Statistica Sinica Preprint No: SS-2019-0416

Title	A UNIFIED FRAMEWORK FOR MINIMUM ABERRATION
Manuscript ID	SS-2019-0416
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0416
Complete List of Authors	Ming-Chung Chang
Corresponding Author	Ming-Chung Chang
E-mail	mcchang0131@gmail.com
Notice: Accepted version subject to English editing.	

A UNIFIED FRAMEWORK FOR MINIMUM ABERRATION

Ming-Chung Chang

National Central University

Abstract: Minimum aberration has been ubiquitously adopted for selecting fractional factorial designs. Much work has been done on its various extensions, from which many fields of experimental design have benefited, including multi-stratum designs, multi-group designs, and multi-platform designs. However, most of these extensions are ad hoc and are developed on case-by-case bases without strong statistical justifications and a unified rationale. In this paper, we provide a new perspective to minimum aberration through a Bayesian approach. Our theory not only features a unified framework for minimum aberration and is easily applied to many situations but also enables experimenters to derive their own aberration criteria. Several theoretical results as well as three numerical illustrations are provided.

Key words and phrases: Bayesian, Fractional factorial, Multi-stratum, Multi-platform, Multi-group, Mixed-level, Blocking, Split-plot, Strip-plot.

1. Introduction

Minimum aberration has been well-known for decades. The first aberration criterion was proposed by Fries and Hunter (1980) and has been popular for assessing fractional factorial designs. It is especially beneficial when experimenters have little knowledge about the potentially important factorial effects. This criterion was originally developed for evaluating *regular* fractional factorial designs with *unstructured experimental units*. Readers can refer to Wu and Hamada (2009) and Cheng (2014) for a comprehensive account.

In the past two decades, several modifications of the aberration criterion in Fries and Hunter (1980) have been proposed for many sophisticated scenarios, including those for nonregular designs, block designs, and split-plot designs (Dean et al., 2015). Sitter, Chen, and Feder (1997), Chen and Cheng (1999), and Cheng and Wu (2002) developed aberration criteria for blocked two-level regular fractional factorial designs. Cheng, Li, and Ye (2004) proposed a version for blocked two-level nonregular fractional factorial designs. Lin (2014) extended the results in Cheng, Li, and Ye (2004) to blocked mixed-level orthogonal arrays. In addition to block designs, the minimum aberration has been used or modified to split-plot designs as well. Huang, Chen, and Voelkel (1998), Bingham and Sitter (1999), and Bingham, Schoen, and Sitter (2004) used it to compare two-level split-plot designs. Tichon, Li, and Mcleod (2012) considered selecting split-plot designs under five scenarios, each associated with a modified aberration criterion. Yang and Lin (2017) utilized the same approach as in Lin (2014) to develop an aberration criterion for mixed-level split-plot designs.

An aberration criterion is mathematically formulated by a *wordlength pattern*, which requires an order of desirability among pertinent words. In the literature, however, most wordlength patterns are ad hoc modifications of that of Fries and Hunter (1980) and lack strong statistical justifications. For block designs, one needs to argue an order between *block defining words* and *treatment defining words*, while three distinct orders were individually proposed by Sitter, Chen, and Feder (1997), Chen and Cheng (1999), and Cheng and Wu (2002). Apart from the difficulty of judging an appropriate order, the lengths of defining words do not provide enough information for ranking designs for many situations, such as blocked nonregular designs, because designs that can estimate the same number of models may have different estimation efficiencies, not to mention to account for the structures of experimental units.

In this paper, we aim at developing a unified theory of aberration criteria for various scenarios in the literature based on a statistically meaningful framework. Moreover, our theory yields a systematic method for experimenters to derive aberration criteria appropriate for specific experimental conditions. Another work relevant to ours is Cheng and Tang (2005), in which the notion of *minimizing contamination* was adopted. However, Cheng and Tang (2005) studied two-level factorial designs with unstructured experimental units. Our theory, by means of a Bayesian approach, has a sound statistical rationale and can be used to assess and compare mixed-level fractional factorial designs with experimental units that have complex structures.

In our work, the treatment factors are allowed to have multiple groups in the sense that those in the same group are assumed to have (nearly) equal importance on the response. This setting has been considered in the literature, such as *control factors* and *noise factors* in robust parameter designs (Taguchi, 1987). Zhu (2003) studied two-level factorial designs with multiple groups of treatment factors. Tichon, Li, and Mcleod (2012) investigated optimal split-plot designs with two groups of treatment factors, separately corresponding to the whole-plot and subplot strata. Recently, an application of multi-group treatment factors was studied in *multi-platform* experiments (Sadeghi, Qian, and Arora, 2016, 2017), where the *sliced factor* itself is in one group and has higher importance than the other factors.

Li, Zhou, and Zhang (2015) and Li, Mee, and Zhou (2018) proposed new aberration criteria for factorial designs with multiple groups of treatment factors. We discuss applying our work to multi-platform experiments in Section S5 of the supplementary material.

This paper is organized as follows. Section 2 provides necessary preliminaries. Section 3 gives the theoretical results of our work and introduces a general aberration criterion with some applications. Section 4 illustrates minimum aberration designs under three settings: unstructured units, blocked mixed-level orthogonal arrays, and three-stage manufacturing processes. Finally, Section 5 concludes this paper. All proofs are deferred to the supplementary material.

2. Preliminaries

2.1 Unit factors and block structures

The experimental units considered in this paper have a structure, which is hereafter referred to as a *block structure.* Many common block structures, such as block designs, split-plot designs, strip-plot designs, and block strip-plot designs, belong to a specific class of block structures: *simple block structures* (Nelder, 1965a,b). A larger class of block structures, covering simple block structures and most block structures commonly encountered in practice, is the *orthogonal block structures* (Speed and Bailey, 1982; Bailey, 1985). Readers can refer to Bailey (2008) and Cheng (2014) for details.

We denote the number of experimental units by N. A block structure can be described by a set of *unit factors* defined on the experimental units. An $n_{\mathcal{F}}$-level unit factor $\mathcal{F}$ can be thought of as a partition of the N units into $n_{\mathcal{F}}$ disjoint subsets. Each subset is called an $\mathcal{F}$-class and consists of units that have the same level of $\mathcal{F}$. A unit factor is said to be *uniform* if all of its classes are of the same size. For two different unit factors $\mathcal{F}_1$ and $\mathcal{F}_2$, we say that $\mathcal{F}_1$ is *nested in* (or *finer than*) $\mathcal{F}_2$, denoted by $\mathcal{F}_1 \prec \mathcal{F}_2$, if two units in the same $\mathcal{F}_1$-class implies that they are in the same $\mathcal{F}_2$-class. The expression $\mathcal{F}_1 \preceq \mathcal{F}_2$ stands for either $\mathcal{F}_1 \prec \mathcal{F}_2$ or $\mathcal{F}_1 = \mathcal{F}_2$. The finest unit factor, denoted by $\mathcal{E}$, has N levels, with each class consisting of one single unit. On the other hand, $\mathcal{U}$ denotes the unit factor that has a single level with all units in the same class. A split-plot design has the block structure $\{\mathcal{U}, \mathcal{P}, \mathcal{E}\}$, where $\mathcal{P}$ partitions the N units into $n_{\mathcal{P}}$ whole-plots. We always include $\mathcal{U}$ and $\mathcal{E}$ into every block structure. A set of unstructured units can be treated as having the block structure $\{\mathcal{U}, \mathcal{E}\}$.

