Optimal Design Construction With Constraints II.

Valery Fedorov
School of Statistics, University of Minnesota*

Technical Report #575

July, 1992

* Valery Fedorov is a visiting professor at the University of Minnesota.

I. Introduction

This paper concerns various numerical procedures which can be used for optimal design construction with constraints. It can be considered as a natural extension of Fedorov (1992), where the main properties and characterization of optimal designs were considered.

All notation coincides with that used in the above mentioned paper. One can find there basic statements and comments on the considered optimization problems.

Section II surveys, very briefly, the basic results for the case without constraints. In Section III the straightforward generalization of these results is considered and discussed. The connection between the optimal design problem with constraints and the optimal design problem for the weighted sum of various standard optimality measures is studied in Section IV. Section V deals with optimal design in the presence of nonlinear constraints.

II. Numerical procedures in the standard case

The main objective of this section is to discuss numerical methods for finding a solution to the following optimization problem

$$\xi^* = \text{Arg}\min_{\xi} \Psi(\xi), \tag{1}$$

where $supp\ \xi \subset X$. We shall use notation $\Psi(\xi)$ for $\Psi[M(\xi)]$, Ψ^* for $\Psi(\xi^*)$ and $\min_x$, $\min_\xi$, $\int$, and so on, instead of $\min_{x \in X}$, $\min_{\xi \in \Xi}$, $\int_X$, respectively, if it does not lead to ambiguity.

For convenience let the basic assumptions from Fedorov (1992) be reproduced here:

(a) X is compact;
(b) f(x) are continuous function in X, $f \in R^m$;
(c) $\Psi(M)$ is a convex function;
(d) there exists q such that

$$\{\xi : \Psi[M(\xi)] \le q \le \infty\} = \Xi\,(q) = \emptyset;$$

(e) for any $\xi \in \Xi(q)$ and $\bar{\xi} \in \Xi$:

$$\Psi[(1-\alpha)M(\xi) + \alpha M(\bar{\xi})] = \Psi[M(\xi)] + \alpha \int \psi(x,\xi)\bar{\xi}(dx), + \tau[\xi(\alpha)],$$

where $\tau[\xi(\alpha)] \le \alpha^2 K_q$, $K_q > 0$.

Notice that assumption (e) is more restrictive than in Fedorov (1992), where $\tau[\xi(\alpha)] = o(\alpha)$.

We shall consider the iterative algorithms of the following type:

$$\xi_{s+1} = (1-\alpha_s)\,\xi + \alpha_s\,\xi^s\,, \tag{2}$$

where

$$\xi^s = \operatorname{Arg}\min_{\xi}\ \Psi[(1-\alpha_s)\xi_s + \alpha\,\xi]. \tag{3}$$

When the first order approximation (see assumption (e)):

$$\xi^s = \operatorname{Arg}\min_{\xi}\ \{\Psi(\xi_s) + \alpha \int \psi(x, \xi_s)\ \xi(dx)\} \tag{4}$$

is used in place of (3), then the corresponding algorithm is called "the first order algorithm".

The simplicity and popularity of the first order algorithms is explained by the fact that a design ξ^s does not depend upon α and

$$\xi^s = \operatorname{Arg}\min_{x}\ \int \psi(x, \xi_s)\ \xi(dx). \tag{5}$$

Optimization problem (5) is linear with respect to ξ and is therefore easier than the original problem (3). Moreover, when there are no constraints imposed on ξ then

$$\xi^s = \xi(x_s),\ x_s = \operatorname{Arg}\min_{x}\ \psi\ (x, \xi_s), \tag{6}$$

where $\xi(x_s)$ is a design measure completely atomized at point x_s.

Approximation (5), (6) allows us to develop a number of first order algorithms. All of them can be imbedded in the following scheme:

(1) There is a design $\xi_s \in \Xi_q$. Find

$$x_s = \operatorname{Arg}\min_{x} \psi(x, \xi_s).$$

(2) Choose $0 \le \alpha_s \le 1$ and construct

$$\xi_{s+1} = (1-\alpha_s)\ \xi_s + \alpha_s\ \xi(x_s).$$

The choice of a sequence $\{\alpha_s\}$ defines a variety of the algorithms (see Cook & Nachtsheim (1989), Fedorov (1972, 1975), Silvey (1980)).

