Technometrics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

Selection of Two-Level Supersaturated Designs for Main Effects Models

Rakhi Singh & John Stufken

To cite this article: Rakhi Singh & John Stufken (2023) Selection of Two-Level Supersaturated Designs for Main Effects Models, Technometrics, 65:1, 96-104, DOI: 10.1080/00401706.2022.2102080

To link to this article: https://doi.org/10.1080/00401706.2022.2102080

TECHNOMETRICS
2023, VOL. 65, NO. 1, 96–104
https://doi.org/10.1080/00401706.2022.2102080

Selection of Two-Level Supersaturated Designs for Main Effects Models

Rakhi Singh and John Stufken

University of North Carolina at Greensboro, Greensboro, NC

ABSTRACT

An extensive literature is available on design selection criteria and analysis techniques for two-level supersaturated designs. The most notable design selection criteria are the popular $E(s^2)$-criterion, $UE(s^2)$-criterion, and more recently, the var(s+)-criterion, while the most notable analysis technique is the Gauss-Dantzig Selector. It has been observed that while the Gauss-Dantzig Selector is often the preferred analysis technique, differences in the screening performance of different designs are not captured well by any of the common design selection criteria. In addition, none of the criteria have any direct connection to the Gauss-Dantzig Selector. We develop two new design selection criteria inspired by large sample desiderata of the Gauss-Dantzig Selector. Then, using a multi-objective Pareto-based coordinate exchange algorithm, we find Pareto eficient designs. The obtained Pareto eficient designs perform better in about 85% of the considered cases as screening designs than the var(s+)-optimal designs, especially when the true signs of effects are known. For the remaining 15% of the cases as well as for the unknown effect signs, the Pareto eficient designs perform at par with the var(s+)-optimal designs.

ARTICLE HISTORY
Received August 2021
Accepted June 2022

KEYWORDS
Active factors; Effect sparsity; Dantzig Selector; Screening designs.

1. Introduction

Screening experiments often use a small number of runs to study a large number of factors of which only a small portion is expected to be important to explain variability in a response variable (effect sparsity). A follow-up study is usually conducted with the selected factors to determine how they affect the response variable. Therefore, it is not overly restrictive to use two-level factors at the screening stage. We consider designs with n runs and m two-level factors under the main effects model such that $n \leq m$. These designs are supersaturated for a main effects model.

The standard design optimality criteria are inadequate for supersaturated designs (SSDs) because they require the model matrix to have a full column rank, which does not hold for SSDs. As a result, there is extensive literature on alternate optimality criteria for SSDs starting with the $E(s^2)$-optimality criterion (Booth and Cox 1962). Several other criteria, such as $UE(s^2)$- and Bayes D-optimality (see, Jones, Lin, and Nachtsheim 2008; Jones and Majumdar 2014, respectively), have been proposed and studied. But none of these criteria are directly related to the screening performance of the designs and can lead to the selection of inferior designs for that task. Weese, Edwards, and Smucker (2017) devised a promising criterion, called var$(s+)$, and Weese et al. (2021) showed that var$(s+)$-optimal designs have excellent screening capability, especially when the effect signs are known.

There is also a growing literature on analysis methods for SSDs. The Gauss-Dantzig Selector (GDS) has surfaced as the best choice in several studies (Phoa, Pan, and Xu 2009; Marley and Woods 2010; Weese, Smucker, and Edwards 2015; Weese, Edwards, and Smucker 2017) that have compared different analysis methods for SSDs. Recently Jones et al. (2020) proposed a two-stage analysis method for special designs to which they refer as group-orthogonal SSDs. The analysis method is based on smaller models that can be fit using least squares. One attractive feature of this method is that there is a direct connection between the design selection and the method of analysis. Weese et al. (2021) compare an improved analysis method of group-orthogonal SSDs to using GDS with their var$(s+)$ designs. They conclude that for identifying active effects while minimizing error rates the var$(s+)$/GDS approach should be used. We propose a new design selection method that, combined with GDS, results in designs that tend to perform better than the var$(s+)$ designs.

Our method, developed in Section 3, uses two criteria motivated by two large sample desiderata of GDS. Adapting a multi-objective Pareto-based coordinate exchange algorithm, we then f ind Pareto ef icient designs based on measures of closeness to meeting these desiderata. We illustrate the success of our method in Section 4 and provide Pareto eficient designs for $5 \leq n \leq 25$ and $n + 1 \leq m \leq 2n$ in the supplementary materials.

CONTACT John Stufken jstufken@gmu.edu University of North Carolina at Greensboro, Greensboro, NC; jstufken@gmu.edu George Mason University, Fairfax, VA.
Rakhi Singh is now at Binghamton University, Binghamton, NY.
John Stufken is now at George Mason University, Fairfax, VA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.

2. Background

2.1. The Model and GDS

Let $\mathbf{1}_n$ be the $n \times 1$ column of 1s, and $\boldsymbol{X}$ be the $n \times m$ design matrix of 1s and -1s corresponding to the m factors. The main effects model is

$$\boldsymbol{y} = \beta_0 \mathbf{1}_n + \boldsymbol{X}\beta + \epsilon, \tag{1}$$

where β_0 is the overall mean, β is a $m \times 1$ vector of parameters corresponding to the m main effects, $\boldsymbol{y} = (y_1, \ldots, y_i, \ldots, y_n)$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_i, \ldots, \epsilon_n)$ are $n \times 1$ vectors of responses and errors, respectively. We assume that y_i's are independent given the design matrix $\boldsymbol{X}$, and that the error terms, ϵ_i, are normally distributed with mean 0 and variance σ^2. Except when discussing existing design criteria in Section 2.2, we will center the response vector and center and normalize the columns of the model matrix $\boldsymbol{X}$. After this transformation we do not need an intercept parameter in the model. We name the transformed $\boldsymbol{y}$ and $\boldsymbol{X}$ as $\boldsymbol{y}_c$ and $\boldsymbol{Z}$, respectively. We retain the notation β in (1) with the new $\boldsymbol{y}_c$ and $\boldsymbol{Z}$.

For reasons discussed in the Section 1, we use GDS to analyze SSDs. GDS (Candès and Tao 2007) was first adopted for SSDs by Phoa, Pan, and Xu (2009). The first stage of GDS, called the Dantzig Selector, uses convex optimization to find parameter estimates. For the centered response $\boldsymbol{y}_c$ and centered and normalized $\boldsymbol{Z}$, the estimator $\hat{\beta}$ is a solution to

$$\min_{\beta \in R^m} ||\beta||_1 \quad \text{subject to} \quad ||\boldsymbol{Z}^T(\boldsymbol{y}_c - \boldsymbol{Z}\beta)||_\infty \le \delta, \tag{2}$$

where $||\beta||_1 = |\beta_1| + \cdots + |\beta_m|$ is the ℓ_1 norm, $||\boldsymbol{b}||_\infty = \max(|b_1|, \ldots, |b_m|)$ is the ℓ_∞ norm, and δ is a tuning parameter. In the second stage, only the effects whose estimates exceed in magnitude a tuning parameter γ are declared active. These effects are then re-estimated using ordinary least squares. This is done for multiple values of $\delta \in (0, ||\boldsymbol{Z}^T\boldsymbol{y}_c||_\infty)$, and model selection criteria like AIC, BIC, adjusted R^2, etc. can be used to select the best value of δ (Phoa, Pan, and Xu 2009; Marley and Woods 2010; Weese et al. 2021). For the choice of γ, recently Weese et al. (2021) evaluated three options and suggested using the data-driven value of $0.1 \times ||\hat{\beta}||_\infty$, where $\hat{\beta}$ corresponds to the solution when $\delta = 0$ in (2). Following Weese et al. (2021), we use this choice of γ and use BIC to tune δ.

