A UNIFIED FRAMEWORK FOR MINIMUM ABERRATION

Ming-Chung Chang

National Central University

Supplementary Material

This supplementary material contains all proofs and additional explanation for the main article. Section S1 introduces the five conditions in Cheng (2014, p. 233) and their importance. Section S2 provides some results for the Bayesian prior. Section S3 contains all lemmas and their proofs. Section S4 presents the proofs of all theorems. Section S5 discusses an application to multi-platform experiments. Section S6 gives the 18-run mixed-level orthogonal array used in the main article. Section S7 studies the relationship between $\mathfrak{W}_1$ and D-efficiency under block designs.

S1. Five conditions in Cheng (2014, p. 233)

Before listing the conditions (i), (ii), (iii), (v), (vi) of Definition 12.4 in Cheng (2014, p. 233), we introduce the notion of *supremum* and *orthogonal*.

Given two unit factors $\mathcal{F}_1$ and $\mathcal{F}_2$, the supremum of $\mathcal{F}_1$ and $\mathcal{F}_2$, denoted by $\mathcal{F}_1 \vee \mathcal{F}_2$, is the unit factor such that (i) $\mathcal{F}_1, \mathcal{F}_2 \preceq \mathcal{F}_1 \vee \mathcal{F}_2$, and (ii) $\mathcal{F}_1 \vee \mathcal{F}_2 \preceq \mathcal{G}$ for all $\mathcal{G}$ such that $\mathcal{F}_1, \mathcal{F}_2 \preceq \mathcal{G}$. We say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are orthogonal if $\mathcal{F}_1$ and $\mathcal{F}_2$ have proportional frequencies in each $(\mathcal{F}_1 \vee \mathcal{F}_2)$-class, i.e., for each $(\mathcal{F}_1 \vee \mathcal{F}_2)$-class Γ, if both the ith $\mathcal{F}_1$-class and the jth $\mathcal{F}_2$-class are contained in Γ, then

$$n_{ij} = \frac{n_{i+}n_{+j}}{|\Gamma|},$$

where n_{i+}, n_{+j}, and n_{ij} are the numbers of units in, respectively, the ith $\mathcal{F}_1$-class, the jth $\mathcal{F}_2$-class, and the intersection of these two classes, and $|\Gamma|$ is the number of units in Γ.

Throughout the main article, we consider block structures $\mathfrak{B}$ that satisfy the following conditions:

$$\text{all the factors in } \mathfrak{B} \text{ are uniform,} \tag{S1.1}$$

$$\mathcal{E} \in \mathfrak{B}, \tag{S1.2}$$

$$\mathcal{U} \in \mathfrak{B}, \tag{S1.3}$$

$$\mathcal{F}, \mathcal{G} \in \mathfrak{B} \Rightarrow \mathcal{F} \vee \mathcal{G} \in \mathfrak{B}, \tag{S1.4}$$

$$\text{all the factors in } \mathfrak{B} \text{ are pairwise orthogonal,} \tag{S1.5}$$

corresponding to the conditions (i), (ii), (iii), (v), (vi) of Definition 12.4 in Cheng (2014, p. 233). We note that most of the block structures encountered in practice satisfy (S1.1)-(S1.5).

The conditions (S1.1)-(S1.5) are essential to all theoretical results in the main article. If a block structure satisfies the five conditions, then, as mentioned in Section 2.3 of the main article, the covariance matrix of the responses has a spectral decomposition characterized by the unit factors in the block structure. An important feature of this spectral decomposition is that all eigenspaces are irrelevant to unknown parameters in the model. Therefore, the aberration criterion proposed in Section 3 of the main article is well defined.

S2. Some results on Bayesian prior

Under certain correlation functions of the Gaussian process, $\boldsymbol{\beta}$ has a zero-mean multivariate normal distribution with

$$\mathrm{var}(\beta_j) = \tau^2 \left\{ \prod_{i:1 \le i \le n, \delta_{ji}=1} r_i \right\}, \tag{S2.6}$$

$$\mathrm{cov}(\beta_j, \beta_{j'}) = 0 \text{ if } j \neq j',$$

where $\tau^2 > 0$, $0 < r_1, ..., r_n < 1$, and $\delta_{ji} = 1$ if β_j involves the ith treatment factor and zero otherwise.

