Experimental design

Friedrich Pukelsheim, A. Wilhelm

F. PUKELSHEIM and A. WILHELM

Universität Augsburg

The authors are to be applauded for emphasizing in their paper the common structure of many instances of the design problem. The good news is that this unifying view is worked out in detail, to an extend yet greater than outlined in the present paper, in the monograph Pukelsheim (1993). There some aspects of the theory are seen in

a different light, and it may be worthwhile to comment on the peculiarities of the two approaches.

1.1 In their display (2), the authors define the design problem as one of *minimizing* a functional Ψ of the information matrix (moment matrix) $M(\xi)$. This is a very formal, mathematical way of stating the problem. The statistical notion of information would call for a maximization, in that an experiment ought to be planned in such a way that the information obtained is as large as possible, while the variability in the data is as small as possible. Hence if goodness of a design is characterized through something like a variance-covariance matrix, minimization would be in order. The idea of choosing the experimental design so as to *maximize* the information in the experiment is the underlying theme in Pukelsheim (1993).

Hence an alternative way of paraphrasing the optimal design problem is to *maximize* a real-valued function $\psi(M(\xi))$.

1.2 Many criteria ψ that are of statistical interest factorize,

$$\psi = \phi \circ C_K,$$

where the information matrix mapping C_K is discussed in detail in Chapter 3 of Pukelsheim (1993). The information matrix mapping yields matrices of the order $s \times s$, for an s-dimensional parameter system of interest given by $K'\theta$, while the larger, original $m \times m$ moment matrix M also reflects the information on the nuisance parameters. It is very helpful to discuss the distinction between parameters of interest and nuisance parameters explicitly, through a composition $\phi \circ C_K$, rather than bury this important statistical concept in a single criterion ψ.

The information matrix mapping is positively homogeneous, superadditive, non-negative, nonconstant, and upper semicontinuous. It is quite natural to demand that the optimality criterion ϕ enjoys the same properties. This leads to the class of information functions of Chapter 5 in Pukelsheim (1993); a statistical interpretation of these properties is given by Pukelsheim (1987). Thus it is *statistical considerations* that define the class of feasible optimality criteria.

In contrast, the present authors choose a more formal, mathematical approach in referring to assumptions $a, b, c, d, b', c', c'', e_0, b^1, d^1$. This hides the statistical issues at the expense of unimportant formal details. The origin for this odd emphasis is the authors' condition (e), of which the full awkwardness would be more visible if the existence of the function $\psi(x, \xi)$ and $\tau(\alpha, \xi, \bar{\xi})$ and the implied convergence as α tends to zero (uniformly in $\xi, \bar{\xi}$?) were properly stated.—A page reference to Ermakov (1983) on how to avoid assumptions (1) and (b) would be appreciated.

Part 1 of Theorem 1 states a bound on the number of support points. A tighter bound has been given by Fellman (1974), Pukelsheim (1980), Chaloner (1984).—Part 4 should read that $\psi(x, \xi^*)$ achieves zero in $\mathrm{supp}\,\xi^*$; the almost everywhere statement refers to the full set X.

1.3 In many applications one seeks to find an optimal design, not within the class of *all* designs, but rather in a restricted subset of competing designs. Since the proper optimization problem refers to the moment matrices $M(\xi)$, and not to the designs ξ themselves, the transition is from the set of *all* moment matrices, to a subset $\mathcal{M}$ of competing moment matrices.

All that the general principal of convex optimization theory require is that the set of competing moment matrices $\mathcal{M}$ is compact and convex, and this is the level on which the design problem is solved in Pukelsheim (1993). There, in Chapter 11, it is illustrated that this approach encompasses such instances as Bayes designs, designs with bounded weights, designs for mixtures of models, designs for mixtures of criteria, and designs with guaranteed efficiencies.

Linear constraints are one way to delineate a specific set $\mathcal{M}$ of competing moment matrices that is compact and convex.

1.4 Nonlinear convex constraints are another way of defining a subset of competing moment matrices $\mathcal{M}$ that is convex and compact. The present Theorem 3 is based on directional derivatives; this concept runs into great difficulties when the criterion function is not differentiable, or when the optimal moment matrix is singular. The counterpart is Theorem 11.20 in Pukelsheim (1993), based on subgradient calculus and also covering the nondifferentiable boundary cases.

2. The numerical challenge of the optimal design problem stems from the fact that the optimum usually is very flat. Hence small perturbations of the support points or the weights do not change the objective function in any significant manner. This is reassuring from a practical point of view, in that the experimenter can adjust the design a little bit without giving away too much in terms of any optimality criterion. However it renders most numerical procedures inefficient since they proceed to the optimum only very slowly.

