ARTICLE IN PRESS

Journal of Statistical Planning and Inference xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Quasi-Newton algorithm for optimal approximate linear regression design: Optimization in matrix space

N. Gaffke *, R. Schwabe

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

ARTICLE INFO

Article history:
Received 19 December 2016
Received in revised form 13 March 2018
Accepted 17 March 2018
Available online xxxx

MSC:
62K05
90C25

Keywords:
Information matrix
Optimality criterion
Local optimal design
Convex hull
Efficiency
Gradient
Local quadratic approximation
Convex quadratic minimization
BFGS update
Multiplicative algorithm

ABSTRACT

Given a linear regression model and an experimental region for the independent variable, the problem of finding an optimal approximate design calls for minimizing a convex optimality criterion over a convex set of information matrices of feasible approximate designs. For numerical solution pure gradient methods are often used by design theorists, as vertex direction, vertex exchange, multiplicative algorithms, or combinations hereof. These methods have two major deficiencies: a slow convergence rate after a quick but rough approximation to the optimum, and often a large support of the obtained nearly optimal design. The latter feature is related to the fact that the methods optimize in the space of design measures which is usually of high or even infinite dimension, whereas the dimension of the information matrices is often small or moderate. For such situations a quasi-Newton method is revisited which was originally established by Gaffke & Heiligers (1996). In the present paper new possibilities of its application are demonstrated. The algorithm optimizes in matrix space. It shows a good global and an excellent local convergence behavior resulting in an accurate approximation of the optimum. A crucial subroutine solves convex quadratic minimization over the set of information matrices via repeated linear minimization over that set, providing thus the quasi-Newton step of the algorithm. This may also be of interest as a tool for computing an approximate design from a given information matrix and such that the support size of the design keeps Carathéodory's bound. Illustrations are given for D- and I-optimality in particular multivariate random coefficient regression models and for T-optimal discriminating design in univariate polynomial models. Moreover, the behavior of the algorithm is tested for cases of larger dimensions: D- and I-optimal design for a third order polynomial model in several variables on a discretized cube.

1. Introduction and basic notation

As a general basis for optimal approximate linear regression design we take the following situation. Let $\mathcal{X}$ denote the experimental region for the independent variable x. To each $x \in \mathcal{X}$ there is given a nonnegative definite $p \times p$ matrix $M(x)$, which we call the elementary information matrix of the design point x. A standard situation is $M(x) = f(x)f(x)^{\mathrm{T}}$ for all $x \in \mathcal{X}$, where f is a given $\mathbb{R}^p$-valued function on $\mathcal{X}$ defining a univariate linear regression model. Situations different from the standard one arise, e. g., in multivariate linear regression (see Section 4), in stratified design (see Harman, 2014) or in size- and cost-constrained design (see Harman and Benkova, 2016). The set $\{M(x) : x \in \mathcal{X}\}$ of elementary information matrices is assumed to be compact.

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

https://doi.org/10.1016/j.jspi.2018.03.005
0378-3758/

An approximate design or design measure ξ is a discrete probability distribution with finite support on the experimental region $\mathcal{X}$,

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_r \\ w_1 & w_2 & \dots & w_r \end{pmatrix},$$

$$\text{where } r \in \mathbb{N},\ x_j \in \mathcal{X} \text{ all distinct, } w_j > 0,\ \sum_{j=1}^{r} w_j = 1\,. \tag{1.1}$$

The set $\operatorname{supp}(\xi) = \{x_1, \dots, x_r\}$ is called the support of ξ. Note that the (finite) size r is not fixed. Different designs may have different numbers of support points.

The information matrices of designs come linearly from the family of elementary information matrices $M(x)$ assigned to the points $x \in \mathcal{X}$. The information matrix of a design ξ from (1.1) is given by

$$M(\xi) = \sum_{j=1}^{r} w_j M(x_j)\,. \tag{1.2}$$

Clearly, $M(\xi)$ is a nonnegative definite $p \times p$ matrix which may be nonsingular (hence positive definite) or singular. Often, when dealing with design optimality, positive definiteness of the information matrix is required and is thus imposed as a feasibility condition on a design. We will employ a more general feasibility condition possibly allowing designs with singular information matrices (see below). But throughout we assume that there exist designs with positive definite information matrices. Denote by $\mathcal{M}$ the set of all information matrices of approximate designs,

$$\mathcal{M} = \big\{M(\xi) : \xi \text{ an approximate design on } \mathcal{X}\big\}\,. \tag{1.3}$$

