

Design Filtering and Reconstruction: A Procedure for Sequentially Locating D-Optimal Design Measures

Mary Paschal Iwundu

Department of Mathematics and Statistics, Faculty of Science, University of Port Harcourt, Choba, Nigeria

Abstract The presence of non-optimal design points in an experimental design measure greatly affects the convergence of a search algorithm to a desired optimum. Filtering and reconstruction is presented as a viable procedure for sequentially locating D-optimal design measures. The method effectively improves experimental design in the search for an optimal design measure. While the procedure is identical to the Wynn's sequential algorithm for constructing D-optimal designs, filtering and reconstruction addresses situations where outlying non-optimal design points had been admitted into the design possibly either by the creation of a poor initial design or by its influence on the next design point(s). By the method, outlying non-optimal design points are removed and the design reconstructed, thus resulting in a significant improvement on the determinant value of information matrix. Approximate solution has been obtained in the construction of D-optimal design measure for a two-dimensional non-interaction polynomial model defined on an irregular continuous design whose boundary is quadrilateral with vertices (2,2), (-1,1), (1,-1) and (-1,-1). Furthermore, bounds have been established for the determinant value of information matrix for each sequentially generated design. The bounds reveal that the D-optimal measure generated by the search procedure is very close to the true unknown D-optimal design measure.

Keywords Filtering, Reconstruction, D-optimal, Sequential algorithm, Non-optimal points, Bounds

1. Introduction

The consideration for construction of optimal design measures has gained top interests among researchers since from early works of Wynn (1970), Mitchell and Miller (1970), Fedorov (1972), Mitchell (1974), etc. In all practical terms, an optimal design enhances efficiency. Choosing an optimal experimental design for a full parameter regression, defined on a regular geometric area, has been given much attention. However difficulty arises with complexity of model or the design region or both. Sometimes it maybe needful to construct optimal designs for models that exist with improper polynomial regression functions and which may be defined on non-regular geometric areas. For whatever settings, the need for optimal designs cannot be played down. A documentary of the usefulness of optimum experimental design has been given by Atkinson (1996). Two very important and frequently used methods of constructing optimal designs are the sequential and exchange methods. Each of these methods has been in one or more of the early works mentioned above. Fundamental among them is the generation of D-optimal design measures. By definition, a design measure is a probability measure ξ defined on the design space \tilde{X} that is a closed compact set in

a Euclidean space of a particular dimension. Moreover, ξ is a member of the set \mathcal{D} , of all measures defined on the Borel Field \mathcal{B} containing all one-point sets such that

$$\int_{\tilde{X}} \xi(dx) = 1.$$

If f_1, f_2, \dots, f_k are k linearly independent functions defined on the design region \tilde{X} , at each point x in \tilde{X} , a random variable $E(Y_x)$ is defined and is such that $E(Y_x) = \theta' f(x)$, where $f(x)$ represents the $k \times 1$ column vector of functions f_i evaluated at x and θ represents the $k \times 1$ column vector of unknown estimable parameters.

For the measure ξ on \tilde{X}

$$m_{ij}(\xi) = \int_{\tilde{X}} f_i(x) f_j(x) \xi(dx).$$

Letting $M(\xi)$ be the $k \times k$ information matrix whose $(ij)^{th}$ entry is $m_{ij}(\xi)$, a discrete design measure may be formed by attaching a mass of $\frac{1}{N}$ to each point of the discrete design such that

$$\frac{X_N' X_N}{N} = M(\xi_N).$$

By discrete design measure, we refer to a design comprising of N points x_1, x_2, \dots, x_N in \tilde{X} , not necessarily distinct.

A design measure ξ^* is said to be D-optimal if

$$\det\{M(\xi^*)\} = \max_{\xi} \det\{M(\xi)\}; \quad \xi \in \mathcal{D}$$

According to Kiefer and Wolfowitz (1960)

$$\det\{M(\xi^*)\} = \max_{\xi} \det\{M(\xi)\}; \quad \xi \in \mathcal{D}$$

* Corresponding author:

mary.iwundu@uniport.edu.ng (Mary Paschal Iwundu)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

is equivalent to

$$\min_{\xi} \max_{\underline{x}} \det(\underline{x}, \xi) = \max_{\underline{x}} \det(\underline{x}, \xi^*) ; \xi \in \mathcal{D}, \underline{x} \in \tilde{X} \\ = k$$

where k is the number of model parameters.

Wynn (1970) presented an algorithm for locating D-optimal design measure. The sequential algorithm of Wynn constructs, as described by Labadi (2013), a converging sequence of discrete (exact) designs. Wynn's algorithm is a procedure that could help in overcoming difficulties that may arise when the model or the design space is sufficiently complicated such as could prevent an immediate evaluation of an optimal design. The Wynn's procedure simply sequentially adds a point of maximum variance of prediction to a given initial design. The process is continued till the design is brought closer to an optimal measure. The initial design point is admissible in the sense that the associated information matrix is nonsingular. Successive addition of design points to the initial design generates a sequence of designs which turn to the D-optimal design measure in the limit. Wynn's method is a one point at a time method. A particular illustration was made using a two-dimensional non-interactive first-order polynomial model defined on an irregular design space whose boundary is quadrilateral with vertices (2,2), (-1,1), (1,-1) and (-1,-1). Since it is possible to have more than one point with maximum variance of prediction, there is a choice of an alternative point which maximizes the variance function. Under such settings, the particular sequence of designs generated will not be unique.

