Proceedings of Manitoba's Undergraduate Science and Engineering Research
Zheng & Mandal, 2018, PMUSER, 4(1): 63–68.

DOI: 10.5203/pmuser.201841670

Research Article

Construction of A-optimal Designs for Linear Models

Xiang Zheng[1], Saumen Mandal[1]

[1]Dept. of Statistics, University of Manitoba, Winnipeg, MB, R3T 2N2

Corresponding Author: X. Zheng (zhengx34@myumanitoba.ca)

Abstract

For estimating parameters of a statistical model, it is important to minimize the variances of the estimators. Efficiency of an estimator increases as its variance becomes smaller. Sometimes instead of minimizing the variances of the individual parameters, it is important to minimize the total or average variance of all the parameter estimators. This refers to A-optimality in the context of optimal experimental design. Motivated by this fact, we construct A-optimal designs for some regression models using a class of algorithms, indexed by a function which depends on the derivatives of the criterion function. We also develop strategies for constructing A-optimal designs and investigate techniques for improving convergence rates by using the properties of the directional derivatives of the criterion function. Computational studies show that convergence of the algorithm improves a great deal when amended by the properties of the directional derivatives. We explored the design construction through some examples including one practical problem arising in chemistry.

Keywords: Average Variance, Directional Derivatives, Multiplicative Algorithms, Optimal Designs, Parameter Estimation

1 Introduction

Optimal designs are constructed according to a statistical criterion for a specific statistical model. The objective is good estimation of the parameters of the model. There are a variety of criteria defining good estimation, the popular ones being D, G, A and linear optimality. In D-optimality, we minimize the determinant of the covariance matrix of the parameter estimators. That is, in this optimality, the generalized variance of the parameter estimators is minimized. Note that because of the reciprocity property of the covariance matrix and the information matrix, minimizing the determinant of the covariance matrix is equivalent to maximizing the determinant of the information matrix. In G-optimality, we minimize the maximum standardized variance of the predicted response over the design space. This optimality may be useful when a researcher is interested in predicting the outcome variable as efficiently as possible over the design space. In this present work, we focus on A-optimal criterion and construct such designs for some models of interest. An A-optimal design seeks to minimize the sum of the variances of the parameter estimators or their average variance. Some motivations for A-optimality are given at the end of this section. A full description of this criterion is given in Section 2.

As the present work is based on optimal design theory, we start with a brief introduction to this area. We first assume a probability model. A probability model is a mathematical representation of a random phenomenon. It is defined by its sample space. A sample space for a probability model is a collection of all possible outcomes of a random experiment. A sample space could be discrete or continuous. In optimal design context, we first assume a probability model of the type

$$y \sim \pi(y|\boldsymbol{x}, \boldsymbol{\theta}, \sigma)$$

where y is the response variable; $\boldsymbol{x}$ are design variables, $\boldsymbol{x} \in \boldsymbol{\mathcal{X}} \subseteq \mathbb{R}^m$, $\boldsymbol{\mathcal{X}}$ is the design space; $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)^\top$ are unknown parameters; σ is a nuisance parameter; and $\pi(.)$ is a probability model. The objective of an optimal design is good estimation of the parameters of the model. As discussed earlier, there are a variety of criteria defining good estimation. When the model is linear in the parameters, we further assume that $y(\boldsymbol{x})$ has an expected value of the explicit form $E(y|\boldsymbol{v}) = \boldsymbol{v}^\top\boldsymbol{\theta}$, where $\boldsymbol{v} \in \mathcal{V}$, $\mathcal{V} = \{\boldsymbol{v} \in \mathbb{R}^k\colon \boldsymbol{v} = \eta(\boldsymbol{x})\}$ with $\eta(\boldsymbol{x}) = (\eta_1(\boldsymbol{x}), \eta_2(\boldsymbol{x}), \ldots, \eta_k(\boldsymbol{x}))^\top$, a vector of k real valued functions defined on the design space $\mathcal{X}$. The space $\mathcal{V}$ is called the induced design space (or design locus) because $\mathcal{V}$ is the image under a set of regression functions η of $\mathcal{X}$. We discuss how we obtain $\mathcal{V}$ from the original design space $\mathcal{X}$ in the following.