It can be seen from (S2.6) that $\mathrm{var}(\beta_j) \ge \mathrm{var}(\beta_{j'})$ if the treatment factors involved in β_j is a subset of those involved in $\beta_{j'}$; that is, (S2.6) satisfies the property in (2.2). Moreover, for a β_j involving the ith treatment factor, $\mathrm{var}(\beta_j)$ increases as r_i increases. Thus, each r_i can be interpreted as a parameter that controls the importance of the ith treatment factor. By letting $r_1 = \cdots = r_n$, we have that $\mathrm{var}(\beta_j) \ge \mathrm{var}(\beta_{j'})$ if the number of treatment factors involved in β_j is less than or equal to that for $\beta_{j'}$, which is consistent with the *effect hierarchy principle* (Wu and Hamada, 2009, p. 172).

S3. Lemmas

Lemma 1. *If the full model matrix* $\mathbf{P}$ *is constructed through (2.1), then for a given* $S \subseteq \{1, ..., n\}$, *tr* $\left[\mathbf{U}_S^T \mathbf{U}_S\right]$ *is a constant for any choice of* N*-run designs.*

Proof. Let $\mathbf{P}_S$ be formed by the columns in $\mathbf{P}$ associated with an $S \subseteq \{1, ..., n\}$. We note that $\mathrm{tr}\left[\mathbf{U}_S^T \mathbf{U}_S\right] = \mathrm{tr}\left[\mathbf{U}_S \mathbf{U}_S^T\right]$. Thus, it suffices to check whether $\mathbf{e}_i^T \mathbf{P}_S \mathbf{P}_S^T \mathbf{e}_i$ is a constant for all $i = 1, ..., \Xi$, where $\mathbf{e}_i$ is the ith column of $\mathbf{I}_\Xi$. We prove this by mathematical induction on $|S|$, the cardinality of S.

It is clearly true when $|S| = 0$; that is, $S = \phi$. Let $|S| = 1$. Without loss of generality, we set $S = \{1\}$. By (2.1), the $\Xi \times p_1$ matrix $\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_S\right]$ can be constructed by

$$\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_S\right] = \mathbf{P}_1 \otimes \frac{1}{\sqrt{p_2}}\mathbf{1}_{p_2} \otimes \cdots \otimes \frac{1}{\sqrt{p_n}}\mathbf{1}_{p_n}.$$

We have

$$\begin{aligned}
\mathbf{e}_i^T \left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_S\right] \left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_S\right]^T \mathbf{e}_i &= \mathbf{e}_i^T \left(\mathbf{P}_1 \otimes \left\{\otimes_{j=2}^n \frac{1}{\sqrt{p_j}}\mathbf{1}_{p_j}\right\}\right) \left(\mathbf{P}_1 \otimes \left\{\otimes_{j=2}^n \frac{1}{\sqrt{p_j}}\mathbf{1}_{p_j}\right\}\right)^T \mathbf{e}_i \\
&= \mathbf{e}_i^T \left(\mathbf{P}_1 \mathbf{P}_1^T\right) \otimes \left(\frac{1}{p_2}\mathbf{1}_{p_2}\mathbf{1}_{p_2}^T\right) \otimes \cdots \otimes \left(\frac{1}{p_n}\mathbf{1}_{p_n}\mathbf{1}_{p_n}^T\right) \mathbf{e}_i \\
&= \frac{1}{p_2 \cdots p_n} \mathbf{e}_i^T \left\{\mathbf{I}_{p_1} \otimes \mathbf{J}_{p_2 \cdots p_n}\right\} \mathbf{e}_i \\
&= \frac{1}{p_2 \cdots p_n}
\end{aligned}$$