The authors restrict attention to the family of numerical algorithms that come with a differentiable optimization problem. The general idea is to use directional derivatives to find a direction of improvement, and then determine an optimal steplength. For the design problem special issues arise in deleting support points, or adding new ones. An overview over existing methods and a unifying approach to them is presented in Gaffke and Mathar (1992).

A solution to this dilemma seems to be provided by the more recent advances in nondifferentiable optimization methods. They generally rely on subgradients rather than directional derivatives. However, they not only use subgradient information at the actual point of iteration, but also the information which has accumulated at previous iteration points. Appropriate versions of these algorithms are proposed by Schramm and Zowe (1988), they are called *bundle trust methods.*

Bundle trust methods merge the bundle concept from nondifferentiable optimization with some features of the trust region approach from differentiable optimization. They avoid the usual line search by using a cutting plane model with an additional trust region term to compute the next iterate as the solution of a quadratic program. The version implemented includes linear constraints and box constraints. General non-linear constraints can be included by a penalty approach.

These methods are designed to solve optimization problems that are convex or nonconvex, constraint or unconstraint, differentiable or nondifferentiable Differenti-ability is replaced by *Lipschitz continuity* of the objective function, and knowledge of any one subgradient in each iteration point. The complete subgradient calculus for the design problem is included in Pukelsheim (1993). In the nonconvex case, the concept of subgradients is replaced by the concept of *generalized gradients* in the sense of Clarke

(1983). An extensive discussion of Lipschitz continuity and the generalized gradient calculus for the design problem are given in Wilhelm (1993a, b).

Affirmative results on the convergence of the bundle trust methods for constraint problems are available only in the convex case. These theoretical shortcomings of the bundle trust methods are more than compensated by the numerical results obtained with the program OPTDES of Wilhelm (1993c). The OPTDES code is a bundle trust method that is adapted to the design problem. The numerical performance of OPTDES is quite impressive, especially for the E-criterion which fails to be differentiable, and the T-criterion where the optimal moment matrices tend to be singular.

There are two other reasons to rely on nonconvex optimization techniques. Firstly, a design is defined through its support points, and its weights. Theorems 1 and 2 of the present paper state that a finite number of support points suffices, whence the problem does remain finite dimensional. Hence there is a fixed number of weights, but at any round of iteration some of them may vanish. The weight vector then comes to lie on the boundary of the probability simplex where differentiability breaks down, compare Fellman (1980).

A practicing statistician usually wants as few support points as possible. The numerical results with OPTDES indicate that the implemented algorithm tends to produce designs with a minimum number of support points, even if the support size of the starting design uses the excessive bound of Theorems 1 and 2.

Secondly, as a function of the support points and the weights, the design problem seizes to be a concave maximization problem. The reason is that the moment matrix is a *convex* function of the support points, which then is composed with a *concave* optimality criterion. For the composition of a concave function with a convex function, global properties such as convexity or concavity cannot generally be asserted, while the local smoothness property of Lipschitz continuity still holds true.

We acknowledge that traditionally the numerical handling of the design problem relies on directional derivatives and line searches. For the reasons and the evidence given above, we find it more profitable to invoke nondifferentiable, nonconvex optimization techniques.

References

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12**, 283–300. "Correction." *Ibidem* **13**, 836.

Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis.* Wiley, New York.

Fellman, J. (1974). On the allocation of linear observations. *Societas Scientiarum Fennica, Commentationes Physico-Mathematicae* **44**, 27–78.

Fellman, J. (1980). On the behaviour of the optimality criterion in the neighborhood of the optimal point. Swedish School of Economics and Business Administration, Helsinki. Working Paper 49, 15 pages.

Gaffke, N. and Mathar, R. (1992). On a class of algorithms from experimental design theory. *Optimization* **24**, 91–126.

Pukelsheim, F. (1980). On linear regression designs which maximize information. *J. Statist. Inference* **4**, 339–364.

Pukelsheim, F. (1987). Information increasing orderings in experimental design theory. *Intern. Statist. Review* **55**, 203–219.

Pukelsheim, F. (1983). *Optimal Design of Experiments.* Wiley, New York.

Schramm, H. and Zowe, J. (1988). A combination of the bundle approach and the trust region concept. In *Mathematical Research 45, Advances in Mathematical Optimization.* (Eds. J. Guddat, B. Blank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, K. Tammer, M. Vlach, K. Zimmermann), Akademie Verlag, Berlin, 196–209.

Wilhelm, A. (1993a). *Optimale Versuchspläne bei differenzierbaren und nichtdifferenzierbaren Zielfunktionen: Konzepte und Algorithmen.* Shaker Verlag, Aachen.

Wilhelm, A. (1993b). Subdifferentiability and Lipschitz continuity in optimal design problems. *In preparation.*

Wilhelm, A. (1993c). OPTDES—Numerical computation of optimal experimental designs using bundle trust methods. *In preparation.*