.

A Position-Based Approach for Design and Analysis of Order-of-Addition Experiments

Zack Stokes and Hongquan Xu

Department of Statistics, University of California, Los Angeles, CA 90095

S1. Proofs

Proof of Theorem 1. (i) By Corollary 6 (i) of Xu (2003), we have $W_1 \geq mr(m-r)/n^2$, with equality if and only if each component appears as equally often as possible in every column. When $n = qm+r$, $\mathbf{F}_{n,m}$ contains q $m \times m$ Latin squares, so each level appears in each column q times in the first qm runs of $\mathbf{F}_{n,m}$. Each column of the last r runs of $\mathbf{F}_{n,m}$ contains each level at most once, meaning that the maximum difference in the number of occurrences of each level per column is 1. Thus, $\mathbf{F}_{n,m}$ has minimum $W_1 = mr(m-r)/n^2$ among all possible designs.

(ii) This is a direct result of (i).

(iii) We show that $\mathbf{F}_{n,m}$ is an orthogonal array of weak strength t for all $t \geq 1$. A design is an orthogonal array of weak strength t if all possible level combinations for any t columns appear as equally often as possible (Xu, 2003). From (i) we know that $\mathbf{F}_{n,m}$ has minimum W_1. Since $\mathbf{F}_{m(m-1),m}$ is a COA with the property that every pair of level combinations shows up exactly once, we know that the sub-design $\mathbf{F}_{n,m}$, $n \leq m(m-1)$, contains each pair of level combinations either 0 or 1 times. Since $n \leq m(m-1)$, $\mathbf{F}_{n,m}$ is an orthogonal array of weak strength t for all $t \geq 1$. Hence, by

Theorems 2 and 3 of Xu (2003), design $\mathbf{F}_{n,m}$ has generalized minimum aberration among all possible designs.

(iv) If $n = m(m-1)$, then the claim is true by the mutual orthogonality of the Latin squares derived in Step 1 of Algorithm 1. The two COA properties of a design given in Definition 1 are invariant with respect to column permutation. Therefore, each $\boldsymbol{C}_i, i = 1, \ldots, (m-2)!$, is also a COA$(m(m-1), m)$ and for any $n > 0$ such that $n/(m(m-1)) = \lambda$ is an integer, concatenating the λ COA$(m(m-1), m)$ designs produces a COA(n, m). $\square$

Proof of Theorem 2. We prove the claim for any COA(n, m) as the full design $\mathbf{F}_m$ is a special case with $n = m!$. For a point $\boldsymbol{b} = (b_1, \ldots, b_m) \in \Omega$, the vector of regression functions under the quadratic model (3.4) is

$$\boldsymbol{f}(\boldsymbol{b}) = (1, p_1(b_1), \ldots, p_1(b_{m-1}), p_2(b_1), \ldots, p_2(b_{m-1}))^{\mathrm{T}},$$

with the first m terms being the regression functions under the first-order model (3.3). For any COA(n, m) the information matrix under the quadratic model and its inverse take the form

$$\begin{aligned}
\boldsymbol{M}(\boldsymbol{\xi}) &= \begin{bmatrix} 1 & \mathbf{0}_{1\times(m-1)} & \mathbf{0}_{1\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \delta\mathbf{J}_{(m-1)} + (1-\delta)\mathbf{I}_{(m-1)} & \mathbf{0}_{(m-1)\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \mathbf{0}_{(m-1)\times(m-1)} & \delta\mathbf{J}_{(m-1)} + (1-\delta)\mathbf{I}_{(m-1)} \end{bmatrix}, \\
\boldsymbol{M}(\boldsymbol{\xi})^{-1} &= \begin{bmatrix} 1 & \mathbf{0}_{1\times(m-1)} & \mathbf{0}_{1\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & -(m\delta)^{-1}(\mathbf{J}_{(m-1)} + \mathbf{I}_{(m-1)}) & \mathbf{0}_{(m-1)\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \mathbf{0}_{(m-1)\times(m-1)} & -(m\delta)^{-1}(\mathbf{J}_{(m-1)} + \mathbf{I}_{(m-1)}) \end{bmatrix},
\end{aligned}$$

where $\delta = -1/(m-1)$, $\mathbf{J}_k$ is a $k \times k$ matrix of 1's, and $\mathbf{I}_k$ is the $k \times k$

identity matrix. The top left 2×2 submatrix in both cases is the equivalent information matrix and inverse under the first-order model. Now we can apply the checking condition (4.2) provided by the equivalence theorem and exploit the properties of the orthogonal polynomial contrasts (3.2).

