Characterization and Optimal Designs for Choice Experiments

Feng-Shun Chai\textsuperscript{1}, Ashish, Das\textsuperscript{2} and Soumen Manna\textsuperscript{2}

\textsuperscript{1}Academia Sinica, Taipei, Taiwan
\textsuperscript{2}Indian Institute of Technology Bombay, Mumbai, India

November 2014

Indian Institute of Technology Bombay
Powai, Mumbai-400 076, India
Characterization and Optimal Designs for Choice Experiments

BY FENG-SHUN CHAI
Academia Sinica, Taipei, Taiwan
ASHISH DAS and SOUMEN MANNA
Indian Institute of Technology Bombay, Mumbai, India

SUMMARY

Street and Burgess (2007) present a comprehensive exposition of designs for choice experiments till then. The choice design involves $n$ attributes (factors) with $i$-th attribute at $l_i$ level, and all choice sets are of size $m$. A choice design comprises $N$ such choice sets. Recently, Demirkale, Donovan and Street (2013) considered the setup of symmetric factorials ($l_i = l$) and obtained $D$-optimal choice designs under main effects model. They provide some sufficient conditions for a designs to be $D$-optimal. In this paper, we first derive a slightly modified Information matrix of a choice design for estimating the factorial effects of a $l_1 \times l_2 \times \cdots \times l_n$ choice experiment. It is seen that such a modification gives the Information matrix the desired additive property and thus, overcomes the existing shortcoming of situations where, with addition of a choice set the information content of the design decreases. While comparing designs with different $N$, we see that one needs to work with the modified information matrix. For a $2^n$ choice experiment, under the main effects model, we give a simple necessary and sufficient condition for the Information matrix to be diagonal. Furthermore, we characterize the structure of the choice sets which gives maximum trace of the Information matrix. Our characterization of such an Information matrix facilitates construction of universally optimal designs with minimal number of choice sets and gives more flexibility for choosing $m$. Finally, we provide universally optimal choice designs for estimating main effects, which are optimal in the class of all designs with given $N$, $n$ and $m$.

Key words and phrases: choice sets; choice design; factorial design; main effects; Hadamard matrix; orthogonal array.

1 INTRODUCTION

Discrete choice experiments are widely used in various areas including marketing, transport, environmental resource economics and public welfare analysis. A choice experiment consists of a number of choice sets, each containing several options (alternatives, profiles or treatment combinations). Respondents are shown each choice set in turn and are asked which option they prefer, as per their perceived utility, in each of the choice sets presented. Each option in a choice set is described by a set of attributes (factors), each with some number of levels. We assume that there are no repeated options in a choice set. We describe the options which are being compared, by $n$ attributes with $i$-th attribute at $l_i$ level ($l_i \geq 2$), and that all the choice sets in a $l_1 \times l_2 \times \cdots \times l_n$ choice experiment have $m$ options. It is ensured that respondents choose one of the options in each choice set (termed forced choice experiment in the literature). A choice design is a collection of choice sets employed in a choice experiment. A choice design comprises $N$ such choice sets.
Street and Burgess (2007) present a comprehensive exposition of designs for choice experiments under multinomial logit (MNL) model. MNL model specifies the probability that an individual will choose one of the $m$ alternatives, say $s_i$, from a choice set $S$ (say). The probability is given as the exponential of the utility of that alternative $s_i$, divided by the sum of all the exponentiated utilities. Mathematically,

$$P(s_i|S) = \frac{e^{V_i}}{\sum_{j=1}^{m} e^{V_j}}, \quad (1.1)$$

where $V_i$ is the utility measure represented by the treatment combination effect for a $l_1 \times l_2 \times \cdots \times l_n$ factorial. For more detailed discussion on MNL model and choice experiments, see Train (2009) and Street and Burgess (2007).

Recently, Demirkale, Donovan and Street (2013) considered the setup of symmetric factorials ($l_i = l$) and obtained $D$-optimal choice designs under main effects model. They provide some sufficient conditions for a designs to be $D$-optimal.

In this paper, we first derive a slightly modified Information matrix of a choice design for estimating the factorial effects. It is seen that such a modification gives the Information matrix the desired additive property and thus, overcomes the existing shortcoming of situations where, with addition of a choice set the information content of the design decreases. While comparing designs with different $N$, we see that one needs to work with the modified information matrix. For a $2^n$ choice experiment, under the main effects model, we give a simple necessary and sufficient condition for the Information matrix to be diagonal. Furthermore, we characterize the structure of the choice sets which gives maximum trace of the Information matrix. Our characterization of such an Information matrix facilitates construction of universally optimal designs with minimal number of choice sets and gives more flexibility for choosing $m$. For $m$ odd, optimal designs are achieved having relatively lesser number of choice sets than those obtained in Demirkale, Donovan and Street (2013). Finally, we provide universally optimal choice designs for estimating main effects, which are optimal in the class of all designs with given $N$, $n$ and $m$.

## 2 Information matrix

In choice experiment we deal with multiple independent populations which have common parameters. In a choice experiment, each choice set represent a different population. We call this set of populations as associated populations. When sampling from such associated populations, Bradley and Gart (1962) have presented related assumptions and asymptotic properties of the ML estimators. Under these assumptions, EI-Helbawy and Bradley (1978) have derived large sample results for paired choice experiments when each choice item is coming from a factorial setup. Later Street and Burgess (2007) generalized the setup for choice set size $m$ and obtained the Information matrix on similar lines. It is seen that their Information matrix is derived using the averaging principle leading to situations where adding more choice sets to a design leads to information matrix with less information content than the information matrix of the original design. In what follows, we adopt an approach different from Bradley and Gart (1962) and EI-Helbawy and Bradley (1978). We derive a slightly modified Information matrix of a choice design for estimating the factorial effects. Such a modification gives the Information matrix the
desired additive property. Our approach addresses a possible lacuna in the current non-additive form of the information matrix. Let \( X_i \) be a random variable over the region \( R_i \), independent of \( \theta = (\theta_1, \theta_2, \ldots, \theta_k)' \), an unknown parameter vector lying on a \( k \)-dimensional interval \( \Omega \). Furthermore let \( f_i(x_i, \theta), i = 1, 2, \ldots, n^* \), be the pdf or pmf of \( X_i \) from \( n^* \) different associated populations. It is not necessary that each \( f_i \) depends on all \( \theta_1, \theta_2, \ldots, \theta_k \). Let \( X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i}) \) be a random sample of size \( n_i \), from \( f_i \). Then the likelihood function corresponding to it is

\[
\mathcal{L}_i = \prod_{j=1}^{n_i} f_i(X_{ij}; \theta) = f_i(X_i; \theta). \tag{2.1}
\]