In this paper, we consider block structures that satisfy conditions (i), (ii), (iii), (v), and (vi) in Definition 12.4 in Cheng (2014, p. 233), which cover orthogonal block structures. To save space, these five conditions, denoted by (S1.1)-(S1.5), and their importance for the theoretical results in our work are given in Section S1 of the supplementary material. We note that the block structures of most experiments encountered in practice, such as blocked, split-plot, or strip-plot factorial experiments, satisfy (S1.1)-(S1.5).

2.2 Treatment factorial effects

Suppose there are n treatment factors with levels $p_1, ..., p_n$, and denote $\prod_{i=1}^{n} p_i$ by Ξ. Let β_0 be the intercept and $\beta_1, ..., \beta_{\Xi-1}$ be the $\Xi - 1$ factorial effects. Denote the $\Xi \times 1$ vector of all β_j's by $\boldsymbol{\beta}$. Let $\boldsymbol{\alpha}$ be the $\Xi \times 1$ vector of the effects of all Ξ treatment combinations. Then, $\boldsymbol{\alpha}$ can be expressed as $\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta}$, where $\mathbf{P}$ is a $\Xi \times \Xi$ full model matrix for a complete factorial experiment with $\mathbf{P}^T\mathbf{P} = \mathbf{I}_\Xi$. It follows that $\mathbf{P}^{-1} = \mathbf{P}^T$ and $\boldsymbol{\beta} = \mathbf{P}^T\boldsymbol{\alpha}$.

The matrix $\mathbf{P}$ can be systematically constructed based on Kurkjian and Zelen (1962) as follows. For each factor $i = 1, ..., n$, define a $p_i \times p_i$ orthogonal matrix $\mathbf{P}_i$ with the first column proportional to the all-one vector. Then, let the remaining $p_i - 1$ columns define $p_i - 1$ treatment contrasts of the main effects of factor i. If $p_1 = 3$, for example, a choice of $\mathbf{P}_1$ is $\begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}$, in which the first column represents the intercept, the second column represents the linear main effect, and the third column represents the quadratic main effect. Once $\mathbf{P}_1, ..., \mathbf{P}_n$ have been constructed, one can obtain $\mathbf{P}$ by

$$\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n, \tag{2.1}$$

where $\otimes$ denotes the Kronecker product operator.

The components of $\boldsymbol{\beta}$ can be divided into 2^n groups in terms of the treatment factors involved. Let S be a subset of $\{1, ..., n\}$, where the empty set is denoted by ϕ. Each S represents one such group and corresponds to certain β_j's. For example, $S = \phi$ corresponds to the intercept; $S = \{i\}$ corresponds to the $p_i - 1$ main effects of factor i; $S = \{i_1, ..., i_k\}$ corresponds to the $(p_{i_1} - 1) \cdots (p_{i_k} - 1)$ k-factor interactions among factors $i_1, ..., i_k$.

In this paper, we adopt a Bayesian framework for $\boldsymbol{\beta}$. To specify the prior distribution of $\boldsymbol{\beta}$, we assume that $\boldsymbol{\beta}$ comprises uncorrelated random variables and follows a zero-mean multivariate normal distribution with $\mathrm{var}(\beta_l) = \mathrm{var}(\beta_j)$ if both β_l and β_j are associated with the same S. Hence, there are at most 2^n distinct values of $\mathrm{var}(\beta_i)$'s. These values are denoted by v_S, $S \subseteq \{1, ..., n\}$. Furthermore, we require

$$v_S \geq v_{S'} \text{ if } S \subset S'. \tag{2.2}$$

This requirement, referred to as the property of *nested decreasing interaction variances* in Kerr (2001), is consistent with the *effect heredity principle*

(Yates, 1935; Wu and Hamada, 2009, p. 172). This Bayesian framework is inspired by Mitchell, Morris, and Ylvisaker (1995), Kerr (2001), Joseph (2006), and Joseph and Delaney (2007). A common technique of their approaches is to induce the prior distribution of $\boldsymbol{\beta}$ from $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is assumed to be a realization of a stationary Gaussian process. Some results of the prior distribution of $\boldsymbol{\beta}$ are given in Section S2 of the supplementary material.

2.3 Statistical model

Suppose N experimental units have a block structure $\mathfrak{B} = \{\mathcal{F}_0, \mathcal{F}_1, ..., \mathcal{F}_m\}$, where $\mathcal{F}_0 = \mathcal{U}$ and $\mathcal{F}_m = \mathcal{E}$. For each $\mathcal{F}_i \in \mathfrak{B}$, let $\mathbf{X}_{\mathcal{F}_i}$ be an $N \times n_{\mathcal{F}_i}$ *incidence* matrix that describes the relationship between the units and the levels of $\mathcal{F}_i$. Each entry of $\mathbf{X}_{\mathcal{F}_i}$ is 0 or 1 such that the ljth entry of $\mathbf{X}_{\mathcal{F}_i}$ is 1 if and only if the lth unit is in the jth $\mathcal{F}_i$-class.

Under a fractional factorial design d with N treatment combinations, let

$$\mathbf{y} = \mathbf{U}\boldsymbol{\beta} + \sum_{i=0}^{m} \mathbf{X}_{\mathcal{F}_i}\boldsymbol{\gamma}^{\mathcal{F}_i},$$

where $\mathbf{y}$ is a vector of responses, $\mathbf{U}$ is the $N \times \Xi$ full model matrix under d (composed of N corresponding rows from $\mathbf{P}$) and $\boldsymbol{\gamma}^{\mathcal{F}_i} = (\gamma_1^{\mathcal{F}_i}, \ldots, \gamma_{n_{\mathcal{F}_i}}^{\mathcal{F}_i})^T$ with $\gamma_j^{\mathcal{F}_i}$ being the effect of the jth level of unit factor $\mathcal{F}_i$ (e.g., block effects, whole-plot effects, and subplot effects). We assume that the $\gamma_j^{\mathcal{F}_i}$'s are uncorrelated, with each $\gamma_j^{\mathcal{F}_i}$ following a zero-mean normal distribution with variance $\sigma^2_{\mathcal{F}_i}$, and that they are independent of $\boldsymbol{\beta}$. Then, the conditional distribution of $\mathbf{y}$ given $\boldsymbol{\beta}$ is the multivariate normal distribution:

$$\mathbf{y}|\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{U}\boldsymbol{\beta}, \sum_{i=0}^{m} \sigma^2_{\mathcal{F}_i}\mathbf{X}_{\mathcal{F}_i}\mathbf{X}^T_{\mathcal{F}_i}). \tag{2.3}$$

Let $\mathbf{V} = \sum_{i=0}^{m} \sigma^2_{\mathcal{F}_i}\mathbf{X}_{\mathcal{F}_i}\mathbf{X}^T_{\mathcal{F}_i}$. If $\mathfrak{B}$ satisfies conditions (S1.1)-(S1.5), then $\mathbf{V}$ has $m+1$ eigenspaces $W_{\mathcal{F}_0}, \ldots, W_{\mathcal{F}_m}$, with one eigenspace associated with each of the $m+1$ unit factors, where $W_{\mathcal{F}_0} = W_{\mathcal{U}}$ is the one-dimensional space consisting of all the vectors with constant entries, and each other eigenvector defines a unit contrast (Cheng, 2014, p. 237). It follows that $\sum_{i=0}^{m} \mathbf{P}_{W_{\mathcal{F}_i}} = \mathbf{I}_N$, where $\mathbf{P}_{W_{\mathcal{F}_i}}$ is the orthogonal projection matrix onto $W_{\mathcal{F}_i}$. Let the corresponding eigenvalues be $\xi_{\mathcal{F}_0}, \ldots, \xi_{\mathcal{F}_m}$. Here, $W_{\mathcal{F}_i}$ and $\xi_{\mathcal{F}_i}$ are called a *stratum* and stratum variance, respectively. It can be shown that $\xi_{\mathcal{F}_i} \leq \xi_{\mathcal{F}_j}$ if $\mathcal{F}_i \preceq \mathcal{F}_j$ (Cheng, 2014, p. 246). The case where $\gamma_1^{\mathcal{F}_i}, \ldots, \gamma_{n_{\mathcal{F}_i}}^{\mathcal{F}_i}$ are unknown constants (fixed effects) can be treated by letting $\sigma^2_{\mathcal{F}_i} = \infty$, leading to $\xi_{\mathcal{F}_j} = \infty$ if $\mathcal{F}_i \preceq \mathcal{F}_j$.