Tsay (1976) gave a general procedure for the sequential construction of D-optimal designs, of which Wynn's procedure is a special case. Robertazzi and Schwartz (1989) presented an accelerated sequential algorithm for producing D-optimal designs. The algorithm has a useful advantage when there is no prior information concerning the structure of the optimal design. An illustration using two dimensional regression function defined on a regular unit square was considered as well as an illustration using three dimensional regression function defined on a regular unit cube. Due to the likely existence of design locations at the interior of the design space, accelerated sequential method uses grid search having discrete grid approximations of a continuous space. In both illustrations, the number of function evaluations was greatly reduced. Hardin and Sloane (1993) presented a super algorithm which finds optimal or near optimal designs for a wide range of low order response surface problems involving large several variables of either the continuous or discrete types or both.

Boon (2007) explored several techniques that could be used to numerically search for exact D-optimal designs. In his paper, several optimization algorithms for generating exact D-optimal design for any regression model were compared. Harman and Benková (2014) considered approximate D-optimal designs on varying experimental situations. Iwundu and Albert-Udochukwuka (2014)

presented an efficient algorithm for constructing N-point D-optimal exact designs on regular as well as irregular design regions.

Removing non-optimal support points in D-optimal design algorithms has been considered an effective way of speeding up algorithms for D-optimal design measures. Pronzota (2003) established a bound which helps to eliminate points from the design space during the search for a D-optimal design. Any point not satisfying the bound is removed from the design space and thus not considered for further investigation. Harman and Pronzota (2007) offered an improvement on the Pronzota (2003) lower bound on the maximum variance of prediction for an optimal point in the search for D-optimal design.

Modifications of the very early algorithms continue to feature in current research works. Very recently, Al Labadi and Wang (2010) considered a two points at a time modification for the Wynn's sequential algorithm for constructing D-optimal design. The modified algorithm adds to an initial design two points that have the same maximum variance of prediction. From the illustrative example of Al Labadi and Wang (2010) the modification achieved a reduction in the computational steps required to reach the D-optimal design by the Wynn's algorithm roughly by one-half. It is obvious that the modification could address the non-uniqueness of the Wynn's generated D-optimal sequence in the presence of only two maximum variance of prediction at each iteration. However, when there is no equal maximum variance of prediction, users of Al Labadi and Wang's algorithm simply would return to the Wynn's one-at-a-time sequential algorithm. Al Labadi (2013) considered the modification of the sequential algorithm and the exchange algorithm due to Fedorov (1972) by respectively, adding or exchanging two or more points at each iterative step of the original algorithm.

In this work a refinement in the construction of D-optimal design using the Wynn's basic sequential procedure is presented. It can be observed from the Wynn's algorithm that in constructing D-optimal design, close attention should be paid to the variance of prediction as the determinant is not monotonically increasing at every phase or vicinity of the search. This point is worthy of note to avoid the error of reporting a false optimum when an improvement is not seen as reflected by the determinant value of information matrix at a current iteration. Understanding that D-optimality is achieved at the point where the variance of prediction approximately equals the number of model parameters is a more helpful rule of thumb for convergence to optimality. The attention however, is the construction of a D-optimal design measure for the two-dimensional no-interaction five parameter polynomial model

$$E(Y_x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2$$

defined on an irregular and continuous space. Specifically, the design region is the irregular quadrilateral defined in Wynn (1970) and having a continuum of support points.

2. Methodology

As there is no information on the likely structure of the optimal design, discrete grid approximations to the continuous space shall be employed and search for the D-optimal design measure shall be within the discretized design space. The grid formations are as uniformly as possible, however due to the irregular nature of the design region, there are some grids that are not uniformly equi-spaced. The fundamental sequential search algorithm shall be based on the algorithm due to Wynn (1970). The search shall commence with an initial design whose design points need not necessarily be optimal points. The only requirement is that the initial design point be admissible in the sense that the associated information matrix of the design is nonsingular.

It is clear that with the Wynn's algorithm, having an initial design with non-optimal points would slow down the process for arriving at the optimal solution. To overcome this limitation, design filtering and reconstruction is proposed at some point in the sequence when cycles of optimal points are formed and it becomes clear which point(s) in the initial design are non-optimal points. By the filtering and reconstruction procedure, non-optimal points (unwanted and outlying points) are removed from the sequence of designs formed at earlier iterations and a replacement with optimal points made. The filtering and reconstruction procedure is supported by the method of experimental design reconstruction which according to Goupy (1996) is the most efficient way of detecting an outlier.

The new procedure centers around handling non-optimal design points that were introduced possibly either by the creation of a poor initial design or by its influence on the next design point(s) added to the initial or existing design. As will be seen in section 3, removal of non-optimal design points

will significantly improve the value of determinant of information matrix at a next iteration. Moreover, the sequence will be certain to converge to the required D-optimal (or near) optimal design measure. Bounds shall be established for the determinant value of information matrix for each sequentially generated design.