According to Fisher (for more details see Rao (1973)), the Information contained in the sample \( X_i \) is denoted by the information matrix \( \mathcal{I}_i = (\mathcal{I}_{i(r)}(\theta))_{k \times k} \), where

\[
\mathcal{I}_{i(r)}(\theta) = \int_{R_i} \frac{\partial \ln f_i}{\partial \theta_r} \frac{\partial \ln f_i}{\partial \theta_s} f_i dx_i = E \left( \frac{\partial \ln f_i}{\partial \theta_r} \frac{\partial \ln f_i}{\partial \theta_s} \right) \tag{2.2}
\]

is non-negative definite.

Now if we take random sample \( X_i \) of size \( n_i \), from each of the \( n^* \) associated populations \( f_i \), then the likelihood function of \( \theta \) for all the samples \( X_1, X_2, \ldots, X_n^* \) can be written as

\[
\mathcal{L} = \prod_{i=1}^{n^*} f_i(X_i; \theta). \tag{2.3}
\]

We define the information for \( \theta \) contained in all the samples \( X_1, X_2, \ldots, X_n^* \); from \( n^* \) associated populations by the Information matrix \( \mathcal{I} = (\mathcal{I}_{rs}(\theta))_{k \times k} \) with

\[
\mathcal{I}_{rs}(\theta) = \sum_{i=1}^{n^*} \mathcal{I}_{i(rs)}(\theta), \tag{2.4}
\]

which is also non-negative definite.

We now derive the expression for the information matrix of a choice design with choice set size \( m \). Consider a \( l_1 \times l_2 \times \cdots \times l_n \) choice experiment with \( L = \prod_{i=1}^{n} l_i \). Let the \( L \) treatments in the choice experiment be denoted by \( T_1, T_2, \ldots, T_L \), where, \( T_i = (i_1 i_2 \ldots i_h \ldots i_k \ldots i_n), i_r = 0, 1, \ldots, l_r - 1; r = 1, 2, \ldots, n; \) is a typical treatment combination. In order to ensure that \( T_i \)’s are arranged in a lexicographic order, let

\[
i = i_1 \prod_{i=2}^{n} l_i + i_2 \prod_{i=3}^{n} l_i + \cdots + i_{n-1} l_n + i_n + 1. \]

In other words, \( i \) is the lexicographic order number of the treatment combination \( T_i \).

Let \( \pi_i = e^{V_i} \) be the parameter associated to the treatment \( T_i \). Our aim is to find the information matrix of certain parametric contrasts involving the parameters \( V_i, i = 1, 2, \ldots, L \). A choice set of size \( m \) is denoted by \( S_m = (T_{j_1}, T_{j_2}, \ldots, T_{j_m}) \), where no two \( j_i \)’s are equal. For a choice set \( S_m \), we represent \( (T_{j_i} \succ \{T_{j_1}, T_{j_2}, \ldots, T_{j_m}\}) \) to mean \( T_{j_i} \) is chosen over \( T_{j_1}, \ldots, T_{j_{i-1}}, T_{j_{i+1}}, \ldots, T_{j_m} \), by the respondent.

Consider an experiment in which there are \( N \) choice sets of size \( m \). We define a set \( A_t \) as

\[
A_t = \{ (j_1, j_2, \ldots, j_m) : (T_{j_1}, T_{j_2}, \ldots, T_{j_m}) \text{ is a choice set in the experiment} \}.  
\]
Also, from (2.7), we get it follows from (2.2) and (2.4) that be a

\[ N_{j_1j_2...j_m} = \begin{cases} 1 & \text{if } (j_1, j_2, \ldots, j_m) \in A_t \\ 0 & \text{if } (j_1, j_2, \ldots, j_m) \notin A_t. \end{cases} \]

Therefore,

\[ N = \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1j_2...j_m}. \tag{2.5} \]

For any \((j_1, j_2, \ldots, j_m) \in A_t\), we can write from (1.1) that

\[ P(T_{j_i} > \{T_{j_1}, T_{j_2}, \ldots, T_{j_m}\}) = \frac{\pi_{j_i}}{\sum_{i=1}^m \pi_{j_i}} \tag{2.6} \]

for \(i = 1, 2, \ldots, m\). Let, \(\pi = (\pi_1, \pi_2, \ldots, \pi_m)'\). Here each choice set \((T_{j_1}, T_{j_2}, \ldots, T_{j_m})\) represent an associate population with parameters \(\pi_{j_1}, \pi_{j_2}, \ldots, \pi_{j_m}\). Therefore, the pmf \(f_{j_1j_2...j_m}\) of the multinomial random variable \((x_{j_1}, x_{j_2}, \ldots, x_{j_m})\) corresponding to the choice set \((T_{j_1}, T_{j_2}, \ldots, T_{j_m})\), is

\[ f_{j_1j_2...j_m}(x_{j_1}, x_{j_2}, \ldots, x_{j_m}; \pi) = \prod_{i=1}^m \left( \frac{\pi_{j_i}}{\sum_{i=1}^m \pi_{j_i}} \right)^{x_{j_i}}, \tag{2.7} \]

where for \(i = 1, 2, \ldots, m\) we define \(x_{j_i} = 1\) if \((T_{j_i} > \{T_{j_1}, T_{j_2}, \ldots, T_{j_m}\})\); and 0 otherwise. To be more precise \(x_{j_i}\) can be written as \(x_{j_i}^{(j_1j_2...j_m)}\), but for notational convenience we retain the notation \(x_{j_i}\) corresponding to the choice set \((T_{j_1}, T_{j_2}, \ldots, T_{j_m})\).

