mathematics

MDPI

Article

A Method for Augmenting Supersaturated Designs with Newly Added Factors

Chun-Wei Zheng [1,2,†], Zong-Feng Qi [1,†], Qiao-Zhen Zhang [2,†] and Min-Qian Liu [2,*]

[1] State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang 471003, China
[2] School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin 300071, China
* Correspondence: mqliu@nankai.edu.cn
† These authors contributed equally to this work.

Abstract: Follow-up experimental designs are popularly used in industry. In practice, some important factors may be neglected for various reasons in the first-stage experiment and they need to be added in the next stage. In this paper, we propose a method for augmenting supersaturated designs with newly added factors and augmented levels using the Bayesian *D*-optimality criterion. In addition, we suggest using the integrated Bayesian *D*-optimal augmented design to plan the follow-up experiment when the newly added factors have been allowed to vary in an appropriate region. Examples and simulation results show that the augmented designs perform well in improving identified rates of latent factor effects.

Keywords: Bayesian design; augmented design; variable selection; newly added factors

MSC: 62K05; 62L05

Citation: Zheng, C.-W.; Qi, Z.-F.; Zhang, Q.-Z.; Liu, M.-Q. A Method for Augmenting Supersaturated Designs with Newly Added Factors. *Mathematics* **2023**, *11*, 60. https://doi.org/10.3390/math11010060

Academic Editor: Elvira Di Nardo

Received: 24 November 2022
Revised: 17 December 2022
Accepted: 19 December 2022
Published: 23 December 2022

1. Introduction

In industrial applications, screening experiments are frequently utilized at the early stages of experiments which remove the negligible, or inactive, factors in further experiments. Due to their economy and flexibility, supersaturated designs (SSDs) are often used to plan a screening experiment when the number of factors is large and the experimental runs are expensive or time-consuming. SSDs introduced by Satterthwaite [1] are defined as having fewer runs than the effects to be estimated; the first systematic construction was provided by Booth and Cox [2] via the computer search. In the literature, most of the SSDs are used for experiments where main effect models are assumed. Since all factorial effects cannot be estimated simultaneously, and a variety of sub-models are identifiable, model selection can be achieved by many analysis methods, such as forward selection, stepwise, all-subsets regression, best-subset selection, simulated annealing model search (Wolters and Bingham [3]), partial least squares methods (Zhang et al. [4]; Yin et al. [5]), shrinkage methods, including smoothly clipped absolute deviation (SCAD, Li and Lin [6]) and Dantzig selector (Phoa et al. [7]), sure independence screening approach (Drosou and Koukouvinos [8]), group screening methods (Jones et al. [9]; Weese et al. [10]) and Bayesian methods (Beattie et al. [11]; Huang et al. [12]). Drosou and Koukouvinos [13] developed a new algorithm, the SVR-RFE, using the minimization of the weight vector as a criterion so as to detect the most significant variables. Some of these methods used in the paper will be introduced briefly in Section 2.2.

Screening experiments often leave some problems unresolved; thus, follow-up experiments are needed in such situations. For SSDs, Gupta et al. [14] augmented two-level designs with additional runs to create a new class of "extended $E(s^2)$-optimal" designs, and the idea was extended by Gupta et al. [15] to *s*-level designs; Qin et al. [16] studied the augmenting method for mixed-level SSDs and their method covers the work

of Gupta et al. [14,15] as two special cases. For uniform designs, Qin et al. [17] and Gou et al. [18] considered the augmenting strategy for two-level and three-level designs under the wrap-around L_2-discrepancy. In general, however, some information would surely have been obtained even if an experiment has failed to achieve the intended goal, and the information gathered should be carefully incorporated to plan the experiment in the next stage. Gutman et al. [19] adopted Bayesian D-optimality to add runs to existing SSDs using information from the initial experiment as prior information, and Zhang et al. [20] suggested using Bayesian D_s-optimality to augment the SSDs. All the above-mentioned studies are merely on row augmented designs.

However, in many follow-up designs, some additional factors may be added in the sequential stage since they may have been neglected in the first stage due to limited resources and knowledge or any other reasons, the aforementioned methods cannot handle this situation. Yang et al. [21] introduced an example from an industrial production showing the necessity of a column augmented design (CAD), the factor of reaction pressure is not considered to be significant initially. Thus, this factor was fixed as the standard atmosphere pressure (0.1 MPa), in order to limit the number of runs for saving cost. However, after analysis of the initial experiment, it was shown that the reaction pressure may be an important factor and need further investigation in the follow-up stage. In a word, there are many disappointing situations in an experiment. For instance, the present main effect model might not capture the nonlinear relationship between factors and response, so the level sizes of the factors should be augmented to at least three in such a situation. In the literature of uniform CADs, Yang et al. [21] proposed mixed two- and three-level uniform CADs under the wrap-around L_2-discrepancy. Liu et al. [22] further discussed the work of Yang et al. [21] under the Lee discrepancy, and Hu et al. [23] proposed augmented uniform q-level designs measured by the average mixture discrepancy. However, these methods did not use the information acquired from the previous experiments and did not augment the factors' level sizes. In this paper, Bayesian D-optimal criterion will be used to augment SSDs with newly added factors, and the analysis results from the initial-stage design can be included as prior information.

The next section reviews the relevant background firstly, then we propose the new augmenting strategies for SSDs using information from the initial runs in Section 3 for two cases; the levels of the newly added factors are fixed at some definite values or allowed to vary in some ranges. Section 4 displays the performances of the augmented designs by highlighting two examples, and Section 5 discuss the simulation performances. Some concluding remarks are provided in Section 6.

2. Preliminaries

The approach for developing Bayesian D-optimal designs in the context of the linear model will be briefly introduced firstly in this section; then, we are going to review some model selection methods that will be used later.

2.1. Bayesian D-Optimality

Consider the linear model

$$\boldsymbol{y} = \boldsymbol{\beta}_0\mathbf{1}_n + \boldsymbol{\beta}_1\boldsymbol{x}_1 + \cdots + \boldsymbol{\beta}_k\boldsymbol{x}_k + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of observations, $\boldsymbol{\beta}_0$ is the intercept term, $\mathbf{1}_n$ is an $n \times 1$ column vector with all elements unity, $\boldsymbol{x}_i$ is an $n \times 1$ vector of settings for the ith variable, $\mathbf{X}$ is the $n \times p$ model matrix with $p = k + 1$, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of coefficients to be estimated. Assume $\boldsymbol{\varepsilon}$ is the noise vector and $\boldsymbol{\varepsilon} \sim N(\mathbf{0_n}, \sigma^2\mathbf{I}_n)$, where $\mathbf{0_n}$ is an $n \times 1$ column vector with all elements zero, and $\mathbf{I}_n$ is an identity matrix of order n. Let the prior distribution of the parameters be $\boldsymbol{\beta} \mid \sigma^2 \sim N(\boldsymbol{\beta}_0, \sigma^2\mathbf{R}^{-1})$, where $\mathbf{R}$ is a prior covariance matrix, and the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{\beta}$ is $\boldsymbol{y} \mid (\boldsymbol{\beta}, \sigma^2) \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$. Then,

the posterior distribution for $\boldsymbol{\beta}$ given $\boldsymbol{y}$ would be $\boldsymbol{\beta} \mid \boldsymbol{y} \sim N\left(\mathbf{b}, \sigma^2(\mathbf{X}'\mathbf{X}+\mathbf{R})^{-1}\right)$, where $\mathbf{b} = (\mathbf{X}'\mathbf{X}+\mathbf{R})^{-1}(\mathbf{X}'\boldsymbol{y}+\mathbf{R}\boldsymbol{\beta}_0)$.