3. Numerical Illustration

To obtain a D-optimal design measure for the two-dimensional non-interaction polynomial model

$$E(Y_x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2$$

defined on the irregular continuous design space in Figure 1, whose boundary is quadrilateral with vertices (2,2), (-1,1), (1,-1) and (-1,-1), we discretize the region and obtain the candidate set

$$\mathcal{C} = \{(-1,1), (-1, \frac{1}{2}), (-1,0), (-1, -\frac{1}{2}), (-1,-1), (-\frac{1}{2}, 1.17), (-\frac{1}{2}, 1), (-\frac{1}{2}, \frac{1}{2}), (-\frac{1}{2}, 0), (-\frac{1}{2}, -\frac{1}{2}), (-\frac{1}{2}, -1), (0,1.33), (0,1), (0, \frac{1}{2}), (0,0), (0, -\frac{1}{2}), (0,-1), (\frac{1}{2}, 1\frac{1}{2}), (\frac{1}{2}, 1), (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, 0), (\frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, -1), (1,1.67), (1,1\frac{1}{2}), (1,1), (1, \frac{1}{2}), (1,0), (1, -\frac{1}{2}), (1, -1), (1\frac{1}{2}, 1.82), (1\frac{1}{2}, 1\frac{1}{2}), (1\frac{1}{2}, 1), (1\frac{1}{2}, \frac{1}{2}), (1.33, 0), (1.17, -\frac{1}{2}), (2, 2), (1.82, 1\frac{1}{2}), (1.67, 1)\}.$$

The candidate set thus defines $\tilde{N} = 39$ design points to be considered in the search for the D-optimal discrete design measure.

The initial design is

$$\xi_5^{(0)} = \begin{pmatrix} -1 & 1 \\ -1 & -1 \\ 0 & 0 \\ 1 & -1 \\ 1\frac{1}{2} & 1.82 \end{pmatrix}$$

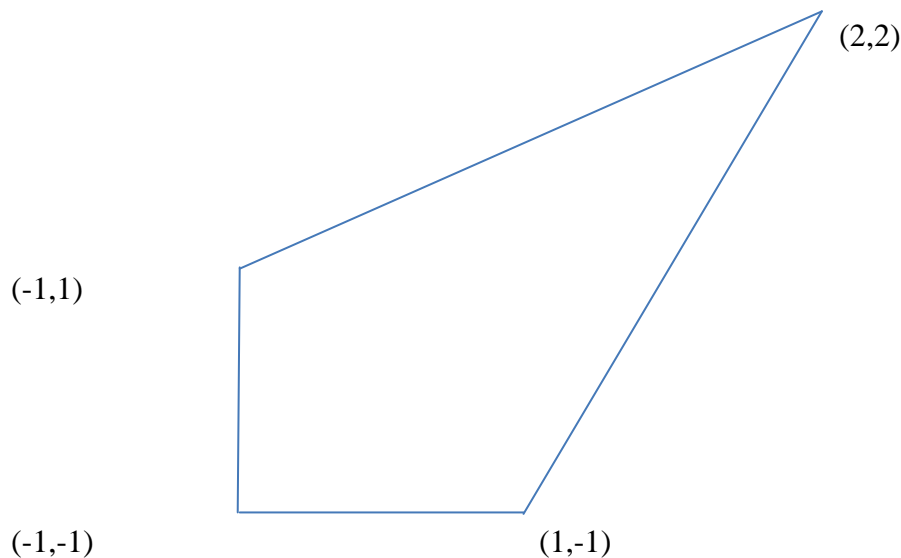


Figure 1. Quadrilateral with vertices (2,2), (-1,1), (1,-1) and (-1,-1)

The generated sequence starting from the initial design yields the statistics in Table 1. The sequence is obtained by adding to the initial design the point of maximum variance of prediction. It is very clearly obvious that the sequence cycles around some design points of which the initial design points (0,0) and $(1\frac{1}{2}, 1.82)$ are not a part. Thus for N=29 the initial design is filtered and reconstructed by replacing the design points (0,0) and $(1\frac{1}{2}, 1.82)$ with (2,2) and $(1\frac{1}{2}, \frac{1}{2})$. This greatly improved the determinant value from the supposed 0.7463 (without filtering and reconstruction) to 0.8407. There is also a replacement of the point (1.67, 1) with (-1, 1) at N=30. The point (1.67, 1) came into the design from the influence of poorly selected initial design. This replacement again allowed a maximal improvement in the determinant value of information matrix from the supposed 0.8414 to 0.8678. The MATLAB Version 2007b was employed in the generation of the sequence of the designs and the outputs are presented in Appendix A for N = 5, 6 and 7 only for space convenience.

