Tatra Mt. Math. Publ. **17** (1999), 149–156

TATRA MOUNTAINS
Mathematical Publications

A COMPARISON OF OPTIMUM DESIGN ALGORITHMS FOR REGRESSIONS WITH CORRELATED OBSERVATIONS – A COMPUTATIONAL STUDY

Ernst Glatzer — Werner G. Müller

ABSTRACT. There exist several algorithms for the calculation of optimum (or improved) designs in the regression case under correlated observations with a given covariance function. One of them was proposed by Brimkulov *et al.* (1980) and it can be interpreted as replacing expressions in the definition of classical (uncorrelated observation) algorithms by their analogous versions in the correlated setup. Unfortunately, the convergence of such an algorithm to the true optimum is not guaranteed and this will be shown in several examples. A comparison with an alternative approach, the recently proposed algorithm by Müller and Pázman (1998), will be given.

1. Introduction

Regard the data y observed at design points $x \in \mathcal{X} \subset \mathbb{R}^2$ as being generated by a parameterized process (random field)

$$y(x) = \eta(x, \beta) + \varepsilon(x)\,, \tag{1}$$

where η denotes the deterministic and ε the stochastic part. The compact set $\mathcal{X}$ is the design space and β is assumed to stem from a finite dimensional parameter space. Additionally to approximating the field by a parametric model η, ε is here assumed to have zero mean $E\big[\varepsilon(x)\big] = 0$, finite variances $E\big[\varepsilon(x)^2\big] = \sigma^2(x) < \infty$ and a (known) covariance function

$$E\big[\varepsilon(x)\varepsilon(z)\big] = c(x, z)\,. \tag{2}$$

We can start by considering algorithms that are suitable for solving the problem

$$\xi^* = \operatorname{Arg}\max_{\xi} \Phi\big[M(\xi)\big]\,,$$

2000 Mathematics Subject Classification: 62K05.
Keywords: exact optimal designs, exchange type algorithms, random fields.

where ξ denotes the so-called design measure (the relative frequency of replications of design points, cf. Kiefer (1959)), and Φ a scalar design criterion. The information matrix (in the uncorrelated case)

$$M(\xi) = \sum_{x \in \mathcal{X}} \dot{\eta}(x)\, \dot{\eta}^T(x) \xi(x)$$

for the parameters β in the model (1) is evaluated at the local guesses β_0, i.e.,

$$\dot{\eta}(x) = \left. \frac{\partial \eta(x,\beta)}{\partial \beta} \right|_{\beta=\beta_0} .$$

Ideally ξ^* consists of fractions of n, the number of observations, to guarantee a feasible (exact) experiment ξ_n.

2. Algorithms in the uncorrelated case

The so-called sensitivity function ϕ completely determines the support of any optimum design. It is constructed by evaluating the variable part of the directional derivative of the design criterion, see, e.g., Fedorov and Hackl (1997). This function 'senses' the influence of a particular point $x^{(0)}$ on the design measure, and $\phi(x^{(0)}, \xi) > c\big[M(\xi)\big]$ indicates that additional experimental effort at this location will pay off with respect to an increase in the criterion. Here, $c\big[M(\xi)\big]$ is a design dependent constant.

It is therefore natural to suggest in a particular situation (given an existing design $\xi^{(0)}$) a construction of a new design $\xi^{(1)}$ by increasing the measure at the point $x^{(0)}$ and reducing it everywhere else, i.e.,

$$\xi^{(1)}(x) = \big(1 - \alpha^{(0)}\big)\xi^{(0)}(x) + \alpha^{(0)}\delta_{x,x^{(0)}} , \qquad \forall x \in \mathcal{X} , \tag{3}$$

where $0 < \alpha^{(0)} < 1$. This can be done iteratively ($s = 0, 1, 2, \dots$) in the hope that eventually no point $x^{(s)}$ with

$$\phi\big(x^{(s)}, \xi^{(s)}\big) > c\big[M(\xi^{(s)})\big]$$

remains, which indicates that $\xi^{(s)}$ has converged to the optimum design ξ^*.