for any $i = 1, ..., \Xi$, where $\mathbf{J}_m$ denotes the $m \times m$ matrix of ones. Since $\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_S\right]\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_S\right]^T = \frac{1}{\Xi}\mathbf{J}_\Xi + \mathbf{P}_S\mathbf{P}_S^T$ and $\mathbf{e}_i^T \frac{1}{\Xi}\mathbf{J}_\Xi \mathbf{e}_i = \frac{1}{\Xi}$, it follows that $\mathbf{e}_i^T\mathbf{P}_S\mathbf{P}_S^T\mathbf{e}_i = \frac{1}{p_2\cdots p_n} - \frac{1}{\Xi}$, which is a constant for all i.

Now suppose it holds for all S with $|S| < k$, where $k \geq 2$. We show that if $|S| = k$, then $\mathbf{e}_i^T\mathbf{P}_S\mathbf{P}_S^T\mathbf{e}_i$ is a constant for all $i = 1, ..., \Xi$. Without loss of generality, we set $S = \{1, ..., k\}$. It follows from a similar argument that the $\Xi \times (p_1 \cdots p_k)$ matrix

$$\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_{\{1\}}, ..., \mathbf{P}_{\{1,...,k\}}\right]$$

can be constructed by

$$\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_{\{1\}}, ..., \mathbf{P}_{\{1,...,k\}}\right] = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_k \otimes \frac{1}{\sqrt{p_{k+1}}}\mathbf{1}_{p_{k+1}} \otimes \cdots \otimes \frac{1}{\sqrt{p_n}}\mathbf{1}_{p_n}.$$

Therefore, we have

$$\begin{aligned}
&\mathbf{e}_i^T\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_{\{1\}}, ..., \mathbf{P}_{\{1,...,k\}}\right]\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_{\{1\}}, ..., \mathbf{P}_{\{1,...,k\}}\right]^T\mathbf{e}_i \\
&= \frac{1}{p_{k+1}\cdots p_n}\mathbf{e}_i^T\left\{\mathbf{I}_{p_1\cdots p_k} \otimes \mathbf{J}_{p_{k+1}\cdots p_n}\right\}\mathbf{e}_i \\
&= \frac{1}{p_{k+1}\cdots p_n}.
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbf{e}_i^T\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_{\{1\}}, ..., \mathbf{P}_{\{1,...,k\}}\right]\left[\frac{1}{\sqrt{\Xi}}\mathbf{1}_\Xi, \mathbf{P}_{\{1\}}, ..., \mathbf{P}_{\{1,...,k\}}\right]^T\mathbf{e}_i \\
&= \frac{1}{\Xi}\mathbf{e}_i^T\mathbf{J}_\Xi\mathbf{e}_i + \left(\sum_{S'\subseteq\{1,...,k\}:1\leq|S'|\leq k-1}\mathbf{e}_i^T\mathbf{P}_{S'}\mathbf{P}_{S'}^T\mathbf{e}_i\right) + \mathbf{e}_i^T\mathbf{P}_S\mathbf{P}_S^T\mathbf{e}_i,
\end{aligned}$$

it follows that $\mathbf{e}_i^T \mathbf{P}_S \mathbf{P}_S^T \mathbf{e}_i$ is a constant for all $i = 1, ..., \Xi$ by the induction hypothesis. □

Lemma 2. *Given a design and an $S \subseteq \{1, ..., n\}$, tr $\left[\mathbf{U}_S^T \mathbf{U}_S\right]$ is a constant for any choice of orthogonal-column-bases in $\mathbf{P}$.*

Proof. Let $\mathbf{P}$ and $\mathbf{P}^*$ be two full model matrices with mutually orthonormal columns. Let $\mathbf{U}$ and $\mathbf{U}^*$ be the resulting model matrices under the given design. We denote $\mathbf{P}_S$, $\mathbf{P}_S^*$, $\mathbf{U}_S$, and $\mathbf{U}_S^*$ the counterparts associated with S. Then, there exists an orthogonal matrix $\mathbf{H}$ such that $\mathbf{P}_S = \mathbf{P}_S^* \mathbf{H}$, yielding $\mathbf{U}_S = \mathbf{U}_S^* \mathbf{H}$. It follows that $\mathbf{U}_S \mathbf{U}_S^T = (\mathbf{U}_S^* \mathbf{H})(\mathbf{U}_S^* \mathbf{H})^T = \mathbf{U}_S^* \mathbf{U}_S^{*T}$, which proves the result. □