Note that \(\sum_{i=1}^m x_{j_i} = 1\). Therefore from equation (2.3), the likelihood function can be written as

\[ \mathcal{L} = \prod_{j_1 < j_2 < \ldots < j_m} \{f_{j_1j_2...j_m}(x_{j_1}, x_{j_2}, \ldots, x_{j_m}; \pi)\}^{N_{j_1j_2...j_m}}. \tag{2.8} \]

Let \(V = (V_1, V_2, \ldots, V_L)'\) be the vector of treatment effects that the researcher can capture for a \(l_1 \times l_2 \times \cdots \times l_n\) choice experiment. Furthermore, let \(\Lambda = ((\lambda_{kl}))\) be a \(L \times L\) matrix representing the information matrix of \(V\). Then, since \(V_i = \ln \pi_i\), it follows from (2.2) and (2.4) that

\[ \lambda_{kl} = \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1j_2...j_m} E \left[ \frac{\partial \ln f_{j_1j_2...j_m}}{\partial V_k} \frac{\partial \ln f_{j_1j_2...j_m}}{\partial V_l} \right] \]

\[ = \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1j_2...j_m} \left[ \frac{\partial \ln f_{j_1j_2...j_m}}{\partial \pi_k} \frac{\partial \ln f_{j_1j_2...j_m}}{\partial \pi_l} \right] \pi_k \pi_l. \tag{2.9} \]

It is clear from (2.9) that if \((k, l)\) does not belong to any element of \(A_t\), then

\[ \lambda_{kl} = 0. \tag{2.10} \]

From (2.7) we note that \((x_{j_1}, x_{j_2}, \ldots, x_{j_m})\) is a multinomial random variable with parameters \(\sum_{i=1}^m \pi_{j_i}\), \(i = 1, 2, \ldots, m\) and \(E(x_{j_i}) = E(x_{j_i}^2) = \frac{\pi_{j_i}}{\sum_{i=1}^m \pi_{j_i}}; i = 1, 2, \ldots, m\). Also, from (2.7), we get

\[ \ln(f_{j_1j_2...j_m}(x_{j_1}, x_{j_2}, \ldots, x_{j_m}; \pi)) = \sum_{i=1}^m x_{j_i} \ln(\pi_{j_i}) - \ln \left( \sum_{i=1}^m \pi_{j_i} \right), \]
and therefore,

$$\frac{\partial \ln f_{j_1, j_2 \ldots j_m}}{\partial \pi_{j_i}} = \frac{x_{j_i}}{\pi_{j_i}} - \frac{1}{\sum_{i=1}^{m} \pi_{j_i}}; \quad i = 1, 2, \ldots, m.$$ 

If \((k, l)\) belongs to an element of \(A_t\), then from (2.9), both the partial derivatives are non-zero for the choice sets \((T_{j_1}, T_{j_2}, \ldots, T_{j_m})\), which contains \(T_k\) and \(T_l\) as options. Thus, without loss of generality, when \((k, l) = (j_1, j_2)\) such that \((j_1, j_2, \ldots, j_m) \in A_t\), we have \(\lambda_{kl} = \lambda_{j_1j_2}\) which is

$$\begin{align*}
= & \sum_{j_3 < j_4 < \cdots < j_m} \left[ \frac{x_{j_1}}{\pi_{j_1}} - \frac{1}{\sum_{i=1}^{m} \pi_{j_i}} \right] \left( \frac{x_{j_2}}{\pi_{j_2}} - \frac{1}{\sum_{i=1}^{m} \pi_{j_i}} \right) \pi_{j_1} \pi_{j_2} \\
= & \sum_{j_3 < j_4 < \cdots < j_m} \left[ \frac{x_{j_1} x_{j_2}}{\pi_{j_1} \pi_{j_2}} - \frac{x_{j_1}}{\pi_{j_1} \sum_{i=1}^{m} \pi_{j_i}} - \frac{x_{j_2}}{\pi_{j_2} \sum_{i=1}^{m} \pi_{j_i}} + \frac{1}{\sum_{i=1}^{m} \pi_{j_i}^2} \right] \pi_{j_1} \pi_{j_2} \\
= & - \sum_{j_3 < j_4 < \cdots < j_m} N_{j_1j_2 \ldots j_m} \frac{\pi_{j_1} \pi_{j_2}}{(\sum_{i=1}^{m} \pi_{j_i})^2}. \quad (2.11)
\end{align*}$$

Also, when \(k = l = j_1\), \(\lambda_{kk} = \lambda_{j_1j_1}\) which is

$$\begin{align*}
= & \sum_{j_2 < j_3 < \cdots < j_m} \left( \frac{x_{j_1}}{\pi_{j_1}} - \frac{1}{\sum_{i=1}^{m} \pi_{j_i}} \right) \pi_{j_1} \pi_{j_1} \\
= & \sum_{j_2 < j_3 < \cdots < j_m} \left[ \frac{x_{j_1}^2}{\pi_{j_1}^2} - \frac{2x_{j_1}}{\pi_{j_1} \sum_{i=1}^{m} \pi_{j_i}} + \frac{1}{(\sum_{i=1}^{m} \pi_{j_i})^2} \right] \pi_{j_1} \pi_{j_1} \\
= & \sum_{j_2 < j_3 < \cdots < j_m} \left[ \frac{1}{\pi_{j_1} \sum_{i=1}^{m} \pi_{j_i}} - \frac{2}{(\sum_{i=1}^{m} \pi_{j_i})^2} + \frac{1}{(\sum_{i=1}^{m} \pi_{j_i})^2} \right] \pi_{j_1} \pi_{j_1} \\
= & \sum_{j_2 < j_3 < \cdots < j_m} N_{j_1j_2 \ldots j_m} \frac{\pi_{j_1} \sum_{i=2}^{m} \pi_{j_i}}{(\sum_{i=1}^{m} \pi_{j_i})^2}. \quad (2.12)
\end{align*}$$