The most common choice for the determination of the candidate point where the measure is increased relative to the other points is

$$x^{(s)} = \arg\max_x \phi\big(x, \xi^{(s)}\big) , \tag{4}$$

although it would be sufficient to choose any $x^{(s)}$ for which the sensitivity function is large.

Algorithms based upon this idea of one-point correction are predominant in the literature, one prototype being that proposed by Wynn (1970). Note that those types of algorithms are only useful for calculating approximate designs. Adaptations for finding exact designs are sometimes straightforward, the simplest is to use $\alpha^{(s)} = \frac{1}{n}$ and subsequent backward and forward excursions. For a detailed account of these refinements see Nguyen and Miller (1992).

3. The correlated setup

Here, the situation is more complex, since the information matrix

$$M(\xi) = \frac{1}{n} \sum_{i,j=1}^{n} \dot{\eta}(x_i) \left[C^{-1}(\xi)\right]_{i,j} \dot{\eta}^T(x_j),$$

with $\left[C(\xi)\right]_{i,j} = c(x_i, x_j)$ is neither convex nor additive in $\xi(x)$. Thus standard theory based upon Kiefer's definition of design measures can be used no longer.

3.1 A heuristic procedure.

It is a straightforward idea to adopt algorithms developed for the independent error case, by replacing expressions in the definition of the sensitivity function by their analogous versions in the correlated setup. This suggestion goes back to Brimkulov *et al.* (1980) who give algorithms for some of the most widely used design criteria. E.g., for the D-criterion $\Phi[M] = -\ln|M|$, where the classical sensitivity function is just the ratio of the prediction variance under the estimated parameter to the prediction variance under the true parameter, cf. Fedorov and Hackl (1997),

$$\phi_D(x, \xi_n) = \frac{E\left[\left(y(x) - \hat{y}(x)\right)^2\right]}{\sigma^2(x)}$$

can be replaced by

$$\tilde{\phi}_D(x, \xi_n) = \frac{\tilde{\sigma}^2(x, \xi_n) + \tilde{\eta}^T(x, \xi_n) M^{-1}(\xi_n) \tilde{\eta}(x, \xi_n)}{\tilde{\sigma}^2(x, \xi_n)}, \tag{5}$$

where

$$\tilde{\sigma}^2(x, \xi_n) = c(x, x) - c^T(x, \xi_n) C^{-1}(\xi_n)\, c(x, \xi_n),$$
$$\tilde{\eta}(x, \xi_n) = \dot{\eta}(x) - \dot{\eta}^T(\xi_n) C^{-1}(\xi_n)\, c(x, \xi_n),$$

$\left[\dot{\eta}(\xi_n)\right]_i = \dot{\eta}(x_i)$, and $\left[c(x, \xi_n)\right]_i = c(x, x_i)$, $i, j = 1, \ldots, n$. The 'sensitivity function' (5) is then maximized (cf. (4)) in every step s of the algorithm.

Originally, (5) was derived from the classical one-point updating formula of the determinant of a symmetric matrix. However, the above statistical interpretation makes clear why an improvement of the design by a correction algorithm based upon $\tilde{\phi}_D$ can be expected. Unfortunately there exists no proof of convergence of this heuristic algorithm (as in the uncorrelated case) and although refinements of the procedure have been suggested, see, e.g., Näther (1985), none of them apparently guarantees perfect performance.

3.2 An approach based upon design measures.

An alternative universal algorithm for finding designs in the presence of correlated errors can be based upon a new interpretation of design measures and correspondingly defined information matrices that was recently proposed by Pázman and Müller (1998). Here we give only a rough description of this algorithm for D-optimality; a general version, details of its derivation and refinements can be found in Müller and Pázman (1999).