By definition (1.2) $\mathcal{M}$ coincides with the convex hull of the set of all elementary information matrices,

$$\mathcal{M} = \operatorname{Conv}\big\{M(x) : x \in \mathcal{X}\big\}\,, \tag{1.4}$$

where 'Conv' stands for 'convex hull'. The assumption of existence of designs with positive definite information matrices can be stated as

$$\mathcal{M} \cap \operatorname{PD}(p) \neq \emptyset\,, \tag{1.5}$$

where $\operatorname{PD}(p)$ denotes the set of all positive definite $p \times p$ matrices. Also, we denote by $\operatorname{NND}(p)$ the set of all nonnegative definite $p \times p$ matrices.

The sets of matrices introduced so far are covered by the linear space $\operatorname{SYM}(p)$ of all symmetric $p \times p$ matrices, which is endowed with the scalar product

$$\langle A, B\rangle = \operatorname{tr}(AB) \quad \text{for all } A, B \in \operatorname{SYM}(p),$$

where $\operatorname{tr}(C)$ denotes the trace of a $p \times p$ matrix C. Obviously, $\operatorname{PD}(p)$ is an open subset of $\operatorname{SYM}(p)$ and a convex cone. The latter means that it is closed under addition and under multiplication by positive scalars. More generally, let $\mathcal{A} \subseteq \operatorname{SYM}(p)$ be a convex cone such that $\operatorname{PD}(p) \subseteq \mathcal{A}$. Note that the latter inclusion implies the inclusion $\operatorname{NND}(p) \subseteq \operatorname{cl}(\mathcal{A})$ and in particular $\mathcal{M} \subseteq \operatorname{cl}(\mathcal{A})$ where $\operatorname{cl}(\mathcal{A})$ denotes the closure of $\mathcal{A}$. We will refer to $\mathcal{A}$ as a 'feasibility cone'. An optimality criterion is a real-valued convex function Φ on some feasibility cone $\mathcal{A}$. A design ξ^* is called Φ-optimal if $M(\xi^*) \in \mathcal{A}$ and

$$\Phi\big(M(\xi^*)\big) \leq \Phi\big(M(\xi)\big) \quad \text{for all } \xi \text{ with } M(\xi) \in \mathcal{A}. \tag{1.6}$$

Popular criteria on $\mathcal{A} = \operatorname{PD}(p)$ are the D-criterion $\Phi_{\mathrm{D}}(M) = \big(\det(M)\big)^{-1/p}$, and an L-criterion $\Phi_{\mathrm{L}}(M) = \operatorname{tr}\big(WM^{-1}\big)$ with a given weight matrix $W \in \operatorname{PD}(p)$. The popular A-criterion is a special L-criterion with $W = I_p$ the p-dimensional identity matrix. We will employ another special L-criterion which is called an I-criterion since it expresses some integral average (w.r.t. Lebesgue measure) of variances of estimated responses. An example of a criterion allowing (certain) singular matrix arguments is given by a 'singular' L-criterion, i. e. the weight matrix $W \in \operatorname{NND}(p) \setminus \{0\}$ is singular and $\Phi_{\mathrm{L}}(M) = \operatorname{tr}\big(WM^{-}\big)$ for all $M \in \mathcal{A}$, where M^{-} denotes any generalized inverse of M, i.e., any $p \times p$ matrix satisfying the equation $MM^{-}M = M$, see Rao and Rao (2001), Ch. 8.2. The feasibility cone is given by $\mathcal{A} = \big\{M \in \operatorname{NND}(p) : \operatorname{range}(W) \subseteq \operatorname{range}(M)\big\}$. In particular, for $W = cc^{\mathrm{T}}$ with a fixed nonzero vector $c \in \mathbb{R}^p$ the (singular) L-criterion is called c-criterion and can be written as $\Phi_c(M) = c^{\mathrm{T}}M^{-}c$ on the feasibility cone $\mathcal{A} = \big\{M \in \operatorname{NND}(p) : c \in \operatorname{range}(M)\big\}$.

The class of T-criteria was introduced by Atkinson and Fedorov (1975) for discriminating between two models. Here we restrict to the case of two nested linear models. Let p be the dimension of the larger model and $p - s$ the dimension of the smaller model, where $1 \leq s < p$. Accordingly, we partition any $M \in \operatorname{SYM}(p)$ as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^{\mathrm{T}} & M_{22} \end{bmatrix},\ \text{with } M_{11}\ (p-s) \times (p-s),\ M_{22}\ s \times s, \tag{1.7}$$