A Note on a Comparison of Starting Points for the Generation of D-optimal Second-order Designs

Ngozi Chizoma Umelo-Ibemere

Department of Statistics, Federal University of Technology, Owerri, Nigeria

Abstract. Usually, line search algorithms begin by an initial guess of the starting point. However, for quick convergence and reduction in computational cost, it is necessary to determine an optimal starting point. In this paper, comparison of three starting points for the generation of second-order $D-$optimal designs using a variance-gradient based exchange algorithm is considered. Quantile plots are used to compare the variances of $D-$optimal designs generated by each of the three starting points. Result obtained reveals that an optimal choice of the starting point improves $D-$efficiency of the generated design points.

Résumé (Abstract in French) Les algorithmes de recherche en ligne commençent par un point donné et il est souhaitable de déterminer ce point de maniére optimale afin d'assurer une convergence plus rapide et de réduire le coût de calculs. Dans ce papier, nous comparons trois méthodes de détermination de trois points de départ pour la génération de second ordre (D-) dans un design optimal, utilisant une méthode de variance-gradient. Des études de simulations illustrent notre méthode

Corresponding Author:ngozi.umelo-ibemere@futo.edu.ng

The author Ngozi Chizoma Umelo-Ibemere, M.Sc., is a lecturer at the Department of Statistics, Federal University of Technology, Owerri, Nigeria.

1. Introduction

In design of experiments, standard statistical designs such as Box Behnken design, central composite design, factorial and fractional factorial designs which are optimal with respect to the $A-, D-$ and $G-$ optimality criteria are usually used. However, these designs are inappropriate in some situations involving non-standard modelling, constrained regions, unusual sample sizes, or 'tailor-made' designs (designs arising just to satisfy the exact needs of the experimenter); see Montgomery (2013, p.511). In such cases, algorithmic construction of optimal designs are useful. Algorithms based on the line search equation have three common parts: the starting point, $\overline{\underline{x}}_0$, the search direction, $\underline{d}$ and the step length, ρ. Such algorithms begin by an initial guess of $\overline{\underline{x}}_0$ and performing some sequences of operations using the $\underline{d}$ and ρ to obtain an iterate, $\underline{x}^*$. A stopping rule is applied and the iterate is tested for convergence depending on the design criterion to be satisfied. The line search equation is given as

$$\underline{x}^* = \overline{\underline{x}}_0 + \rho\underline{d} \tag{1}$$

At the jth step, the line equation is,

$$\underline{x}_j^* = \underline{x}_{j-1}^* + \rho_j\underline{d}_j \tag{2}$$

Many algorithms have been developed for the construction of $D-$optimal designs, see for example, Fedorov (1972), Mitchell (2000), Atkinson and Donev (1992). Onukogu (2002) proposed a Super Convergent Line Search algorithm that converges to $D-$optimality criterion. Ugbe and Chigbu (2013) applied the algorithm in non-overlapping segmentation for solving constrained optimization problems. Onukogu and Nsude (2014) modified Onukogu (2002) algorithm to generate $D-$optimal designs for second order models defined in the experimental space [-1,1].

A 'good' choice of the starting point helps to attain quick convergence in few iterations. Gertz, Nocedal and Sartenar (2004) noted the poor performance (or even failure) in linear and quadratic programming when the starting point is unfavourable. Based on that, they presented a starting point strategy in non-linear programming. An appropriate choice of an initial starting point reduces computational cost and improves the speed of convergence, see Torabi and Hosseini (2018). $\overline{\underline{x}}_0$ can be chosen in some arbitrary manner but should be a reasonable estimate of the solution (Nocedal and Wright, 1999). Hence, there is need for optimal choice of the starting point.

In this paper, three starting points shall be compared using the gradient-based algorithm proposed by Onukogu and Nsude (2014). Each starting point is applied to the algorithm to generate $D-$optimal second-order designs in the experimental

space [-1,1]. The optimal design points obtained are compared graphically and with $D-$efficiency. For $D-$efficiency, the optimal points are compared with Box and Draper (1974) minimum point second-order designs. Comparison of the starting points are carried out for a second order model defined in the experimental space [-1,1].

