Statistica Sinica

Relationship between orthogonal and baseline parameterizations and its applications to design constructions

Cheng-Yu Sun and Boxin Tang

Simon Fraser University, Canada

Abstract: When studying two-level factorial designs, factorial effects are usually defined as a set of orthogonal treatment contrasts, which we refer to as the orthogonal parameterization (OP). While most design results and analysis strategies have been developed and understood within the scope of the OP, a more appropriate alternative in some situations is the baseline parameterization (BP). In this study, we examine the relationship between the OP and the BP, which allows us to better understand the relatively unexplored BP. In addition to being insightful, this relationship is useful in design construction. The design properties considered here are estimability, optimality, and robustness. We find that a general class of Rechtschaffner designs exhibit robust properties under the BP.

Key words and phrases: Effect hierarchy, efficiency criterion, minimum aberration, orthogonal array, Rechtschaffner design, robust design.

1. Introduction In many industrial and scientific investigations, the objective is to build a model that can adequately describe how the response of a system changes when the levels of the input factors change. The impact on the mean response caused by changing the levels of one or more factors is called a factorial effect. The most commonly adopted definition of factorial effects for a 2^m factorial, given by Box and Hunter (1961), is a set of mutually orthogonal treatment contrasts, called the orthogonal parameterization (OP). Despite having received less attention, a more appropriate alternative in some situations is the baseline parameterization (BP). Under the BP, experimenters are more interested in the effects when non-involved factors are kept at their intrinsic baseline levels.

The BP is relatively underexplored, but is becoming more import. Yang and Speed (2002), Kerr (2006), and Banerjee and Mukerjee (2008) investigated factorial designs under the BP in the context of cDNA microarray experiments. More recently, Mukerjee and Tang (2012) proposed a minimum K-aberration criterion to sequentially minimize the bias in the estimation of main effects caused by non-negligible interactions, in the order of importance given by the effect hierarchical principle (Wu and Hamada (2011), pp.172–3). The construction of minimum K-aberration designs is further considered in Li, Miller, and Tang (2014), Miller and Tang (2016), and Mukerjee and Tang (2016).

Because the factorial effects under the OP and BP are both treatment contrasts, there must exist a linear relationship between them. What cannot be foreseen is the special way one set of effects depends on the other. This special pattern in the linear relationship has some important implications in the construction of baseline designs. We aim to derive this relationship and explore its applications to design construction under the BP in terms of estimability, optimality, and robustness.

The rest of this paper is organized as follows. In Section 2, we first provide formal definitions of factorial effects under the OP and the BP. Then, we derive the linear relationship between the two types of parameterization and examine its implications. Section 3 shows how to use the results in Section 2 to find designs under the BP. Here we show that certain orthogonal arrays continue to be optimal under the BP. General Rechtschaffner designs are introduced, and are shown to enjoy a robust property under the BP. Section 4 concludes the paper. All proofs are given in the appendix.

2. The relationship between the OP and the BP

Consider a factorial experiment involving m two-level factors $F_1, F_2, \ldots, F_m$, each at levels zero and one. Let τ_g denote the mean response at the treatment combination $g = (g_1, g_2, \ldots, g_m)$, with $g_i = 0$ or 1 $(i = 1, 2, \ldots, m)$, and let $\mathcal{G}$ be the collection of all 2^m treatment combinations. Because the treatment combination $(1, 1, 0, \ldots, 0)$ corresponds to the subset $\{1, 2\}$ of $S = \{1, 2, \ldots, m\}$, we use τ_{12} and $\tau_{(1,1,0,\ldots,0)}$ interchangeably, depending on which one is more convenient within the context. Under the OP, for a subset $v = \{i_1, i_2, \cdots, i_k\}$ of S, the k-factor interaction $F_{i_1} F_{i_2} \cdots F_{i_k}$ (the main effect if $k = 1$) is given by

$$\beta_v = \frac{1}{2^m} \sum_{g \in \mathcal{G}} \tau_g (-1)^{\sum_{h=1}^{k} g_{i_h}}. \tag{2.1}$$

We let $\beta_\phi = 2^{-m} \sum_{\mathcal{G}} \tau_g$, which is the grand mean. Under the BP, the main effect of F_i is given by $\theta_i = \tau_i - \tau_\phi$, and the two-factor interaction $F_i F_j$ is given by $\theta_{ij} = \tau_{ij} - \tau_i - \tau_j + \tau_\phi$. More generally, for a subset $w = \{i_1, i_2, \cdots, i_k\}$ of S, the k-factor interaction $F_{i_1} F_{i_2} \cdots F_{i_k}$ under the BP is given by

$$\theta_w = \sum_{u \subseteq w} \tau_u (-1)^{|w| - |u|}, \tag{2.2}$$

where $|\cdot|$ stands for the cardinality of a set.

