Construction of Search Designs from Orthogonal Arrays

K. Chatterjee, P. Angelopoulos and C. Koukouvinos

Visva-Bharati University, INDIA and National Technical University of Athens, Greece

Search Linear Model

Consider an experiment involving m factors $F_1, F_2, \cdots, F_m$ respectively, where each factor has two different levels 0 and 1. Corresponding to an n run design d, Srivastava (1975), proposed the following Search Linear Model.

$$y = D_1\theta_1 + D_2\theta_2 + \epsilon$$

- $y \rightarrow$ Vector of observations.
- $D_1, D_2 \rightarrow$ Known design matrices.
- $\theta_1, \theta_2 \rightarrow$ Vectors of unknown factorial effects.
- $\epsilon \rightarrow$ Vector of random error components.
- We have prior information that only a few factorial effects belonging to θ_2 are possibly significant (non-negligible).
- No knowledge about which factorial effects?

A Necessary and Sufficient Condition (Srivastava (1975)

- Suppose θ_1 is an $m_1 \times 1$ vector.
- Suppose θ_2 is an $m_2 \times 1$ vector.
- k ($k << m_2/2$) parameters belonging to θ_2 are possibly significant.

Theorem 1. A design d with m factors in n runs will solve our objective in the noiseless case (error variance $\sigma^2 = 0$) if and only if

$$\text{rank}[D_1, D_2^*] = m_1 + 2k$$

- D_2^* is any $n \times 2k$ submatrix of D_2.
- Even in the noisy case (error variance $\sigma^2 > 0$), the above the rank condition remains necessary.

Search Designs from Orthogonal Arrays

- Search designs are considered to be a very useful tool when there is some prior information on the number of possibly non-negligible effects.
- Literature review reveals that most of the work in this field dealt with the cases where there are at most one or two two-factor interaction effects considered non-negligible.
- Moreover, this is the first attempt to provide search designs based on orthogonal arrays.
- The main objective of this presentation is to show how to construct search designs that are capable of estimating the general mean, all the main effects and allow the search and estimation of few two-factor interaction effects based on orthogonal arrays.
- We focus our work in two-level orthogonal arrays with small number of runs which are optimal from the point of view of estimating the general mean and the main effects.

Table 1: The 12-runs Plackett-Burman design

Row	Level combinations	Row	Level combination
1	1 1 1 1 1 1 1 1 1 1 1	7	0 0 0 1 0 0 1 0 1 1 1
2	0 1 0 1 1 1 0 0 0 1 0	8	1 0 0 0 1 0 0 1 0 1 1
3	0 0 1 0 1 1 1 0 0 0 1	9	1 1 0 0 0 1 0 0 1 0 1
4	1 0 0 1 0 1 1 1 0 0 0	10	1 1 1 0 0 0 1 0 0 1 0
5	0 1 0 0 1 0 1 1 1 0 0	11	0 1 1 1 0 0 0 1 0 0 1
6	0 0 1 0 0 1 0 1 1 1 0	12	1 0 1 1 1 0 0 0 1 0 0

Case of 12-runs

- Out of the $\binom{11}{4} = 330$ 12×4 possible subdesigns of the Placket-Burman design, there exists only one non isomorphic orthogonal array that can be used as a search design capable of searching and estimating any one, two or three two-factor interaction effect(s) along with the general mean and all the main effects.
- Moreover, for $k = 2, 3$, the design consist of small number of level combinations and hence is economical.
- If we consider any 12×5 subdesign of the Placket-Burman design, we can obtain a search design which will be capable of searching and estimating any one two-factor interaction effect along with general mean and all the main effects.
- We must mention that there exist only two non isomorphic orthogonal arrays with five columns. Both of them satisfy the rank condition and can be efficiently used as search designs.
- There are two non isomorphic projections of the 12 run Placket-Burman design into six columns. Only one of them satisfies the rank condition of Shrivastava for searching one two-factor interaction and can be used as a search design for exploring one two factor interaction.
- The search design can be obtained by selecting the first 6 columns of the Placket-Burman design.
- There are no other projections of the 12-runs Plackett-Burman design that can serve as search designs for searching non negligible two-factor interactions.

