Journal of Statistical Planning and Inference 154 (2014) 141–155

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

# Algorithmic construction of optimal designs on compact sets for concave and differentiable criteria

Luc Pronzato [a,*], Anatoly A. Zhigljavsky [b]

[a] Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis, Bât. Euclide, Les Algorithmes, 2000 route des lucioles, BP 121, 06903 Sophia Antipolis cedex, France

[b] School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4YH, UK

ARTICLE INFO

*Article history:*
Received 5 July 2013
Received in revised form
6 February 2014
Accepted 30 April 2014
Available online 14 May 2014

*MSC:*
62K05
90C30
90C26

*Keywords:*
Approximate design
Optimum design
Construction of optimal designs
Global optimization

ABSTRACT

We consider the problem of construction of optimal experimental designs (approximate theory) on a compact subset $\mathcal{X}$ of $\mathbb{R}^d$ with nonempty interior, for a concave and Lipschitz differentiable design criterion $\phi(\cdot)$ based on the information matrix. The proposed algorithm combines (a) convex optimization for the determination of optimal weights on a support set, (b) sequential updating of this support using local optimization, and (c) finding new support candidates using properties of the directional derivative of $\phi(\cdot)$. The algorithm makes use of the compactness of $\mathcal{X}$ and relies on a finite grid $\mathcal{X}_\ell \subset \mathcal{X}$ for checking optimality. By exploiting the Lipschitz continuity of the directional derivatives of $\phi(\cdot)$, efficiency bounds on $\mathcal{X}$ are obtained and $\epsilon$-optimality on $\mathcal{X}$ is guaranteed. The effectiveness of the method is illustrated on a series of examples.



## 1. Introduction and motivation

A design measure $\xi$ on a finite set $\mathcal{X} \subset \mathbb{R}^d$ with $\ell$ elements is characterized by the $\ell$-dimensional vector of weights $W$ (nonnegative and summing to one) allocated to the $\ell$ elements of $\mathcal{X}$. The determination of an optimal measure $\xi^*$ which maximizes a concave differentiable criterion $\phi(\cdot)$ then forms a finite-dimensional convex problem for which many optimization algorithms are available, see, *e.g.*, Hiriart-Urruty and Lemaréchal (1993), den Hertog (1994), Nesterov and Nemirovskii (1994), Ben-Tal and Nemirovski (2001), Boyd and Vandenberghe (2004), and Nesterov (2004) for recent developments on convex programming. In particular, the cutting plane method of Kelley (1960) is considered by Sibson and Kenny (1975), and a variant of it (with a modified version of the Equivalence Theorem) by Gribik and Kortanek (1977), see also Pronzato and Pázman (2013, Chapter 9). However, design problems are usually such that (a) the cardinality $\ell$ of the design space, which determines the dimension of the optimization problem to be solved, is large, and (b) there always exists an optimal measure $\xi^*$ with a few support points only, *i.e.*, such that only a few components of $W$ are positive (say $m$, with $m \ll \ell$). These particularities have motivated the development of specific methods which happen to be competitive

* Corresponding author.
*E-mail addresses:* pronzato@i3s.unice.fr (L. Pronzato), ZhigljavskyAA@cf.ac.uk (A.A. Zhigljavsky).

http://dx.doi.org/10.1016/j.jspi.2014.04.005
0378-3758/

compared to general-purpose convex-programming algorithms. One may distinguish three main families, see Pronzato and Pázman (2013, Chapter 9) for a recent survey:

(i) *Vertex-direction* methods only increase one component of $W$ at each iteration, all other components being multiplied by the same factor smaller than one, see Wynn (1970), Fedorov (1972) and also Frank and Wolfe (1956) for a method originally proposed in a more general context. A significant improvement can be achieved by decreasing some individual components of $W$ only, which allows us to set some components of $W$ to zero, *i.e.*, to remove support points from a poorly chosen initial design and sometimes to exchange two components of $W$ (vertex-exchange methods), see Atwood (1973), St. John and Draper (1975), Böhning (1985, 1986), and Molchanov and Zuyev (2001, 2002).
(ii) In *gradient* methods the gradient direction projected on the set of design measures is used as the direction of movement at each iteration, see Wu (1978a, 1978b) and Atwood (1976) for an extension to a Newton-type method.
(iii) At each iteration of a *multiplicative* method, each component of $W$ is multiplied by a suitably chosen positive function, see, *e.g.*, Titterington (1976), Silvey et al. (1978), Torsney (1983), Fellman (1989), Dette et al. (2008), Yu (2010a,b), and Torsney (2009) for a historical review.