Both β_v and θ_w measure the impact on τ_g caused by level changing of the involved factor(s). However, the former considers an overall effect, whereas the latter focuses on the situation in which all non-involved factors are set at level zero, the baseline level. For example, consider $v = w = \{1\}$ in (2.1) and (2.2). Let $\mathcal{G}^* = \{(g_2, g_3, \ldots, g_m) : g_i = 0, 1\}$. The main effects of F_1 under the OP and the BP can be written as $\beta_1 =$

$(1/2^m)\sum_{g^*\in\mathcal{G}^*}\left(\tau_{(0,g^*)}-\tau_{(1,g^*)}\right)$ and $\theta_1=\tau_{(1,0,\ldots,0)}-\tau_{(0,0,\ldots,0)}$, respectively. Up to a constant, β_1 averages out the effects of F_1 conditional on every $g^*\in\mathcal{G}^*$, while θ_1 computes only the effect of F_1 when all other factors are set at their baseline levels.

The BP arises naturally when each factor has a null state or a baseline level. For example, in a toxicological study, each factor is a toxin, and each treatment combination is a mix of several toxins. The absence and presence of a particular toxin can be represented by levels zero and one, respectively. In an agricultural experiment, two kinds of fertilizers may be applicable, serving as the two levels of a factor. Then level zero can stand for the currently used fertilizer, and level one for the new fertilizer.

By combining (2.1) and (2.2), we obtain a linear relationship between the OP and BP, as stated in the following theorem.

Theorem 1. *We have that*

(i) $\beta_v=\sum_{w\supseteq v}a_w\theta_w$, *with* $a_w=(-1)^{|v|}2^{-|w|}$,

(ii) $\theta_w=\sum_{v\supseteq w}c_v\beta_v$, *with* $c_v=(-2)^{|w|}$.

In Theorem 1, the θ_w's in the expression of β_v are those with w containing v. A similar phenomenon occurs in the expression of θ_w in terms of β_v. It is this special pattern in the linear relationship between θ_w and β_v that makes it useful in the construction of baseline designs, which we examine in Section 3. Proposition 2 in Mukerjee and Tang (2012), which states that an orthogonal array is universally optimal for estimating the main effects under the BP, is established based on the simple fact that $\theta_i = -2\beta_i$, for $i = 1, 2, \ldots, m$, if $\beta_v = 0$ for all $|v| \geq 2$. A more important implication is that the absence of interactions under the OP yields the same result under the BP, and vice versa. We now consider a situation that is more general than the absence of interactions. For a collection $\mathcal{C}$ of subsets of S, we say it is *echelon* if for any s collected by $\mathcal{C}$, all subsets of s are also collected. Then, Theorem 1 implies the following result.

Corollary 1. *Let $\mathcal{C}$ be echelon. Then, $\beta_v = 0$ for all $v \notin \mathcal{C}$, if and only if $\theta_w = 0$ for all $w \notin \mathcal{C}$. As a special case, the absence of factorial effects of order k or higher is invariant to the choice of the parameterization.*

If a collection of factorial effects, say $\{\beta_v : v \in \mathcal{C}\}$ or $\{\theta_w : w \in \mathcal{C}\}$, are believed to be active, the corresponding models under the OP and BP are, respectively,

$$\tau_g = \sum_{v \in \mathcal{C}} \beta_v \prod_{k \in v} (1 - 2g_k) \ (g \in \mathcal{G}); \tag{2.3}$$

$$\tau_g = \sum_{w \in \mathcal{C}} \theta_w \prod_{k \in w} g_k \ (g \in \mathcal{G}). \tag{2.4}$$

We say that models (2.3) and (2.4) are, respectively, the OP and the BP models associated with $\mathcal{C}$, and are called echelon if $\mathcal{C}$ is echelon. Corollary 1 states that these two models are equivalent if $\mathcal{C}$ is echelon. The main-effect-only model and the models that contain all of the main effects, plus some/all of the two-factor interactions, are most often used in practice, all of which are echelon models. We end this section with two toy examples that illustrate Theorem 1 and Corollary 1.

