

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



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	ix
List of Symbols	xiii
Abstract	xvii
1 Introduction	1
1.1 Background	1
1.1.1 Terminology	3
1.1.2 The Model	5
1.1.3 Measures of Goodness	12
1.1.4 Experimenter's Objectives	15
1.1.5 Isomorphism	16
1.2 Notation and Useful Results	17
1.2.1 Properties of OAs	18
1.2.2 Matrix Algebra	20
1.3 Literature Review	21
1.3.1 Optimality Criteria	22
1.3.2 Structural Properties	23
1.3.3 Adjoined or Unavailable Runs	27
1.4 Research Questions	29
2 Main Effects Only Models	33
2.1 The Class of Competing Designs	34
2.1.1 Augmenting p Runs	34

2.1.2	Missing t Runs	34
2.1.3	Efficiency	35
2.2	Information Matrix for the Altered Design	35
2.2.1	Augmenting p Runs	35
2.2.2	Missing t Runs	36
2.2.3	A Bound on the Determinant	37
2.3	Augmenting OAs by Adjoining Two Runs	45
2.3.1	Optimal Hamming Distance	45
2.3.2	Designs with $N \leq 100$	48
2.3.3	Further Construction Considerations	51
2.4	Augmenting OAs by Adjoining Three Runs	53
2.4.1	Using the Same Hamming distance For All Pairs	55
2.4.2	Further run construction considerations	57
2.5	Augmenting OAs by Adjoining Four Runs	58
2.5.1	Using the Same Hamming Distance For All Pairs	58
2.5.2	Further Run Construction Considerations	60
2.6	Augmenting OAs by Adjoining p Runs	62
2.7	Choosing Runs From Within the OA	64
3	Geometric Isomorphism of Ternary OAs in 18 Runs	71
3.1	Geometric Versus Combinatorial Isomorphism	71
3.2	Complete Enumeration	75
3.2.1	Three Factors	76
3.2.2	Four Factors	82
3.2.3	m Factors	85
4	Models Containing One Linear-by-Linear Interaction	87
4.1	Information Matrix for the OA	88
4.1.1	Inverse of the Information Matrix for the OA	93
4.2	Information Matrix of the Altered Design	94
4.2.1	Augmenting p Runs	94

4.2.2	Missing t Runs	95
4.2.3	The Structure of $\mathbf{AM}^{-1}\mathbf{A}'$	96
4.2.4	A Bound on the Determinants	97
4.3	Entries in $\mathbf{u}_{(ab)}$	106
4.3.1	Linear Terms	107
4.3.2	Quadratic Terms	110
4.3.3	Properties of $ \gamma_{abc} $ and $ \omega_{(ab)c} $	115
4.4	Empirical Results	118
4.4.1	Three Factors	119
4.4.2	More Than Three Factors	128
4.4.3	Model-Robustness Criteria	147
5	Models Containing More Than One Linear-by-Linear Interaction	149
5.1	Information Matrix for the OA	150
5.1.1	Inverse of the Information Matrix for the OA	152
5.1.2	Information Matrix for the Altered Design	153
5.1.3	Bound on ν	154
5.2	Empirical Results For $\nu = 2$	155
5.2.1	Adjoining $p = 1$ Run, $\nu = 2$	156
5.2.2	Missing $t = 1$ Run, $\nu = 2$	158
5.2.3	Adjoining $p = 2$ Runs, $\nu = 2$	159
5.2.4	Missing $t = 2$ Runs, $\nu = 2$	160
5.3	Model-Robustness Vs. Model-Discrimination	161
6	Concluding Remarks	165
6.1	Review	165
6.2	Possible Future Research	166
6.2.1	Q_B Criterion	167
6.2.2	Pure Error	168
6.2.3	Compound Criteria	169

Appendix	171
Bibliography	177

List of Tables

1.1.1 Data for the punching experiment	2
1.1.2 Two representations of \mathbf{R} for the data in Table 1.1.1	4
1.1.3 Polynomial contrast coding for two-level and three-level factors . . .	6
1.1.4 Three design matrices with 3 three-level factors in 14 runs	12
1.1.5 A -, D -, and E -criteria for each of the designs in Table 1.1.4	13
1.1.6 Two designs isomorphic to \mathbf{R}_1 from Table 1.1.4	17
2.2.1 $\text{OA}[32, 2^5 \times 4^7]$	41
2.2.2 Each plausible set of (d_1, d_2, d_3) for an $\text{OA}[48, 2^1 \times 3^1 \times 4^2]$	42
2.2.3 $\text{OA}[48, 2^1 \times 3^1 \times 4^2]$	44
2.3.1 Optimal $p = 2$ runs to append to $\text{OA}[32, 2^4 \times 4^4]$	47
2.3.2 Optimal $p = 2$ runs to append to $\text{OA}[32, 2^4 \times 4^4]$	47
2.3.3 $\text{OA}[32, 4^8 \times 8^1]$ from Kuhfeld (2006)	49
2.3.4 Expansive replacement of an 8-level factor	50
2.3.5 Expansive replacement of a 4-level factor	50
2.3.6 Best $ q $ when $p = 2$ for all OAs with $N \leq 100$	50
2.3.7 Optimal (d_1, d_2) for $(s_1, s_2) = (2, 4)$ and a select set of (m_1, m_2) . .	52
2.4.1 Number of designs where we can extend a pair of runs with $\mathbf{a}_1 \cdot \mathbf{a}'_2 = 0$ to a triple of runs in which each pair has $\mathbf{a}_x \cdot \mathbf{a}'_y = 0$	54
2.4.2 Hamming distance vectors when $p = 3$	55
2.5.1 Hamming distance vectors when $p = 4$	59
2.7.1 Distribution of Hamming distances between all pairs of runs for each of the 4 $\text{OA}[18, 3^3]$ combinatorial isomorphism classes	65

2.7.2 Distribution of Hamming distances between all pairs of runs for each of the 12 $\text{OA}[18, 3^4]$ combinatorial isomorphism classes	67
2.7.3 $\text{OA}[18, 3^4]$ from combinatorial isomorphism class 12	68
2.7.4 Distribution of Hamming distances between all pairs of runs for each of the 10 $\text{OA}[18, 3^5]$ combinatorial isomorphism classes	69
2.7.5 Distribution of Hamming distances between all pairs of runs for each of the 8 $\text{OA}[18, 3^6]$ combinatorial isomorphism classes	69
2.7.6 Distribution of Hamming distances between all pairs of runs for each of the 3 $\text{OA}[18, 3^7]$ combinatorial isomorphism classes	69
3.1.1 Two $\text{OA}[18, 3^3]$ s from combinatorial isomorphism class 1	73
3.1.2 Efficiency [†] of OA minus 2 run designs when one linear-by-linear interaction is included in the model for each of the $\text{OA}[18, 3^3]$ s in Table 3.1.1	73
3.1.3 Efficiency [†] of OA plus 2 run designs when one linear-by-linear inter- action is included in the model for each of the $\text{OA}[18, 3^3]$ s in Table 3.1.1	74
3.2.1 Contingency table of levels of \mathbf{c}_3 vs. pairs of levels in $[\mathbf{c}_1 \ \mathbf{c}_2]$	77
3.2.2 Squares within Table 3.2.1	77
3.2.3 Squares within Table 3.2.2 superimposed	78
3.2.4 All valid squares	78
3.2.5 \mathbf{c}_3 remains unchanged after reversing its levels	79
3.2.6 \mathbf{c}_3 changes after reversing its levels	80
3.2.7 Append \mathbf{c}_4 to a three-factor OA	83
3.2.8 Contingency table of levels of \mathbf{c}_4 (either \mathbf{c}_{4_1} or \mathbf{c}_{4_2}) vs. pairs of levels in $[\mathbf{c}_1 \ \mathbf{c}_2]$	83
3.2.9 Number of geometric isomorphism classes for $\text{OA}[18, 3^m]$ s	86
4.2.1 $\text{OA}[18, 3^4]$ s from geometric isomorphism class 3	100
4.3.1 Consider the number of triples without the middle level	108
4.3.2 Effect of reversing levels on the value of γ_{abc}	109
4.3.3 Values of $ \gamma_{abc} $ for each $\text{OA}[18, 3^3]$ geometric isomorphism class . .	110