Table 1. The generated sequence of D-optimal discrete exact designs

| Design size | Design points added | Point of maximum variance of prediction | Determinant of information matrix |
|-------------|-----------------------------|---|-----------------------------------|
| 5 | $\xi_5^{(0)}$ | 461.9177 | 0.0058 |
| 6 | 1.67, 1 | 11.8344 | 0.2169 |
| 7 | 2, 2 | 13.0825 | 0.2983 |
| 8 | $\frac{1}{2}, 1\frac{1}{2}$ | 7.6342 | 0.4389 |
| 9 | $1\frac{1}{2}, \frac{1}{2}$ | 6.0904 | 0.4760 |
| 10 | -1, $\frac{1}{2}$ | 6.6836 | 0.4712 |
| 11 | -1, -1 | 7.1765 | 0.4882 |
| 12 | 2, 2 | 7.4134 | 0.5221 |
| 13 | $\frac{1}{2}, 1\frac{1}{2}$ | 7.2944 | 0.5660 |
| 14 | $\frac{1}{2}, -1$ | 6.3668 | 0.6100 |
| 15 | $1\frac{1}{2}, \frac{1}{2}$ | 5.8724 | 0.6285 |
| 16 | 2, 2 | 5.8424 | 0.6334 |
| 17 | -1, 1 | 6.0738 | 0.6386 |
| 18 | $\frac{1}{2}, 1\frac{1}{2}$ | 6.1291 | 0.6513 |
| 19 | -1, -1 | 6.1091 | 0.6662 |
| 20 | 1, -1 | 6.0096 | 0.6813 |
| 21 | $1\frac{1}{2}, \frac{1}{2}$ | 5.8884 | 0.6942 |
| 22 | 2, 2 | 5.6872 | 0.7044 |
| 23 | $\frac{1}{2}, 1\frac{1}{2}$ | 5.5208 | 0.7098 |
| 24 | $\frac{1}{2}, -1$ | 5.6647 | 0.7115 |
| 25 | -1, $\frac{1}{2}$ | 5.6494 | 0.7170 |
| 26 | $1\frac{1}{2}, \frac{1}{2}$ | 5.7496 | 0.7225 |
| 27 | -1, -1 | 5.8661 | 0.7306 |
| 28 | 2, 2 | 5.5875 | 0.7414 |
| 29 | $\frac{1}{2}, 1\frac{1}{2}$ | 5.3852 | 0.8407 |
| 30 | $\frac{1}{2}, -1$ | 5.4158 | 0.8678 |

| | | | |
|----|-----------------------------|--------|--------|
| 31 | $1\frac{1}{2}, \frac{1}{2}$ | 5.4524 | 0.8696 |
| 32 | $\frac{1}{2}, 1\frac{1}{2}$ | 5.3977 | 0.8724 |
| 33 | -1, -1 | 5.4029 | 0.8742 |
| 34 | 2, 2 | 5.1671 | 0.8762 |
| 35 | $1\frac{1}{2}, \frac{1}{2}$ | 5.2141 | 0.8732 |
| 36 | -1, 1 | 5.2928 | 0.8715 |
| 37 | $\frac{1}{2}, 1\frac{1}{2}$ | 5.3101 | 0.8716 |
| 38 | $\frac{1}{2}, -1$ | 5.3363 | 0.8723 |
| 39 | 2, 2 | 5.3433 | 0.8736 |
| 40 | -1, -1 | 5.2421 | 0.8752 |
| 41 | $1\frac{1}{2}, \frac{1}{2}$ | 5.1631 | 0.8750 |
| 42 | $\frac{1}{2}, 1\frac{1}{2}$ | 5.2725 | 0.8733 |

4. Discussion

A refinement has been provided for a more effective use of the Wynn's sequential algorithm in the construction of D-optimal design measure. The refinement addresses situations where non-optimal design points have been introduced into the search at an early stage of experimentation. The non-optimal design points may have been introduced possibly either by the creation of a poor initial design or by its influence on the next design point(s) added to the initial or existing design. Experimental designs in the presence of one or more non-optimal design points do not allow maximal possible improvement in the determinant value of information matrix at any iteration. Such non-optimal design points certainly do not reflect similar characteristics of the other points and hence are treated as unwanted or outlying design points of which Pronzota (2003) suggests should be removed from the design space and not considered for further experimentation.

The filtering and reconstruction procedure allows a maximal improvement in the determinant value of information matrix. This is not surprising because an experimental design with non-optimal point(s) offers low determinant value of information matrix. By having points with similar characteristics in the design should yield a maximal improvement in the determinant value of information matrix.

For the construction of D-optimal design measure for the two-dimensional no-interaction five parameter polynomial model defined on an irregular and continuous space, the algorithm converged at the design measure, ξ_{34}^* where the maximum variance of prediction was closest to the number of model parameters. There was no noticeable improvement in the determinant value of information matrix nor in the maximum variance of prediction two steps after N=34. The convergence to the D-optimal design measure was certain as N increased. Wynn's mathematical justification supports this. Bounds for the determinants of information matrices associated with each design at each iterative step have been computed and are presented in Table 2. The bounds as

established by Wynn (1970) are given by

$$A_N \leq \det\{M(\xi^*)\} \leq B_N$$

where

$$A_N = \det\{M(\xi_N)\} \left\{ \frac{\bar{d}(\xi_N)}{k} \right\}^k \left\{ \frac{k-1}{\bar{d}(\xi_N)-1} \right\}^{k-1}$$

$$B_N = \det\{M(\xi_N)\} \exp\{\bar{d}(\xi_N) - k\}$$

Here $\bar{d}(\xi_N)$ represents the maximum variance of prediction using the design measure, ξ_N . The bounds show how close the search is to the optimum design at any stage.

Table 2. Bounds for the determinants of information matrices

| Design size | Design points added | Determinant of information matrix | A_N | B_N |
|-------------|-----------------------------|-----------------------------------|---------|--------------------------|
| 5 | $\xi_5^{(0)}$ | 0.0058 | 0.2214 | 1.5859×10^{196} |
| 6 | 1.67, 1 | 0.2169 | 0.0095 | 201.5588 |
| 7 | 2, 2 | 0.2983 | 0.4394 | 965.6915 |
| 8 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.4389 | 0.4813 | 8.2542 |
| 9 | $1\frac{1}{2}, \frac{1}{2}$ | 0.4760 | 0.4867 | 1.4163 |
| 10 | -1, $\frac{1}{2}$ | 0.4712 | 0.4933 | 2.5374 |
| 11 | -1, -1 | 0.4882 | 0.5231 | 4.3037 |
| 12 | 2, 2 | 0.5221 | 0.5661 | 5.8328 |
| 13 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.5660 | 0.6100 | 5.6139 |
| 14 | $\frac{1}{2}, -1$ | 0.6100 | 0.6302 | 2.3929 |
| 15 | $1\frac{1}{2}, \frac{1}{2}$ | 0.6285 | 0.6380 | 1.5038 |
| 16 | 2, 2 | 0.6334 | 0.6424 | 1.4707 |
| 17 | -1, 1 | 0.6386 | 0.6525 | 1.8688 |
| 18 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.6513 | 0.6668 | 2.0144 |
| 19 | -1, -1 | 0.6662 | 0.6816 | 2.0199 |
| 20 | 1, -1 | 0.6813 | 0.6946 | 1.8698 |
| 21 | $1\frac{1}{2}, \frac{1}{2}$ | 0.6942 | 0.7050 | 1.6878 |
| 22 | 2, 2 | 0.7044 | 0.71129 | 1.2672 |
| 23 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.7098 | 0.7140 | 1.1949 |
| 24 | $\frac{1}{2}, -1$ | 0.7115 | 0.7180 | 1.3831 |
| 25 | -1, $\frac{1}{2}$ | 0.7170 | 0.7233 | 1.3726 |
| 26 | $1\frac{1}{2}, \frac{1}{2}$ | 0.7225 | 0.7308 | 1.5289 |
| 27 | -1, -1 | 0.7306 | 0.7415 | 1.7371 |
| 28 | 2, 2 | 0.7414 | 0.7468 | 1.3341 |
| 29 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.8407 | 0.8435 | 1.2358 |
| 30 | $\frac{1}{2}, -1$ | 0.8678 | 0.8711 | 1.3152 |
| 31 | $1\frac{1}{2}, \frac{1}{2}$ | 0.8696 | 0.8735 | 1.3671 |
| 32 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.8724 | 0.8755 | 1.2985 |
| 33 | -1, -1 | 0.8742 | 0.8774 | 1.3079 |
| 34 | 2, 2 | 0.8762 | 0.8768 | 1.0356 |
| 35 | $1\frac{1}{2}, \frac{1}{2}$ | 0.8732 | 0.8741 | 1.0817 |
| 36 | -1, 1 | 0.8715 | 0.8732 | 1.1680 |
| 37 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.8716 | 0.8735 | 1.1885 |
| 38 | $\frac{1}{2}, -1$ | 0.8723 | 0.8745 | 1.2210 |
| 39 | 2, 2 | 0.8736 | 0.8759 | 1.2314 |
| 40 | -1, -1 | 0.8752 | 0.8764 | 1.1149 |
| 41 | $1\frac{1}{2}, \frac{1}{2}$ | 0.8750 | 0.8756 | 1.0300 |
| 42 | $\frac{1}{2}, 1\frac{1}{2}$ | 0.8733 | 0.8748 | 1.1469 |

From Table 2, the bounds $0.8768 \leq \det\{M(\xi^*)\} \leq 1.0356$ associated with the design ξ_{34} whose determinant value of information matrix is 0.8762, shows that ξ_{34} is very close to the true optimum design. The maximum determinant value of information matrix in the sequence generated is also associated with the design, ξ_{34} . Furthermore, this design has a maximum variance of prediction approximately equal to k , the number of model parameters. For eight experimental runs after $N=34$ no significant improvement was seen in the search. On the basis of the grid search with the 39 grid points, ξ_{34} is thus reported as approximately D-optimal. The design measure is as in Figure 2. The associated information matrix is

$$M(\xi_{34}) = \begin{pmatrix} 1.0000 & 0.5735 & 0.5882 & 1.6397 & 1.6618 \\ 0.5735 & 1.6397 & 1.0147 & 2.0404 & 1.7647 \\ 0.5882 & 1.0147 & 1.6618 & 1.8015 & 2.0662 \\ 1.6397 & 2.0404 & 1.8015 & 4.5570 & 3.8235 \\ 1.6618 & 1.7647 & 2.0662 & 3.8235 & 4.5846 \end{pmatrix}$$

The weights associated with the D-optimal points are distributed as in Table 3. It is possible that having more grid points could yield a better approximation to the true unknown D-optimal design measure whose determinant value of information matrix would more satisfactorily meet the bounds provided by Wynn.