Table 2: A 16×8 orthogonal array based search design

Row	Level combination	Row	Level combination
1	1 1 1 0 0 0 1 1	9	0 1 1 1 1 1 0 0
2	1 1 0 1 0 1 1 0	10	0 1 1 1 0 0 0 0
3	1 1 0 0 1 1 0 1	11	0 1 1 0 0 1 1 1
4	1 1 0 0 1 0 0 0	12	0 1 0 1 1 0 1 1
5	1 0 1 1 1 0 1 0	13	0 0 1 0 1 0 0 1
6	1 0 1 1 0 1 0 1	14	0 0 0 1 1 1 1 1
7	1 0 1 0 1 1 1 0	15	0 0 0 0 0 1 0 0
8	1 0 0 1 0 0 0 1	16	0 0 0 0 0 0 1 0

Case of 16-runs

- There are 11 non isomorphic orthogonal arrays with 5 columns. Three of them satisfy the rank condition when we consider the model with main effects and 2 two-factor interactions.
- Such a search design can be obtained if we consider the first 5 columns of the orthogonal array. No more than 2 two-factor interactions can be searched using 16 run orthogonal arrays.
- There exist 27 non isomorphic orthogonal arrays with 6 columns. Out of the 27 non isomorphic classes 8 of them satisfy the rank condition and can serve as search designs when we consider a model with the general mean, the main effects model with 1 two-factor interaction.
- There exist no 16 run orthogonal arrays with 6 factors that can serve as search designs for searching more than one non-negligible two-factor interaction.
- Similarly for the case with 16 runs and 7 columns there exist 55 non isomorphic orthogonal arrays. 10 non isomorphic classes satisfy the rank condition when considering the general mean , the main effects model and 1 two-factor interactions and can serve as search designs.
- When considering the case with 16 runs and 8 factors there exist 80 non isomorphic orthogonal arrays. Three of them satisfy the rank condition of equation and can be used to search for 1 non negligible two-factor interaction.

Table 3: A 20×8 orthogonal array based search design

Row	Level combination	Row	Level combination
1	1 1 1 1 1 1 1 1 1 1	11	1 0 0 1 1 0 0 0 0 1
2	0 1 0 1 1 1 1 0 0 1	12	0 0 1 1 0 0 0 0 1 0
3	1 0 1 1 1 1 0 0 1 0	13	0 1 1 0 0 0 0 1 0 1
4	0 1 1 1 1 0 0 1 0 0	14	1 1 0 0 0 0 1 0 1 0
5	1 1 1 1 0 0 1 0 0 1	15	1 0 0 0 0 1 0 1 0 1
6	1 1 1 0 0 1 0 0 1 1	16	0 0 0 0 1 0 1 0 1 1
7	1 1 0 0 1 0 0 1 1 0	17	0 0 0 1 0 1 0 1 1 1
8	1 0 0 1 0 0 1 1 0 0	18	0 0 1 0 1 0 1 1 1 1
9	0 0 1 0 0 1 1 0 0 0	19	0 1 0 1 0 1 1 1 1 0
10	0 1 0 0 1 1 0 0 0 0	20	1 0 1 0 1 1 1 1 0 0

Case of 20-runs

- When considering the model with 5 main effects and up to 5 two-factor interactions, there are 4 non isomorphic classes that satisfy the rank condition and can be used as search designs. A search design can be obtained if we select the first five columns.
- No more than 5 two-factor interactions can be identified using 20 runs orthogonal arrays.

Preliminaries on Searching Probabilities

- Keeping in view of the main objective of search designs, it is important to note that these designs should be able to identify the true model with high probability.
- Shirakura, Takahashi and Srivastava (1996) initiated the work on the probability of correct searching for search designs.
- The following notations will be helpful to obtain the probability of correct searching.
- Let U be the set of all sets consisting of k out of $m(m-1)/2$ possible two-factor interaction effects.
- For any $u \in U$, let $M(u)$ be the model containing the general mean, all main effects and k two-factor interaction effects contained in u.
- Also let $S^2(u)$ be the sum of squares due to error for the model $M(u)$.

