

Computing A D_s -Optimal Design Augmentation Scheme Using PROC IML (And Why SAS/QC® Can't Do It)

Paul T. Savarese, East Tennessee State University, Johnson City, TN

Raymond H. Myers, Virginia Polytechnic Institute and State University, Blacksburg, VA

ABSTRACT

In the design of experiments it is often desired to augment a given design with additional runs. The augmentation is done with respect to some optimality criterion such as D-optimality. The first such algorithm was DETMAX developed by the late Toby Mitchell which augmented points in a D-optimal fashion. D_s -optimality involves optimizing with respect to a subset of the regressor variables, a scenario with particular applications in robust design. This paper will discuss a method for augmenting an experimental design in a D_s -optimal fashion and illustrates an algorithm using PROC IML to implement the method. An explanation of why SAS/QC cannot perform a D_s -optimal augmentation scheme will be given.

1. INTRODUCTION

In this paper we discuss the augmentation of an existing response-surface experimental design with additional runs (a "run" being a single measurement of the response variable(s) at one set of experimental conditions) selected in order to optimize some criterion. Response-surface experiments usually involve regressors that are controlled by the experimenter and not observed, as might be the case in a more general regression setting. Two related problems present themselves; these are augmentation of an existing design and design-building or producing a design from scratch. In either problem design points (runs) are chosen for augmentation or exchange one at a time with regard to some optimality criterion. We are concerned chiefly with augmentation, but both problems are attacked in similar fashion. An augmentation algorithm will be developed that adds to a given experimental design those runs, selected from a candidate list, that provide maximal improvement with regard to a particular optimality criteria.

Contemplate a scenario where an experimenter is constrained, perhaps by cost or time considerations, to at most N runs. Augmentation algorithms can be quite useful here especially if no standard design of the required size exists. For example, suppose at most $N=25$ runs may be used to estimate $\underline{\beta}$ in the linear model $\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon}$. Suppose the proposed relationship is a first-order plus two-way interaction multiple-regression model (often called a response-surface model when $\underline{X}$ is controlled) with six main-effects. A full factorial ($2^6 = 64$ runs) far exceeds the number of allotted runs. Furthermore, typical half or quarter-fraction fractional-factorial designs of, respectively, $2^{6-1} = 32$ runs and $2^{6-2} = 16$ runs are either too large or unnecessarily small. Clearly an automated augmentation or design-building algorithm could prove useful to "optimally" utilize all 25 runs available.

We consider two similar optimality criteria: D-optimality and D_s -optimality. D-optimality is an optimality criterion that addresses the entire complement of regressors in the model. D_s -optimality, on the other hand, addresses a similar optimality criteria with regard to a *subset* of the regressors. (Hence the subscript "s" in D_s -optimality.) Details on these two criteria are given below. Additionally, we explain why SAS/QC, though it allows augmentation in a D-optimal sense cannot augment in a D_s -optimal fashion. Later we present a flowchart of our algorithm.

A word about notational conventions is in order. Throughout this paper vectors are underlined, matrices are indicated in bold, and transposes are indicated with primes. Examples of these conventions are, respectively, vector $\underline{y}$, matrix $\underline{X}$, and its transpose $\underline{X}'$.

2. A D-OPTIMAL STRATEGY

In the linear model $\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon}$, $\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$ is the OLS estimator of $\underline{\beta}$. Assuming homogeneous variances, i.e. - $VAR(\underline{\epsilon}) = \sigma^2\mathbf{I}$, results in $VAR(\hat{\underline{\beta}}) = (\underline{X}'\underline{X})^{-1}\sigma^2$. For details see Graybill (1976). Note that $\underline{X}$ is an $n \times p$ matrix where n is the number of runs in the design (also referred to as design points, points, or observations); p is the number of parameters in the model including an intercept, i.e. - the number of columns of $\underline{X}$.

Definition 1 D-optimality

An experimental design is said to be D-optimal if it minimizes the determinant of $(X'X)^{-1}$, denoted $DET[(X'X)^{-1}]$. Equivalently, a design is D-optimal if it maximizes $DET[X'X]$.

