# NUMERICAL EVALUATION OF THE D-EFFICIENCY CRITERION

**Hélcio Vieira Junior**
Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes 50, Vila das Acácias
12228-900 - São José dos Campos – SP
junior_hv@yahoo.com.br

**Karl Heinz Kienitz**
Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes 50, Vila das Acácias
12228-900 - São José dos Campos - SP
kienitz@ita.br

**Mishel Carmen Neyra Belderrain**
Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes 50, Vila das Acácias
12228-900 - São José dos Campos - SP
carmen@ita.br

## RESUMO

O critério D-otimalidade é relacionado à precisão da regressão linear. Colocado em outras palavras, é determinado o impacto de utilizar delineamentos com diferentes níveis de D-otimalidade na precisão dos coeficientes e da predição da regressão.

É mostrado que não é possível relacionar o critério D-otimalidade a nenhuma das medidas de eficiência (MDE) utilizadas, i.e., não há uma única MDE que seja uma função monotônica absoluta no intervalo estudado.

**PALAVARAS CHAVE. Delineamento de Experimentos, Superfície de Resposta, Estatística.**

**Área principal: Estatística**

## ABSTRACT

We related the D-optimality criterion to the precision of the linear regression. Put in other words, we assessed the impact of using designs with different D-optimality levels in the precision of the regression coefficients and predictions.

It was shown that it is not possible to relate the D-efficiency criterion to any of the measures of efficiency (MOE) used, i.e., there is not a single MOE that is an absolutely monotonic function in the studied interval.

**KEYWORDS. Design of Experiments, Response Surface, Statistics.**

**Main area: Statistics**

## 1. Introduction

"There is not a single area of science and engineering that has not successfully employed statistically designed experiments" (Montgomery, 2005).

The Design of Experiments (DoE) was born in the 1920's through the pioneering work of Fisher (1958) in the agriculture arena. As noted by Montgomery (2005), it has a vast number of success stories.

DoE can be divided in two classifications:

- Classical Designs: designs based on orthogonal arrays (see Hedayat, Sloane and Stufken, 1999); and
- Alphabetic Optimal Designs: designs based on a single efficiency criterion.

The latter were created because classical designs may require an excessive number of design points (also known as runs) for certain combination of factors. For example: a problem with 6 numerical discrete factors with 2, 2, 5, 7, 7 and 11 levels requires 10780 runs (the only classical DoE suitable for this problem is the Full Factorial), which makes its utilization impractical.

One of the approaches utilized to solve this problem, also known as "the curse of dimensionality", was the development of criteria in which several designs could be evaluated with the objective of selecting the best.

The most popular criteria are given bellow, where $X$ is the design matrix which has $n$ rows and $p$ columns, and $x$ is a $1\times p$ matrix that describes a feasible combination of levels of the $p$ design factors:

- **A-Optimality**: deals with the variances of the regression coefficients by minimizing the trace of the design information matrix: $Min\ trace\left(\left(X^T X\right)^{-1}\right)$;
- **G-Optimality**: deals with the variance of the prediction by minimizing the maximum scaled prediction variance: $Min\ Max\ Nx^T\left(X^T X\right)^{-1}x$;
- **D-Optimality**: deals with the volume of the joint confidence region on the vector of regression coefficients by minimizing the determinant of the information matrix: $Min\left|\left(X^T X\right)^{-1}\right|$.

According to Montgomery (2005, p. 441), "[p]erhaps the most widely used [criterion] is the D-optimality".

The objective of this paper is to relate the D-optimality to the precision of the regression. Put in other words, it is desired to assess the impact of using designs with different D-optimality levels in the precision of the regression coefficients and predictions.

The organization of this paper is the following. The next section will describe the D-efficiency criterion and the methodology of our study. Numerical results will be presented in Section 3, and Section 4 summarizes our conclusions.