A systematic method to construct $\mathbf{P}_{W_{\mathcal{F}}}$ is as follows. Define $V_{\mathcal{F}}$ as the column space of $\mathbf{X}_{\mathcal{F}}$ for each $\mathcal{F} \in \mathfrak{B}$. The orthogonal projection matrix onto $V_{\mathcal{F}}$ is $\mathbf{P}_{V_{\mathcal{F}}} = \mathbf{X}_{\mathcal{F}}(\mathbf{X}^T_{\mathcal{F}}\mathbf{X}_{\mathcal{F}})^{-1}\mathbf{X}^T_{\mathcal{F}}$. It can be shown that $\mathbf{P}_{W_{\mathcal{F}}} = \mathbf{P}_{V_{\mathcal{F}}} - \sum_{\mathcal{G}\in\mathfrak{B}:\mathcal{F}\prec\mathcal{G}} \mathbf{P}_{W_{\mathcal{G}}}$. Thus, one can obtain every $\mathbf{P}_{W_{\mathcal{F}}}$ by starting from $\mathbf{P}_{W_{\mathcal{U}}} =$

$\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$. More details can be found in Cheng (2014, p. 243).

3. A general aberration criterion

In this section, we propose an aberration criterion for design assessment and selection based on the Bayesian approach. This criterion is capable of handling mixed-level treatment factors as well as complex structures of experimental units. In addition, it is easily modified according to experimenters' beliefs about important factorial effects. Sections 3.1 to 3.3 illustrate its three common applications.

From (2.2) and (2.3), the posterior distribution $\boldsymbol{\beta}|\mathbf{y}$ is multivariate normal with a mean vector and the covariance matrix $\text{cov}(\boldsymbol{\beta}|\mathbf{y}) = \boldsymbol{\Sigma}_\beta - \boldsymbol{\Sigma}_\beta\mathbf{U}^T(\mathbf{U}\boldsymbol{\Sigma}_\beta\mathbf{U}^T + \mathbf{V})^{-1}\mathbf{U}\boldsymbol{\Sigma}_\beta$, where $\boldsymbol{\Sigma}_\beta$ is the (prior) covariance matrix of $\boldsymbol{\beta}$. Let $\mathbf{M} = \text{cov}(\boldsymbol{\beta}|\mathbf{y})^{-1}$. A commonly used design selection criterion, *Bayesian D-optimality*, is to maximize $\det[\mathbf{M}]$. While the D-optimality has good statistical interpretation, it is not easily manageable. A good surrogate for the D-optimality, referred to as the *(M.S)-optimality* due to Eccleston and Hedayat (1974), is to first maximize $\text{tr}[\mathbf{M}]$ and then minimize $\text{tr}[\mathbf{M}^2]$ among the designs that maximize $\text{tr}[\mathbf{M}]$.

For each $S \subseteq \{1, ..., n\}$, let $\mathbf{U}_S$ be composed of the columns in $\mathbf{U}$ associated with S. If $S = \{1, 2\}$ with $p_1 = 2$ and $p_2 = 3$, for example, then

$\mathbf{U}_S$ consists of $(2-1)(3-1) = 2$ columns, each representing a treatment contrast of the two-factor interaction between factors 1 and 2 under the given design.

Define

$$\Phi_1(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^{m} \sum_{S \subseteq \{1,\dots,n\}} \frac{v_S}{\xi_{\mathcal{F}_i}} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right],$$

$$\Phi_2(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^{m} \frac{1}{\xi_{\mathcal{F}_i}^2} \mathrm{tr}\left[(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U})^2\right]$$
$$+ 2 \sum_{0 \le l < s \le m} \frac{1}{\xi_{\mathcal{F}_l} \xi_{\mathcal{F}_s}} \mathrm{tr}\left[(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_l}} \mathbf{U})(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_s}} \mathbf{U})\right],$$

where $\mathbf{v}$ and $\boldsymbol{\xi}$ are the vectors of v_S's and $\xi_{\mathcal{F}_i}$'s. We have the following result for the Bayesian (M.S)-optimality.

Theorem 1. *The Bayesian (M.S)-optimality involves to first maximize* $\Phi_1(d; \boldsymbol{\xi}, \mathbf{v})$ *and then minimize* $\Phi_2(d; \boldsymbol{\xi}, \mathbf{v})$ *among the designs that maximize* $\Phi_1(d; \boldsymbol{\xi}, \mathbf{v})$.

To get a more structured form of $\Phi_1(d; \boldsymbol{\xi}, \mathbf{v})$, we need Lemmas 1 and 2 in Section S3 of the supplementary material, which jointly state that $\mathrm{tr}\left[\mathbf{U}_S^T \mathbf{U}_S\right]$ does not depend on the choice of designs and orthogonal-column-bases of the column space of $\mathbf{P}$. We summarize these as a theorem.

Theorem 2. *For an* $S \subseteq \{1, \dots, n\}$, *tr* $\left[\mathbf{U}_S^T \mathbf{U}_S\right]$ *is a constant for any choice of* N*-run designs as well as for any choice of orthogonal-column-bases in*

P.

With Theorem 2 and the property $\sum_{i=0}^{m} \mathbf{P}_{W_{\mathcal{F}_i}} = \mathbf{I}_N$, maximizing $\Phi_1(d;\boldsymbol{\xi},\mathbf{v})$ is reduced to minimizing

$$\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v}) = \sum_{i=0}^{m-1} \sum_{S\subseteq\{1,\ldots,n\}} v_S \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}}\right) \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right].$$

by replacing $\mathbf{P}_{W_{\mathcal{F}_m}}$ with $\mathbf{I}_N - \sum_{i=0}^{m-1} \mathbf{P}_{W_{\mathcal{F}_i}}$.

In addition to the choice of designs, $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ and $\Phi_2(d;\boldsymbol{\xi},\mathbf{v})$ depend on unknown parameters $\mathbf{v}$ and $\boldsymbol{\xi}$. The following result serves as a useful tool for searching for optimal designs with respect to minimizing $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ for all *feasible* $\mathbf{v}$ and $\boldsymbol{\xi}$. In this paper, $\mathbf{v}$ and $\boldsymbol{\xi}$ are said to be feasible if $\mathbf{v}$ satisfies (2.2) and $\mathcal{F}_i \prec \mathcal{F}_j$ implies $\xi_{\mathcal{F}_i} \leq \xi_{\mathcal{F}_j}$.