In a D-optimal augmentation scheme a design point, say x_i , is chosen that yields minimal improvement in $DET[(X'X)^{-1}]$ or gives maximal improvement in $DET[X'X]$. D-optimality has an easily understandable physical interpretation: a design that is D-optimal for a given model is one that minimizes the volume of the simultaneous confidence hyper-ellipsoid on the coefficient vector β . Historically, algorithms that implement a D-optimal strategy did not rely directly on the computation of determinants. Rather, D-optimal algorithms involve searching for points that give maximal improvement in an equivalent prediction variance entity, with the reason being computational efficiency. Prediction variance entities involve inverses and are more efficient to compute than determinants – especially determinants of large orders. We will not, in this paper, give detailed proofs of the equivalence of using prediction variances rather than determinants. However, some required definitions and theorems will be stated.

The first such algorithm, DETMAX of Mitchell (1974) operated on a prediction variance criterion. The prediction variance, given homoscedasticity, is $VAR(\hat{y}) = X(X'X)^{-1}X' \sigma^2 = H \sigma^2$, where $H = X(X'X)^{-1}X'$. H is the well-known projection matrix or *hat matrix*. Let x_i ($p \times 1$) correspond to the i^{th} row ($i = 1, 2, \dots, n$) of X in column form. Then the diagonals of the hat matrix, known as *hat diagonals* are $h_{ii} = x_i'(X'X)^{-1}x_i$. Well known properties of the hat diagonals are:

$$i) \quad \sum_{i=1}^n h_{ii} = p \quad (2.1)$$

$$ii) \quad 1/n \leq h_{ii} \leq 1 \quad (2.2)$$

For details on the above properties see Myers (1990).

Definition 2 Usual Prediction Variance Criterion

The usual prediction variance criterion at location x_i is

$$VAR[\hat{y}(x_i)] / \sigma^2 = x_i'(X'X)^{-1}x_i = h_{ii} \quad (2.3)$$

For a D-optimal strategy we wish to identify the candidate location x_i that provides maximum increase in the prediction variance. Admittedly, this seems contrary to intuitive notions of optimality. The mystery will be resolved shortly. Actually, DETMAX used the so-called *deleted* prediction variance. Recall that the augmentation scenario, as we have defined it, involves selecting new or additional design points. In this vein we may think of x_i as some new candidate location. Let X denote the updated X matrix including the candidate point x_i , and let X_{-i} denote the X matrix from the previous stage. That is X_{-i} can be thought of as the updated X matrix with the candidate point removed. Now, we define the *deleted* prediction variance used by DETMAX.

Definition 3 Deleted Prediction Variance Criterion

The *deleted* prediction variance criterion is defined to be

$$VAR[\hat{y}(x_{i,-i})] / \sigma^2 = x_i'(X'_{-i}X_{-i})^{-1}x_i = h_{ii,-i} \quad (2.4)$$

where x_i is a new candidate point and $X'_{-i}X_{-i}$ is $X'X$ from the previous stage, i.e., without the candidate point.

An computational expression for $X'_{-i}X_{-i}$ is

$$X'_{-i}X_{-i} = [X'X - x_i x_i'] \quad (2.5)$$

Details of the expression (2.5) can be found in Appendix B.4 of Myers (1990).

In other words, $h_{ii,-i}$ is the prediction variance at the new point (x_i) using $X'X$ from the previous stage. Mitchell indicated this was an appropriate form in which to use the prediction variance as the D-optimal augmentation criterion.

Though the original DETMAX algorithm relied on the deleted prediction variance $h_{ii,-i}$ it can be shown that maximizing the customary prediction variance h_{ii} is equivalent to maximizing $h_{ii,-i}$. The next theorem establishes the

relationship between $h_{ii,i}$ and h_{ii} . Essentially this allows one to use either the updated matrix X or the previous stage X matrix denoted X_{-i} in the prediction variance criterion. For our eventual D_s -optimal strategy, presented in the next section, it will be advantageous to utilize the updated X matrix. The advantages of using the updated X matrix and usual prediction variance will become clearer when we present the D_s -optimal algorithm. Now we give the relationship between h_{ii} and $h_{ii,i}$.

$$\text{Theorem 1} \quad h_{ii,i} = [h_{ii} / (1 - h_{ii})] \quad (2.6)$$

Proof: (omitted) For details see Savarese(1992).