The following three sequences $\{\alpha_s\}$ are most popular (especially in the theoretical considerations):

(1) $\lim\limits_{s\to\infty} \alpha_s = 0, \quad \sum\limits_{s=0}^{\infty} \alpha_s = \infty;$

(2) $\alpha_s = \operatorname{Arg}\min\limits_{x} \Psi[M(\xi_{s+1}(\alpha))]$, where $\xi_{s+1}(\alpha) = (1-\alpha)\,\xi_s + \alpha\,\xi(x_s)$

(3) $\alpha_s = \begin{cases} \alpha_{s-1} & \text{if } \Psi[\xi_s(\alpha_{s-1})] < \Psi(\xi_{s-1}) \\ \alpha_{s-1}/\gamma \quad \gamma>1 & \text{otherwise.} \end{cases}$

Proving the convergence of the algorithm corresponding to any of these sequences is a rather standard thing (see Fedorov (1975, 1986), Wu & Wynn (1978)) for the optimization theory.

For instance, let us prove the (weak) convergence of the algorithm for $\{\alpha_s\}$ defined by rule (2).

Lemma 1. If (a) – (e) hold and $\xi, \bar{\xi} \in \Xi_q$,

then

$$\min_x \psi(x, \xi) \le \int \psi(x,\xi)\bar{\xi}(dx) \le \Psi[M(\bar{\xi})] - \Psi[M(\xi)].$$

Proof. The result is a corollary of the convexity of $\Psi(M)$.

Theorem 1. If (a) – (e) hold then for $\{\alpha_s\}$ defined by rule (2) the iterative procedure (1), (2) converges:

$$\lim_{s\to\infty} \Psi(\xi_s) = \Psi^*. \qquad (7)$$

Proof. By the definition $\{\Psi(\xi_s)\}$ is monotonously decreasing and therefore it converges:

$$\lim_{s\to\infty} \Psi(\xi_s) = \Psi_* \qquad (8)$$

Assume that Ψ_* is not the optimal value of the objective function. Then

$$\Psi_* - \Psi^* = \delta > 0.$$

It means that for any s

$$\min_x \psi(x,\xi_s) \le -\delta. \qquad (9)$$

Rule (2) together with assumption (e) provide that

$$\Psi(\xi_s) - \Psi(\xi_{s+1}) \geq \delta/4K_q > 0$$

and subsequently

$$\lim_{s \to \infty} \Psi(\xi_s) = \infty .$$

which contradicts (8), proving (7).

Iterative procedure (1), (2) with any rule (1) – (3) guarantees the convergence of the corresponding algorithms, but they are rather slow and serve more to the theory than to the practice.

A significant improvement in the rate of convergence is realized, when the iterative procedure (1), (2) is modified in the following way (see Atwood (1973), Fedorov (1975)):

$$x_s = \text{Arg min } [\psi(x_s^+, \xi_s), - \psi(x_s^-, \xi_s)], \qquad (10)$$

where

$$x_s^+ = \text{Arg} \min_{x \in X} \psi(x, \xi_s) \text{ and } x_s^- = \text{Arg} \min_{x \in X_s} \psi(x, \xi_s)\ , \ X_s = \text{supp } \xi_s,$$

and

$$\alpha_s = \begin{cases} \gamma_s & x_s = x_s^+ \\ -\min[\gamma_s,\ p_{si} / (1-p_{si})] \end{cases} \qquad (11)$$

where $\{\gamma_s\}$ has to obey one of the rules (1) – (3) and p_{si} is a measure of a point $x_{si} \in X_s$.

The iterative procedure (1), (2) & (10), (11) is usually complemented by some clustering rule, for instance:
all supporting points $x_i \in X_{(s-1)}$ satisfying the inequality

$$\| x_i - x_s^+ \| \leq d \qquad (12)$$

have to be merged with the point x_s^+, i.e. all corresponding measures have to be transferred to x_s^+.

One can introduce "forward" and "backward" excursions when correspondingly n^+ steps operating with "new" points x_s^+ and n^- steps handling points x_s^- are doing subsequently; the numbers n^+ and n^- are called a "length" of an excursion (see Mitchell (1974)).

