Open Journal of Applied Sciences, 2026, 16(1), 367-382
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

Identify the Optimal Baseline Design from the Plackett-Burman Design

Mingze Xin

School of Mathematics and Statistics, Shandong Normal University, Jinan, China
Email: 2640228674@qq.com

How to cite this paper: Xin, M.Z. (2026) Identify the Optimal Baseline Design from the Plackett-Burman Design. *Open Journal of Applied Sciences*, **16**, 367-382.
https://doi.org/10.4236/ojapps.2026.161023

Received: December 24, 2025
Accepted: January 19, 2026
Published: January 22, 2026

Abstract

In recent years, baseline designs have garnered extensive attention due to their broad applicability across numerous fields. In order to select excellent baseline designs, Mukerjee and Tang (2012) proposed the K-aberration criterion. The Plackett-Burman design represents a classic category of non-regular designs and has been relatively less studied as a baseline design. This paper, starting from the Plackett-Burman design, investigates the properties it exhibits when employed as a baseline design. We have obtained an important property: The sub-designs constructed from the column sets of the Plackett-Burman design with equivalent distance vectors possess the same K-aberration sequence. This property improves the efficiency of our search for the optimal K-aberration sub-designs of Plackett-Burman design.

Keywords

Baseline Design, Plackett-Burman Design, Cyclic Generation Method

1. Introduction

Previous research on regular and nonregular designs has largely been based on orthogonal parameterization. In recent years, baseline designs based on baseline parameterization have garnered significant attention due to their wide applicability. Below, we will provide specific examples to explain what baseline parameterization and orthogonal parameterization are.

For an orthogonal design with N rows and m columns, where the factor levels are 0 and 1, select any $s, 2 \le s \le m$ columns $g_1, \cdots, g_s, g_1, \cdots, g_s \in \{1, \cdots, m\}$ from this design to form a set S. Let $\phi(S)$ denote the vector obtained by summing the corresponding elements of the columns in S $(\bmod 2)$. Let $\varphi(S)$ denote the sum of the elements in the vector $\phi(S)$. Let $J_s(S) = |2\varphi(S) - N|$. It is important to emphasize that when the factor levels of the design under discussion are −1 and 1, the definition of $J_s(S)$ undergoes the following corresponding changes: $J_s(S) = \left|\sum_{i=1}^{N} j_{ig_1} j_{ig_2} \cdots j_{ig_s}\right|$. Where j_{io} represents the element in the i-th row and $o, o \in \{g_1, \cdots, g_s\}$-th column of the design. Regardless of whether the factor levels in the orthogonal design are 0, 1, or 1, −1. If $J_s(S) = 0$, then the s columns are said to be completely orthogonal. If $J_s(S) = N$, then the s columns are said to be completely confounded. If $0 < J_s(S) < N$, then the s columns are said to be partially confounded. What has been introduced above is orthogonal parameterization. The specific details can be found in Deng and Tang (1999) [1]. The left table in **Table 1** represents an orthogonal design with factor levels of 0 and 1, while the middle table in **Table 1** represents an orthogonal design with factor levels of −1 and 1. The first and second columns of both tables are completely orthogonal, while the first, second, and fourth columns are completely confounded. The difference between baseline parameterization and orthogonal parameterization lies in the fact that, for a baseline design with N rows and m columns, the m factors do not exhibit orthogonality or confounding. Furthermore, in a baseline design, the factor levels are restricted to 0 and 1. When selecting s factors from the m factors in the baseline design, the interaction effects among these s factors can be represented by the product of the corresponding elements in the s columns. The specific details can be found in Mukerjee and Tang (2012) [2]. In **Table 1**, columns 1 to 5 in the right-hand table represent the design matrix of the baseline design, while column 6 represents the interaction between columns 1 and 2, and column 7 represents the interaction between columns 1, 2, and 3. Baseline design is a design where factor levels are restricted to 0 and 1, and the number of 0s and 1s in its design matrix can vary arbitrarily. In contrast, orthogonal design under orthogonal parameterization, the factor levels can be 0, 1, or −1, 1 and the design matrix must satisfy orthogonality between any two columns.

Table 1. The table on the left and the table in the middle represent orthogonal designs, while the table on the right represents the baseline design.

1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	6	7
0	0	0	0	0	−1	−1	−1	−1	−1	0	0	0	0	0	0	0
0	0	1	0	1	−1	−1	1	−1	1	0	0	1	0	1	0	0
0	1	0	1	1	−1	1	−1	1	1	0	1	0	1	1	0	0
0	1	1	1	0	−1	1	1	1	−1	0	1	1	1	0	0	0
1	0	0	1	0	1	−1	−1	1	−1	1	0	0	1	0	0	0
1	0	1	1	1	1	−1	1	1	1	1	0	1	1	1	0	0
1	1	0	0	1	1	1	−1	−1	1	1	1	0	0	1	1	0
1	1	1	0	0	1	1	1	−1	−1	1	1	1	0	0	1	1

We use D to denote a baseline design with N rows and m columns. The matrix D is a matrix with elements 0 and 1, 0 represents the baseline level, while 1 denotes the test level. Let $\Omega_s(D)$ be the set of all $N \times s$ submatrices of D. Unless otherwise specified, we will denote this set by Ω_s. Here is an example of a baseline design with 8 rows and 5 columns.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}$$

