Optimal Design Criteria Standardized by the Coefficient of Variation

J. López–Fidalgo and B. Fernández–Garzón

Departmento de Estadística. Universidad de Salamanca.
Plaza de los Caídos s/n, 37002–Salamanca (Spain).
Tel: 34 923 294458, Fax: 34 023 294514,
e-mail: fidalgo@usal.es, http://web.usal.es/~fidalgo

ABSTRACT
Many of the usual criteria for optimal experimental design do not take into account the different scale of the variance of the parameters. Dette (1997) provides a standardization based on the efficiencies for estimating each of the parameters. This leads to designs with similar efficiencies for all of the parameters.
In this paper a new way of standardization by the coefficient of variation is given. For this approach even in the linear model the information matrix will depend on the parameters. Locally and Bayesian optimal designs are computed.

Keywords and phrases: A–optimality, Bayesian Designs, Locally Optimal Designs, Nonlinear Models, Rounding off Methods, Standardized Criteria.
Classification AMS: 62K05.

1 Introduction
The response of an experiment performed under some conditions can be expressed with a statistical model,

\[ E[y] = \eta(x, \theta), \quad \text{var}[y] = \sigma^2, \quad x \in X, \]

where \( X \) is the design space with all of the possible conditions of the experiment, \( \theta \) are unknown parameters to be estimated and the variance is assumed constant.
The simplest case is the linear model,

\[ E[y] = f^T(x)\theta, \quad \text{var}[y] = \sigma^2, \quad X, \]
with \( f^T(x) = (f_1(x), \ldots, f_k(x)) \) known continuous linearly independent functions and \( \theta^T = (\theta_1, \ldots, \theta_k) \).

If the conditions of the experiment, \( x \), are under the control of the practitioner a design of the experiment can be made in advance. Thus, an experimental design will be a sequence of experimental conditions \( x_1, \ldots, x_n \), where some of them may be repeated, from a compact set \( X \).

Let \( Y^T = (y_1, \ldots, y_n) \) be the vector of observations at these points and \( X \) be the design matrix,

\[
X = (x_{ij}), \quad x_{ij} = f_j(x_i), \quad i = 1, \ldots, n; j = 1, \ldots, k.
\]

The inverse of the information matrix, \( X^T X \), is proportional to the covariance matrix of the least square estimates. Thus, an experimental design should, in some sense, be found “minimizing” the inverse of the information matrix.

Assuming that only \( m \) of the points are different, for example \( x_1, \ldots, x_m \) without loss of generality, a probability measure determines the design. If each point appears \( n_i \) times (\( \sum_{i=1}^m n_i = n \)) in the design, \( p_i = n_i/n \) will be the probability of \( x_i \), that is the proportion of experiments to be made under these condition.

Using this idea Kiefer (1971) gives a more general definition of design (approximate design) as any probability distribution or at least as a probability distribution, \( \xi \), with a finite support \( supp(\xi) \),

\[
\xi = \begin{bmatrix}
  x_1 & x_2 & \ldots & x_n \\
p_1 & p_2 & \ldots & p_n
\end{bmatrix}
\]

where \( \xi(x_i) = p_i \). This new definition has nicer properties, especially the equivalence theorem that gives a good tool for constructing optimal designs. The problem of founding an efficient exact design from an approximate design turns out very important for low sample size (Pukelsheim and Rieder, 1992; Imhof et al, 2001).

Let \( \Xi \) be the convex set of the approximate designs. The information matrix, that coincides with the Fisher information matrix, under the normality assumption, is defined as:

\[
M(\xi) = \int_X f(x) f^T(x) \xi(dx).
\]

For nonlinear models the Fisher information matrix can be used in the usual way taking into account also the dependence on the parameters,

\[
M(\xi, \theta) = E \left[ \left( \frac{\partial \eta(x, \theta)}{\partial \theta} \right) \left( \frac{\partial \eta(x, \theta)}{\partial \theta} \right)^T \right].
\]

The set of the information matrices, \( \mathcal{M} \), is also a convex and a compact set. Caratheodory’s theorem says that given an information matrix there is always a design with at most \( \frac{1}{2} k(k + 1) + 1 \) points in its support associated to it.

An optimality criterion for choosing a design is given by a function \( \Phi : \mathcal{M} \rightarrow \mathbb{R} \cup \{+\infty\} \), such that:
1. The function $\Phi$ is non-increasing in the sense that if $M_1 - M_2$ is a non-negative definite matrix then $\Phi(M_1) \leq \Phi(M_2)$.

2. $\Phi$ is a convex function.