Computations became simpler when in (10) one looks for x_s^+ which is only an approximate solution of the corresponding optimization problem, i.e.

$$\psi(x_s^+, \xi_s) = \min_x \psi(x, \xi_s) + \delta_s , \qquad (13)$$

where $\lim_{s\to\infty} \delta_s = 0$.

All these changes or improvements do not destroy the convergence of the iterative procedure.

III. Linear constraints. Direct first order algorithm.

As it was discussed in Fedorov (1992) it is quite natural in experimental practice to search for an optimal design under constraints:

$$\xi^* = \text{Arg} \min_\xi \Psi(\xi) \ , \qquad (14)$$

$$\text{s.t. } \int \zeta(x)\xi(dx) \leq c \ , \ \zeta \in R^p.$$

Assumptions (a) – (e) have to be slightly modified. To assumption (b) one has to add the continuity of $\zeta(x)$ (it will be referred to as assumption (b^1). Assumption (d) has to be replaced by assumption (d^1): there exists q such that

$$\{\xi: \ \Psi(\xi) \leq q < \infty, \int \zeta(x)\xi(dx) \leq c\} = \Xi(q) \neq \emptyset.$$

An iterative procedure that is very similar to procedure (1), (2) (with all improvements discussed in Section II if one needs them) can be used to construct ξ^*:

(1c) There is a design $\xi_s \in \Xi_q$. Find

$$\xi^s = \text{Arg} \min_\xi \int \psi(x,\xi_s) \ \xi(dx), \qquad (15)$$

$$\text{s.t. } \int \zeta(x)\xi(dx) \leq c.$$

(2c) Choose $0 \leq \alpha_s \leq 1$ and construct

$$\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s \ \xi^s \qquad (16)$$

It is important to point out that the transition from (5) to (6) is not valid now. Therefore, one has to work with (15) or (5), and that is much more complicated than to work with (6). In (15) the optimization has to be made in the space of probability measures ξ, while in (6) one looks for a minimum over X.

Assume also that problem (15) can be solved. Then one can use all results of the previous section starting from Lemma 1, and the following theorem results:

Theorem 2. If (a) – (e), (b^1), (d^1) hold then for $\{\alpha_s\}$ defined by rule (2) the iterative procedure (1c), (2c) converges:

$$\lim_{s\to\infty} \Psi[M(\xi_s)] = \Psi^* .$$

The proof is identical to the proof of Theorem 1 with the obvious substitution

$$\min_x \psi(x, \xi_s) \text{ by } \int\psi(x,\xi_s)\xi^s(dx).$$

Thus one does not face any difficulties in optimal design construction if optimization problem (15) can be easily solved. Unfortunately, it is often not the case in practice. Some simplification can be obtained via Note 1, Section III from Fedorov (1992). Then (15) can be reduced to a finite dimensional problem:

$$\xi^s = \begin{Bmatrix} x_1^s \dots x^s_{l+1} \\ p_1^s \dots p^s_{l+1} \end{Bmatrix} = \text{Arg} \min_{p_j, x_j} \sum_{j=1}^{l+1} p_j \psi(x_j, \xi_s) \tag{17}$$

$$\text{s.t.} \quad \sum_{j=1}^{l+1} p_j \zeta(x_j) \le c .$$

This problem can be considered as practically solvable when the number of constraints and the dimension of X is reasonably small.. Some numerical methods to solve (17) based on the cutting–plane technique are considered by Gaivoronski (1986). They frequently assume extra iterating within every s–th iteration, and the algorithms become impractical.

Obviously some corrections or amendments can be made to simplify the iterative procedure (1c), (2c).

For instance, when an initial design ξ_0 is inside of Ξ_q, i.e.

$$\int\zeta(x)\xi(dx) < c - \delta, \quad \delta_\alpha > 0, \alpha = 1, \dots, l \tag{18}$$

then the iterations (1), (2) defined in the previous section can be used. Naturally one has continuously to verify that (18) is satisfied.

The convergence of the iterative procedure takes still place when instead of solution of (15) or (17) one would look for any design ξ^s such that

$$\int \psi(x,\xi_s)\xi^s(dx) \geq \gamma_s > 0 \quad \text{and} \quad \int \zeta(x)\xi^s(dx) \leq c, \tag{19}$$

where $\{\gamma_s\}$ some diminishing sequence.