Mukerjee and Tang (2012) [2] proposed a theory related to the optimality of baseline designs. They focused on main effects designs and proved that designs with strength 2 have universal optimality in estimating main effects. When estimating main effects, the presence of active interaction effects can cause bias in the estimation of the main effects. Consider the baseline design D, where the m factors of D are denoted as $F_1, \cdots, F_m$. Let $g_1, \cdots, g_s, 1 \le g_1 < \cdots < g_s \le m$ be any s numbers selected from $1, \cdots, m$. We denote $\theta_{g_1,\cdots,g_s}$ as the interaction effects of the s factors $F_{g_1}, \cdots, F_{g_s}$. Let $\alpha(g_1, \cdots, g_s)$ denote the number of occurrences of the row vector $(1,1,\cdots,1)$ in the submatrix of D formed by the $g_1, \cdots, g_s$-th columns. Here, we denote $\mu(g_1, \cdots, g_s)$ as an $m \times 1$ column vector. We then define the i-th element of the vector $\mu(g_1, \cdots, g_s)$ as follows. When $i \in \{g_1, \cdots, g_s\}$, let the i-th element of $\mu(g_1, \cdots, g_s)$ be $\alpha(g_1 \cdots g_s)$; when $i \notin \{g_1, \cdots, g_s\}$, let the i-th element of $\mu(g_1, \cdots, g_s)$ be $\alpha(\langle i g_1 \cdots g_s \rangle)$, where $\langle i g_1 \ldots g_s \rangle$ represents the ascending order of $i, g_1, \cdots, g_s$. Mukerjee and Tang(2012) [2] pointed out that if there are interaction effects among $F_{g_1}, \cdots, F_{g_s}$, these interaction effects will contribute a bias of magnitude $\xi(g_1, \cdots, g_s)\theta_{g_1,\cdots,g_s}$ when estimating the main effects, where $\xi(g_1, \cdots, g_s) = (2/N)(2\mu(g_1, \cdots, g_s) - \alpha(g_1, \cdots, g_s)1_m)$. Here, 1_m represents a column vector of size $m \times 1$, where all elements are ones. According to the principle that interaction effects of the same order are of equal importance, they provided that $K_s = \sum_{\Omega_s} \xi(g_1, \cdots, g_s)' \xi(g_1, \cdots, g_s)$. It is evident that K_s measures the bias effect of all s-th order interaction effects on the estimation of the main effects. Since lower-order interaction effects are more important than higher-order interaction effects, we aim to find a design that sequentially minimizes $(K_2, K_3, \cdots)$. This is precisely the minimum K-aberration criterion proposed by Mukerjee and Tang (2012) [2]. The minimum K-aberration criterion is defined as follows in Definition 1.

Definition 1. *Consider two baseline designs D_1 and D_2 with the same number of rows and columns, both of which are orthogonal designs with strength 2. Let j' be the smallest integer at which the values of K_j for D_1 and D_2 first differ. If the value of $K_{j'}$ for D_1 is smaller than that for D_2, then D_1 is said* *to have a smaller* K *-aberration than* D_2*. A minimum* K *-aberration design is one in which no design exists with a smaller* K *-aberration.*

Mukerjee and Tang (2012) [2] proved the following expression, where $s = 2,\cdots,m-1$, and let $\alpha(\omega)$ denote the number of rows in $\omega \in \Omega_s$ in which all elements are equal to 1.

$$K_s = 4/N^2\left(sT_1 + T_2\right), \tag{1}$$

where $T_1 = \sum_{\omega\in\Omega^s}\left(\alpha(\omega)\right)^2$ and $T_2 = \sum_{\omega^*\in\Omega^{s+1}}\sum_{\omega^\circ\in\Omega^s\left(\omega^*\right)}\left(2\alpha\left(\omega^*\right)-\alpha\left(\omega^\circ\right)\right)^2$. The derivation details of this sequence can be found in the literature by Mukerjee and Tang (2012) [2].

In order to reduce the computational effort required to find optimal baseline designs, Mukerjee and Tang (2016) [3] provided an equivalent transformation form Z_s for $K_s, 2 \le s \le m$ based on the original approach, and referred to Z_s as a moment confounding of a design. The paper points out that the sequential minimization of a design's $K_2,\cdots,K_m$ is equivalent to the sequential minimization of the design's $Z_2,\cdots,Z_m$. The method of finding an optimal baseline design by minimizing the sequence of $Z_2,\cdots,Z_m$ is applicable to both regular and non-regular designs. The definition of Z_s is $Z_s = N^{-2}tr\left(ZZ'\right)^{[s]}WW'$. The matrix $Z = \left(z_{uj}\right), 1 \le u \le N, 1 \le j \le m$ is a baseline design matrix of size $N \times m$ with elements 0 and 1. $W = J_{Nm} - 2Z$, where J_{Nm} is an $N \times m$ matrix with all elements equal to 1.

This paper starts with the Plackett-Burman design to identify designs with good K -aberration properties. Chapter 2 introduces the relevant symbols and definitions of the Plackett-Burman design. Chapter 3 introduces the properties of baseline designs when studying them starting from the Plackett-Burman design. Chapter 4 presents the application of the theory, where the minimum K -aberration subdesign with 7 columns and 24 rows was identified for a Plackett-Burman design with 24 rows and 23 columns.

2. Symbols and Definitions Related to Plackett-Burman Design

In this section, we introduce some symbols and definitions related to Plackett-Burman design that will be used in the derivations of subsequent lemmas or theorems.