When the design of an experiment is used to collect data and perform estimation, the process can contain three stages generally. Firstly, select an optimal design D under some criteria; then, perform the experiment in terms of the design D and collect response $\boldsymbol{y}$; finally, consider an appropriate model and estimate the parameters $\boldsymbol{\beta}$ where $\mathbf{X}$ is the corresponding model matrix. If the model assumed is valid, the quality of design D determines the accuracy of parameters estimation. In this way, if we aim to find an optimal design to reduce $\mathrm{Var}(\boldsymbol{\beta} \mid \boldsymbol{y}) = \sigma^2(\mathbf{X}'\mathbf{X}+\mathbf{R})^{-1}$, the posterior variances of the parameter estimates in model (1), which can be accomplished by maximizing $|\mathbf{X}'\mathbf{X}+\mathbf{R}|$, and then the design acquired this way, is called a Bayesian D-optimal design.

2.2. Variable Selection Methods

Four variable selection methods will be employed to screen the significant effects later: the least absolute shrinkage and selection operator (LASSO) proposed by Tibshirani [24], the SCAD by Fan and Li [25], the minimax concave penalty (MCP) by Zhang [26] and the Dantzig selector (DS) by Candes and Tao [27].

The LASSO, MCP and SCAD are all members of the penalized least squares. A form of penalized least squares is defined as

$$(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}) + \sum_{j=1}^{k} p_\lambda(|\beta_j|),$$

where $p_\lambda(\cdot)$ is a penalty function, and λ is a tuning parameter.

The l_1 penalty $p_\lambda(|\beta|) = \lambda|\beta|$ results in LASSO. Although the LASSO has many attractive properties, the shrinkage introduced by the LASSO results in a significant bias toward zero for large regression coefficients, SCAD and MCP are proposed to diminish this bias. The SCAD defined on $[0,\infty)$ is given by

$$p'_\lambda(\theta) = \begin{cases} \lambda, & \text{if } \theta \le \lambda, \\ \frac{\gamma\lambda-\theta}{\gamma-1}, & \text{if } \lambda < \theta \le \gamma\lambda, \\ 0, & \text{if } \theta > \gamma\lambda, \end{cases}$$

for $\lambda \ge 0$ and $\gamma > 2$. The MCP defined on $[0,\infty)$ is given by

$$p'_{\lambda,\gamma}(\theta) = \begin{cases} \lambda - \frac{\theta}{\gamma}, & \text{if } \theta \le \gamma\lambda, \\ 0, & \text{if } \theta > \gamma\lambda, \end{cases}$$

for $\lambda \ge 0$ and $\gamma > 1$.

In the simulations, we select tuning parameter λ using a data-driven approach, 10-fold cross-validation (CV), and set the tuning parameter $\gamma = 3.7$ for SCAD and $\gamma = 3$ for MCP, the values suggested for linear regression in Fan and Li [25] and Breheny et al. [28]. In this paper, we apply coordinate descent algorithms to LASSO, SCAD and MCP regression models. We implement these algorithms through the publicly available R package "ncvreg".

The DS constitutes a popular shrinkage type of the variable selection method. While under some conditions the estimates from LASSO and DS may have similarities (James et al. [29]), they differ conceptually in that the LASSO is based on a regularized likelihood function, whereas the DS is based on an estimating equation. The estimator $\hat{\boldsymbol{\beta}}$ is the solution to

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \left\|\mathbf{X}^T(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})\right\|_\infty \le \delta, \tag{2}$$

where $\|\boldsymbol{\beta}\|_1 = \sum_{j=0}^{k}|\beta_j|$ is the l_1 norm, $\|\boldsymbol{a}\|_\infty = \max(|a_0|, \ldots, |a_k|)$ is the l_∞ norm, and δ is a tuning constant.

Candes and Tao [27] suggested rewriting (2) as a linear program solved by linear program solvers in the following generic form with inequality constraints:

$$\max_x (c + \lambda\bar{c})^\top x \quad \text{s.t.} \quad Ax \leq b + \lambda\bar{b}, \quad x \geq 0, \tag{3}$$

where $A = \begin{pmatrix} \mathbf{X}^\top\mathbf{X} & -\mathbf{X}^\top\mathbf{X} \\ -\mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{X} \end{pmatrix}$, $b = \begin{pmatrix} \mathbf{X}^\top\boldsymbol{y} \\ -\mathbf{X}^\top\boldsymbol{y} \end{pmatrix}$, $c = -\mathbf{1}$, $\bar{b} = \mathbf{1}$, $\bar{c} = \mathbf{0}$, $x = \begin{pmatrix} \boldsymbol{\beta}^+ \\ \boldsymbol{\beta}^- \end{pmatrix}$, $\boldsymbol{\beta}_j^+ = \boldsymbol{\beta}_j \cdot \mathbb{I}\left(\boldsymbol{\beta}_j > 0\right)$ and $\boldsymbol{\beta}_j^- = \boldsymbol{\beta}_j \cdot \mathbb{I}\left(\boldsymbol{\beta}_j < 0\right)$, and $\mathbb{I}(\cdot)$ is the indicator function.

In this paper, we solve the DS by applying the parametric simplex method (Pang et al. [30]) to the linear program and set the regularization factor $\lambda = 0.01$ in (3). The algorithm is conducted by function "dantzig" and "dantzig.selector" in R package "fastclime" where the number of the maximum path length is set to be 100. Note that, for the Dantzig selector variable selection method, the values $\lambda = 0.01$ and maximum path length 100 are based on multiple attempts; we also tried other values of λ such as 0.05, 0.1, 0.5 and so on, and the results are similar.

3. Optimal CADs

Generally, there may be a few factors influencing the performance of a system or process, however some factors may not be explored in the present design because of limited expenditure or time constraint, these factors have been set to a constant value or allowed to vary in an appropriate region temporarily. In fact, these potential factors may exert some effects on the response; the research on their impact would be restarted when the condition of follow-up experiments is conducive to adding new factors such as the availability of adequate funds. Section 3.1 deals with the CAD for held-constant factors, and Section 3.2 proposes an augmenting strategy to plan the next-stage experiment with new factors allowed to vary in some ranges separately.

Assume $D_1(n_1, k_1)$ is an initial two-level design with n_1 runs and k_1 factors adopted for the screening experiment in the first stage, each factor setting can be coded as 1 or -1, as usual. There are $r = k_2 - k_1$ new factors which need to be explored in the next experiment, as suggested by experts; $D_2(n_2, k_2)$ denotes the design that would be used to plan the experiment in the next stage with n_2 runs and k_2 factors, where $k_2 > k_1$.

3.1. Bayesian D-Optimal CADs

Firstly, a common situation is that each new factor is fixed at a certain value in the experiment of the previous stage. Therefore, it is reasonable to assume that researchers would explore their effects on the response in the region near the fixed value in the next-stage experiment; thus, the code "0" can be used to denote the fixed value, then the initial design can be written as $D_1^*(n_1, k_2) = (D_1, \mathbf{0}_{n_1 \times (k_2 - k_1)})$. Without further prior information, these new factors are set at three levels in the following design, which means that the levels of each new factor are at an equal distance around the fixed value.