The equivalence theorem (see Th. 2, Fedorov 1992) contains a hint for another modification of the iterative procedure. Particularly part 4 of this theorem states that the dual function

$$q(x, u, \xi) = \psi(x, \xi) + u^T\zeta(x) - u^Tc$$

achieves zero almost everywhere in supp ξ^*.

Therefore at stage (1c) one can do the following:

(1c') There is a design $\xi_s \in \Xi_q$. Let supp $\xi_s = \{x_1{}^s, ..., x_{n_s}{}^s\}$.

Step (a). Solve the linear programming:

$$u_s = \operatorname{Arg}\max_{\underline{u}} \; u_{l+1}, \tag{20}$$

$$\psi(x_i^s, \xi_s) + u^T\zeta(x_i^s) - u^Tc - u_{l+1} \geq 0, \quad i=1, ..., n_s$$

$$\underline{u}^T = (u^T, u_{c+1}), \quad u_\alpha > 0, \quad \alpha = 1, ..., l \;;$$

Step (b). Find

$$\xi^s = \xi(x_s) = \operatorname{Arg}\min_x \; q(x, u_s, \xi_s).$$

IV. Linear constraints. Lagrangian approach.

On an intuitive level it is clear that minimizing function

$$\Phi(\xi) = \Psi(\xi) + \lambda^T Z(\xi), \tag{21}$$

where $\lambda_\alpha > 0$, $\alpha = 1, ..., l$ and $Z(\xi) = \int \zeta(x)\xi(dx)$, one guarantees that $Z(\xi) \leq c_\lambda$.

The problem consists of finding a vector λ such that c_λ is close to the vector c defined in (14). The following theorem can be useful for this purpose.

Theorem 3. If (a) – (e), (b') hold, then a design:

$$\xi_\lambda = \operatorname{Arg}\min \{\Psi(\xi) + \lambda^T Z(\xi)\} \tag{22}$$

is a solution of the constrained optimization problem (14) with

$$c = c_\lambda = Z(\xi_\lambda).$$

Proof. Let $\bar{\xi} = (1-\alpha)\,\xi + \alpha\,\bar{\xi}$ where $\xi \in \Xi(q)$ and $\bar{\xi} \in \Xi$ (see assumption (e), Section II). It is easy to check that

$$\partial\,\Phi\,(\bar{\xi})/\partial\,\alpha = \int \psi(x, \xi)\,\bar{\xi}(dx) + \lambda^T\,[\,\zeta(x) - Z(\xi)] = \int \tilde{\psi}\,(x, \lambda, \xi)\,\bar{\xi}(dx)$$

and therefore assumption (e) takes place for $\Phi\,(\xi)$ if it does for $\Psi(\xi)$. Similarly one can check the validity of all assumptions (a) – (d) for $\Phi\,(\xi)$ if they are valid for $\Psi\,(\xi)$. Thus all the conditions of the equivalence theorem for the unconstrained case (see Fedorov, 1992) hold.

Thus the necessary and sufficient condition for a design ξ_λ to be optimal (i.e. providing minimum for $\Phi(\xi)$) is fulfillment of the inequality

$$\min_x \tilde{\psi}\,(x, \lambda, \xi_\lambda) \geq 0. \qquad (23)$$

But $\tilde{\psi}\,(x, \lambda, \xi_\lambda)$ coincides with the function $q(x, u, \xi_\lambda)$ defined in Theorem 2 from Fedorov (1992) if one sets $u = \lambda$ and $\phi(x) = \zeta(x) - c_\lambda$, $c_\lambda = Z(\xi_\lambda)$. Thus it follows that

$$\min_x \max_u \{\, \psi(x, \xi_\lambda) + u^T\,[\zeta(x) - c_\lambda]\,\} \geq$$

$$\geq \min_x \{\, \psi(x, \xi_\lambda) + \lambda[\zeta(x) - c_\lambda]\} \geq 0 \qquad (24)$$

This inequality is sufficient to assert that

$$\xi_\lambda = \operatorname{Arg}\min_\xi \Psi(\xi),$$

$$\text{s.t.} \int \zeta(x)\xi(dx) \leq c_\lambda\,,$$

and it completes the proof.