In multiplicative methods all initial weights must be positive, those which should be zero at the optimum decrease continuously along iterations but typically stay strictly positive, *i.e.*, never achieve exactly zero. The convergence is inevitably slow close to the optimum. The same phenomenon occurs for vertex-direction methods, unless the decrease of some particular individual components of $W$ is allowed at some iterations, so that poor initial support points can be removed. Even in this case, if $\ell$ is large many iterations are required to identify the components of $W$ which should be positive at the optimum. Gradient-type methods are also not efficient when $\ell$ is large.

Several authors tried combinations of different methods to make use of their respective advantages. A very sensible algorithm is proposed in Wu (1978a,b); it combines a gradient method (always working in a small dimensional subspace) with a vertex-direction method (which allows a suitable updating of this subspace). A mixture of multiplicative and vertex-exchange algorithms is proposed in Yu (2011) for $D$-optimum design; it includes a nearest-neighbor exchange strategy which helps in apportioning weights between adjacent points in $\mathcal{X}$ and has the property that poor support points are quickly removed from the support of the initial measure. Attractive performance is reported. All the methods above, however, are restricted to the case where the design space $\mathcal{X}$ is finite with $l$ being not too large and, for some of them, to particular design criteria ($D$-optimality, for instance).

When one is interested in the determination of an optimal design on a compact subset $\mathcal{X}$ of $\mathbb{R}^d$ with nonempty interior, the usual practice consists in discretizing $\mathcal{X}$ into a finite set $\mathcal{X}_\ell$ with $\ell$ elements and applying one of the methods above. When a precise solution is required, then $\ell$ is necessarily very large (in some cases, $\ell = 10^6$ should be considered as a small number) and none of the methods above is efficient. Refining iteratively a finite grid contained in $\mathcal{X}$ is a possible option, see Wu (1978a), but the search for the optimal design is still performed in a discrete set.

In contrast, the algorithm we propose makes use of the compactness of $\mathcal{X}$ and looks for the optimal support in the whole set $\mathcal{X}$. A finite grid $\mathcal{X}_\ell \subset \mathcal{X}$ is only used to check optimality on $\mathcal{X}_\ell$ and, using the Lipschitz continuity of directional derivatives of $\phi(\cdot)$, to construct an efficiency bound on $\mathcal{X}$. A key ingredient in the algorithm is the separation between the determination of the support points of $\xi^*$ (knowing that there are at most $m$ of them) from the determination of the associated weights (an $m$-dimensional convex problem).

The determination of the support of $\xi^*$ is a non-convex problem, usually multimodal, for which the straightforward application of a global search method (such as the simulated annealing or one of the genetic algorithms) cannot be recommended. Indeed, these heuristic methods do not use the crucial information about the objective function provided by the Lipschitz constants, derivatives and convexity of the weight-optimization problem, and most of them do not provide any indication of the closeness of the returned solution to an optimum, global or local. On the other hand, by using properties of the directional derivative of $\phi(\cdot)$, we can easily locate good candidates for the support of $\xi^*$ and thereby we do not need to perform a global search in the $m \times d$ dimensional space $\mathcal{X}^d$. The aim of this paper is to show how these properties can be combined with algorithms for convex optimization which are classically used in the design for the determination of optimal weights to yield an efficient method of construction of optimal designs on compact sets.

The paper is organized as follows. Section 2 defines the problem and introduces the notation. Section 3 presents the algorithm and proves its convergence. A few illustrative examples are given in Section 4 where the results obtained with the proposed algorithm are discussed. Section 5 concludes and indicates some remaining open issues. A few technical aspects are collected in the Appendix.

## 2. Notation and problem statement

Let the design space $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$ with nonempty interior (typically $\mathcal{X} = [-1, 1]^d$) and denote by $\Xi(\mathcal{X})$ the set of probability measures on $\mathcal{X}$. Any element $\xi \in \Xi(\mathcal{X})$ will be called design measure, or shortly design.