In optimal design theory, an approximate design is characterized by a probability measure, say p, defined on the design space $\mathcal{X}$ and hence on $\mathcal{V}$. In practice we must discretize these spaces. Suppose that we discretize the design space $\mathcal{X}$ into J distinct points, say, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_J$. Thus, $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_J\}$, where $\boldsymbol{v}_j = \eta(\boldsymbol{x}_j)$, $j = 1, 2, \ldots, J$.

At this point we specify p by a set of weights or proportions p_j satisfying $p_j \geq 0$, $j = 1, 2, \ldots, J$, $\sum p_j = 1$. We assign weight p_j to $\boldsymbol{v}_j$. We wish to choose the vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_J)^\top$ optimally. If $\hat{\boldsymbol{\theta}}$ is the least squares estimator of $\boldsymbol{\theta}$, then the covariance matrix $cov(\hat{\boldsymbol{\theta}}) \propto \boldsymbol{M}^{-1}(\boldsymbol{p})$, where $\boldsymbol{M}(\boldsymbol{p})$ is the per observation information matrix. The matrix $\boldsymbol{M}(\boldsymbol{p})$ is given by

$$\boldsymbol{M}(\boldsymbol{p}) = \sum_{j=1}^{J} p_j \boldsymbol{v}_j \boldsymbol{v}_j^\top = \boldsymbol{V} P \boldsymbol{V}^\top \tag{1}$$

where $\boldsymbol{V} = [\boldsymbol{v}_1\, \boldsymbol{v}_2\, \ldots\, \boldsymbol{v}_J]$ and $\boldsymbol{P} = \text{diag}(p_1, p_2, \ldots, p_J)$. In order to ensure good estimation of $\boldsymbol{\theta}$, we wish to choose the proportion p_j of observations taken at $\boldsymbol{x}_j$ by optimizing some criterion, say $\phi(\boldsymbol{p})$. There are many useful books in optimal design [1, 2, 3, 4, 5].

The aim of this paper is to construct A-optimal designs in linear regression models. In A-optimality, we minimize the sum of the variances of the parameter estimators or their average variance. This criterion was introduced by Elfving[6]. There is an extensive literature available for this criterion. The alphabetical nomenclature for different design criteria was introduced by Kiefer[7]. As the trace is the sum of the main diagonal elements of a matrix, the A-optimal criterion minimizes the trace of the covariance matrix of the parameter estimators. A-optimality is important in the sense that we always try to minimize the variances of the parameters of a statistical model. Instead of minimizing the variances of the individual parameters, this optimality minimizes the total or average variance of all the parameter estimators. We construct A-optimal designs using a class of multiplicative algorithms, indexed by a function which depends on the derivatives of the A-criterion function. The function is positive, increasing and may depend on a free positive parameter. The goal is also to develop strategies for constructing A-optimal designs and investigate techniques for improving convergence rates by using the properties of the directional derivatives of the criterion function.

We take the criterion function $\phi(\boldsymbol{p})$ to be the A-optimality criterion. We maximize $\phi(\boldsymbol{p})$ subject to $p_j \geq 0$ and $\sum_{j=1}^{J} p_j = 1$. So we consider an example of the following general problem.

$$\text{Maximize } \phi(\boldsymbol{p}) \text{ over } \mathcal{P} \equiv \{\boldsymbol{p} = (p_1, p_2, \ldots, p_J) : \tag{2}$$

$$p_j \geq 0, \sum_{j=1}^{J} p_j = 1\}$$

The equality constraint $\sum p_j = 1$ renders the problem a constrained optimization problem. Note also that $\mathcal{P}$ is a probability simplex. The probabilities are nonnegative and sum to one. The set is closed and bounded. By the definition of convexity, the full constraint region is convex.

2 Methods

Our general problem is to maximize a criterion $\phi(\boldsymbol{p})$ subject to $p_j \geq 0$, $j = 1, 2, \ldots, J$ and $\sum p_j = 1$. In order to solve this problem, we first need to determine the optimality conditions.