Therefore, in terms of \(\pi_i\)'s, \(\Lambda\) can be rewritten as

$$\begin{align*}
\lambda_{kl} = \begin{cases} 
\sum_{j_2 < j_3 < \cdots < j_m} N_{j_1j_2 \ldots j_m} \frac{\pi_{j_1} (\sum_{i=2}^{m} \pi_{j_i})}{(\sum_{i=1}^{m} \pi_{j_i})^2} & \text{if } k = l = j_1 \\
- \sum_{j_3 < j_4 < \cdots < j_m} N_{j_1j_2 \ldots j_m} \frac{\pi_{j_1} \pi_{j_2}}{(\sum_{i=1}^{m} \pi_{j_i})^2} & \text{if } k = j_1, l = j_2 \\
0 & \text{otherwise.}
\end{cases} \quad (2.13)
\end{align*}$$

Since \(P(T_{j_1} > \{T_{j_1}, T_{j_2}, \ldots, T_{j_m}\}) = \frac{\pi_{j_1}}{\sum_{i=1}^{m} \pi_{j_i}}\) is not dependent on parameter scale, we assume a convenient scale determining constraint

$$\sum_{i=1}^{L} V_i = 0. \quad (2.14)$$
Let
\[ B_{(p+q)} = \begin{pmatrix} B_{(p)} \\ B_{(q)} \end{pmatrix} \]  
(2.15)
be a partition of the orthonormal contrast matrix of order \((L-1) \times L\), with \(p + q = L - 1\). Here, our interest lies in finding the information matrix of \(\Theta_1 = B_{(p)}V\), while \(\Theta_0 = B_{(q)}V\) are the nuisance parameters. Under the assumption
\[ \Theta_0 = B_{(q)}V = 0_q, \]  
(2.16)
we first find the information matrix of \(\Theta_1 = (\theta_1, \theta_2, \ldots, \theta_p)'\). Let \(I_p\) denote an identity matrix of order \(p\). Also, let \(G' = \begin{bmatrix} \sqrt{L-1} & B'_{(p)} & B'_{(q)} \end{bmatrix}\), where 1 is a column vector of all ones. Then \(G\) is an orthogonal matrix of order \(L \times L\), and \(GG' = G'G = I_L\).

Therefore,
\[ B'_{(p)}B_{(p)} = I_L - \frac{11'}{L} - B'_{(q)}B_{(q)}. \]  
(2.17)
Now, \(\Theta_1 = B_{(p)}V\) and using (2.14), (2.16) and (2.17), we have
\[ B'_{(p)}\Theta_1 = B'_{(p)}B_{(p)}V \]
\[ \Rightarrow B'_{(p)}\Theta_1 = [I_L - \frac{11'}{L} - B'_{(q)}B_{(q)}]V \]
\[ \Rightarrow B'_{(p)}\Theta_1 = I_2V = V. \]  
(2.18)

Let \(B_{(p)} = ((b_{1r}, b_{2r}))\). Also, let the \(p \times p\) information matrix of \(\Theta_1\) be denoted by \(C_{(p)} = ((c_{rs}))\). Then from (2.2) and (2.4), and using (2.18), we have
\[ c_{rs} = \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1j_2\ldots j_m} E \left[ \frac{\partial \ln f_{j_1j_2\ldots j_m}}{\partial \theta_r} \frac{\partial \ln f_{j_1j_2\ldots j_m}}{\partial \theta_s} \right] \]
\[ = \sum_k \sum_l \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1j_2\ldots j_m} E \left[ \frac{\partial \ln f_{j_1j_2\ldots j_m}}{\partial V_k} \frac{\partial \ln f_{j_1j_2\ldots j_m}}{\partial V_l} \right] b_{rk}b_{sl} \]
\[ = \sum_k \sum_l \lambda_{kl}b_{rk}b_{sl}. \]  
(2.19)
Thus,
\[ C_{(p)} = B_{(p)}\Lambda B'_{(p)}, \]  
(2.20)
where \(B_{(p)}\Lambda B'_{(p)}\) is non-negative definite. For notational convenience we denote \(C_{(p)}\) by \(C\). A choice design for estimating \(\Theta_1\), is said to be connected if \(\text{rank}(C) = p\). We restrict ourselves to the class of all connected designs. When a design is connected, it ensures the estimably of \(\Theta_1\). In general \(\Theta_1\) is estimable if and only if \(\text{rank}(C) = p\). Thus, the information matrix of \(B_{(p)}V\), under main effects model involves \(B_{(p)}\) which is the contrast matrix corresponding to the complete set of factorial effects involving single factors.
3 C-MATRIX UNDER MAIN EFFECTS MODEL

For the purpose of optimal choice design, as in the literature, we assume that the options are equally attractive i.e., \( \pi_1 = \pi_2 = \cdots = \pi_L = (\pi_0, \text{say}) \).

Then from (2.13), \( \Lambda \) turns out to be

\[
\lambda_{kl} = \begin{cases} 
\frac{m-1}{m^2} \sum_{j_2<j_3<\cdots<j_m} N_{j_1,j_2\cdots,j_m} & \text{if } k = l = j_1 \\
-\frac{1}{m^2} \sum_{j_3<j_4<\cdots<j_m} N_{j_1,j_2\cdots,j_m} & \text{if } k = j_1, l = j_2 \\
0 & \text{otherwise},
\end{cases}
\]

(3.1)

Let \( M^{(j_1,j_2\cdots,j_m)} = (m_{st}) \) be a \( L \times L \) matrix corresponding to a choice set \((T_{j_1}, T_{j_2}, \ldots, T_{j_m})\), where

\[
m_{st} = \begin{cases} 
m - 1 & \text{if } s = t, (s, t) \in \{j_1, j_2, \ldots, j_m\} \\
-1 & \text{if } s \neq t, (s, t) \in \{j_1, j_2, \ldots, j_m\} \\
0 & \text{otherwise}.
\end{cases}
\]

Then for any choice experiment with \( N \) choice sets, we can write

\[
\Lambda = \sum_{j_1<j_2<\cdots<j_m} L_{j_1,j_2\cdots,j_m} M^{(j_1,j_2\cdots,j_m)}. 
\]

(3.2)

We now concentrate on \( 2^n \) choice experiments \((l_i = 2, i = 1, 2, \ldots, n)\). Then, the main effect contrast belonging to factor \( F_h \) is defined by \( P_h \), where

\[
P_h = \otimes_{i=1}^{h-1} \left( \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \right) \otimes \left( \frac{1}{\sqrt{2}} \frac{-1}{\sqrt{2}} \right) \otimes_{i=h+1}^{n} \left( \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \right).
\]

(3.3)

Under the main effects model, let \( B = \sqrt{2^n} \left( P_1' \cdots P_h' \cdots P_k' \cdots P_n' \right)' \) be the \( n \times 2^n \) matrix with rows representing the orthogonal contrast vectors corresponding to the \( n \) main effects. We assume that all the other factorial effects are zero. Then by (2.20), the \( C \)-matrix for estimating the main effects is \( C = 2^{-n} BAB' \), where, as before, for a given choice design \( d \), \( \Lambda \) is a \( 2^n \times 2^n \) matrix defined under equally attractive options.