We use P to denote a Plackett-Burman design of size $N \times m$, where the elements are either 0 or 1. Plackett-Burman design is typically generated by a special cyclic method [4]. Let the first row of P be α_1. α_1 is a row vector containing only 0 s and 1 s, with at least one occurrence of both 0 and 1. We represent α_1 as $\left(a_1,\cdots,a_m\right)$, The element $a_i, i \in \{1,\cdots,m\}$ refers to the i -th element in the first row of P. Let us assume that α_2 represents the second row of P. Then, the 2nd to the m -th elements of α_2 are respectively equal to the 1st to the $(m-1)$ -th elements of α_1. The first element of α_2 is equal to the m -th element of α_1. That is, $\alpha_2 = \left(a_m, a_1,\cdots,a_{m-1}\right)$. Let us assume that α_3 represents the third row of P. Then, the 2nd to the m-th elements of α_3 are respectively equal to the 1st to the $(m-1)$-th elements of α_2. The first element of α_3 is equal to the m-th element of α_2. That is, $\alpha_3 = (a_{m-1}, a_m, a_1, \cdots, a_{m-2})$. By the same logic, we can derive $\alpha_4, \cdots, \alpha_m$ from α_1. Let α_{m+1} represent the $(m+1)$-th row of P. Here, we define $\alpha_{m+1} = (0, 0, \cdots, 0)$ or $(1, 1, \cdots, 1)$ (This depends on the first line of the Plackett-Burman design). Therefore, P can be expressed as follows.

$$P = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \\ \alpha_{m+1} \end{pmatrix}$$

Here is an example of Plackett-Burman design. Consider an $N \times m$ Plackett-Burman design, where its first row is

$$(1,1,0,1,1,1,0,0,0,1,0).$$

According to the cyclic generation method of Plackett-Burman design, we can obtain its design matrix as follows.

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

3. The Properties of Plackett-Burman Design

Select an arbitrary s-column submatrix $(w_1, \cdots, w_s)(1 \le w_1 < \cdots < w_s \le m)$ from an $N \times m$ Plackett-Burman design. We let d_i represents the number of columns between the w_i-th column and the w_{i+1}-th column within the m columns of Plackett-Burman design, in this case, $i = 1, \cdots, s-1$, it is evident that $d_i = w_{i+1} - w_i - 1$. Specifically, let d_s represent the number of columns between the w_s-th column and the w_1-th column within the m columns of Plackett-Burman design. Specifically, d_s is the sum of the number of columns after the w_s-th column and the number of columns before the w_1-th column within the m columns of Plackett-Burman design. It is evident that $d_s = m - w_s + w_1 - 1$. $(d_1, \cdots, d_s)$ is referred to as the distance vector of $(w_1, \cdots, w_s)$. Let the set of distance vectors $(d_1, \cdots, d_s), (d_2, \cdots, d_s, d_1), \cdots, (d_s, d_1, \cdots, d_{s-1})$ be denoted as D_s.

To ensure the smooth progress of the proofs of the lemma and theorem in this section, we now provide the definitions of function $a_{i,j}$ and function C_i, respectively. The definition of function $a_{i,j}$ can be found in Equation (2), and the definition of function C_i can be found in Equation (3).

$$a_{i,j} = \begin{cases} a_{i-j}, & \text{if } j < i; \\ a_{m-j+i}, & \text{if } i \le j \le m+i-1; \\ a_{2m-j+i}, & \text{if } m+i \le j \le 2m+i-1. \end{cases} \tag{2}$$

The Function $a_{i,j}$ is a binary function, with independent variables i and j, where $i \in \{1,\cdots,m\}$, $j \in \{0,\cdots,2m+x-1\}$. In Equation (2), a_{i-j}, a_{m-j+i} and a_{2m-j+i} represent the $(i-j)$-th, $(m-j+i)$-th and $(2m-j+i)$-th elements in the first row of Plackett-Burman design, respectively.

$$C_i = \begin{cases} i, & \text{if } 1 \le i \le N-1; \\ i-N+1, & \text{if } N \le i \le 2N-2. \end{cases} \tag{3}$$

The function C_i is a unary function, with i as its independent variable, where $i \in \{1,\cdots,2N-2\}$.

Lemma 1. *Select two* s*-column submatrices,* $(g_1,\cdots,g_s)$ *and* $(g_1',\cdots,g_s')$*, from an* $N \times m$ *Plackett-Burman design. Let the distance vectors of the two* s*-column submatrices be denoted as* $(d_1,\cdots,d_s)$ *and* $(d_1',\cdots,d_s')$*, respectively. If* $d_i = d_i'(i=1,\cdots,s)$*, then* $(g_1,\cdots,g_s)$ *and* $(g_1',\cdots,g_s')$ *have the same* $\alpha(\omega)$*.*

Proof. Select two s-column submatrices, $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$, from an $N \times m$ Plackett-Burman design. Let the distance vectors of the two s-column submatrices be denoted as $(d_1,\cdots,d_s)$ and $(d_1',\cdots,d_s')$, respectively. Now, let $d_i = d_i'(i=1,\cdots,s)$. By considering the distance vector of $(g_1,\cdots,g_s)$, $g_1,\cdots,g_s$ can be further expressed in the form of Equation (4).

$$g_i = \begin{cases} g_i, & \text{if } i = 1; \\ g_1 + \sum_{j=1}^{i-1} d_j + i - 1, & \text{if } i = 2,\cdots,s. \end{cases} \tag{4}$$

Similarly, by considering the distance vector of $(g_1',\cdots,g_s')$, $g_1',\cdots,g_s'$ can be further expressed in the form of Equation (5).

$$g_i' = \begin{cases} g_i', & \text{if } i = 1; \\ g_1' + \sum_{j=1}^{i-1} d_j' + i - 1, & \text{if } i = 2,\cdots,s. \end{cases} \tag{5}$$

(i) $g_1 = g_1'$. From Equations (4) and (5), it can be observed that when $g_1 = g_1'$, $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ are two identical s-column submatrices selected from an $N \times m$ Plackett-Burman design. Therefore, it is evident that $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ have the same $\alpha(\omega)$.