Theorem 3 provides a solution for (14) if one can determine the vector λ such that $Z(\xi_\lambda) \leq c$.

Because only the simplest problems can be solved analytically (see Cook & Wong, (1992), where the $\lambda \in R^1$ was considered in detail), the following numerical approach looks very promising for most applications.

It will be assumed that for any given λ, problem (22) can be numerically solved at no signficant expense. The likelihood of this assumption is confirmed by the experience reported by various authors (see Fedorov (1975), Mitchell (1974), Nachtsheim (1987)) who had experimented with numerous modifications of the algorithms considered in section II. Additionally we shall assume that a practitioner will be satisfied if the vector $c_\lambda = Z(\xi_\lambda)$ would be "close" to c (but not exactly equal), for instance, a squared distance

$$y_\lambda = (c_\lambda - c)^T A(c_\lambda - c) \ , \ A > 0$$

would be small.

Thus, one can consider the "empirical" optimization problem:

$$\lambda^* = \operatorname{Arg} \min_{\lambda} (c_\lambda - c)^T A(c_\lambda - c) \ , \tag{26}$$

$$\text{s.t. } \lambda_\alpha \geq 0 \ , \ \ \alpha = 1, \ldots, l \ ,$$

Considering y_λ as a response function depending upon λ, one can apply to the empirical optimization technique which is the famous milestone in the experimental design.

V. Nonlinear constraints

All the basic results of sections III and IV can be generalized for the case with nonlinear constraints (see Section IV, Fedorov, (1992)):

$$\xi^* = \operatorname{Arg} \min_{\xi} \Psi[M(\xi)] \ , \tag{27}$$

$$\text{s.t. } \Phi(\xi) \leq 0 \ , \ \ \Phi \in R^l .$$

It is assumed additionally to (a) – (e), (b') that

(c') $\Phi(\xi)$ are convex functions,

(e') $\Phi[\bar{\xi}(\alpha)] = \Phi(\xi) + \alpha \int \phi(x, \xi) \bar{\xi}(dx) + \gamma[\underline{\xi}(\alpha)]$,

where $\underline{\xi}(\alpha) = (1 - \alpha)\xi + \alpha \bar{\xi}$, $\xi \in \Xi_q$, $\underline{\xi} \in \Xi$,

and $\gamma[\underline{\xi}(\alpha)] \leq \alpha^2 Q_q$, $Q_q > 0$ uniformly for all $\underline{\xi}(\alpha) \in \Xi(q)$.

By linearizing $\Phi(\xi)$ in the "vicinity" of optimal design (see Section IV, Fedorov, 1992) one can check that all results of Section III stay valid if the vector $\phi(x, \xi)$ substitutes everywhere $\zeta(x) - c$ is replaced by $\phi(x, \xi)$ throughout.

The approach proposed in Section IV looks less elegant from a mathematical point of view but it is more practical. In this case instead of problem (27) one has to consider as a sequence of optimization problems defined by (26) for a "compound" criteria of optimality (we use the terminology introduced in Cook & Wong (1992)).

$$\xi_\lambda = \mathrm{Arg}\min_{\xi} \{\Psi(\xi) + \lambda^T \Phi(\xi)\}\,, \tag{28}$$

$$\lambda_\alpha > 0,\ \alpha = 1, \ldots, l\,,$$

where very frequently $\Phi(\xi)$ is a vector of auxiliary optimality criteria (see Section IV, Fedorov (1992), Cook & Weng (1992)).

Together with the equivalence theorem (Theorem 3, Fedorov (1992)) for infinitely many constraints the approach from Section IV gives a new insight into the problem of optimal experimental design for nonlinear models.

Let

$$M(\xi, \theta) = \int f(x, \theta)\, f^T(x, \theta)\, \xi(dx),$$

where $f(x, \theta) = \partial\, \eta\, (x, \theta)/\partial\, \theta$ and $\theta \in \Omega \subset R^l$.