## 2. Methodology

### 2.1 D-Efficiency

To allow for the comparison of different designs, the D-Efficiency criterion was created as given in (1), where $X_i$ is the $i^{th}$ design matrix, $|M|$ is the determinant of matrix $M$, and $p$ is the number of model parameters.

$$D_e = \left( \frac{\left| \left( X_1^T X_1 \right)^{-1} \right|}{\left| \left( X_2^T X_2 \right)^{-1} \right|} \right)^{\frac{1}{p}} \tag{1}$$

The D-Efficiency describes the relative efficiency of one design in relation to another. If, for example, $D_e = 0.5$, this implies that the second design $\left( X_2 \right)$ must be replicated $1/D_e = 2$ times to have the same precision the first has (Montgomery, 2005).

**2.2 Measures of Efficiency**

In order to reach our objective, we compare several designs with different D-Efficiency levels by using the following measures of efficiency:

MOE 1. Mean Euclidian Distance $E_D$ value;

MOE 2. Probability of yielding a smaller Euclidian Distance $E_D$ value;

MOE 3. Mean Missing Coefficients Number $M_c$ value;

MOE 4. Probability of yielding a smaller Missing Coefficients Number $M_c$ value;

MOE 5. Mean Incorrect Coefficients Number $I_c$ value;

MOE 6. Probability of yielding a smaller Incorrect Coefficients Number $I_c$ value;

MOE 7. Mean Prediction Variance $P_v$ value;

MOE 8. Probability of yielding a smaller Prediction Variance $P_v$ value;

MOE 9. Maximum Prediction Variance $P_v$ value;

MOE 10. Mean $R_{adj}^2$ value;

MOE 11. Probability of yielding a greater $R_{adj}^2$ value.

To be able to measure these efficiencies, the following measures of performance are defined:

MP 1. Euclidian Distance $E_D$

The Euclidian distance between the estimated regression coefficients and the true regression coefficient. This measure is given by (2), where $\beta$ is the estimated regression coefficient vector and $\beta$ is the true model coefficient vector.

$$E_D = \sqrt{\left( \beta - \beta \right)^T \left( \beta - \beta \right)} \tag{2}$$

MP 2. Missing Coefficients Number $M_c$

The number of coefficients that are in the true model and were estimated as having a $p-$value (the probability of obtaining a test statistic at least as extreme as the one that was actually observed given that the null hypothesis is true) greater than 5%.

MP 3. Incorrect Coefficients Number $I_c$

The number of coefficients that are not in the true model and were estimated as having a $p-$value smaller than or equal 5%.

MP 4. Prediction Variance $P_v$

The average squared deviation from the true mean value measured in a uniform

$11^p$ grid in space $V=[-1,1]^p$, where $p$ is the number of model parameters.

An example of the first three measures of performance is given bellow. Imagine an experiment with two factors and that the nature behaves according to the model given by (3), where $\varepsilon$ is the random noise which is independent and identically distributed (I.I.D.) Normally distributed with zero mean and variance $\sigma^2$.

$$y=10+2x_1+x_2^2+\varepsilon \quad (3)$$

If the experimenter tries to fit a full second order model $\left(y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1x_2+\beta_4x_1^2+\beta_2x_2^2\right)$ to the data and has $\beta=\left(9.5 \quad 2.2 \quad 0 \quad 0.1 \quad 0.5 \quad 0\right)^T$ as estimated regression coefficients with $p-$values smaller than or equal 5%, then: $E_D \cong 1.245$, $M_c=1$ (due to $x_2^2$) and $I_c=2$ (due to $x_1x_2$ and $x_1^2$). We remark that the estimated coefficients with value zero are the ones that have $p-$values greater than 5%.

### 2.3 Model

We use the model adopted in Goel et al. (2008), which is described by (4).

$$y=10(1+x_1+x_3+x_2^2+x_4^2)+\varepsilon \quad (4)$$

The random noise $\varepsilon$ is I.I.D. Normally distributed with zero mean and unitary variance $\left(\varepsilon \sim N(0,1)\right)$.

### 2.4 Designs

We compare 5 different designs for 4 factors with 25 design points: (1) a Latin-Hypercube Design for second-order models (LHD-2); (2) a Face-centered Central Composite Design (FCCD); (3) a Maxmin Latin-Hypercube Design (LHS); (4) a computer generated D-Optimal Design; and a Box-Behnken Design (BBD).

All the designs are listed in the appendix A.

## 3. Numerical Evaluation

We generated a random noise vector $\varepsilon$ and, by using it, created the observed simulated outputs $y$ for all the designs. After this, we fitted a second order model to all designs by using a standard stepwise regression with maximum and minimum $p-$values of 5% for, respectively, a term to be added or removed from the model. The results of these stepwise regressions were stored and the whole process was repeated 10,000 times.

We used the function *stepwisefit*, from the software MATLAB R2007b, to perform these 10,000 macro-replications.

Tables 1 and 2 and Figures 1 and 2 summarizes the results of our study. Observe that the MOEs 3 and 4 are not listed in any Table or Figure. The reason is that all designs correctly identified all the true model coefficients, so $\overline{M_c}=0$ for all designs.

As the design 4 (D-optimal Design) is the one with greatest D-Efficiency value, the MOEs 2, 6, 8 and 11 were defined as the probability the other designs yield a better (smaller in MOEs 2, 6 and 8 and greater in MOE 11) value than the design 4.

Table 1. D-Efficiency and MOEs 1, 5, 7, 9 and 10.

| Design | $D_{eff}$ | $\overline{E_D}$ | $\overline{I_c}$ | $\overline{P_v}$ | $Max(P_v)$ | $\overline{R^2_{adj}}$ |
|---|---|---|---|---|---|---|
| **Des. 1** | 0.798 | 1.324 | 0.650 | 100.395 | 133.852 | 0.991 |
| **Des. 2** | 0.961 | 1.051 | 0.644 | 100.258 | 123.212 | 0.996 |
| **Des. 3** | 0.799 | 1.443 | 0.566 | 100.465 | 136.742 | 0.986 |
| **Des. 4** | 1.000 | 1.130 | 0.651 | 100.364 | 122.092 | 0.996 |
| **Des. 5** | 0.863 | 1.062 | 0.632 | 100.270 | 126.850 | 0.994 |

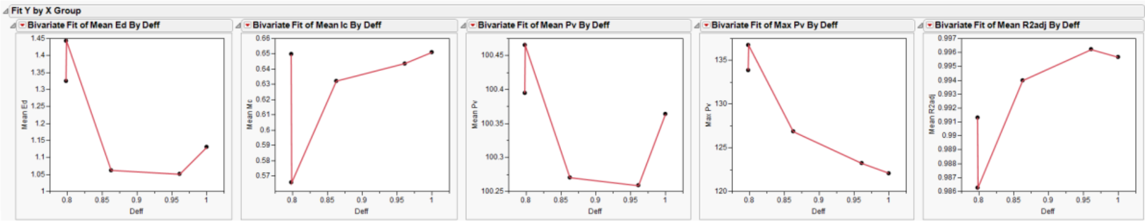


Figure 1. Bivariate plot of D-Efficiency versus $\left\{\overline{E_D}, \overline{I_c}, \overline{P_v}, Max(P_v), \overline{R^2_{adj}}\right\}$

Table 2. Pairwise probability of yielding a best value than Design 4 (MOEs 2, 6, 8 and 11).

| Design | $D_{eff}$ | $E_D$ | $I_c$ | $P_v$ | $R^2_{adj}$ |
|---|---|---|---|---|---|
| **Des. 1** | 0.798 | 39.110% | 29.720% | 49.010% | 3.820% |
| **Des. 2** | 0.961 | 54.530% | 30.260% | 49.880% | 66.270% |
| **Des. 3** | 0.799 | 35.000% | 31.030% | 48.600% | 0.330% |
| **Des. 5** | 0.863 | 53.450% | 71.120% | 49.530% | 16.640% |