2.1 Optimality Conditions and a Class of Algorithms

We determine the optimality conditions in terms of point to point directional derivatives. We use differential calculus and exploit the directional derivative of Whittle[8]. The directional derivative $F_\phi\{\boldsymbol{p}, \boldsymbol{q}\}$ of a criterion function $\phi(.)$ at $\boldsymbol{p}$ in the direction of $\boldsymbol{q}$ is defined as

$$F_\phi(\boldsymbol{p}, \boldsymbol{q}) = \lim_{\varepsilon \downarrow 0} \frac{\phi\{(1-\varepsilon)\boldsymbol{p} + \varepsilon \boldsymbol{q}\} - \phi(\boldsymbol{p})}{\varepsilon}. \tag{3}$$

The derivative $F_\phi\{\boldsymbol{p}, \boldsymbol{q}\}$ exists even if the criterion function $\phi(.)$ is not differentiable. If $\phi(.)$ is differentiable, (3) can be simplified as: $F_\phi(\boldsymbol{p}, \boldsymbol{q}) = (\boldsymbol{q} - \boldsymbol{p})^\top \partial\phi/\partial\boldsymbol{p}$. Let $F_j = F_\phi(\boldsymbol{p}, \boldsymbol{e}_j)$, where $\boldsymbol{e}_j$ is the j^{th} unit vector in $\mathbb{R}^J$. So, F_j can be simplified as

$$F_j = d_j - \sum_{i=1}^{J} p_i d_i \tag{4}$$

where $d_j = \partial\phi/\partial p_j$, $j = 1, 2, \ldots, J$. As F_j is the directional derivative of the criterion function $\phi(.)$ at $\boldsymbol{p}$ in the direction of the extreme vertex $\boldsymbol{e}_j$, we call F_j the vertex directional derivative of the criterion $\phi(.)$ at $\boldsymbol{p}$.

Now, if $\phi(.)$ is differentiable at an optimizing distribution $\boldsymbol{p}^*$, then the first-order conditions for $\phi(\boldsymbol{p}^*)$ to be a local maximum of $\phi(.)$ in the feasible region of the problem are

$$F_j^* = F_\phi\{\boldsymbol{p}^*, \boldsymbol{e}_j\} \begin{cases} = 0 & \text{for } p_j^* > 0 \\ \leq 0 & \text{for } p_j^* = 0. \end{cases} \tag{5}$$

If the criterion $\phi(.)$ is concave on the feasible region, then the first-order conditions (5) are both necessary and sufficient for optimality, a result known as the general equivalence theorem in optimal design [8, 19].

It is typically not possible to evaluate an optimal solution explicitly. So, we often require an algorithm in order

to construct the optimizing distribution. A class of algorithms which neatly satisfy the basic constraints of the optimal weights take the form

$$p_j^{(r+1)} \quad \propto \quad p_j^{(r)} f(d_j^{(r)}) \tag{6}$$

where $d_j^{(r)} = \partial\phi/\partial p_j$ at r^{th} iterate $\boldsymbol{p} = \boldsymbol{p}^{(r)}$ and the function $f(.)$ satisfies the conditions that it's positive and strictly increasing. The function $f(.)$ may depend on a free positive parameter δ. Torsney[9] first proposed this type of iteration by taking the function $f(d) = d^\delta$, $\delta > 0$. Silvey, Titterington, and Torsney[10] studied the choice of δ when the function $f(d)$ takes the same form as above, i.e., d^δ. Torsney[11] considered the choice $f(d) = e^{\delta d}$, where the partial derivatives could be both positive and negative. Torsney[12] explored the monotonicity property of particular values of the free parameter δ. Titterington[13] describes a proof of monotonicity of $f(d) = d$ for constructing D-optimal designs. Chowdhury, Chen, and Mandal[14] used the above algorithm and considered a class of optimization problems on minimizing variance based criteria in respect of parameter estimators. Mandal, Torsney, and Carriere[15], and Mandal and Torsney[16] further developed the algorithm based on a constrained optimization problem and a clustering approach, respectively. Mandal, Torsney, and Chowdhury[17] used a Lagrangian approach and constructed optimal designs by minimizing a covariance criterion. They established optimality conditions for a non-standard criterion function. The conditions are given in terms first and second order conditions.