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{b})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{\xi})^{-1}\boldsymbol{f}(\boldsymbol{b}) &= 1 - \frac{2}{m\delta}\sum_{k=1}^{m-1} p_1^2(b_k) - \frac{2}{m\delta}\sum_{k=1}^{m-2}\sum_{l>k}^{m-1} p_1(b_k)p_1(b_l) \\ &\quad - \frac{2}{m\delta}\sum_{k=1}^{m-1} p_2^2(b_k) - \frac{2}{m\delta}\sum_{k=1}^{m-2}\sum_{l>k}^{m-1} p_2(b_k)p_2(b_l).\end{aligned}$$

Because $\boldsymbol{b} = (b_1, \ldots, b_m)$ is a permutation of $\{1, \ldots, m\}$, using (3.2) and some algebra, for $j = 1, 2$, we have

$$2\sum_{k=1}^{m-1} p_j^2(b_k) + 2\sum_{k=1}^{m-2}\sum_{l>k}^{m-1} p_j(b_k)p_j(b_l) = m.$$

Therefore,

$$\boldsymbol{f}(\boldsymbol{b})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{\xi})^{-1}\boldsymbol{f}(\boldsymbol{b}) = 1 - \frac{1}{\delta} - \frac{1}{\delta} = 1 + (m-1) + (m-1) = 2m - 1.$$

As the quadratic model has $p = 2m - 1$ parameters, the equality in (4.2) holds for any $\boldsymbol{b} \in \Omega$. By the equivalence theorem, every COA(n, m) is D-optimal for the quadratic model. The proof of D-optimality for the first-order model is simpler. This completes the proof. $\square$

Proof of Theorem 3. For the full design $\mathbf{F}_m$ and the second-order position model (3.5), let $\boldsymbol{X}$ be the $n \times p$ model matrix and $\boldsymbol{M} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}/n$ be the $p \times p$ information matrix with $n = m!$ and $p = (m-1)(m+2)/2$. Let $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ be the hat matrix. To prove the D-optimality, we need to show that the equality in (4.2) holds for any $\boldsymbol{b} \in \Omega$. For any $\boldsymbol{b} \in \Omega$, by the standard linear model theory, the variance of the fitted value when $\boldsymbol{x} = \boldsymbol{b}$ is $Var(\hat{y}(\boldsymbol{x})) = \sigma^2 \boldsymbol{f}(\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{b}) = n^{-1}\sigma^2\boldsymbol{f}(\boldsymbol{b})^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{f}(\boldsymbol{b})$. Because every $\boldsymbol{b}$ is a row of the full design $\mathbf{F}_m$, it is sufficient to show that each of the diagonal elements of the hat matrix $\boldsymbol{H}$ is p/n.

To do this, we consider the extended second-order model

$$y = \beta_0 + \sum_{k=1}^{m} p_1(b_k)\beta_k + \sum_{k=1}^{m} p_2(b_k)\beta_{kk} + \sum_{1\le k<l\le m} p_1(b_k)p_1(b_l)\beta_{kl} + \varepsilon, \quad \text{(S1.1)}$$

which includes all m first-order, m pure quadratic and $m(m-1)/2$ bilinear (or interaction) terms. The extended second-order model has $q = (m+1)(m+2)/2$ parameters. Let $\boldsymbol{Z}$ be the $n \times q$ model matrix for the full design $\mathbf{F}_m$. Due to the constraints on the orthogonal polynomials and the fact that each row is a permutation, $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ has rank p and its inverse does not exist, so we consider its Moore-Penrose generalized inverse $(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-}$. By the standard linear model theory (Seber and Lee, 2003), the projection matrix $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-}\boldsymbol{Z}^{\mathrm{T}}$ of the extended second-order model (S1.1) is identical to the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ of the second-order position model (3.5) because columns of $\boldsymbol{Z}$ and $\boldsymbol{X}$ span the same linear space. Under the extended model (S1.1), all variables are exchangeable; therefore, the variances of the fitted values are the same for all rows of the full design. This is equivalent to saying that the diagonal elements of projection matrix $\boldsymbol{P}$ are the same. Since $\boldsymbol{P}$ is idempotent and has rank p, its trace is equal to its rank. Therefore, all of the diagonal elements of $\boldsymbol{P}$, and hence $\boldsymbol{H}$, are equal to p/n. This completes the proof. $\square$