2. $D-$optimality Criterion

Kiefer and Wolfowitz (1959) greatly contributed to the development of $D-$optimality criterion. This criterion maximizes the determinant of Fisher's information matrix. It also minimizes the generalized variance of the regression parameter estimates and the volume of the confidence ellipsoid of the parameter; see Atkinson and Donev (1992). Consider the usual linear model,

$$y = f^{'}(\underline{x})\beta + e; \tag{3}$$

defined in an experimental space,$\{\tilde{X}, F_x, \Sigma_x\}$ where $\tilde{X}$ is a $k-$dimensional compact, continuous and metric space of all possible experimental trials, F_x is a finite dimensional linear space of continuous functions and Σ_x is a non-negative continuous space of random error associated with the response at each point $\underline{x} \in \tilde{X}$: see Onukogu (1997). $f(\underline{x})$ is a $p-$ parameter polynomial of degree m, β is a $p-$ vector of unknown parameters to be estimated by the experimenter, $\underline{x} = (x_1, x_2, \ldots, x_k)$ are non-stochastic independent variables in $\tilde{X}$, $e \in \Sigma_x$. The model in (3) can be written in matrix form as

$$Y = X\beta + e; \tag{4}$$

where Y is an $n \times 1$ response vector, X is an $n \times p$ design matrix of rank p, β is a $p \times 1$ vector of unknown parameters, $e \sim N(0, \sigma^2 I)$ random error term. The least square estimate of β is $\hat{\beta} = (X^{'}X)^{-1}X^{'}Y$.

An optimal design generated based on this criterion for a specified model is called $D-$optimal design. Therefore, $D-$optimal design maximizes the determinant of $(X'X)$, that is $max|X'X|$. For a specified $p-$parameter linear model defined in an experimental space, an $n-$point $D-$optimal design is given as $|M(\underline{x}, \xi_n^*)| = max \frac{|X^{'}X|}{n^p}$, see Heredia-Langner et al(2003). $D-$efficiency is one of the properties of $D-$optimum designs. It is defined as

$$D - eff = \left(\frac{|M(\underline{x}, \xi_n^*)|}{\underset{\xi}{max}|M(\underline{x}, \xi_n)|} \right)^{\frac{1}{p}}$$

where, the denominator is the maximum of all available $D-$optimal designs.

3. Method

The three starting points shall be applied to the algorithm using a second order model with two independent variables. The model is shown in (5)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + e \tag{5}$$

3.1. The Algorithm

The algorithm as presented by Onukogu and Nsude (2014) is summarized in the following steps;

1. Construct N grid points $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N)$ for the factors $\underline{x} = (x_1, x_2, \ldots, x_k)$ spread over $\tilde{X}$ such that $X(\xi_N) = \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_N \end{pmatrix}$ is an $N \times p$ extended design matrix and $M(\underline{x}, \xi_N^{(0)}) = (X'(\xi_N)X(\xi_N))N^{-1}$ is a $p \times p$ normalized Fisher's information matrix.

2. Define the measure $\xi_N^{(0)} = \{\frac{x_1}{N}, \frac{x_2}{N}, \ldots, \frac{x_N}{N}\}$

3. Take a sample, $n(\leq N)$ support points, thus define the initial design measure
$\xi_n^{(0)} = \begin{pmatrix} \underline{x}_1 & w \\ \underline{x}_2 & w \\ \vdots & \vdots \\ \underline{x}_n & w \end{pmatrix}$; $w = \frac{1}{n}$ such that $\Delta_{(0)} = |M(\underline{x}, \xi_n^{(0)})| \neq 0$

4. Define the variance function, $V(\underline{x}, \xi_N^{(0)}) = \underline{x}' M^{-1}(\underline{x}, \xi_N^{(0)})\underline{x}$, a $2m$ degree polynomial and compute $V(\underline{x}, \xi_N^{(0)})_{i=1}^N$

5. From step 4, obtain the gradient vector, $g_{ik} = \frac{\delta V(\underline{x}, \xi_N^{(0)})}{\delta \underline{x}_{ik}}$, a $2m - 1$ degree polynomial evaluated at all $i = 1, 2 \ldots, N$.