$$\xi_{34} = \begin{pmatrix} -1 & 1 \\ -1 & -1 \\ 2 & 2 \\ 1 & -1 \\ 1.5 & 0.5 \\ -1 & 1 \\ 2 & 2 \\ 0.5 & 1.5 \\ 1.5 & 0.5 \\ -1 & 0.5 \\ -1 & -1 \\ 2 & 2 \\ 0.5 & 1.5 \\ 0.5 & -1 \\ 1.5 & 0.5 \\ 2 & 2 \\ -1 & 1 \\ 0.5 & 1.5 \\ -1 & -1 \\ 1 & -1 \\ 1.5 & 0.5 \\ 2 & 2 \\ 0.5 & 1.5 \\ 0.5 & -1 \\ -1 & 0.5 \\ 1.5 & 0.5 \\ -1 & -1 \\ 2 & 2 \\ 0.5 & 1.5 \\ 0.5 & -1 \\ 1.5 & 0.5 \\ 0.5 & 1.5 \\ -1 & -1 \\ 2 & 2 \end{pmatrix}$$

Figure 2. 39 point approximately D-optimal design measure

Table 3. Weights for the 34 point D-optimal design measure

| Design point | Weight |
|-----------------------------|----------------|
| -1, 1 | $\frac{3}{34}$ |
| -1, -1 | $\frac{5}{34}$ |
| 2, 2 | $\frac{7}{34}$ |
| 1, -1 | $\frac{2}{34}$ |
| $1\frac{1}{2}, \frac{1}{2}$ | $\frac{6}{34}$ |
| $\frac{1}{2}, 1\frac{1}{2}$ | $\frac{6}{34}$ |
| $-1, \frac{1}{2}$ | $\frac{2}{34}$ |
| $\frac{1}{2}, -1$ | $\frac{3}{34}$ |

5. Conclusions

A refinement in the construction of D-optimal designs using the Wynn's basic sequential procedure has been presented. The refinement considers filtering and reconstruction as a viable procedure for sequentially locating D-optimal design measures. The method effectively improves experimental design in the search for an optimal design measure. While the procedure is identical to the Wynn's sequential algorithm for constructing D-optimal designs, filtering and reconstruction addresses situations where outlying non-optimal design points had been admitted into the design possibly either by the creation of a poor initial design or by its influence on the next design point(s). By the method, outlying non-optimal design points are removed and the design reconstructed, thus resulting in a significant improvement on the determinant value of information matrix.

In particular, an approximate solution has been obtained in the construction of D-optimal design measure for a two-dimensional non-interaction polynomial model defined on an irregular continuous design whose boundary is quadrilateral with vertices (2,2), (-1,1), (1,-1) and (-1,-1). Furthermore, bounds have been established for the

determinant value of information matrix for each sequentially generated design. The bounds reveal that the D-optimal measure generated by the search procedure is very close to the true unknown D-optimal design measure.

Although D-optimality criterion is determinant-based, it is important to pay close attention to the variance of prediction when constructing D-optimal designs. This is because the determinant of information matrix may not monotonically increase at every phase or vicinity of the search. Since D-optimality is achieved at the point where the variance of prediction approximately equals the number of model parameters, paying close attention to the variance of prediction becomes a helpful rule of thumb for convergence to optimality.

APPENDIX A

MATLAB OUTPUT_ Design Filtering and Reconstruction Starting design

N=5

```
>> det((A'*A)/5)= 0.0058
>> B*(5*inv(A'*A))*B' = 5.0000
>> C*(5*inv(A'*A))*C'=43.3011
>> D*(5*inv(A'*A))*D'=61.7426
>> E*(5*inv(A'*A))*E'=30.2587
>> F*(5*inv(A'*A))*F'=5.0000
>> G*(5*inv(A'*A))*G'=69.1991
>> H*(5*inv(A'*A))*H'=33.8518
>> I*(5*inv(A'*A))*I'
    3.5937
>> J*(5*inv(A'*A))*J'
    3.7380
>> K*(5*inv(A'*A))*K'
    4.2187
>> L*(5*inv(A'*A))*L'
    61.1866
>> M*(5*inv(A'*A))*M'
    169.1710
>> N*(5*inv(A'*A))*N'
    59.3537
>> O*(5*inv(A'*A))*O'
    9.9851
>> P*(5*inv(A'*A))*P'
    5.0000
>> Q*(5*inv(A'*A))*Q'
    14.3325
>> R*(5*inv(A'*A))*R'
    94.1334
>> S*(5*inv(A'*A))*S'
    181.8960
>> T*(5*inv(A'*A))*T'
    24.4992
>> U*(5*inv(A'*A))*U'
    5.4687
>> V*(5*inv(A'*A))*V'
```

```
    8.5222
>> W*(5*inv(A'*A))*W'
    3.5937
>> X*(5*inv(A'*A))*X'
    46.8340
>> Y*(5*inv(A'*A))*Y'
    92.1282
>> Z*(5*inv(A'*A))*Z'
    34.0297
>> (ZA)*(5*inv(A'*A))*(ZA)'
    15.0000
>> (ZB)*(5*inv(A'*A))*(ZB)'
    75.7564
>> (ZC)*(5*inv(A'*A))*(ZC)'
    100.0163
>> (ZD)*(5*inv(A'*A))*(ZD)'
    57.7140
>> (ZE)*(5*inv(A'*A))*(ZE)'
    5.0000
>> (ZF)*(5*inv(A'*A))*(ZF)'
    5.0000
>> (ZG)*(5*inv(A'*A))*(ZG)'
    54.2187
>> (ZH)*(5*inv(A'*A))*(ZH)'
    259.2856
>> (ZI)*(5*inv(A'*A))*(ZI)'
    449.2774
>> (ZJ)*(5*inv(A'*A))*(ZJ)'
    313.6530
>> (ZK)*(5*inv(A'*A))*(ZK)'
    127.2938
>> (ZL)*(5*inv(A'*A))*(ZL)'
    155.0000
>> (ZM)*(5*inv(A'*A))*(ZM)'
    320.8080
>> (ZN)*(5*inv(A'*A))*(ZN)'
    461.9177
```