S2. A-optimality

Using a popular metheuristic algorithm, Differential Evolution (Storn and Price 1997; Chakraborty 2008), we have found nearly A-optimal designs for the position models for several values of m. Table 6 shows the relative efficiency of $\mathbf{F}_m$ to these designs and indicates that the full design is indeed sub-optimal in most cases, with its efficiency growing worse as m increases. Furthermore, we have found that A-optimal designs under our models depend on specifically which component effects are removed from the model to make it estimable. Since our decision to remove the effect of component m was in large part arbitrary, we hope to explore this phenomenon further and produce A-optimal designs which are robust to this choice.

Table 6: Relative A-efficiency of $\mathbf{F}_m$ to near-optimal designs.

m	first-order	quadratic	second-order
3	0.951	1	0.951
4	0.909	0.987	0.885
5	0.879	0.934	0.812

S3. Additional Permutation Robustness Results

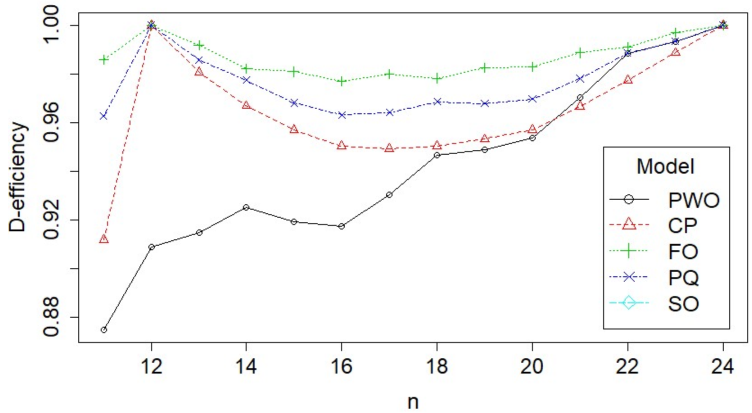
For each sample size n in Figure 2 up to five unique designs are necessary to obtain the largest D-efficiency under each model. Instead, we could consider a single design $\mathbf{F}^*_{n,m}$ for each combination of n and m that is derived from maximizing the geometric mean efficiency of the estimable models, akin to those presented in Table 5. The results of this analysis are presented in Figure 3. Generally, we see that many of the efficiencies are on par with what we observed when maximizing the value for each model individually. A notable exception is the efficiency of the maximal geometric mean design under the PWO model, which for some combinations of n and m is slightly lower than the efficiency of the design that solely optimizes performance for the PWO model; see Figure 2.

While Figures 2 and 3 substantiate our claim that our algorithm produces efficient designs, they do not inherently show how much there is to gain or lose by selection of permutations in Algorithm 1. They also do not consider the effect of level or $\boldsymbol{C}_i$ permutations. To remedy this, Table 7 gives the maximal D_{PWO} and D_{SO} values obtained through a brute force search over all choices of the three permutations and the improvement relative to the values of the $\mathbf{F}_{n,m}$ designs in Table 5. We do not include the other efficiencies since many of the designs are optimal under the CP, first-order and quadratic models and show minimal improvement with column permutations. The notation $\mathbf{F}^{+}_{n,m}$ is used to represent the resulting designs. Δ_{PWO} and Δ_{SO} give the difference in D-efficiency under the specified model between the best permuted design and the design $\mathbf{F}_{n,m}$ given in Table 5. Cases for which the same set of permutations generate the best design for both models are indicated by †.

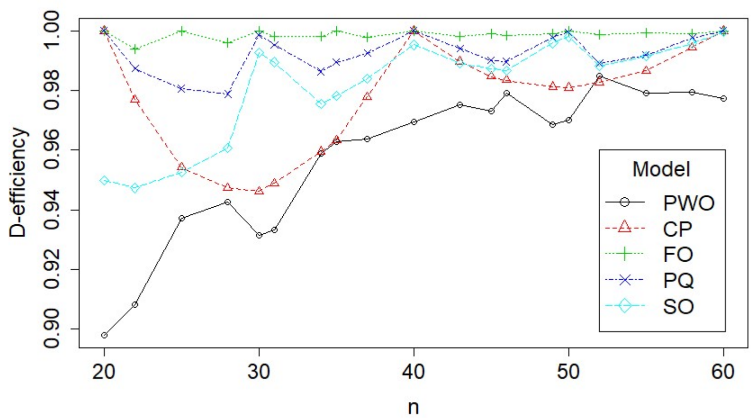
Table 7: Maximal D-efficiencies of designs for the PWO and second-order models under permutation.

