j, \quad (j \in J). \end{aligned} \tag{6}$$

Here, $\lambda_1, \ldots, \lambda_G \in \mathbb{R}$ are tuning parameters. The problem of (6) is a second order cone programming and can be solved by the interior point methods.

Note that the formulation of (6) contains linear constraints and some specialized algorithms which solves the group lasso problem may not applicable. We can also consider the Lagrangian relaxation problem of (6) as follows:

$$\min_{\{\boldsymbol{\beta}_{j\circ} | j\in J\}} \quad \sum_{j\in J} (||\boldsymbol{\beta}_{j\circ}||_2^2 + \kappa_j ||M\boldsymbol{\beta}_{j\circ} - \boldsymbol{e}_j||_2^2) + \sum_{g=1}^{G} \lambda_g ||\boldsymbol{\beta}_{\circ g}|| \tag{7}$$

where, κ_j, $(j \in J)$ is a tuning parameter. The formulation of (7) is the same as the group lasso and thus the specialized algorithms for the group lasso is applicable. In particular, the solution of the problem of (7) does not accurately satisfy the constraints of the unbiasedness when κ_j's are not so large. This means that, from the problem of (7), we can obtain the solution when we allow confounding among the factors. Therefore, the formulation of (7) works even when the number of the elements of J is larger than G or the number of non-zero $\boldsymbol{\beta}_{\circ g}$'s.

As the following example shows, in the formulation of (6) we have to determine the values of $\lambda_1, \ldots, \lambda_G \in \mathbb{R}$ carefully to obtain the sparse solution.

Example 4. *As in Examples 1 and 2, we assume that there are $F = 3$ factors and each factor has two levels 1 or -1. Furthermore, we assume that the design matrix C is given by* (3)*. We consider the main effect model: $R = \gamma_0 + \sum_{j=1}^{3} \gamma_j a_j + \epsilon$. Then the model matrix is given by* (5)*. Suppose that we want to estimate γ_1, γ_2 and γ_3, i.e. $J = \{1, 2, 3\}$. To this model, traditional*

Table 1: L_4 orthogonal array

Run	1	4	6	7
a_1	1	1	-1	-1
a_2	1	-1	1	-1
a_3	1	-1	-1	1

Table 2: The orthogonal array $(-1) \times L_4$

Run	2	3	5	8
a_1	-1	-1	1	1
a_2	-1	1	-1	1
a_3	-1	1	1	-1

design of experiments evaluates that L_4 orthogonal array in Table 1 is fine. Note that Table 1 is transposed against the traditional notation of L_4 orthogonal array, i.e., each column corresponds to a run of the experiment and the rows indicate the levels of a_1, a_2 and a_3. The following is one of feasible solutions of (6)*:*

$$\begin{aligned} \boldsymbol{\beta}_{1\circ} &= \frac{1}{4}(1, 0, 0, 1, 0, -1, -1, 0)^T, \\ \boldsymbol{\beta}_{2\circ} &= \frac{1}{4}(1, 0, 0, -1, 0, 1, -1, 0)^T, \\ \boldsymbol{\beta}_{3\circ} &= \frac{1}{4}(1, 0, 0, -1, 0, -1, 1, 0)^T. \end{aligned} \tag{8}$$

Note that the 2nd, 3rd, 5th and 8th elements of $\boldsymbol{\beta}_{1\circ}, \boldsymbol{\beta}_{2\circ}$ and $\boldsymbol{\beta}_{3\circ}$ are all zero in (8)*. This means that only the 1st, 4th, 6th and 7th experimental run in* (3) *are used for the estimation of γ_1, γ_2 and γ_3, and hence the solution* (8) *gives L_4 orthognal array given in Table 1. If the values of tuning parameters $\lambda_1, \ldots, \lambda_G \in \mathbb{R}$ are properly given, then we obtain the solution* (8) *as the optimal solution of* (6)*. However, as shown below, we need to pay a lot of attention to the values of tuning parameters $\lambda_1, \ldots, \lambda_G \in \mathbb{R}$. If we set $\lambda_1 = \cdots = \lambda_8 = \lambda$ for any non-negative real number λ, then we can not obtain the solution* (8)*, i.e. L_4 orthogonal array, as the optimal solution of* (6)*. To see this, let us consider the following feasible solution:*

$$\begin{aligned} \boldsymbol{\beta}_{1\circ} &= \frac{1}{4}(0, 1, 1, 0, -1, 0, 0, -1)^T, \\ \boldsymbol{\beta}_{2\circ} &= \frac{1}{4}(0, 1, -1, 0, 1, 0, 0, -1)^T, \\ \boldsymbol{\beta}_{3\circ} &= \frac{1}{4}(0, -1, 1, 0, 1, 0, 0, -1)^T. \end{aligned} \tag{9}$$