2.2 Optimizing Distribution and the A-optimality

Our general problem is given in (2), where $\phi(\boldsymbol{p})$ is a criterion function of interest. In the present work, we take $\phi(\boldsymbol{p})$ as the A-optimality criterion. We mentioned earlier that instead of minimizing the variances of the individual parameters of a model, it is also important to minimize the total or average variance of all the parameter estimators. In order to minimize the total or average variance of all the parameter estimators, we need to minimize the A-optimality criterion. In Section 1, we have seen that the covariance matrix of $\hat{\boldsymbol{\theta}}$ is actually the inverse of the information matrix $M(\boldsymbol{p})$. Because of this reciprocity property, minimizing the variance corresponds to maximizing the information. In terms of a maximization problem, the A-optimality criterion is defined by

$$\phi_A(\boldsymbol{p}) \quad = \quad \psi_A\{\boldsymbol{M}(\boldsymbol{p})\} \quad = \quad -Trace\{\boldsymbol{M}^{-1}(\boldsymbol{p})\}. \tag{7}$$

The above criterion has some properties. The criterion is concave and an increasing function over $\mathcal{M}$, where $\mathcal{M}$ is the set of all positive definite symmetric matrices. The criterion is differentiable whenever it is finite, and the first derivative is given by

$$\frac{\partial\phi_A}{\partial p_j} \quad = \quad \boldsymbol{v}_j^T\, \boldsymbol{M}^{-2}(\boldsymbol{p})\, \boldsymbol{v}_j. \tag{8}$$

The A-optimality criterion was considered by Elfving[6] and Chernoff[18] and subsequently studied [1, 2, 3, 5, 19, 20, 21]. Recently Chowdhury, Chen, and Mandal[14] considered a class of optimization problems on minimizing variance based criteria in respect of parameter estimators of a linear model. They did not consider the A-optimality directly. Instead of considering all the parameters, they considered minimizing the total variance of the estimators of some parameters of interest.

It can be easily shown that A-optimality is a special case of Linear optimality in which we minimize the criterion $\phi_L(\boldsymbol{p}) = \text{Trace}\{\boldsymbol{M}^{-1}(\boldsymbol{p})\boldsymbol{L}\}$, where $\boldsymbol{L}$ is a $k \times k$ matrix of coefficients. Suppose the matrix $\boldsymbol{L}$ is of rank s, where $s \leq k$. Then $\boldsymbol{L}$ can be expressed as: $\boldsymbol{L} = \boldsymbol{A}^\top\boldsymbol{A}$, where $\boldsymbol{A}$ is a $s \times k$ matrix of rank s. Then the criterion $\phi_L(\boldsymbol{p})$ can be written as $\phi_L(\boldsymbol{p}) = \text{Trace}\{\boldsymbol{M}^{-1}(\boldsymbol{p})\boldsymbol{A}^\top\boldsymbol{A}\} = \text{Trace}\{\boldsymbol{A}\boldsymbol{M}^{-1}(\boldsymbol{p})\boldsymbol{A}^\top\}$. When $\boldsymbol{A} = \boldsymbol{c}^\top$, where $\boldsymbol{c}$ is a $k \times 1$ vector, $\phi_L(\boldsymbol{p})$ corresponds to the c-optimality criterion. When $\boldsymbol{A}$ or $\boldsymbol{L}$ is an identity matrix, $\phi_L(\boldsymbol{p})$ corresponds to the A-optimality criterion.

2.3 Construction of A-optimal Designs

As we discussed earlier, problem (2) has a set of constraints on the design weights, namely, $p_j \geq 0$ and $\sum_{j=1}^J p_j = 1$. An iteration which neatly submits to these constraints is given in (6). As we mentioned, this type of iteration was first proposed by Torsney[9]. The function $f(.)$ in the algorithm may depend on a positive parameter δ. The full form of the algorithm is given by

$$p_j^{(r+1)} \quad = \quad \frac{p_j^{(r)} f(x_j^{(r)},\, \delta)}{\sum_{j=1}^{J} p_j^{(r)} f(x_j^{(r)},\, \delta)} \tag{9}$$