6. Define the k-component direction vector $\underline{d} = \sum_{i=1}^N \theta_i \underline{g}_i$; $\theta_i \geq 0$, $\sum_{i=1}^N \theta_i = 1$ with variance $V(\underline{d}) = \sum_{i=1}^N \theta_i^2 V(\underline{x})$. Normalize to $\underline{d}_0 = (\underline{d}'\underline{d})^{-\frac{1}{2}}\underline{d}$.

7. Obtain θ_i from the equation $\frac{\delta V(\underline{d})}{\delta \theta_i} = 2\sum_{i=1}^N \theta_i V(\underline{x}) = 0$

8. Obtain the starting point, $\overline{\underline{x}}_0$.

9. Compute the step length, ρ from the equation $\frac{\delta V(\underline{x}, \xi_N^{(0)})}{\delta \rho} = 0$.

10 Move to the point $\underline{x}_1^* = \overline{\underline{x}}_0 + \rho_0 \underline{d}_0$ and at the jth step move to
$\underline{x}_j^* = \underline{x}_{j-1}^* + \rho_j \underline{d}_j$

11. If $V(\underline{x}_j^*, \xi_N) > V(\underline{x}_{min}, \xi_N)$, exchange $\underline{x}_j^*$ with $\underline{x}_{(min)}$
else, stop and check for convergence.
Is $\|\Delta_{(j)} - \Delta_{(j-1)}\| < \epsilon \geq 0$?
Yes: stop, $\xi_N^{(j)}$ is $D-$ optimal
No: $\xi_N^{(j)} = \xi_N^{(j-1)} + \underline{x}_j^*$ and continue from step 2.

3.2. Starting Points

The starting points to be considered for the construction of second order $D-$optimal designs are given in equations (6), (7) and (8).

$$\overline{\underline{x}}_0 = \frac{\sum_{i=1}^N \underline{x}_i}{N} \qquad (6)$$

$$\overline{\underline{x}}_0 = \sum_{i=1}^N \theta_i \underline{x}_i \qquad (7)$$

where θ_i is the weighting factor for each $\underline{x}$.

$$\overline{\underline{x}}_0 = \sum_{i=1}^{N} \alpha_i \underline{x}_i \qquad (8)$$

where, $\alpha_i = \frac{(\underline{x}_i'\underline{x}_i)^{-1}}{\sum_{i=1}^{N}(\underline{x}_i'\underline{x}_i)^{-1}}$

4. Discussion of Results

Eleven support points chosen from $\tilde{X}$ are;

$\{(-1,-1); (-\frac{1}{2},-1); (0,-1); (1,-1); (-1,0); (-\frac{1}{2},0); (0,0); (1,0); (-1,1); (0,1); (1,1)\}$.
$V(\underline{x}, \xi_N^{(0)}), g_{ik}, \underline{d}_0$ and ρ are computed from the algorithm presented in section 3.1. The algorithm was implemented using MATLAB, a software that handles matrices efficiently. The three starting points are separately applied to the variance-exchange algorithm to obtain a minimum point design in the experimental space; $\tilde{X} = \{\underline{x} : x_1, x_2; -1 \le x_1, x_2 \le 1\}$ for the model in (5). This implies that from the starting points in (6), a move is made to $\underline{x}_j^* = \underline{x}_{j-1}^* + \rho_j \underline{d}_j$ and $\underline{x}^*$ is obtained. The same procedure is carried out using (7) and (8) separately but with the same support points and $\xi_n^{(0)}$. At convergence, the design points and determinants obtained for each of the starting points are shown in Table 1. For clarity, (6), (7) and (8) shall be referred to as starting points I, II, and III respectively. The $|M(\underline{x}, \xi_n^*)|$ is calculated for each starting point and the optimal designs generated are compared with Box and Draper (1974) $D-$optimal minimum point designs using $D-$efficiency. Results obtained show that the $D-$optimal designs generated compare well with those of Box and Draper (1974) but starting point III has the highest $D-$efficiency value. In addition, the combined quantile plots are used to compare the performance of designs generated by the starting points. For more details on the combined quantile plots, see, Khuri, Kim and Um (1996), Chigbu and Umelo-Ibemere (2015). Graphical comparison reveals that starting point III generated design points have the lowest quantile values for radius,$r = 0.3, 0.7$ but have equivalent quantile values for $r = (1.1, 1.4)$. Information obtained from the plots reveal that the three starting points perform well. However, starting point III slightly performs better than I and II.