Theorem 3. *Suppose $\mathfrak{B}$ is a block structure satisfying conditions (S1.1)-(S1.5). Then, a necessary and sufficient condition for a design to minimize $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$ is that it minimizes*

$$\sum_{S\in\mathfrak{S}} \sum_{i:\mathcal{F}_i\in\mathfrak{G}} tr\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right]$$

for all nonempty subsets $\mathfrak{S} \subseteq 2^{\{1,\ldots,n\}} \setminus \{\phi\}$ and $\mathfrak{G} \subseteq \mathfrak{B} \setminus \{\mathcal{F}_m\}$ such that

$$S \in \mathfrak{S}, S' \in 2^{\{1,\ldots,n\}} \setminus \{\phi\},\ \textit{and } S' \subset S \Rightarrow S' \in \mathfrak{S}, \tag{3.4}$$

$$\mathcal{F} \in \mathfrak{G}, \mathcal{F}' \in \mathfrak{B},\ \textit{and } \mathcal{F} \prec \mathcal{F}' \Rightarrow \mathcal{F}' \in \mathfrak{G}. \tag{3.5}$$

We illustrate Theorem 3 with a simple scenario. Suppose $n = 2$ and $\mathfrak{B} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}$ with $\mathcal{F}_2 \prec \mathcal{F}_1 \prec \mathcal{F}_0$. All subsets of $2^{\{1,2\}} \setminus \{\phi\} = \{\{1\}, \{2\}, \{1,2\}\}$ that satisfy (3.4) are $\mathfrak{S}_1 = \{\{1\}\}$, $\mathfrak{S}_2 = \{\{2\}\}$, $\mathfrak{S}_3 = \{\{1\}, \{2\}\}$, $\mathfrak{S}_4 = \{\{1\}, \{2\}, \{1,2\}\}$. Likewise, all subsets of $\mathfrak{B} \setminus \{\mathcal{F}_2\}$ that satisfy (3.5) are $\mathfrak{G}_1 = \{\mathcal{F}_0\}$, $\mathfrak{G}_2 = \{\mathcal{F}_0, \mathcal{F}_1\}$. By Theorem 3, if a design minimizes $\sum_{S \in \mathfrak{S}_i} \sum_{j:\mathcal{F}_j \in \mathfrak{G}_l} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_j}} \mathbf{U}_S\right]$ for $i = 1, ..., 4$ and $l = 1, 2$, then it minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$.

Theorem 3 extends Theorem 5.1 in Chang and Cheng (2018) in two ways. First, Chang and Cheng (2018) is limited to two-level designs, while here we are able to deal with the case of mixed-level treatment factors. Second, Theorem 3 provides a sufficient and necessary condition for a design to be optimal for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$, while Theorem 5.1 in Chang and Cheng (2018) requires the values of $\mathbf{v}$.

Similar to Chang and Cheng (2018), Theorem 3 is able to eliminate inferior designs. For two designs d_1 and d_2, if $\sum_{S \in \mathfrak{S}} \sum_{i:\mathcal{F}_i \in \mathfrak{G}} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right]$ of d_1 is no greater than that of d_2 under every combination of $\mathfrak{S}$ and $\mathfrak{G}$, with strict inequality for at least one combination, then d_2 is worse than d_1 and is said to be *inadmissible*. Eliminating inadmissible designs yields a considerable reduction of designs that need to be considered. If there remains one design (up to isomorphism), it minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$. Usually, using $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ is enough to distinguish designs. If there remain more than one nonisomorphic designs, one can use either $\Phi_2(d;\boldsymbol{\xi},\mathbf{v})$ or the actual Bayesian D-optimal criterion to assess them.

In the remaining part of this section, we illustrate equivalent forms of minimizing $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ under several specific scenarios. Some are reduced to well-known aberration criteria. To define an aberration criterion, one needs a desirability order about the importance of factorial effects. This can be achieved under appropriate settings of the values of $\mathbf{v}$.

If it is known that the 2^n subsets of $\{1,...,n\}$ can be divided into J groups $\mathfrak{H}_1,...,\mathfrak{H}_J$, such that $v_S = v_{S'}$ for S, S' in the same group and $v_S > v_{S'}$ for $S \in \mathfrak{H}_l$ and $S' \in \mathfrak{H}_{l'}$ with $l < l'$, then, since $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ is linear in v_S's, the following wordlength pattern is induced:

$$\sum_{i=0}^{m-1}\left\{\left(\frac{1}{\xi_{\mathcal{F}_m}}-\frac{1}{\xi_{\mathcal{F}_i}}\right)\left(\sum_{S\in\mathfrak{H}_1}\operatorname{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right],\ldots,\sum_{j\in\mathfrak{H}_J}\operatorname{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right]\right)\right\}. \tag{3.6}$$

An aberration criterion can be defined as sequentially minimizing this wordlength pattern. Since $\operatorname{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right] = \operatorname{tr}\left[\mathbf{U}_S\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\right]$, it follows from the proof of Lemma 2 (in the supplementary material) that (3.6) does not depend on orthogonal bases in $\mathbf{P}$.

If, on the other hand, the information about important factorial effects is vague, then the effect hierarchy principle in Wu and Hamada (2009,

p. 172) is often assumed, especially for screening experiments (Dean and Lewis, 2006). Under the Bayesian framework, this principle is basically consistent with choosing $\mathfrak{H}_l = \{S \subseteq \{1, ..., n\} : |S| = l\}$, $l = 1, ..., n$; or equivalently,

$$\text{(i) } v_S = v_{S'} \text{ if } |S| = |S'|,$$
$$\text{(ii) } v_S > v_{S'} \text{ if } |S| < |S'|. \tag{3.7}$$

It is obvious that (3.7) satisfies (2.2). By replacing "$S' \subset S$" in (3.4) with "$v_{S'} \geq v_S$", we can establish another version of Theorem 3, which is tailored to the setting in (3.7).

Theorem 4. *Suppose $\mathfrak{B}$ is a block structure satisfying conditions (S1.1)-(S1.5). Then, under (3.7), a necessary and sufficient condition for a design to minimize $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\mathbf{v}$ that satisfy (3.7) and feasible $\boldsymbol{\xi}$ is that it minimizes*

$$\sum_{S \in \mathfrak{G}} \sum_{i : \mathcal{F}_i \in \mathfrak{G}} tr\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right]$$

for all nonempty subsets $\mathfrak{G} \subseteq \mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (3.5) and $\mathfrak{S} \subseteq 2^{\{1,\ldots,n\}} \setminus \{\phi\}$ satisfying

$$S \in \mathfrak{S}, S' \in 2^{\{1,\ldots,n\}} \setminus \{\phi\}, \text{ and } v_{S'} \geq v_S \Rightarrow S' \in \mathfrak{S}. \tag{3.8}$$

For $n = 3$, all nonempty subsets of $2^{\{1,2,3\}} \backslash \{\phi\} = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$ that satisfy (3.8) are $\mathfrak{S}_k = \{S \subseteq \{1, ..., n\} : 0 < |S| \leq k\}$, $k = 1, 2, 3$, each corresponding to main effects, effects up to two-factor interactions, or effects up to the three-factor interaction.

When (3.7) holds, with an additional requirement that $v_S \gg v_{S'}$ if $|S| < |S'|$ (i.e., lower-order effects are much more important than higher-order ones), minimizing $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ is equivalent to sequentially minimizing

$$\mathfrak{W} = \sum_{i=0}^{m-1} \left\{ \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) \left(\sum_{S:|S|=1} \operatorname{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right], \ldots, \sum_{S:|S|=n} \operatorname{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right] \right) \right\}.$$

The $\mathfrak{W}$ can be regarded as a wordlength pattern and induces an aberration criterion for complex block structures. This criterion, not an ad hoc one, is developed based on good properties of a statistical model. If $\boldsymbol{\xi}$ are known, their values can be inserted. Otherwise, based on Theorem 4, a design sequentially minimizes $\mathfrak{W}$ for all feasible $\boldsymbol{\xi}$ provided that it sequentially minimizes

$$\mathfrak{W}_{\mathfrak{G}} = \left(\sum_{i:\mathcal{F}_i \in \mathfrak{G}} \sum_{S:|S|=1} \operatorname{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right], \ldots, \sum_{i:\mathcal{F}_i \in \mathfrak{G}} \sum_{S:|S|=n} \operatorname{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right] \right)$$

for all $\mathfrak{G} \subseteq \mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (3.5).