This solution (9) *corresponds to the orthogonal array in Table 2. It can be easily seen that the value of the objective function of* (6) *for* (8) *and that for* (9) *are the same. Furthermore, note that the problem of* (6) *is strictly convex and thus the minimum is attained at a unique point. Therefore, the feasible solution* (8) *is not the optimal solution of* (6)*, if we set* $\lambda_1 = \cdots = \lambda_8 = \lambda$. □

The above example implies that the presence of the symmetries in the formulation of (6) prevents the sparseness of the solution. In Example 4, if we set $\lambda_1 = \cdots = \lambda_8 = \lambda$, then there exists a symmetry with respect to the transformation $\boldsymbol{\beta}_{\circ 1} \leftrightarrow -\boldsymbol{\beta}_{\circ 8}$, $\boldsymbol{\beta}_{\circ 2} \leftrightarrow -\boldsymbol{\beta}_{\circ 7}$, $\boldsymbol{\beta}_{\circ 3} \leftrightarrow -\boldsymbol{\beta}_{\circ 6}$, $\boldsymbol{\beta}_{\circ 4} \leftrightarrow -\boldsymbol{\beta}_{\circ 5}$. Therefore, for example, if there exists a feasible solution satisfying $\boldsymbol{\beta}_{\circ 1} = \mathbf{0}$, then there exists a feasible solution satisfying $\boldsymbol{\beta}_{\circ 8} = \mathbf{0}$ having the same value of the objective function. Because the minimum of the objective function is attained at a unique point, a feasible solution satisfying $\boldsymbol{\beta}_{\circ 1} = \mathbf{0}$ and $\boldsymbol{\beta}_{\circ 8} \neq \mathbf{0}$ is not the optimal solution. A feasible solution satisfying $\boldsymbol{\beta}_{\circ 1} = \mathbf{0}$ and $\boldsymbol{\beta}_{\circ 8} = \mathbf{0}$ might be an optimal solution, but we can not reduce the number of experimental runs any more due to the linear constraints.

3 Numerical examples

In this section, we mainly consider the case where each factor has two levels $\{1, -1\}$. For our experiment, a laptop with an 1.20 GHz CPU and 8GB RAM was used. You can find the file of the program written in Python used in the following examples at `https://github.com/tanaken-basis/explasso`.

As Example 4 shows, we need to carefully determine the values of $\lambda_1, \ldots, \lambda_G$. Therefore, it is important to investigate how the optimal solution depends on the values of tuning parameters. In the following Examples 5, 6 and 7, the specified values are used as the values of tuning parameters. In Examples 8 and 9, we use uniform random numbers between 0 and 100 as the values of tuning parameters.

Example 5. *As in Examples 1 and 2, we assume that there are* $F = 3$ *factors and each factor has two levels* 1 *or* -1*, and the model and model matrix are given by* (4) *and* (5) *respectively. Suppose that we want to estimate* γ_1, γ_2 *and* γ_3*. Furthermore, we use the following values as the values of the tuning parameters:* $\lambda_1 = \lambda_4 = \lambda_6 = \lambda_7 = 1$*,* $\lambda_2 = \lambda_3 = \lambda_5 = \lambda_8 = 10$*. Then,*

by solving (6), we obtain the optimal solution $\boldsymbol{\beta}_{1\circ}, \boldsymbol{\beta}_{2\circ}, \boldsymbol{\beta}_{3\circ}$ *as follows:*

$$\begin{aligned} \boldsymbol{\beta}_{1\circ} &= \frac{1}{4}(1,0,0,1,0,-1,-1,0)^T, \\ \boldsymbol{\beta}_{2\circ} &= \frac{1}{4}(1,0,0,-1,0,1,-1,0)^T, \\ \boldsymbol{\beta}_{3\circ} &= \frac{1}{4}(1,0,0,-1,0,-1,1,0)^T. \end{aligned} \tag{10}$$

The computation time was 0.039 seconds. Table 3 is the optimal design matrix which consists of the $1st$*,* $4th$*,* $6th$ *and* $7th$ *experimental run in* (3)*, i.e.* $2nd$*,* $3rd$*,* $5th$ *and* $8th$ *elements of* $\boldsymbol{\beta}_{1\circ}, \boldsymbol{\beta}_{2\circ}$ *and* $\boldsymbol{\beta}_{3\circ}$ *in* (10) *are all zero. Note that Table 3 is identical with the* L_4 *orthogonal array in Table 1.* □

Table 3: The optimal design matrix in Example 5

Run	1	4	6	7
a_1	1	1	-1	-1
a_2	1	-1	1	-1
a_3	1	-1	-1	1

In the above Example 5, the results are almost obvious because we choose the values of the tuning parameters such that the L_4 orthognal array is obtained as the optimal design and only the experimental runs whose values of tuning parameters are 1 are chosen in the optimal design matrix. In the next example, we see an example where an experimental run with a large tuning parameter value is sometimes chosen. That means that the optimal design matrix is not determined only by the values of tuning parameters.

Example 6. *Assume that there are* $F = 4$ *factors* a_1, a_2, a_3, a_4 *and each factor has two levels* -1 *or* 1*. Furthermore, we assume that the relation between the response variable and the factors is formulated as*

$$R = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_3 + \gamma_4 a_4 + \gamma_5 a_1 a_2 + \gamma_6 a_1 a_3 + \gamma_7 a_1 a_4 + \epsilon. \tag{11}$$

In this case, there are $G = 2^4 = 16$ *experimental runs in the full factorial design and the model*

matrix M of the full factorial design is given as follows:

$$\begin{array}{c} const. \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_1a_2 \\ a_1a_3 \\ a_1a_4 \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix}. \quad (12)$$

Suppose that we want to estimate $\gamma_1, \ldots, \gamma_7$. Furthermore, we use the following values as the values of the tuning parameters:

$$(\lambda_1, \ldots, \lambda_{16}) = (1, 40, 45, 10, 45, 15, 5, 40, 45, 10, 5, 30, 5, 45, 40, 50). \quad (13)$$

Then, by solving (6), we obtain the optimal solution $\boldsymbol{\beta}_{1\circ}, \ldots, \boldsymbol{\beta}_{7\circ}$ as follows:

$$\begin{aligned} \boldsymbol{\beta}_{1\circ} &= \frac{1}{8}(1,0,0,1,0,1,1,0,0,-1,-1,0,-1,0,0,-1)^T, \\ \boldsymbol{\beta}_{2\circ} &= \frac{1}{8}(1,0,0,1,0,-1,-1,0,0,1,1,0,-1,0,0,-1)^T, \\ \boldsymbol{\beta}_{3\circ} &= \frac{1}{8}(1,0,0,-1,0,1,-1,0,0,1,-1,0,1,0,0,-1)^T, \\ \boldsymbol{\beta}_{4\circ} &= \frac{1}{8}(1,0,0,-1,0,-1,1,0,0,-1,1,0,1,0,0,-1)^T, \\ \boldsymbol{\beta}_{5\circ} &= \frac{1}{8}(1,0,0,1,0,-1,-1,0,0,-1,-1,0,1,0,0,1)^T, \\ \boldsymbol{\beta}_{6\circ} &= \frac{1}{8}(1,0,0,-1,0,1,-1,0,0,-1,1,0,-1,0,0,1)^T, \\ \boldsymbol{\beta}_{7\circ} &= \frac{1}{8}(1,0,0,-1,0,-1,1,0,0,1,-1,0,-1,0,0,1)^T. \end{aligned} \quad (14)$$

The computation time was 0.196 seconds. Table 4 shows the optimal design matrix given by the solution (14)*. Note that Table 4 is equivalent to the L_8 orthogonal array. Furthermore, note that the 16th experimental run with the greatest tuning parameter $\lambda_{16} = 50$ is chosen in this setting.* □

In the next example, we consider the case where a non-orthognal array is generated.