Example 1. Consider a three-factor system A, with mean responses given by

$$\text{System A: } (\tau_{000}, \tau_{001}, \tau_{010}, \tau_{011}, \tau_{100}, \tau_{101}, \tau_{110}, \tau_{111}) = (1, 1, 1, 1, 2, 2, 5, 5).$$

By equation (2.2), there are only two active factorial effects under the BP: $\theta_1 = 1$ and $\theta_{12} = 3$. However, by equation (2.1), there are three active factorial effects under the OP: $\beta_1 = -1.25$, $\beta_2 = -0.75$, and $\beta_{12} = 0.75$. The OP model that contains only β_1 and β_{12} fails to characterize the mean response structure, because $\mathcal{C} = \{\phi, \{1\}, \{1, 2\}\}$ is not an echelon collection. Applying part (i) of Theorem 1, $\beta_{12} = 0.25\theta_{12} + 0.125\theta_{123} = 0.75$. One can compute β_v similarly for other v.

Example 2. A second system has the following mean responses:

$$\text{System B: } (\tau_{000}, \tau_{001}, \tau_{010}, \tau_{011}, \tau_{100}, \tau_{101}, \tau_{110}, \tau_{111}) = (1, 1, -1, -1, 2, 2, 3, 3).$$

Under the BP, $(\theta_1, \theta_2, \theta_{12}) = (1, -2, 3)$, and all other θ_w are zero. Because the model is associated with an echelon collection $\mathcal{C} = \{\phi, \{1\}, \{2\}, \{1, 2\}\}$, by Corollary 1, the OP model that contains only β_1, β_2, and β_{12} is true as well. Using equation (2.1) to verify this, we find that $(\beta_1, \beta_2, \beta_{12}) = (-1.25, 0.25, 0.75)$, and all other β_v are zero.

3. Finding baseline designs

3.1 Preliminary results

Suppose N experimental runs are allowed in a design $\mathcal{D}$, and let $(g_{i1}, g_{i2}, \cdots, g_{im})$ denote the ith run $(i = 1, 2, \ldots, N)$. Under design $\mathcal{D}$, the OP and BP models associated with $\mathcal{C}$ are, respectively,

$$E(Y_i) = \sum_{v \in \mathcal{C}} \beta_v \prod_{j \in v} (1 - 2g_{ij}) \ (i = 1, 2, \ldots, N); \tag{3.5}$$

$$E(Y_i) = \sum_{w \in \mathcal{C}} \theta_w \prod_{j \in w} g_{ij}, \ (i = 1, 2, \ldots, N), \tag{3.6}$$

where Y_i is the response of the ith run. Let $X_{\mathcal{C}}$ and $W_{\mathcal{C}}$ be the model matrices of (3.5) and (3.6), respectively. A design is said to be able to estimate model (3.5) (respectively, model (3.6)) if $X'_{\mathcal{C}} X_{\mathcal{C}}$ (respectively, $W'_{\mathcal{C}} W_{\mathcal{C}}$) is invertible.

Theorem 2. *If a design is able to estimate an echelon OP model, it is able to estimate its counterpart BP model, and vice versa.*

Theorem 2 allows the estimability of certain BP models to be established with little effort. One example is that the full kth-order model, the model that contains all factorial effects of order k or lower, can be estimated under an orthogonal array of strength $2k$. Another interesting application of Theorem 2 is given in the next example.

Example 3. Cheng (1995) showed that an N-run orthogonal array, if N is not a multiple of eight, can estimate the full second-order model when projected onto any four factors. This projection property, by Theorem 2, holds regardless of the parameterization.

