Journal of Statistical Planning and Inference 139 (2009) 478 – 485

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

A hybrid SAGA algorithm for the construction of $E(s^2)$-optimal cyclic supersaturated designs

Christos Koukouvinos*[,1], Kalliopi Mylona[2], Dimitris E. Simos[1]

Department of Mathematics, National Technical University of Athens, Zografou Campus, Zografou, 15773 Athens, Greece

ARTICLE INFO

Article history:
Received 4 January 2007
Received in revised form
16 March 2008
Accepted 14 May 2008
Available online 22 May 2008

MSC:
62K10
62K15
90C27
90C59

Keywords:
Cyclic generator
$E(s^2)$-optimality
Factorial designs
Genetic algorithms
Hybrid
Metaheuristics
Simulated annealing
Supersaturated designs

ABSTRACT

In this paper, we propose a hybrid simulated annealing genetic algorithm (SAGA) for generating cyclic structured supersaturated designs. The hybrid SAGA combines features such as the power of the GA and the speed of a local optimizer such as SA, merging the previous metaheuristics into a powerful hybrid optimization algorithm. This class of hybrid metaheuristics enabled us to build supersaturated designs for $q = 2, 3, \ldots, 14$ generators. Comparisons are made with previous works and it is shown that the hybrid SAGA is a powerful tool for the construction of $E(s^2)$-optimal supersaturated designs.

1. Introduction

Supersaturated design is a factorial design in which the number of experimental runs n is lower than the number of factors m, that is $n \leqslant m$. For each factor of a two-level supersaturated design there are two possible settings known as levels, which can be coded as ± 1. Any combination of the levels of all factors under consideration is called a treatment combination. Let $\mathbf{X} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m]$ be the design matrix of the experiment in which, each row represents the n treatment combinations and each column gives the sequence of factor levels. For each factor, both level values are of equal interest and each experimental result should have equal influence. Thus we consider designs with the equal occurrence property, where all columns consist of $n/2$ elements equal to 1 and $n/2$ elements equal to -1, when n is even. The designs with the equal occurrence property are called balanced designs. The last row of the designs presented in this paper is a row of $+1$'s. If we omit this row, we obtain designs for odd number of rows, with the number of $+1$'s to be one less than the number of -1's in each column. These designs are non-balanced designs.

* Corresponding author. Tel.: +30 210 7721706.
E-mail addresses: ckoukouv@math.ntua.gr (C. Koukouvinos), kmylona@central.ntua.gr (K. Mylona), dsimos@math.ntua.gr (D.E. Simos).

[1] This research was financially supported from the General Secretariat of Research and Technology by a Grant PENED 03ED740.
[2] This research was financially supported by Greek State Scholarships Foundation (IKY).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.05.006

Industrial experiments are generally very expensive and usually is needed to look at many factors simultaneously. A supersaturated design can save considerable cost when the number of factors is large and a small number of runs is desired or available. The usefulness of these designs relies upon the realism of effect sparsity (Box and Meyer, 1986), namely, that the number of dominant active factors is small. Therefore for situations where there really is no prior knowledge of the effects of factors, but a strong belief in factor sparsity, and where the aim is to find out if there are any dominant factors and to identify them, experimenters should seriously consider using supersaturated designs. For more details regarding the usage of supersaturated designs see Holcomb et al. (2007) and Gilmour (2006).

2. Optimality criteria for supersaturated designs

Orthogonality between all pairs of columns of the model matrix, which is formed from the design matrix by appending a column of 1's as the first column, is required to estimate all factor effects. This condition cannot be satisfied for all pair of columns in a supersaturated design where $m \geqslant n$. Therefore we try to construct designs as near orthogonal as possible. We present here the three optimality criteria which we applied for the construction and evaluation of supersaturated designs.