N=6

```
>> det((A'*A)/6)
    0.2169
>> B*(6*inv(A'*A))*B'
    4.1808
>> C*(6*inv(A'*A))*C'
    4.4356
>> D*(6*inv(A'*A))*D'
    4.4858
>> E*(6*inv(A'*A))*E'
    3.8654
>> F*(6*inv(A'*A))*F'
    4.9040
>> G*(6*inv(A'*A))*G'
    6.1498
>> H*(6*inv(A'*A))*H'
    5.2170
>> I*(6*inv(A'*A))*I'
    4.1484
```

| | |
|--------------------------------------|--------------------------------------|
| >> $J*(6*\text{inv}(A'*A))*J'$ | 11.8344 |
| 3.3957 | >> $(ZM)*(6*\text{inv}(A'*A))*(ZM)'$ |
| >> $K*(6*\text{inv}(A'*A))*K'$ | 5.7864 |
| 2.4931 | >> $(ZN)*(6*\text{inv}(A'*A))*(ZN)'$ |
| >> $L*(6*\text{inv}(A'*A))*L'$ | 5.9357 |
| 3.7699 | |
| >> $M*(6*\text{inv}(A'*A))*M'$ | N=7 |
| 10.2186 | >> $\det((A'*A)/7)$ |
| >> $N*(6*\text{inv}(A'*A))*N'$ | 0.2983 |
| 7.5252 | >> $B*(7*\text{inv}(A'*A))*B'$ |
| >> $O*(6*\text{inv}(A'*A))*O'$ | 4.8750 |
| 6.0477 | >> $C*(7*\text{inv}(A'*A))*C'$ |
| >> $P*(6*\text{inv}(A'*A))*P'$ | 5.1277 |
| 5.0259 | >> $D*(7*\text{inv}(A'*A))*D'$ |
| >> $Q*(6*\text{inv}(A'*A))*Q'$ | 5.1769 |
| 3.9939 | >> $E*(7*\text{inv}(A'*A))*E'$ |
| >> $R*(6*\text{inv}(A'*A))*R'$ | 4.5095 |
| 5.2811 | >> $F*(7*\text{inv}(A'*A))*F'$ |
| >> $S*(6*\text{inv}(A'*A))*S'$ | 5.5055 |
| 11.3405 | >> $G*(7*\text{inv}(A'*A))*G'$ |
| >> $T*(6*\text{inv}(A'*A))*T'$ | 6.9058 |
| 6.8927 | >> $H*(7*\text{inv}(A'*A))*H'$ |
| >> $U*(6*\text{inv}(A'*A))*U'$ | 5.6426 |
| 5.9208 | >> $I*(7*\text{inv}(A'*A))*I'$ |
| >> $V*(6*\text{inv}(A'*A))*V'$ | 3.9665 |
| 5.1636 | >> $J*(7*\text{inv}(A'*A))*J'$ |
| >> $W*(6*\text{inv}(A'*A))*W'$ | 3.0496 |
| 4.1551 | >> $K*(7*\text{inv}(A'*A))*K'$ |
| >> $X*(6*\text{inv}(A'*A))*X'$ | 2.3789 |
| 5.2247 | >> $L*(7*\text{inv}(A'*A))*L'$ |
| >> $Y*(6*\text{inv}(A'*A))*Y'$ | 4.3343 |
| 8.7227 | >> $M*(7*\text{inv}(A'*A))*M'$ |
| >> $Z*(6*\text{inv}(A'*A))*Z'$ | 11.4717 |
| 6.2603 | >> $N*(7*\text{inv}(A'*A))*N'$ |
| >> $(ZA)*(6*\text{inv}(A'*A))*(ZA)'$ | 7.8120 |
| 3.8544 | >> $O*(7*\text{inv}(A'*A))*O'$ |
| >> $(ZB)*(6*\text{inv}(A'*A))*(ZB)'$ | 5.4885 |
| 4.3026 | >> $P*(7*\text{inv}(A'*A))*P'$ |
| >> $(ZC)*(6*\text{inv}(A'*A))*(ZC)'$ | 4.2445 |
| 4.3436 | >> $Q*(7*\text{inv}(A'*A))*Q'$ |
| >> $(ZD)*(6*\text{inv}(A'*A))*(ZD)'$ | 3.5671 |
| 3.5114 | >> $R*(7*\text{inv}(A'*A))*R'$ |
| >> $(ZE)*(6*\text{inv}(A'*A))*(ZE)'$ | 5.8361 |
| 4.1357 | >> $S*(7*\text{inv}(A'*A))*S'$ |
| >> $(ZF)*(6*\text{inv}(A'*A))*(ZF)'$ | 13.0825 |
| 5.8178 | >> $T*(7*\text{inv}(A'*A))*T'$ |
| >> $(ZG)*(6*\text{inv}(A'*A))*(ZG)'$ | 7.2298 |
| 2.7613 | >> $U*(7*\text{inv}(A'*A))*U'$ |
| >> $(ZH)*(6*\text{inv}(A'*A))*(ZH)'$ | 5.5405 |
| 3.6925 | >> $V*(7*\text{inv}(A'*A))*V'$ |
| >> $(ZI)*(6*\text{inv}(A'*A))*(ZI)'$ | 4.6086 |
| 6.4753 | >> $W*(7*\text{inv}(A'*A))*W'$ |
| >> $(ZJ)*(6*\text{inv}(A'*A))*(ZJ)'$ | 3.9211 |
| 5.6474 | >> $X*(7*\text{inv}(A'*A))*X'$ |
| >> $(ZK)*(6*\text{inv}(A'*A))*(ZK)'$ | 5.8578 |
| 3.9087 | >> $Y*(7*\text{inv}(A'*A))*Y'$ |
| >> $(ZL)*(6*\text{inv}(A'*A))*(ZL)'$ | 10.0678 |


```

>> Z*(7*inv(A*A))*Z'
7.2941
>> (ZA)*(7*inv(A*A))*(ZA)'
4.3218
>> (ZB)*(7*inv(A*A))*(ZB)'
4.5483
>> (ZC)*(7*inv(A*A))*(ZC)'
4.5675
>> (ZD)*(7*inv(A*A))*(ZD)'
3.8664
>> (ZE)*(7*inv(A*A))*(ZE)'
4.8250
>> (ZF)*(7*inv(A*A))*(ZF)'
4.7119
>> (ZG)*(7*inv(A*A))*(ZG)'
2.2603
>> (ZH)*(7*inv(A*A))*(ZH)'
4.0923
>> (ZI)*(7*inv(A*A))*(ZI)'
7.5160
>> (ZJ)*(7*inv(A*A))*(ZJ)'
6.5600
>> (ZK)*(7*inv(A*A))*(ZK)'
4.5097
>> (ZL)*(7*inv(A*A))*(ZL)'
4.6450
>> (ZM)*(7*inv(A*A))*(ZM)'
3.6701
>> (ZN)*(7*inv(A*A))*(ZN)'
6.1930