The procedure is also a one-point correction algorithm and is based on a directional derivative (a 'sensitivity function' $\bar{\phi}_D(\xi, x)$). Unfortunately this derivative is not available in a simple functional form. A case distinction is required and therefore at every step s, the candidate points x_A and x_B need to be calculated as follows:

$$x_A = \arg \max_{\{x:\xi^{(s)}(x) \leq \frac{1}{n}\}} \bar{\phi}_A(\xi^{(s)}, x),$$

$$x_B = \arg \max_{\{x:\xi^{(s)}(x) > \frac{1}{n}\}} \bar{\phi}_B(\xi^{(s)}, x),$$

where

$$\bar{\phi}_A(\xi^{(s)}, x) = \frac{g(\xi^{(s)}, x)}{\xi^{(s)}(x)},$$

and

$$\bar{\phi}_B(\xi^{(s)}, x) = \frac{1}{\xi^{(s)}(x) - \frac{1}{n}} \left\{ g(\xi^{(s)}, x) - \left[\sum_{x' \in \mathcal{X}} g(\xi^{(s)}, x') \right] h(x) \right\}.$$

The following definitions are needed (here $\rho > 0$ is an auxiliary parameter for computational purposes):

$$(1) \quad g(\xi^{(s)}, x) = \bar{\eta}^T(x) \left[\sum_{x', z \in \mathcal{X}} \dot{\eta}(x') \left[C(\xi^{(0)}) + \rho V(\xi^{(s)}) \right]^{-1}_{x', z} \dot{\eta}^T(z) \right]^{-1} \bar{\eta}(x),$$

$$(2) \quad \bar{\eta}(x) = \sum_{z \in \mathcal{X}} \left[C(\xi^{(0)}) + \rho V(\xi^{(s)}) \right]^{-1}_{x, z} \dot{\eta}(z),$$

$$(3) \quad V(\xi^{(s)}) = \operatorname{diag}\left\{ \ln \frac{\min\{\max_x \xi^{(s)}(x), \frac{1}{n}\}}{\min\{\xi^{(s)}(x), \frac{1}{n}\}} \; ; \; x \in \mathcal{X} \right\},$$

$$(4)\quad h(x) = \begin{cases} \frac{1}{n_{\max}} & \text{if } \xi^{(s)}(x) = \max\limits_{x' \in \mathcal{X}} \xi^{(s)}(x')\,, \\ 0 & \text{else}\,, \end{cases}$$

with $n_{\max}$ being the number of points x such that $\xi^{(s)}(x) = \max\limits_{x' \in X} \xi^{(s)}(x')$.

The candidate point for the one-point correction is then given by the rule

$$x^{(s)} = \begin{cases} x_A & \text{if } \begin{array}{l} \bar{\phi}_A(\xi^{(s)}, x_A) > 0\,, \\ \bar{\phi}_A(\xi^{(s)}, x_A) = 0\,, \quad \bar{\phi}_B(\xi^{(s)}, x_B) < 0\,, \end{array} \\ x_B & \text{if } \begin{array}{l} \bar{\phi}_A(\xi^{(s)}, x_A) < 0\,, \\ \bar{\phi}_A(\xi^{(s)}, x_A) = 0\,, \quad \bar{\phi}_B(\xi^{(s)}, x_B) > 0\,. \end{array} \end{cases} \tag{6}$$

Note the similarity of the algorithm with (5) in only slightly correlated (or uncorrelated) schemes with uniform design measure.

4. The practical performance of the algorithms

For detection of failures and for comparison of the efficiency of the two algorithms presented above, a number of examples from the literature were chosen for which an exact solution of the design problem was available.

(1) Example 1, Näther (1985): The simple linear regression

$$y(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}^T \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \varepsilon(x)\,, \quad x \in [-1, 1]$$

with covariance function

$$c(x,z) = \begin{cases} 1 - |x - z|\,; & |x - z| < 1\,, \\ 0\,; & |x - z| \geq 1\,. \end{cases}$$

Here, the optimal design consists of the three points $\{-1, 0, 1\}$.