3. $\Phi$ is a positive homogeneous function in the sense that
   \[ \Phi(\delta M) = \frac{1}{\delta} \Phi(M), \quad \delta \geq 0 \]

A $\Phi$-optimal design is a design $\xi^*$ minimizing $\Phi$. The efficiency of a design $\xi$ for with respect to the criterion $\Phi$ will be

\[ \text{eff}_\Phi(\xi) = \frac{\Phi(M(\xi^*))}{\Phi(M(\xi))}. \]

Thus, if a design has 50% efficiency then a half of the observations with the optimal design would produce the same results in the sense of the criterion $\Phi$.

The criterion used in this paper is $A$-optimality that considers the mean of the variances of the estimates,

\[ \Phi_A(M(\xi)) = \frac{1}{k} \text{tr}M^{-1}(\xi). \]

The problem with this type of criteria arises when the scales of the parameters are very different. Then the criterion pays too much attention to the parameters with high absolute variability. For solving this point Dette (1997) proposes the standardization of optimality criteria by the efficiencies of a design for estimating each of the parameters. This criterion provides designs with quite similar efficiencies for all of the parameters. However, usually the main goal in a regression model is to detect significant parameters and thus significant influence of explanatory variables in the response. Therefore, a clearly significant parameter does not need much attention in the designing of the experiment. The opposite happens with a parameter that demands some specific observations to detect whether it is significant or not.

In order to compute efficiencies of a design for each parameter the graphic method proposed by Elfving (1952) for $c$–optimality is very convenient. A $c$–optimal design minimizes the variance of the estimate of a particular linear combination of the parameters, $c^T \theta$,

\[ \Phi_c(M(\xi)) = c^T M^{-1}(\xi)c. \]

This is the theorem providing a graphic method for constructing $c$–optimal designs.

**Theorem 1.1.** An exact design $\xi^*$ of size $n$, $x_1, \ldots, x_n$, is $c$–optimal if and only if there exists a value $\gamma > 0$, such that $\gamma c$ is on the boundary of the convex hull of the set $\{f(x), x \in \mathcal{X}\} \cup \{-f(x), x \in \mathcal{X}\}$ and $\gamma c = \sum_{i=1}^{n} z_i \xi^*(x_i)f(x_i)$ for some $z_i \in \{-1, 1\}, i = 1, \ldots, n$. Moreover $c^T M^{-1}(\xi^*)c = \gamma^{-2}$.
Figure 1 shows the convex hull of the Elfving’s set for logistic regression in a big compact interval. Wald (1943) proved that a c–optimal design maximizes the power function for $c^T \theta$ in a linear model. Kiefer and Wolfowitz (1960) proved a particular equivalence theorem, generalized by Whittle in 1973.

**Theorem 1.2.** The following conditions are equivalent:

1. $\xi^*$ is $\Phi$–optimal.
2. $\xi^*$ maximizes $\inf_\eta \partial \Phi(M(\xi),M(\eta))$.
3. $\inf_\eta \partial \Phi(M(\xi^*),M(\eta)) = 0$.

where $\partial \Phi(M,N)$ is the directional derivative of Fréchet of the function $\Phi$ at $M$ in the direction of $N$,

$$\partial \Phi(M,N) = \lim_{\alpha \to 0^+} \frac{1}{\alpha} \{ \Phi((1-\alpha)M + \alpha N) - \Phi(M) \}, \quad M,N \in \mathcal{M}$$

If $\Phi$ is differentiable the so called sensitive function is defined as:

$$\psi(x,\xi) = f^T(x) \nabla \Phi(M(\xi)) f(x) - \text{tr}M(\xi) \nabla \Phi(M(\xi)),$$

and $\psi(x,\xi)$ vanishes at each point of $\text{supp}(\xi)$. A design $\xi^*$ is $\Phi$–optimal if and only if $\psi(x,\xi^*) \geq 0$, $x \in \mathcal{X}$. Thus, the points of the support of the $\Phi$–optimal design that are in the interior of the set $\mathcal{X}$ satisfy:

$$\frac{\partial \psi(x,\xi^*)}{\partial x} = 0.$$
For A–optimality $\psi(x, \xi) = \text{tr}M^{-1}(\xi^*) - f^T(x_i)M^{-2}(\xi^*)f(x_i)$.

There are two approaches for computing optimal designs, when the information matrix depends on the unknown parameters to be estimated. A locally $\Phi$-optimal design will be computed for nominal values of the parameters. That could be considered as an initial estimation. Sometimes an explicit design depending on general initial values of the parameters can be computed, but usually numerical methods have to be used for computing optimal designs for specific nominal values.