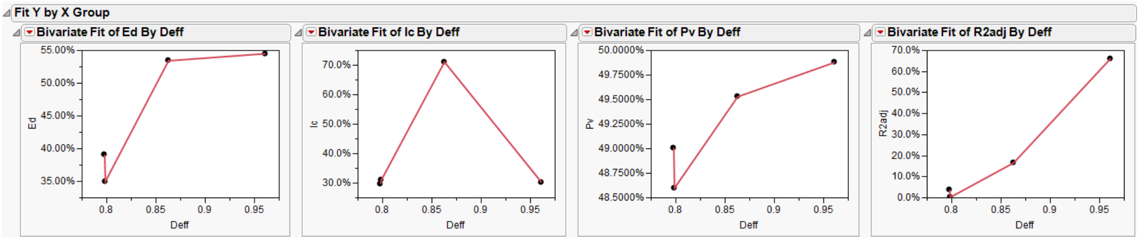


Figure 2. Bivariate plot of D-Efficiency versus $\left\{E_D, I_c, P_v, R^2_{adj}\right\}$

As it can be seen by Tables 1 and 2 and Figures 1 and 2, there is not a single MOE that is an absolutely monotonic function in the studied interval.

## 4. Conclusions

We related the D-optimality criterion to the precision of the linear regression. Put in other words, we assessed the impact of using designs with different D-optimality levels on the precision of the regression coefficients and predictions.

The second section described the D-efficiency criterion and the methodology of our study. Numerical results were presented in Section 3

It was shown that it is not possible to relate the D-efficiency criterion to any of the measures of efficiency used, i.e, there is not a single MOE that is an absolutely monotonic function in the studied interval.

Future research should focus on: (1) trying to relate other efficiency criteria (A- and G-efficiency) to the precision of the linear regression; and (2) the use of other models (third order model, for example) in the study.

## References

**Fisher, R. A.** *Statistical Methods for Research Workers*. Oliver and Boyd, 1958.
**Goel, T., R. T. Haftka, W. Shyy, e L. T. Watson**. (2008), Pitfalls of using a single criterion for selecting experimental designs. *International Journal for Numerical Methods in Engineering*, 127-155.
**Hedayat , A. S., N. J. Sloane, e J. Stufken**. *Orthogonal Arrays: Theory and Applications.* Springer-Verlag, 1999.
**Montgomery, Douglas C**. *Design and Analysis of Experiments.* John Wiley & Sons, 2005.