where $x_j^{(r)} = d_j^{(r)}$, the partial derivatives evaluated at $\boldsymbol{p}^{(r)}$. The function $f(\cdot,\cdot)$ is positive and strictly increasing in x. The function depends on a free positive parameter δ. When the partial derivatives are positive, a typical choice of $f(\cdot,\cdot)$ is d^δ[9]. When the partial derivatives are both positive and negative, a choice of $f(\cdot,\cdot)$ is $e^{d\delta}$[11]. Mandal and Torsney[16] considered the choice of $f(\cdot,\cdot)$ as d^δ, and further developed the algorithm based on a clustering approach. Algorithm (9) possess several nice properties. Any iterate $\boldsymbol{p} = \boldsymbol{p}^{(r)}$ is always feasible. An iterate $\boldsymbol{p}^{(r)}$ is a fixed point of the iteration if the

derivatives $\partial\phi/\partial p_j^{(r)}$ corresponding to nonzero $p_j^{(r)}$ are all equal.

However, convergence of the algorithm could be slow if we do not choose the function $f(\cdot,\cdot)$ in an objective way. We need to develop strategies for better convergence of the algorithm for constructing designs that optimize the A-optimal criterion. We attempt to improve the convergence by considering the first argument of the function $f(\cdot,\cdot)$ as the vertex directional derivatives of the A-optimal criterion function. Recall that from equation (5) the first order conditions for optimality are $F_j = 0$ for $p_j^* > 0$, and $F_j \leq 0$ for $p_j^* = 0$. Recall also that from equation (4) we have $F_j = d_j - \sum_{j=1}^{J} p_j d_j$. So the vertex directional derivatives are both positive and negative. Using equation (4), we can prove that $\sum_{j=1}^{J} p_j F_j = 0$. This suggests that we can improve the convergence of the algorithm if we choose a function which is centred at zero and also changes quickly about the value $F = 0$. It is also important that we should treat positive and negative directional derivatives symmetrically. One choice of $f(x,\delta)$ with the potential to satisfy these requirements is the normal cumulative distribution function. That is, $f(x,\delta) = \Phi(\delta x)$. This function changes quickly at $x = F = 0$. If we take x as the partial derivatives of the A-optimal criterion, this choice of $f(x,\delta)$ could be bad because the partial derivatives of A-optimal criterion are not centred at zero. The convergence of the algorithm will also depend on the choice of the parameter δ. Depending on the numerical values of the partial and directional derivatives, we need to choose the values of δ carefully.

3 Examples, Results, & Discussion

3.1 Example 1 — Quadratic Regression

We first construct A-optimal design to the quadratic regression model. This is a polynomial regression model in one variable. In polynomial regression, the regression function $E(Y|x)$ is nonlinear in the design variable x. However, the regression function is linear in the parameters. Therefore polynomial regression is considered to be a special case of multiple linear regression. Quadratic regression is a polynomial regression of order two. The model is given by

$$E(Y|x) = \theta_1 + \theta_2 x + \theta_3 x^2$$

with the design interval [-1, 1]. The design space is continuous. So we discretize the design space to be in some form of uniform grid on the continuous design space. In particular, we approximate the design interval by a grid of points equally spaced at intervals of 0.01. We report the performance of algorithm (9), by taking the first argument of the function $f(x,\delta)$ as both the partial and directional derivatives of the A-optimal criterion. As discussed earlier, we take $f(x,\delta) = \Phi(\delta x)$. We record, for $n = 1, 2, \ldots, 6$, the number of iterations needed to achieve $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-n}$, where F_j are the vertex directional derivatives. We take the initial design to be $p_j^{(0)} = 1/J$, $j = 1, 2, \ldots, J$. Results are reported in Table 1 for $x = d$ and in Table 2 for $x = F$. Results for the best choices of δ are given in bold numbers. Note that the best choices of δ are determined by the least number of iterations at $n = 6$. Note that, for $x = F$ (Table 2), the design did not converge beyond the value of $\delta = 0.15$. This is why the best choice is given in the last row of this table. The algorithm converges to a solution having three support points, namely -1, 0 and 1 with corresponding weights (0.25, 0.50, 0.25). The directional derivatives corresponding to the above three support points are zero and are negative towards all zero weighted remaining design points. Therefore the design satisfies the first-order optimality conditions (5).

Table 1: *Number of iterations needed to achieve* $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-n}$ *for the Quadratic Regression Model with* $f(x,\delta) = \Phi(\delta x)$, $x = d$.