Table 1: Comparison of the $D-$optimal designs for each starting point

starting point	$\underline{x}^*$	$|M(\underline{x}^*, \xi)|$	$D-$eff
I	(-1,-1);(-0.0813,0.0499); (1,-1);(-1,1);(1,0.4);(0.4,1)	0.55×10^{-2}	0.9929
II	(-1,-1);(-0.1403,0.0444); (1,-1);(-1,1);(1,0.4);(0.4,1)	0.55×10^{-2}	0.9929
III	(-1,-1);(-0.1867,-0.0952); (1,-1);(-1,1);(1,0.4);(0.4,1)	0.57×10^{-2}	0.9988
Box and Draper	(-1,-1);(1,-1);(-1,1);(-0.1315,-0.1315); (1,0.3945);(0.3945,1)	0.574×10^{-2}	

5. Conclusion

The choice of optimal starting point in line search methods influence the optimal design points generated at convergence. Although the starting points considered are reasonable estimates of the mean of the support points, starting point III has the highest determinant of the information matrix. The algorithm converged using the three starting points. Graphical comparison of the variances of the optimal design points generated also reveal convergence to $D-$optimality with the lowest variation in starting point III generated designs. In comparison with Box and Draper (1974) minimum point design, the design points generated with starting point III are more efficient than those of the other starting points.

References

Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs*. University Press, Oxford.

Box, M.J., and Draper, N.R. (1974). On minimum-point second-order designs. *Technometrics*. 16(4), 613-616.

Chigbu, P.E. and Umelo-Ibemere, N.C. (2015). On slope estimation capabilities of some second order response surface designs. *J. Stat. Appl. Pro.* 4, No.3, 361 - 366.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*, Academic Press Inc., New York.

Gertz, M., Nocedal, J. and Sartenaer, A.(2004). A starting point strategy for nonlinear interior methods. *Applied Mathematics Letters*, 17, 945-952.

Heredia-Langner, A., Carlyle, W. M., Montgomery, D.C., Borror, C.M., and Runger, G.C. (2003) Genetic algorithms for the construction of $D-$optimal designs. *Journal of Quality Technology* 35(1), 28-46.

Kiefer, J., and Wolfowitz, J.(1959). Optimal designs in regression problems. *Ann Math*, 12,363-366.

Khuri, A.I., Kim, H.J., and Um, Y. (1996). Quantile plots of the prediction variance of response surface designs. *Computational Statistics and Data Analysis* 22, 395-407.

Mitchell, T.J. (2000). An algorithm for the construction of $'D-$ optimal$'$ experimental designs. *Technometrics* 42 (1),48-54.