S4. Proofs of all theorems

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 $\Phi_1(d; \boldsymbol{\xi}, \mathbf{v})$.*

Proof. We note that another expression of $\text{cov}(\boldsymbol{\beta}|\mathbf{y})$ is $\left(\mathbf{U}^T \mathbf{V}^{-1} \mathbf{U} + \boldsymbol{\Sigma}_\beta^{-1}\right)^{-1}$. The Bayesian D-optimality is to maximize

$$\det[\mathbf{M}] = \det[\boldsymbol{\Sigma}_\beta^{-1}] \det[\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{V}^{-1} \mathbf{U} + \mathbf{I}_\Xi].$$

Since $\det[\mathbf{\Sigma}_\beta^{-1}]$ is a constant irrelevant to designs, based on the (M.S)-optimality, one needs to maximize $\mathrm{tr}[\mathbf{\Sigma}_\beta \mathbf{U}^T\mathbf{V}^{-1}\mathbf{U} + \mathbf{I}_\Xi]$ and then minimize $\mathrm{tr}[(\mathbf{\Sigma}_\beta \mathbf{U}^T\mathbf{V}^{-1}\mathbf{U} + \mathbf{I}_\Xi)^2]$ among the designs that maximize $\mathrm{tr}[\mathbf{\Sigma}_\beta \mathbf{U}^T\mathbf{V}^{-1}\mathbf{U} + \mathbf{I}_\Xi]$. Then, the result is proved by the use of a similar derivation as in Chang and Cheng (2018). □

Theorem 2. *For an* $S \subseteq \{1, ..., n\}$, *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* $\mathbf{P}$.

Proof. This result is proved by Lemmas 1 and 2. □

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*

$$S \in \mathfrak{S}, S' \in 2^{\{1,\ldots,n\}} \setminus \{\phi\},\ \textit{and } S' \subset S \Rightarrow S' \in \mathfrak{S}, \tag{S4.7}$$

$$\mathcal{F} \in \mathfrak{G}, \mathcal{F}' \in \mathfrak{B},\ \textit{and } \mathcal{F} \prec \mathcal{F}' \Rightarrow \mathcal{F}' \in \mathfrak{G}. \tag{S4.8}$$

Proof. We first prove the necessity part. Suppose a design d minimizes

$\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$. For any $\mathfrak{S}$ satisfying (S4.7), let

$$v_S = v_{S'} > 0 \text{ for any } S, S' \in \mathfrak{S}, \text{ and } v_{S''} = 0 \text{ for all } S'' \notin \mathfrak{S}. \qquad \text{(S4.9)}$$

Similarly, for any $\mathfrak{G}$ satisfying (S4.8), let $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_0}$ for any $\mathcal{F} \in \mathfrak{G}$, and $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_m}$ for any $\mathcal{F} \notin \mathfrak{G}$. The $\mathbf{v}$ and $\boldsymbol{\xi}$ in this setting are obviously feasible. Under this setting, minimizing $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ is reduced to minimizing $\sum_{S\in\mathfrak{S}}\sum_{i:\mathcal{F}_i\in\mathfrak{G}} \mathrm{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right]$. Since d minimizes $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$, it must minimize $\sum_{S\in\mathfrak{S}}\sum_{i:\mathcal{F}_i\in\mathfrak{G}} \mathrm{tr}\left[\mathbf{U}_S^T\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{U}_S\right]$.