For a design $\mathcal{D}$ and an OP model associated with $\mathcal{C}$, we define its $D_{\mathcal{C}}$-efficiency as $\det(X_{\mathcal{C}}'X_{\mathcal{C}})$, and its $A_{\mathcal{C}}$-efficiency as $\mathrm{trace}(X_{\mathcal{C}}'X_{\mathcal{C}})^{-1}$. We say a design is $D_{\mathcal{C}}$-optimal (respectively, $A_{\mathcal{C}}$-optimal) if it maximizes $\det(X_{\mathcal{C}}'X_{\mathcal{C}})$ (respectively, minimizes $\mathrm{trace}(X_{\mathcal{C}}'X_{\mathcal{C}})^{-1}$) among all competing designs. Similarly, we can define the $D_{\mathcal{C}}$- and $A_{\mathcal{C}}$-optimality criteria under the BP by replacing $X_{\mathcal{C}}$ with $W_{\mathcal{C}}$.

Proposition 1. *Let $\mathcal{C}$ be an echelon collection. If a design is $D_{\mathcal{C}}$-optimal under the OP, it is $D_{\mathcal{C}}$-optimal under the BP, and vice versa.*

Proposition 1 is an implication of a more general result given by Proposition 2, which can be derived directly from Theorem 1. Note that Propositions 1 and 2 are both special cases of Lemma 6 in Stallings and Morgan (2015), though stated in a different context.

Proposition 2. *If $\mathcal{C}$ is echelon, then det$(X'_{\mathcal{C}}X_{\mathcal{C}})$ is proportional to det$(W'_{\mathcal{C}}W_{\mathcal{C}})$. The ratio does not depend on the design, but on $\mathcal{C}$ alone.*

We conclude this subsection with a corollary. Its implication will be discussed after Theorem 3 in the next subsection.

Corollary 2. *Let $\mathcal{C}$ be an echelon collection. The $D_{\mathcal{C}}$-efficiency of a design remains unchanged under level switching of one or more factors, regardless of the parameterization.*

3.2 Designs from orthogonal arrays

Cheng (1980) showed that an orthogonal array is universally optimal under the main-effect-only model. As another example, a design given by an orthogonal array of strength $2k$ is A- and D-optimal under the full kth-order model. These results are all obtained all under the OP. In this subsection, we generalize a result of Moriguti (1954) to baseline designs. We also comment on generating baseline designs with robust properties.

Consider the OP model associated with $\mathcal{C}$, and let $\hat{\beta}_v$ be the least squares estimator of β_v. We assume, as usual, that all observations are uncorrelated and have a common variance. Moriguti (1954) proved that a design in which the model matrix $X_{\mathcal{C}}$ has mutually orthogonal columns minimizes $\mathrm{Var}(\hat{\beta}_v)$ for each $v \in \mathcal{C}$ among all competing designs. The next theorem states that a similar result holds for the BP if $\mathcal{C}$ is echelon.

Theorem 3. *Under an OP model associated with $\mathcal{C}$, a design $\mathcal{D}$ minimizes Var$(\hat{\beta}_v)$ for each $v \in \mathcal{C}$ among all competing designs if $X_{\mathcal{C}}$ is orthogonal. Furthermore, if $\mathcal{C}$ is echelon, then under the counterpart BP model, $\mathcal{D}$ also minimizes Var$(\hat{\theta}_w)$ among all competing designs for every w in $\mathcal{C}$ that is not contained by another u in $\mathcal{C}$.*

For convenience, we call θ_w a *cap effect* if w is not contained by another u in $\mathcal{C}$. Then, Theorem 3 establishes the optimality for every cap effect under the stated conditions. Cap effects should be tested first for their significance when seeking a simpler model in the analysis stage. We consider some useful cases. If the main-effects model is considered with the inclusion of an intercept, then all the main effects are cap effects. Therefore, Theorem 3 generalizes a result of Mukerjee and Tang (2012), who established the optimality for every main effect. For a model consisting of all main effects and all two-factor interactions, the two-factor interactions are cap effects.

In a model of all main effects plus some two-factor interactions, these two-factor interactions are cap effects, as are the main effects not involved in these two-factor interactions.

Because switching the two levels does not affect the orthogonality of $X_{\mathcal{C}}$, Theorem 3 also suggests a simple strategy for generating an efficient baseline design that is robust to non-negligible effects. While a full investigation of this problem is beyond the scope of this study, we give an example to illustrate the idea.