```

security technology department. The Johns Hopkins University Applied Physics laboratory, Laurel, MD 20723. SpringSim Vol. 2, pp. 121-126.

- [5] Fedorov, V. V. (1972). Theory of Optimal Experiments. *New York. Academic Press.*
- [6] Goupy, J. (1996) Outliers and Experimental designs. *Chemometrics and Intelligent Laboratory Systems*, Vol. 35, pp. 145-156.
- [7] Hardin, R. H. and Sloane, N. J. A. (1993) A new approach to the construction of optimal designs. *Journal of Statistical Planning and Inference*. Vol. 37, pp. 339-369.
- [8] Harman, R. and Benková, E. (2014) Department of Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 84248 Bratislava, Slovakia. Preprint submitted to Elsevier. <https://arxiv.org/pdf/1408.2698>
- [9] Harman, R. and Pronzota, L. (2007) Improvements on removing nonoptimal support points in D-optimum design algorithms. *Statistics and Probability letters*, Vol. 77, pp. 90-94.
- [10] Iwundu, M. P. And Albert-Udochukwuka, E. B. (2014). An efficient algorithm for D-optimal designs. *Journal of Applied Sciences*, 14: 3547-3554.
- [11] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extreme problems. *Canad. Journal of Math.* Vol. 12, pp. 363-366.
- [12] Mitchell and Miller (1970). Use of design repair to construct designs for special linear models. Report ORNL-4661, pp. 130-131. Mathematics Division, Oak Ridge National Laboratory, Oak Ridge.
- [13] Mitchell, T. J. (1974). An Algorithm for the Construction of D-optimal Experimental Designs. *Technometrics*, 16, pp 203-210.
- [14] Pronzota, L. (2003). Removing non-optimal support points in D-optimum design algorithms. *Statistics and Probability letters*, Vol. 63, pp. 223-228.
- [15] Robertazzi, T. G. and Schwartz, S. C. (1989). An accelerated sequential algorithm for producing D-optimal designs. *SIAM J. SCI. STAT. COMPUT.* Vol. 10, No. 2, pp. 341-358.
- [16] Tsay, J. (1976). On the sequential construction of D-optimal designs. *Journal of American Statistical Association*, Vol. 71, Issue 355, pp. 671-674.
- [17] Wynn, H. P. (1970) The sequential generation of D-optimal experimental designs. *The Annals of Mathematical Statistics*, Vol. 41, No. 5, pp. 1655-1664.

REFERENCES

- [1] Al Labadi, L. and Wang, Z. (2010) Modified Wynn's sequential algorithm for constructing D-optimal designs: Adding two points at a time. *Communications in Statistics-Theory and Methods*, Vol. 39, pp. 2818-2828.
- [2] Al Labadi, L. (2013). Some refinements on Fedorov's algorithm for constructing D-optimal designs. *Brazilian Journal of Probability and Statistics*. DOI: 10.1214/13-BJPS228.
- [3] Atkinson, A. C. (1996). The usefulness of Optimum Experimental Designs. *Journal of Royal Statistical Society, Series B*, Vol. 58, No. 1, pp. 59-76.
- [4] Boon, J. E. (2007) Generating Exact D-optimal designs for polynomial models. Operations assessment group, National