Optimum Experimental Design package

and web application

Juan M. Rodríguez Díaz y Guillermo Sánchez León

A package gathering some functions about Optimal Experimental Design subject has been developed, including the last contributions of the authors in some of the top journals of the statistical area. It provides tools and functions that compute optimal (or nearly optimal) designs for different situations and optimality criteria. The goodness of the software has been checked by means of the role it has played in the published papers, and it has also been proved to be very useful for educational purposes, for instance when teaching the subject in a doctorate course. This fact has especially been taken into account when creating the corresponding fully - detailed help structure. The work has been completed with the adaptation of the package, using webMathematica, to be available through Internet, with an on-line help. It can be run directly in the web site
http : // www3.enusa.es/webMathematica/OED/OED - index.html

■ Introduction

Very often when performing scientific experiments, the model used depends on unknown parameters. In order to estimate these parameters, some samples are taken, selecting values of the independent variables (a design) and observing the corresponding values of the dependent variable. With the information given by the set of samples (that is, the information given by the design) the parameters can be estimated, always with an imprecision that can be measured by the variance of the estimators.

Of course, what is always desirable is an accurate estimation of the parameters of the model, that is, a estimation with small variance. In this sense, a design is optimal when minimizes the variance of the estimators of the parameters.

During the last century, Optimal Experimental Design theory has developed very fast, especially in the last 50 years. Some computer routines have been created, in most of cases trying to solve specific problems related to a particular research. But to date there is not a general package of functions that can be used for a great variety of subjects and situations. This work tries to be the origin of such a package, gathering a bunch of general-purpose functions and some others related mainly with the research of the authors. The Optimal Design package can help researchers, professors and students dealing with this subject. Never complete, new functions can (and should) be added to the initial ones in due course.

■ Optimal Design of Experiments

A design is a set of points where to take observations (exact design), but it can also be seen as a probability measure with finite support (approximate design). The main objective of the theory is to search for designs providing the 'best' estimation of the parameters of a regression model, where 'best' stands for the estimation with less variance.

When working with non - linear models the best design will depend on the values of the unknown parameters. To solve this problem some kind of additional information is needed, either an initial value for the parameters or a priori distribution for them. In any case, the optimal design will be a function of these initial values or distributions (locally or Bayesian optimal designs).

It is easy to decide which is the best design when there is only one unknown parameter. But in the general case, when dealing with several parameters, 'variance' means the covariance matrix of the estimators of the parameters. However, there is not a unique way to minimize a matrix, and we could prefer to pay attention to the determinant, the trace, the eigenvalues or some other properties. Depending on the choice, we talk of different optimality criteria. The most widely used is D-optimality, that centers on the determinant of the covariance matrix. Another important one is c-optimality, that tries to minimize the variance of the estimator of a linear combination of the parameters given by the vector c. A-optimality tries to minimize the trace of the covariance matrix, that is, the average of the variances of the parameters, some others pay attention to the eigenvalues, etc.

The main tool in Optimal Experimental Design is the information matrix, that has been proved to be (asymptotically) proportional to the inverse of the covariance matrix. For this reason, many optimality criteria try to minimize the inverse of the information matrix.

The efficiency of a design for a particular criterion function measures how good the design is, compared to the optimal design for that criterion. The efficiency expresses the proportion of observations that the optimal design needs to get a estimation of the parameters as good as the one provided by the given design. The equivalence theorems are very important tools to check whether a design is optimal respect to a specific optimality criteria. An interesting result derived from them is to see how close a design is to the optimal, obtaining a lower bound of the efficiency.

Optimal Designs are not easy to find analytically. Very often, an algorithmic approach is the best (and maybe the only) way to work. The most popular algorithm adds a point (with a respective weight) in each step of the procedure. The condition derived from the equivalence theorems provide a stopping rule for the algorithmic process: when the design obtained has an efficiency greater than a value initially fixed.

Some general manuals on Optimal Experimental Design are for instance the books of Pázman (1986), Atkinson and Donev (1992) or Pukelsheim (1993, 2006).