From Theorem 1 combined with h_{ii} positive and bounded by 1 from (2.2) we observe that maximizing h_{ii} is equivalent to maximizing $h_{ii,i}$. Next we present results that tie the determinant of $X'X$ to h_{ii} .

$$\text{Theorem 2} \quad \text{DET}[X'X] = \text{DET}[X'_{-i}X_{-i}] / (1 - h_{ii}) \quad (2.7)$$

Proof: (omitted) For details see Savarese(1992).

Theorem 2 provides evidence of the appropriateness of maximizing prediction variances rather than determinants. Note that in an algorithmic scenario $X'_{-i}X_{-i}$ is just $X'X$ from a previous stage. Thus, $\text{DET}[X'_{-i}X_{-i}]$ is a constant with respect to the candidate point x_i . Clearly from (2.7), with h_{ii} positive and bounded by 1, maximization of h_{ii} at any candidate location x_i gives a maximum increase in $\text{DET}[X'X]$.

In summary, in an augmentation algorithm, maximizing the deleted prediction variance of Mitchell ($h_{ii,i}$) is equivalent to maximizing the usual prediction variance (h_{ii}) and is equivalent to maximizing $\text{DET}(X'X)$ which is, by definition, a D -optimal augmentation.

3. A D_s -OPTIMAL STRATEGY

Suppose we are interested in a subset of the regressors in X . Thus we partition $X = [X_1 | X_2]$ and we have the linear model $y = X_1\beta_1 + X_2\beta_2 + \epsilon$, where again $\epsilon \sim N(0, \sigma^2 I)$. Our interest is focused on the regressors in X_2 . For example, this might occur in a parameter design situation where one has X_1 as controllable variables and X_2 as noise variables – see Taguchi (1979). The process variance $\text{VAR}(y|x_i)$, at a given location x_i , is affected by the settings of the noise variables; so X_2 (and β_2) would be of primary concern.

Analogous to the D -optimal situation discussed above we have $\text{VAR}(\beta_2) = M^{-1} \sigma^2$, where

$$M = [X'_2 X_2 - X'_2 X_1 (X'_1 X_1)^{-1} X'_1 X_2] = X'_2 [I - X_1 (X'_1 X_1)^{-1} X'_1] X_2 \quad (3.1)$$

and we wish to augment X with the location x_i that minimizes $\text{DET}[M^{-1}]$, or equivalently maximizes $\text{DET}[M]$. This choice of x_i can be shown to minimize the volume of the joint confidence hyper-ellipsoid on the vector β_2 – for details consult Savarese(1992). Let $H_1 = X_1 (X'_1 X_1)^{-1} X'_1$ which is a Hat matrix in X_1 . Note that H_1 is idempotent, i.e. $H_1^2 = H_1$, as is $I - H_1$. Thus,

$$\begin{aligned} M &= X'_2 [I - H_1] X_2 \\ &= X'_2 [I - H_1]^2 X_2 \\ &= X'_2 [I - H_1] [I - H_1] X_2 \\ &= W'W \text{ where } W = [I - H_1] X_2 \end{aligned} \quad (3.2)$$

Note that $\text{VAR}(\beta_2) = (W'W)^{-1} \sigma^2$, which parallels $\text{VAR}(\beta) = (X'X)^{-1} \sigma^2$ in the previous section. This leads to a definition of D_s -optimality that closely resembles that for D -optimality.

Definition 4 D_s -optimality

An experimental design is D_s -optimal if it maximizes $\text{DET}[W'W]$.

D_s -optimality was originally proposed by Kiefer (1961) whose work was extended by Karlin and Studden (1965). A treatment of D_s -optimality in the response surface arena is given in Wardrop and Myers (1990).

Maximizing $\text{DET}[\mathbf{W}'\mathbf{W}]$ is very similar to the D -optimality scenario of maximizing $\text{DET}[\mathbf{X}'\mathbf{X}]$. However, there are some artful differences. These subtleties involve expressing the determinant criteria (maximizing $\text{DET}[\mathbf{W}'\mathbf{W}]$) in terms of "prediction variance" type criteria, i.e. maximizing an h_{ii} type animal.