(ii) $g_1 \ne g_1'$. Without loss of generality, let $g_1 < g_1'$. Let the number of columns between the g_1-th and g_1'-th columns in the m columns of Plackett-Burman design be d, thus, we have $g_1' = g_1 + d + 1$. From this, $g_1',\cdots,g_s'$ can be further expressed as $g_i' = g_i + d + 1(i=1)$, $g_i' = g_1 + d + 1 + \sum_{j=1}^{i-1} d_j + i - 1(i=2,\cdots,s)$. The first row of $(g_1,\cdots,g_s)$ can be represented as

$$\left(a_{g_1}, a_{g_1+d_1+1}, \cdots, a_{g_1+\sum_{j=1}^{s-1} d_j+s-1}\right). \tag{6}$$

By considering the Plackett-Burman design generation method and the definition of Equation (2), it can be concluded that the $(v+1)(v=0,\cdots,N-2)$-th row of $(g_1,\cdots,g_s)$ can be represented as

$$\left(a_{g_1,v}, a_{g_1+d_1+1,v}, \cdots, a_{g_1+\sum_{j=1}^{s-1} d_j+s-1,v}\right). \tag{7}$$

The first row of $(g_1',\cdots,g_s')$ can be represented as

$$\left(a_{g_1+d+1}, a_{g_1+d+1+d_1+1}, \cdots, a_{g_1+d+1+\sum_{j=1}^{s-1} d_j+s-1}\right). \tag{8}$$

By considering the Plackett-Burman design generation method and the definitions of Equations (2) and (3), the C_{v+d+2}-th row of $(g_1',\cdots,g_s')$ can be represented as

$$\left(a_{g_1+d+1,v+d+1}, a_{g_1+d+1+d_1+1,v+d+1}, \cdots, a_{g_1+d+1+\sum_{j=1}^{s-1} d_j+s-1,v+d+1}\right). \tag{9}$$

The element at the z-th position in the C_{v+d+2}-th row of $(g_1',\cdots,g_s')$ is $a_{g_1+d+1+\sum_{j=1}^{z-1} d_j+z-1,v+d+1}$, $(z=1,\cdots,s)$. From Equation (2), we can deduce that $a_{g_1+\sum_{j=1}^{z-1} d_j+z-1,v} = a_{g_1+d+1+\sum_{j=1}^{z-1} d_j+z-1,v+d+1}$. Since $a_{g_1+\sum_{j=1}^{z-1} d_j+z-1,v}$ is the element in the $(v+1)$-th row and z-th column of $(g_1,\cdots,g_s)$, it follows that the element in the C_{v+d+2}-th row and z-th column of $(g_1',\cdots,g_s')$ is equal to the element in the $(v+1)$-th row and z-th column of $(g_1,\cdots,g_s)$.

By considering all values of z, it follows that the C_{v+d+2}-th row of $(g_1',\cdots,g_s')$ is identical to the $(v+1)$-th row of $(g_1,\cdots,g_s)$. Since the last row of the Plackett-Burman design consists entirely of zeros or ones, by traversing v, it can be concluded that $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ have the same $\alpha(\omega)$. □

Lemma 2. *Select two s-column submatrices, $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$, from an $N\times m$ Plackett-Burman design. Let the distance vectors of the two s-column submatrices be denoted as $(p_1,\cdots,p_s)$ and $(p_1',\cdots,p_s')$, respectively. If both $(p_1,\cdots,p_s)$ and $(p_1',\cdots,p_s')$ belong to the set D_s, then $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ have the same $\alpha(\omega)$.*

Proof. Select two s-column submatrices, $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$, from an $N\times m$ Plackett-Burman design. Let the distance vectors of the two s-column submatrices be denoted as $(p_1,\cdots,p_s)$ and $(p_1',\cdots,p_s')$, respectively. Let both $(p_1,\cdots,p_s)$ and $(p_1',\cdots,p_s')$ belong to D_s. Similar to the discussion in Lemma 1, $(g_1,\cdots,g_s)$ can be further expressed as $g_i = g_i(i=1)$, $g_i = g_1 + \sum_{j=1}^{i-1} p_j + i - 1(i=2,\cdots,s)$. $g_1',\cdots,g_s'$ can be further expressed as $g_i' = g_i'(i=1)$, $g_i' = g_1' + \sum_{j=1}^{i-1} p_j + i - 1(i=2,\cdots,s)$.

(i) $(p_1,\cdots,p_s) = (p_1',\cdots,p_s')$. In this case, $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ have identical distance vectors. By Lemma 1, it follows that, at this point, $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ have the same $\alpha(\omega)$.

(ii) $(p_1,\cdots,p_s) \neq (p_1',\cdots,p_s')$. Let $(p_1,\cdots,p_s) = (d_1,\cdots,d_s)$ and

$(p_1',\cdots,p_s')=(d_2,\cdots,d_s,d_1)$. Discuss the values of $\alpha(\omega)$ for $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ under these two distance vectors, respectively. As stated in Lemma 1, we only need to consider the case where both g_1 and g_1' are equal to 1. Now, let both g_1 and g_1' be equal to 1, so $g_2,\cdots,g_s$ can be further represented as $g_i=\sum_{j=1}^{i-1}d_j+i\,(i=2,\cdots,s)$. $g_2',\cdots,g_s'$ can be further represented as $g_i'=\sum_{j=2}^{i}d_j+i\,(i=2,\cdots,s)$. The first row of $(g_1,\cdots,g_s)$ is

$$\left(a_1,a_{2+d_1},\cdots,a_{s+\sum_{j=1}^{s-1}d_j}\right). \tag{10}$$

Based on the Plackett-Burman design generation method and the definition of Equation (2), it can be concluded that the $(v+1)$-th row of $(g_1,\cdots,g_s)$ can be represented as

$$\left(a_{1,v},a_{2+d_1,v},\cdots,a_{s+\sum_{j=1}^{s-1}d_j,v}\right). \tag{11}$$