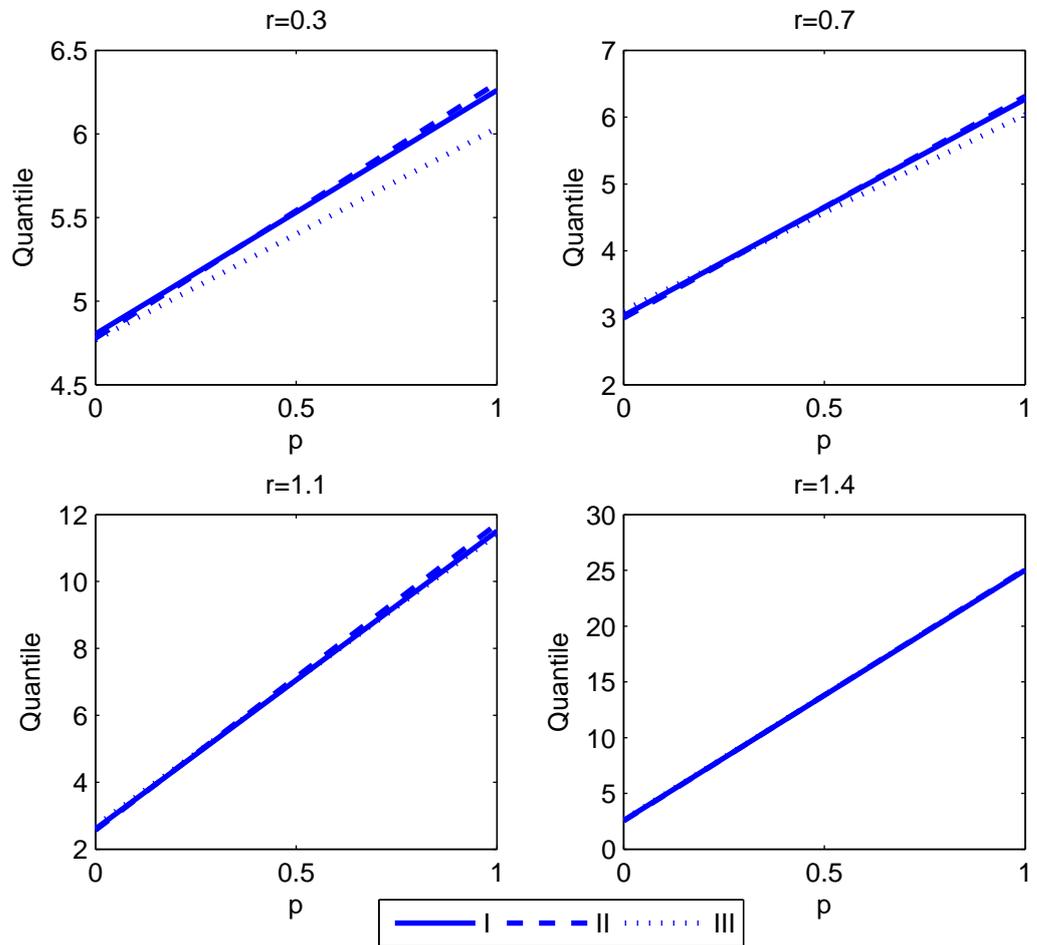Montgomery, D.C. (2013). Design and Analysis of Experiments. John Wiley and Sons, Inc.

Fig. 1. quantiles of the variance functions, $V(\underline{x}, \xi_n^{(*)})$ for the $D-$optimal designs using starting points I, II,and III

Nocedal, J. and Wright, S.J. *Numerical Optimization*, Springer-Verlag, New York.

Onukogu, I.B. (1997). *Foundations of Optimal Exploration of Response Surfaces*. Ephrata press, Nsukka.

Onukogu, I.B. (2002). Super convergent lines series: algorithms and properties. In: *Super Convergent Lines Series in Optimal Design of Experiment and Mathematical Programming*, (Edited by I.B. Onukogu and P.E. Chigbu),112-114, AP Express Publishing Company, Nsukka.

Onukogu, I.B. and Nsude, F. (2014). A global convergent sequence for construction of $D-$optimal exact designs in an infinite space of trials. *Journal of Statistics: Advances in Theory and Applications* 11(1), 39-52.

Torabi, M. and Hosseini, M. (2018). A new descent algorithm using the three-step discretization method for solving unconstrained optimization problems. *Mathematics* 6(4),6

N. C. Umelo-Ibemere, African Journal of Applied Statistics, Vol. 7 (2), 2020, 1001 - 1008. A Note on Comparison of Starting Points for the Generation of D-optimal Second-order Designs. 1008

Ugbe, T.A. and Chigbu, P.E. (2013). On non overlapping segmentation of the response surfaces for solving constrained programming problems through super convergent lines series. *Communications in Statistics-Theory and Methods* 43 (2), 306-320.