



Algorithms for approximate linear regression design with application to a first order model with heteroscedasticity



N. Gaffke^{a,*}, U. Graßhoff^b, R. Schwabe^a

^a University of Magdeburg, Faculty of Mathematics, PF 4120, D-39016 Magdeburg, Germany

^b University of Münster, Germany

ARTICLE INFO

Article history:

Received 29 May 2012

Received in revised form 18 July 2013

Accepted 20 July 2013

Available online 27 July 2013

Keywords:

Information matrix

Optimality criterion

Steepest descent

quasi-Newton method

Efficiency

Invariant design

ABSTRACT

The basic structure of algorithms for numerical computation of optimal approximate linear regression designs is briefly summarized. First order methods are contrasted to second order methods. A first order method, also called a vertex direction method, uses a local linear approximation of the optimality criterion at the actual point. A second order method is a Newton or quasi-Newton method, employing a local quadratic approximation. Specific application is given to a multiple first order regression model on a cube with heteroscedasticity caused by random coefficients with known dispersion matrix. For a general (positive definite) dispersion matrix the algorithms work for moderate dimension of the cube. If the dispersion matrix is diagonal, a restriction to invariant designs is legal by equivariance of the model and the algorithms also work for large dimension.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The concept of approximate linear regression design can be summarized as follows, where for shortness we omit the details on the linear regression model. Let \mathcal{X} be a given design region. For every $x \in \mathcal{X}$ there is given an information matrix $M(x)$ which is a nonnegative definite $p \times p$ matrix. The set $\{M(x) : x \in \mathcal{X}\}$ is assumed to be compact. An approximate design ξ consists of a finite subset of \mathcal{X} called the support of ξ and of corresponding weights,

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_r \\ w_1 & w_2 & \dots & w_r \end{pmatrix}, \quad \text{where } r \in \mathbb{N}, x_i \in \mathcal{X}, w_i > 0, \sum_{i=1}^r w_i = 1.$$

The information matrix of ξ is defined to be

$$M(\xi) = \sum_{i=1}^r w_i M(x_i).$$

In particular, the set of all information matrices of approximate designs constitutes a convex and compact set which is the convex hull of the one-point information matrices,

$$\mathfrak{M} = \{M(\xi) : \xi \text{ an approximate design}\} = \text{Conv} \{M(x) : x \in \mathcal{X}\}.$$

Design optimality refers to an optimality criterion Φ . Here we deal only with optimality criteria whose domain is the set $\text{PD}(p)$ of all positive definite $p \times p$ matrices, and moreover we assume Φ to be convex,

$$\Phi : \text{PD}(p) \longrightarrow \mathbb{R} \quad \text{convex.}$$

* Corresponding author. Tel.: +49 3916758306; fax: +49 3916711172.

E-mail addresses: norbert.gaffke@ovgu.de (N. Gaffke), rainer.schwabe@ovgu.de (R. Schwabe).

An approximate design ξ^* is said to be Φ -optimal, if its information matrix $M^* = M(\xi^*)$ is positive definite and minimizes $\Phi(M)$ over all $M \in \mathfrak{M} \cap \text{PD}(p)$. In particular, only designs with nonsingular information matrices are considered for optimality. Statistically this means that a design should allow the estimation of all p regression coefficients. Popular optimality criteria are the D-criterion and L-criteria given by

$$\Phi_D(M) = (\det(M))^{-1/p}, \quad \Phi_L(M) = \text{tr}(W M^{-1}) \quad \text{for all } M \in \text{PD}(p),$$

where Φ_L employs a given “weight matrix” $W \in \text{PD}(p)$. We note that especially for the D-criterion other (equivalent) formulations are used, e.g., $-\log \det(M)$. The version we prefer has the advantage of being positive and positively homogeneous of order -1 , so that a natural notion of efficiency of an information matrix (or of a design) can be used (see Section 2). The problem of computing a Φ -optimal design constitutes a convex minimization problem on the (convex) set $\mathfrak{M} \cap \text{PD}(p)$. For practical programming purposes as well as for gaining more generality we consider minimization problems in a vector variable m (instead of a matrix variable M) of the following (more general) type.

$$\text{minimize } \phi(m) \quad \text{over } m \in \mathcal{M} \cap \mathcal{A}, \tag{1.1}$$

where $\emptyset \neq \mathcal{A} \subseteq \mathbb{R}^q$, \mathcal{A} convex and open, $\phi : \mathcal{A} \rightarrow \mathbb{R}$ convex,

and $\emptyset \neq \mathcal{M} \subseteq \text{cl}(\mathcal{A})$, \mathcal{M} convex and compact,

where $\text{cl}(\mathcal{A})$ denotes the closure of \mathcal{A} . A condition which obviously ensures the existence of an optimal solution to (1.1) is that for some $m_0 \in \mathcal{M} \cap \mathcal{A}$ the set $\{m \in \mathcal{M} \cap \mathcal{A} : \phi(m) \leq \phi(m_0)\}$ is compact. This is usually satisfied for optimal design problems on D- or L-optimality, as well as for many other optimality criteria, when stated like (1.1). Moreover, strict convexity of the D- and L-criteria implies uniqueness of the optimal solution to (1.1).