Let $\mathbf{X_1}$ be a model matrix corresponding to the initial design $D_1^*(n_1, k_2)$ with n_1 runs and response vector $\boldsymbol{y_1}$, and $\mathbf{X_2}$ be the additional n_2 rows with response vector $\boldsymbol{y_2}$ for the second design D_2. That is

$$\mathbf{X} = \begin{pmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \end{pmatrix}. \tag{4}$$

Once the data from the first stage have been collected, variable selection methods can be employed to identify active factors and the information from the analysis may be used as a prior to plan the next-stage design. Gutman et al. [19] pointed out that the experimenters can classify a factor as a primary term (highlighted by an analysis method or several methods), secondary term (if there is an indication that the term may be active, but it is not predominant), or potential term (with little evidence to suggest it is active). Furthermore,

the prior covariance matrix $\mathbf{R}$ would be set in terms of the classification. Considering the situation that some existing experimental factors' level sizes may be augmented to three as well as new factors needing to be added, we modify the classification method to make it appropriate for mixed-level designs; the grouping is applied to the effects, not for factors, then quadratic effects can be involved.

The classification information for effects comes from two ways, data analysis for the initial design and advice given by experts or operators. For the former, we refer to the practice of Zhang et al. [20]: an effect is specified to the primary group which is considered as active with a high possibility if its identified rate is no less than 75% with multiple data-analysis methods. Those effects that have not been identified by any methods would be classified into the potential group, and the others whose identified rates belong to $(0, 75\%)$ constitute the secondary group. The threshold and analysis methods can be adjusted according to experience and background knowledge. For the second kind of information, some concern the factors included in the initial design and the others are for the new factors, and a blocking factor might be added to describe the potential system difference between the two stages. Anyway, group classification for effects would be enriched and adjusted, effects of new factors and those factors augmented to three levels suggested by experts would be specified to the primary group, some effects advised strongly by experts would be put into appropriate groups too, even if initial data analysis fails to give such an indication. Then, the prior variance can be set similarly as DuMouchel and Jones [31]; the coefficients for effects in the primary group are specified to have a diffuse prior variance tending to infinity, which implies that they are likely to be much different from zero. On the other hand, effects in the potential group are unlikely to be significant, and it is proper to assume that they have a relative small variance. For those terms in the secondary group, they may or may not be active, so their prior variance should be finite, but larger than that for the potential terms. We assume that the primary group contains an intercept and its size is p_1, and p_2 and p_3 are the sizes for the secondary group and potential group, respectively. We point out that the blocking effect should be classified in an appropriate group following advice from experts. When effects corresponding to $\mathbf{X}$ have been reordered in accordance with the sequence: the primary group, the secondary group and the potential group, then the prior covariance matrix $\mathbf{R}$ would be

$$\mathbf{R} = \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} & \mathbf{0}_{p_1 \times p_3} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{I}_{p_2}/\gamma^2 & \mathbf{0}_{p_2 \times p_3} \\ \mathbf{0}_{p_3 \times p_1} & \mathbf{0}_{p_3 \times p_2} & \mathbf{I}_{p_3}/\tau^2 \end{pmatrix}, \tag{5}$$

where $\tau^2 < \gamma^2$. Gutman et al. [19] set $\tau^2 = 5$ and $\gamma^2 = 100$.

Therefore, a Bayesian D-optimal column augmented design (BD-CAD) is created by choosing D_2 to maximize

$$|\mathbf{X_1}^T\mathbf{X_1} + \mathbf{X_2}^T\mathbf{X_2} + \mathbf{R}|. \tag{6}$$

The coordinate-exchange algorithm can be used to find the local or approximately global optimal design to maximize (6); the details are given in Algorithm 1.

Algorithm 1 Search for BD-CAD

1: Analyze initial design $D_1(n_1, k_1)$ under the linear main effect model, screen active factors with multiple methods, for example, the four variable selection methods mentioned above.
2: Combine professional advice and experiences with the results from the data analysis, determine which factors need to be expanded to three levels, classify effects into three groups and set the prior variance matrix $\mathbf{R}$ of parameters.
3: Obtain model matrix $\mathbf{X_1}$ for $D_1^*(n_1, k_2)$ under the model adjusted.
4: Generate a uniform random number from $[-1, 1]$ for each coordinate x_{ij} in $D_2(n_2, k_2)$, and obtain its corresponding model matrix $\mathbf{X_2}$. Then, compute the value of (6); this is the Bayesian D-optimal criterion value that we will improve by changing the design matrix $D_2(n_2, k_2)$ element by element.
5: Improve the resulting starting design on a coordinate-by-coordinate basis. Go through the n_2k_2 coordinates one by one, find the value in the set $\{-1, +1\}$ for two-level factors and $\{-1, 0, +1\}$ for factors augmented to three levels that optimize the Bayesian D-optimality criterion value in Equation (6). If the value in (6) is improved, the old value of the coordinate is replaced with the new one. Repeat this process until no more exchanges are made.
6: Repeat Step 4 and Step 5 M times; the design with the largest value of (6) in Step 5 is an approximate BD-CAD.
7: **return** $D_2(n_2, k_2)$, the design is a good choice to plan the next-stage experiment under the Bayesian D-optimal criterion.

3.2. Integrated Bayesian D-Optimal CADs

In some practical experiments, some factors that have a significant impact on the response may have not been controlled strictly at a certain level; for these allowed-to-vary factors, the procedure proposed in last subsection is not suitable. We assume these factors vary in a feasible range χ to make the first-stage experiment proceed successfully.

Though the levels of the $r = k_2 - k_1$ factors in the first stage are unknown, we can use $D_1^{**}(n_1, k_2) = (D_1, D_1^u)$ to denote the initial design, where

$$D_1^u = \begin{bmatrix} z_1 & \cdots & z_r \\ \vdots & \ddots & \vdots \\ z_{r(n_1-1)+1} & \cdots & z_{rn_1} \end{bmatrix} \tag{7}$$

with the elements unknown. However, z_i can be regarded as a random number in an interval. For each fixed setting $\mathbf{z} = (z_1, z_2, \ldots, z_{rn_1})$ corresponding to D_1^u, we can obtain a model matrix $\mathbf{X_1}(\mathbf{z})$ from the "design" D_1^{**} under some appropriate assumptions for the model; then, maximizing (6) under this setting can achieve an optimal CAD according to the procedure of the last subsection. However, what we are looking for is an augmented design with global optimality in a more general context, which should be relatively robust and independent of settings for D_1^u. Thus, we propose an integrated Bayesian D-optimal criterion over the region χ for new factors added,

$$\frac{\int \cdots \int_\chi f(z_1, z_2, \ldots, z_{rn_1}) \mathrm{d}z_1 \cdots \mathrm{d}z_{rn_1}}{\int \cdots \int_\chi \mathrm{d}z_1 \cdots \mathrm{d}z_{rn_1}}, \tag{8}$$

where

$$f(z_1, z_2, \ldots, z_{rn_1}) = |\mathbf{X_1}(\mathbf{z})^T \mathbf{X_1}(\mathbf{z}) + \mathbf{X_2}^T \mathbf{X_2} + \mathbf{R}| \tag{9}$$

is the determinant of the posterior covariance inverse matrix under a setting for D_1^u. The design D_2 maximizing (8) is an appropriate choice to plan the follow-up experiment for the situation of adding allowed-to-vary factors; the resulting final design can be called an integrated Bayesian D-optimal column augmented design (IBD-CAD).