2.2. Existing Design Criteria

Most existing design criteria are formulated in terms of the main effects model in Equation (1). Moreover, based on the premise that orthogonal columns are good for disentangling active main effects, albeit not possible in the supersaturated design setting, the focus is on finding designs that minimize some function of the absolute pairwise correlations between the columns of $\boldsymbol{X}$ or of $[\mathbf{1}_n \quad \boldsymbol{X}]$. The first criterion, the $E(s^2)$-optimality, surfaced in Booth and Cox (1962) where the authors proposed to find a design that minimizes the average squared inner-products between all columns of $\boldsymbol{X}$, that is, to minimize

$$E(s^2) = \frac{1}{\binom{m}{2}} \sum_{1 \le i < j \le m} (\boldsymbol{x}_i^T \boldsymbol{x}_j)^2, \tag{3}$$

among all balanced designs, where a balanced design is one that has an equal number of ± 1s for each factor. By definition, balanced designs only exist for even n. Lower bounds to the $E(s^2)$-criterion have been obtained and designs satisfying the lower bounds have been constructed (Lin 1993; Nguyen 1996; Tang and Wu 1997; Cheng 1997; Butler et al. 2001; Bulutoglu and Cheng 2004; Nguyen and Cheng 2008). Readers are referred to Georgiou (2014) and references therein for an excellent review. The identification of $E(s^2)$-optimal designs for general n and m has turned out to be challenging. Dropping the requirement of balance, Marley and Woods (2010) extended this definition by including the sum of squares of the inner products of all columns with the column $\mathbf{1}_n$ to (3). This extended definition was named the $UE(s^2)$-criterion by Jones and Majumdar (2014) who then also obtained the optimality bounds and constructed optimal designs for this criterion. Thus, a design is a $UE(s^2)$-optimal design if it minimizes

$$UE(s^2) = \frac{1}{\binom{m+1}{2}} \left[\sum_{i=1}^{m} (\mathbf{1}^T \boldsymbol{x}_i)^2 + \sum_{1 \le i < j \le m} (\boldsymbol{x}_i^T \boldsymbol{x}_j)^2 \right] \tag{4}$$

among all SSDs with the same values of n and m. Jones and Majumdar (2014) and Cheng et al. (2018) suggested a few criteria to further discriminate between $UE(s^2)$-optimal designs. In addition, SSDs have also been studied under other criteria, such as Bayes D-optimality (Jones, Lin, and Nachtsheim 2008), a class of B-optimality criteria (Deng, Lin, and Wang 1996), the resolution-rank of a design (Deng, Lin, and Wang 1999), the estimation capacity based model-robust criterion (Jones et al. 2009), and maximizing the minimum power for different designs (Weese, Smucker, and Edwards 2015). The var(s) criterion was introduced in Weese, Smucker, and Edwards (2015) where

$$\text{var}(s) = UE(s^2) - (UE(s))^2, \tag{5}$$

where

$$UE(s) = \frac{1}{\binom{m+1}{2}} \left[\sum_{i=1}^{m} (\mathbf{1}^T \boldsymbol{x}_i) + \sum_{1 \le i < j \le m} (\boldsymbol{x}_i^T \boldsymbol{x}_j) \right],$$

and a var(s)-optimal design is one that minimizes var(s) among all possible SSDs subject to a specified $UE(s^2)$ eficiency. Further, with the additional constraint that $UE(s) > 0$, Weese, Edwards, and Smucker (2017) called it the constrained-*positive* var(s), or var$(s+)$, criterion. The use of s in $E(s^2)$, $UE(s^2)$, var(s), and var$(s+)$ is attributed to the fact that traditionally the inner products $\boldsymbol{x}_i^T\boldsymbol{x}_j$ were denoted by s_{ij}'s for $i,j = 1, \ldots, m$.

Assessment of SSDs should be based on their performance in effect screening. The screening performance can be studied via simulations. Two common criteria to assess the performance are

- power: the proportion of active effects correctly identified in a specified scenario, and
- (Type 1) error: the proportion of inactive effects declared to be important in a specified scenario.

Distribution of power and error are studied over multiple iterations for each scenario. Following previous studies, we consider a design to be a good design if it has high average power and low average error across a range of different simulation scenarios. In other contexts, power and error are also referred to as true positive rate and false positive rate, respectively, however, for consistency with the screening design literature, we retain the terms power and error throughout this work. Several studies comparing the screening performance of different designs have concluded that optimal and eficient designs perform approximately equally well, no matter which optimality criterion is used (Marley and Woods 2010; Weese, Smucker, and Edwards 2015). Nonetheless, design choice matters. Random designs (±1s being assigned to each entry of $\boldsymbol{X}$ with probability 0.5) generally do not perform well, making it important to select a design judiciously. Weese et al. (2021) concluded that, while using GDS, (a) var(s+) designs have better screening performance than other SSDs when the effect signs are known, and (b) the var(s+) designs and $UE(s^2)$-optimal designs have identical performance when the ef fect signs are unknown. The claim in (b) is generally true but there are situations when some $UE(s^2)$-optimal designs have an inferior screening performance (see examples for $(n, m) = (12, 14), (20, 37)$ in the supplementary materials).

3. Design Criteria based on GDS Analysis

Weese et al. (2021) concluded that var$(s+)$ designs are strong performers especially if the signs of the effects are known. While they used the GDS analysis as part of their screening strategy, there is no explicit connection between the analysis and the design selection criterion. In this section we build a design selection method that is motivated by the GDS analysis. We will use the transformed versions of y and $\boldsymbol{X}$, that is, $\boldsymbol{y_c}$ and $\boldsymbol{Z}$, respectively, and the true parameter vector β has m components corresponding to m main effects. For a given δ and n-run design, let $\hat{\beta}^n(\delta) = (\beta_1^n(\delta), \ldots, \beta_m^n(\delta))$ be the GDS estimate of β.

3.1. When Does GDS Work Well?

A good GDS estimate $\hat{\beta}^n(\delta)$ of the true parameter β should possess two large sample properties:

(a) Estimation consistency: The estimation consistency asks that there exists a $\delta (\geq 0)$ such that

$$||\hat{\beta}^n(\delta) - \beta||_2 = o_p(1),$$

where $||\boldsymbol{b}||_2 = (b_1^2 + \cdots + b_m^2)^{0.5}$ denotes the ℓ_2 norm of $\boldsymbol{b}$.

(b) Model selection consistency: The model selection consistency asks that there exists a $\delta (\geq 0)$ such that

$$P(\{i \in 1, \ldots, m : \hat{\beta}_i^n(\delta) = 0\} = \{i \in 1, \ldots, m : \beta_i = 0\}) \to 1 \quad \text{as} \quad n \to \infty.$$

It is not too hard to see that one property does not necessarily imply the other and that it would be desirable for $\hat{\beta}_n(\delta)$ to have both properties for the given δ. As is common in the literature, we will replace (b) by the stronger condition of weak sign consistency, namely,

(c) Weak sign consistency: The weak sign consistency asks that

$$P(\exists \delta \geq 0, \text{sign}(\hat{\beta}^n(\delta)) = \text{sign}(\beta)) \to 1 \quad \text{as} \quad n \to \infty,$$

where $\text{sign}(\beta) = (\text{sign}(\beta_1), \ldots, \text{sign}(\beta_m))$ and $\text{sign}(\beta_i) = 1, -1, 0$ if $\beta_i > 0$, $\beta_i < 0$, and $\beta_i = 0$, respectively, $i = 1, \ldots, m$.

Condition (a) is satisfied if the matrix $\boldsymbol{Z}$ satisfies the uniform uncertainty principle (Candès and Tao 2007) (which states that the matrix $\boldsymbol{Z}$ obeys the restricted isometry principle defined in the next section), whereas Condition (c) is satisfied *only if* $\boldsymbol{Z}$ satisfies the weak irrepresentable conditions (Gai, Zhu, and Lin 2013). These conditions are further discussed in the following sections. Neither of these conditions is easy to verify, and our new design selection method will therefore be based on surrogates for these two conditions.

3.2. Estimation Consistency

Candès and Tao (2005, 2007) showed that GDS correctly estimates the true β with large probability provided that β is sufficiently sparse and that the matrix $\boldsymbol{Z}$ satisfies the Restricted Isometry Principle (RIP) with small values of the isometry constants. For a positive integer $k (\leq m)$, let F_k denote the set of all subsets of $\{1, 2, \ldots, m\}$ of size k, and let f denote an element in F_k. Rephrasing Candès and Tao (2007), the k-restricted isometry constant τ_k of $\boldsymbol{Z}$ is the smallest quantity such that

$$(1 - \tau_k) \sum_{i \in f} a_i^2 \leq \left|\left| \sum_{i \in f} a_i \boldsymbol{z}_i \right|\right|^2 \leq (1 + \tau_k) \sum_{i \in f} a_i^2 \qquad (6)$$

for all $f \in F_k$ and for any k constants (a_i), $i \in f$. Here $\boldsymbol{z}_i$ denotes the ith column of $\boldsymbol{Z}$. Note that if, for some $f \in F_k$, the vectors $\boldsymbol{z}_i$, $i \in f$, are orthogonal, then for that f it would be fine for τ_k to be 0. But we cannot have orthogonality for all $f \in F_k$, so that τ_k in (6) will have to be positive. With a smaller value of τ_k so that (6) holds for all $f \in F_k$ and all $(a_i), i \in f$, the columns of $\boldsymbol{Z}$ can be considered being "closer" to orthogonality. Design choice matters to make this happen. Moreover, it assures, through the RIP results, that the Dantzig Selector can, with high probability, correctly recover active effects under sparsity. For example, with $\tau_k + \tau_{2k} + \tau_{3k} < 1$, the Dantzig Selector is very accurate if there are no more than k active effects (Candès and Tao 2007). While this condition can be satisfied for larger n and m (e.g., for $n = 1000$ and $m = 5000$, even for random designs it is often satisfied for k up to 30), it fails to hold for typical sizes of n and m in screening experiments. Nonetheless, as we will see, it pays to aim for designs with relatively small isometry constants τ_k for a range of values for k.

3.2.1. The Estimation Consistency-based Criterion

Let $\boldsymbol{Z}_f$ be the $n \times k$ matrix obtained by only keeping the columns of $\boldsymbol{Z}$ corresponding to the indices in f. The term $\left|\left| \sum_{i \in f} a_i \boldsymbol{z}_i \right|\right|^2$ in (6) is equal to $\boldsymbol{a}^T \boldsymbol{Z}_f^T \boldsymbol{Z}_f \boldsymbol{a}$, where $\boldsymbol{a}$ is a $k \times 1$ vector of the constants (a_i), $i \in f$. With $\lambda_{f1} \leq \lambda_{f2} \leq \cdots \leq \lambda_{fk}$ as the k ordered eigenvalues of $\boldsymbol{Z}_f^T \boldsymbol{Z}_f$, we know that,

$$\lambda_{f1} \boldsymbol{a}^T \boldsymbol{a} \leq \boldsymbol{a}^T \boldsymbol{Z}_f^T \boldsymbol{Z}_f \boldsymbol{a} \leq \lambda_{fk} \boldsymbol{a}^T \boldsymbol{a}, \qquad (7)$$

for any $\boldsymbol{a}$ and $f \in F_k$. Also, for any f, $\lambda_{f1} \leq 1$, $\lambda_{fk} \geq 1$, and $\lambda_{f1} = \lambda_{fk} = 1$ implies pairwise orthogonality of the columns in $\boldsymbol{Z}_f$. Define $\tau^f = \max(1-\lambda_{f1}, \lambda_{fk}-1)$ for each $f \in F_k$. From (7), we see that (6) holds for $\tau_k = \max_f \tau^f$. With the GDS analysis in mind, if we expect k active effects, we should aim for designs with relatively small values of τ_k, τ_{2k} and τ_{3k}. In other words, we should consider a range of values for k. In addition, rather than just focusing on $\max_f \tau^f$ it is beneficial to look at the tail of the distribution of τ^f over f. For a given k, the distribution of the τ^f_k's over f can vary considerably for two designs. Comparing these designs just based on $\max_f \tau^f_k$ is not a very robust measure. A design with relatively few τ^f_k's near its maximum, would allow a smaller value for the isometry constant except for a few choices of $f \in F_k$. After considering different options through simulations, we found that the 95th percentile of the distribution of the τ^f_k's offered a more meaningful comparison of designs in terms of screening performance than the maximum. With ζ_k as the 95th percentile of the τ^f_k's for a given design, we propose to use the following estimation consistency-based criterion.

EC-optimality: For fixed n and m, a design is EC-optimal if it minimizes

$$\text{EC} = \text{mean}(\zeta_2, \ldots, \zeta_{\lfloor n/2 \rfloor}). \tag{8}$$

A larger range than $2, \ldots \lfloor n/2 \rfloor$ could have been considered, but the required increase in computational expense offered little in return. One could also consider other functions than the mean of the ζ_k's, but there was no compelling reason to make a different choice. Finally, computing ζ_k exactly for a given design requires the computation of $\binom{m}{k}$ values for τ^f_k. We return to this in Section 3.4.