The other approach is the Bayesian theory for optimal designs. Chaloner and Verdinelli (1995) provides a nice review. By providing a prior distribution for the parameters and a utility function related to the criterion function a Bayesian design can be computed. This approach is also used for linear models. Let us assume a normal distribution on the observations

$$Y/\sigma^2 \sim N(X\theta, \sigma^2 I),$$

with known $\sigma^2$ and a normal multivariate prior distribution,

$$\theta/\sigma^2 \sim N(\theta_0, \sigma^2 R^{-1}),$$

with $R$ a known $k \times k$ matrix. Then, the posterior distribution will be a normal distribution with covariance matrix $D(X) = \sigma^2(X^TX + R)^{-1}$. From this matrix the usual criteria can be considered. But in this case the information matrix depends on the sample size, $X^TX + R = n(M + n^{-1}R)$. As usual in Bayesian theory a big sample size or a small prior information leads to classic designs. The choice of the loss function is “equivalent” to the criterion function in the classic theory. Thus, the quadratic loss function,

$$L(\xi) = \int (\theta - \hat{\theta})^T K(\theta - \hat{\theta})p(y, \theta/\xi) dy d\theta$$

leads to

$$\Phi(\xi) = \text{tr}\{KD(X)\}$$

that is Bayesian A–optimality for $K = I$ and Bayesian $c$–optimality for $K = cc^T$.

In the case of nonlinear models an approximation of the posterior distribution is considered,

$$\theta/\sigma^2, y \sim N(\hat{\theta}, \sigma^2 M(\hat{\theta}, \xi)^{-1}).$$

Then, the expected loss leads to the Bayesian A–optimality as follows:

$$\Phi(\xi) = E_\theta(\text{tr}KM^{-1}(\xi, \theta)).$$

A similar equivalence theorem can be used in this context.
2 Standardized Optimality Criteria

Dette (1997) suggests re-scaling the variance of the estimates of the parameters using the efficiencies of the design for estimating each parameter. For instance, A–optimality becomes SA–optimality,

$$\Phi_{SA}(M(\xi)) = \sum_{j=1}^{k} \frac{\{M^{-1}(\xi)\}_{jj}}{\{M^{-1}(\xi_j^*)\}_{jj}}$$

where $\xi_j^*$ is the best design to estimate $\theta_j$ and $M^−$ denotes a generalized inverse matrix of $M$.

Cook and Fedorov (1995) suggest another alternative restricting the search to optimal designs with a lower bound of efficiency for individual parameters. The A–optimal design for the linear regression model in a general interval $[a,b]$ (see e.g. Torsney and López–Fidalgo, 1995) is a 2–point design at $a$ and $b$ with weight at $a$,

$$p_A = \begin{cases} 
1/2 & \text{if } |a| = |b| \\
\frac{b^2+1-\sqrt{(b^2+1)(a^2+1)}}{b^2-a^2} & \text{if } |a| \neq |b|.
\end{cases}$$

Dette (1997) gives a 2–point SA–optimal design for the linear regression model in a general interval $[a,b]$ where the weight at $a$ is now,

$$p_{SA} = \begin{cases} 
1/2 & \text{if } |a| = |b| \\
\frac{4|a|^2+(|b|+|a|)^2-\sqrt{(4|a|^2+(|b|+|a|)^2)(4|a|^2+(|b|+|a|)^2)}}{4(b^2-a^2)} & \text{if } |a| \neq |b|.
\end{cases}$$

Some results for different values of $a$ and $b$ are shown in Table 1.