In practical application, the multiple integration can be computationally complex for large rn_1; in such a situation, one can take N Halton sequences (Halton [32]), $\mathbf{z}_i = (z_1, z_2, \ldots, z_{rn_1})$ for $i = 1, \ldots, N$, to obtain an approximation of (8)

$$\frac{1}{N}\sum_{i=1}^{N} |\mathbf{X_1}(\mathbf{z}_i)^T \mathbf{X_1}(\mathbf{z}_i) + \mathbf{X_2}^T \mathbf{X_2} + \mathbf{R}| \tag{10}$$

using the quasi-Monte Carlo method. That is, an approximate integrated Bayesian D-optimal augmented design can be obtained by choosing D_2 to maximize (10); the final design produced by such a procedure can be called an approximate IBD-CAD.

4. Explanatory Examples

When it is discovered that a few factors ignored for some reasons in the preliminary experiment may influence the performance of a process or system, we would start to consider them in the next-stage experiment when some conditions are available. Choosing in which design to arrange the next-stage experiment depends on the type of factors added, held-constant or allowed-to-vary; in the previous trials, BD-CAD is suitable for the first case, and it is better to choose IBD-CAD for the latter. Two small-scale simulated examples would be used to demonstrate the procedure to find a BD-CAD and IBD-CAD with the proposed methods, respectively; the screening accuracy for the resulting CADs would also be explored. The number of factors ignored in the previous experiments are often assumed to be relatively small, so in the examples we consider that there are two held-constant or allowed-to-vary factors that may exert some effects on the response.

Example 1 (CAD for the held-constant factors)**.** *Assume that two factors are held-constant and the remaining thirteen factors in Table 1 are planned by a two-level $E(s^2)$-optimal SSD$(8, 13)$ presented in Gutman et al. [19] in the initial experiment. Table 1 is for the first-stage experiment where the code "0" denotes the specified value for two held-constant factors.*

Table 1. Initial design with thirteen factors planned by $E(s^2)$-optimal SSD$(8, 13)$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	y
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	37.79
2	1	1	1	−1	−1	1	−1	−1	−1	1	−1	−1	1	0	0	1	−8.31
3	1	−1	−1	−1	1	−1	1	−1	1	−1	−1	1	1	0	0	1	3.87
4	1	−1	1	1	1	−1	−1	−1	−1	−1	1	−1	−1	0	0	1	37.60
5	−1	1	−1	1	−1	−1	−1	1	1	−1	−1	−1	1	0	0	1	7.67
6	−1	1	−1	−1	1	−1	−1	1	−1	1	1	1	−1	0	0	1	19.34
7	−1	−1	1	−1	−1	1	1	1	1	−1	1	−1	−1	0	0	1	8.66
8	−1	−1	−1	1	−1	1	1	−1	−1	1	−1	1	−1	0	0	1	7.10

Suppose that the true model between response and the factors is

$$y = 8x_4 + 6x_5 + 9x_{11} + 7x_{14} + 10x_{11}^2 + \delta x_{16} + \varepsilon, \qquad \varepsilon \sim N(0, 1), \tag{11}$$

where δ denotes the blocking effect which illustrates the shift between the two stages when one expects the response to experience a drift over time, and x_{16} equals $+1$ in the initial design and -1 in the augmented design. Consider there is a moderate blocking effect, $\delta = 4$, here. Note the factor x_{14} is assumed to be active in the true model, though it is not considered as a design factor in the initial experiment. In addition, assume the curvature exists in the relationships between the 11th factor and the response. The responses generated from model (11) are listed in Table 1 too.

Since the initial design is an SSD, the main effect model would be applied to analyze the design. Certainly, the blocking effect, the significance of two held-constant factors and the quadratic effect of x_{11}^2 cannot be explored temporarily. Four methods mentioned in

Section 2 can be used to screen active factors from this SSD$(8, 13)$. It turns out that three factors x_4, x_5, x_{11} are identified as significant by all four methods, and none of the other factors have been detected.

It is assumed that after a period of time, the conditions for starting the next-stage experiment have been met. Suppose some experts combine the results of data analysis with the experience of operators and give advice as follows: for the 3rd and 11th factor, at least three settings are required in order to detect the curvature on the system. Moreover, the two held-constant factors may have a significant effect on the response, so they should be handled as design factors and their effects in the area near their specified values should be detected. In addition, there may exist a blocking effect and we classify it into the secondary group. Considering these demands and the screening results from the initial design, the effects are divided into three groups,

$$\begin{aligned}&\text{primary group: } \{\text{intercept}, \beta_3, \beta_4, \beta_5, \beta_{11}, \beta_{14}, \beta_{15}, \beta_3^2, \beta_{11}^2, \beta_{14}^2, \beta_{15}^2\},\\ &\text{secondary group: } \{\delta\},\\ &\text{potential group: } \{\beta_1, \beta_2, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{12}, \beta_{13}\}.\end{aligned}$$

Based on the classification, the prior covariance matrix $\mathbf{R}$ can be determined.

The main effect model would still be used to analyze the CAD; note that for three-level factors, the quadratic effects would also be included in the model as well as the linear effects,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{15} x_{15} + \beta_{3,3} x_3^2 + \beta_{11,11} x_{11}^2 + \beta_{14,14} x_{14}^2 + \beta_{15,15} x_{15}^2 + \delta x_{16} + \varepsilon. \quad (12)$$

We employ Algorithm 1 with 1000 initial random designs to add $n_2 = 7$ runs to obtain the BD-CAD, the augmented seven runs are shown in Table 2. In Table 2, factors x_3 and x_{11} are already at three levels in the second stage, and the settings of these two factors are relatively symmetrical, each with two high and low levels and three intermediate levels, while for the newly added factors x_{14} and x_{15}, the intermediate levels are relatively less . Considering that the two factors x_{14} and x_{15} were fixed at the intermediate level in the previous experiment, from the formal point of view, this additional design is quite reasonable.

Table 2. Additional seven runs of the BD-CAD for the SSD$(8, 13)$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	y
9	−1	−1	1	−1	1	1	−1	1	1	1	−1	−1	−1	1	−1	−1	2.33
10	1	1	0	1	−1	−1	1	−1	−1	−1	1	1	1	0	−1	−1	16.19
11	−1	−1	1	−1	1	1	−1	1	1	1	0	−1	−1	0	1	−1	−6.14
12	1	1	−1	1	−1	−1	1	−1	−1	−1	1	1	1	1	1	−1	24.38
13	−1	−1	0	−1	1	1	−1	1	1	1	−1	−1	−1	−1	1	−1	−10.82
14	−1	−1	0	−1	1	1	−1	1	1	1	0	−1	−1	1	0	−1	2.08
15	1	1	−1	1	−1	−1	1	−1	−1	−1	0	1	1	−1	−1	−1	−8.02

After obtaining the augmented seven runs of CAD, a set of random numbers is generated as observations for response according to the model (11). Then, for all 15 "observations", the four methods mentioned in Section 2.2 are applied to select significant effects; all methods identify effects $\beta_4, \beta_5, \beta_{11}, \beta_{14}, \beta_{11,11}, \delta$ (blocking effect) as significant effects, and no other effect is selected by any method.

Example 2 (CAD for the allowed-to-vary factors)**.** *Assume x_{14}, x_{15} are allowed-to-vary in the initial stage; the initial design and model are still the same as those in Example 1. Note that one can observe the response when there are "uncontrollable" factors in practice, but we cannot obtain the random numbers as observations for response without specific values of these factors when doing simulations. Here, for the sake of simplicity, we use mean interpolation, that is, we take a fixed value*

of 0 as the value of the two allowed-to-vary factors in the initial stage; thus, we can still use the effect classification result to augment the row and column directly.

The assumptions of advice from experts and operators are similar to before except that we suppose that no mean shift exists here. The responses in the first-stage design are 32.53, −14.22, −1.68, 34.63, 3.06, 16.18, 6.62 and 4.05, in order. According to Section 3.2, we can add rows and columns according to the IBD-CAD criterion; the augmented seven runs with $M = 200$ and $N = 100$ are shown in Table 3.

Table 3. Additional seven runs of the IBD-CAD for the SSD$(8, 13)$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	y
9	1	−1	−1	−1	−1	−1	1	−1	−1	−1	1	1	1	−1	−1	−0.17
10	−1	1	1	−1	1	−1	1	−1	1	−1	−1	1	−1	−1	1	6.07
11	1	−1	−1	1	1	1	−1	−1	1	1	0	−1	1	1	1	21.54
12	−1	1	0	−1	1	−1	1	1	1	−1	−1	1	−1	1	−1	−8.98
13	−1	−1	0	1	1	1	−1	1	1	1	0	−1	−1	−1	0	13.37
14	−1	1	1	−1	−1	−1	1	−1	−1	−1	0	1	1	0	−1	−22.14
15	1	1	0	−1	−1	−1	1	−1	−1	−1	1	1	1	0	1	11.34

After obtaining the IBD-CAD and following the analysis procedures like before, the truly significant effects $\beta_4, \beta_5, \beta_{11}, \beta_{14}, \beta_{11,11}$ are identified as significant effects by all the four methods and no other effect is selected by any method.

In practical application, the appropriate augmenting scheme is selected according to the value of the newly added factors. When each newly added factor has been fixed at a value accurately in the first-stage experiment, the pattern of BD-CAD is a great fit. However, in many cases, the newly added factors may change in some regions, or we can not know their values in the first stage for various reasons, or there may be some measurement errors in some practical experiments; the IBD-CAD may be more proper for such situation. That is, the IBD-CAD has a wider range of applications and a simple scale simulation can show that its performance is also satisfying.

Assume the initial experiment with thirteen factors has been planned by the design given in Example 1 and the model is the same; however, when the next phase of the experiment can be carried out, the initial settings of the newly added factors have been lost. Now one can have three choices to arrange the next-stage experiment, column augmenting with D-optimal criterion (D-CAD), which is to maximize $|\mathbf{X_1}^T\mathbf{X_1} + \mathbf{X_2}^T\mathbf{X_2}|$ for model (12), BD-CAD and IBD-CAD. All these criteria are aimed at optimizing $\mathbf{X_2}$. In the analysis of significant effects, the settings of the newly added factors in the initial stage are all plugged in as 0. A total of 300 repetitions are performed and the effects identification results from the comprehensive designs are given in Table 4. Here, we take $M = 1000$ to obtain the D-CAD and BD-CAD, and $M = 100$ and $N = 100$ to obtain the IBD-CAD.

Table 4. Summary of simulation results from different CADs with seven runs added.

	AEIR	**IEIR**
D-CAD	0.433	0.163
BD-CAD	0.750	0.030
IBD-CAD	0.653	0.023

In Table 4, two measurements are employed to evaluate the performances: average of active effect identified rate (AEIR) and average of inactive effect identified rate (IEIR). We use the four aforementioned screening methods to identify significant effects. When an effect is selected by three or four methods, we regard it as significant. We calculate the AEIR by the average of the frequencies of the regarded significant effects which are all truly

significant effects in the model, and calculate the IEIR by the average of the frequencies of regarded significant effects which are not in the true model. Clearly, AEIR is the larger the better, IEIR is the smaller the better. The performance of the D-CAD, as expected, is worse compared with the other two procedures since this augmenting method does not make the most of the information from the initial design, like the BD-CAD and IBD-CAD. The BD-CAD has the highest AEIR since the augmenting plan is appropriate for this situation where new factors are fixed at some values. However, when using IBD-CAD, we believe that there is some uncertainty about the preliminary values of the new factors, so here the approximated IBD-CAD has no advantage over the identification of active factors when new factors have been fixed exactly. However, the IBD-CAD has the smallest IEIR among the three CADs. Certainly, the AEIR for the approximated IBD-CAD is also acceptable and it is 22% higher than the D-CAD.

Remark 1. *Here, we emphasize that quadratic effects have been classified into the primary group since they matter greatly to those factors augmented to three levels. That is also quite natural. Because we want to explore the existence of curvatures, the corresponding factors are expanded to three levels. In fact, if we put them into the secondary or potential group in the simulation, it is likely that the middle level 0 will appear less or even none. That may be because if we classify the quadratic effects into the secondary or potential group, it is considered that there is no evidence or suggestion that the quadratic effects are important, so few middle level 0s will be allocated to the augmented design to detect the quadratic effects.*

Remark 2. *For the run size of the augmented design n_2, suppose that in the second phase of the experiment, the number of parameters to be estimated in the model is m. Based on massive simulations, we recommend that n_2 is not less than $m - n_1$; fewer runs would result in undesirable identification. If we add more runs in Example 2, Table 4 would have more delicate results.*

Remark 3. *In coordinate-exchange algorithms, the numbers of random starting designs M and Halton sequences N need to be determined. For BD-CADs, the algorithm runs for 0.82 seconds if $M = 100$ and 6.49 seconds if $M = 1000$; for IBD-CADs, the algorithm runs for 10.5 seconds when $M = 10$ and 52.34 seconds when $M = 100$ if we fix $N = 100$. The larger M, the more accurate the criteria values (6) and (10); considering the time cost, we also carried out some simulations for other values in the example: for the two kinds of augmented designs, we find that $M = 100$ is enough and the impact of M on the identification of significant effects is not apparent. Similarly, the larger N, the more accurate the integral approximation, but too large an N is a waste of time, so we suggest $N = 100$. The system used for the simulations was an AMD Ryzen 7 4800H using parallel execution provided by R package, parallel and foreach. The OS configuration was Windows 10.*

5. Simulation Results

In this section, we carry out simulations to study the performances of the BD-CAD and IBD-CAD, several analysis methods are used together to judge the power of various designs in the simulations. The details of the general simulation protocol will be described firstly.

1. The initial designs. Two $E(s^2)$-optimal designs, the SSD$(8, 13)$ in Table 1 and SSD$(7, 15)$ presented in Gutman et al. [19], are taken to be the first-stage designs; the latter is shown in Table 5.

Table 5. $E(s^2)$-optimal SSD$(7,15)$ presented in Gutman et al. (2014) [19].