The results of the previous section allow us to express the "prediction variance" criterion as $w_{ii} = \mathbf{w}'_i (\mathbf{W}'\mathbf{W})^{-1} \mathbf{w}_i$ where $\mathbf{w}_i$ is a row of the $\mathbf{W}$ matrix in column form. Note that row $\mathbf{w}_i$ corresponds to the candidate point $\mathbf{x}_i$ in the foregoing discussion. $\mathbf{W}$ being updated from a previous stage by the addition of the successful candidate point. Hence w_{ii} plays the role that the hat diagonal h_{ii} plays for a D -optimal scheme. Appealing to former results again we could equivalently maximize the *deleted* "prediction variance" entity $w_{ii,-i} = \mathbf{w}'_i (\mathbf{W}'_{-i} \mathbf{W}_{-i})^{-1} \mathbf{w}_i$ which also maximizes $\text{DET}[\mathbf{W}'\mathbf{W}]$ and yields D_s -optimal augmentation. $\mathbf{W}_{-i}$, of course, is the $\mathbf{W}$ matrix from the previous stage of the augmentation. Refer to the discussion following Theorem 2 above.

A point worth mentioning here regards why the undeleted "prediction variance" w_{ii} is used with an updated $\mathbf{W}$ matrix rather than something like $w_{ii,-i}$ similar to DETMAX. Since the updated $\mathbf{W}$ must be computed anyway we use the undeleted form of the prediction variance w_{ii} rather than $w_{ii,-i}$ in our D_s -optimal scheme. This is analogous to using h_{ii} over $h_{ii,-i}$ in a D -optimal scheme. We have demonstrated that using h_{ii} or $h_{ii,-i}$ makes no difference in regards to D -optimality. This principle holds for using w_{ii} rather than $w_{ii,-i}$ as well; it is sheerly a matter of programming convenience.

4. WHY SAS/QC CAN'T PERFORM D_s -OPTIMAL AUGMENTATION

The SAS/QC package has been developed to provide many statistical quality control and experimental design conveniences to experimenters. In particular, SAS/QC can augment an existing design or build a design from scratch relative to several standard optimality criteria including D -optimality. However, we are not aware of any software package that can perform a D_s -optimal augmentation in a straightforward manner. One might ask whether if only the subset of variables of interest in $\mathbf{X}_2$ (from $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2]$) are specified in the SAS/QC D -optimal algorithm (PROC OPTEX) would this provide a D_s -optimal algorithm? The answer, regrettably, is no and the explanation involves rather subtle logic.

Upon examination one finds that specifying the regressors in $\mathbf{X}_2$ alone would, in a D -optimal algorithm, choose points in accordance with maximum increase in $\text{DET}[\mathbf{X}'_2 \mathbf{X}_2]$. Recall that the D_s -optimal criteria does not maximize $\text{DET}[\mathbf{X}'_2 \mathbf{X}_2]$ but $\text{DET}[\mathbf{M}] = \text{DET}[\mathbf{X}'_2 \mathbf{X}_2 - \mathbf{X}'_2 \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2] = \text{DET}[\mathbf{W}'\mathbf{W}]$ where $\mathbf{W}$ is given in (3.2). Furthermore, neither SAS/QC nor any other commercially available computer package has provision for partitioning $\mathbf{X}$ then computing and updating an entity such as $\mathbf{W}$. No modifications nor manipulations of SAS/QC procedures will allow for a D_s -optimal criteria to be used in choosing points for inclusion in the design. A new algorithm is called for. Our algorithm, presented in the next section implements the correct artifice to obtain a D_s -optimal design augmentation.

5. AN AUGMENTATION ALGORITHM FOR A D_s -OPTIMAL STRATEGY

An algorithm was developed to perform augmentation of an existing design via a D_s -optimal augmentation scheme. That is, the candidate point chosen for augmentation is that point which yields the greatest increase in the D_s -optimal criteria.

