Abstracts of the Ninth Workshop on Simulation (2018) 1-2

A Quasi-Newton Algorithm for Optimal Approximate Linear Regression Design

Norbert Gaffke [1]

1 Introduction

Given a linear regression model and an experimental region for the independent variable, a problem of optimal approximate design leads to minimizing a convex criterion function $\Phi(M)$ over the set of all information matrices $M(\xi)$ of feasible approximate designs ξ. The set $\mathcal{M} = \big\{M(\xi) \,:\, \xi \text{ any approximate design}\big\}$ is typically given as a convex hull of the set of all information matrices of elementary designs,

$$\mathcal{M} \,=\, \mathrm{Conv}\,\big\{M(x) \,:\, x \in \mathcal{X}\big\}, \tag{1}$$

where $\mathcal{X}$ denotes the experimental region and $M(x)$ is the elementary information matrix at the design point x which is a nonnegative definite $p \times p$ -matrix. The optimization problem reads as

$$\text{minimize} \quad \Phi(M) \quad \text{over} \quad M \in \mathcal{M} \cap \mathcal{A}\,, \tag{2}$$

where $\mathcal{A}$ is a given 'feasibility cone' constituting the domain of Φ, i.e., $\mathcal{A}$ is a convex cone of symmetric $p \times p$ -matrices containing all positive definite $p \times p$ -matrices. It is assumed that the generating set $\{M(x) \,:\, x \in \mathcal{X}\}$ is compact, its convex hull $\mathcal{M}$ contains some positive definite matrix, and the (convex) criterion function Φ is twice continuously differentiable on $\mathrm{int}(\mathcal{A})$, the interior of $\mathcal{A}$. Moreover, the algorithm requires that linear minimization over $\mathcal{M}$, or equivalently over its generating set, can easily be done, i.e., a subroutine is available to solve the problem

$$\text{minimize} \quad \mathrm{tr}\big(A\,M(x)\big) \quad \text{over} \quad x \in \mathcal{X}\,, \tag{3}$$

for any given symmetric $p \times p$ -matrix A. Note that for a finite experimental region $\mathcal{X}$ linear minimization is trivial, unless $\mathcal{X}$ is tremendously large. The quasi-Newton algorithm for solving (2) was originally established in [2]. The recent paper [3] has demonstrated new possibilities of applications of the algorithm.

2 Outline of the algorithm

As a main tool the algorithm employs a subroutine which provides minimization over $\mathcal{M}$ of any given convex quadratic function via repeatedly solving linear minimization problems (3). The subroutine is an adaptation of a more general method

[1] University of Magdeburg, Faculty of Mathematics, PF 4120, D-39016 Magdeburg, Germany E-Mail: `norbert.gaffke@ovgu.de`

in [4] and therefore we call it the 'Higgins-Polak subroutine'. The outline given next of the quasi-Newton algorithm for solving (2) employs $p(p+1)/2$-dimensional column vectors m obtained by vectorization of symmetric $p \times p$ -matrices M which is convenient in view of quadratic approximations and BFGS-updates. By $g(m)$ we denote the gradient of Φ at $m \in \mathcal{M} \cap \mathrm{int}(\mathcal{A})$.

Quasi-Newton algorithm:
(o) *Initialization*: choose any $m_1 \in \mathcal{M} \cap \mathrm{int}(\mathcal{A})$; compute $g_1 = g(m_1)$; choose any $B_1 \in \mathrm{PD}\big(p(p+1)/2\big)$; set $t = 1$. Go to step (i).

(i) *Quasi-Newton step*: apply the Higgins-Polak subroutine to compute an optimal solution $\overline{m}_t \in \mathcal{M}$ to the convex quadratic minimization problem

$$\text{minimize} \quad (g_t - B_t m_t)^{\mathrm{T}} m + \tfrac{1}{2} m^{\mathrm{T}} B_t m \quad \text{over} \quad m \in \mathcal{M}.$$

Go to step (ii).

(ii) *Line search*: apply an adaptation of Fletcher's line search procedure ([1], Chapter 2.6) which computes a suitable $\alpha_t \in (0, \alpha_{\max}]$, where $\alpha_{\max}$ is a predefined constant in $(0, 1)$ usually close to 1, e. g., $\alpha_{\max} = 0.99$. Set $m_{t+1} = (1 - \alpha_t) m_t + \alpha_t \overline{m}_t$ and compute the gradient $g_{t+1} = g(m_{t+1})$. Go to step (iii).

(iii) *BFGS update*: let $\delta_t = m_{t+1} - m_t$ and $\gamma_t = g_{t+1} - g_t$. Set

$$B_{t+1} = B_t + (\gamma_t^{\mathrm{T}} \delta_t)^{-1} \gamma_t \gamma_t^{\mathrm{T}} - (\delta_t^{\mathrm{T}} B_t \delta_t)^{-1} B_t \delta_t \delta_t^{\mathrm{T}} B_t, \quad \text{if } \gamma_t^{\mathrm{T}} \delta_t > 0,$$

and set $B_{t+1} = B_t$ otherwise, i. e., if $\gamma_t^{\mathrm{T}} \delta_t = 0$.
Go to step (i) with t replaced by $t + 1$.

References

[1] Fletcher R. *Practical Methods of Optimization.* 2nd Edition, Wiley, New York, 1987.

[2] Gaffke N., Heiligers B. *Second order methods for solving extremum problems from optimal linear regression design.* Optimization, 1996, v. 36, p. 41-57.

[3] Gaffke N., Schwabe R. *Quasi-Newton algorithm for optimal approximate linear regression design: optimization in matrix space.* Submitted to: Journal of Statistical Planning and Inference, 2018.

[4] Higgins J.E., Polak E. *Minimizing pseudoconvex functions on convex compact sets.* Journal of Optimization Theory and Applications, 1990, v. 65, p. 1-27.