We note that each $\mathfrak{W}_{\mathfrak{G}}$ can be regarded as a wordlength pattern and induces an aberration criterion for the block structure $\mathfrak{G} \cup \{\mathcal{F}_m\}$, where all unit effects are fixed effects; that is, $\xi_{\mathcal{F}} = \infty$ if $\mathcal{F} \in \mathfrak{G}$, because under the block structure $\mathfrak{G} \cup \{\mathcal{F}_m\}$,

$$\lim_{\xi_{\mathcal{F}} \to \infty : \mathcal{F} \in \mathfrak{G}} \Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v}) \propto \sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{S \subseteq \{1,\ldots,n\}} v_S \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right].$$

Consequently, if a design has minimum aberration under each case of fixed unit effects (i.e., $\mathfrak{W}_{\mathfrak{G}}$ with $\mathfrak{G}$ satisfying (3.5)), then it has minimum aberration under random unit effects (i.e., $\mathfrak{W}$).

The aberration criterion induced by $\mathfrak{W}$ can be applied to any block structure that satisfies conditions (S1.1)-(S1.5). In Sections 3.1 to 3.3, we introduce three common applications.

As a remark, if a finer hierarchy exists among β_j's such that all the β_j's can be divided into K groups $\mathfrak{I}_1, ..., \mathfrak{I}_K$, with those in the same group having equal variance and $\mathrm{var}(\beta_j) > \mathrm{var}(\beta_{j'})$ for $\beta_j \in \mathfrak{I}_l$ and $\beta_{j'} \in \mathfrak{I}_{l'}$ with $l < l'$, then a more flexible version of (3.6) is

$$\sum_{i=0}^{m-1} \left\{ \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) \left(\mathrm{tr}\left[\mathbf{U}_1^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_1\right], \ldots, \mathrm{tr}\left[\mathbf{U}_K^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_K\right] \right) \right\}, \qquad (3.9)$$

where $\mathbf{U}_l$ is composed of the columns in $\mathbf{U}$ associated with the β_j's belonging to $\mathfrak{I}_l$. The above one is useful for certain situations, such as multi-platform experiments and experiments with quantitative treatment factors.

3.1 Unstructured units

For unstructured experimental units, the block structures are denoted by $\{\mathcal{F}_0, \mathcal{F}_1\}$ with $\mathcal{F}_0 = \mathcal{U}$ and $\mathcal{F}_1 = \mathcal{E}$. Since $W_{\mathcal{F}_0}$ is spanned by the vector of ones, we have $\mathbf{P}_{W_{\mathcal{F}_0}} = \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$ and $\mathbf{P}_{W_{\mathcal{F}_1}} = \mathbf{I}_N - \mathbf{P}_{W_{\mathcal{F}_0}}$. It follows that sequentially minimizing $\mathfrak{W}$ is equivalent to sequentially minimizing

$$\mathfrak{W}_0 = \left(\sum_{S:|S|=1} (\mathbf{1}_N^T\mathbf{U}_S)(\mathbf{1}_N^T\mathbf{U}_S)^T, \ldots, \sum_{S:|S|=n} (\mathbf{1}_N^T\mathbf{U}_S)(\mathbf{1}_N^T\mathbf{U}_S)^T \right). \quad (3.10)$$

As given by Cheng (2014, p. 340), the wordlength pattern of the *generalized aberration criterion* proposed by Xu and Wu (2001) takes the form: $\frac{\Xi}{N^2}\sum_{S:|S|=k}(\mathbf{1}_N^T\mathbf{U}_S)(\mathbf{1}_N^T\mathbf{U}_S)^T$, $k = 1, ..., n$. Thus, it is equivalent to sequentially minimizing $\mathfrak{W}_0$. Moreover, it follows from Theorem 4 that if a design minimizes $\sum_{S:0<|S|\le k}(\mathbf{1}_N^T\mathbf{U}_S)(\mathbf{1}_N^T\mathbf{U}_S)^T$ for all $k = 1, ..., n$, then it minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\mathbf{v}$ satisfying (3.7); based on this, a generalized minimum aberration design must not be inadmissible. The following result implies that a design cannot minimize $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\mathbf{v}$ satisfying (3.7) if it has replication.

Theorem 5. *If an N-run design consists of m replicates, then*

$$\sum_{k=0}^{n} \sum_{S:|S|=k} (\mathbf{1}_N^T\mathbf{U}_S)(\mathbf{1}_N^T\mathbf{U}_S)^T = N + 2m.$$

Theorem 5 discloses a disadvantage of using designs with replicates in terms of estimating factorial effects. By Theorem 5, for two designs with the same run size, the one with more replicates has a larger value of $\sum_{k=0}^{n}\sum_{S:|S|=k}(\mathbf{1}_N^T\mathbf{U}_S)(\mathbf{1}_N^T\mathbf{U}_S)^T$. Thus, it does not reach the necessary and sufficient condition in Theorem 4 and cannot minimize $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ for all $\mathbf{v}$ satisfying (3.7). This is not surprising since replicates do not provide any information about factorial effects.

3.2 A chain of nested unit factors

In many real applications, the experimental units are partitioned by a chain of nested unit factors, such as block designs, split-plot designs, and split-split plot designs.

Without loss of generality, suppose the block structure is $\{\mathcal{F}_0,\mathcal{F}_1,...,\mathcal{F}_m\}$ with $\mathcal{F}_i\prec\mathcal{F}_j$ if $i>j$, where block designs or split-plot designs correspond to $m=2$ and split-split plot designs correspond to $m=3$. Since the $\mathfrak{G}$'s that satisfy (3.5) are $\{\mathcal{F}_0\}$, $\{\mathcal{F}_0,\mathcal{F}_1\}$,..., $\{\mathcal{F}_0,\mathcal{F}_1,\mathcal{F}_2,...,\mathcal{F}_{m-1}\}$, it follows from Theorem 4 that a design sequentially minimizes $\mathfrak{W}$ for all feasible $\boldsymbol{\xi}$ provided that it sequentially minimizes

$$\mathfrak{W}_l=\left(\sum_{i=0}^{l}\sum_{S:|S|=1}\mathrm{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right],\ldots,\sum_{i=0}^{l}\sum_{S:|S|=n}\mathrm{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right]\right)$$

for all $l=0,1,...,m-1$.

For block or split-plot experiments, we have $m=2$ and $\mathcal{F}_1$ partitions the units into blocks or whole-plots. In this case, we have $\mathbf{P}_{W_{\mathcal{F}_0}}=\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$,

$\mathbf{P}_{W_{\mathcal{F}_1}} = \mathbf{P}_{V_{\mathcal{F}_1}} - \mathbf{P}_{W_{\mathcal{F}_0}}$, $\mathbf{P}_{W_{\mathcal{F}_2}} = \mathbf{I}_N - (\mathbf{P}_{W_{\mathcal{F}_0}} + \mathbf{P}_{W_{\mathcal{F}_1}})$, and

$$\mathfrak{W} = \sum_{i=0,1} \left\{ \left(\frac{1}{\xi_{\mathcal{F}_2}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) \left(\sum_{S:|S|=1} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right], \ldots, \sum_{S:|S|=n} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right] \right) \right\}.$$

Then, we have that if a design sequentially minimizes $\mathfrak{W}_0$ and $\mathfrak{W}_1$, respectively, then it sequentially minimizes $\mathfrak{W}$ for all feasible $\boldsymbol{\xi}$.