Example 4. Consider the model associated with $\mathcal{C} = \{\phi, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}\}$ and an eight-run design $\mathcal{D}$, displayed in transposed form below:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Design $\mathcal{D}$ is a resolution-IV regular design. Because the design has an orthogonal model matrix $X_{\mathcal{C}}$, it has the optimal properties given in Theorem 3. Let $\mathcal{D}^*$ be the design obtained from $\mathcal{D}$ by level switching the fourth factor. Then, $\mathcal{D}^*$ has the same optimality properties as $\mathcal{D}$. To further distinguish one design from the other, we compute the bias caused by non-negligible effects. Assume θ_{24} is the only non-negligible effect. Following the idea of the minimum K-aberration, the design with smaller value of $\|(W'_{\mathcal{C}}W_{\mathcal{C}})^{-1}W'_{\mathcal{C}}W_{24}\|$ is preferred, where $W_{\mathcal{C}}$ is the model matrix under the BP, W_{24} is the Hadarmard product of the second and fourth factors in the design matrix, and $\|\cdot\|$ denotes the Euclidean norm. Because $\|(W'_{\mathcal{C}}W_{\mathcal{C}})^{-1}W'_{\mathcal{C}}W_{24}\|$ is equal to 2 for $\mathcal{D}$ and 0.816 for $\mathcal{D}^*$, $\mathcal{D}^*$ is preferred.

3.3 Rechtschaffner designs

Consider the full second-order model associated with the collection $\mathcal{C}_2 = \{s \subseteq S : |s| \leq 2\}$. Based on the aforementioned one-to-one correspondence between a subset and a treatment combination, $\mathcal{C}_2$ corresponds to a design consisting of $(1 + m + m(m-1)/2)$ different treatment combinations, which is known as the Rechtschaffner design, denoted by $\mathcal{D}_{\mathcal{C}_2}$. Using the same correspondence, we define $\mathcal{D}_{\mathcal{C}}$ similarly for any $\mathcal{C}$, and still call it a Rechtschaffner design. Design $\mathcal{D}_{\mathcal{C}_2}$ was first presented by Rechtschaffner (1967), who suggested its use under the full second-order model. The estimability of $\mathcal{D}_{\mathcal{C}_2}$ under the OP was later proved by several authors, with generalizations to echelon models for mixed-level and/or higher-order situations. We state a result for the two-level situation, which is a special case of Theorem 15.25 in Cheng (2014).

Proposition 3. *For an echelon collection $\mathcal{C}$, the OP model associated with $\mathcal{C}$ is estimable under the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$.*

Under the BP, the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$ has a stronger property.

Theorem 4. *For any collection $\mathcal{C}$, the BP model associated with $\mathcal{C}$ is estimable under the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$.*

Compared with Proposition 3, Theorem 4 does not assume that $\mathcal{C}$ is echelon. A special case of Rechtschaffner designs is $\mathcal{D}_{\mathcal{C}_1}$ with $\mathcal{C}_1 = \{s \subseteq S : |s| \leq 1\}$. This design, commonly known as a one-factor-at-a-time design, was discussed in Mukerjee and Tang (2012) for its following robust property: non-negligible interactions never cause bias in the estimation of the main effects under the BP. This property, in fact, holds for any Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$ with an echelon $\mathcal{C}$.

Theorem 5. *Let $\mathcal{C}$ be an echelon collection. Then, the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$ allows an unbiased estimation of the BP model associated with $\mathcal{C}$, even if the effects outside the model are non-negligible.*

Example 5. Consider the model associated with $\mathcal{C} = \{\phi, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}\}$ and the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$, displayed in transposed form below:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

If θ_{24} is a non-negligible effect, the bias it causes can be found using

$(W'_{\mathcal{C}}W_{\mathcal{C}})^{-1}W'_{\mathcal{C}}W_{24}\theta_{24}$. It is clear that W_{24} is an all-zero vector; hence, θ_{24} does not cause bias in $\hat{\theta}_w$, for all $w \in \mathcal{C}$. The same argument can be made for all other effects outside the model.

Though the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$ enjoys a nice property of robustness, it is not very efficient. We now consider a class of N-run Rechtschaffner designs based on $\mathcal{D}_{\mathcal{C}}$, where $\mathcal{C} = \{s_0 = \phi, s_1, s_2, \ldots, s_p\}$, by allowing each run in $\mathcal{D}_{\mathcal{C}}$ to appear multiple times. Let f_j be the number of times the treatment combination corresponding to s_j appears in $\mathcal{D}_{\mathcal{C}}$, for $j = 0, 1, \ldots, p$, where $N = \sum_{j=0}^{p} f_j$. The next result gives an optimal allocation.