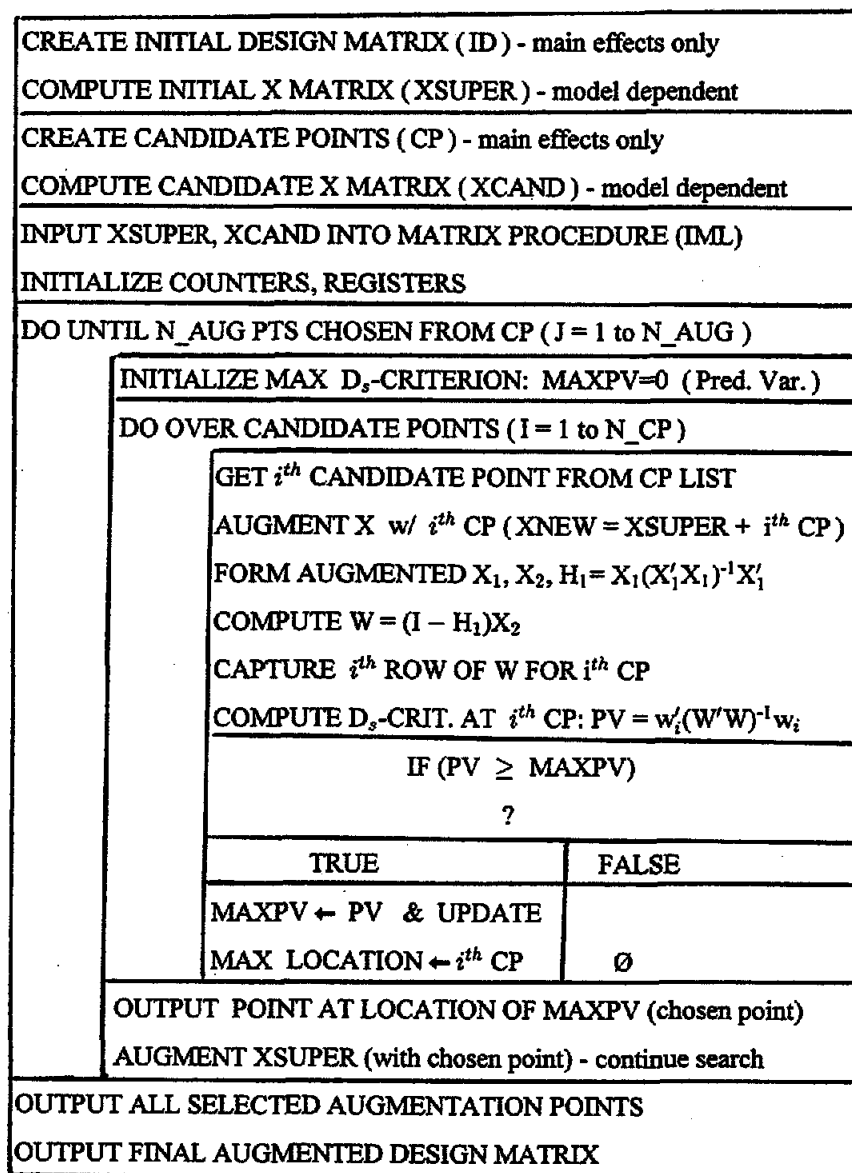
The matrix based computer language, SAS/IML[®] (SAS[®] Interactive Matrix Language) was chosen to implement the algorithm. Also, the SAS Macro Language was used to provide additional flexibility in varying the initial design, candidate points and number of points to be chosen for augmentation. The power of the SAS software system exerts itself here. The DATA STEP was used to easily form and manipulate various initial designs and candidate lists which were then passed to PROC IML inside a macro. The matrix row and column addressing features of PROC IML proved very useful in forming the required vectors, matrices updated matrices involved. A Nassi-Schneiderman (N-S) Flowchart of the algorithm is shown below. The N-S Flowchart depicts a computer program as a single enclosed box, with a sequence of events (implied rather than explicitly stated) proceeding from top to bottom. A brief discussion of the flowchart/algorithm ensues.