Under a block design, $\mathfrak{W}_1$ defines an aberration criterion for models with fixed block effects. By letting

$$\mathfrak{W}_{1,i} = \left(\sum_{S:|S|=1} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right], \ldots, \sum_{S:|S|=n} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S \right] \right), i = 0, 1,$$

we have $\mathfrak{W}_1 = \mathfrak{W}_{1,0} + \mathfrak{W}_{1,1}$. It can be seen that $\mathfrak{W}_{1,0} = \frac{1}{N}\mathfrak{W}_0$, which is proportional to the generalized wordlength pattern; also, $\mathfrak{W}_{1,1}$ defines a wordlength pattern proportional to the *block wordlength pattern* in the literature (e.g., Cheng, Li, and Ye (2004)). Thus, $\mathfrak{W}_1$ combines the treatment wordlength pattern and block wordlength pattern through $\mathfrak{W}_1 = \mathfrak{W}_{1,0} + \mathfrak{W}_{1,1}$, which is different from those in previous works, such as Chen and Cheng (1999); Cheng, Li, and Ye (2004); Lin (2014). For example, Cheng, Li, and Ye (2004) and Lin (2014) proposed two aberration criteria for blocked nonregular designs by arguing two types of desirability between treatment defining words and block defining words. The two wordlength patterns in Cheng, Li, and Ye (2004) are proportional to

$$W_1 = (\delta_{1,0}, \delta_{2,0}, \delta_{1,1}, \delta_{3,0}, \delta_{4,0}, \delta_{2,1}, \delta_{5,0}, \delta_{6,0}, \delta_{3,1}, \delta_{7,0}, \ldots),$$

$$W_2 = (\delta_{1,0}, \delta_{1,1}, \delta_{2,0}, \delta_{3,0}, \delta_{2,1}, \delta_{4,0}, \delta_{5,0}, \delta_{3,1}, \delta_{6,0}, \delta_{7,0}, \ldots),$$

with $\delta_{k,i} = \sum_{S:|S|=k} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right]$. Those defined in Lin (2014) possess the same patterns but were under (3.9) with $\mathfrak{I}_l$ consisting of the β_j's of the same polynomial degree l. It can be seen that $\delta_{k,1}$ precedes $\delta_{2k,0}$ in W_2, while $\delta_{2k,0}$ precedes $\delta_{k,1}$ in W_1. Since $\mathfrak{W}_1 \propto \lim_{\xi_{\mathcal{F}_1} \to \infty} \mathfrak{W}$, we expect $\mathfrak{W}_1$ would produce similar designs to W_2 than to W_1 because W_2 regards confounding treatments with blocks more severe than W_1. However, deciding to use which of W_1 and W_2 heavily relies on subjective judgment. In our work, the use of $\mathfrak{W}_1$ is justified by the Bayesian (M.S)-optimality. In addition, it can be shown that $\mathfrak{W}_1$ tends to maximize D-efficiency under certain fixed-effect models. More details can be found in Section S7 of the supplementary material. A numerical comparison of $\mathfrak{W}_1$, W_1, and W_2 is given in Section 4.2.

3.3 Experiments with multiple processing stages

For experiments with multiple processing stages, the experimental units are partitioned into disjoint classes at each stage. For the treatment factors at some stage, the levels of each of them are randomly assigned to the classes of the partition, with the same level assigned to all the units in the same class. Many industrial experiments have a sequence of processing stages (Mee and Bates, 1998; Butler, 2004; Bingham et al., 2008; Antolino et al., 2009a,b; Ranjan, Bingham, and Dean, 2009; Cheng and Tsai, 2011; Yuangyai and Lin, 2013).

In an experiment with multiple processing stages, the partition of the experimental units at the ith stage defines a unit factor $\mathcal{F}_i$. As mentioned in Cheng and Tsai (2011), the resulting block structure may not satisfy conditions (S1.1)-(S1.5). Cheng and Tsai (2011) proved that if the $\mathcal{F}_i$'s (except $\mathcal{U}$ and $\mathcal{E}$) are uniform, mutually orthogonal, and are not nested in one another, then the resulting block structure satisfies the five conditions if and only if these $\mathcal{F}_i$'s define an orthogonal array of strength two.

Here we consider block structures $\mathfrak{B} = \{\mathcal{U}, \mathcal{E}, \mathcal{F}_1, ..., \mathcal{F}_h\}$, where $\mathcal{F}_1, ..., \mathcal{F}_h$ define an orthogonal array of strength two on the experimental units. Because $\mathcal{E} \prec \mathcal{F}_1, ..., \mathcal{F}_h \prec \mathcal{U}$ and the $\mathcal{F}_i$'s are not nested in one another, the $\mathfrak{G}$'s that satisfy (3.5) are $\{\mathcal{U}\}$, $\{\mathcal{U}, \mathcal{F}_i\}$ with $1 \leq i \leq h$, $\{\mathcal{U}, \mathcal{F}_i, \mathcal{F}_j\}$ with $1 \leq i, j \leq h$,..., $\{\mathcal{U}, \mathcal{F}_1, ..., \mathcal{F}_h\}$. There are 2^h such subsets to be considered. It follows that $\mathbf{P}_{W_\mathcal{U}} = \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$ and $\mathbf{P}_{W_{\mathcal{F}_i}} = \mathbf{P}_{V_{\mathcal{F}_i}} - \mathbf{P}_{W_\mathcal{U}}$ for $i = 1, ..., h$.

The *split-lot* designs in Mee and Bates (1998) belong to this category. Suppose 16 batches of material are to be arranged into four groups of equal size at each of three stages ($h = 3$). From Theorem 4, a design sequentially minimizes $\mathfrak{W}$ for all feasible $\boldsymbol{\xi}$ provided that it sequentially minimizes

$$\mathfrak{W}_{\mathfrak{I}} = \left(\sum_{\mathcal{F} \in \mathfrak{I}} \sum_{S:|S|=1} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}}} \mathbf{U}_S\right], \ldots, \sum_{\mathcal{F} \in \mathfrak{I}} \sum_{S:|S|=n} \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}}} \mathbf{U}_S\right] \right) \tag{3.11}$$

for all $\mathfrak{I} = \{\mathcal{U}\}$, $\{\mathcal{U}, \mathcal{F}_1\}$, $\{\mathcal{U}, \mathcal{F}_2\}$, $\{\mathcal{U}, \mathcal{F}_3\}$, $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_2\}$, $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_3\}$, $\{\mathcal{U}, \mathcal{F}_2, \mathcal{F}_3\}$, and $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$. We note that this is a scenario of orthogonal block structures but not simple block structures.

4. Examples: minimum aberration designs under three scenarios

In this section, we apply the aberration criteria developed in Section 3 under (3.7) to three block structures. For the situations where $\mathbf{v}$ does not satisfy (3.7), it is easy to derive appropriate aberration criteria based on the results in Section 3 (e.g., (3.9)).

4.1 Eighteen-run nonregular designs

Suppose there are 18 unstructured experimental units. We have the block structure $\{\mathcal{U}, \mathcal{E}\}$. Consider a three-level 18-run orthogonal array of strength two in columns 2 to 8 of Table 8C.2 of Wu and Hamada (2009), also given in Section S6 of the supplementary material. Many three-level 18-run nonregular designs with fewer factors can be obtained by column deletion from the array.