Proposition 4. *Let $\mathcal{C}$ be an echelon collection. An N-run Rechtschaffner design based on $\mathcal{D}_{\mathcal{C}}$ is $A_{\mathcal{C}}$-optimal under the BP if $f_j = Nq_j^{1/2} / \sum_{j=0}^{p} q_j^{1/2}$, for $j = 0, 1, \ldots, p$, where q_j is the number of subsets in $\mathcal{C}$ that contain s_j.*

4. Conclusion

We have derived a linear relationship between the OP and the BP. From its special pattern, we conclude that an echelon model has the same form under the two types of parameterization. We further discuss its implications for the estimability, optimality, and robustness of baseline designs. In particular, we show that certain orthogonal arrays continue to be optimal under the BP. We introduce general Rechtschaffner designs, showing they enjoy a robust property that is only available under the BP.

There are two possible future research directions. The first is illustrated by Example 5, in which we find the level permutations that minimize the bias caused by non-negligible effects. Under the main-effect-only model, this has been investigated by Mukerjee and Tang (2012) and Li, Miller, and Tang (2014). However, it would be useful to obtain results for more general echelon models. The second is to consider a compromise between robust and optimal designs, which can be done by adding runs to a Rechtschaffner design. The compromise designs are expected to enjoy in-between performance in terms of both efficiency and robustness, as demonstrated for the main-effect model of Karunanayaka and Tang (2017).

Acknowledgments

This research was supported by the Natural Science and Engineering Research Council of Canada.

A. Appendix: Proofs

A.1 Proof of Theorem 1

Let τ be a column vector with componenets $\tau_\phi, \tau_1, \tau_2, \tau_{12}, \ldots, \tau_{12\cdots m}$ in Yates order. Vectors θ and β are similarly defined. Let H_m be the m-fold Kronecker product of H and L_m the m-fold Kronecker product of L, where

$$H = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \text{ and } L = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

We then have $\beta = H_m\tau$ and $\tau = L_m\theta$. Therefore $\beta = H_mL_m\theta$ and $\theta = (H_mL_m)^{-1}\beta$. Theorem 1 follows by noting that H_mL_m is the m-fold Kronecker product of HL and $(H_mL_m)^{-1}$ is the m-fold Kronecker product of $(HL)^{-1}$ and the special forms of HL and $(HL)^{-1}$ as given by

$$HL = \begin{bmatrix} 1 & 1/2 \\ 0 & -1/2 \end{bmatrix} \text{ and } (HL)^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & -2 \end{bmatrix}.$$

A.2 Proof of Theorem 2

This result follows immediately from Proposition 2.

A.3 Proof of Corollary 2

For a design $\mathcal{D}$, let $\mathcal{D}_\pi$ be the design obtained from $\mathcal{D}$ by level switching one or more factors. We use W and W_π to denote the model matrices under $\mathcal{D}$ and $\mathcal{D}_\pi$ for the BP, respectively. Matrices X and X_π are defined similarly for the OP. By Proposition 2, the ratio $(\det(X'X)/\det(W'W)) = (\det(X_\pi'X_\pi)/\det(W_\pi'W_\pi))$ is a constant which only depends on the model. Since $\det(X'X) = \det(X_\pi'X_\pi)$, we conclude that $\det(W'W) = \det(W_\pi'W_\pi)$.

A.4 Proof of Theorem 3

Due to a result by Moriguti (1954), $\mathrm{Var}(\hat{\beta}_v)$ attains its minimal value for each $v \in \mathcal{C}$ if $X_{\mathcal{C}}$ is orthogonal. If $\mathcal{C}$ is echelon, by Theorem 1 and Corollary 1, we have that $\theta_w = \sum_{v \supseteq w, v \in \mathcal{C}} c_v \beta_v$. If w is not contained by another u in $\mathcal{C}$, then $\theta_w = c_w \beta_w$. Thus, $\mathrm{Var}(\hat{\theta}_w) = c_w^2 \mathrm{Var}(\hat{\beta}_w)$ is minimized.

