Journal of Statistical Planning and Inference 161 (2015) 81–90

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Column-orthogonal and nearly column-orthogonal designs for models with second-order terms

S. Stylianou [a], K. Drosou [b], S.D. Georgiou [a,*], C. Koukouvinos [b]

[a] School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, VIC 3001, Australia
[b] Department of Mathematics, National Technical University of Athens, Zografou, 15773, Athens, Greece

ARTICLE INFO

*Article history:*
Received 23 May 2014
Received in revised form 20 January 2015
Accepted 21 January 2015
Available online 30 January 2015

*Keywords:*
Screening design
Aliased structure
Generalized orthogonal design
Linear model
Column-orthogonal designs
Nearly column-orthogonal design

ABSTRACT

In this paper, we study U-type, column-orthogonal and nearly column-orthogonal designs. Nearly column-orthogonal designs are very useful in cases where column-orthogonal designs are not known. New designs are obtained which are suitable for screening experiments. In certain cases, the constructed designs are shown to be optimal with respect to their aliased structure. The aliased structures, of U-type, column-orthogonal and nearly column-orthogonal designs, are calculated and presented in closed form. This fact makes the new approach innovating and enables the construction of designs that are different from the designs constructed in the literature. An extended multiplication theorem and new infinite families of column-orthogonal designs are presented using periodic Golay pairs.



## 1. Introduction and preliminary results

Designs for computer experiments have extensively developed in the recent literature: see for example Georgiou and Efthimiou (2014), Hernandez et al. (2012), Lin et al. (2010, 2009), Sun et al. (2010, 2009), Yang and Liu (2012), Yin and Liu (2013). A large class of designs for computer experiments includes the well known and commonly used Latin hypercube designs (LHDs). These designs have $n$ uniformly spaced levels on $n$ runs. Recently, Sun et al. (2011), following Bingham et al. (2009), relaxed the LHDs' restriction that the run size equals the number of levels, and introduced an alternative class of designs, the so-called "column-orthogonal designs for computer experiments". When all levels are equally replicated and uniformly distributed in each column, the design is called a $U$-type design. When each column is orthogonal to the mean effect, the design is called a *mean orthogonal* design. A design with $n$ runs and $m$ factors having $z$ zeros and $q-1$ non-zero equally replicated distinct levels in each factor is denoted by $D(n; z; q^m)$. When $z = 0$, then the design is a $U$-type design and $z$ can be omitted from the notation, i.e. $D(n; (q-1)^m)$. When $z = n/q$, then the design is a $U$-type design and $z$ can be omitted from the notation, i.e. $D(n; q^m)$. When $z \neq 0$ and $z \neq n/q$, then the derived designs are mean orthogonal, but not of U-type, and such designs will be suitable only for quantitative factors. A design is called a column-orthogonal design, denoted by COD$(n; z; q^m)$, if the inner product of any two columns is zero. A nearly column-orthogonal design is denoted by NCOD$(n; z; q^m)$. Note that a column-orthogonal design COD$(n; z; q^m)$ is not always of U-type. A column-orthogonal design COD$(n; z; q^m)$ that is also of U-type, will be denoted as *U-type* COD$(n; z; q^m)$. A column-orthogonal design is called

* Corresponding author.
*E-mail address:* stelios.georgiou@rmit.edu.au (S.D. Georgiou).

http://dx.doi.org/10.1016/j.jspi.2015.01.002

$\ell$-orthogonal if the sum of the elementwise products of any $\ell$ columns (whether they are distinct or not) is zero. When $\ell = 3$, then the designs are called 3-orthogonal (see Bingham et al., 2009). The design $T$ is said to be a full fold-over (or fold-over) of the design $D$ if $T = \begin{pmatrix} D \\ -D \end{pmatrix}$. In Georgiou et al. (2014b), it was shown that any full fold-over design $T$ is 3-orthogonal.

A matrix is said to be *circulant* if each row vector is rotated one element to the right relative to the preceding row vector. A circulant matrix $A = \text{circ}(B)$ is fully specified by one vector $B$, which appears as the first row of the matrix. The remaining rows of $A$ are each cyclic permutations of the vector $B$ with an offset equal to the row index. Circulant matrices with an opposite direction of shift are called *back-circulants*. If no confusion is caused, the circulant matrix and the corresponding generator row vector will both be abbreviated by the same letter, i.e. $A = \text{circ}(A)$. The elements of the information matrix of a circulant matrix can be calculated using the periodic autocorrelation function of its first row. A set of matrices $\bigcup_{j=1}^{\ell}\{B_j\}$ is said to be *disjoint* if $B_i * B_j = 0$ for all $i \neq j,\ i, j = 1, 2, \dots, \ell$ where $*$ denotes the Hadamard product. Let $A = \{A_j : A_j = \{a_{j,0}, a_{j,1}, \dots, a_{j,n-1}\},\ j = 1, \dots, \ell\}$, be a set of $\ell$ vectors of length $n$. These vectors are said to be a set of *disjoint vectors* if the set of the corresponding circulant matrices $B_j = \cup_j\{\text{circ}(A_j)\}, j = 1, \dots, \ell$ is disjoint. The periodic autocorrelation function $P_A(s)$ (in abbreviation PAF) and the non-periodic autocorrelation function $N_A(s)$ (in abbreviation NPAF) are defined, reducing $i + s$ modulo $n$, as