Computational Procedure

- **Step 1**. For any $u \in U$, generate a vector y_u of n observations assuming that $M(u)$ is the true model.
- **Step 2**. Using the generated observational vector y_u, calculate $S^2(u')$ for all $u' \in U$.
- **Step 3**. Repeat the above steps N (we considered $N = 1000$) times.
- **Step 4**. Obtain the proportion of times $S^2(u) < S^2(u')$ for all $u' \in U/u$.
- Repeat the above steps for all $u \in U$.
- For any $u \in U$, let P_u denote the proportions of times $S^2(u) < S^2(u')$ for all $u' \in U/u$.
- For any search design d, if P_d denotes the probability of correct model identification, then according to Srivastava (1975), the minimum value of P_u provides an estimate of P_d , where the minimum is taken over all possible models for $u \in U$.
- The effect size of the interaction is express through ρ_i, that is defined as $\rho_i = \theta_{2i}/\sigma, 1 \leq i \leq m(m-1)/2$.
- When one two-factor interaction effect is considered in the model ($k = 1$), we use ρ to denote the ρ_i value of the two-factor interaction effect included in the model.

Application of the Algorithm

- Consider a model that includes, apart from the general mean and all main effects, 2 two-factor interaction effects for a design consisting of the first four columns of the design given in Table 1. The searching probabilities are presented below.
- We observe that the searching probabilities are very high even for very small effect sizes.

ρ_1 ρ_2	0.5	1.0	1.5	2.0	2.5	3.0
0.5	0.2560	0.4445	0.4810	0.4840	0.4855	0.4865
1.0	0.4390	0.7750	0.8535	0.8735	0.8795	0.8855
1.5	0.4690	0.8570	0.9715	0.9840	0.9855	0.9865
2.0	0.4845	0.8545	0.9810	0.9975	0.9985	0.9990
2.5	0.4890	0.8680	0.9835	0.9975	1.0000	1.0000
3.0	0.4965	0.8905	0.9845	0.9975	1.0000	1.0000

Searching Probabilities for designs with 12 and 16 runs

Design ρ	0.5	1.0	1.5	2.0	2.5	3.0
d_1_12.5.1	0.295	0.708	0.938	0.994	0.999	0.1
d_2_12.5.1	0.307	0.691	0.915	0.978	0.997	0.999
d_1_16.5.2	0.342	0.935	0.998	1	1	1
d_2_16.5.2	0.133	0.529	0.848	0.962	0.995	1
d_1_16.6.1	0.241	0.662	0.914	0.987	0.999	1
d_2_16.6.1	0.146	0.426	0.669	0.86	0.943	0.97
d_1_16.7.1	0.183	0.555	0.856	0.977	0.994	0.999
d_2_16.7.1	0.104	0.304	0.553	0.791	0.903	0.967
d_1_16.8.1	0.14	0.487	0.795	0.947	0.988	0.996
d_2_16.8.1	0.055	0.198	0.411	0.649	0.816	0.92

discussion

- In this presentation we initiate a research for the use of orthogonal arrays in the correct model identification problem, in the presence of few non-negligible two-factor interactions.
- Search designs are an important class of experimental designs that can very efficiently be used when there is some prior knowledge on the number of non-negligible effects in an experimental plan.
- From the simulation study we conclude that orthogonal arrays provide experimental plans with very high searching probabilities and can be successfully employed in the search problem, with an economy in the run size.
- Furthermore, we have considered cases where the number of significant interactions is up to five.
- In this framework we believe that the use of orthogonal arrays as search designs should be further explored, to include more than two-level designs and even mixed level orthogonal arrays.

References

- Chatterjee, K. and Mukerjee, R. (1986). Some search designs for symmetric and asym-metric factorials. J. Statist. Plann. Inf., 13, 357-363
- Sarkar, A. and Chatterjee, K. (2010). An MEP.2 plan in 3^n factorial experiment and its probability of correct identification.*Journal of Statistical Planning and Inference*, 140, 3531-3539.
- Shirakura, T., Takahashi, T. and Srivastava, J.N. (1996). Searching probabilities for nonzero effects in search designs for the noisy case.*Annals of Statistics*, 24, 2560-2568.
- Srivastava, J. N. (1975). Designs for searching non-negligible effects. In: *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.), 507-519, North-Holland, Amsterdam.