In the remaining part of the proof, we prove the sufficiency part. Let $\boldsymbol{\xi}$ be fixed and

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

$S \subseteq \{1,...,n\}$. Then, we have $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v}) = \sum_{S\subseteq\{1,...,n\}} v_S C_S$. We show that a design that minimizes $\sum_{S\in\mathfrak{S}} C_S$ for all nonempty subsets $\mathfrak{S}$ of $\mathfrak{A}' \subseteq 2^{\{1,...,n\}}$ satisfying (S4.7), with $2^{\{1,...,n\}}$ replaced with $\mathfrak{A}'$, also minimizes $\sum_{S\subseteq\{1,...,n\}} v_S C_S$ for all feasible $\mathbf{v}$.

Similar to the proof of Theorem 5.1 in Chang and Cheng (2018), we apply mathematical induction on the number of elements in $\mathfrak{A}'$. It is clearly true when $\mathfrak{A}'$ consists of one element. Now suppose that it holds for all subsets of $2^{\{1,...,n\}}$ with fewer than s elements, $s \geq 2$; we need to show that if $|\mathfrak{A}'| = s$ and a design d^* minimizes $\sum_{S\in\mathfrak{S}} C_S$ for all subsets $\mathfrak{S}$

of $\mathfrak{A}'$ satisfying (S4.7) with $2^{\{1,\ldots,n\}}$ replaced with $\mathfrak{A}'$, then d^* minimizes $\sum_{S\subseteq\{1,\ldots,n\}} v_S C_S$ for all feasible $\mathbf{v}$. Under the given assumption, by taking $\mathfrak{S} = \mathfrak{A}'$, we have that

$$d^* \text{ minimizes } \sum_{S\in\mathfrak{B}'} C_S. \tag{S4.10}$$

If all the v_S's for which $S \in \mathfrak{A}'$ are equal, say they are all equal to v, then, by (S4.10), d^* minimizes $v \sum_{S\in\mathfrak{A}'} C_S = \sum_{S\in\mathfrak{A}'} v_S C_S$. On the other hand, suppose not all the v_S's for which $S \in \mathfrak{A}'$ are equal. Let v be the smallest value of such v_S's and let $\mathfrak{A}^* = \{S \in \mathfrak{A}' : v_S > v\}$. Then, $\mathfrak{A}^*$ is nonempty and $|\mathfrak{A}^*| < s$. For each $S \in \mathfrak{A}^*$, we have $v_S - v > 0$. Moreover, $v_S - v \geq v_{S'} - v$ if $S \subset S'$ and $S, S' \in \mathfrak{A}^*$. Furthermore,

$$\sum_{S\in\mathfrak{A}'} v_S C_S = \sum_{S\in\mathfrak{A}^*} (v_S - v)\, C_S + v \sum_{S\in\mathfrak{A}'} C_S.$$

By (S4.10), it suffices to show that d^* minimizes $\sum_{S\in\mathfrak{A}^*} (v_S - v)\, C_S$. Since $|\mathfrak{A}^*| < s$, by the induction hypothesis, it remains to show that d^* minimizes $\sum_{S\in\mathfrak{S}} C_S$ for all nonempty subsets $\mathfrak{S}$ of $\mathfrak{A}^*$ satisfying the following condition:

$$S \in \mathfrak{S}, S' \in \mathfrak{A}^* \text{ and } S' \subset S \Rightarrow S' \in \mathfrak{S}. \tag{S4.11}$$

Suppose a subset $\mathfrak{S}$ of $\mathfrak{A}^*$ satisfies (S4.11). By the assumption on $\mathfrak{A}'$, d^* minimizes $\sum_{S\in\mathfrak{S}} C_S$ provided that $\mathfrak{S}$ also satisfies (S4.7) with $2^{\{1,\ldots,n\}}$

replaced with $\mathfrak{A}'$. That is, given $S \in \mathfrak{S}$, $S' \in \mathfrak{A}'$, and $S' \subset S$, we want to show $S' \in \mathfrak{S}$. Because $S \in \mathfrak{S} \subseteq \mathfrak{A}^*$, by the definition of $\mathfrak{A}^*$, we have $v_S > v$. Moreover, $v_{S'} \geq v_S$ since $S' \subset S$. Thus we have $v_{S'} \geq v_S > v$, which leads to $S' \in \mathfrak{A}^*$. Then by (S4.11), $S' \in \mathfrak{S}$. Therefore, given $\boldsymbol{\xi}$, we have proved that a design that minimizes $\sum_{S \in \mathfrak{S}} C_S$ for all nonempty subsets $\mathfrak{S}$ of $\mathfrak{A}' \subseteq 2^{\{1,\ldots,n\}}$ satisfying (S4.7) also minimizes $\sum_{S \subseteq \{1,\ldots,n\}} v_S C_S$ for all feasible $\mathbf{v}$.

We now let $\mathbf{v}$ be fixed and

$$B_i = \sum_{S \subseteq \{1,\ldots,n\}} v_S \mathrm{tr}\left[\mathbf{U}_S^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}_S\right],$$

$i = 0, ..., m-1$. Then, we have $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}}\right) B_i$. Based on a similar argument, we can show that a design that minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_i$ for all nonempty subsets $\mathfrak{G}$ of $\mathfrak{A}' \subseteq \mathfrak{B}$ satisfying (S4.8) also minimizes $\sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}}\right) B_i$ for all feasible $\boldsymbol{\xi}$.

It turns out that if a design has minimized $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\boldsymbol{\xi}$ with $\mathbf{v}$ satisfying (S4.9) under each $\mathfrak{S}$ that satisfies (S4.7), then for each feasible $\boldsymbol{\xi}$, this design minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\mathbf{v}$; that is, it simultaneously minimizes $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\mathbf{v}$ and $\boldsymbol{\xi}$. Thus, to prove the sufficiency part, it suffices to check if a design minimizes

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

for all nonempty subsets $\mathfrak{S}$ satisfying (S4.7) and $\mathfrak{G}$ satisfying (S4.8).

□

Theorem 4. *Suppose $\mathfrak{B}$ is a block structure satisfying conditions (S1.1)-(S1.5). Then, under (3.10), a necessary and sufficient condition for a design to minimize $\Phi_1^*(d;\boldsymbol{\xi},\mathbf{v})$ for all $\mathbf{v}$ that satisfy (3.10) and feasible $\boldsymbol{\xi}$ is that it minimizes*

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

for all nonempty subsets $\mathfrak{G}\subseteq\mathfrak{B}\setminus\{\mathcal{F}_m\}$ satisfying (S4.8) and $\mathfrak{S}\subseteq 2^{\{1,\ldots,n\}}\setminus\{\phi\}$ satisfying

$$S\in\mathfrak{S}, S'\in 2^{\{1,\ldots,n\}}\setminus\{\phi\},\ \text{and } v_{S'}\geq v_S \Rightarrow S'\in\mathfrak{S}. \tag{S4.12}$$

Proof. The proof is done by replacing "$S'\subset S$" with $v_{S'}\geq v_S$ in the proof of Theorem 3. □

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

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

Proof. We have

$$\sum_{k=0}^{n} \sum_{S:|S|=k} (\mathbf{1}_N^T \mathbf{U}_S)(\mathbf{1}_N^T \mathbf{U}_S)^T = \sum_{S \subseteq \{1,\ldots,n\}} (\mathbf{1}_N^T \mathbf{U}_S)(\mathbf{1}_N^T \mathbf{U}_S)^T$$
$$= \mathrm{tr}\left[(\mathbf{1}_N^T \mathbf{U})(\mathbf{1}_N^T \mathbf{U})^T\right]$$
$$= \mathrm{tr}\left[(\mathbf{1}_N \mathbf{1}_N^T)(\mathbf{U}\mathbf{U}^T)\right]$$
$$= \sum_{1 \le i,j \le N} \mathbf{u}_i^T \mathbf{u}_j,$$

where $\mathbf{u}_l^T$ is the lth row of $\mathbf{U}$. Since the full model matrix $\mathbf{P}$ satisfies $\mathbf{P}\mathbf{P}^T = \mathbf{I}_\Xi$, we have $\mathbf{u}_i^T \mathbf{u}_j = 1$ if $i = j$ and zero otherwise. It follows that $\sum_{1 \le i,j \le N} \mathbf{u}_i^T \mathbf{u}_j = N + 2m$ for an N-run design with m replicates. □

S5. Application to multi-platfrom experiments

Sadeghi, Qian, and Arora (2016) and Sadeghi, Qian, and Arora (2017) discussed design selection for multi-platform experiments with unstructured units, in which the sliced factor was deemed much more important than the other treatment factors. They defined a *sliced effect hierarchy principle*; based on which a *sliced aberration criterion* was proposed.