(2) Example 2, Müller (1995): Consider again a simple linear regression and that now

$$c(x,z) = \begin{cases} \sigma^2, & x = z\,, \\ 0.4\sigma^2, & 0 < |x - z| \leq 1.8\,, \\ 0\,, & \text{else}\,. \end{cases}$$

The optimum three-point designs on a grid $\mathcal{X} = [-1, -0.875, \ldots, 1]$ are either $\{-1, -0.75, 1\}$ or $\{-1, 0.75, 1\}$.

(3) Example 3, Sacks and Ylvisaker (1968): The quadratic regression

$$\eta(x) = \beta x^2,$$

with Brownian motion noise, with the consequent covariance-function

$$c(x,z) = \min(x,z), \qquad x,z \in [0,1].$$

The optimal n-point design $\{x_1^*, \dots, x_n^*\}$ is given by

$$x_i^* = \frac{i}{n}.$$

We consider the cases $n = 6$ and $n = 8$ on a 24 point grid.

(4) Example 4, Bischoff (1992): Consider now the trigonometric regression

$$\dot{\eta}^T(x) = (1, \sin x, \cos x),$$

with

$$c(x,z) = 1 + \delta_{x,z}, \qquad x,z \in [0, 2\pi).$$

Here the optimum n-point design is given by

$$x_i^* = \frac{2\pi(i-1)}{n},$$

which can be realized on an equidistant 24 point grid for our choices $n = 4$ and $n = 6$.

(5) Example 5, Rasch (1997): The uncorrelated setting, where replication of observations at the same design points are not allowed is a special case of ours with $c(x,z) = \delta_{x,z}$. As an example we chose here

$$\eta(x) = \beta_1 + \beta_2 e^{\beta_3 x}, \qquad x \in [0, 25],$$

with $\beta_0^T = (1, -1.4, -0.2)$, for which the 9-point replication-free D-optimal design on a 26 point grid is known to be $\{0, 1, 4, 5, 6, 22, 23, 24, 25\}$.

Both algorithms contain user-specified components. In the case of (5) this is only the initial design $\xi_n^{(0)}$, whereas for (6) we have the real-valued parameter $\rho \geq 1$ and the initial measure $\xi^{(0)}$ (with support on the whole $\mathcal{X}$). Thus in an automated version of the algorithms, which is needed for the simulations, we can in (5) simply choose a random collection of n points out of the region $\mathcal{X}$ as the starting design.

For (6) an automated version of the initial stage needs a little more thought. It turns out that a good initial measure seems to be the uniform on $\mathcal{X}$, probably because no particular point is a priori favoured. In a sequence of initiation runs with increasing $\rho = 10^j$, $j = 0, 1, 2, \dots$, the best ρ is determined and then actual runs are performed with this ρ first with a uniform, subsequently with a random starting measure if the optimum was not found immediately.

The average D-efficiencies for the results from the two algorithms (always started from the same random design) in 1000 runs each are reported in Table 1.

TABLE 1. Average D-efficiencies (in %) of competing designs in various examples, * efficiency can be increased to 100 % if started by uniform design.

Example	n	random	Brimkulov *et al.*	Müller & Pázman
1: Näther	3	31.0	100.0	100.0
2: Müller	3	36.4	99.2	97.4*
3: Sacks & Ylvisaker	6	72.4	99.9	98.7
	8	80.3	99.9	99.9*
4: Bischoff	4	42.1	98.8	89.1
	6	63.2	99.7	94.5*
5: Rasch	9	25.0	100.0	89.0*

The improvements over a random design, where n support points are selected independently and with equal probabilities from all grid points, are considerable for all examples with algorithm (5) performing slightly better.

If the exact design is needed rather than a reasonable approximation, however, both algorithms can have severe problems of finding the global optimum. To evaluate this property it is useful to record the average number of runs that are necessary for finding the global optimum and relate them to the computation speed of the respective algorithms. All computations were carried out on a Pentium PC with 233 MHz and the results are reported in Table 2.