The first row of $(g_1',\cdots,g_s')$ is

$$\left(a_1,a_{2+d_2},\cdots,a_{s+\sum_{j=2}^{s}d_j}\right). \tag{12}$$

Based on the generation method of Plackett-Burman design and the definitions of Equations (2) and (3), we can deduce that the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ can be represented as

$$\left(a_{1,v+s+\sum_{j=2}^{s}d_j-1},a_{2+d_2,v+s+\sum_{j=2}^{s}d_j-1},\cdots,a_{s+\sum_{j=2}^{s}d_j,v+s+\sum_{j=2}^{s}d_j-1}\right). \tag{13}$$

The $(c-1)$-th element in the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ is $a_{c-1+\sum_{j=2}^{c-1}d_j,v+s+\sum_{j=2}^{s}d_j}$, $(3\le c\le s)$. From Equation (2), we can deduce that $a_{c-1+\sum_{j=2}^{c-1}d_j,v+s+\sum_{j=2}^{s}d_j}=a_{c+\sum_{j=1}^{c-1}d_j,v}$. That is, the $(c-1)$-th element in the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ is equal to the c-th element in the $(v+1)$-th row of $(g_1,\cdots,g_s)$. In particular, the s-th element in the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ is equal to $a_{s+\sum_{j=2}^{s}d_j,v+s+\sum_{j=2}^{s}d_j-1}$. From Equation (2), we can deduce that $a_{s+\sum_{j=2}^{s}d_j,v+s+\sum_{j=2}^{s}d_j-1}=a_{1,v}$. That is, the s-th element in the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ is equal to the first element in the $v+1$-th row of $(g_1,\cdots,g_s)$. The first element in the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ is denoted as $a_{1,v+s+\sum_{j=2}^{s}d_j-1}$. From Equation (2), we can deduce that $a_{1,v+s+\sum_{j=2}^{s}d_j-1}=a_{2+d_1,v}$. That is, the first element in the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$ is equal to the second element in the $(v+1)$-th row of $(g_1,\cdots,g_s)$.

By iterating over all possible values of v, it can be observed that the $(v+1)$-th row of $(g_1,\cdots,g_s)$ corresponds to the $C_{v+s+\sum_{j=2}^{s}d_j}$-th row of $(g_1',\cdots,g_s')$, with both rows containing the same number of 0 s and 1 s. Therefore, when the distance vectors of $(g_1,\cdots,g_s)$ and $(g_1',\cdots,g_s')$ are equal to $(d_1,\cdots,d_s)$ and $(d_2,\cdots,d_s,d_1)$ respectively, the $\alpha(\omega)$ of $(g_1,\cdots,g_s)$ is identical to that of $(g_1',\cdots,g_s')$. From the scenario where $(p_1,\cdots,p_s)=(d_2,\cdots,d_s,d_1)$ and $(p_1',\cdots,p_s')=(d_3,\cdots,d_s,d_1,d_2)$, extending all the way to $(p_1,\cdots,p_s)=(d_s,d_1,\cdots,d_{s-1})$ and $(p_1',\cdots,p_s')=(d_1,\cdots,d_s)$, the same conclusion holds as in the case where $(p_1,\cdots,p_s)=(d_1,\cdots,d_s)$, $(p_1',\cdots,p_s')=(d_2,\cdots,d_s,d_1)$.

By the principle of transitivity, if both $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$ are derived from D_s, then the $\alpha(\omega)$ of $(g_1,\cdots,g_s)$ is identical to that of $(g'_1,\cdots,g'_s)$. □

Theorem 1. *Select two s-column submatrices, $(g_1,\cdots,g_s)$ and $(g'_1,\cdots,g'_s)$, from an $N\times m$ Plackett-Burman design. Let the distance vectors of the two s-column submatrices be denoted as $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$, respectively. If both $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$ belong to the set D_s, then, the s-column sub-design constructed from $(g_1,\cdots,g_s)$ and the s-column sub-design constructed from $(g'_1,\cdots,g'_s)$ possess the same K-aberration sequence.*

Proof. Select two s-column submatrices, $(g_1,\cdots,g_s)$ and $(g'_1,\cdots,g'_s)$, from an $N\times m$ Plackett-Burman design. Let the distance vectors of the two s-column submatrices be denoted as $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$, respectively. Select any u columns from $g_1,\cdots,g_s$. Let the selected u columns be denoted as $h_1,\cdots,h_u$, and it holds that $h_1<\cdots<h_u$. Let the e-th column $h_e\,(e=1,\cdots,u)$ among the u columns $h_1,\cdots,h_u$ be the $c_e, c_e\in\{1,\cdots,s\}$-th column among the s columns $g_1,\ldots,g_s$. Let $(n_1,\cdots,n_u)$ denote the distance vector of $(h_1,\cdots,h_u)$. For the selected columns $h_1,\cdots,h_u$, we choose a corresponding set of u specific columns $h'_1,\cdots,h'_u$ from $g'_1,\cdots,g'_s$. Let $h'_e\,(e=1,\cdots,u)$ be the $c'_e, c'_e\in\{1,\cdots,s\}$-th column among the s columns $g'_1,\cdots,g'_s$. Let $(n'_1,\cdots,n'_u)$ denote the distance vector of $(h'_1,\cdots,h'_u)$.

(i) $(p_1,\cdots,p_s)=(p'_1,\cdots,p'_s)$.