■ The OED Package

The Package is a set of functions that can be grouped together depending on the work they perform and the subject they are specially designed for. The main groups are:

- Basic (essential) functions, containing general-purpose functions
- Minimize functions, for one and several variables, with or without restrictions
- Algorithms for finding (quasi)-optimal designs
- c-optimality: computes c-optimal designs based on the generalization of Elfving's method (Elfving, 1952)
- Spatial Designs: spatial criterion and algorithms for computing optimal designs.
- Generalized Exponential, computing D- and c-optimal designs for the Generalized Exponential Model.

and the scheme of the package is the following :

- Basics
  - Information Matrix: **inforMat**
  - Fisher Matrix: **fisherMat**
  - Weight Average: **weightAv**
- Minimize functions
  - One Variable : **minFunction**
  - Multivariate with restrictions

    Computing criterion first : **minFunctions1**

    Checking restrictions first: **minFunctions2**
- Algorithms
  - Algorithm for Characteristic criteria: **charactAlgorithm**
  - General Fedorov - Wynn algorithm: **fedorovWynnAlgorithm**

- c - optimality
    - Elfving' s method: **cOptimElfving**
- Spatial Designs
    - Normal - approximation criterion: **moransINormalApprox**
    - Algorithms

        Full Enumeration**: fullEnumeration**

        Simple Search: **simpleSearch**

        Fedorov Exchange: **fedExch**
- Generalized Exponential
    - D - optimal designs: **gEDopt**
    - c - optimal designs: **gEcopt**

plus some auxiliary functions hidden in the Private context.

A web application has been created, using *webMathematica* in order to allow the executing of the functions through Internet. It follows this same scheme and, as the local package, it contains a complete help structure containing many examples showing the use of the functions for different situations, including a concise and convenient bibliography for each function, as well as links to other related functions.

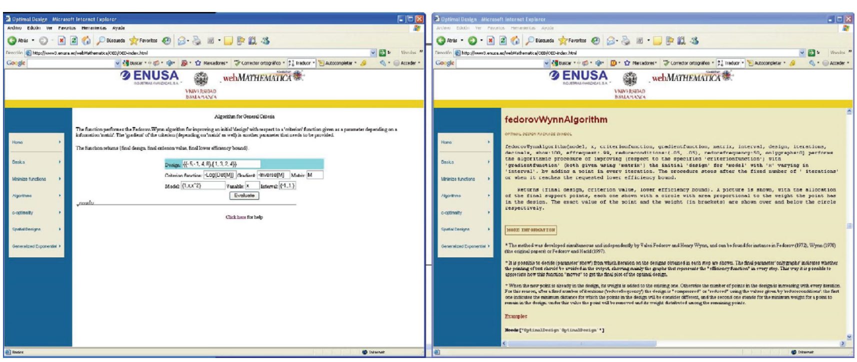


Figure 1. Screen shots of function and help structure of the web application

The application is hosted in a server of ENUSA Industrias Avanzadas S.A.,
http : // www3.enusa.es/webMathematica/OED/OED - index.html
and supported by one of the authors, Sánchez León, the creator of the well known package Biokmod (2007), that is hosted in the same server.

■ Examples of use

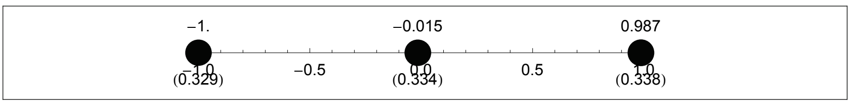
Functions that compute minimum points :

```
minFunction[Sin[x] Cos[2 x], x, {-Pi, Pi}, 3]
```

```
{-1., 1.571}
```

The main algorithms are due to Fedorov (1972) and Wynn (1972).An adaptation by Rodríguez-Diaz and López-Fidalgo (2003) is implemented in the package:

```
charactAlgorithm[{1, x, x^2}, x, 3,
 {-1, 1}, {{-.5, -.1, .4, .8}, {.1, .3, .2, .4}},
 100, 3, 101, .99, {.1, .05}, 24]
```