The quadratic model in [-1,1] has a common A– and SA–optimal design,

$$\xi = \begin{bmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}.$$ 

3 Standardized criteria through the coefficient of variation

The main contribution of this paper is to propose a new way of standardizing optimality criteria by using the coefficient of variation. The advantage of this standardization is that the criterion based on the coefficient of variation pays more attention to those parameters which need it more in terms of significance than the standardized criteria by the efficiency does. On the other hand the way of restricting the search to optimal designs with a lower bound of efficiency for individual parameters (Cook and Fedorov, 1995) means the always difficult task of choosing a specific lower bound.
Table 1: A– and SA-optimal designs for the linear regression model in an interval [a, 1] for several values of \(a \in [-5, 0]\). Second and third column are the weights of the A– and SA–optimal designs at point \(a\), respectively. From the fourth to the seventh column the efficiencies of the A– and SA–optimal designs for estimating the parameters \(\theta_0\) and \(\theta_1\) are shown.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(p_A)</th>
<th>(p_{SA})</th>
<th>(\text{eff}(\xi_A, \theta_0))</th>
<th>(\text{eff}(\xi_{SA}, \theta_0))</th>
<th>(\text{eff}(\xi_A, \theta_1))</th>
<th>(\text{eff}(\xi_{SA}, \theta_1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>0.217</td>
<td>0.352</td>
<td>0.985</td>
<td>0.870</td>
<td>0.680</td>
<td>0.912</td>
</tr>
<tr>
<td>-4</td>
<td>0.255</td>
<td>0.363</td>
<td>0.984</td>
<td>0.897</td>
<td>0.761</td>
<td>0.925</td>
</tr>
<tr>
<td>-3</td>
<td>0.309</td>
<td>0.383</td>
<td>0.984</td>
<td>0.931</td>
<td>0.854</td>
<td>0.945</td>
</tr>
<tr>
<td>-2</td>
<td>0.387</td>
<td>0.419</td>
<td>0.988</td>
<td>0.971</td>
<td>0.949</td>
<td>0.974</td>
</tr>
<tr>
<td>-1</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0.586</td>
<td>0.691</td>
<td>0.586</td>
<td>0.691</td>
<td>0.971</td>
<td>0.854</td>
</tr>
<tr>
<td>-1/2</td>
<td>0.559</td>
<td>0.581</td>
<td>0.955</td>
<td>0.971</td>
<td>0.986</td>
<td>0.974</td>
</tr>
<tr>
<td>-1/3</td>
<td>0.573</td>
<td>0.617</td>
<td>0.886</td>
<td>0.931</td>
<td>0.979</td>
<td>0.945</td>
</tr>
<tr>
<td>-1/4</td>
<td>0.579</td>
<td>0.637</td>
<td>0.832</td>
<td>0.897</td>
<td>0.975</td>
<td>0.925</td>
</tr>
<tr>
<td>-1/5</td>
<td>0.581</td>
<td>0.648</td>
<td>0.793</td>
<td>0.870</td>
<td>0.974</td>
<td>0.912</td>
</tr>
</tbody>
</table>

Thus, a A-optimal standardized by the coefficient of variation (CA-optimal design) minimizes the function

\[
\Phi_{CA}(M(\xi)) = \sum_{j=1}^{k} \frac{\{M^{-1}(\xi)\}_{jj}}{\theta_j^2}.
\]

A general definition for any criterion can be given at this point. Let \(\Psi\) be a function non-negative, positively homogeneous, convex, lower semi-continuous in the set

\[
\mathbb{R}_{\geq}^m = \{a = (a_1, \ldots, a_m) \in \mathbb{R}^m : a_j \geq 0, j = 1, \ldots, m\}
\]

and positive in

\[
\mathbb{R}_{>}^m = \{a \in \mathbb{R}_{>}^m : a_j > 0, j = 1, \ldots, m\} \quad (m \leq k).
\]

Let \(e_j\) the unit vector in \(\mathbb{R}^k\).

**Definition 3.1.** A design \(\xi\) is \(C\Psi\)-optimal (\(\Psi\)-optimal standardized by the coefficient of variation) for the parameters \(\theta_1, \ldots, \theta_m\) in the model \(y = f^T(x)\theta + \epsilon\), if \(e_j \in \text{Im}(M(\xi))\), for \(j = 1, \ldots, m \) and \(\xi\) minimizes the function

\[
\Psi(M(\xi)) = \Psi \left[ \frac{e_1^TM^{-1}(\xi)e_1}{\theta_1^2}, \ldots, \frac{e_m^TM^{-1}(\xi)e_m}{\theta_m^2} \right].
\]

CA-optimality is a particular case of Definition 3.1, using the function \(\sum_{i=1}^{k} a_i\). These criteria are equivalent to the ordinary criteria over the matrices \(B^{-2}M^{-1}(\xi)\) with \(B = \text{diag}(\theta_1, \ldots, \theta_m, 0, \ldots, 0)\), instead of the usual \(M^{-1}(\xi)\). Thus the matrix depends on the parameters and the treatment is similar to the nonlinear models.
### 3.1 Locally CA-optimality

The criterion of CA–optimality looks for the minimization of

$$\sum_{j=1}^{k} \frac{\text{var}(\hat{\theta}_j)}{\theta_j^2} \propto \text{tr}(B^{-2}M^{-1}(\xi)),$$

with \( B = \text{diag}(\theta_1, \ldots, \theta_k) \). Assume there is a nominal value \( \theta^{(0)} \) for the parameters \( \theta \).

