Journal of Statistical Planning and Inference 155 (2014) 127–134

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

# Design of variable resolution for model selection

C. Devon Lin [a,*], Susanne Morrill [b,1]

[a] Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

[b] Department of Computer Science at University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

ARTICLE INFO

*Article history:*
Received 10 March 2013
Received in revised form 10 June 2014
Accepted 11 June 2014
Available online 15 July 2014

*Keywords:*
Bayesian variable selection
Dantzig selector
Generalized minimum aberration
Hadamard matrix
Interaction
Orthogonal array

ABSTRACT

Designs of variable resolution were introduced for practical situations where interactions arise only within certain factors. Optimal designs of variable resolution based on the minimum $G_2$-aberration criterion and design efficiency are provided. A simulation study is carried out to demonstrate the merit of designs of variable resolution in model selection. Some comments are made on the performance of optimal designs of variable resolution in using certain model selection methods.



## 1. Introduction

A weld repaired cast fatigue experiment (Hunter et al., 1982) was used to study the effects of seven factors on fatigue life of weld repaired casting. The seven factors are (A) initial structure ($\beta$ treat or as received), (B) bead size (large or small), (C) pressure treat (HIP or none), (D) heat treat (solution treat/age or anneal), (E) cooling rate (rapid or slow), (F) polish (mechanical or chemical) and (G) final treat (peen or none). Hunter et al. (1982) identified factors D and F as significant. Hamada and Wu (1992) declared factor F and two-factor interaction FG as significant effects based on an analysis strategy that entertains interactions as well as main effects. Hamada and Wu (1992) showed the model with factor F and two-factor interaction FG provides more accurate predictions. If we put factors A–E into one group and factors F and G into the other group, the important interaction in the model suggested by Hamada and Wu (1992) arises only within factors in one group. This is an example of a practical situation that potentially important interactions arise only within factors in each of disjoint groups. Another example is the experiment presented by Wu and Chen (1992) for a surface-mounting process that consists of two independent steps, epoxy application and component placement, with no interaction between factors across steps. The potentially important interactions are between factors within each step.

For such a situation, Lin (2012) proposed *design of variable resolution*, which consists of groups of columns of higher resolution than the overall resolution of the design. Statistical justifications of this type of designs were provided by Lin (2012) from the estimation and design efficiency perspectives. Such designs provide a good class for finding optimal ones that minimize the contamination of non-negligible two-factor interactions on the estimation of main effects. In addition, they tend to achieve higher design efficiency.

* Corresponding author. Tel.: +1 613 533 2412; fax: +1 613 533 2964.
*E-mail addresses:* cdlin@mast.queensu.ca (C.D. Lin), susannemorrill@gmail.com (S. Morrill).

1 Tel.: +1 604 822 3061; fax: +1 604 822 5485.

http://dx.doi.org/10.1016/j.jspi.2014.06.001
0378-3758/

The purpose of this article is to provide catalog of optimal designs of variable resolution based on the minimum $G_2$-aberration criterion and design efficiency, and examine the performance of such designs in model selection when there is prior knowledge that important interactions arise only among factors within each of disjoint groups. A simulation study is carried out to demonstrate that such optimal designs enable more accurate model selection than optimal designs obtained without using the prior knowledge on important interactions.

Various methods have been proposed to build models for designs of experiments. Chipman et al. (1997) proposed a Bayesian variable selection method for analyzing data from experimental designs with complex aliasing. The method entertains interactions in addition to main effects. We apply this method for model selection when designs of variable resolution are used for planning experiments. The other method considered here is the Dantzig selector (Candes and Tao, 2007) which was introduced for variable selection particularly when the number of factors is much larger than the run size. The Dantzig selector was previously demonstrated as an effective factor screening method by Phoa et al. (2009) and Marley and Woods (2010). For other model selection methods, see Hamada and Wu (1992), Lewis and Dean (2001), Yuan et al. (2007), Phoa (2012, 2014) and references therein.