If one wishes to ensure that for any θ some optimality criteria does not exceed a given level, then the following optimization problem can be considered:

$$\xi^* = \mathrm{Arg}\min_{\xi} \Phi[M(\xi, \theta_o)]\,, \tag{29}$$

$$\text{s.t.}\quad \Phi[M(\xi, \theta)] \le c \qquad \text{for all } \theta \in \Omega,$$

where vector θ_o could be, for instance, a prior estimate of the values of the parameters to be estimated.

From Theorem 3, (Fedorov 1992) it follows that there exist $(l+1)$ vectors θ_k^* such that (29) can be reduced to the following finite dimensional problem:

$$\xi^* = \mathrm{Arg}\min_{\xi} \Phi[M(\xi, \theta_0)], \tag{30}$$

$$\text{s.t.}\quad \phi[M(\xi, \theta_k^*)] \le c\,,\quad k=1, \ldots, l+1$$

Theorem 3 tells us that the "compound" optimization problem

$$\xi_\lambda^* = \text{Arg}\min_\xi \{\Phi[M(\xi, \theta_o)] + \sum_{k=1}^{l+1} \lambda_k \ \Phi[M(\xi, \theta_k^*)]\} \tag{31}$$

is closely related to (30). If one knows θ_k^* then all the results of section IV can be used without any changes. Unfortunately θ_k^* must often be searched for and the author failed to find any reasonable numerical procedure to do this. On the common sense level it is clear that θ_k^* ought to be some "worst" point in Ω, corresponding to a "smallest" information matrix.

If there is no prior preference to any point from Ω instead of (30) one may consider another very similar problem:

$$\xi^* = \text{Arg} \min_\xi c\ , \tag{32}$$

$$\text{s.t.} \quad \phi[M(\xi, \theta_k^*)] \le c, \ k=1, \ldots, l+1,$$

which, or course, can be considered as a particular case of (27) and therefore all results related to (30) remain valid. For instance, (32) can be solved by multiple application to the solution of the "compound" optimization problem:

$$\xi^* = \text{Arg}\min_\xi \sum_{k=1}^{l+1} \lambda_k \ \Phi[M(\xi, \theta_k^*)]\ , \tag{33}$$

where again no strict recipes how to find out set of θ_k^*, k=1, ..., l+1 are known to the authors.

We can conclude this section mentioning that (32) is equivalent to the following minimax problem:

$$\xi^* = \text{Arg}\min_\xi \max_{\theta\subset\Omega} \Phi[M(\xi, \theta_k)] \tag{34}$$

Acknowledgement:

My long discussions with Dennis Cook have triggered the main results of sections IV and V. The constructive critiques of Chris Nachtsheim contributed in clarifying some statements and made the manuscript more readable.

REFERENCES

Atwood, C.L. (1973) Sequences Converging to D–optimal Design of Experiments, *Ann. Stat.* 1, 342–352.

Cook, R.D. & Nachtsheim, C.J. (1989) Computer–aided Blocking for Factorial and Response–Surface Designs, *Technometrics* 31, 339–346.

Cook, R.D. & Wong, W.K. (1992). On the equivalence of constrained and compound optimal designs. Submitted for publication.

Fedorov, V.V. (1992) Optimal Design Construction with Constraints I, Technical Report No. 573, University of Minnesota, p 11.

Fedorov, V.V. & Uspensky, A.B. (1975) Numerical Aspects of Design and Analysis of Experiments, Moscow State University, Moscow, pp 167.

Fedorov, V.V. (1986) Optimal Design of Experiments: Numerical Methods, WP–86–55, IIASA, Laxenburg, Austria, p 10.

Gaivoronski, A. (1986) Linearization Methods of Optimization of Functionals which Depend on Probability Measures. *Mathematical Programming Study*, 28, 157–181.

Lauter, E. (1976) Optimal Multipurpose Designs for Regression Models. *Math. Operations for Sch. u. Statist.*, 7, 51–68.

Mitchell, T. J. (1974) An Algorithm for the Construction of D–optimal Experimental Designs, *Technometrics*, 16, 203–211.

Nachtsheim, C.J. (1987) Tools for Computer–Aided Design of Experiments, *Journal of Quality Technology*, 19, 132–160.

Wu, C.F. & Wynn, H. (1978) The Convergence of General Step–Length Algorithms for Regular Optimum Design Criteria, *Ann. Statist.*, 6, 1273–1285.