Without loss of generality, let $(p_1,\cdots,p_s)=(p'_1,\cdots,p'_s)=(d_1,\cdots,d_s)$. When $1\le e\le u-1$, it follows that $n_e=\sum_{f=c_e}^{c_{e+1}-1}d_f+c_{e+1}-c_e-1$. In particular, $n_u=\sum_{f=c_u}^{s}d_f+\sum_{r=1}^{c_1-1}d_r+s-c_u+c_1-1$. By setting $c'_e=c_e$, we obtain $(n'_1,\cdots,n'_u)=(n_1,\cdots,n_u)$. From Lemma 1, it follows that the $\alpha(\omega)$ of $(h_1,\cdots,h_u)$ and $(h'_1,\cdots,h'_u)$ is identical under this condition. Therefore, when $(p_1,\cdots,p_s)=(p'_1,\cdots,p'_s)$, for every $(h_1,\cdots,h_u)$, there exists a corresponding $(h'_1,\cdots,h'_u)$ such that the $\alpha(\omega)$ of $(h_1,\cdots,h_u)$ and $(h'_1,\cdots,h'_u)$ is identical.

(ii) $(p_1,\cdots,p_s)\ne(p'_1,\cdots,p'_s)$.

Let $(p_1,\cdots,p_s)=(d_1,\cdots,d_s)$ and $(p'_1,\cdots,p'_s)=(d_2,\cdots,d_s,d_1)$. When $1\le e\le u-1$, it follows that $n_e=\sum_{f=c_e}^{c_{e+1}-1}d_f+c_{e+1}-c_e-1$. In particular, $n_u=\sum_{f=c_u}^{s}d_f+\sum_{r=1}^{c_1-1}d_r+s-c_u+c_1-1$. When $1\le e\le u-2$, let $c'_e=c_{e+1}-1$ and let $n'_e=\sum_{f=c_{e+1}}^{c_{e+2}-1}d_f+c_{e+2}-c_{e+1}$. When $e=u$ and $c_1>1$, let $c'_u=c_1-1$ and $n'_u=\sum_{f=c_1}^{c_2-1}d_f+c_2-c_1-1$. When $e=u$ and $c_1=1$, let $c'_u=s$ and $n'_u=\sum_{f=1}^{c_2-1}d_f+c_2-2$. We can discover that $(h'_1,\cdots,h'_u)$ precisely corresponds to a u-column submatrix of $(g'_1,\cdots,g'_s)$, with the distance vector $(n'_1,\cdots,n'_u)$ of $(h'_1,\cdots,h'_u)$ being equal to $(n_2,\cdots,n_u,n_1)$. From Lemma 2, it follows that the $\alpha(\omega)$ of $(h_1,\cdots,h_u)$ and $(h'_1,\cdots,h'_u)$ is identical under this condition. Therefore, when $(p_1,\cdots,p_s)=(d_1,\cdots,d_s)$ and $(p'_1,\cdots,p'_s)=(d_2,\cdots,d_s,d_1)$, for every $(h_1,\cdots,h_u)$, there exists a corresponding $(h'_1,\cdots,h'_u)$ such that the $\alpha(\omega)$ of

$(h_1,\cdots,h_u)$ and $(h'_1,\cdots,h'_u)$ is identical.

Let the t-th order aliasing of the sub-design formed by $(g_1,\cdots,g_s)$ and the sub-design formed by $(g'_1,\cdots,g'_s)$ be denoted as K_t and K'_t, respectively. Furthermore, let $T_1 = T_1$ and $T_2 = T_2$ for K_t, and $T_1 = T'_1$ and $T_2 = T'_2$ for K'_t. Through a discussion of the two distinct scenarios involving $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$, we can deduce that for any t columns $h_1,\cdots,h_t$ selected from $g_1,\cdots,g_s$, there exist uniquely corresponding t columns $h'_1,\cdots,h'_t$ in $g'_1,\cdots,g'_s$, such that $(h_1,\cdots,h_t)$ and $(h'_1,\cdots,h'_t)$ have the same $\alpha(\omega)$. That is, $T_1 = T_1^{'}$.

For any $t+1$ columns $h_1,\cdots,h_{t+1}$ of $g_1,\cdots,g_s$, there exist unique $t+1$ columns $h'_1,\cdots,h'_{t+1}$ of $g'_1,\cdots,g'_s$ such that $(h_1,\cdots,h_{t+1})$ and $(h'_1,\cdots,h'_{t+1})$ have the same $\alpha(\omega)$. For cases (i) and (ii) concerning $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$, the distance vector $(n_1,\cdots,n_{t+1})$ of $(h_1,\cdots,h_{t+1})$ and the distance vector $(n'_1,\cdots,n'_{t+1})$ of $(h'_1,\cdots,h'_{t+1})$ respectively satisfy the following two scenarios.

(I) $(n_1,\cdots,n_{t+1}) = (n'_1,\cdots,n'_{t+1})$.

(II) $(n_1,\cdots,n_{t+1}) = (n'_{t+1},n'_1,\cdots,n'_t)$.

Similarly to the discussion on $(g_1,\cdots,g_s)$ and $(g'_1,\cdots,g'_s)$. For any t-column submatrix of $(h_1,\cdots,h_{t+1})$, there exists a unique t-column submatrix of $(h'_1,\cdots,h'_{t+1})$ such that the value of $\alpha(\omega)$ is identical for these two t-column submatrices, which consequently leads to $T_2 = T'_2$. Given that $T_1 = T'_1$ and $T_2 = T'_2$, it ultimately follows that $K_t = K'_t$. Therefore, when $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$ satisfy conditions (i) and (ii), the s-column-subdesign formed by $(g_1,\cdots,g_s)$ and the s-column-subdesign formed by $(p'_1,\cdots,p'_s)$ have the same K-aberration sequence.