Example 7. *Assume that there are $F = 4$ factors a_1, a_2, a_3, a_4 and each factor has two levels* 1 *or* -1*. Furthermore, we assume that the relation between the response variable and the factors*

Table 4: The optimal design matrix in Example 6

Run	1	4	6	7	10	11	13	16
a_1	1	1	1	1	-1	-1	-1	-1
a_2	1	1	-1	-1	1	1	-1	-1
a_3	1	-1	1	-1	1	-1	1	-1
a_4	1	-1	-1	1	1	-1	-1	1
a_1a_2	1	1	-1	-1	-1	-1	1	1
a_1a_3	1	-1	1	-1	-1	1	-1	1
a_1a_4	1	-1	-1	1	-1	1	1	-1

is formulated as

$$R = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_3 + \gamma_4 a_4 + \gamma_5 a_1 a_2 + \gamma_6 a_1 a_3 + \gamma_7 a_1 a_4 + \gamma_8 a_2 a_3 + \epsilon.$$

In this case, there are $G = 2^4 = 16$ *experimental runs in the full factorial design and the model matrix* M *of the full factorial design is given as follows:*

$$\begin{array}{c} const. \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_1a_2 \\ a_1a_3 \\ a_1a_4 \\ a_2a_3 \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}.$$

Suppose that we want to estimate $\gamma_1, \ldots, \gamma_8$. *Then, by solving (6) with the tuning parameters*

given in (13)*, we obtain the optimal solution* $\boldsymbol{\beta}_{1\circ}, \ldots, \boldsymbol{\beta}_{8\circ}$ *as follows:*

$$\begin{aligned}
\boldsymbol{\beta}_{1\circ} &= \frac{1}{8}(1,0,0,1,0,1,1,0,0,-1,-0.115,-0.885,-1,0,-0.885,-0.115)^T,\\
\boldsymbol{\beta}_{2\circ} &= \frac{1}{8}(1,0,0,1,0,-1,-1,0,0,1,1.868,-0.868,-1,0,-0.868,-0.132)^T,\\
\boldsymbol{\beta}_{3\circ} &= \frac{1}{8}(1,0,0,-1,0,1,-1,0,0,1,-0.115,-0.885,1,0,-0.885,-0.115)^T,\\
\boldsymbol{\beta}_{4\circ} &= \frac{1}{8}(1,0,0,-1,0,-1,1,0,0,-1,1.868,-0.868,1,0,-0.868,-0.132)^T,\\
\boldsymbol{\beta}_{5\circ} &= \frac{1}{8}(1,0,0,1,0,-1,-1,0,0,-1,-1.868,0.868,1,0,0.868,0.132)^T,\\
\boldsymbol{\beta}_{6\circ} &= \frac{1}{8}(1,0,0,-1,0,1,-1,0,0,-1,0.115,0.885,-1,0,0.885,0.115)^T,\\
\boldsymbol{\beta}_{7\circ} &= \frac{1}{8}(1,0,0,-1,0,-1,1,0,0,-1,-1.932,2.932,1,0,-1.068,0.068)^T,\\
\boldsymbol{\beta}_{8\circ} &= \frac{1}{8}(0,0,0,0,0,0,0,0,0,2,0.064,-2.064,-2,0,1.936,0.064)^T,
\end{aligned} \tag{15}$$

The computation time was 0.211 seconds. Table 5 shows the optimal design matrix given by the solution (15)*. Note that the first* 6 *columns of Table 5 are similar to Table 4.* □

Table 5: The optimal design matrix in Example 7

Run	1	4	6	7	10	11	12	13	15	16
a_1	1	1	1	1	-1	-1	-1	-1	-1	-1
a_2	1	1	-1	-1	1	1	1	-1	-1	-1
a_3	1	-1	1	-1	1	-1	-1	1	-1	-1
a_4	1	-1	-1	1	-1	1	-1	1	1	-1
a_1a_2	1	1	-1	-1	-1	-1	-1	1	1	1
a_1a_3	1	-1	1	-1	-1	1	1	-1	1	1
a_1a_4	1	-1	-1	1	1	-1	1	-1	-1	1
a_2a_3	1	-1	-1	1	1	-1	-1	-1	1	1

In the above Examples 5, 6 and 7, the specified values are used as the values of tuning parameters. In the next example, we consider what happens if the values of tuning parameters are randomly chosen.

Example 8. *As in Example 6, we assume that there are* $F = 4$ *factors and each factor has two levels* -1 *or* 1*, and the model and model matrix are given by* (11) *and* (12) *respectively. In this case, there are* $G = 2^4 = 16$ *experimental runs in the full factorial design and the model matrix*

Table 6: Frequency table of number of the experimental runs

# of experimental runs	Frequency
8	51
9	186
10	299
11	257
12	138
13	52
14	16
15	1
16	0

M of the full factorial design is given by (12)*. Suppose that we want to estimate $\gamma_1, \ldots, \gamma_7$. In this example, we investigate the number of the experimental runs of the optimal design given by solving (6) with randomy chosen tuning parameters. We use uniform random numbers between 0 and 100 as the values of the tuning parameters. We repeated the following computations 1000 times and obtained the results in Table 6.*