If \( M(\xi) \) is nonsingular then the function \( \Phi_{CA}(M(\xi)) = \text{tr}(B^{-2}M^{-1}(\xi)) \) is differentiable and the gradient of \( \Phi_{CA} \) is

$$\nabla \Phi_{CA}(M(\xi)) = -M^{-1}(\xi)B^{-2}M^{-1}(\xi).$$

The equivalence theorem says that a necessary and sufficient condition for a design \( \xi^* \), with nonsingular information matrix \( M(\xi^*) \), to be CA–optimal is that

$$f^T(x)M^{-1}(\xi^*)B^{-2}M^{-1}(\xi^*)f(x) \leq \text{tr}(B^{-2}M^{-1}(\xi^*)), \quad x \in X.$$

Then, this is the case of a linear criterion with and unknown matrix, \( L = B^{-2} \). From the results of Pukelsheim and Torsney (1991) a formula can be provided to compute CA-optimal designs given a design support such that the corresponding regression vectors for the support points are linearly independent. Thus, the CA–optimal design for the points \( x_i, i = 1, \ldots, k \), has the following weights:

$$\xi(x_i) = \frac{\sqrt{c_{ii}}}{\sum_{j=1}^{k} \sqrt{c_{jj}}},$$

where \( c_{jj}, j = 1, \ldots, k \), are the elements of the diagonal of the matrix \( C = UU^T \) with \( U = (XX^T)^{-1}XB^{-1} \).

If all of the parameters are of the same absolute variability then this criteria are equivalent to the classic criteria. Otherwise they will provide designs that discover significant parameters near zero not detected by the usual criteria or even the standardized by the efficiencies. This situation is being considered in the next sections for the simple linear regression model.

### 3.2 Locally CA-optimal designs for simple linear regression

Let the model be

$$E[y] = \theta_0 + \theta_1 x, \quad x \in [a, b],$$

and let \( \theta_0^{(0)} \) and \( \theta_1^{(0)} \) be the nominal values of the parameters. Then the locally CA-optimal approximate design is

$$\xi_{CA} = \left\{ \frac{a}{p_{CA}}, \frac{b}{1 - p_{CA}} \right\}, \quad \text{where } p_{CA} = \left\{ \begin{array}{ll} 1/2 & \text{si } |a| = |b| \\ \frac{\sqrt{\theta_0^{(0)}b^2 + \theta_1^{(0)}a^2}}{\sqrt{\theta_0^{(0)}b^2 + \theta_1^{(0)}a^2 + \theta_0^{(0)}b^2}} & \text{si } |a| \neq |b|. \end{array} \right.$$
In the case of $|a| = |b|$, it is the A-optimal and the efficiency for both parameters is 1 (Table 1). If $|a| \neq |b|$ the efficiencies of this design for each parameter are:

$$\text{eff}(\xi_{CA}, \theta_0) = \frac{(a - b)^2}{b^2(1 + \frac{\sqrt{a^2+r^2}}{\sqrt{b^2+r^2}}) + a^2(1 + \frac{\sqrt{b^2+r^2}}{\sqrt{a^2+r^2}})}$$

$$\text{eff}(\xi_{CA}, \theta_1) = 4\sqrt{a^2 + r^2}\sqrt{b^2 + r^2}$$

where $r = \frac{\theta_0^{(0)}}{\theta_1^{(0)}}$. Also the efficiencies depend on the nominal values of the parameters. In this particular case the dependence is through the ratio $r$.

For specific values of $a$ and $b$, the efficiency for estimating $\theta_0$ will increase as $|r|$ decreases. That is, for small values of $|\theta_0^{(0)}|$ relative to $|\theta_1^{(0)}|$, the CA–optimal design will be more efficient for estimating $\theta_0$. On the other hand, if $|r|$ increases the efficiency for estimating $\theta_1$ increases too. This is true since the derivative,

$$\frac{\partial \text{eff}(\xi_{CA}, \theta_0)}{\partial r} = -\frac{(a - b)^2(a^2 - b^2)^2r^3}{uv(b^2(r^2 + uv) + a^2(2b^2 + r^2 + uv))^2},$$

where $u = \sqrt{a^2 + r^2}$ and $v = \sqrt{b^2 + r^2}$, is negative for $r > 0$ (the case $r < 0$ is symmetric). In a similar way $\text{eff}(\xi_{CA}, \theta_1)$ is non increasing for $|r|$. Figure 2 shows the efficiencies for the interval $[-5, 1]$.