Thus the \( C \)-matrix for estimating the main effects is

\[
C = 2^{-n} BAB' = 2^{-n} B \left( \frac{1}{m^2} \sum_{j_1<j_2<\cdots<j_m} N_{j_1,j_2\cdots,j_m} M^{(j_1,j_2\cdots,j_m)} \right) B' \\
= \frac{1}{2^n m^2} \sum_{j_1<j_2<\cdots<j_m} N_{j_1,j_2\cdots,j_m} \{ BM^{(j_1,j_2\cdots,j_m)} B' \} = \frac{1}{2^n m^2} C^*,
\]

(3.4)

where

\[
C^* = \sum_{j_1<j_2<\cdots<j_m} N_{j_1,j_2\cdots,j_m} \{ BM^{(j_1,j_2\cdots,j_m)} B' \}.
\]

(3.5)
Furthermore, $C^*$ (and also $C$) is symmetric.

We can consider the matrix $M^{(j_1j_2\ldots j_m)}$ as the contribution of the choice set $(T_{j_1}, T_{j_2}, \ldots, T_{j_m})$ to $\Lambda$ (and thus the $C^*$-matrix). The definition of $M^{(j_1j_2\ldots j_m)}$ suggests that we can write

$$M^{(j_1j_2\ldots j_m)} = \sum_{j_r < j_{r'}} M^{(j_{j_r}j_{j_{r'}})}$$

(3.6)

where, $j_r, j_{r'} \in \{j_1, j_2, \ldots, j_m\}$.

This means, the contribution of the choice set $(T_{j_1}, T_{j_2}, \ldots, T_{j_m})$ to the $C^*$-matrix is equal to the sum of the individual contributions of the \(\binom{m}{2}\) different component pairs that it contains. Therefore, $C^*$ corresponding to choice sets of size $m$ can be translated in terms of $C^*$ corresponding to component sets of size 2.

**Lemma 3.1.** Let $B_h = (x_{h1}, \ldots, x_{hj_r}, \ldots, x_{hj_r}, \ldots, x_{h2^n})$ and $B_k = (x_{k1}, \ldots, x_{kj_r}, \ldots, x_{kj_r}, \ldots, x_{kj_r})$ be row vectors with $2^n$ real elements. Then for a given component pair $(T_{j_r}, T_{j_{r'}})$, the value of $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = (x_{hj_r} - x_{hj_r}) (x_{kj_r} - x_{kj_r})$.

**Proof.** We observe that $M^{(j_{j_r}j_{j_{r'}})}$ is a $2^n \times 2^n$ matrix with all elements 0 except $M^{(j_{j_r}j_{j_{r'}})} = M^{(j_{j_r}j_{j_{r'}})} = 1$ and $M^{(j_{j_r}j_{j_{r'}})} = -1$. Then

$$B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = (0, \ldots, (x_{hj_r} - x_{hj_r}), \ldots, (x_{hj_r} - x_{hj_r}), \ldots, 0) B'_k$$

$$= (x_{hj_r} - x_{hj_r}) x_{kj_r} - (x_{hj_r} - x_{hj_r}) x_{kj_r}$$

$$= (x_{hj_r} - x_{hj_r})(x_{kj_r} - x_{kj_r}).$$

(3.7)

\[\square\]

From Lemma 3.1, it follows that for a component pair $(T_{j_r}, T_{j_{r'}})$, the possible realized values of $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k$ are:

(P1) If $x_{hj_r} = x_{hj_r}$ or $x_{kj_r} = x_{kj_r}$, then $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = 0$.

(P2) If $x_{hj_r} = -x_{hj_r} = \pm 1$ and $x_{kj_r} = -x_{kj_r} = \pm 1$, then $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = 4$.

(P3) If $x_{hj_r} = -x_{hj_r} = \pm 1$ and $x_{kj_r} = -x_{kj_r} = \mp 1$, then $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = -4$.

(3.10)

With treatment combination $T_i = (i_1 i_2 \ldots i_h \ldots i_k \ldots i_n)$, $i_r = 0, 1; r = 1, 2, \ldots, n$; the $h^{\text{th}}$ and $k^{\text{th}}$ positions of the treatment $T_i$ is denoted by $(i_h, i_k)_{hk}$ and for the component pair $(T_{j_r}, T_{j_{r'}})$, the $h^{\text{th}}$ and $k^{\text{th}}$ positions of the treatment pair is denoted by $(i_h i_k, j_h j_k)_{hk}$. The following Lemma provides a converse result of Lemma 3.1 in the sense that it establishes possible component pairs that gives rise to specific values of $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k$.

**Lemma 3.2.** Three exhaustive cases leading to possible values of $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k$ are

- Case 1: $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = -4$ when $(i_h i_k, j_h j_k)_{hk} \equiv (01, 10)_{hk}$.
- Case 2: $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = 4$ when $(i_h i_k, j_h j_k)_{hk} \equiv (00, 11)_{hk}$.
- Case 3: $B_h M^{(j_{j_r}j_{j_{r'}})} B'_k = 0$ for all other situations.

**Proof.** From (3.3) and the fact that $T_i$’s are arranged in a lexicographic order, for any treatment combination $T_j$, $x_{kj} = 1,-1$ if and only if $j_h = 0, 1$ respectively ($h = 1, 2, \ldots, n$). The proof then follows from (3.7). \[\square\]
4 CHARACTERIZATION OF $C$-MATRIX

In what follows, we first find conditions under which the $C$-matrix has off-diagonal elements zero.