Without loss of generality, suppose the first treatment factor is the sliced factor. In our approach, we can adopt the prior in (2.3), with $r_1 > r_2 = r_3 = \cdots = r_n$, to fit this scenario. The resulting order of $\mathrm{var}(\beta_j)$'s follows two rules: (a) $\mathrm{var}(\beta_j) > \mathrm{var}(\beta_{j'})$ if β_j involves fewer factors than $\beta_{j'}$, and (b) for the β_j and $\beta_{j'}$ involving the same number of factors, $\mathrm{var}(\beta_j) > \mathrm{var}(\beta_{j'})$ if β_j involves the sliced factor but $\beta_{j'}$ does not. This order is consistent with their sliced effect hierarchy principle. If we require $r_1 \approx 1$ and $r_2 = r_3 = \cdots = r_n \approx 0$, then minimizing $\Phi_1^*(d; \boldsymbol{\xi}, \mathbf{v})$ would be nearly equivalent to optimizing the sliced aberration criterion.

In general, if the treatment factors are divided into g groups, where those in the same group are of equal importance, one can assign

$$\mathrm{var}(\beta_j) = \tau^2 r_1^{t_1} \cdots r_g^{t_g}, \tag{S5.13}$$

where t_l is the number of treatment factors in the lth group involved in β_j. When $g = 2$, $t_1 \in \{0, 1\}$, and $t_2 \in \{0, 1, ..., n-1\}$, (S5.13) corresponds to the setting in Sadeghi, Qian, and Arora (2016, 2017). Another scenario of multi-group treatment factors is in Tichon, Li, and Mcleod (2012), in which we have $g = 2$, $t_1 =$ "the number of whole-plot treatment factors involved in β_j", and $t_2 =$ "the number of subplot treatment factors involved in β_j".

S6. 18-run orthogonal array

The following table is the 18-run mixed-level orthogonal array used in Sections 4.1 and 4.2. It comes from Table 8C.2 of Wu and Hamada (2009) except that the 7th column is after the permutation: $(0, 1, 2) \to (2, 0, 1)$.

18-run orthogonal array							
0	0	0	0	0	0	2	0
0	0	1	1	1	1	0	1
0	0	2	2	2	2	1	2
0	1	0	0	1	1	1	2
0	1	1	1	2	2	2	0
0	1	2	2	0	0	0	1
0	2	0	1	0	2	0	2
0	2	1	2	1	0	1	0
0	2	2	0	2	1	2	1
1	0	0	2	2	1	0	0
1	0	1	0	0	2	1	1
1	0	2	1	1	0	2	2
1	1	0	1	2	0	1	1
1	1	1	2	0	1	2	2
1	1	2	0	1	2	0	0
1	2	0	2	1	2	2	1
1	2	1	0	2	0	0	2
1	2	2	1	0	1	1	0

S7. $\mathfrak{W}_1$ and D-efficiency under block designs

Suppose $\mathfrak{B} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}$, where $\mathcal{F}_1$ partitions N experimental units into blocks. Consider a statistical model similar to that in Section 2.3:

$$\mathbf{y} = \mathbf{U}\boldsymbol{\beta} + \mathbf{X}_{\mathcal{F}_0}\boldsymbol{\gamma}^{\mathcal{F}_0} + \mathbf{X}_{\mathcal{F}_1}\boldsymbol{\gamma}^{\mathcal{F}_1} + \mathbf{X}_{\mathcal{F}_2}\boldsymbol{\gamma}^{\mathcal{F}_2}.$$