![Figure 2: Efficiencies for $\theta_1$ (continuous line) and $\theta_0$ (dash line) of the CA–optimal design in $[-5, 1]$ for different values of $r$.](image)

If $r$ approaches 0 the CA–optimal design approaches the optimal design for estimating $\theta_0$,

$$\xi_{\theta_0} = \begin{cases} a & \frac{|b|}{|b| + |a|}, \\ b & \frac{|a|}{|b| + |a|} \end{cases}.$$
Otherwise for large values of $r$ the CA–optimal designs approaches the optimal design for estimating $\theta_1$,

$$\xi_{\theta_1} = \left\{ \begin{array}{c} a \\ 1/2 \\ b \\ 1/2 \end{array} \right\}. $$

### 3.3 Statistical interpretation of the criteria standardized by the coefficient of variation

The linear regression model is considered here with $x \in [-5, 1]$. Let assume $\theta_0 = 1$ and $\theta_1 = 0.05$ as the real values. There is a large difference of magnitude between them. These values will be assumed as nominal values as well, but it will be seen that with other close up values the results remain valid. It will be seen that the CA–optimal design produces more efficient estimates for $\theta_1$ than A– and SA–optimal designs.

A–, SA– and CA–optimal designs for this model are:

$$\xi_A = \left\{ \begin{array}{c} -5 \\ 0.217 \\ 0.783 \end{array} \right\}, \quad \xi_{SA} = \left\{ \begin{array}{c} -5 \\ 0.352 \\ 0.648 \end{array} \right\}, \quad \xi_{CA} = \left\{ \begin{array}{c} -5 \\ 0.493 \\ 0.507 \end{array} \right\}. $$

The efficiencies of these designs with respect to each parameter are shown in Table 2.

<table>
<thead>
<tr>
<th>eff($\xi_x, \theta_i$)</th>
<th>$\theta_0$</th>
<th>$\theta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_A$</td>
<td>0.985</td>
<td>0.680</td>
</tr>
<tr>
<td>$\xi_{SA}$</td>
<td>0.870</td>
<td>0.912</td>
</tr>
<tr>
<td>$\xi_{CA}$</td>
<td>0.702</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 2: Efficiencies of A–, SA–, CA–optimal designs for estimating $\theta_0$ and $\theta_1$

The A–optimal design pays less attention to the efficiency of the estimate of $\theta_1$. The SA–optimal design equalizes both efficiencies, but it could be said that the parameter $\theta_1$ needs more attention in order to detect its significance. This is what the CA–optimal design provides, keeping both efficiencies still over 70%. Thus both hypothesis tests will be more easily significant with CA–optimality. A simulation study is carried out in the next section to show this idea.

### 3.4 Simulation

The simulation is carried out with 100 observations. For rounding off an approximate design to an exact design of size $n$, the procedure described by Pukelsheim and Rieder (1992) will be used. The rounded off exact designs with 100 observations of the A–, SA– and CA–optimal approximate designs take 22, 35 and 49 of the observations at $-5$ and the rest at 1 respectively.
Three simulations for the linear regression model \( y = 1 + 0.05x + \epsilon \) according to the designs given before were made assuming a normal distribution of mean 0 and variance 1 for the experimental error \( \epsilon \). Table 3 shows the estimates of the parameters as well as the p–values for both parameters and the 3 designs.

<table>
<thead>
<tr>
<th>Design</th>
<th>estimate ( \theta_1 )</th>
<th>p-value</th>
<th>estimate ( \theta_0 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A–optimal</td>
<td>0.061</td>
<td>0.104</td>
<td>1.191</td>
<td>0.000</td>
</tr>
<tr>
<td>SA–optimal</td>
<td>0.069</td>
<td>0.060</td>
<td>0.887</td>
<td>0.000</td>
</tr>
<tr>
<td>CA–optimal</td>
<td>0.063</td>
<td>0.025</td>
<td>1.092</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 3: Estimates and p-values for the parameters and the corresponding tests \( H_0 : \theta_1 = 0 \) and \( H_0 : \theta_0 = 0 \) for a sample size of \( n = 100 \)

The three designs can detect easily the significance of \( \theta_0 \), but \( \theta_1 \) is only significant with the CA–optimal design. Of course, this is just an example to show how this criterion works.

### 4 Relative efficiency of the CA–optimal design

In this section the sensitivity of CA–optimal designs to wrong choices of the nominal values of the parameters is considered. Assume \( \theta_0 = 1 \) and \( \theta_1 = 0.05 \) are again the real values of the parameters and set nominal values near them. Two kind of relative efficiencies are considered here. On one hand, efficiencies for estimating the real values of the parameters and then efficiencies of the wrong CA–optimal designs with respect to the true CA–optimal will be computed.

Tables 4 and 5 show the efficiencies for estimating \( \theta_1 \) and \( \theta_0 \) respectively for different nominal values around the real values.