Let $F_h$ and $F_k$ be any two main effects and let

- $N_{hk}^+ =$ Total number of component pairs of the type $(00, 11)_{hk}$ corresponding to $h^{th}$ and $k^{th}$ positions across all $\binom{m}{2}$ possible pairs of a choice set of size $m$ and among all such sets in the choice experiment.
- $N_{hk}^- =$ Total number of component pairs of the type $(01, 10)_{hk}$ corresponding to $h^{th}$ and $k^{th}$ positions across all $\binom{m}{2}$ possible pairs of a choice set and among all such sets in the choice experiment.

**Theorem 4.1.** The $(h, k)^{th}$ position of $C$-matrix will be zero if and only if $N_{hk}^+ = N_{hk}^-$. 

**Proof.** Following (3.6), since $C^*$ corresponding to choice sets of size $m$ can be translated in terms of $C^*$ corresponding to component sets of size 2, therefore the proof follows from Lemma 3.2 on noting the contribution towards the $(h, k)^{th}$ position of the $C$-matrix by $N$ choice sets through its $\binom{m}{2}$ possible component pairs. The three exhaustive cases of Lemma 3.2 leads to

- $N_{hk}^- =$ Total number of component pairs falling under Case 1.
- $N_{hk}^+ =$ Total number of component pairs falling under Case 2.
- $N - (N_{hk}^+ + N_{hk}^-) =$ Total number of choice pairs falling under Case 3.

Let $c_{hk}$ denote the $(h, k)^{th}$ element of $C^*$. Then, it follows from (3.5) that

\[
\begin{align*}
c_{hk} &= \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1 j_2 \ldots j_m} \{ B_h M^{(j_1 j_2 \ldots j_m)} B_k' \} \\
&= \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1 j_2 \ldots j_m} \left\{ B_h \left( \sum_{j_r < j_r'} M^{(j_r j_r')} \right) B_k' \right\} \\
&= \sum_{j_1 < j_2 < \ldots < j_m} N_{j_1 j_2 \ldots j_m} \sum_{j_r < j_r'} \{ B_h M^{(j_r j_r')} B_k' \} \\
&= \left\{ (4N_{hk}^+ - 4N_{hk}^-) + 0(N - (N_{hk}^+ + N_{hk}^-)) \right\}.
\end{align*}
\]

Thus $c_{hk} = 0$ if and only if $N_{hk}^+ = N_{hk}^-$. \qed

We will now find the contribution of each choice set $S_m$ of size $m$ to the diagonal positions of $C^*$ as defined in (3.5).

**Lemma 4.2.** Every component pair adds a value 4 in the $(h, h)^{th}$ position of the $C^*$, if and only if the pair has a change of level at the $h^{th}$ position of its treatment combinations.
Proof. Every component pair \((T_{jr}, T_{jr}')\) is adding a value \(B_h M^{(j,r,r')} B'_h\) at \(c_{hh}\). From (P2) of (3.9) it follows that this value will be 4 if and only if there is a change of level in the \(h^{th}\) position of the component pair.

Let \(n_h \in \{0, 1, 2, \ldots, m\}\) be the number of treatment combinations which have zero at the \(h^{th}\) position in the choice set \(S_m\).

**Lemma 4.3.** Every \(S_m\) adds a value \(4n_h(m - n_h)\) to the \((h,h)^{th}\) position of \(C^*\).

**Proof.** Lemma 4.2 says that every component pair adds a value 4 to \(c_{hh}\), if and only if the pair has a change of level at the \(h^{th}\) position of its treatment combinations. There are a total of \(\binom{m}{2}\) component pairs possible from \(S_m\). The contribution of \(S_m\) to \(c_{hh}\) is the same as the sum of contributions of all the \(\binom{m}{2}\) component pairs corresponding to \(S_m\). Now there are \(n_h\) treatment combinations in \(S_m\) which have a 0 at the \(h^{th}\) position. We call this subset as \(A\). Therefore the set \(\bar{A}\) contains all treatment combinations which have a 1 at the \(h^{th}\) position. Every component pair which has one treatment from \(A\) and another treatment from \(\bar{A}\), adds a value 4 to \(c_{hh}\). There are a total of \(n_h(m - n_h)\) such pairs and they all together add a value \(4n_h(m - n_h)\) to \(c_{hh}\). Hence the theorem.

**Corollary 4.4.** Every \(S_m\) adds a \(4 \sum_{h=1}^{n} n_h(m - n_h)\) value to the \(\text{trace}(C^*)\).

We will now find out the expression of \(\text{trace}(C^*)\) when there are \(N\) choice sets. For this purpose we will use the following notations.

- \(S_{m_i}\) = the \(i^{th}\) choice set, \(i = 1, 2, \ldots, N\).
- \(n_{h_i}\) = number of treatment combinations which have zero at the \(h^{th}\) position in the choice set \(S_{m_i}\).

**Lemma 4.5.** For \(N\) choice sets \(S_{m_1}, \ldots, S_{m_N}\), \(\text{trace}(C^*) = 4 \sum_{i=1}^{N} \sum_{h=1}^{n} n_{h_i}(m - n_{h_i})\).

**Proof.** From Corollary (4.4), every choice set \(S_{m_i}\) adds a value \(4 \sum_{h=1}^{n} n_{h_i}(m - n_{h_i})\) to \(\text{trace}(C^*)\). Therefore, for \(N\) choice sets \(\text{trace}(C^*) = 4 \sum_{i=1}^{N} \sum_{h=1}^{n} n_{h_i}(m - n_{h_i})\).