A.5 Proof of Theorem 4

Consider the matrix $W_m = L_m$ in the proof of Theorem 1, which is the model matrix of the full model. Let W_m^* be the $N \times N$ submatrix of W_m, obtained by deleting all rows and columns except for the j_1-, j_2-,..., j_N-th rows and columns. It is sufficient to show that W_m^* is non-singular. Note that $j_1 = 1$ since a Rechtschaffner design always contains $g = (0, \ldots, 0)$ and the model always contains the intercept. The non-singularity of W_m^* is easily seen since W_m is a lower triangular matrix with all diagonals being one, which the case is because $W_m = W_{m-1} \otimes W_1$ and W_1 has the same pattern.

A.6 Proof of Theorem 5

Let $\mathcal{C} = \{s_0 = \phi, s_1, s_2, \ldots, s_p\}$. Without loss of generality, let the i-th run $g_i = (g_{i1}, \ldots, g_{im})$ correspond to s_i, $i = 0, 1, \ldots, p$. The fitted model can be written as $E(Y) = W_{\mathcal{C}}\theta_{\mathcal{C}}$, where $E(Y) = (\tau_{s_0}, \tau_{s_1}, \ldots, \tau_{s_p})'$ and $\theta_{\mathcal{C}} = (\theta_{\phi}, \theta_{s_1}, \ldots, \theta_{s_p})'$. Since there may exist some non-negligible effects θ_w with $w \notin \mathcal{C}$, we let the true model be $E(Y) = W_{\mathcal{C}}\theta_{\mathcal{C}} + \sum_{w \notin \mathcal{C}} W_w \theta_w$, where W_w is a $(p+1) \times 1$ column vector with the i-th entry equal to $\prod_{j \in w} g_{ij}$.

Let $\hat{\theta}_{\mathcal{C}}$ be the least square estimator from the fitted model. Then, $E(\hat{\theta}_{\mathcal{C}}) = (W'_{\mathcal{C}}W_{\mathcal{C}})^{-1}W'_{\mathcal{C}}E(Y) = \theta_{\mathcal{C}} + \sum_{w \notin \mathcal{C}}(W'_{\mathcal{C}}W_{\mathcal{C}})^{-1}W'_{\mathcal{C}}W_w\theta_w$. Thus, if we can show that for each $w \notin \mathcal{C}$, W_w is an all-zeros column vector, then the proof is completed. This is evident because $\prod_{j \in w} g_{ij}$ is one if s_i contains w as a subset, and zero otherwise. However, due to the fact that $\mathcal{C}$ is echelon, no s_i can contain w as a subset.

A.7 Proof of Proposition 4

Let model (3.6) under the Rechtschaffner design $\mathcal{D}_{\mathcal{C}}$ (i.e., $f_j = 1$ for $j = 0, 1, \ldots, p$.) be $E(Y) = W_{\mathcal{C}}\theta_{\mathcal{C}}$, where $E(Y) = (\tau_{s_0}, \tau_{s_1}, \ldots, \tau_{s_p})'$ and $\theta_{\mathcal{C}} = (\theta_{\phi}, \theta_{s_1}, \ldots, \theta_{s_p})'$. Consider an N-run Rechtschaffner design and let E be the $(p+1) \times (p+1)$ identity matrix. The model matrix can be written as $AW_{\mathcal{C}}$, where A is an $N \times (p+1)$ matrix. The first f_0 rows of A are the first row of E, the following f_1 rows are the second row of E, and so on. The $A_{\mathcal{C}}$-efficiency is

$$\mathrm{tr}\left((AW_{\mathcal{C}})'(AW_{\mathcal{C}})\right)^{-1} = \mathrm{tr}\left(W_{\mathcal{C}}^{-1}(A'A)^{-1}(W'_{\mathcal{C}})^{-1}\right) = \mathrm{tr}\left((A'A)^{-1}(W'_{\mathcal{C}})^{-1}(W_{\mathcal{C}})^{-1}\right)$$

It is evident that $(A'A)^{-1} = diag(f_0^{-1}, f_1^{-1}, \ldots, f_p^{-1})$, so the $A_{\mathcal{C}}$-efficiency is $\sum_{j=0}^{p} q_j f_j^{-1}$, where q_j is the (j,j)-th element of $(W_{\mathcal{C}}')^{-1}(W_{\mathcal{C}})^{-1}$, for $j = 0, 1, \ldots, p$. By Cauchy-Schwarz inequality, subject to $\sum_{j=0}^{p} f_j = N$, $\sum_{j=0}^{p} q_j f_j^{-1}$ is minimized if $f_j = N \left({q_j}^{0.5} / \sum_{j=0}^{p} {q_j}^{0.5} \right)$, so the proof can be completed by showing q_j is the number of subsets in $\mathcal{C}$ that contain s_j.