For $n = 3$, Wang and Wu (1995) showed that there are three nonisomorphic designs. Xu and Wu (2001) gave their generalized wordlength patterns, which are $(0, 0, 0.5)$, $(0, 0, 1)$, $(0, 0, 2)$. The first one has generalized minimum aberration and, by Theorem 4, minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\mathbf{v}$ satisfying (3.7). Moreover, since the sums of the generalized wordlengths of the three designs are $0.5, 1, 2$, respectively, it follows from $(\mathbf{1}_N^T \mathbf{U}_\phi)(\mathbf{1}_N^T \mathbf{U}_\phi)^T = \frac{N^2}{\Xi} = 12$ and Theorem 5 that $\frac{3^3}{18^2}\{(18 + 2m) - 12\} = l$ with $l = 0.5, 1, 2$ for the three designs. We have $m = 0, 3, 9$, respectively. Therefore the first design does not have replicates, while the other two designs separately have 3 and 9 replicates.

For $n = 4$, Xu and Wu (2001) gave the generalized wordlength patterns of the only four nonisomorphic designs, which are $(0, 0, 2, 1.5)$, $(0, 0, 2.5, 1)$, $(0, 0, 3.5, 0)$, $(0, 0, 3.5, 0)$. The first one has generalized minimum aberration and, by Theorem 4, minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\mathbf{v}$ satisfying (3.7). The sums of the generalized wordlengths are all equal to 3.5. By Theorem 5, we have $\frac{3^4}{18^2}\{(18 + 2m) - 4\} = 3.5$. Thus $m = 0$ and these four designs have no replicates.

4.2 Blocked mixed-level orthogonal arrays

Lin (2014) studied blocked mixed-level orthogonal arrays and listed several minimum aberration designs in terms of W_1 and W_2. We consider a scenario in their study: 18-run blocked orthogonal arrays with 3 blocks of size 6 and 4 treatment factors, consisting of 1 two-level factor and 3 three-level factors. Each blocked orthogonal array is constructed by selecting five columns in Table 8C.2 of Wu and Hamada (2009), also given in Section S6 of the supplementary material, where one is the two-level column, one is a three-level column for blocking, and the others are three-level columns. There are $7 \times C_3^6 = 140$ candidate designs.

A complete search shows that no design has minimum aberration with respect to both $\mathfrak{W}_0$ and $\mathfrak{W}_1$. The minimum aberration design with respect to $\mathfrak{W}_1$, denoted by d^*, is constructed by selecting the 8th column for blocking, the 1st column for the two-level treatment factor, the 2nd, 4th, 5th columns for the three-level treatment factors. It has $\mathfrak{W}_1 = (0, 0.125, 0.708, 1, 0.75, 0.042, 0)$. Figure 1 gives the ranking of all 140 candidate designs in terms of $\mathfrak{W}_1$, where each point represents a design and the x-axis is their rank values (average if tied, smaller the better). The black dot is d^*, with rank value 1. The red and blue dots represent those with minimum aberration in terms of W_1 and W_2, respectively. We can see that the three minimum aberration designs under the three different aberration criteria do not coincide. As what we suspect in Section 3.2, the one obtained using W_2 is closer to that using $\mathfrak{W}_1$. Besides, d^* has maximum D-efficiency under certain fixed-effect models. Refer to Section S7 of the supplementary material for details.

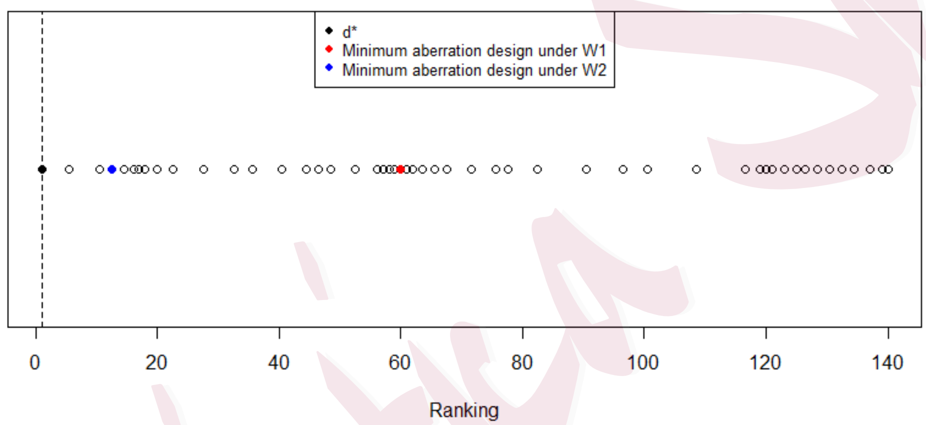

Figure 1: Comparison of $\mathfrak{W}_1$, W_1, and W_2

4.3 Three-stage manufacturing process

Butler (2004) mentioned a three-stage manufacturing process with a few treatment factors in each stage. Suppose there are 36 experimental units and each stage consists of 2 three-level treatment factors. The 36 units are divided into 6 groups of equal size in each stage. We have the block structure $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{E}\}$, where each $\mathcal{F}_i$ is a unit factor for one stage and partitions the 36 units into 6 classes. We also require $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ define an orthogonal array of strength two that can be represented by the following Latin square (Wu and Hamada, 2009, p. 151):

A	B	C	D	E	F
B	C	F	A	D	E
C	F	B	E	A	D
D	E	A	B	F	C
E	A	D	F	C	B
F	D	E	C	B	A

where each row, column, and letter represent a group of the first, second, and third stages, respectively. To reduce the computational burden, we assume that all the interactions of treatment factors across different stages are negligible.

A complete search shows that the design given in Table 1 has minimum aberration with respect to (3.11) for all $\mathfrak{I} = \{\mathcal{U}\}$, $\{\mathcal{U}, \mathcal{F}_1\}$, $\{\mathcal{U}, \mathcal{F}_2\}$, $\{\mathcal{U}, \mathcal{F}_3\}$, $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_2\}$, $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_3\}$, $\{\mathcal{U}, \mathcal{F}_2, \mathcal{F}_3\}$, and $\{\mathcal{U}, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$, with wordlength patterns $(0, 6)$, $(16, 20)$, $(16, 20)$, $(16, 20)$, $(32, 34)$, $(32, 34)$, $(32, 34)$, and $(48, 48)$. Thus it has minimum aberration with respect to $\mathfrak{W}$ for all feasible $\boldsymbol{\xi}$. The three stages share the same design settings, balance and without replicates.

Stage 1		Stage 2		Stage 3	
0	0	0	0	0	0
0	1	0	1	0	1
1	0	1	0	1	0
1	2	1	2	1	2
2	1	2	1	2	1
2	2	2	2	2	2

Table 1: Minimum aberration design: three-stage manufacturing process

5. Concluding remarks

In this paper, we develop a unified theory of aberration criteria through a Bayesian perspective. Our theory provides various applications to the cases of mixed-level/multi-group treatment factors, nonregular designs, as well as orthogonal block structures. Given design situations, experimenters can create suitable aberration criteria based on our theory. In addition, we provide a useful result to screen out inadmissible designs.

The block structures we consider in this paper require uniform unit factors. In real applications, however, this may not be feasible. For instance, this is impossible if the number of experimental units is not a multiple of the number of levels of some unit factor. Since this assumption is crucial to our theory, developing a more general theory is needed and will be considered in future work.