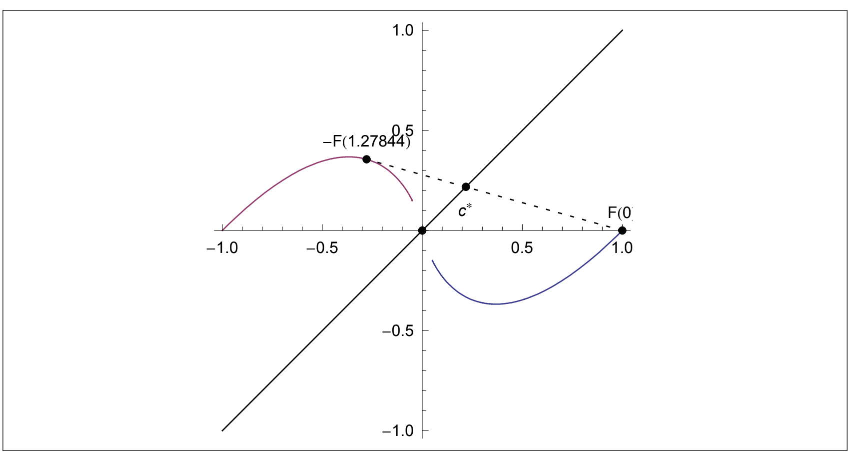



```
{{{-1., -0.015, 0.987}, {0.329, 0.334, 0.338}},
 7.00063, 0.922135}
```

Figure 2. Algorithmic computation of D-optimal designs based on

The computation of c-optimal designs had been mainly reduced to two-parameter models, for which the Elfving's set was easy to built. López-Fidalgo and Rodríguez-Díaz (2004) extended the procedure to higher-dimension models.

```
cOptimElfving[E^(-b x) {1, -a x} /. {a → 1, b → 1},
 x, {0, 3}, {1, 1}, 3, 2]
```

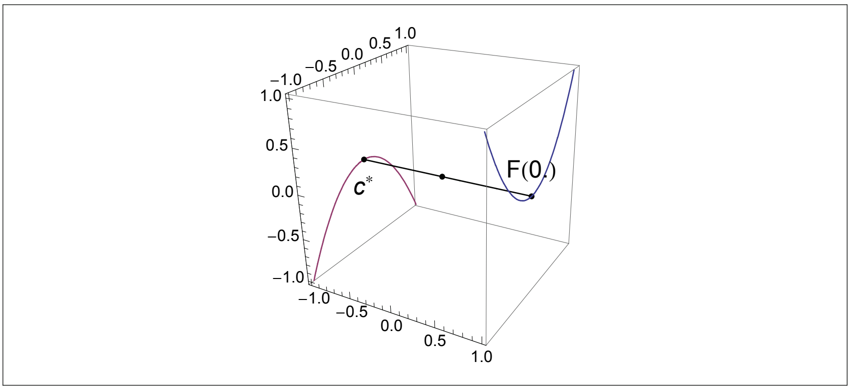



```
{{{0, 1.27844}, {0.388184, 0.611816}}, 21.0784}
```

Figure 3. Elfving's set for two-parameter model

Quatratic model : Example of 1, 2 and 3-point c-optimal designs :

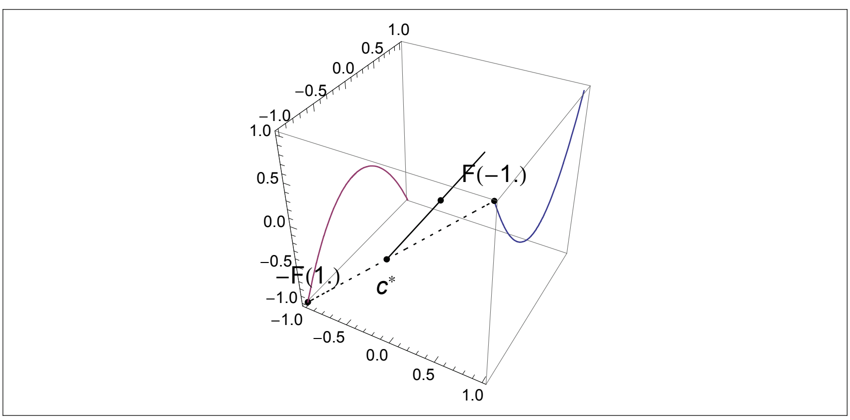
```
cOptimElfving[{1, x, x^2}, x, {-1, 1}, {1, 0, 0}, 3, 2]
```




```
{{{0.}, {1.}}, 1.}
```

