The Performance of Orthogonal Arrays with Adjoined or Unavailable Runs

Emily Bird

Doctor of Philosophy

2016

School of Mathematical and Physical Sciences
University of Technology Sydney

UNIVERSITY OF TECHNOLOGY SYDNEY

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student: ________________________

Date: ________________________

I have already submitted some of this work for publication. A majority of the work in Chapter 2 on OAs with adjoined runs has appeared in:

Bird, E. M. and Street, D. J. (2016). D-optimal asymmetric orthogonal array plus p run designs. *Journal of Statistical Planning and Inference*, 170:64-76.

Acknowledgements

I would like to give a very big thank you to my supervisor, Prof Deborah Street, for all her guidance, advice and patience throughout this project. This thesis simply wouldn't be possible without the support of someone so understanding and adapting to my research style. And so, hence, she is thus a very thorough editor, therefore. Thank you.

I would also like to thank my co-supervisor, Dr Stephen Bush, for his proof-reading and feedback, as well as his friendly encouragement along the way.

Thanks to my family for their continued love and support through every crazy venture I throw myself into, even a PhD. Special thanks to Dad, partly for his proof-reading of early stages, but mostly for listening to my endless musings over the years.

Finally, a list of acknowledgements would not be complete without paying tribute my faithful hound, Maxie, for keeping me sane. With her relentless love and enthusiasm she constantly reminded me that no matter how compelling my excuses may be, we still have to go for a walk. She will be sorely missed.

Contents

List of Tables

List of Symbols

m	Number of factors, $m = \sum_{i=1}^{k} m_i$
k	Number of distinct factor-levels
m_i	Number of factors with s_i levels, $1 \leq i \leq k$
s_i	Number of levels for each of the m_i factors, $1 \leq i \leq k$
n	Number of runs in a design
N	Number of runs in an OA
α	Number of terms in a main effects only model, $\alpha = 1 + \sum_i (s_i - 1) m_i$
ν	Number of linear-by-linear interaction terms in a model, $0 \leq \nu \leq \binom{m}{2}$
υ	Total number of terms in a model, $\upsilon = \alpha + \nu$
$\mathbf{R}$	Design matrix of order $n \times m$
$\mathbf{r}_i$	ith row of $\mathbf{R}$; row vector of level combinations of length m, $1 \leq i \leq n$
r_{ij}	Entry in the ith row and jth column of $\mathbf{R}$; jth entry in $\mathbf{r}_i$
$\mathbf{0}_s$	Column vector of length s with all elements equal to 0
$\mathbf{1}_s$	Column vector of length s with all elements equal to 1
$\mathbf{I}_s$	Identity matrix of order s
$\mathbf{P}_s$	Contrast matrix of order $(s-1) \times s$
$P_s(l)$	$i$the column of $\mathbf{P}_s$ where $i = l + 1$, $0 \leq l \leq s - 1$
$\mathbf{p}_{sh}$	hth row of $\mathbf{P}_s$, $1 \leq h \leq s - 1$
$\mathbf{p}_{sl}$	$\mathbf{p}_{s1}$, the first row of $\mathbf{P}_s$
$\mathbf{X}$	Model matrix of order $n \times \upsilon$
$\mathbf{x}_i$	The ith row of $\mathbf{X}$, $1 \leq i \leq n$
F_j	The jth factor, $1 \leq j \leq m$
$\mathbf{f}_{jh}$	The hth column within the set of $s_i - 1$ columns in $\mathbf{X}$ corresponding to the main effects of F_j
$\mathbf{f}_{jl}$	The column in $\mathbf{X}$ corresponding to the linear component of F_j
$\mathbf{f}_{jq}$	The column in $\mathbf{X}$ corresponding to the quadratic component of F_j