4.3.4 Calculation of γ_{abc} and $\omega_{(ab)c}$ for an $\text{OA}[18, 3^3]$ from the first geometric isomorphism class	112
4.3.5 Effect of reversing levels on the value of $\omega_{(12)3}$	113
4.3.6 $ \gamma_{abc} $ and $ \omega_{(ab)c} $ for each $\text{OA}[18, 3^3]$ geometric isomorphism class . .	114
4.3.7 Effect on γ_{abc} and $\omega_{(ab)c}$ of reversing levels of factors	115
4.4.1 Subset of results from Table 4.3.6	121
4.4.2 $ \mathbf{a}_{ME_i} \mathbf{u}_{(ab)} $ for each of the classes in Table 4.4.1	121
4.4.3 Optimal $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 1 run designs	129
4.4.4 Best average $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 1 run designs	130
4.4.5 Optimal $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 1 run designs	131
4.4.6 Best average $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 1 run designs	135
4.4.7 Optimal $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 2 runs designs	139
4.4.8 Best average $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 2 runs designs	139
4.4.9 Optimal $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 2 runs designs	140
4.4.10 Best average $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 2 runs designs	144
5.2.1 Optimal $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 1 run designs	157
5.2.2 Best average $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 1 run designs	157
5.2.3 Optimal $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 1 run designs	158
5.2.4 Best average $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 1 run designs	158
5.2.5 Optimal $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 2 runs designs	159
5.2.6 Best average $ \mathbf{M}_A $ for $\text{OA}[18, 3^m]$ plus 2 runs designs	159
5.2.7 Optimal $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 2 runs designs	160
5.2.8 Best average $ \mathbf{M}_B $ for $\text{OA}[18, 3^m]$ minus 2 runs designs	160
5.3.1 $\text{OA}[18, 3^6]$ from the 85th geometric isomorphism class	161
A1 Optimal (d_1, d_2) for $(s_1, s_2) = (2, 3)$	171
A2 Optimal (d_1, d_2) for $(s_1, s_2) = (2, 5)$	172
A3 Optimal (d_1, d_2) for $(s_1, s_2) = (3, 4)$	172
A4 Optimal (d_1, d_2) for $(s_1, s_2) = (3, 5)$	172
A5 All possible \mathbf{c}_3	173

E1	Geometric isomorphism classes for $\text{OA}[18, 3^3]\text{s}$	175
E2	Geometric isomorphism classes for $\text{OA}[18, 3^4]\text{s}$	175
E3	Geometric isomorphism classes for $\text{OA}[18, 3^5]\text{s}$	175
E4	Geometric isomorphism classes for $\text{OA}[18, 3^6]\text{s}$	175
E5	Geometric isomorphism classes for $\text{OA}[18, 3^7]\text{s}$	175

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, $0 \leq \nu \leq \binom{m}{2}$
v	Total number of terms in a model, $v = \alpha + \nu$
\mathbf{R}	Design matrix of order $n \times m$
\mathbf{r}_i	i th row of \mathbf{R} ; row vector of level combinations of length m , $1 \leq i \leq n$
r_{ij}	Entry in the i th row and j th column of \mathbf{R} ; j th 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 $(s - 1) \times s$
$\mathbf{P}_s(l)$	i th column of \mathbf{P}_s where $i = l + 1$, $0 \leq l \leq s - 1$
\mathbf{p}_{sh}	h th row of \mathbf{P}_s , $1 \leq h \leq s - 1$
\mathbf{p}_{s1}	\mathbf{p}_{s1} , the first row of \mathbf{P}_s
\mathbf{X}	Model matrix of order $n \times v$
\mathbf{x}_i	The i th row of \mathbf{X} , $1 \leq i \leq n$
F_j	The j th factor, $1 \leq j \leq m$
\mathbf{f}_{jh}	The h th 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 i th row of \mathbf{A}
\mathbf{b}_i	The i th row of \mathbf{B}
q	Compact notation for $\mathbf{a}_i \cdot \mathbf{a}'_j$ or $\mathbf{b}_i \cdot \mathbf{b}'_j$ 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; i th 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 $\text{OA}[18, 3^m]$ 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 $\text{OA}[18, 3^m]$ 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 .