The article is organized as follows. Section 2 reviews designs of variable resolution and design criteria, and provides catalogs of 16-run optimal designs of variable resolution with the overall resolution III and higher, 24-run optimal designs of variable resolution with the overall resolution IV, all with respect to the minimum $G_2$-aberration criterion and design efficiency. Section 2 also provides the optimal 24-run designs of variable resolution with the overall resolution III in a restricted class. Section 3 presents a simulation study and summarizes the results. Section 4 provides concluding remarks.

## 2. Design of variable resolution and design catalog

We consider an experiment of $n$ runs for $p$ factors each at two levels. These $p$ factors can be divided into $k$ groups with $p_i$ factors in the $i$th group ($i = 1, \ldots, k$). Suppose that the experiment has prior knowledge that two-factor interactions arise only within factors in each of $k$ groups. Three-factor and higher order interactions are assumed to be negligible. Let $d = (d_{ij})$ with entries $d_{ij} = \pm 1$ be the design matrix and $D_i$ be the design matrix for the $i$th group of factors. That is, $d = (D_1, \ldots, D_k)$.

### 2.1. Design of variable resolution

A design $d$ of $n$ runs for $p$ factors can be represented by a set of $p$ columns $\{d_1, \ldots, d_p\}$. For any $m$-subset $s = \{d_{j_1}, \ldots, d_{j_m}\}$ of $d$, Deng and Tang (1999) defined

$$J(s) = J(d_{j_1}, \ldots, d_{j_m}) = \left|\sum_{i=1}^{n} d_{ij_1} \cdots d_{ij_m}\right|.$$

Let $r$ be the smallest integer such that $\max_{|s|=r} J(s) > 0$, where the maximization is over all the $r$-subsets of a design $d$. Let $\mathrm{D}(n, p, r)$ represent a two-level design of $n$ runs, $p$ factors and resolution $r$. Lin (2012) defined a $\mathrm{D}(n, p, r)$ to be a design of variable resolution if its columns can be partitioned into $k$ groups with the $i$th group $D_i$ being a $\mathrm{D}(n, p_i, r_i)$ which satisfies (i) $r < r_i \le p_i + 1$ for $i = 1, \ldots, k-1$; and (ii) either $r < r_k \le p_k + 1$ or $r_k = r$.

### 2.2. Design catalog

We consider searching for optimal designs of variable resolution with respect to the minimum $G_2$-aberration criterion and design efficiency. The minimum $G_2$-aberration criterion aims to find designs of variable resolution that sequentially minimizes $C_3$, $C_4$ and $C_5$ given by

$$C_3 = B_3 + 2\sum_{\substack{i=1,\\ p_i>2}}^{k} B_{(i,3)} - \sum_{\substack{k\ge 3,\\ i\ne j\ne l}} B_{(i,1)(j,1)(l,1)}, \tag{1}$$

$$C_4 = \sum_{\substack{i=1,\\ p_i\ge 3}}^{k} \Big\{\sum_{i'\ne i} B_{(i',1)(i,3)} + 4B_{(i,4)} I_{(p_i>3)}\Big\}, \tag{2}$$

$$C_5 = \sum_{\substack{i=1,\\ p_i\ge 4}}^{k} \Big[\sum_{i'\ne i} B_{(i',1)(i,4)} + \{5B_{(i,4)} + (p_i-3)B_{(i,3)}\} I_{(p_i>4)} + B_{(i,3)} I_{(p_i=4)}\Big], \tag{3}$$

where

$$B_3 = n^{-2} \sum_{c_1,c_2,c_3\in(D_1,\ldots,D_k)} J(c_1,c_2,c_3)^2, \qquad B_{(i,1)(j,1)(l,1)} = n^{-2} \sum_{\substack{c_1\in D_i, c_2\in D_j,\\ c_3\in D_l}} J(c_1,c_2,c_3)^2,$$

$$B_{(i',1)(i,j)} = n^{-2} \sum_{\substack{c_1\in D_{i'},\\ \{c_2,\ldots,c_{j+1}\}\in D_i}} J(c_1,\ldots,c_{j+1})^2, \qquad B_{(i,j)} = n^{-2} \sum_{\{c_1,\ldots,c_j\}\in D_i} J(c_1,\ldots,c_j)^2, \tag{4}$$