<table>
<thead>
<tr>
<th>( \theta_0^{(0)} )</th>
<th>( \theta_1^{(0)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.025</td>
<td>0.5</td>
</tr>
<tr>
<td>0.05</td>
<td>1.0</td>
</tr>
<tr>
<td>0.075</td>
<td>1.5</td>
</tr>
<tr>
<td>0.1</td>
<td>1.000</td>
</tr>
<tr>
<td>0.15</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4: Efficiencies of the CA–optimal design, \( \xi_{CA} \), for estimating \( \theta_1 \) for different values of \( \theta_0^{(0)} \) and \( \theta_1^{(0)} \)

Even considering these wrong designs the efficiencies for estimating \( \theta_1 \) are higher than the efficiencies with respect to the true SA– and A–optimal designs. Thus, the CA–optimal will produce more significant tests for \( \theta_1 \) than the SA– and A–optimal designs do.
Table 5: Efficiencies of the CA–optimal design, $\xi_{CA}$, for estimating $\theta_0$ for different values of $\theta_0^{(0)}$ and $\theta_1^{(0)}$. 

<table>
<thead>
<tr>
<th>$\theta_0^{(0)}$</th>
<th>0.025</th>
<th>0.05</th>
<th>0.075</th>
<th>0.1</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.702</td>
<td>0.726</td>
<td>0.759</td>
<td>0.793</td>
<td>0.853</td>
</tr>
<tr>
<td>1.0</td>
<td>0.695</td>
<td>0.702</td>
<td>0.712</td>
<td>0.726</td>
<td>0.759</td>
</tr>
<tr>
<td>1.5</td>
<td>0.693</td>
<td>0.697</td>
<td>0.702</td>
<td>0.708</td>
<td>0.726</td>
</tr>
</tbody>
</table>

4.1 Robustness of CA–optimal designs

On the other hand Table 6 shows efficiencies of the wrong CA–optimal designs with respect to the true CA–optimal for different values of $\theta_0^{(0)}$ and $\theta_1^{(0)}$. 

<table>
<thead>
<tr>
<th>$\theta_0^{(0)}$</th>
<th>0.025</th>
<th>0.05</th>
<th>0.075</th>
<th>0.1</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.000</td>
<td>0.998</td>
<td>0.991</td>
<td>0.978</td>
<td>0.901</td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.992</td>
</tr>
<tr>
<td>1.5</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
</tr>
</tbody>
</table>

Table 6: Efficiencies of the wrong CA–optimal designs for the true CA–optimal for different values of $\theta_0^{(0)}$ and $\theta_1^{(0)}$ when the real values of the parameters are $\theta_0 = 1$ and $\theta_1 = 0.05$.

Since the CA-optimal in a simple linear regression model depends on the value of the parameters through $r = \theta_0^{(0)}/\theta_1^{(0)}$, the efficiency of these designs can be expressed as a function of $r$. Let $\xi_{CA}^r$ be the CA–optimal for $\theta_0 = 1$ and $\theta_1 = 0.05$ and let $\xi_{CA}(r)$ be the CA-optimal design for a specific value of $r$, then

$$
\text{eff}(\xi_{CA}(r), \xi_{CA}^r) = \frac{45.879\sqrt{1 + r^2}\sqrt{25 + r^2}}{290.278 + 22.944(r^2 + \sqrt{1 + r^2}\sqrt{25 + r^2})}
$$

that is a symmetric function, that is shown in Figure 3. Between 10 and 30 the efficiency is over 0.9985. For the case $r = 1$, that is, when the CA–optimal design is A–optimal, the efficiency is 0.691. The limit of the efficiency is 0.9998 as $r$ goes to $\infty$ (optimal design for estimating $\theta_1$) and 0.6912 as $r$ goes to 0 (optimal design for estimating $\theta_0$).

But this is not the best graphic since the interval $[0, 20]$ represents the same, but inverse, ratios than the interval $[20, \infty]$. Thus it would be a better way to draw a graphic doing a one to one mapping of the intervals $[0, 20]$ and $[20, \infty]$ into $[-1, 0]$ and $[0, 1]$, respectively. Now Figure 4 is much easier to interpret taking into account that 1 means a perfect matching with the real parameters.

Figure 5 shows similar curves for different real values of $r$ shown within each graphic.
5 CA–optimal Bayesian design for simple linear regression

Assuming a prior distribution on the parameters $\theta$ the criterion function can be expressed as

$$\Phi(\xi) = E_\theta(\text{tr}(B^{-2}M^{-1}(\xi))) = \text{tr}(E_\theta(B^{-2})M^{-1}(\xi)),$$

where $E_\theta(\cdot)$ is the expectation over the prior distribution of $\theta$. Thus, the problem is reduced to compute $E_\theta(1/\theta_i^2)$, for each $i$. The final result is equivalent to take $\theta_i^{(0)} = 1/\sqrt{E_\theta(1/\theta_i^2)}$ as nominal values in the locally optimal designs obtained above. For simple linear regression let $\theta_0$ and $\theta_1$ follow independent uniform distributions in some pair of intervals. Thus, if $\theta_i$ is uniform on $(a,b)$, then

$$E_\theta(1/\theta_i^2) = \int_a^b \frac{1}{(b-a)\theta_i^2} d\theta_i = \frac{1}{ab}.$$