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	−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	1
3	−1	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	−1
5	1	−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	1
7	1	−1	1	1	−1	1	1	1	1	−1	1	−1	1	1	1

2. The choice of significant effects in the true model.
   (a) When the SSD$(8,13)$ is assumed as the initial design, we consider the true model without interaction terms. Three of these 13 factors are randomly selected as active factors, and their linear effects and one randomly selected quadratic effect are assumed significant. Among the r newly added factors, one factor is randomly selected and its linear effect is assumed significant.
   (b) When SSD$(7,15)$ is assumed as the initial design, a significant two-factor interaction will be included in the true model here. Similar to the last scenario, three factors are randomly selected as active factors, and their linear effects are assumed significant. For the significant interaction effect, at least one of its parent main effects should be significant under the effect heredity principle. Thus, here, its significant parent effect is selected randomly from the three active factors and the other insignificant parent effect is randomly selected from the rest factors. In addition, among the newly added factors, assuming a randomly selected quadratic effect is significant.
3. The experts' advice. Based on the analysis results of the initial experiment and their experience, the experts may give advice in the following aspects: which factors need to be studied at three levels because one may expect substantial curvature, which new factors need to be added in the next-stage experiment, and whether the response experiences a drift over time. We believe that most of the experts' suggestions are reasonable, but they are not all entirely correct. In view of this, we set the following experts' recommendations.
   (a) For 2(a), besides the new factors, two factors need to be studied at three levels: one is active, the other one is inactive. Both are selected randomly.
   (b) For 2(b), we consider a more complicated situation where several experts give different advice about two-factor interactions. First, they advise that the two-factor interaction term in the true model is significant. Then, a few two-factor interactions, each with one factor whose main effect is neither active nor significantly identified, are suggested to be explored in the next-stage experiment, as for the other factors' effects among these two-factor interactions, two of them inherit a different active parent main effect, respectively, and the others inherit the effects in the primary group as parent main effects. Therefore, the number of two-factor interactions advised by the experts is three plus the number of effects in the primary group. In addition, one main effect, which is in fact inactive, is suggested to be significant. A randomly selected active factor and the new factors need to be augmented to three levels.
   (c) The blocking effect needs to be detected and it is classified into the primary group.
4. The number of newly added factors, r. We add $r = 1,2,3,4$ three-level factors, respectively, in the second stage.
5. The number of added runs n_2. For 2(a), the model is simple and there is no interaction, so we add $n_2 = 6,9,12$ runs in the augmented designs in the simulations. However,

for 2(b), $n_2 = 9, 12$, and 15, follow-up runs will be planned in the next-stage designs since the model is more complicated than the former one.

6. The size of significant effects including the blocking effect. The regression coefficients are drawn from $\boldsymbol{\beta} \sim N(\mu, 1)$, $\mu = 4, 6, 8$, respectively.
7. Analysis methods. The four methods aforementioned in Section 2.2 are used to identify effects. Effects selected by three or four methods are regarded significant.
8. The blocking effect is significant.

In our simulations, the levels of the newly added factors are set to be 0 in the first-stage experiment, and we think it is reasonable for the following two reasons. If the levels of the newly added factors are determined to be 0 in the initial experiment, certainly, the first augmenting scheme BD-CAD is more reasonable. However, considering the measurement error or disturbance in the experiment, the second follow-up scheme IBD-CAD can also be used. On the other hand, when the accurate levels of the newly added factors in the first stage are unknown, the second augmenting procedure is more reasonable, but the first scheme is also feasible in consideration of interpolation with the middle value in its range of variation.

In general, the BD-CAD or approximated IBD-CAD searching with coordinate-exchange algorithm is a locally optimal design. To make the likelihood of finding the globally optimal design large, one should repeat the exchange algorithm a large number of times. For our simulation, only a few factors are involved and the number of follow-up runs is small; too many random starts are not necessary. In our simulation, the number of random starts $M = 100$ and the number of Halton sequences $N = 100$ can produce desirable outcomes, so here we do not make further investigation. Certainly, for general cases, to ensure that either we find the globally optimal design or a design with an optimal criterion value that is very close to the global optimum, a large number of random starts is needed.

For the sake of understanding and avoiding redundancy, we only take the settings of the SSD$(8, 13)$ in 2(a), 3(a) as an example to elaborate the simulations. In each of the 300 iterations, the whole procedure is as follows.

1. From the columns of the SSD$(8, 13)$, three columns are randomly assigned as the active factors; then, one column from the three is randomly selected to compute the quadratic term. The blocking factor equals $+1$ in the initial design and -1 in the augmented design. The coefficients for three linear main effects, one quadratic effect and a blocking effect are obtained by sampling from $N(\mu, 1)$.
2. The remaining 10 columns (inactive effects) in the SSD$(8, 13)$ are assigned a coefficient of exactly zero.
3. r column vectors are added with all elements zero to the SSD$(8, 13)$, one column is randomly selected and its coefficient for linear effect is obtained by sampling from $N(\mu, 1)$.
4. The response vector is generated from the model in (1) with errors ϵ_i generated from $N(0, 1)$.
5. The analysis methods introduced in Section 2.2 are used to identify the active factors, then the effects grouped according to the results and suggestions from experts.
6. The prior covariance matrix R is assigned, then n_2 runs are generated as described in Sections 3.1 and 3.2; we can obtain the BD-CAD and approximated IBD-CAD with two procedures.
7. The response vector for the second-stage design is generated as before.
8. The four methods are used to identify the significant effects for all observations from the 'big' design, which combines the initial design with the augmented design, the declared 'significant' effects are determined by at least three analysis methods.

At the end of 300 iterations, the performances are reported in Figures 1–3. Since, AEIR is the-larger-the-better, and IEIR is the-smaller-the-better, in the figures, the higher solid line is the better, since it means that the corresponding augmenting procedure can identify all the active effects with a higher probability under the same analysis methods. However,

the lower dotted line is the better, the design corresponding to the lower dotted line may screen out the inactive effects with a lower probability.

Figure 1 shows the screening results for augmented experimental data with r, with the number of newly added factors in the second stage increasing. First of all, we note that there is no significant difference between two augmenting methods, especially for the SSD$(8, 13)$. Furthermore, as expected, the difficulty of identifying active effects from more factors will increase with the same number of experimental runs; both procedures perform worse in terms of AEIR when there are more factors added in the next-stage design providing the same magnitude of coefficients. However, IEIR begins to show a weak upward trend when r increases from 2 to 4, meaning that the screening procedure is a bit cautious for $r = 1$; to avoid missing some active effects, IEIR tends to include more factors in the model. For the SSD$(7, 15)$ with a bigger model including two-factor interactions, we take the settings as 2(b), 3(b) and the simulation procedure is similar to the procedure of the SSD$(8, 13)$. The second augmenting plan shows a number of advantages, though they are not particularly obvious. At the same time, we note that in order to achieve the same level for AEIR, a relatively complex model and economical initial design may require more follow-up runs, especially as here, many suggestions from experts are improper and the model considered for the augmented design may not follow the effect sparsity principle. Too many inappropriate suggestions create significant uncertainty; the IBD-CAD performs a little bit better.