TABLE 2. Computation times (in hundredths of seconds) per run and average number of hits (true global optimum) per minute of competing designs in various examples, (ρ) indicates knowledge of best value for ρ.

Example	n	Müller & Pázman			Brimkulov *et al.*	
		hsec	h/min (ρ)	h/min	h/min	hsec
1: Näther	3	4.07	1474	1474	2765	2.17
2: Müller	3	0.88	6818	3409	1233	3.80
3: Sacks & Ylvisaker	6	45.09	12	10	80	11.40
	8	17.63	340	340	52	11.53
4: Bischoff	4	1.76	37	31	318	5.70
	6	11.59	518	29	380	9.50
5: Rasch	9	21.86	274	91	254	23.62

5. Conclusions

Probably due to their simplicity and their lack of restrictive assumptions on η or c, algorithms based on the heuristic idea presented in Section 3.1 are the most popular in applications of correlated design problems (see, e.g., Rabinowitz and Steinberg, 1990). However, it seems that there are many instances, where an alternative method, such as the one from Section 3.2, offers complimentary backup for the calculated designs or sometimes even leads to the detection of nonconvergence. In a number of cases the new algorithm seems to be more competitive with respect to computation time. An investigation on larger scale problems seems to be appropriate.

REFERENCES

[1] BISCHOFF, W.: *On exact D-optimal designs for regression models with correlated observations*, Ann. Inst. Statist. Math. **44** (1992), 229–238.

[2] BRIMKULOV, U. N.—KRUG, G. K.—SAVANOV, V. L.: *Numerical construction of exact experimental designs when the measurements are correlated*, Zavodskaya Laboratoria **36** (1980), 435–442. (In Russian.)

[3] FEDOROV, V. V.—HACKL, P.: *Model-Oriented Design of Experiments*, Lecture Notes in Statistics, Vol. 125, Springer Verlag, Berlin, 1997.

[4] KIEFER, J.: *Optimal experimental designs (with discussion)*, J. Roy. Statist. Soc. **B** (1959), 272–319.

[5] MÜLLER, W. G.: *An example of optimal design for correlated observations*, Österreichische Zeitschrift für Statistik **24** (1995), 9–15. (In German.)

[6] MÜLLER, W. G.—PÁZMAN, A.: *An algorithm for the computation of optimum designs under a given covariance structure*, Comput. Statistics **14** (1999).

[7] NÄTHER, W.: *Effective Observations of Random Fields*, Teubner-Texte zur Mathematik, Leipzig, 1985.

[8] NGUYEN, N.—MILLER, A. J.: *A review of some exchange algorithms for constructing discrete D-optimal designs*, Comput. Statist. Data Anal. **14** (1992), 489–498.

[9] PÁZMAN, A.—MÜLLER, W. G.: *A new interpretation of design measures*, in: MODA5 (Atkinson, A. C., Pronzato, L., and Wynn, H. P., eds.), Physica Verlag, 1998, pp. 239–246.

[10] RABINOWITZ, N.—STEINBERG, D. M.: *Optimal configuration of a seismographic network: A statistical approach*, Bull. Seismol. Soc. Amer. **80** (1990), 187–196.

[11] RASCH, D. A. M. K.—HENDRIX, E. M. T.—BOER, E. P. J.: *Replicationfree optimal designs in regression analysis*, Comput. Statistics **12** (1997), 19–32.

[12] SACKS, J.—YLVISAKER, D.: *Design for regression problems with correlated errors*, Ann. Math. Statistics **37** (1966), 66–89.

[13] WYNN, H. P. (1970): *The sequential generation of D-optimum experimental designs*, Ann. Math. Statistics **41** (1970), 1655–1664.

Received March 15, 1998

Department of Statistics
University of Economics
Augasse 2–6
A–1090 Vienna
AUSTRIA

E-mail: ernst.glatzer@wu-wien.ac.at
werner.mueller@wu-wien.ac.at