3.3. Weak Sign Consistency

For an integer k, $1 \leq k \leq m$, let $f \in F_k$ and $\boldsymbol{Z}_f$ be defined as in Section 3.2. We write $\bar{f}$ for the complement of f in $\{1, \ldots, m\}$, and use $\boldsymbol{C}_{f_1,f_2}$ to denote the matrix $\boldsymbol{Z}_{f_1}^T \boldsymbol{Z}_{f_2}$ for $f_1, f_2 \in \{1, \ldots, m\}$, where the two subsets can be of different sizes. Finally, we use $f^* \in F$ to denote the (unknown) subset corresponding to k active effects. For f^*, the Weak Irrepresentable Conditions (WICs) for Dantzig Selector (Gai, Zhu, and Lin 2013) require that there exists a set $f \in F_k$ so that $\boldsymbol{C}_{f^*,f}$ is invertible and

$$\begin{aligned} |\boldsymbol{C}_{\bar{f}^*,f}\boldsymbol{C}^{-1}_{f^*,f}\text{sign}(\beta_{f^*})| &< \boldsymbol{1}_{m-k} \\ |\boldsymbol{C}_{f,f^*}\boldsymbol{C}^{-1}_{f,f^*}\text{sign}(\boldsymbol{C}^{-1}_{f^*,f}\text{sign}(\beta_{f^*}))| &< \boldsymbol{1}_{m-k} \end{aligned} \tag{9}$$

hold element-wise in absolute value. The WICs are necessary for weak sign consistency of the Dantzig Selector (Gai, Zhu, and Lin 2013).

Since, typically, we know neither the value of k nor the active effects nor the signs of the active effects nor the set f, if any, for which the WICs are satisfied, these inequalities are virtually impossible to check. Even if one is willing to specify a value for k, for each of the $\binom{m}{k}$ choices for $f^* \in F_k$ and each of the 2^k choices for $\text{sign}(\beta_{f^*})$ we would have to find an $f \in F_k$ so that the WICs are satisfied. For larger m, this is computationally infeasible for even modest values of k. Therefore, we now devise a criterion based on simplified WICs to distinguish between designs.

3.3.1. The Weak Sign Consistency-based Criterion

We simplify the WICs by assuming that: (a) the true signs of active effects are known, and (b) the set f is the same as f^*. Based on our simulations, the assumption in (b) is not all that restrictive. In fact, the WICs for LASSO use $f = f^*$, so, our designs would also work well if LASSO was instead used as analysis tool. Note that if (a) holds and it is known that an active effect has a negative sign, then we can simply multiply the corresponding column in $\boldsymbol{Z}$ by -1 to make the sign of that effect positive. So, without loss of generality, when (a) holds we can take the signs for all active effects to be positive. The assumption in (a) is slightly more restrictive but it is not without precedence (Weese et al. 2021). As we will see in Section 4, the designs that we identify with these simplifications also perform well when the signs are unknown. For the set of true active effects f^* of size k, the simplified WICs require that $\boldsymbol{C}_{f^*,f^*}$ is invertible and

$$\begin{aligned} |\boldsymbol{C}_{\bar{f}^*,f^*}\boldsymbol{C}^{-1}_{f^*,f^*}\boldsymbol{1}_k| &< \boldsymbol{1}_{m-k} \\ |\boldsymbol{C}_{f^*,f^*}\boldsymbol{C}^{-1}_{f^*,f^*}\text{sign}(\boldsymbol{C}^{-1}_{f^*,f^*}\boldsymbol{1}_k)| &< \boldsymbol{1}_{m-k} \end{aligned} \tag{10}$$

hold element-wise in absolute value. For a given design, let γ_k denote the proportion of sets f^* of size k for which the simplified WICs are satisfied. Then, using the same range of k as for EC-optimality, we define the following weak sign consistency-based criterion:

For f^*, the Weak Irrepresentable Conditions (WICs) for Dantzig Selector (Gai, Zhu, and Lin 2013) require that there exists a set $f \in F_k$ so that $\boldsymbol{C}_{f^*,f}$ is invertible and

WSC-optimality: For fixed n and m, a design is WSC-optimal if it maximizes

$$\text{WSC} = \text{mean}(\gamma_2, \ldots, \gamma_{\lfloor n/2 \rfloor}). \tag{11}$$

Ideally for any given design and value of k, we would like to evaluate ζ_k in (8) and γ_k in (11). But that is impossible given the rapid growth of $\binom{m}{k}$, especially when searching for good designs. Instead, for a given m and k, we will use $\min(s, \binom{m}{k})$ subsets of size k. If $s \geq \binom{m}{k}$, we use all subsets of size k and otherwise randomly select s subsets. Writing $S^s = (S_2, \ldots, S_{\lfloor n/2 \rfloor})$ for the collection of sets used, we approximate EC and WSC based on S^s.

3.4. Pareto Optimalty

While EC-optimal designs and WSC-optimal designs often appear to perform well as screening designs, based on empirical evidence we recommend using both criteria for finding good screening designs. One approach is to optimize a weighted linear combination of the two criteria. But optimal designs would depend heavily on the non-interpretable weights (Lu, Anderson-Cook, and Robinson 2011). Another approach is to find designs that optimize one criterion meeting a threshold for the other criterion, as was done with var(s+)-optimality. But there is neither a natural threshold nor a natural choice for the roles of the two criteria in our case. Instead, we will search for Pareto optimal designs using both criteria (see, Lu, Anderson-Cook, and Robinson 2011; Cao, Smucker, and Robinson 2017). This searches for designs that are not dominated by other designs. Design d_1 dominates design d_2 if d_1 is at least as good as

d_2 on both criteria and is strictly better on at least one of them. The collection of all such designs form the Pareto front. A recent computationally eficient state-of-the-art algorithm for finding designs on the Pareto front is the hybrid elitist Pareto optimality-based coordinate exchange algorithm (called EPCEA) of Cao, Smucker, and Robinson (2017).

The hybrid EPCEA algorithm uses three operators: the coordinate exchange operator, the enhanced elitism operator, and the multi-start operator. The coordinate exchange operator takes design d and the number of maximum coordinate exchanges *niter* as inputs and outputs a set of nondominated designs at the Pareto front. It sequentially exchanges each element of d with another in the set $\{-1, 1\}$ either until a set of non-dominated designs can no longer be improved by coordinate exchanges or the *niter* number of possible exchanges are considered. The details of this operator remain the same as in Cao, Smucker, and Robinson (2017) and are provided in the supplementary materials. Before discussing our modifications, we first present the EPCEA algorithm of (Cao, Smucker, and Robinson 2017):

(1) Set $i = 1$.
(2) Start with an initial design d_i.
(3) Find the set of designs at the Pareto front (say, $P^0_{d_i}$) using the coordinate exchange operator with inputs d_i and *niter*.
(4) **Enhanced elitism operator:**
   (4.1) Perform the coordinate exchange operator on each design that belongs to $P^g_{d_i}$ but not in $P^{g-1}_{d_i}$ for $g = 0, 1, 2, \ldots$ with $P^{-1}_{d_i}$ being the empty set. Find all non-dominated designs among the resultant designs and the designs in $P^g_{d_i}$ and call the new set $P^{g+1}_{d_i}$.
   (4.2) Stop when $P^g_{d_i} = P^{g-1}_{d_i}$ (Condition A) is met.
(5) **Multi-start operator:** Set $i = i + 1$. Repeat steps (2)–(4) for a different initial design d_i as long as $i \leq ndes$.
(6) Find all non-dominated designs among all the resultant designs obtained in step (5) for $i = 1, \ldots, ndes$ and produce a final set of non-dominated designs at the Pareto front.