			D-efficiency		Change	
n	m	Design	D_{PWO}	D_{SO}	Δ_{PWO}	Δ_{SO}
†12	4	$\mathbf{F}^{+}_{12,4}$	0.909	1	0	0
†16	4	$\mathbf{F}^{+}_{16,4}$	0.917	0.953	0	0
†20	4	$\mathbf{F}^{+}_{20,4}$	0.954	0.961	0	0
20	5	$\mathbf{F}^{+}_{20,5}$	0.898	0.959	0.898	0
24	5	$\mathbf{F}^{+}_{24,5}$	0.926	0.961	0.381	0.012
40	5	$\mathbf{F}^{+}_{40,5}$	0.969	0.999	0.08	0
†60	5	$\mathbf{F}^{+}_{60,5}$	0.977	0.999	0	0.013

(a) $m = 4$



(b) $m = 5$



(c) $m = 7$

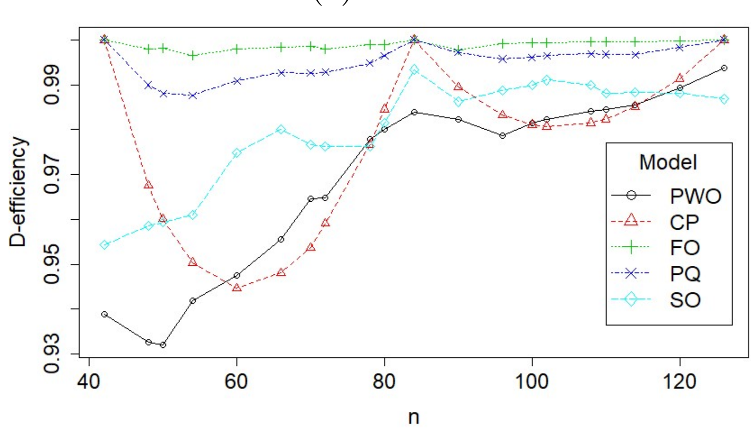


Figure 3: The D-efficiency of $\mathbf{F}^*_{n,m}$ which maximizes the geometric mean efficiency for variable run sizes for (a) $m = 4$, (b) $m = 5$, and (c) $m = 7$.

There are several interesting observations we can make from Table 7. First, as expected, we see that the D_{PWO} values of our designs in situations that were previously troubling greatly improve with this manipulation, closing the gap between our designs and Voelkel's. We also see minor improvements in the D_{SO} values for some cases with $m = 5$. Most importantly, considering all three types of permutations did not lead to designs that substantially outperform the $\mathbf{F}^*_{n,m}$ designs that only considered column permutations.

We are also interested in knowing the worst efficiency attainable under our algorithm. Table 8 summarizes this effect in the same manner as before, this time using $\mathbf{F}^-_{n,m}$ to denote the worst design. In this case we see that compared to the small gains in D_{SO} of Table 7, the loss of efficiency due to poor selection of permutations is relatively large when $m = 5$. On the other hand, the minimal value of D_{PWO} is often not a substantial decrease from the values found without permutations. For larger m, the brute force approach is limited by the same combinatorial explosion that motivates order-of-addition designs.

Table 8: Minimal D-efficiencies of designs for the PWO and second-order models under permutation.

			D-efficiency		Change	
n	m	Design	D_{PWO}	D_{SO}	Δ_{PWO}	Δ_{SO}
†12	4	$\mathbf{F}^-_{12,4}$	0.909	1	0	0
†16	4	$\mathbf{F}^-_{16,4}$	0.917	0.953	0	0
†20	4	$\mathbf{F}^-_{20,4}$	0.954	0.961	0	0
20	5	$\mathbf{F}^-_{20,5}$	0	0.428	0	-0.531
24	5	$\mathbf{F}^-_{24,5}$	0.545	0.581	0	-0.368
†40	5	$\mathbf{F}^-_{40,5}$	0.784	0.735	-0.105	-0.264
†60	5	$\mathbf{F}^-_{60,5}$	0.861	0.875	-0.116	-0.111