**Theorem 4.6.** Maximum of \(\text{trace}(C^*)\) is attained when

\[
n_{h_i} = \begin{cases} 
\frac{m}{2} & \text{if } m \text{ even} \\
\frac{m - 1}{2} \text{ or } \frac{m + 1}{2} & \text{if } m \text{ odd}
\end{cases}
\]

for every position \(h\) and for every choice set \(S_{m_i}\).
Proof. Maximum of $\text{trace}(C^\ast)$ is attained when every choice set $S_{m_i}$ in the experiment contributes maximum value towards $\text{trace}(C^\ast)$. Each choice set $S_{m_i}$ will contribute maximum value if and only if every $h^{th}$ position of its treatments contributes maximum value to $c_{hk}$. Lemma 4.3 says that if $S_{m_i}$ has $n_{hi}$ zeros at the $h^{th}$ position of its treatments then it will add a value $4n_{hi}(m - n_{hi})$ to $\text{trace}(C^\ast)$. We want to maximize $4n_{hi}(m - n_{hi})$ for $n_{hi}$. Let $f(n_{hi}) = 4n_{hi}(m - n_{hi})$ and let $k_0$ be the point at which the function attains its maximum. Then,

$$f(k_0 - 1) \leq f(k_0) \quad \Rightarrow \quad 4(k_0 - 1)(m - (k_0 - 1)) \leq 4k_0(m - k_0)$$

$$\Rightarrow \quad m(k_0 - 1) - k_0(k_0 - 1) \leq mk_0 - k_0^2$$

$$\Rightarrow \quad 2k_0 - m - 1 \leq 0$$

$$\Rightarrow \quad k_0 \leq \frac{m + 1}{2} \quad (4.1)$$

and $f(k_0) \geq f(k_0 + 1)$

$$\Rightarrow \quad k_0 \geq \frac{m - 1}{2} \quad (4.2)$$

Since $k_0$ only takes integer value, therefore from (4.1) and (4.2) we conclude that $f(n_{hi})$ is maximum when (i) $n_{hi} = \frac{m - 1}{2}$ or $n_{hi} = \frac{m + 1}{2}$ (for $m$ odd) and (ii) $n_{hi} = \frac{m}{2}$ (for $m$ even). Hence the proof. \hfill \Box

**Theorem 4.7.** For $N$ choice sets of size $m$, the upper bound to $\text{trace}(C^\ast)$ is

$$\text{trace}(C^\ast) \leq \begin{cases} Nnm^2 & \text{for } m \text{ even} \\ Nn(m^2 - 1) & \text{for } m \text{ odd} \end{cases}$$

*Proof.* From Lemma 4.5 and Theorem 4.6 we can say that $\text{trace}(C^\ast)$ will be maximum if and only if each $n_{hi} = k_0$ for every $h$ and every $i$. Therefore, for $m$ even, $k_0 = \frac{m}{2}$ and

$$\text{trace}(C^\ast) \leq 4 \sum_{i=1}^{N} \sum_{h=1}^{n} \frac{m}{2} \left( m - \frac{m}{2} \right) = 4 \sum_{i=1}^{N} \sum_{h=1}^{n} \frac{m^2}{4} = Nnm^2.$$  

Also, for $m$ odd, $k_0 = \frac{m - 1}{2}$ or $\frac{m + 1}{2}$ and

$$\text{trace}(C^\ast) \leq 4 \sum_{i=1}^{N} \sum_{h=1}^{n} \frac{m \pm 1}{2} \left( m - \frac{m \pm 1}{2} \right) = 4 \sum_{i=1}^{N} \sum_{h=1}^{n} \frac{m^2 - 1}{4} = Nn(m^2 - 1).$$

\hfill \Box

**Corollary 4.8.** For $N$ choice sets of size $m$, the upper bound to $\text{trace}(C)$ is

$$\text{trace}(C) \leq \begin{cases} \frac{Nn}{2^m} & \text{for } m \text{ even} \\ \frac{Nn(m^2 - 1)}{2^nm^2} & \text{for } m \text{ odd} \end{cases}$$
Proof. Follows from Theorem 4.7 on noting that
\[
\max(\text{trace}(C)) = \frac{1}{2^n m^2} \max(\text{trace}(C^*)),
\]
\]

Remark 4.9. For given \(N\) and \(n\), with respect to maximum of \(\text{trace}(C)\), (i) all designs with \(m\) even are equivalent and (ii) a design with \(m\) odd is always inferior to a design with \(m\) even.

5 Construction of universally optimal designs

The criteria of *universal optimality* was introduced by Kiefer (1975) and is a strong family of optimality criteria which includes \(A-, D-\), and \(E-\) criteria as particular cases.

Let \(W_p\) denote the class of positive definite symmetric matrices of order \(p\). A design \(d^* \in \mathcal{D}\) is universally optimal over \(\mathcal{D}\) if \(d^*\) minimizes \(\phi(C_d), d \in \mathcal{D}\) for any \(\phi : W_p \rightarrow (-\infty, \infty]\) satisfying

1. \(\phi\) is matrix convex, i.e., \(\phi(aC_1 + (1 - a)C_2) \leq a\phi(C_1) + (1 - a)\phi(C_2)\) for \(C_i \in W_p, i = 1, 2\) and \(0 \leq a \leq 1\),

2. \(\phi(bC)\) is non increasing in the scalar \(b \geq 0\) for each \(C \in W_p\),

3. \(\phi\) is invariant under each simultaneous permutation of rows and columns of \(C\) in \(W_p\).

Kiefer (1975) obtained the following sufficient condition for universal optimality.

Suppose \(d^* \in \mathcal{D}\) and \(C^*_d\) satisfies (a) \(C^*_d\) is scalar multiple of \(I_p\), i.e., \(C^*_d = \alpha I_p\), and (b) \(\text{trace}(C^*_d) = \max_{d \in \mathcal{D}} \text{trace}(C_d)\), then \(d^*\) is universally optimal in \(\mathcal{D}\).

We now provide a simple method for constructing universally optimal designs for a \(2^n\) choice experiment with choice set size \(m\). Our characterization of the Information matrix facilitates construction of such optimal designs with minimal number of choice sets and gives more flexibility for choosing \(m\). For \(m\) odd, optimal designs are achieved having relatively lesser number of choice sets than those obtained in Demirkale, Donovan and Street (2013). Let \(\mathcal{D}_{N,m}\) be the class of all connected choice designs involving \(N\) choice sets of size \(m\) each.