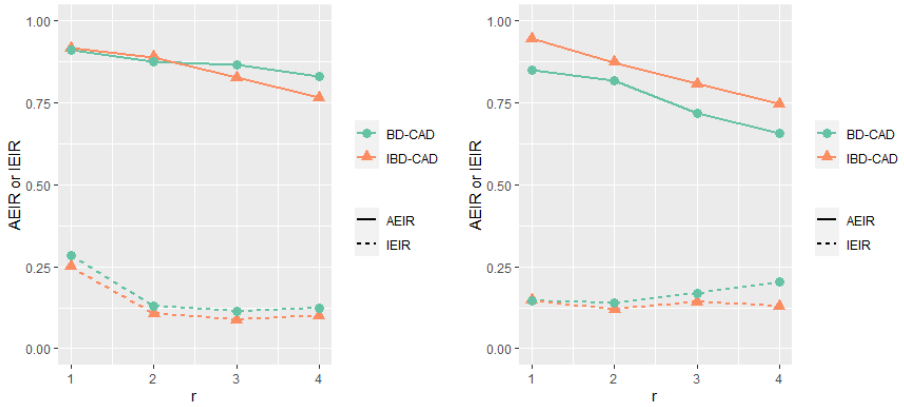
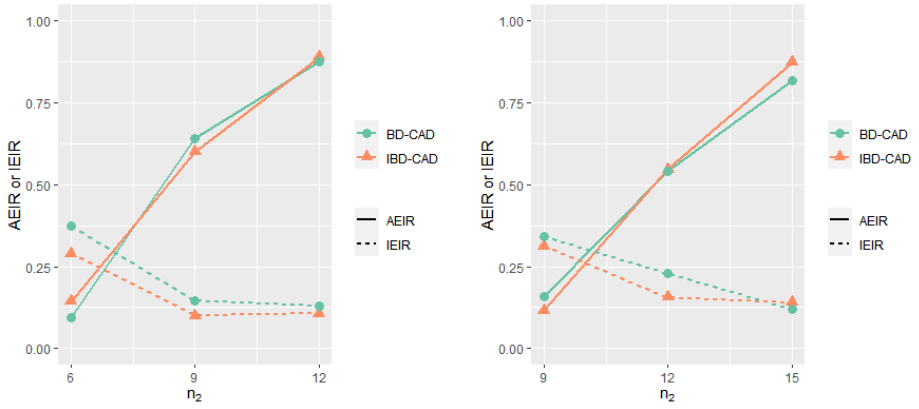


Figure 1. AEIR and IEIR versus the number of newly added factors r for $\boldsymbol{\beta} \sim N(6, 1)$. (**a**) SSD$(8, 13)$, $n_2 = 12$. (**b**) SSD$(7, 15)$, $n_2 = 15$.

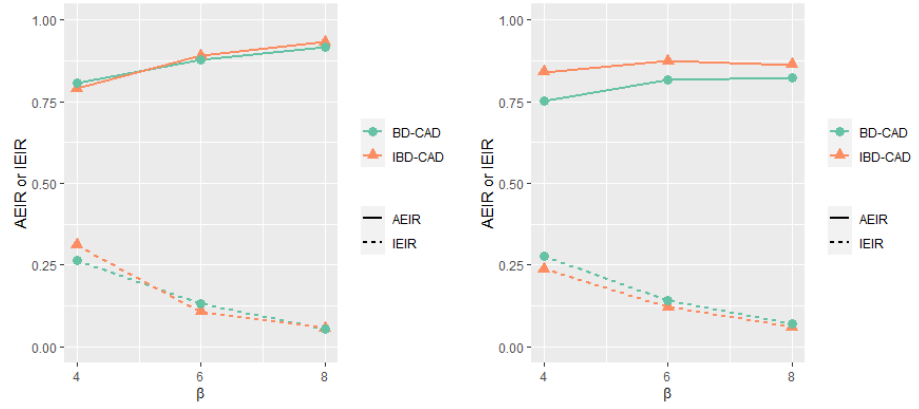
Figure 2 shows that the performances of the two augmenting methods tend to be better with more runs added and they perform similarly since two augmented designs do not show obvious differences from AEIR and IEIR. Considering the simplicity of calculation and the fact it is time-consuming, the BD-CAD is more suitable. For the SSD$(8, 13)$, both augmenting plans perform badly in terms of AEIR or IEIR when adding six runs; the active effects missed, in most cases, are quadratic terms; sometimes, the effects of the newly added factors cannot be screened out. The reason is that if we just augment six runs on the basis of the SSD$(8, 13)$, the degrees of freedom are much less than the number of effects we need to identify which include all the main effects in the initial analysis, four quadratic effects and a blocking effect. However, the screening results improve soon when $n_2 = 9$, and Table 4 also tells us that the two augmenting plans performs well for coefficients a little bit larger, when $n_2 = 7$. The situation is roughly the same for the SSD$(7, 15)$; adding fewer runs makes it difficult to miss none of the active effects; more runs need to be added since we are facing a more complicated model.



(**a**) (**b**)

Figure 2. AEIR and IEIR versus the number of augmented design runs n_2 for $\boldsymbol{\beta} \sim N(6,1)$, $r = 2$. (**a**) SSD$(8,13)$. (**b**) SSD$(7,15)$.

Figure 3 shows the influence of the model coefficients. The screening performances reveal that our augmenting methods are effective since both AEIR and IEIR are at the acceptable levels. In addition, AEIR and IEIR for the two SSDs show different trends with the model coefficients increasing, AEIR grows faster and IEIR falls faster too for the SSD$(8,13)$. The assumption of the model for the design is simple; thus, the screening results improve quickly when the coefficients increase a little. However, for the SSD$(7,15)$, the model size is large and we assume several second-order interaction effects which are not significant to be included in the model, so although AEIR improves with the increasing of the coefficients, it is not as fast as that for the SSD$(8,13)$.



(**a**) (**b**)

Figure 3. AEIR and IEIR versus the size of significant effects $\boldsymbol{\beta}$ for $r = 2$. (**a**) SSD$(8,13)$, $n_2 = 12$. (**b**) SSD$(7,15)$, $n_2 = 15$.

As far as IEIR is concerned, IEIR values obtained by IBD-CADs are generally lower in three figures; thus, we conclude that the IBD-CAD is preferable to BD-CAD in most circumstances. Especially, in Figures 1b and 3b, all AEIR values of IBD-CADs are higher than those obtained by BD-CADs and all IEIR values of IBD-CADs are lower than those obtained by BD-CADs, so the IBD-CAD is more suitable for the second-stage experiment in the case of the initial design SSD$(7,15)$ with the true model including a two-factor interaction.

6. Concluding Remarks

Traditional SSDs can provide the low-cost identification for main effects models under the effect sparsity assumption, that is the reason that we use them as the initial designs. However, the screening probably results in a less-than-ideal performance which cannot identify the active factors correctly.

Further, some additional factors ignored or not considered as design factors in the early experiments may be added now. Therefore, it is necessary to proceed with the column-augmented follow-up design. Comparing with previous researches, we first develop a method to generate follow-up designs in terms of row augmenting, column augmenting, augmenting-factors' level sizes and using the information acquired from the previous experiments to guide the follow-up designs. The previous researches only considered one or two perspectives. In this paper, we propose the BD-CAD and IBD-CAD in light of different situations: that the newly added factors have been fixed at certain values and allowed to vary in ranges, respectively. In addition, we augment the levels of some specified factors to detect the curvature effects. The initial design and augmented design can be integrated into a prior information-guided factor screening and factor addition experiment which can be applied to the identification of significant effects, including hereditary effects and blocking effects under the second-order model, for the sake of reducing cost.

In practice, we apply BD-CAD or IBD-CAD according to different contexts, as discussed in Section 3: BD-CAD is applicable for held-constant factors and IBD-CAD is for planning next-stage experiment when factors vary in an interval or range separately. However, in the simulations, when the new factors are allowed to vary in some ranges separately, we cannot simulate this situation exactly, so we take a fixed value of 0 as the values of the allowed-to-vary factors in the initial stage, just like the held-constant factors. As shown by the simulation results, for different evaluation indicators, the preferable designs are different. If the time cost is what we consider, BD-CAD is more suitable; if the IEIR is what we consider, IBD-CAD is more suitable.

For the IBD-CAD, the multiple integration is approximated by the quasi-Monte Carlo method in terms of N Halton sequences. However, a uniform design and Latin hypercube design can also help to realize the approximation of integration and the work is in progress.