In keeping with the standard operation of an augmentation algorithm an initial design (ID) is chosen. An $\mathbf{X}$ matrix (XSUPER) which expresses the response surface model is specified. The initial design is to be augmented with a predetermined number of points (N_{AUG}) selected from a candidate list (XCAND) which is usually a large or very large experimental design, but need not be. Initialization is performed where the D_s -optimality criterion, i.e. - the maximum prediction variance criterion (MAXPV) is set to zero. The algorithm passes through the candidate list considering each point in turn and chooses for augmentation the candidate point that provides maximal increase in the D_s -optimality criterion. As stated earlier this criterion is either $w_{ii} = \mathbf{w}'_i (\mathbf{W}'\mathbf{W})^{-1} \mathbf{w}_i$ or, equivalently, $w_{ii,-i} = \mathbf{w}'_i (\mathbf{W}'_{-i} \mathbf{W}_{-i})^{-1} \mathbf{w}_i$.

In the programming of the D_s -optimal algorithm we must compute $\mathbf{w}_i$ which is not quite as straightforward as using a candidate point $\mathbf{x}_i$ and computing h_{ii} . It is simplest to update $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2]$ with the candidate point $\mathbf{x}_i$, compute $\mathbf{H}_1 = \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1$ and $\mathbf{W} = [\mathbf{I} - \mathbf{H}_1] \mathbf{X}_2$ from updated $\mathbf{X}_1$ and $\mathbf{X}_2$ matrices and then "pull off" row $\mathbf{w}_i$. This avoids having to

deal with $W'_{-i} W_{-i}$ from the previous stage and is easier from a programming standpoint. For programming simplicity and since the updated W would have had to be computed anyway to find $W'_{-i} W_{-i}$, we use the undeleted form of the prediction variance w_{ii} rather than $w_{ii,-i}$ in our D_s -optimal scheme. After the necessary entities are updated with the chosen point, the procedure iterates until the required number of points have been chosen.

The selected points remain in the candidate list in order to allow the algorithm to choose a candidate point more than once. Also, the candidate list usually contains, at minimum, the initial design in order to allow for possible replication of any initial design point. Finally, the chosen list of points and the final design are output. A complete listing of the source code, applying the algorithm to an example design, is available from the author and is given in Savarese (1992). Space limitations require that the source code be given elsewhere.



N-S Flowchart D_s -Optimal Augmentation Algorithm

6. CONCLUSION

We feel that our D_s -optimal algorithm will prove useful in situations where design augmentation is advantageous and primary interest lies with a subset of regressors. It has been shown that such an algorithm can be implemented using the powerful tools available in the SAS software system. A formal D_s -optimal design-building algorithm was not programmed, but would be based on augmenting or exchanging points in accordance with the criteria developed in this paper.

7. REFERENCES

- Graybill, F.A. (1976) *Theory and Application of the Linear Model*, Duxbury Press, Boston
- Karlin, S. and Studden, W.J. (1965) "Optimal Experimental Designs", *Annals of Mathematical Statistics*, 37, 793-815
- Kiefer, J. (1961) "Optimal Designs in Regression Problems II", *Annals of Mathematical Statistics*, 32, 298-328
- Mitchell, T.J. (1974) "An Algorithm for the Construction of D-optimal Experimental Designs", *Technometrics*, 16, 203-210
- Myers, R.H. (1990) *Classical and Modern Regression with Applications*, 2nd Edition, PWS-KENT Publishing Company, Boston
- Savarese, P.T. (1992) "New Design Comparison Criteria in Taguchi's Robust Parameter Design", Ph.D. Dissertation, Department of Statistics, Virginia Polytechnic Institute and State University
- Taguchi, G. (1986) *Introduction to Quality Engineering: Designing Quality into Products and Processes*, Asian Productivity Organization, Tokyo, Japan
- Wardrop, D. and Myers, R.H. (1990) "Some Response Surface Designs for Finding Optimal Conditions", *Journal of Statistical Planning and Inference*, 16, 203-210

SAS, SAS/QC, and SAS/IML are registered trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration.

Paul T. Savarese, Ph.D.
Assistant Professor
Department of Mathematics
East Tennessee State University
P.O. Box 70663
Johnson City, TN 37614
(423) 439-5594
SAVARESE@ETSU.EAST-TN.EDU

Raymond H. Myers, Ph.D.
Professor
Department of Statistics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24060