Figure 4. Figure captionQuadratic model: Elfving's set for c=(1,0,0) ^{t}

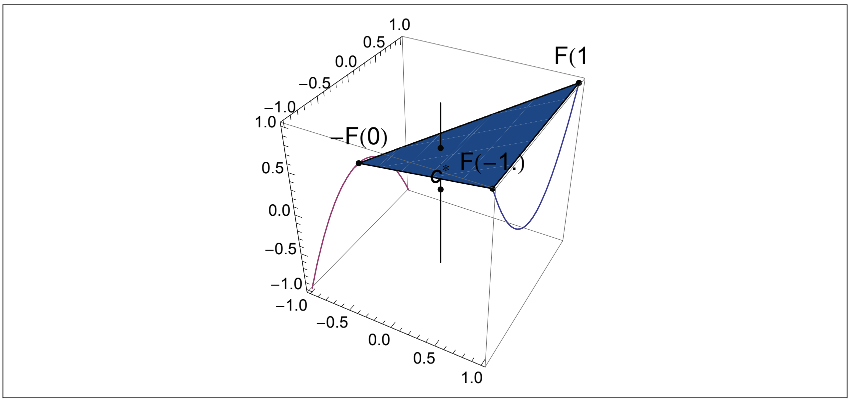
```
cOptimElfving[{1, x, x^2}, x, {-1, 1}, {0, 1, 0}, 3, 2]
```




```
{{{-1., 1.}, {0.5, 0.5}}, 1.}
```

Figure 5. Figure captionQuadratic model: Elfving's set for c=(0,1,0) ^{t}

```
cOptimElfving[{1, x, x^2}, x, {-1, 1}, {0, 0, 1}, 3, 4, 2]
```




```
{{{-1., 1., 0}, {0.25, 0.25, 0.5}}, 4.}
```

Figure 6. Figure captionQuadratic model: Elfving's set for c=(0,0,1) ^{t}

■ Discussion and further work

A new package on Optimal Design of Experiments is introduced, trying to be the origin of a repository of functions related to the subject. Nowadays, a set of general functions as well as some others related to the work of the authors is contained in the package, that should be continuously increased with new implementations. In fact, right now the group of Spatial Design Functions is going to be completed with new creations, mainly related to different types of grids.

A web application using *webMathematica* has been developed to make the package available through Internet. It should also be increased as the package is, what means a permanent work for both the programmers and the administrator of the web site.

■ Acknowledgment

This work has been partially supported for Junta de Castilla y León and grants MTM 2007 - 672111 - C03 - 01 (Ministerio de Educación y Ciencia) and HU2007 - 0007 (Acciones Integradas Program).

References

[1] A.C. Atkinson and A.N.Donev, *Optimum Experimental Designs,* New York: Oxford Science Publications, 1992.

[2] G. Elfving, "Optimum allocation in linear regression theory", *The annals of Mathematical statistics*, **23**, 1952 pp. 255-262.

[3] V.V. Fedorov, *Theory of Optimal Designs*, New York: Academic Press, 1972.

[4] J. López-Fidalgo and J.M. Rodríguez-Díaz , "Elfving method for m-dimensional models", *Metrika* **59**(3), 2004 pp. 235-244.

[5] A. Pázman, *Foundations of Optimum Experimental Design*, Dordrecht: Reidel, 1986.

[6] F. Pukelsheim, *Optimal Design of Experiments*, New York: John Wiley and Sons, 1993.

[7] F. Pukelsheim, *Optimal Design of Experiments*, Philadelphia: John Wiley and Sons, 2006.

[8] J.M. Rodríguez-Díaz and J. López-Fidalgo, "A bidimensional class of optimality criteria involving Φ_p and Characteristic Criteria", *Statistics* **37**(4), 2003 pp. 325-334.

[9] G. Sánchez G., “Fitting bioassay data and performing uncertainty analysis with BIOKMOD”, *Health Physics* **92**(1), 2007 pp. 64-72.

[10] H. P. Wynn, “The sequential generation of D-optimum experimental designs”, *The annals of mathematical statistics*, **5**, 1970 pp. 1655-1664.

About the Authors

Juan M. Rodríguez Díaz
University of Salamanca
juanmrod@usal.es

Guillermo Sánchez León
E.N.U.S.A. Industrias Avanzadas S.A.
guillermo@usal.es