By definition (2.2), for any $w \in \mathcal{C}$, $\theta_w = \sum_{u \subseteq w} \tau_u (-1)^{|w|-|u|}$, which is equal to $\sum_{u \in \mathcal{C}, u \subseteq w} \tau_u (-1)^{|w|-|u|}$ since $\mathcal{C}$ is echelon. It is then implied that $\theta_{\mathcal{C}} = W_{\mathcal{C}}^{-1} E(Y)$ gives the definition back, and thus the j-th column of $W_{\mathcal{C}}^{-1}$ is

$$\left((-1)^{|s_0|-|s_j|} I(s_0 \supseteq s_j), (-1)^{|s_1|-|s_j|} I(s_1 \supseteq s_j), \ldots, (-1)^{|s_p|-|s_j|} I(s_p \supseteq s_j) \right)',$$

where $I(s_i \supseteq s_j) = 1$ if s_i contains s_j as a subsets, and 0 otherwise. Now we can find that the (j,j)th element of $(W_{\mathcal{C}}')^{-1}(W_{\mathcal{C}})^{-1}$, which is the squared length of the jth column vector of $W_{\mathcal{C}}^{-1}$, is $\sum_{i=0}^{p} \{(-1)^{|s_i|-|s_j|} I(s_i \supseteq s_j)\}^2 = \sum_{i=0}^{p} I(s_i \supseteq s_j)$ $(j = 0, \ldots, p)$.

References

Banerjee, T. and Mukerjee, R. (2008). Optimal factorial designs for cDNA microarray experiments. *The Annals of applied statistics*, **2**(1), 366-385.

Box, G. E. and Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs. *Technometrics*, **3**(3), 311-351.

Cheng, C. S. (1980). Orthogonal arrays with variable numbers of symbols. *The Annals of Statistics*, **8**(2), 447-453.

Cheng, C. S. (1995). Some projection properties of orthogonal arrays. *The Annals of Statistics*, 1223-1233.

Cheng, C. S. (2014). *Theory of factorial design: Single-and multi-stratum experiments.* CRC Press.

Karunanayaka, R. C. and Tang, B. (2017). Compromise designs under baseline parameterization. *Journal of Statistical Planning and Inference*, **190**, 32-38.

Kerr, K. F. (2006). Efficient 2^k factorial designs for blocks of size 2 with microarray applications. *Journal of quality technology*, **38**(4), 309-318.

Li, P., Miller, A., and Tang, B. (2014). Algorithmic search for baseline minimum aberration designs. *Journal of Statistical Planning and Inference*, **149**, 172-182.

Miller, A. and Tang, B. (2016). Using regular fractions of two-level designs to find baseline designs. *Statistica Sinica*, 745-759.

Moriguti, S. (1954). Optimality of orthogonal designs. *Rep. Statist. Appl. Res. Un. Japan Sci. Engrs*, **3**, 75-98.

Mukerjee, R. and Tang, B. (2012). Optimal fractions of two-level factorials under a baseline parameterization. *Biometrika*, **99**(1), 71-84.

Mukerjee, R. and Tang, B. (2016). Optimal two-level regular designs under baseline parametrization via cosets and minimum moment aberration. *Statistica Sinica*, 1001-1019.

Rechtschaffner, R. L. (1967). Saturated fractions of 2^n and 3^n factorial designs. *Technometrics*, **9**(4), 569-575.

Stallings, J. W. and Morgan, J. P. (2015). General weighted optimality of designed experiments. *Biometrika*, **102**(4), 925-935.

Wu, C. J. and Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization* (Vol. 552). John Wiley & Sons.

Yang, Y. H. and Speed, T. (2002). Design issues for cDNA microarray experiments. *Nature Reviews Genetics*, **3**(8), 579-588.

Simon Fraser University, Canada

E-mail: chengyus@sfu.ca

Simon Fraser University, Canada

E-mail: boxint@sfu.ca