$$P_A(s) = \sum_{j=1}^{\ell}\sum_{i=0}^{n-1} a_{j,i}a_{j,i+s}, \quad \text{and} \quad N_A(s) = \sum_{j=1}^{\ell}\sum_{i=0}^{n-s-1} a_{j,i}a_{j,i+s}, \quad s = 0, \dots, n-1$$

respectively. The set of row vectors $A$ is said to have zero periodic autocorrelation function (zero PAF) if $P_A(s) = 0$ and zero non-periodic autocorrelation function (zero NPAF) if $N_A(s) = 0 (s = 1, \dots, n-1)$. A pair of matrices $(A, B)$ is said to be amicable (anti-amicable) if $AB^T - BA^T = 0 (AB^T + BA^T = 0)$. Following Kharaghani (2000), a set $\{B_1, B_2, \dots, B_{2n}\}$ of real square matrices is said to be *amicable* if

$$\sum_{i=1}^{n}\left(B_{2i-1}B_{2i}^T - B_{2i}B_{2i-1}^T\right) = 0. \tag{1}$$

Clearly, a set of mutually amicable matrices is amicable but the converse is not generally true. A set of matrices $\{B_1, B_2, \dots, B_n\}$ of order $m$ satisfies the *additive property* if

$$\sum_{i=1}^{n} B_iB_i^T = fI_m.$$

**Notation 1.** The following notation is also needed.

- $\{.\}$ is used to denote a set, while $[.]$ is used to denote a multiset (the same element is allowed to exist multiple times in a multiset). For example, the set $A = \{a_1, a_2, a_3\}$ is the same as set $B = \{a_1, a_1, a_2, a_2, a_2, a_3\}$; both have the three distinct entries $a_1, a_2, a_3$. On the other hand, the multiset $C = [a_1, a_2, a_3]$ is not the same as the multiset $D = [a_1, a_1, a_2, a_2, a_2, a_3]$ since the first has the three elements $a_1, a_2, a_3$ while the second has the six elements $a_1, a_1, a_2, a_2, a_2, a_3$, even though both have the same three distinct elements $a_1, a_2, a_3$.
- $\#(x, S)$ is used to denote the number of $x$'s in multiset $S$. Example: $\#(a_2, D) = 3$ and $\#(a_3, D) = 1$.
- $\{S\}$ denotes the set of the distinct elements of multiset $S$. Example: $\{C\} = \{D\} = A$.
- $|x|$ denotes the absolute value of the element $x$.

We review some necessary definitions and properties of generalized orthogonal designs. Generalized orthogonal designs and their properties were introduced in Georgiou et al. (2004).

**Definition 1.** Let $D$ be an $n \times m$ matrix on the commuting variables $x_1, \dots, x_t$ where each variable appears in each column in one of the two forms $\pm a_{ij}x_i$, for each $i = 1, \dots, t, j = 1, \dots, u_i$, and $\sum_{i=0}^{t} u_i = n$, where $u_0$ is the number of zeros in each column. Set $s_i = \sum_{j=1}^{u_i} a_{ij}^2$. Then $D$ is a *generalized orthogonal design (GOD)* iff $D^TD = \left(\sum_{i=1}^{t} s_ix_i^2\right) I_m$. Design $D$ will be denoted as $D = \text{GOD}(n; m; a_{1,1}, \dots, a_{1,u_1}; a_{2,1}, \dots, a_{2,u_2}; \dots; a_{t,1}, \dots, a_{t,u_t})$. One alternative notation of a generalized orthogonal design will be $D = \text{GOD}(n; m; \langle k_{1,1}, a_{1,1}\rangle, \dots, \langle k_{1,u_1}, a_{1,u_1}\rangle; \dots; \langle k_{t,1}, a_{t,1}\rangle, \dots, \langle k_{t,u_t}, a_{t,u_t}\rangle)$ where $k_{i,j}$ denotes how many times the variable $x_i$ has the coefficient $a_{i,j}$. If $k_{i,j} = 1$, we usually write $(\dots, a_{ij}, \dots)$, otherwise we write $(\dots, \langle k_{i,j}, a_{ij}\rangle, \dots)$.

Let $A_i (i = 1, \dots, 8)$ be circulant matrices of order $n$ and $R_n$ be the back diagonal identity matrix of order $n$. The following structures are known:

$$C_2 = \begin{pmatrix} A_1 & A_2R_n \\ -A_2R_n & A_1 \end{pmatrix}, \qquad C_4 = \begin{pmatrix} A_1 & A_2R_n & A_3R_n & A_4R_n \\ -A_2R_n & A_1 & A_4^TR_n & -A_3^TR_n \\ -A_3R_n & -A_4^TR_n & A_1 & A_2^TR_n \\ -A_4R_n & A_3^TR_n & -A_2^TR_n & A_1 \end{pmatrix},$$