Theorem 5.1. Let \(n = 4t - j\), where \(t\) is a positive integer and \(j = 0, 1, 2, 3\). Also, given a Hadamard matrix \(H\) of order \(4t\), let for \(u = 1, 2, \ldots, 4t\), \(H_u\) be the Hadamard matrix derived from \(H\) by multiplying the \(u\)-th column of \(H\) by \(-1\). Let \(Z_1 = H, Z_2 = -H, Z_{2u+1} = H_u, Z_{2u+2} = -H_u\). For \(w = 1, 2, \ldots, 2n + 2\), let \(A_w\) be respective matrices obtained by replacing \(-1\)'s by \(0\) and deleting rightmost \(j\) columns from \(Z_w\), where \(j = 4t - n, j \in \{0, 1, 2, 3\}\). Consider each row of \(A_w\) as treatment combination. Then \(D_1 = (A_1, A_2), D_2 = (A_1, A_2, A_3), D_3 = (A_1, A_2, A_3, A_4), \ldots, D_{2n+1} = (A_1, A_2, A_3, A_4, \ldots, A_{2n+2})\) are universally optimal \(2^n\) choice design in \(\mathcal{D}_{4t,m}\) for \(m = 2, 3, 4, \ldots, 2n + 2\), respectively.
Remark 5.3. minimizes repetitive sets of options within the constructed choice sets. such a flexibly may allow having

\[ N \] of all connected choice designs involving \( 8 \) choice sets of size 4 each. The

Example 5.1. optimal derived from \( H \) further increase in the choice set size by considering distinct Hadamard matrices \( H \).

Remark 5.2. Proof. To prove that this construction gives universally optimal choice design, we

will show that the \( C \)-matrix of the design is of the form \( \alpha I_n \), where \( \alpha \) is a constant and \( \text{trace}(C) \) is maximum. Therefore, first we show that every \( (h,k) \)-th position of

the design \( D_w \) consists of the component pair designs \( \{ (A_\delta, A_{\delta'}), 1 \leq \delta < \delta' \leq w + 1 \} \). We denote the component pair designs of \( D_w \) by \( D_{\delta\delta'}^w \), \( 1 \leq \delta < \delta' \leq w + 1 \). We will now calculate \( N_{hk}^+ \) and \( N_{hk}^- \) for the design \( D_w, w = 1, 2, \ldots, 2n + 1 \).

Since \( H \) is a Hadamard matrix of order \( 4t \), for any two columns \( h \) and \( k \) of \( A_w \), the combinations from the set \( \{ (00)_{hk}, (11)_{hk} \} \) and from the set \( \{ (10)_{hk}, (01)_{hk} \} \) occurs equally often. Therefore, in every component pair design \( D_{\delta\delta'}^w \), it is easy to see that \( N_{hk}^+ = N_{\delta\delta'}^-(hk), 1 \leq \delta < \delta' \leq w + 1 \), where \( N_{\delta\delta'}^-(hk) \) is the total number of pairs of the type \( (00,11)_{hk} \) corresponding to \( h^\text{th} \) and \( k^\text{th} \) positions among all choice pairs in \( D_{\delta\delta'}^w \), and \( N_{\delta\delta'}^+(hk) \) is the total number of pairs of the type \( (01,10)_{hk} \) corresponding to \( h^\text{th} \) and \( k^\text{th} \) positions among all choice pairs in \( D_{\delta\delta'}^w \). In other words, for the design \( D_w, N_{hk}^+ = N_{hk}^- \). Using the result of Theorem 4.1 it thus follows that the \( C \)-matrix has off-diagonal elements zero for the design \( D_w, w = 1, 2, \ldots, 2n + 1 \).

The construction also ensures that \( n_{hk} = m/2 \) for \( D_w \)'s with \( m \) even, and \( n_{hk} = (m-1)/2 \) or \( (m+1)/2 \) for \( D_w \)'s with \( m \) odd, for every position \( h \) and for every choice set. Therefore using Theorem 4.6 and Corollary 4.8 we can say that the diagonal elements of \( C \)-matrix are equal and \( \text{trace}(C) \) is maximum for the design. Thus the designs \( D_w, w = 1, 2, \ldots, 2n+1 \) are universally optimal design for \( m = 2, 3, \ldots, 2n+2 \) respectively for a \( 2^n \) choice experiment.

Remark 5.2. The construction as provided in Theorem 5.1 can be extended to allow further increase in the choice set size by considering distinct Hadamard matrices \( H_u \) derived from \( H \) by multiplying any \( s \) columns of \( H \) by \(-1, s = 1, 2, \ldots, 2t \). Though such a flexibility may allow having \( m \) large, it is desirable to select those \( H_u \) which minimizes repetitive sets of options within the constructed choice sets.

Remark 5.3. In view of Remark 4.9, for given \( N \) and \( n \), it follows that a universally optimal \( 2^n \) choice design in \( D_{N,2} \) is also universally optimal in a more broader class of all connected choice designs involving \( N \) choice sets and arbitrary \( m \).

Example 5.1. Consider a \( 2^{8-j} \) choice experiment \( (j = 0, 1, 2, 3) \) conducted through \( 8 \) choice sets of size 4 each. The \( 2^8 \) \((j = 0)\) choice design \( d^* \) which is universally optimal in \( D_{8,4} \) is

\[
d^* = \begin{pmatrix}
(11111111, & 00000000, & 01111111, & 10000000)
(10101010, & 01010101, & 00101010, & 11010101)
(11010100, & 00110011, & 01001100, & 10110011)
(10011001, & 01100110, & 00011001, & 11100110)
(11110000, & 00011111, & 01110000, & 10011111)
(10100101, & 01011010, & 00100101, & 11011010)
(11000011, & 00111100, & 01000011, & 10111100)
(10010110, & 01101001, & 00010101, & 11101001)
\end{pmatrix}
\]

Deleting the last \( j \) factors, we get the corresponding universally optimal design in \( D_{8,4} \) for a \( 2^{8-j} \) experiment. Finally, taking the first \( m \) elements from each choice set \( (m = 2, 3) \), we get universally optimal designs in \( D_{8,m} \) for a \( 2^{8-j} \) experiment \((j = 0, 1, 2, 3)\).
References