Notice that the designs obtained in step (6) of the EPCEA are only Pareto eficient since the true Pareto front is not known. Finding global Pareto optimal designs is nearly impossible except perhaps for small values of n and m. Also, since $n \leq m$, correlations between factors cannot be completely avoided. But, typically, SSDs with two columns having a perfect correlation of ± 1 are avoided as are the columns with all 1s or all -1s. In the same spirit, we ignore potential exchanges in the coordinate exchange operator that would result in a perfect correlation. Additionally, in Algorithm 1, we adapt the EPCEA algorithm for SSDs by making the following three changes:

(i) Two further conditions are added to step (4.2) in the EPCEA. Condition B requires step (4.1) to stop if $g \leq 100$, that is, the enhanced elitism operator is used at most $g = 100$ times. Condition C requires that step (4.1) be stopped if the cardinality of the set $P^g_{d_i}$ reaches 100, that is, we allow a maximum number of 100 designs at the Pareto front at each iteration. These additions are added purely due to computational considerations. The new step (4.2) becomes "stop when any of Conditions A, B, or C is met."

Algorithm 1: Modified EPCEA for SSDs

inputs: the number of sets s, the maximum number of coordinate exchanges *niter*, the number of choices for S^s *ntry*, the number of initial designs *ndes* and designs $d_1, \ldots, d_{ndes}$.

```
for i = 1 → ndes do
    for j = 1 → ntry do
        For the jth choice of the collection S^s, say S^s_j, design
        d_i, and niter exchanges, run steps (2)-(4) of the
        EPCEA to obtain non-dominated designs at the
        Pareto front, say P_j;
    end
    For the collection S_i^{ntry×s} (= ⨁_{j=1}^{ntry} S_j^s) of size (ntry × s),
    retain the non-dominated designs from ⨁_{j=1}^{ntry} P_j to
    produce a set P^i;
end
For a different collection S^{5s} of size 5s, combine all
resultant designs from P^1,..., P^{ndes} to produce a final set
of Pareto eficient designs P_F;
```

output: output designs in P_F

(ii) Steps (2)–(4) of the EPCEA produce a set of non-dominated designs at the Pareto front when starting from design d_i. While different designs with the same criteria values may result in different screening performance, using all of them further in EPCEA results in a large computational cost. For each starting design, we retain one arbitrary design among all designs obtained in steps (2)–(4) with the same values of the two criteria. Similarly, we retain designs with the unique values of both criteria in step (6).

(iii) Recall, from the discussion at the end of Section 3.3.1, that the values of our criteria are dependent on S^s. For a given design d_i and a choice of s, we repeat steps (2)–(4) of the EPCEA for *ntry* different choices of the collection S^s. Similar to (ii), we also retain unique designs for each collection. This step assesses the sensitivity of the EPCEA to the choice of S^s. In the next section, we demonstrate that the Pareto fronts are not very sensitive to the choice of S^s. Therefore, we could use a small value of *ntry* to keep the computation time manageable.

Algorithm 1 presents the EPCEA with these three modifications. Despite massive simplifications, Algorithm 1 remains computationally expensive. For example, starting with $ndes = 1$ and $ntry = 2$ for $n = 16, m = 31$, Algorithm 1) terminated in about 26 hr on a Desktop with an Intel Xeon CPU @ 3.70GHz processor with 32GB RAM. Nevertheless, finding simplified Pareto eficient designs is a worthwhile exercise since, as demonstrated in the next section, they perform better than the state-of-the-art designs in about 85% of the considered cases, and equivalent in the remaining 15% cases, when the effect signs are known.

The final choices for the inputs in Algorithm 1 are all motivated by computational costs. Algorithm 1 could converge in less than *niter* coordinate exchanges, but if it does not, setting

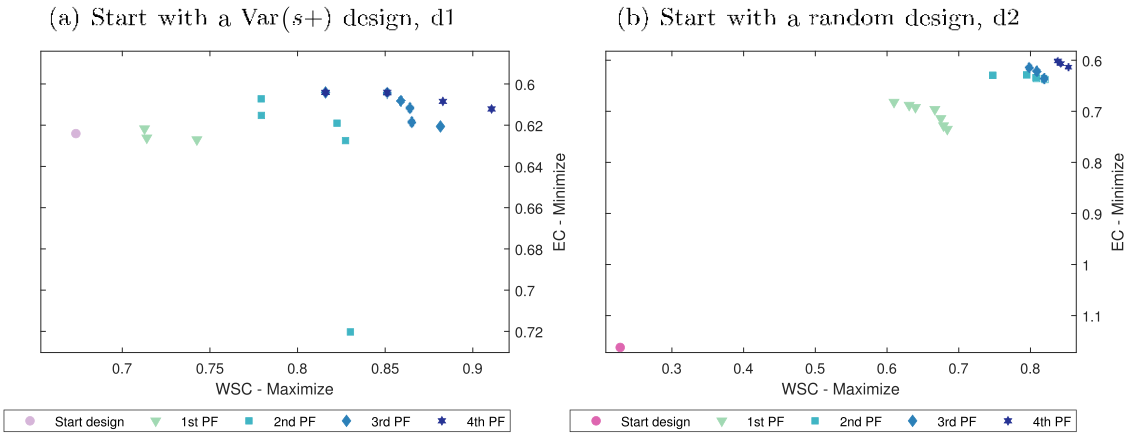

Figure 1. Pareto fronts starting from two different designs for $n = 9$ and $m = 10$. Each point corresponds to a different design, while different symbols indicate the stage at which the design was found.

a higher value of *niter* increases computational costs. Recall that for every coordinate exchange, EC and WSC are both calculated over $s(\lceil n/2 \rceil - 1)$ sets. Therefore, the higher the value of s, the larger the computational cost. Line 7 of Algorithm 1 uses a collection of size $5s$. Though the choice is somewhat arbitrary, it is used to achieve more stable criteria values for producing a final set of non-dominated designs. Next, a large value of *ntry* does not necessarily result in better values of EC and WSC. However, using $ntry = 1$ could lead to poor results with a bad choice of S^s. We suggest using a few different values of *ntry* to get stable results. Finally, starting from a good design often leads to faster and better solutions. As a result, we deploy var(s+)-optimal designs as one of the initial designs.

4. Pareto Eficient Designs

We first illustrate Algorithm 1 by means of two examples: (a) $n = 9, m = 10$, and (b) $n = 14, m = 24$. We use the same 12 simulation scenarios as in Weese et al. (2021), the details of which are now provided for the sake of completeness. The number of active ef fects are taken to be $\lceil 0.25n \rceil$, $\lceil 0.50n \rceil$, or $\lceil 0.75n \rceil$, with magnitude generated as Exp(1) + SN, where SN = 1 or 3. Ef fect signs are assumed to be known (all ef fect signs positive) or unknown (effect signs randomly selected from ± 1 with probability 0.5), denoted by K and U, respectively. The magnitudes of inactive effects are generated by taking the absolute value of $N(0, 6^{-2})$ and errors are independent $N(0,1)$. The responses are then generated using model (1). We denote the simulation scenarios with a triplet: (SN, Sparsity, Direction). For example, the triplet $(1, 0.25n, K)$ implies that the true model has $\lceil 0.25n \rceil$ active effects with SN=1, and signs of effects are assumed to be known.

Example 4.1. Let $n = 9$ and $m = 10$. For $k = 2$ to $k = \lceil n/2 \rceil = 5$, since $\binom{m}{5} = 252$ is small, we run Algorithm 1 with $ntry = 1$ and $s_k = \binom{m}{k}$. The value of *niter* is set to 5000. Additionally, set $ndes = 2$, and use a var(s+)-optimal design (d1) and a randomly created design (d2) as initial designs. In Figure 1, dots labeled "Start design" (light and dark pink in panels (a) and (b), respectively) correspond to the two initial designs. Triangles, squares, diamonds, and hexagons correspond to $g = 1, 2, 3, 4$,

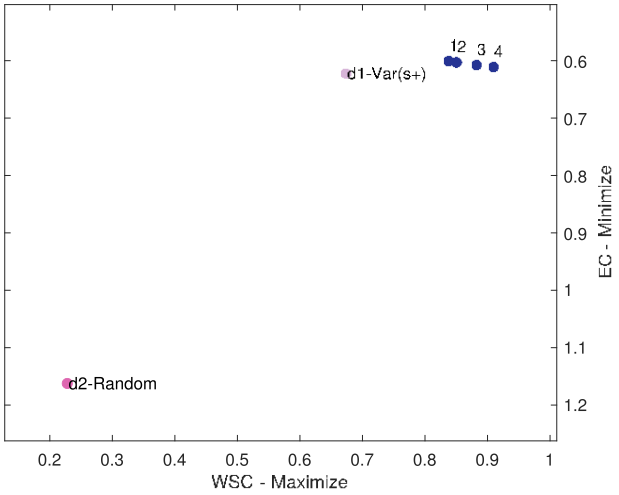

Figure 2. Final designs at the Pareto Front for $n = 9$ and $m = 10$. Each point corresponds to a different design. The designs labeled 1 through 4 are the n Pareto eficient designs found when using a var(s+) and random design as starting designs.

respectively, in step (4) of the EPCEA. The four hexagons in panel (a) and three hexagons in panel (b) represent the Pareto eficient designs obtained from the initial designs d1 and d2, respectively. Changes in criteria values are much higher for d2 than for d1 indicating that d2 had a lot of room for improvement. For this example, Algorithm 1 converges in 4 steps for both starting designs. Step 7 of Algorithm 1 results in four Pareto eficient designs in P_F displayed as blue dots in Figure 2. Designs 2, 3, and 4 were generated by d1, while design 1 was generated by d2. In Figure 3, we show the performance of the leftmost and rightmost designs on the Pareto Front, labeled, respectively, as 1 and 4 in Figure 2. Power and error for the other two designs, labeled as 2 and 3 in Figure 2, falls between power and error for designs 1 and 4. As shown in Figure 3, the Pareto eficient designs have a better screening performance than d1 and d2 when effects signs are known (panel (a)) and perform equivalently to the var(s+)-optimal design when effect signs are unknown (panel (b)). The standard deviations of power and error for each design-scenario combination in Figure 3 are provided in the supplementary materials.

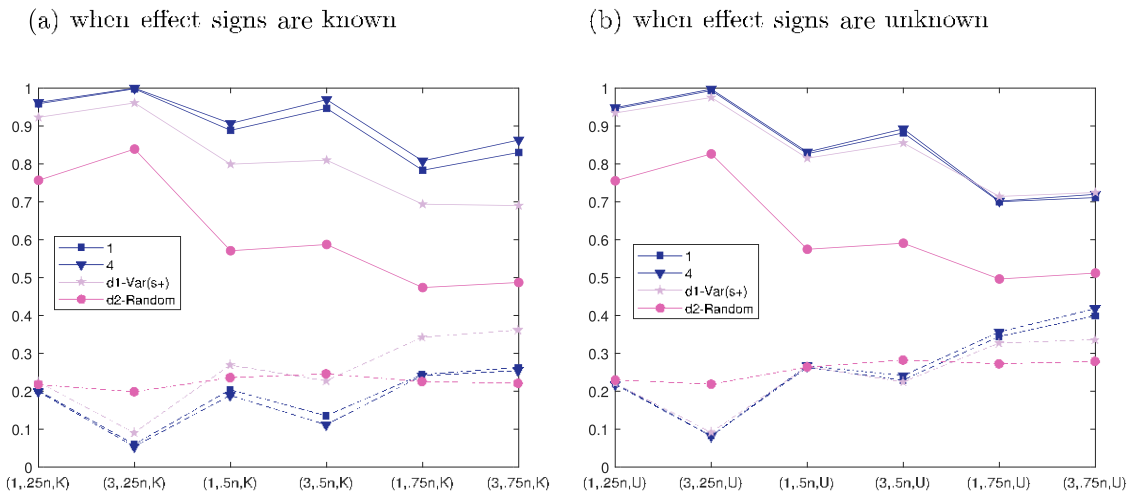

Figure 3. Average screening performance (power (solid lines) and error (broken lines)) of the four selected designs from Figure 2 over 10,000 iterations for $n = 9$ and $m = 10$.

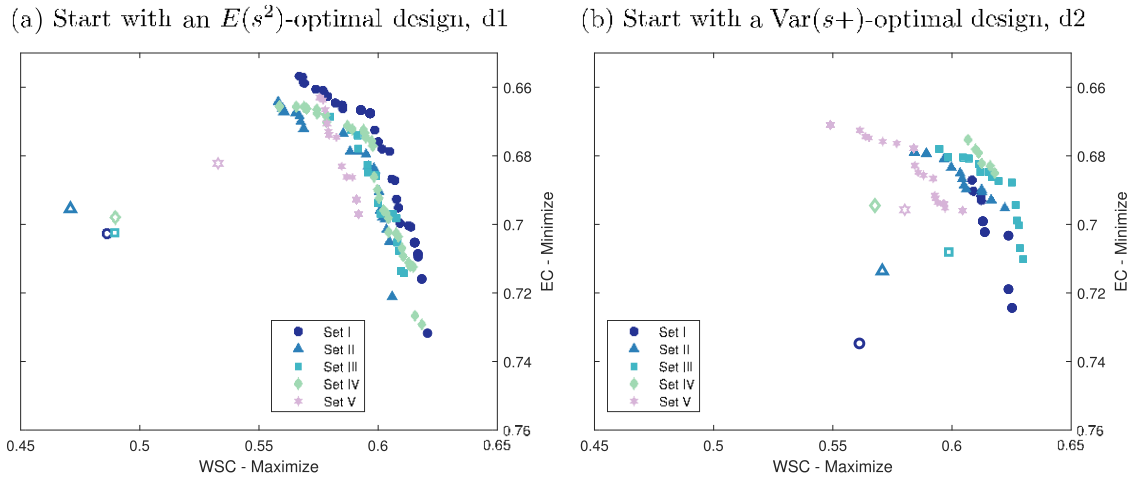

Figure 4. Pareto Fronts for 10 different starting sets for $n = 14$ and $m = 24$, five for d1 and another five for d2. The hollow symbols correspond to the values of the starting design on the five different sets, and the filled symbols correspond to Pareto eficient designs for these different sets.