Supplementary Material

The supplementary material contains all proofs and additional explanation for this paper.

Acknowledgments

We would like to thank the referees for their helpful and constructive comments, which have led to an improved paper. This work was financially supported by the Ministry of Science and Technology (MOST) in Taiwan with grant number 107-2118-M-008-001-MY2 and the Young Scholar Fellowship Program (Einstein Program) under Grant 108-2636-M-008-004-.

References

Antolino, N., Hayes, G., Kirkpatrick, R., Muhlstein, C. L., Frecker, M. I., Mockensturm, E. M., and Adair, J. H. (2009a). Lost mold rapid infiltration forming of mesoscale ceramics: part 1, fabrication. *Journal of American Ceramic Society 92*, pp. S63–S69.

Antolino, N., Hayes, G., Kirkpatrick, R., Muhlstein, C. L., Frecker, M. I., Mockensturm, E. M., and Adair, J. H. (2009b). Lost mold rapid infiltration forming of mesoscale ceramics: part 2, geometry and strength improvements. *Journal of American Ceramic Society 92*, pp. S70–S78.

Bailey, R. A. (1985). Factorial design and Abelian groups. *Linear Algebra Appl. 70*, pp. 349–368.

Bailey, R. A. (2008). *Design of Comparative Experiments.* Cambridge University Press, Cambridge, U.K.

Bingham, D. R. and Sitter, R. R. (1999). Minimum-aberration two-level fractional factorial split-plot designs. *Technometrics 41*, pp. 62–70.

Bingham, D. R., Schoen, E. D., and Sitter, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors. *Appl. Statist. 53*, pp. 325–339.

Bingham, D. R., Sitter, R. R., Kelly, E., Moore, L., and Olivas, J. D. (2008). Factorial designs with multiple levels of randomization. *Statist. Sinica. 18*, pp. 493–513.

Butler, N. A. (2004). Construction of two-level split-lot fractional factorial designs for multistage processes. *Technometrics 46*, pp. 445–451.

Chang, M. C. and Cheng, C. S. (2018). A Bayesian approach to the selection of two-level multi-stratum factorial designs. *Ann. Statist 46*, pp. 1779–1806.

Chen, H. and Cheng, C. S. (1999). Theory of optimal blocking of 2^{n-m} designs. *Ann. Statist 27*, pp. 1948–1973.

Cheng, C. S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments.* CRC Press.

Cheng, C. S. and Tang, B. (2005). A general theory of minimum aberration and its applications.

Ann. Statist 33, pp. 944–958.

Cheng, C. S. and Tsai, P. W. (2011). Multistratum fractional factorial designs. *Statist. Sinica 21*, pp. 1001–1021.

Cheng, S. W., Li, W., and Ye, K. Q. (2004). Blocked nonregular two-level factorial designs. *Technometrics 46*, pp. 269–279.

Cheng, S. W. and Wu, C. F. J. (2002). Choice of optimal blocking schemes in two-level and three-level designs. *Technometrics 44*, pp. 269–277.

Dean, A., Morris, M., Stufken, J., and Bingham, D. (2015). *Handbook of Design and Analysis of Experiments*. CRC Press.

Dean, A. M. and Lewis, S. (2006). *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*. Springer, New York.

Eccleston, J. A. and Hedayat, A. (1974). On the theory of connected designs: Characterization and optimality. *Ann. Statist 2*, pp. 1238–1255.

Fries, A. and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics 22*, pp. 601–608.

Huang, P., Chen, D., and Voelkel, J. O. (1998). Minimum aberration two-level split-plot designs. *Technometrics 40*, pp. 314–326.

Joseph, V. R. (2006). A Bayesian approach to the design and analysis of fractionated experiments. *Technometrics 48*, pp. 219–229.

Joseph, V. R. and Delaney, J. D. (2007). Functionally induced priors for the analysis of experiments. *Technometrics 49*, pp. 1–11.

Kerr, M. K. (2001). Bayesian optimal fractional factorials. *Statist. Sinica 11*, pp. 605–630.

Kurkjian, B. and Zelen, M. (1962). A calculus for factorial arrangements. *Ann. Math. Statist. 33*, pp. 600–619.

Li, W., Mee, R. W., and Zhou, Q. (2018). Using individual factor information in fractional factorial designs. *Technometrics 61*, pp. 38–49.

Li, W., Zhou, Q., and Zhang, R. C. (2015). Effective designs based on individual word length patterns. *J. Statist. Plann. Inference 163*, pp. 43–47.

Lin, C. Y. (2014). Optimal blocked orthogonal arrays. *J. Statist. Plann. Inference 145*, pp. 139–147.

Mee, R. W. and Bates, R. L. (1998). Split-lot designs: experiments for multistage batch processes. *Technometrics 40*, pp. 127–140.

Mitchell, T. J., Morris, M. D., and Ylvisaker, D. (1995). Two-level fractional factorials and Bayesian prediction. *Statist. Sinica 5*, pp. 559–573.

Nelder, J. A. (1965a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. R. Soc. Lond. Ser. A 283*, pp. 147–162.

Nelder, J. A. (1965b). The analysis of randomized experiments with orthogonal block structure.

II. Treatment structure and the general analysis of variance. *Proc. R. Soc. Lond. Ser. A 283*, pp. 163–178.

Ranjan, P., Bingham, D. R., and Dean, A. M. (2009). Existence and construction of randomization defining contrast subspaces for regular factorial designs. *Ann. Statist. 37*, pp. 3580–3599.

Sadeghi, S., Qian, P. Z. G., and Arora, N. (2016). Sliced designs for multi-platform online experiments. *manuscript*.

Sadeghi, S., Qian, P. Z. G., and Arora, N. (2017). Sliced minimum aberration designs for four-platform experiments. *manuscript*.

Sitter, R. R., Chen, H., and Feder, M. (1997). Fractional resolution and minimum aberration in blocked 2^{n-k} designs. *Technometrics 39*, pp. 382–390.

Speed, T. P. and Bailey, R. A. (1982). On a class of association schemes derived from lattices of equivalence relations. *Algebraic Structures and Applications*, pp. 55–74.

Taguchi, G. (1987). *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*. UNIPUB/Kraus International Publications, White Plains, New York and Dearborn, Michigan.

Tichon, J. G., Li, W., and Mcleod, R. G. (2012). Generalized minimum aberration two-level split-plot designs. *J. Statist. Plann. Inference 142*, pp. 1407–1414.

Wang, J. C. and Wu, C. F. J. (1995). A hidden projection property of Plackett-Burman and related designs. *Statist. Sinica 5*, pp. 235–250.

Wu, C. F. J. and Hamada, M. (2009). *Experiments: Planning, Analysis, and Optimization* (2 ed.). Wiley, New York.

Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for assymmetrical fractional factorial designs. *Ann. Statist 29*, pp. 1066–1077.

Yang, P. and Lin, C. Y. (2017). Optimal split-plot orthogonal arrays. *Aust. N. Z. J. Stat. 59*, pp. 81–94.

Yates, F. (1935). Complex experiments. *Supplement to the J. R. Stat. Soc. Ser. 2*, pp. 181–274.

Yuangyai, C. and Lin, D. K. J. (2013). Understanding multistage experiments. *Int. J. Experimental Design and Process Optimisation 3*, pp. 384–409.

Zhu, Y. (2003). Structure function for aliasing patterns in 2^{n-l} design with multiple groups of factors. *Ann. Statist 31*, pp. 995–1011.

Graduate Institute of Statistics, National Central University

E-mail: mcchang0131@gmail.com

Phone: 886-3-422-7151 ext. 65451