Three different intervals are considered for one parameter and two for the other. The intervals for $\theta_0$ have 5 as middle point with different lengths in order to consider
different degrees of information about the parameters. For $\theta_1$, intervals with different middle points, 2 and 7, with the same length are considered in order to observe the effect of the magnitude of the parameter on the CA–optimal design. Going on with the example for $x \in [-5, 1]$,

$$\xi_{BCA} = \left\{ \begin{array}{ll}
-5 & 
\text{if } 2 \\
p_{BCA} & 
\text{if } 7 \\
1 - p_{BCA} & \end{array} \right\}$$

Table 7 shows the weight $p_{BCA}$ of the CA–optimal Bayesian design for different distributions. The efficiency for estimating one parameter increases for larger intervals, meaning less prior information, and when the middle point is smaller. Therefore, the CA–optimal Bayesian design provides more efficient estimations and more significant tests for small parameters and for parameters with less prior information.
Table 7: For each joined distribution it is shown the value of $q$ for the CA–optimal design at point $-5$ and the efficiencies for estimating each parameter

<table>
<thead>
<tr>
<th>distribution of $\theta_0$</th>
<th>distribution of $\theta_1$</th>
<th>$p_{BCA}$</th>
<th>$\text{eff}_{\theta_0}(\xi)$</th>
<th>$\text{eff}_{\theta_1}(\xi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{U}(4.5, 5.5)$</td>
<td>$\mathcal{U}(1, 3)$</td>
<td>0.345</td>
<td>0.876</td>
<td>0.904</td>
</tr>
<tr>
<td>$\mathcal{U}(4, 6)$</td>
<td>$\mathcal{U}(1, 3)$</td>
<td>0.343</td>
<td>0.870</td>
<td>0.901</td>
</tr>
<tr>
<td>$\mathcal{U}(1, 9)$</td>
<td>$\mathcal{U}(1, 3)$</td>
<td>0.274</td>
<td>0.945</td>
<td>0.796</td>
</tr>
<tr>
<td>$\mathcal{U}(4.5, 5.5)$</td>
<td>$\mathcal{U}(6, 8)$</td>
<td>0.196</td>
<td>0.995</td>
<td>0.630</td>
</tr>
<tr>
<td>$\mathcal{U}(4, 6)$</td>
<td>$\mathcal{U}(6, 8)$</td>
<td>0.195</td>
<td>0.995</td>
<td>0.628</td>
</tr>
<tr>
<td>$\mathcal{U}(1, 9)$</td>
<td>$\mathcal{U}(6, 8)$</td>
<td>0.178</td>
<td>0.999</td>
<td>0.586</td>
</tr>
</tbody>
</table>

6 Discussion

After the results given above it can be said that $\Phi$–optimality gives designs paying more attention to the parameters that “need” it the most. Here the coefficient of variation is used for the standardization of the usual criteria. The computation of this coefficient is not possible when the parameter is zero. But still the criterion can be used with small, but non zero, nominal values and will provide good designs for the estimation of this parameter. Nevertheless the power of the corresponding test can decrease considerably. Thus, another kind of control can be done using constraint criteria jointly with these using the results of Cook and Wong (1994). There are now some open lines of research of this standardization for other models and criteria, such as CMV–optimality. Thus, CMV–optimal designs can be compared with the MV- and SMV-optimal designs given by Dette (1997) for the quadratic model or for binary response models given by Dette and Sahm (1997) and López–Fidalgo and Tommasi (2003). For the quadratic model in $[-1, 1]$ the CA–optimal design is,

$$\xi = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ p & 1 - 2p & p \end{array} \right\}, \quad \text{with} \quad p = \frac{\sqrt{\frac{1}{\sigma_0^{11}} + \frac{1}{\sigma_2^{11}}} - \sqrt{\frac{1}{\sigma_0^{11}} + \frac{1}{\sigma_2^{11}}}}{2 \left( \sqrt{\frac{1}{\sigma_0^{11}} + \frac{1}{\sigma_2^{11}}} + \sqrt{\frac{1}{\sigma_0^{11}} + \frac{1}{\sigma_2^{11}}} \right)}.$$

It remains also the computation of Bayesian designs for other distributions. The choice of the uniform distribution in this paper made easy to compare the results.

References