Example 4.2. Let $n = 14$ and $m = 24$. We set *niter* $= 5000$, and *ndes* $= 2$, once starting with an $E(s^2)$-optimal design (d1) and once with a var$(s+)$-optimal design (d2). Since $\binom{m}{\lfloor dn/2 \rfloor}$ is huge, we use $s = 1000$. Recall that $s_k = s$ if $s < \binom{m}{k}$, and $\binom{m}{k}$, otherwise. Therefore, for $S^s = S^{1000} = (S_2, S_3, S_4, S_5, S_6, S_7)$, components S_2 and S_3 use all $\binom{m}{k}$ sets whereas S_4 to S_7 use 1000 randomly selected sets. We use five different choices of S^{1000} setting *ntry* $= 5$. In panel (a) of Figure 4, starting from d1, five different colors represent Pareto eficient designs for five different choices of S^s. Panel (b) shows the same development when starting from a var$(s+)$-optimal design. The final values of g are $(13, 9, 14, 16, 9)$ for the five different choices in panel (a) versus $(10, 10, 12, 13, 10)$ for panel (b). Thus, for some collection S^s, the Pareto front P_F is reached faster than for others. However, Figures 4–6 give some assurance that the choice of collection and starting design lead to only minor differences in the criteria values for the designs and, more importantly, the screening performance of the designs.

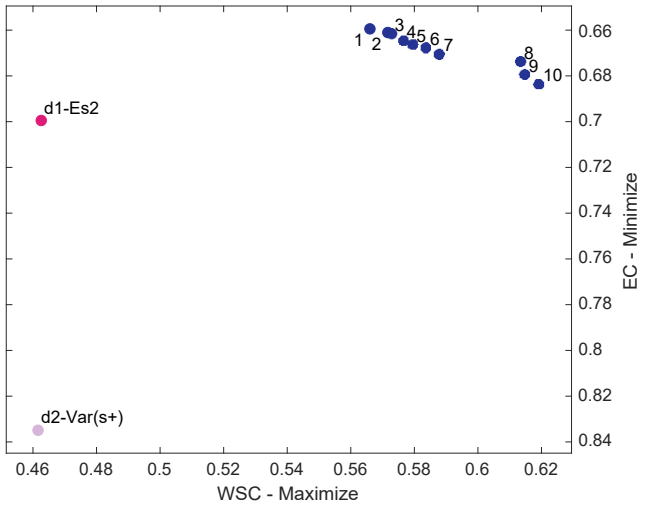

Figure 5. Final designs at the Pareto Front for $n = 14$ and $m = 24$.

Figure 5, similar to Figure 2, shows the f inal 10 Pareto ef i-cient designs (in blue) that constitute the P_F as well as two initial starting designs (in dark and light pink). Designs numbered 1–7 are the ones obtained from d1, whereas designs marked 8–10 are obtained from d2. This clear divide based on the choice of the initial design is uncommon (see the Supplementary Material for other examples). Similar to Example 4.1, in Figure 6, we

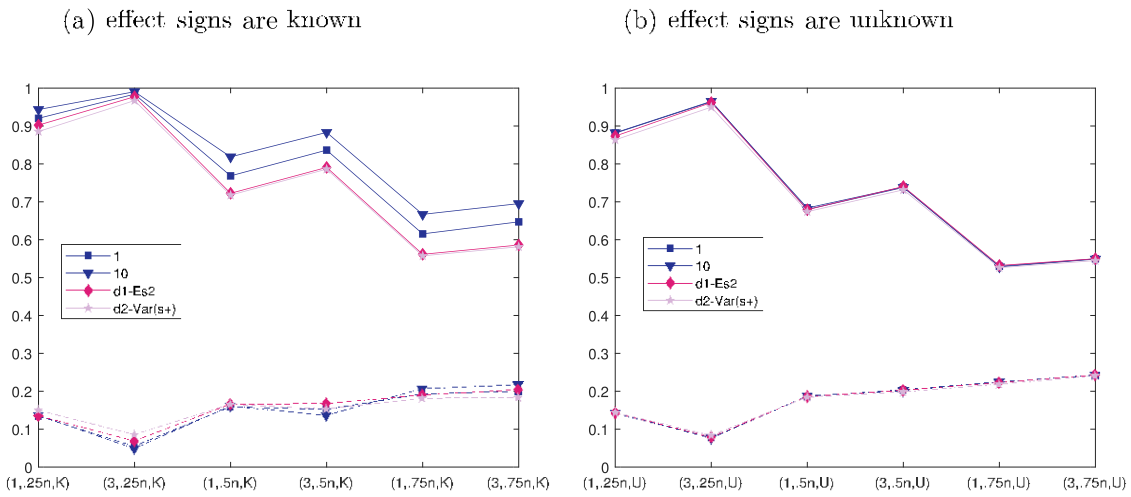

Figure 6. Average screening performance (power (solid lines) and error (broken lines)) of the four selected designs from Figure 5 over 10,000 iterations for $n = 14$ and $m = 24$.

show the performance of the leftmost and rightmost designs on the Pareto Front, labeled as 1 and 10 in Figure 5. The power and error of the designs labeled 8 and 9 were similar to those of design 10, whereas the performances of designs labeled 2 to 7 were similar to that of design 1. Similar to Example 4.1, panel (a) of Figure 6 shows that for known effect signs, Pareto eficient designs perform better than both the $E(s^2)$-optimal design as well as the var(s+)-design. In panel (b) of Figure 6, all designs have equivalent screening performance for unknown effect signs. The standard deviations of power and error iterations for each design-scenario combination in Figure 6 are provided in the supplementary materials.

Panel (a) in Figure 6 shows that design 10 performs slightly better than design 1. Design 10 is at the right end of the Pareto front in Figure 5. The behavior that designs toward the right end of Pareto Front perform better is also reflected in Figure 3, design 1 performs worse than design 4, though the differences are small. We evaluate this further in the next example.

Example 4.3. Let $n = 16$ and $m = 28$. We use $niter = 5000$, $ntry = 5$, $s = 1000$, and $ndes = 2$ with a UE(s^2)-optimal and a var(s+)-optimal design. Algorithm 1 provides twelve Pareto eficient designs (numbered 1 to 12 from left to right on the Pareto front). We observe that, for known effect signs, with small increases in the error for all eight designs, the power increases as we move from left to right on the Pareto front, except for the first two scenarios where the power seems constant across different designs. The figures that illustrate this as well as the figure that compares Pareto eficient designs with the starting designs are provided in the supplementary materials. Based on these three examples as well as several others shown in the supplementary materials, designs at the right end of the Pareto front empirically seem to have a better screening performance.

In addition to the examples discussed here, the screening performance of Pareto eficient designs obtained using $ntry = 2$ and var(s+) design as the initial design, for $5 \le n \le 25$ and $n + 1 \le m \le 2n$ are provided in the supplementary materials. For each n and m, at most two Pareto eficient designs on the right end of the Pareto front are also provided. The screening performance and designs corresponding to the examples from Weese et al. (2021) are also separately provided in the Supplementary Material. As a summary, when the effect signs are known, Pareto eficient designs perform better than the var(s+)-optimal designs in about 85% of the considered cases and perform equivalently in all other cases except for one. When the effect signs are unknown, Pareto eficient designs perform at par with the var(s+)-optimal designs in all cases. It is also worth noting that our criterion is not a surrogate for the var(s+) criterion. Of the 315 Pareto eficient designs, only 2 have a smaller var(s) value than the corresponding var(s+)-optimal design.