## Appendix A

| LHS for second-order | | | | FCCD | | | | Maxmin LHS | | | | D-optimal | | | | BBD | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.134 | -0.178 | -0.975 | 0.392 | -1 | -1 | 1 | 1 | -1.000 | -0.385 | 0.051 | -0.867 | -1 | -1 | -1 | -1 | -1 | 1 | 0 | 0 |
| 0.442 | -1.000 | 0.543 | -0.306 | 1 | 1 | -1 | 1 | 0.698 | 0.102 | 0.750 | 0.521 | -1 | -1 | -1 | 1 | -1 | 0 | 0 | 1 |
| 0.585 | 0.025 | 1.000 | -0.751 | 0 | 0 | 0 | 1 | -0.097 | -0.269 | -1.000 | -0.711 | -1 | -1 | 0 | -1 | -1 | 0 | -1 | 0 |
| -0.392 | 0.785 | 0.887 | 0.537 | 1 | 1 | -1 | -1 | -0.596 | 0.660 | 1.000 | -0.577 | -1 | -1 | 1 | 0 | 0 | -1 | 0 | -1 |
| -0.906 | 0.537 | -0.608 | -0.441 | 1 | 1 | 1 | 1 | 0.821 | -0.711 | -0.893 | 0.348 | -1 | -1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 1.000 | -0.058 | -0.327 | -0.777 | -1 | -1 | -1 | -1 | 0.337 | 0.267 | -0.669 | 0.031 | -1 | 0 | 0 | 1 | 1 | -1 | 0 | 0 |
| 0.841 | -0.823 | -0.808 | 0.219 | 0 | 0 | 1 | 0 | 0.934 | -0.088 | -0.229 | -0.768 | -1 | 0 | 1 | -1 | -1 | 0 | 1 | 0 |
| -0.862 | -0.723 | 0.734 | -0.474 | -1 | -1 | -1 | 1 | -0.147 | 0.002 | 0.932 | 0.261 | -1 | 1 | -1 | -1 | 0 | 0 | 1 | 1 |
| 0.407 | 0.599 | -0.107 | -0.057 | -1 | 1 | -1 | 1 | -0.700 | 0.418 | -0.549 | 0.709 | -1 | 1 | -1 | 1 | -1 | -1 | 0 | 0 |
| 0.675 | 0.333 | -0.576 | 1.000 | 1 | -1 | 1 | 1 | 0.439 | 0.249 | 0.582 | -0.962 | -1 | 1 | 1 | -1 | -1 | 0 | 0 | -1 |
| -0.333 | 0.755 | 0.348 | -0.918 | 1 | -1 | -1 | -1 | 0.483 | 0.067 | 0.638 | -0.190 | -1 | 1 | 1 | 1 | 0 | -1 | 1 | 0 |
| 0.315 | 0.700 | -1.000 | -0.622 | -1 | 1 | 1 | 1 | 0.232 | -0.520 | 0.531 | -0.016 | 0 | -1 | -1 | 1 | 0 | 1 | 0 | -1 |
| 0.783 | -0.483 | 0.721 | 0.742 | 1 | 1 | 1 | -1 | -0.232 | 0.882 | 0.153 | 0.888 | 0 | -1 | 1 | -1 | 1 | 0 | 0 | -1 |
| -0.548 | -0.918 | -0.108 | 0.896 | 1 | -1 | 1 | -1 | -0.739 | -0.872 | -0.141 | 0.391 | 0 | 0 | -1 | -1 | 1 | 0 | 1 | 0 |
| -0.525 | -0.308 | 0.807 | 0.443 | -1 | -1 | 1 | -1 | -0.388 | 1.000 | -0.740 | -0.506 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| -0.614 | 0.225 | 0.308 | -0.548 | -1 | 1 | -1 | -1 | 0.001 | -0.402 | -0.115 | -1.000 | 1 | -1 | -1 | -1 | 1 | 1 | 0 | 0 |
| -0.778 | -0.634 | -0.804 | -0.340 | -1 | 1 | 1 | -1 | -0.482 | -0.924 | -0.637 | -0.290 | 1 | -1 | -1 | 0 | 0 | 0 | 0 | 0 |
| -0.073 | -0.642 | -0.446 | -1.000 | 0 | -1 | 0 | 0 | 1.000 | -0.596 | 0.446 | -0.452 | 1 | -1 | 0 | 1 | 0 | -1 | 0 | 1 |
| 0.212 | -0.048 | 0.108 | 0.108 | 0 | 0 | 0 | 0 | 0.709 | 0.553 | -0.354 | 1.000 | 1 | -1 | 1 | -1 | 0 | -1 | -1 | 0 |
| -0.308 | 1.000 | -0.527 | 0.641 | 0 | 0 | -1 | 0 | 0.608 | -0.209 | 0.003 | 0.469 | 1 | -1 | 1 | 1 | 0 | 1 | -1 | 0 |
| 0.055 | -0.275 | 0.110 | -0.129 | 1 | -1 | -1 | 1 | 0.066 | 0.449 | 0.265 | -0.316 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| -1.000 | 0.362 | -0.058 | 0.601 | 0 | 0 | 0 | -1 | 0.164 | -1.000 | -0.970 | -0.101 | 1 | 1 | -1 | -1 | 0 | 0 | -1 | -1 |
| 0.141 | -0.359 | -0.242 | 0.067 | -1 | 0 | 0 | 0 | -0.865 | 0.704 | 0.344 | 0.623 | 1 | 1 | -1 | 1 | 1 | 0 | -1 | 0 |
| 0.192 | 0.251 | 0.548 | 0.795 | 1 | 0 | 0 | 0 | -0.944 | -0.746 | 0.823 | 0.821 | 1 | 1 | 1 | -1 | 0 | 0 | 1 | -1 |
| 0.883 | 0.938 | 0.363 | -0.025 | 0 | 1 | 0 | 0 | -0.338 | 0.813 | -0.443 | 0.126 | 1 | 1 | 1 | 1 | 0 | 0 | -1 | 1 |