δ	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
0.05	162	1755	17729	64702	106506	147568
0.10	130	1352	13565	49469	81419	112802
0.11	**131**	**1354**	**13567**	**49470**	**81418**	**112799**
0.12	134	1373	13736	50073	82408	114169
0.20	237	2166	21310	77542	127568	176705

Table 2: *Number of iterations needed to achieve* $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-n}$ *for the Quadratic Regression Model with* $f(x,\delta) = \Phi(\delta x)$, $x = F$.

δ	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
0.10	56	616	6241	22781	37500	51957
0.12	46	514	5201	18984	31250	43298
0.14	39	441	4459	16273	26786	37113
0.15	**59**	**412**	**4162**	**15189**	**25001**	**34639**

As we discussed earlier, we attempt to improve the convergence of the algorithm by using the directional derivatives of the A-optimal criterion as the first argument of the function $f(x,\delta) = \Phi(\delta x)$. Results in Tables 1 and 2 clearly illustrate that convergence is improved considerably. For example, with $x = d$, $\delta = 0.11$ and $n = 6$, the number of iterations needed to converge to the A-optimal design is 112799 (Table 1), whereas by using $x = F$ and $\delta = 0.15$, this number reduces to 34639 (Table 2).

3.2 Example 2 — A Practical Problem in Chemistry

We now consider a model which is used in a practical problem arising in Chemistry. The regression model describes the relationship between the viscosity y and the concentration x of a chemical solution. Viscosity is the response. The model is given by

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2$$

with the design interval restricted to (0.0, 0.2]. This model was considered[22] for constructing designs for minimally dependent observations.

Note that there is no intercept in this model. The design space is continuous. As we discussed in Section 1, we discretize the design space to be in some form of uniform grid of points equally spaced at intervals of 0.01. We report the performance of algorithm (9), by taking the first argument of $f(x, \delta)$ as both the partial and directional derivatives of the A-optimal criterion. We record the number of iterations needed to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-n}$. We take the initial design to be $p_j^{(0)} = 1/J, j = 1, 2, \ldots, J$. Results are reported in Table 3 for $x = d$ and Table 4 for $x = F$. Note that, for this particular model, both the partial and directional derivatives are numerically very large, so we needed smaller values of the parameter δ compared to the previous model. The algorithm converges to a solution having three support points, namely 0.01, 0.12 and 0.20 with corresponding weights (0.413419, 0.380949, 0.205632). Here also the directional derivatives corresponding to the above three support points are zero and are negative towards all zero weighted remaining design points. Therefore the design satisfies the first-order optimality conditions (5).

In this model, with $x = d$, $\delta = 7 \times 10^{-6}$ and $n = 6$, the number of iterations needed to converge to the A-optimal design is 9927 (Table 3), whereas by using $x = F$ and $\delta = 1.005 \times 10^{-05}$, this number reduces to 2863 (Table 4). Thus we see that convergence of the algorithm is improved considerably by using the directional derivatives of the A-optimal criterion function.

Table 3: *Number of iterations needed to achieve* $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-n}$ *for the Viscosity Model with* $f(x, \delta) = \Phi(\delta x)$, $x = d$.

$\delta \times 10^{-6}$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
3	5918	7458	8998	10537	12077	13615
6	4357	5489	6622	7754	8886	10018
7	**4319**	**5441**	**6562**	**7684**	**8806**	**9927**
8	4410	5555	6699	7844	8989	10133
10	4947	6230	7513	8796	10078	11360

Table 4: *Number of iterations needed to achieve* $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-n}$ *for the Viscosity Model with* $f(x, \delta) = \Phi(\delta x)$, $x = F$.

$\delta \times 10^{-05}$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
1.0	1277	1610	1943	2277	2609	2942
1.004	1233	1547	1921	2263	2597	2927
1.005	**1491**	**1763**	**2037**	**2311**	**2589**	**2863**
1.006	1919	2273	2629	2983	3337	3693
1.008	4723	5637	6553	7467	8385	9309

4 Conclusions

In the present work, we addressed an important problem of optimal design and statistical inference. The objective was good estimation of the parameters. For a statistical model, it is important to estimate the parameters with minimum variance.