5. Conclusions

Supersaturated designs for main effects models have been extensively studied. Several optimality criteria and several analysis methods have been studied in the supersaturated design literature. While GDS is often the preferred method of analysis, until recently, it has been observed that no available design criterion is able to discriminate between good and bad designs from a screening perspective. Recently, the var(s+)-optimal designs are empirically shown to have a better performance than other designs, especially when effect signs are known. However, none of these criteria are directly connected to GDS.

With the aim of finding a connection between design selection and GDS, we propose two new criteria inspired by the large sample properties of GDS. For GDS to perform well, small values of k-restricted isometry constants and a large proportion of sets satisfying the weak irrepresentable conditions are desirable. A design that performs poorly for one or a small number of models with k active effects may not be a bad design if it performs relatively well for most other models. Therefore, we consider the upper tail behavior of the isometry constants through the EC-optimality. In addition, since evaluating WICs is computationally expensive, we evaluate the proportion of sets satisfying the simplified WICs through the WSC-optimality. We then find Pareto eficient designs using an advanced multi-

objective coordinate exchange algorithm. For known signs, we see that the Pareto eficient designs have better screening performance than var(s+)-optimal designs for about 85% of the considered cases. For the remaining 15% cases as well as for the unknown effect signs, the Pareto eficient designs perform at par with the var(s+)-optimal designs.

Despite simplifications, the two criteria remain computationally expensive. Given that Pareto eficient designs are directly connected to GDS and have a superior screening performance, an eficient algorithm could be a useful contribution to the SSD literature.

Supplementary Materials

All designs used here are provided in the supplementary materials along with (a) the performance of additional examples, (b) Pareto eficient designs and their screening performance for $5 \le n \le 25$ and $n + 1 \le m \le 2n$, and (c) MATLAB code implementation of Algorithm 1.

Acknowledgments

The authors are grateful to the Editor, an Associate Editor, and two referees for detailed comments that resulted in a significant improvement of the initial version of the article.

Disclosure Statement

The authors declare that they have no conf lict of interest.

Funding

JS gratefully acknowledge support through NSF grant DMS-1935729.

References

Booth, K. H. V., and Cox, D. R. (1962), "Some Systematic Supersaturated Designs," *Technometrics*, 4, 489–495. [96,97]

Bulutoglu, D. A., and Cheng, C.-S. (2004), "Construction of *E(s²)*-optimal Supersaturated Designs," *Annals of Statistics* 32, 1662–1678. [97]

Butler, N. A., Mead, R., Eskridge, K. M., and Gilmour, S. G. (2001), "A General Method of Constructing *E(s²)*-Optimal Supersaturated Designs," *Journal of the Royal Statistical Society*, Series B, 63, 621–632. [97]

Candès, E., and Tao, T. (2005), "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, 51, 4203–4215. [98]

——— (2007), "The Dantzig Selector: Statistical Estimation when p is much Larger than n," *Annals of Statistics*, 35, 2313–2351. [97,98]

Cao, Y., Smucker, B. J., and Robinson, T. J. (2017), "A Hybrid Elitist Pareto-based Coordinate Exchange Algorithm for Constructing Multi-Criteria Optimal Experimental Designs," *Statistics and Computing*, 27, 423–437. [99,100]

Cheng, C.-S. (1997), "*E(s²)*-optimal Supersaturated Designs," *Statistica Sinica*, 7, 929–939. [97]

Cheng, C.-S., Das, A., Singh, R., and Tsai, P.-W. (2018), "E(s^2)-and UE(s^2)-Optimal Supersaturated Designs," *Journal of Statistical Planning and Inference*, 196, 105–114. [97]

Deng, L. Y., Lin, D. K. J., and Wang, J. (1996), "A Measurement of Multi-Factor Orthogonality," *Statistics & Probability Letters*, 28, 203–209. [97]

——— (1999), "A Resolution Rank Criterion for Supersaturated Designs," *Statistica Sinica*, 9, 605–610. [97]

Gai, Y., Zhu, L., and Lin, L. (2013), "Model Selection Consistency of Dantzig Selector," *Statistica Sinica*, 23, 615–634. [98,99]

Georgiou, S. D. (2014), "Supersaturated Designs: A Review of their Construction and Analysis," *Journal of Statistical Planning and Inference*, 144, 92–109. [97]

Jones, B., Lekivetz, R., Majumdar, D., Nachtsheim, C. J., and Stallrich, J. W. (2020), Construction, Properties, and Analysis of Group-Orthogonal Supersaturated Designs," *Technometrics*, 62, 403–414. [96]

Jones, B. A., Li, W., Nachtsheim, C. J., and Kenny, Q. Y. (2009), "Model-Robust Supersaturated and Partially Supersaturated Designs," *Journal of Statistical Planning and Inference*, 139, 45–53. [97]

Jones, B. A., Lin, D. K. J., and Nachtsheim, C. (2008), "Bayesian D-optimal Supersaturated Designs," *Journal of Statistical Planning and Inference*, 138, 86–92. [96,97]

Jones, B. A., and Majumdar, D. (2014), "Optimal Supersaturated Designs," *Journal of the American Statistical Association*, 109, 1592–1600. [96,97]

Lin, D. K. J. (1993), "A New Class of Supersaturated Designs," *Technometrics*, 35, 28–31. [97]

Lu, L., Anderson-Cook, C. M., and Robinson, T. J. (2011), "Optimization of Designed Experiments based on Multiple Criteria Utilizing a Pareto Frontier," *Technometrics*, 53, 353–365. [99]

Marley, C. J., and Woods, D. C. (2010), "A Comparison of Design and Model Selection Methods for Supersaturated Experiments," *Computational Statistics & Data Analysis*, 54, 3158–3167. [96,97,98]

Nguyen, N.-K. (1996), "An Algorithmic Approach to Constructing Supersaturated Designs," *Technometrics*, 38, 69–73. [97]

Nguyen, N.-K., and Cheng, C.-S. (2008), "New *E(s²)*-Optimal Supersaturated Designs Constructed from Incomplete Block Designs," *Technometrics*, 50, 26–31. [97]

Phoa, F. K., Pan, Y. H., and Xu, H. (2009), "Analysis of Supersaturated Designs via the Dantzig Selector," *Journal of Statistical Planning and Inference*, 139, 2362–2372. [96,97]

Tang, B., and Wu, C. F. J. (1997), "A Method for Constructing Super-Saturated Designs and its Es^2 Optimality," *Canadian Journal of Statistics*, 25, 191–201. [97]

Weese, M. L., Edwards, D. J., and Smucker, B. J. (2017), "A Criterion for Constructing Powerful Supersaturated Designs when Effect Directions are Known," *Journal of Quality Technology*, 49, 265–277. [96,97]

Weese, M. L., Smucker, B. J., and Edwards, D. J. (2015), "Searching for Powerful Supersaturated Designs," *Journal of Quality Technology*, 47, 66–84. [96,97,98]

Weese, M. L., Stallrich, J. W., Smucker, B. J., and Edwards, D. J. (2021), "Strategies for Supersaturated Screening: Group Orthogonal and Constrained Var(s) Designs," *Technometrics*, 63, 443–455. [96,97,98,99,101,103]