Step 1. *Generate uniform random numbers between 0 and 100 as the values of tuning parameters $\lambda_1, \ldots, \lambda_{16}$.*

Step 2. *Solve (6) with tuning parameters given in Step 1.*

Step 3. *Count the number of the experimental runs of the optimal design given in Step 2.*

As shown in this results, though we chose the values of tuning parameters randomly, the number of the experimental runs of the optimal design was less than or equal to 10 in more than half of the cases. □

Finally, we give an example which shows how the computation time increases as the number of factors increases.

Table 7: The number of factors and computation time in seconds

# of factors	time [sec]
1	0.023
2	0.017
3	0.051
4	0.117
5	0.335
6	1.533
7	8.440
8	60.096
9	560.237

Example 9. *In this example, we observe the computation time to solve the problem of* (6)*, changing the number of factors* F *from* 1 *to* 9*. We assume that each factor has two levels* 1 *or* -1 *and the relation between the response variable and the factors is formulated as the main effect model (with no interaction terms) i.e.*

$$R = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2 + \cdots + \gamma_F a_F + \epsilon.$$

Suppose that we want to estimate $\gamma_1, \ldots, \gamma_F$*. Furthermore, we use uniform random numbers between 0 and 100 as the values of the tuning parameters. Table 7 implies the computation time increases exponentially as the number of factors increases.* □

4 Concluding remarks

We apply the group lasso to construct fractional factorial designs. Though, in the examples treated in this paper, we mainly treat the case where each factor has two levels and the full factorial design A which consists of all combinations of the levels of the factors without repetition, our approach, described in Section 2.2, has the following features.

- Each factor can have two or more levels.
- By duplicating the columns of the design matrix, we can treat the case of repeated measurements.

- Assume that we have already observed the responses at the experimental runs of $\boldsymbol{a}_{g'_1}, \ldots, \boldsymbol{a}_{g'_Q}$ where $g'_1, \ldots, g'_Q \in \{1, \ldots, G\}$. Then, by setting $\lambda_{g'_1} = \cdots = \lambda_{g'_Q} = 0$, we can choose additional experimental runs given $R_{g'_1}, \ldots, R_{g'_Q}$.

Furthermore, we do not need to consider all combinations of the levels of the factors in A. If the number of the levels or the factors increases, then the number of the combinations increases explosively. This means that the number of the variables in the formulation of (6) increases explosively, and hence it becomes difficult to solve the problem. Therefore, if the number of the levels or the factors is large, then it is needed to reduce the number of the experimental runs of A in advance to bound the number of variables in (6). For future work, we will investigate how to choose the experimental runs in advance. We will also investigate how to determine the values of $\lambda_1, \ldots, \lambda_G \in \mathbb{R}$.

Kentaro Tanaka (Associate Professor, Faculty of Economics, Seikei University)

Masami Miyakawa (Professor, Department of Industrial Engineering and Economics, Tokyo Institute of Technology)

This work was supported by JSPS KAKENHI Grant Number JP16K00056.

References

Belevitch, V. (1950). Theorem of 2n-terminal networks with application to conference telephony. Electronic Communications, 26, pp. 231–244.

Jones, B., & Nachtsheim, C. J. (2011). A class of three level designs for definitive screening in the presence of second-order effects. Journal of Quality Technology, 43(1), pp. 1–15.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), pp. 267–288.

Vieira, H. Jr., Sanchez, S. M., Kienitz, K. H., & Belderrain, M. C. N. (2011a). Generating and improving orthogonal designs by using mixed integer programming. European Journal of Operational Research, 215, pp. 629–638.

Vieira, H. Jr., Sanchez, S. M., Kienitz, K. H., & Belderrain, M. C. N. (2011b). Improved efficient, nearly orthogonal, nearly balanced mixed designs. Proceedings of the 2011 Winter Simulation Conference, pp. 3605–3616.

Xiao, L., Lin, D. K. J., & Bai, F. (2012). Constructing definitive screening design using conference matrices. Journal of Quality Technology, 44(1), pp. 2–8.

Yokoya, D., & Yamada, T. (2011). A mathematical programming approach to the construction of bibds. International Journal of Computer Mathematics, 88(5), pp. 1067–1082.

Yuan, M., & Lin, Y. (2007). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B, 68(1), pp. 49–67.