$\mathbf{M}$	Information matrix of order $v \times v$ $\mathbf{M} = \mathbf{X}'\mathbf{X}$
p	Number of adjoined runs
t	Number of missing runs
$\mathbf{A}$	Model matrix of order $p \times v$ for the p adjoined runs
$\mathbf{B}$	Model matrix of order $t \times v$ for the t missing runs
$\mathbf{a}_i$	The ith row of $\mathbf{A}$
$\mathbf{b}_i$	The ith row of $\mathbf{B}$
q	Compact notation for $\mathbf{a}_i.\mathbf{a}'_j$ or $\mathbf{b}_i.\mathbf{b}'_j$ under a main effects only model
$\mathbf{a}_{ME_i}$	$\mathbf{a}_i$ truncated to the first α entries (i.e. $\mathbf{a}_i$ for a main effects only model)
$\mathbf{b}_{ME_i}$	$\mathbf{b}_i$ truncated to the first α entries (i.e. $\mathbf{b}_i$ for a main effects only model)
$\mathbf{X}_A$	Model matrix of order $(N+p) \times v$ for an OA plus p runs design
$\mathbf{X}_B$	Model matrix of order $(N-t) \times v$ for an OA minus t runs design
$\mathbf{M}_A$	Information matrix of order $v \times v$ for an OA plus p runs design $\mathbf{M}_A = \mathbf{X}'\mathbf{X} + \mathbf{A}'\mathbf{A}$
$\mathbf{M}_B$	Information matrix of order $v \times v$ for an OA minus t runs design $\mathbf{M}_B = \mathbf{X}'\mathbf{X} - \mathbf{B}'\mathbf{B}$
$\mathbf{\Omega}_A$	Matrix of order $p \times p$ to be optimised for an OA plus p runs design under a main effects only model $\mathbf{\Omega}_A = N\mathbf{I}_p + \mathbf{A}\mathbf{A}'$
$\mathbf{\Omega}_B$	Matrix of order $t \times t$ to be optimised for an OA minus t runs design under a main effects only model $\mathbf{\Omega}_B = \mathbf{B}\mathbf{B}' - N\mathbf{I}_p$
$\mathbf{\Psi}_A$	Matrix of order $p \times p$ to be optimised for an OA plus p runs design under a main effects and linear-by-linear interactions model $\mathbf{\Psi}_A = \mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$
$\mathbf{\Psi}_B$	Matrix of order $t \times t$ to be optimised for an OA minus t runs design under a main effects and linear-by-linear interactions model $\mathbf{\Psi}_B = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}' - \mathbf{I}_t$
d_i	Hamming distance between a pair of runs for the m_i factors with s_i levels
$\mathbf{D}_{kp}$	Matrix of Hamming distances of order $k \times \binom{p}{2}$
$\mathbf{d}_i$	Distance vector; ith row of $\mathbf{D}_{kp}$
$d_{i(xy)}$	Hamming distance between $\mathbf{r}_x$ and $\mathbf{r}_y$ for the m_i factors with s_i levels; Entries in $\mathbf{d}_i$

$\mathbf{z}_{(ab)}$ The column in $\mathbf{X}$ corresponding to the linear-by-linear interaction
$\mathbf{z}_{(ab)} = \mathbf{f}_{al} \circ \mathbf{f}_{bl}$

$\mathbf{u}_{(ab)}$ $\alpha \times 1$ vector $\mathbf{X}'_{ME}\mathbf{z}_{(ab)}$

$u_{(ab)cl}$ Entry in $\mathbf{u}_{(ab)}$ corresponding to $\mathbf{f}'_{cl}\mathbf{z}_{(ab)}$
$u_{(ab)cl} = \sum_{i=1}^{n} f_{al_i} f_{bl_i} f_{cl_i}$

$u_{(ab)cq}$ Entry in $\mathbf{u}_{(ab)}$ corresponding to $\mathbf{f}'_{cq}\mathbf{z}_{(ab)}$
$u_{(ab)cq} = \sum_{i=1}^{n} f_{al_i} f_{bl_i} f_{cq_i}$

γ_{abc} Coefficient for $u_{(ab)cl}$
$u_{(ab)cl} = \gamma_{abc} \frac{3}{2}\sqrt{\frac{3}{2}}$

$\omega_{(ab)c}$ Coefficient for $u_{(ab)cq}$
$u_{(ab)cq} = \omega_{(ab)c} \frac{3\sqrt{2}}{4}$

$\mathbf{Z}$ $N \times \nu$ matrix where each column is a vector $\mathbf{z}_{(ab)}$ for an interaction in the model

$\mathbf{U}$ $\alpha \times \nu$ matrix $\mathbf{X}'_{ME}\mathbf{Z}$

Abstract

Orthogonal arrays are a class of fractional factorial designs that are optimal according to a range of optimality criteria. This makes it tempting to construct fractional factorial designs by adjoining additional runs to an OA, or by removing runs from an OA, when the number of runs available for the experiment is only slightly larger or smaller than the number in the OA.

In this thesis we examine the performance of *OA plus p runs designs* and *OA minus t runs designs* in the context of D-optimality and model-robustness. Although we attempt to make general observations where possible, our primary goal is to inform the use of quantitative factors, hence we focus on factors at more than two levels to allow for some curvature in the model.

We begin by considering the performance of these designs under a main effects only model, and show that optimality depends only on the pairwise Hamming distance of the adjoined (or removed) runs. We present an algorithm for finding optimal Hamming distances and provided general methods for constructing optimal sets of runs once the optimal pairwise Hamming distances have been identified.

In order to consider the performance of these designs when interaction terms are included in the model, it transpired that we require a complete set of geometrically non-isomorphic designs to study. Thus, we enumerate all geometric isomorphism classes for symmetric OAs with ternary factors and 18 runs, and prove that these classes cover the entire $\mathrm{OA}[18, 3^m]$ space.

We then consider the inclusion of a subset of linear-by-linear interactions in the model, and derive matrices to be optimised under this setting. We conduct an empirical study on the $\mathrm{OA}[18, 3^m]$ we have enumerated and give examples of D-optimal and model-robust designs for each of the design spaces, that is, for each of $m = 3$, 4, 5, 6 and 7.