S4. Additional Model Misspecification Results

We fix $(n, m) = (12, 4)$ and consider the PWO and CP models. We define our confidence that the PWO model is indeed the true model as $\alpha \in [0, 1]$ and similarly our confidence in the CP model as $1 - \alpha$. We then use the discrete form of Differential Evolution (Cuevas et al. 2011) to find 12-run designs for various values of α that maximize the desirability function given by

$$\bar{G}^{(\alpha)}(\boldsymbol{\xi}) = D^{\alpha}_{PWO}(\boldsymbol{\xi}) D^{1-\alpha}_{CP}(\boldsymbol{\xi}). \tag{S4.1}$$

Differential Evolution is inspired by principles of natural selection, mutation and genetic crossover and has been shown to work well for finding optimal designs while only depending on the choice of a few parameters (Paredes-García and Castaño-Tostado, 2017). We implement it using the R package DEoptim (Ardia et al. 2011) and after choosing appropriate settings for the parameters, it is able to quickly locate the global maximum for all $\alpha = 0, 0.1, \ldots, 1$.

Figure 4 shows the unweighted efficiencies of the designs found by the search. In this plot, α gives our confidence in the PWO model. When $\alpha = 0$, we assume that the data follow the CP model with high confidence $(1 - \alpha = 1)$. In this case the algorithm finds a design that is isomorphic to $\mathbf{F}_{12,4}$ from Algorithm 1, with the efficiencies matching our results from Table 5. Designs with this property are represented in the plot by the "F" symbol. As we then increase α and split our confidence between this model and the PWO model, $\mathbf{F}_{12,4}$ continues to have maximal $\bar{G}^{(\alpha)}$. In fact, it is not until we increase α from 0.7 to 0.8 that this changes. For $\alpha \geq 0.8$ the algorithm finds a design equivalent to Voelkel.12a. These designs are denoted with the "V" symbol.

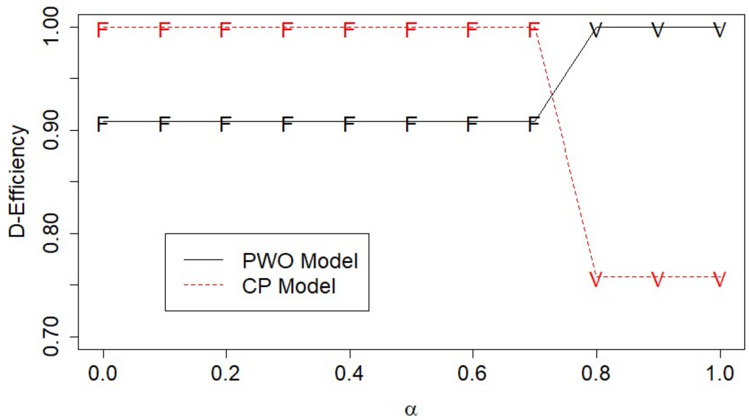


Figure 4: D-efficiencies of designs that maximize (S4.1).

This result demonstrates that our designs are indeed robust to model misspecification in this case. In addition to the model's form, there is also the underlying assumption made by each of these models as to whether the relative positions or absolute positions are important in determining the response. By demonstrating that our designs are robust to model misspecification under this pair of models, we have also shown that they are robust to this assumption.

References

Ardia, D., Mullen, K.M., Ulrich, J., Peterson, B.G. (2011). DEoptim. Package version 2.2-5.

Chakraborty, U. K. (2008). *Advances in Differential Evolution.* Springer.

Cuevas, E., Zaldivar, D., Pérez-Cisneros, M., and Ramírez-Ortegón, M. (2011). Circle detection using discrete differential evolution optimization. *Pattern Analysis and Applications* **14**(1), 93–107.

Paredes-García, W.J. and Castaño-Tostado, E. (2017). Differential evolutionary algorithm in the construction process of optimal experimental designs. *Communications in Statistics – Simulation and Computation* **46**(10), 7733–7743.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis*, 3rd edition, Wiley.

Storn, R. and Price, K. (1997). Differential evolution–A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11**, 341–359.

Xu, H. (2003). Minimum moment aberration for nonregular designs and supersaturated designs. *Statistica Sinica* **13**, 691–708.