In the case where $(p_1,\cdots,p_s) \neq (p'_1,\cdots,p'_s)$, the same conclusion holds when $(p_1,\cdots,p_s) = (d_2,\cdots,d_s,d_1)$, $(p'_1,\cdots,p'_s) = (d_3,\cdots,d_s,d_1,d_2)$, $\cdots$, $(p_1,\cdots,p_s) = (d_s,d_1,\cdots,d_{s-1})$ and $(p'_1,\cdots,p'_s) = (d_1,\cdots,d_s)$. From transitivity, it follows that if both $(p_1,\cdots,p_s)$ and $(p'_1,\cdots,p'_s)$ belong to D_s, then the s-column-subdesign formed by $(g_1,\cdots,g_s)$ and the s-column-subdesign formed by $(g'_1,\cdots,g'_s)$ have the same K-aberration sequence. □

Theorem 2. *The conclusion provided by Theorem 1 improves the search efficiency for the optimal baseline sub-design of the Plackett-Burman design by a factor of* $m-1$*, independent of the number of columns in the selected sub-design.*

Proof. Select s columns, where $s < m$, from the m columns of the Plackett-Burman design, denoted as $g_1,\cdots,g_s, 1 \le g_1 < \cdots < g_s \le m$. Let $g_1 = 1$, then the distance vector of $(g_1,\cdots,g_s)$ can be represented as $(g_2-2, g_3-g_2-1,\cdots,g_s-g_{s-1}-1, m-g_s)$. Thus, according to Theorem 1, the s-column sub-designs formed by $(g_1+i,\cdots,g_s+i), i = \{1,\cdots,m-r\}$ and $(g_1,\cdots,g_s)$ have the same K-aberration sequence.

Let the distance vector of $(g'_1,\cdots,g'_s), 1 \le g'_1 < \cdots < g'_s \le m$ be denoted as $(g_3-g_2-1, g_4-g_3-1,\cdots,m-g_s, g_2-2)$. Let $g'_1 = 1$. Thus, by Theorem 1, the s-column sub-designs formed by $(g'_1,\cdots,g'_s)$ and $(g_1,\cdots,g_s)$ have the same K-aberration sequence. The s-column sub-designs formed by

$(g_1'+i,\cdots,g_s'+i), i=\{1,\cdots,g_2-2\}$ and $(g_1',\cdots,g_s')$ have the same K-aberration sequence.

Let the distance vector of $(g_1',\cdots,g_s')$ be denoted as $(g_4-g_3-1, g_5-g_4-1,\cdots,g_2-2, g_3-g_2-1)$. Let $g_1'=1$. Thus, by Theorem 1, the s-column sub-designs formed by $(g_1',\cdots,g_s')$ and $(g_1,\cdots,g_s)$ have the same K-aberration sequence. The s-column sub-designs formed by $(g_1'+i,\cdots,g_s'+i), i=\{1,\cdots,g_3-g_2-1\}$ and $(g_1',\cdots,g_s')$ have the same K-aberration sequence.

Similarly, Let the distance vector of $(g_1',\cdots,g_s')$ be denoted as $(m-g_s, g_2-2,\cdots,g_{s-1}-g_{s-2}-1, g_s-g_{s-1}-1)$. Let $g_1'=1$. Thus, by Theorem 1, the s-column sub-designs formed by $(g_1',\cdots,g_s')$ and $(g_1,\cdots,g_s)$ have the same K-aberration sequence. The s-column sub-designs formed by $(g_1'+i,\cdots,g_s'+i), i=\{1,\cdots,g_s-g_{s-1}-1\}$ and $(g_1',\cdots,g_s')$ have the same K-aberration sequence.

Therefore, each time we select an s-column sub-design, there will be $(m-s)+(1+g_2-2)+(1+g_3-g_2-1)+\cdots+(1+g_s-g_{s-1}-1)=m-1$ distinct s-column sub-designs that have the same K-aberration sequence as the selected one. Next, we select $s, s<m$ columns from the m columns of the Plackett-Burman design, denoted as $g_1'',\cdots,g_s''$. If the distance vector of $g_1'',\cdots,g_s''$ does not belong to D_s, then the sub-design formed by $(g_1'',\cdots,g_s'')$ will have the same K-aberration sequence as $m-1$ other s-column sub-designs. In summary, we conclude that the result provided by Theorem 1 improves the search efficiency for the optimal baseline sub-designs of the Plackett-Burman design by a factor of $m-1$, independent of the number of columns in the selected sub-design. □

4. Application

Wu and Hamada (2000) [4] presented a Plackett-Burman design with 24 runs and 23 columns in their work. We will identify the minimal K-aberration design among all the 24-run, 7-column sub-designs based on the theory presented in this paper. According to our theory, after excluding all sub-designs with identical K-aberration sequences, we have obtained a total of 10,659 sub-designs. By calculating the Z_2 for these sub-designs, we identified two sub-designs with the smallest Z_2, which are determined by the columns 1,234,567 and 1,234,579, respectively. By calculating Z_3 for these two sub-designs, we found that the sub-design determined by columns 1,234,579 has the smallest Z_3. Therefore, the minimal aberration design among all the 24-run, 7-column sub-designs of this 24-run, 23-column Plackett-Burman design is the sub-design determined by columns 1,234,579. If we were to evaluate the Z-values for all possible 24-run, 7-column sub-designs, we would need to assess C_{23}^7, which equals 245,157 designs. However, the theory presented in this paper significantly reduces the search space.

We implemented the results presented in the application using R software. The R code is provided in Appendix A. The specific results are presented in **Table 2**.

Table 2. The two sub-designs, Z_2 and Z_3.

Z-value	1,234,567	1,234,579
Z_2	7.5	7.5
Z_3	32.01042	31.34375

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Deng, L.-Y. and Tang, B. (1999) Generalized Resolution and Minimum Aberration Criteria for Plackett-Burman and Other Nonregular Factorial Designs. *Statistica Sinica*, **9**, 1071-1082.