As highlighted by one referee, one idea is to take Laplace distribution as the noise vector $\boldsymbol{\varepsilon}$'s distribution instead of normal distribution. We considered that the noise vector takes Laplace distribution with location parameter 0 and scale parameter 1; there is no significant difference between Laplace distribution and normal distribution in terms of AEIR and IEIR. For example, by taking the above Laplace distribution, we set $r = 12$, $n_2 = 12$, $\boldsymbol{\beta} = 6$ and iterate 300 times. If we use BD-CADs, AEIR = 0.83 and IEIR = 0.26. If we use IBD-CADs, AEIR = 0.88 and IEIR = 0.22.

The coordinate-exchange algorithm runs in polynomial time, which means that the time it needs to find an optimal design does not increase when the size of the design and the number of factors increase. Besides the cost of the coordinate-exchange algorithm, for the BD-CAD, it takes a little time to screen active factors in the initial stage which can be ignored; for the IBD-CAD, it takes $n_1k_2^2 + n_2k_2^2$ times more than the BD-CAD since we need to calculate (9) using Halton sequences. The memory is available for storing the matrix $\mathbf{X_1}$ and $\mathbf{X_2}$. We provide the codes at https://github.com/apple369/CADesign accessed on 17 November 2022.

Author Contributions: C.-W.Z.: conceptualization, methodology, formal analysis, software, visualization, writing—original draft preparation; Z.-F.Q.: resources, project administration, funding acquisition; Q.-Z.Z.: methodology, formal analysis, writing—review and editing, supervision; M.-Q.L.: methodology, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the State Key Laboratory of CEMEE (CEMEE2020K0301A), the National Natural Science Foundation of China (Grant Nos. 12131001 and 12226343), National Ten Thousand Talents Program, and the 111 Project B20016.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Satterthwaite, F.E. Random balance experimentation. *Technometrics* **1959**, *1*, 111–137. [CrossRef]
2. Booth, K.H.; Cox, D.R. Some systematic supersaturated designs. *Technometrics* **1962**, *4*, 489–495. [CrossRef]
3. Wolters, M.A.; Bingham, D. Simulated annealing model search for subset selection in screening experiments. *Technometrics* **2011**, *53*, 225–237. [CrossRef]
4. Zhang, Q.Z.; Zhang, R.C.; Liu, M.Q. A method for screening active effects in supersaturated designs. *J. Stat. Plan. Inference* **2007**, *137*, 2068–2079. [CrossRef]
5. Yin, Y.H.; Zhang, Q.Z.; Liu, M.Q. A two-stage variable selection strategy for supersaturated designs with multiple responses. *Front. Math. China* **2013**, *8*, 717–730. [CrossRef]
6. Li, R.; Lin, D.K.J. Data analysis in supersaturated designs. *Stat. Probab. Lett.* **2002**, *59*, 135–144. [CrossRef]
7. Phoa, F.K.; Pan, Y.H.; Xu, H. Analysis of supersaturated designs via the Dantzig selector. *J. Stat. Plan. Inference* **2009**, *139*, 2362–2372. [CrossRef]
8. Drosou, K.; Koukouvinos, C. Sure independence screening for analyzing supersaturated designs. *Commun. Stat.-Simul. Comput.* **2009**, *48*, 1979–1995. [CrossRef]
9. Jones, B.; Lekivetz, R.; Majumdar, D.; Nachtsheim, C.J.; Stallrich, J.W. Construction, properties, and analysis of group-orthogonal supersaturated designs. *Technometrics* **2020**, *62*, 403–414. [CrossRef]
10. Weese, M.L.; Stallrich, J.W.; Smucker, B.J.; Edwards, D.J. Strategies for supersaturated screening: Group orthogonal and constrained var(*s*) designs. *Technometrics* **2020**, *63*, 443–455. [CrossRef]
11. Beattie, S.D.; Fong, D.K.H.; Lin, D.K.J. A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics* **2002**, *44*, 55–63. [CrossRef]
12. Huang, H.; Yang, J.; Liu, M.Q. Functionally induced priors for componentwise Gibbs sampler in the analysis of supersaturated designs. *Comput. Stat. Data Anal.* **2014**, *72*, 1–12. [CrossRef]
13. Drosou, K.; Koukouvinos, C. A new variable selection method based on SVM for analyzing supersaturated designs. *J. Qual. Technol.* **2019**, *51*, 21–36. [CrossRef]
14. Gupta, V.K.; Singh, P.; Kole, B.; Parsad, R. Addition of runs to a two-level supersaturated design. *J. Stat. Plan. Inference* **2010**, *140*, 2531–2535. [CrossRef]
15. Gupta, V.K.; Chatterjee, K.; Das, A.; Kole, B. Addition of runs to an *s*-level supersaturated design. *J. Stat. Plan. Inference* **2012**, *142*, 2402–2408. [CrossRef]
16. Qin, H.; Chatterjee, K.; Ghosh, S. Extended mixed-level supersaturated designs. *J. Stat. Plan. Inference* **2015**, *157*, 100–107. [CrossRef]
17. Qin, H.; Gou, T.; Chatterjee, K. A new class of two-level optimal extended designs. *J. Korean Stat. Soc.* **2016**, *45*, 168–175. [CrossRef]
18. Gou, T.; Qin, H.; Chatterjee, K. Efficient asymmetrical extended designs under wrap-around L_2-discrepancy. *J. Syst. Sci. Complex.* **2018**, *31*, 1391–1404. [CrossRef]
19. Gutman, A.J.; White, E.D.; Lin, D.K.J.; Hill, R.R. Augmenting supersaturated designs with Bayesian *D*-optimality. *Comput. Stat. Data Anal.* **2014**, *71*, 1147–1158. [CrossRef]
20. Zhang, Q.Z.; Dai, H.S.; Liu, M.Q.; Wang, Y. A method for augmenting supersaturated designs. *J. Stat. Plan. Inference* **2019**, *199*, 207–218. [CrossRef]
21. Yang, F.; Zhou, Y.D.; Zhang, A. Mixed-level column augmented uniform designs. *J. Complex.* **2019**, *53*, 23–39. [CrossRef]
22. Liu, J.; Ou, Z.; Hu, L.; Wang, K. Lee discrepancy on mixed two- and three-level uniform augmented designs. *Commun. Stat.-Theory Methods* **2019**, *48*, 2409–2424. [CrossRef]
23. Hu, Z.; Liu, J.; Li, Y.; Li, H. Uniform augmented *q*-level designs. *Metrika* **2020**, *84*, 969–995. [CrossRef]
24. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [CrossRef]
25. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
26. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [CrossRef]
27. Candes, E.; Tao, T. The Dantzig selector: Statistical estimation when *p* is much larger than *n*. *Ann. Stat.* **2007**, *35*, 2313–2351.
28. Breheny, P.; Huang, J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **2011**, *5*, 232–253. [CrossRef]
29. James, G.M.; Radchenko, P.; Lv, J. DASSO: Connections between the Dantzig selector and lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2009**, *71*, 127–142. [CrossRef]

30. Pang, H.; Vanderbei, R.J.; Liu, H.; Zhao, T. Parametric simplex method for sparse learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 188–197.
31. DuMouchel, W.; Jones, B. A simple Bayesian modification of *D*-optimal designs to reduce dependence on an assumed model. *Technometrics* **1994**, *36*, 37–47.
32. Halton, J.H. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.* **1960**, *2*, 84–90. [CrossRef]