Here we assume $\sigma^2_{\mathcal{F}_0} = 0$ (i.e., no random intercept) and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}^{\mathcal{F}_1}$ are vectors of unknown constants. Since $\mathbf{X}_{\mathcal{F}_2} = \mathbf{I}_N$, by replacing the notation $\boldsymbol{\gamma}^{\mathcal{F}_2}$ with $\boldsymbol{\epsilon}$, we have

$$\mathbf{y} = \mathbf{U}\boldsymbol{\beta} + \mathbf{X}_{\mathcal{F}_1}\boldsymbol{\gamma}^{\mathcal{F}_1} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\gamma}^{\mathcal{F}_1}$ represents fixed block effects and $\boldsymbol{\epsilon}$ is a vector of uncorrelated homoskedastic random errors. To estimate factorial effects $\boldsymbol{\beta}$, we eliminate block effects by projecting $\mathbf{y}$ onto $W_{\mathcal{F}_2}$ and obtain

$$\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{y} = \mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}\boldsymbol{\beta} + \mathbf{P}_{W_{\mathcal{F}_2}}\boldsymbol{\epsilon}$$

because $\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{X}_{\mathcal{F}_1}$ is a zero matrix. Then, the information matrix of $\boldsymbol{\beta}$ is given by $\mathbf{U}^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}$ (Dean et al., 2015, p. 80).

Assume that β_j's can be divided into K groups $\mathfrak{I}_1, ..., \mathfrak{I}_K$ such that the β_j's belonging to $\mathfrak{I}_l$ are more likely to be important than those belonging to $\mathfrak{I}_{l'}$ for $l < l'$. Let $\mathbf{U} = [\mathbf{U}_1, ..., \mathbf{U}_K]$ according to $\mathfrak{I}_1, ..., \mathfrak{I}_K$. Then, we can sequentially maximize the information of the β_j's belonging to $\mathfrak{I}_l$ from $l = 1$ to $l = K$ through D-efficiencies. This leads to sequentially maximizing

$$\left(\det\left[\mathbf{U}_1^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_1\right], \det\left[\mathbf{U}_2^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_2\right], \ldots, \det\left[\mathbf{U}_K^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_K\right]\right).$$

Based on the (M.S)-optimality, a one-step surrogate of which is to sequentially maximize

$$\left(\mathrm{tr}\left[\mathbf{U}_1^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_1\right], \mathrm{tr}\left[\mathbf{U}_2^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_2\right], \ldots, \mathrm{tr}\left[\mathbf{U}_K^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_K\right]\right).$$

Since $\mathbf{P}_{W_{\mathcal{F}_2}} = \mathbf{I}_N - (\mathbf{P}_{W_{\mathcal{F}_0}} + \mathbf{P}_{W_{\mathcal{F}_1}})$, it yields sequentially minimizing $\mathfrak{W}_1$ given the condition that $\mathrm{tr}\left[\mathbf{U}_l^T\mathbf{U}_l\right]$ is a constant for all l.

In Section 4.2, let $\mathbf{P}$ in (2.1) be constructed through orthogonal polynomial contrasts and $\mathfrak{I}_l$ be the set of the β_j's associated with the orthogonal polynomial contrasts of degree l. Then $\mathbf{U}_1$ consists of four columns, representing all four linear main effects; $\mathbf{U}_2$ has nine columns, representing six linear-by-linear interactions and three quadratic main effects. We find that the 140 candidate blocked mixed-level orthogonal arrays have the same $\det\left[\mathbf{U}_1^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_1\right]^{\frac{1}{4}} = 0.333$, while d^* has the largest $\det\left[\mathbf{U}_2^T\mathbf{P}_{W_{\mathcal{F}_2}}\mathbf{U}_2\right]^{\frac{1}{9}} = 0.294$.

References

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.

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

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

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*.

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.

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

Graduate Institute of Statistics, National Central University

E-mail: mcchang0131@gmail.com

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