[2] Mukerjee, R. and Tang, B. (2011) Optimal Fractions of Two-Level Factorials under a Baseline Parameterization. *Biometrika*, **99**, 71-84. https://doi.org/10.1093/biomet/asr071

[3] Mukerjee, R. and Tang, B. (2016) Optimal Two-Level Regular Designs under Baseline Parametrization via Cosets and Minimum Moment Aberration. *Statistica Sinica*, **26**, 1001-1019. https://doi.org/10.5705/ss.202015.0214

[4] Wu, C.F.J. and Hamada, M.S. (2011) Experiments: Planning, Analysis, and Optimization. John Wiley & Sons.

Appendix

A Implementation in R Code

 # Generate a 24-row, 23-column Plackett-Burman design matrix p, and generate the subsequent matrix p1 to be used later.

```
first_row <- c(1,1,1,1,1,-1,1,-1,1,1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,-1,-1)

matrix_rows <- list()

matrix_rows[[1]] <- first_row

for (i in 2:23) {

    prev_row <- matrix_rows[[i-1]]

    shifted_row <- c(prev_row[length(prev_row)], prev_row[-length(prev_row)])

    matrix_rows[[i]] <- shifted_row

}

matrix_rows[[24]] <- rep(-1, length(first_row))


p <- do.call(rbind, matrix_rows)

p1 <- p

p1[p1 == 1] <- 0

p1[p1 == -1] <-1

#Used to find the column set of non-equivalent distance vectors.

install.packages("combinat")      #The package does not need to be installed if you already have it.

library(combinat)

all_combinations <- combn(2:23, 6)

S <- list()

compute_vector <- function(nums) {
    sorted_nums <- sort(c(1, nums))
    a <- sorted_nums[2]
    b <- sorted_nums[3]
    c <- sorted_nums[4]
    d <- sorted_nums[5]
    e <- sorted_nums[6]
    f <- sorted_nums[7]

    v1 <- c(a - 2, b - a - 1, c - b - 1, d - c - 1, e - d - 1, f - e - 1, 23 - f)
    v2 <- c(v1[2], v1[3], v1[4], v1[5], v1[6], v1[7], v1[1])
    v3 <- c(v2[2], v2[3], v2[4], v2[5], v2[6], v2[7], v2[1])
    v4 <- c(v3[2], v3[3], v3[4], v3[5], v3[6], v3[7], v3[1])
```

```
        v5 <- c(v4[2], v4[3], v4[4], v4[5], v4[6], v4[7], v4[1])
        v6 <- c(v5[2], v5[3], v5[4], v5[5], v5[6], v5[7], v5[1])
        v7 <- c(v6[2], v6[3], v6[4], v6[5], v6[6], v6[7], v6[1])

        list(v1 = v1, v2 = v2, v3 = v3, v4 = v4, v5 = v5, v6 = v6, v7 = v7)
    }

    is_new_vector <- function(new_vec, existing_vectors) {
        for (vec in existing_vectors) {
            if (identical(new_vec, vec)) {
                return(FALSE)
            }
        }
        return(TRUE)
    }

    existing_vectors <- list()
    for (i in 1:ncol(all_combinations)) {
        selected_nums <- all_combinations[, i]
        full_sequence <- sort(c(1, selected_nums))
        digit_string <- paste(full_sequence, collapse = "")
        vectors <- compute_vector(selected_nums)
        new_sequence <- TRUE
        for (j in 1:7) {
            vec_j <- vectors[[j]]
            if (!is_new_vector(vec_j, existing_vectors)) {
                new_sequence <- FALSE
                break
            }
        }
        if (new_sequence) {
            S <- c(S, digit_string)
            existing_vectors <- c(existing_vectors, list(vectors$v1, vectors$v2, vectors$v3, vectors$v4, vec-
tors$v5, vectors$v6, vectors$v7))
        }
    }

    S <- unique(unlist(S))
    print(S)

    # Find the minimal sub-design for Z2 .

    results <- list()

    for (digit_str in S) {

        digits <- as.numeric(strsplit(digit_str, "")[[1]])

        mat <- p[, digits]

        mat1 <- p1[, digits]

        J=mat1%*%t(mat1)
        J1=J^2
        J2=mat%*%t(mat)
        J3=J1%*%J2
```

```
        Z2=(1/24^2)*sum(diag(J3))

        results[[digit_str]] <- list(
        Z2 = Z2
        )
}

if (length(results) == 0)
{
        cat("No results are available for processing.\n")
} else {

        Z2_values <- unlist(lapply(results, function(x) x$Z2))

        min_Z2 <- min(Z2_values)

        min_strings <- names(results)[which(Z2_values == min_Z2)]

        cat("The minimum Z2 value is:", min_Z2, "\n")

        cat("The corresponding string is:", paste(min_strings, collapse = ", "), "\n")
}

# Find the minimal Z3 for two sub-designs.

S1<-c("1234567","1234579")

results <- list()

for (digit_str in S1) {

        digits <- as.numeric(strsplit(digit_str, "")[[1]])

        mat <- p[, digits]

        mat1 <- p1[, digits]

        J=mat1%*%t(mat1)
        J1=J^3
        J2=mat%*%t(mat)
        J3=J1%*%J2
        Z3=(1/24^2)*sum(diag(J3))
        Z3

        results[[digit_str]] <- list(
        Z3 = Z3
        )
}

if (length(results) == 0)
{
        cat("No results are available for processing.\n")
} else {

        Z3_values <- unlist(lapply(results, function(x) x$Z3))
```

```
min_Z3 <- min(Z3_values)

min_strings <- names(results)[which(Z3_values == min_Z3)]

cat("The minimum Z3 value is:", min_Z3, "\n")

cat("The corresponding string is:", paste(min_strings, collapse = ", "), "\n")
```