$E(s^2)$-criterion: Let s_{ij} be the element in the ith row and jth column of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Booth and Cox (1962) proposed as a criterion for comparing designs the minimization of average of s_{ij}^2, denoted by ave(s^2) or $E(s^2)$, where

$$E(s^2) = \sum_{1 \leqslant i<j \leqslant m} s_{ij}^2 \Big/ \binom{m}{2}. \tag{1}$$

The term s_{ij} measures the degree of non-orthogonality between two factors i and j. If $s_{ij} = 0$, the factors i and j are orthogonal. If n is even but not a multiple of 4 (i.e. $n \equiv 2 (\mathrm{mod}\, 4)$) then s_{ij} cannot be equal to 0. In these cases, factors i and j are called near orthogonal if s_{ij} is close to 0. When $s_{ij} = \pm n$ then $c_i = \pm c_j$ and c_i and c_j are completely depended. Designs with any completely depended factors are usually rejected.

It is known (Nguyen, 1996; Nguyen and Cheng, 2008) that the sum of squares of the elements of $\mathbf{XX}^{\mathrm{T}}$ and $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ reaches the minimum if $\mathbf{XX}^{\mathrm{T}}$ is of the form $(m-x)\mathbf{I}_n + x\mathbf{J}_n$, where $x = -m/(n-1)$ for even n and $-m/n$ for odd n, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{J}_n$ is the $n \times n$ matrix with all its elements equal to 1.

In this case $E(s^2)$ can be shown to be

$$n(m^2 + (n-1)x^2 - mn)/(m(m-1)). \tag{2}$$

This quantity can be used as a lower bound for $E(s^2)$. Nguyen (1996), Tang and Wu (1997) independently showed that

$$E(s^2) \geqslant \frac{n^2(m-n+1)}{(n-1)(m-1)}. \tag{3}$$

This is equal to (2) for even n. The lower bound (3) is attainable when $m = q(n-1)$, where q is a positive integer and $n \equiv 0(\mathrm{mod}\, 4)$. It is also attainable when q is even and $n \equiv 2(\mathrm{mod}\, 4)$. The relation (2) also provides a new lower bound for $E(s^2)$ when n is odd

$$E(s^2) \geqslant \frac{m(n^2+n-1) - n^3}{n(m-1)}. \tag{4}$$

Butler et al. (2001) derived some lower bounds for $E(s^2)$ for supersaturated designs with n runs and $m = q(n-1) + k$ factors ($|k| < n/2$, q positive for $n \equiv 0(\mathrm{mod}\, 4)$, q even for $n \equiv 2(\mathrm{mod}\, 4)$). Recently, Bulutoglu and Cheng (2004) presented some improved lower bounds for $E(s^2)$, which apply to all cases. These bounds were improved by Ryan and Bulutoglu (2007). We note that during our metaheuristic search for the $E(s^2)$-optimal supersaturated design with 182 factors in 14 runs, we identified that there does not exist suitable q in order to use the lower bounds proved by Ryan and Bulutoglu (2007). Thus, we give their theorem with a slight modification in order to adjust it, also to our case.

Notation: For any $x \in \mathbb{R}$, $\lfloor x \rfloor^+ = \max\{0, \lfloor x \rfloor\}$, and $\lceil x \rceil^+ = \max\{0, \lceil x \rceil\}$ where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions, respectively.

Theorem 1. *Suppose m is a positive integer such that $m > n-1$. Then there is a unique positive integer q (which depends on n and m) such that $-2n+2 \leqslant m - q(n-1) < 2n-2$ and $(m+q) \equiv 2(\mathrm{mod}\, 4)$. Let $g(q) = (m+q)^2 n - q^2 n^2 - mn^2$.*

(1) *If $n \equiv 0(\mathrm{mod}\, 4)$, then*

$$E(s^2) \geqslant \begin{cases} \dfrac{g(q) + 2n^2 - 4n}{m(m-1)} & \text{when } |m - q(n-1)| < n-1, \\ \dfrac{g(q) - 2n^2 + 4n + 4n|m-q(n-1)|}{m(m-1)} & \text{when } n-1 < |m-q(n-1)| \leqslant \frac{3}{2}n - 2, \\ \dfrac{g(q) + 4n^2 - 4n}{m(m-1)} & \text{when } |m-q(n-1)| > \frac{3}{2}n - 2. \end{cases}$$