We considered A-optimal designs and minimized the total or average variance of all the parameter estimators. In order to solve this optimization problem, we minimized the trace of the covariance matrix of the parameter estimators. Because of the reciprocity property of the covariance matrix and the information matrix, minimizing the variance corresponds to maximizing the information. We determined the optimality conditions in terms of point to point directional derivatives. In particular, we expressed the optimality conditions in terms of vertex directional derivatives of the criterion function. We constructed the A-optimal designs by using a class of algorithms which neatly fit the basic constraints of our optimization problem. We then developed techniques for improving convergence rates by using the properties of the directional derivatives of the criterion function. Computational studies show that convergence of the algorithm improves a great deal when amended by the proposed approach based on using the properties of the directional derivatives. We constructed optimal designs for some models including one practical model which describes the relationship between the viscosity and the concentration of a chemical solution.

5 Acknowledgements

The authors sincerely thank the editor and two referees for their comments and suggestions that have resulted in an improved version of this manuscript. This research was done while Xiang Zheng was completing his 2018 summer research project under the supervision of Professor Saumen

Mandal. Mr. Zheng expresses his sincerest gratitude to Prof. Mandal for his continuous support, encouragement and advice. The research was supported by Prof. Mandal's Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

1. Fedorov, V. V. 1972. *Theory of optimal experiments*. Academic Press, New York.
2. Silvey, S. D. 1980. *Optimal design: an introduction to the theory for parameter estimation*. Chapman and Hall, London.
3. Atkinson, A. C., Donev, A. N., & Tobias, R. D. 2007. *Optimum experimental designs, with SAS*. Oxford University Press, Oxford.
4. Pukelsheim, F. 2006. *Optimal design of experiments*. Society for Industrial and Applied Mathematics, Philadelphia.
5. Berger, M. P. F. & Wong, W. K. 2009. *An introduction to optimal designs for social and biomedical research*. John Wiley and Sons, West Sussex.
6. Elfving, G. 1952. *Annals of Mathematical Statistics*, 23: 255–262.
7. Kiefer, J. 1959. *Journal of the Royal Statistical Society: Series B*, 21: 272–319.
8. Whittle, P. 1973. *Journal of the Royal Statistical Society: Series B*, 35: 123–130.
9. Torsney, B. 1977. *Journal of the Royal Statistical Society: Series B*, 39: 26–27.
10. Silvey, S. D., Titterington, D. M., & Torsney, B. 1978. *Communications in Statistics — Theory and Methods*, 7: 1379–1389.
11. Torsney, B. 1988. *In: Optimal Design and Analysis of Experiments*, (Edited by Y. Dodge, V. V. Fedorov, & H. P. Wynn), Elsevier Science Publishers B.V., North Holland, 361–370.
12. Torsney, B. 1983. *In: Proceedings of the International Symposium on Semi-Infinite Programming and Applications, Lecture Notes in Economics and Mathematical Systems*, volume 215, (Edited by K. O. Kortanek & A. V. Fiacco), Springer, Berlin, 249–260.
13. Titterington, D. M. 1976. *In: Proceedings of the 1976 Conference on Information Sciences and Systems, John Hopkins University, Baltimore*, volume 3, 213–216.
14. Chowdhury, M., Chen, M., & Mandal, S. 2018. *Communications in Statistics — Simulation and Computation*. (accepted).
15. Mandal, S., Torsney, B., & Carriere, K. C. 2005. *Journal of Statistical Planning and Inference*, 128: 609–621.
16. Mandal, S. & Torsney, B. 2006. *Journal of Statistical Planning and Inference*, 136: 1120–1134.
17. Mandal, S., Torsney, B., & Chowdhury, M. 2017. *Australian and New Zealand Journal of Statistics*, 59: 255–273.
18. Chernoff, H. 1953. *Annals of Mathematical Statistics*, 24: 586–602.
19. Kiefer, J. 1974. *Annals of Statistics*, 2: 849–879.
20. Wu, C. F. J. 1978. *Communications in Statistics — Theory and Methods*, 7: 1399–1412.
21. Fellman, J. 1989. *Journal of Statistical Planning and Inference*, 21: 85–92.
22. Torsney, B. & Alahmadi, A. M. 1995. *Statistica Sinica*, 5: 499–514.