Typically in the context of optimal design, the set \mathcal{M} is given as the convex hull of a compact family $m(x) \in \mathbb{R}^q$, $x \in \mathcal{X}$, i.e.,

$$\mathcal{M} = \text{Conv} \{m(x) : x \in \mathcal{X}\}, \quad \text{where } \{m(x) : x \in \mathcal{X}\} \text{ is compact.} \tag{1.2}$$

Then, one is interested in an optimal design rather than in an optimal vector m^* (an optimal solution to (1.1)) which puts the additional problem of finding a decomposition of the vector $m^* \in \mathcal{M}$, according to (1.2):

$$\text{Find } r \in \mathbb{N}, x_1, \dots, x_r \in \mathcal{X}, \text{ and } w_1, \dots, w_r > 0, \quad \sum_{i=1}^r w_i = 1, \text{ s.t. } m^* = \sum_{i=1}^r w_i m(x_i).$$

Then a corresponding design ξ^* has support points x_i and weights w_i , $i = 1, \dots, r$. The problem of finding a decomposition of a given vector $m \in \mathcal{M} = \text{Conv} \{m(x) : x \in \mathcal{X}\}$ will be solved, where the number r will not exceed Caratheodory's bound $q + 1$, (see Section 2). In our algorithms the iteration process employs q -dimensional vectors m instead of designs ξ , which has the advantage of a finite dimension q . In particular, no heuristics in the iteration process are needed for removing support points with small weights or substitution of a cluster of support points by a single point. A design is constructed in a final step after the iteration process. This is a novel feature of our methods.

Further viewpoints presented in the paper might be interesting, though not completely new. Classical algorithms for optimal design require at each step minimization of a linear function over the “vertex set” $\{m(x) : x \in \mathcal{X}\}$. Here a reduction to “ c -fractional minimization” is introduced which proves to be useful in applications to a particular model in Section 3. A first order method using vertex directions (“ c -fractional steepest descent”) is contrasted to a second order method of quasi-Newton type. As a main difference in the performance of both methods a greater accuracy of the latter in approaching the optimum is observed. An important ingredient of our quasi-Newton method is a modification of an algorithm of Higgins and Polak (1990) to minimize a convex quadratic function over \mathcal{M} . This may be interesting by itself since it is one of the very few algorithms from optimization theory dealing with a feasible region defined as a convex hull like (1.2), which is the typical situation in optimal approximate linear regression design. Also it constitutes a tool for computing a decomposition of a point from the convex hull to obtain a corresponding design.

The paper is arranged as follows. In Section 2 the basic structure of the algorithms for solving (1.1) is outlined. In Section 3 we focus to a particular model: multiple first order regression on the cube $[-1, 1]^k$ with heteroscedasticity caused by random coefficients as studied previously by Graßhoff et al. (2012). We apply the algorithms for moderate dimensions $k \leq 10$. The case of a diagonal dispersion matrix of the random coefficients considered in Section 4 allows a reduction by invariance, which reduces the computational effort considerably. A brief discussion of the results is given in Section 5. For convenience the modified Higgins–Polak algorithm is outlined in Appendix A, and a further Appendix B refers to the particular model from Section 3 and describes how to find the global minimum of a given quadratic function (in general, neither convex nor concave) over the cube $[-1, 1]^k$. This is to achieve c -fractional linear minimization within the algorithms applied to the particular model from Section 3, but it forces restriction to moderate k .

2. Structure of the algorithms

In what follows, we assume (1.2) and that the (convex) optimality criterion ϕ is differentiable or even twice continuously differentiable. So non-differentiable optimality criteria, as e.g. the E-criterion, will be excluded.

First order methods using vertex directions date back to Wynn (1970, 1972), Fedorov (1972), and others (see Gaffke and Mathar, 1992, pp. 91–92, for further references). Second order methods were established firstly by Atwood (1976) and

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات