

Submodularity and Randomized rounding techniques for Optimal Experimental Design

Mustapha Bouhtou**, Stéphane Gaubert*, Guillaume Sagnol*

*INRIA Saclay & CMAP, Ecole Polytechnique, Palaiseau, France

**Orange Labs, Issy-les-Moulineaux, France

Emails: mustapha.bouhtou@orange-ftgroup.com
{guillaume.sagnol, stephane.gaubert}@inria.fr

Abstract

We review recent results obtained by the authors on the approximability of a family of combinatorial problems arising in *optimal experimental design*. We first recall a result based on submodularity, which states that the greedy approach always gives a design within $1 - 1/e$ of the optimal solution. Then, we present a new result on the design found by rounding the solution of the continuous relaxed problem, an approach which has been applied by several authors: When the goal is to select n out of s experiments, the D -optimal design may be rounded to a design for which the dimension of the observable subspace is within $\frac{n}{s}$ of the optimum.

Keywords: Optimal design of experiments, Kiefer's p -criterion, D -optimality, Polynomial-time approximability, Discrete optimization.

1 Introduction and Statement of the problem

The theory of *optimal design of experiments* plays a central role in statistics. It explains how to best select experiments in order to estimate a set of parameters. In this paper, we assume that each experiment allows one to measure the value of one or several linear combinations of the parameters. This setting is known in the literature as the optimal design of *multiresponse experiments*. For more details on the subject, the reader is referred to the monographs of

Fedorov[6] and Pukelsheim[11].

In a number of real-world applications, the design variables are discrete, since the experimenter can only choose the experiments to conduct from a finite set, and eventually how many times to perform them. The resulting problem is NP hard [13], and several authors proposed to use heuristics, such as the greedy algorithm [13,14] or roundings of the continuous solution [3,15].

The purpose of this article is to investigate the legitimacy of these heuristic techniques. We will briefly recall in Section 2 the results announced by the authors on the greedy algorithm in [13]. Then, we give in Section 3 a new result concerning the rounding algorithm for D -optimal designs. The proofs of all results can be found in the companion paper of the third author [12].

We denote vectors by bold-face lowercase letters and use capital letters to denote matrices. We set $[s] := \{1, \dots, s\}$. We denote by $\boldsymbol{\theta} \in \mathbb{R}^m$ the vector of the parameters that we want to estimate. In accordance with the classical linear regression model, we assume that the experimenter disposes of a collection of s experiments, each one providing a (multidimensional) observation which is a linear function of the parameters, up to a noise on the measurement (the variance of which is set to 1 for simplicity sake). In other words, for each experiment $i \in [s]$, we have

$$\mathbf{y}_i = A_i \boldsymbol{\theta} + \boldsymbol{\epsilon}_i, \quad \mathbb{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\epsilon}_i) = I, \quad (1)$$

where \mathbf{y}_i is the vector of measurement of length l_i , A_i is a $(l_i \times m)$ -observation matrix, and I denotes the identity matrix. The errors on the measurements are assumed to be mutually independent, i.e. $\mathbb{E}(\boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_j) = 0$ for all $i \neq j \in [s]$.

The aim is to choose a well suited subset $\mathcal{I} \subseteq [s]$ of experiments that one will conduct in order to estimate the parameters. We therefore define the *design* variable \mathbf{w} as the 0/1 vector of length s , where w_k takes the value 1 if and only if $k \in \mathcal{I}$. Given the collection of observations $(\mathbf{y}_i)_{i \in \mathcal{I}}$, it is known (see e.g. Pukelsheim [11]) that the best linear unbiased estimator of $\boldsymbol{\theta}$ is given by least square theory (Gauss-Markov theorem), and the inverse of its variance is given by

$$M_F(\mathbf{w}) := A(\mathbf{w})^T A(\mathbf{w}) = \sum_{k=1}^s w_k A_k^T A_k. \quad (2)$$

The classical experimental design approach consists in choosing the set of experiments \mathcal{I} (or the design w) in order to maximize the matrix $M_F(\mathbf{w})$ with respect to the Löwner ordering [11]. This can be interpreted as the problem of finding the design for which the confidence ellipsoids of the best estimator are the smallest (for the inclusion relation). Since the Löwner ordering on symmetric matrices is only a partial ordering, the problem consisting in

maximizing $M_F(\mathbf{w})$ is ill-posed. Kiefer [7] proposed to use the class of matrix means Φ_p , which map the cone of semidefinite positive matrices \mathbb{S}_m^+ onto the real line and satisfy natural properties, such as monotonicity with respect to Löwner ordering and concavity. For a real number $p \in]-\infty, 1] \setminus \{0\}$, and a positive definite matrix M with eigenvalues $0 < \lambda_1 \leq \dots \leq \lambda_m$,

$$\Phi_p(M) := \left(\frac{1}{m} \text{trace } M^p\right)^{1/p} = \left(\frac{1}{m} \sum_{j=1}^m \lambda_j^p\right)^{1/p} \quad (3)$$

Interestingly, Φ_p can also be defined for $p = -\infty$ and $p = 0$ as limiting cases, i.e. $\Phi_{-\infty}(M) = \lambda_1$ and $\Phi_0(M) = (\det(M))^{1/m}$, which are the classical optimality criteria referred in the literature as E - and D -optimality respectively. The definition of Φ_p is also extended to singular positive semidefinite matrices M (i.e. $\lambda_1 = 0$) by continuity, so that $\Phi_p(M) = 0$ for all $p \leq 0$, and $\Phi_p(M)$ is defined by Equation (3) for all $p \in [0, 1]$.

We can finally give a mathematical formulation to the problem of selecting the best subset of experiments. If the maximal number of experiments is n , the Φ_p -Optimal Design problem is:

$$\max_{w \in \{0,1\}^s, \sum_k w_k \leq n} \Phi_p\left(M_F(\mathbf{w})\right). \quad (4)$$

We note that the problem of maximizing $M_F(\mathbf{w})$ with respect to the Löwner ordering remains meaningful even when $M_F(\mathbf{w})$ is not of full rank. This case does arise in under-instrumented situations, in which some constraints may not allow one to conduct a number of experiments which is sufficient to infer all the parameters. In this case, positive values of p must be used, and Problem (4) is equivalent to:

$$\max_{w \in \{0,1\}^s, \sum_k w_k \leq n} \varphi_p(\mathbf{w}) := \text{trace} \left(\sum_k w_k A_k^T A_k \right)^p \quad (P_p)$$

In the under-instrumented situation, an interesting and natural idea is to choose the design which maximizes the rank of the information matrix $M_F(\mathbf{w})$. The first order expansion of φ_p as $p \rightarrow 0^+$ shows that (P_p) may be thought as a regularization of this *rank optimization* problem, and φ_p can be seen as a deformation of the rank criterion for $p \in]0, 1]$:

$$\text{trace } M^p = \sum_{k=1}^r \lambda_k^p = r + p \log\left(\prod_{k=1}^r \lambda_k\right) + \mathcal{O}(p^2) \quad (5)$$

Consequently, $\text{trace } M^0$ will stand for $\text{rank}(M)$ in the sequel, so that $\varphi_0(\mathbf{w}) = \text{rank}(M_F(\mathbf{w}))$, and the *rank maximization* problem is (P_0) .

2 Submodularity and the greedy algorithm

In this section, we recall some results announced in [13], which state that Problem (P_p) is equivalent to the maximization of a *nondecreasing submodular* function [8,9,4,16]. Note that there is no point to consider multiplicative approximation factors for the Φ_p -optimal problem when $p \leq 0$, since the criterion takes the value 0 as long as the information matrix is singular. For $p \leq 0$ indeed, the instances of the Φ_p -optimal problem where no feasible design lets $M_F(\mathbf{w})$ be of full rank have an optimal value of 0. For all other instances, any polynomial-time algorithm with a positive approximation factor would necessarily return a design of full rank. Provided that $P \neq NP$, this would contradict the NP-hardness of the rank optimization problem (P_0) [13]. So, we investigate approximation algorithms only in the case $p \geq 0$, and approximation factors are given with respect to the objective function φ_p . Whenever it is necessary, we will identify the function $\varphi_p : \{0, 1\}^s \rightarrow \mathbb{R}$ with the set function $\varphi_p : 2^{[s]} \rightarrow \mathbb{R}$.

Definition 2.1 A real-valued set function $f : 2^E \rightarrow \mathbb{R}$ such that $f(\emptyset) = 0$ is called *nondecreasing submodular* if it satisfies the following conditions :

- $f(I) \leq f(J)$ whenever $I \subseteq J \subseteq E$;
- $f(I) + f(J) \geq f(I \cup J) + f(I \cap J)$ for all $I, J \subseteq E$.

The next results show that for all $p \in [0, 1]$, Problem (P_p) is $1 - e^{-1}$ -approximable in polynomial time. This can be attained with the help of the greedy algorithm, whose principle is to start from $\mathcal{G}_0 = \emptyset$ and to construct sequentially the sets $\mathcal{G}_k := \mathcal{G}_{k-1} \cup \operatorname{argmax}_{i \in [s]} \varphi_p(\mathcal{G} \cup i_k)$, for $k = 1, \dots, n$.

Theorem 2.2 (Submodularity of φ_p [12]) For all $p \in [0, 1]$, φ_p is a nondecreasing submodular set function.

The proof of this result relies on a matrix inequality of independent interest:

$$\forall X, Y, Z \in \mathbb{S}_m^+, \forall p \in [0, 1], \\ \operatorname{trace}(X + Y + Z)^p + \operatorname{trace} Z^p \leq \operatorname{trace}(X + Z)^p + \operatorname{trace}(Y + Z)^p.$$

Corollary 2.3 (Approximability of φ_p -Optimal Design) Let $p \in [0, 1]$. The greedy algorithm yields a $1 - e^{-1}$ approximation factor for Problem (P_p) .

Proof. Nemhauser, Wolsey and Fisher [9] proved it for any nondecreasing submodular function over a uniform matroid. Moreover, when the maximal number of experiments which can be selected is n , this approximation factor can be improved to $1 - (1 - 1/n)^n$. \square

3 Approximation Factor for the randomized rounding

The optimal design problem has a natural continuous convex relaxation which has been extensively studied [2,5,17]. As mentioned in the introduction, several authors proposed to solve this convex relaxation and to round the solution to obtain a near-optimal discrete design. In this paper, we investigate the legitimacy of this technique. We show in Theorem 3.6 that the D -optimal design may be rounded to a random discrete design which approximates the optimum of the rank optimization problem (P_0) by an average factor of $\frac{n}{s}$. While this result may look rather worse than the greedy approximation factor presented in Section 2, it is (almost) optimal since there are some instances for which the average ratio of approximation is $\frac{n}{s-1}$ (Remark 3.7).

Another motivation for this section is the recent results from Calinescu, Chekuri, Pál and Vondrák [4,16] who showed that the problem of maximizing a nondecreasing submodular function over an arbitrary matroid is $(1 - e^{-1})$ -approximable, by first *approaching the maximum* of a continuous extension of the submodular function, and then using the pipage rounding of Ageev and Sviridenko [1] to return a discrete solution which achieves the $(1 - e^{-1})$ -approximation factor. For our problem, the greedy algorithm of Section 2 is preferable to obtain a $(1 - e^{-1})$ -approximation factor, but the ideas of Calinescu and his coauthors are useful to establish the approximability factor of the rank optimization problem (P_0) by rounding algorithms.

The continuous relaxation of the D -optimal problem is obtained by removing the integer constraint $\mathbf{w} \in \{0, 1\}^s$:

$$\max_{\mathbf{w} \in (\mathbb{R}^+)^s, \sum_k w_k \leq n} \det \left(\sum_k w_k A_k^T A_k \right). \quad (6)$$

We assume without loss of generality that $\text{rank}(\sum_{i=1}^s A_i^T A_i) = m$, so that (6) is well posed. Otherwise, we must project each observation matrix onto an appropriate observable subspace, as in Pukelsheim [11]. The function $\log(\det(\cdot))$ is strictly concave on the interior of S_m^+ , and Problem (6) can be solved by a projected gradient algorithm, interior point techniques, or multiplicative algorithms [2,5,17]. In this paper, we focus on the rounding techniques only, and we assume that an optimal solution \mathbf{w}^* of Problem (6) is readily known. We also denote a discrete solution of Problem (P_0) by S^* . Since $M_F(\mathbf{w}^*)$ is of full rank, we have: $m = \varphi_0(\mathbf{w}^*) = \text{rank}(M_F(\mathbf{w}^*)) \geq \varphi_0(S^*)$.

Definition 3.1 We say that an algorithm approximates the optimal solution of the rank optimization problem (P_0) by a factor α if for all possible instances,

it returns a feasible random subset \hat{S} such that:

$$\mathbb{E}(\varphi_0(\hat{S})) \geq \alpha \varphi_0(S^*).$$

We study two kinds of randomized roundings: the pipage rounding algorithm of Ageev and Sviridenko [1] and another rounding algorithm which we call *proportional rounding*. We will make use of the extension by expectation [4] of the submodular set function φ_0 , which is defined by

$$F_0(\mathbf{y}) = \mathbb{E}[\varphi_0(\hat{S})] \tag{7}$$

where \hat{S} is a random set of $[s]$ which contains $\{i\}$ independently with probability y_i . If \mathbf{y} is the 0/1-vector associated to S , we have $F_0(\mathbf{y}) = \varphi_0(S)$, which tells us that F_0 is an extension of the rank function φ_0 indeed. Note that F_0 can be defined only if all coordinates of \mathbf{y} are smaller than 1.

The idea of the randomized pipage rounding [4], as reduced to the simple case of uniform matroids, is the following: Let \mathbf{y} be a vector such that $\sum_i y_i = n$, and let i, j be two indices for which \mathbf{y} is fractional. The function $t \rightarrow F_0(y + te_i - te_j)$ is convex, so that F_0 is increasing when we move in one of the directions $(e_i - e_j)$ or $(e_j - e_i)$. We choose between these directions with probabilities that are proportional to the values of t for which y_i or y_j becomes a 0 or a 1, and we set $y \leftarrow y + te_i - te_j$.

Lemma 3.2 (Calinescu et al [4]) *Given a vector $\mathbf{y} \in [0, 1]^s$ such that $\sum_i y_i = n$, the randomized pipage rounding returns in $O(s)$ iterations a random set S of cardinality n , of expected value $\mathbb{E}[\varphi_0(S)] \geq F_0(\mathbf{y})$.*

This lemma shows that if we can prove that $F_0(\mathbf{w}^*) \geq \alpha m (\geq \alpha \varphi_0(S^*))$, then the randomized pipage rounding algorithm approximates the solution of (P_0) by a factor α .

We now present another rounding, which can be used even if some coordinates of \mathbf{w}^* are larger than 1. The principle of this rounding is to start with $S_0 = \emptyset$, and, for $k = 1, \dots, n$, we construct S_k from S_{k-1} by adding in it exactly one new element, namely $i \in [s] \setminus S_{k-1}$ with probability $\frac{w_i^*}{\sum_{j \notin S_{k-1}} w_j^*}$. If at some point, all the remaining coordinates $(w_j)_{j \notin S_{k-1}}$ are equal to 0, uniform probabilities are used. We denote by $S_n(\mathbf{w})$ the random set of cardinality n obtained by this procedure, which we call *proportional rounding* of vector \mathbf{w} .

We next give a proposition which shows how we can bound the coordinates w_i^* of the D -optimal design. This is a generalization of a result of Atwood [2] who obtained $w_i^* \leq \frac{1}{m}$ in the case where all the observations are scalar, i.e. when the observation matrices are row vectors. The proof is omitted due to

lack of space, and is presented in [12]. It is inspired from a fix point equation from Pukelsheim [10], which must be satisfied by the optimal designs in the scalar observations case.

Proposition 3.3 *Let \mathbf{w}^* be a D -optimal design and S be an arbitrary subset of $[s]$. We have the following bound on the optimal coordinate w_i^* :*

$$\frac{\sum_{i \in S} w_i^*}{n} \leq \frac{\text{rank}(\sum_{i \in S} M_i)}{\text{rank}(\sum_{i=1}^n M_i)} = \frac{\varphi_0(S)}{m}. \quad (8)$$

Before we give the approximability factor that one can guarantee by using our rounding procedures, we need these two technical lemmas, which are proved in [12] by elementary induction.

Lemma 3.4 *Let $\alpha\Delta_s$ denote the simplex $\{x \in (\mathbb{R}^+)^s \mid \sum_i x_i = \alpha\}$. We define the random variable $W_n(\mathbf{w}) = \sum_{i \in S_n(\mathbf{w})} w_i$, where $S_n(\mathbf{w})$ is the random subset of $[s]$ obtained by proportional rounding. Then, we have*

$$\forall \mathbf{w} \in \alpha\Delta_s, \quad \mathbb{E}[W_n(\mathbf{w})] \geq \mathbb{E}[W_n(\frac{\alpha}{s}, \dots, \frac{\alpha}{s})] = n \frac{\alpha}{s}.$$

Lemma 3.5 *For all vector $w \in [0, 1]^s$, the following equality holds:*

$$\sum_{S \subset \{1, \dots, s\}} \left(\sum_{i \in S} w_i \right) \prod_{i \in S} w_i \prod_{i \notin S} (1 - w_i) = \sum_{i=1}^s w_i^2.$$

The main result of this section follows from these two lemmas, by expanding the expressions $\mathbb{E}[\varphi_0(S_n(\mathbf{w}^*))]$ and $F_0(\mathbf{w}^*)$ [12].

Theorem 3.6 (Rounding Approximability Factor) *Let \mathbf{w}^* be a continuous D -optimal design. The proportional rounding of \mathbf{w}^* approximates the solution of the rank optimization problem (P_0) by $\frac{n}{s}$. If \mathbf{w}^* is bounded by 1, then the pipage rounding yields the same $\frac{n}{s}$ approximation factor.*

Remark 3.7 The inequalities $\mathbb{E}[\varphi_0(S_n(\mathbf{w}^*))] \geq \frac{n}{s}m$ and $F_0(\mathbf{w}^*) \geq \frac{n}{s}m$ are optimal. To see this, consider the $s \times s$ -information matrices M_1, \dots, M_s , where each M_i has a single nonzero entry on the i^{th} term of its diagonal. Moreover, if $s > n + 1$, we exhibit in [12] some information matrices for which the ratios $\frac{\mathbb{E}[\varphi_0(S_n(\mathbf{w}^*))]}{\varphi_0(S^*)}$ and $\frac{F(\mathbf{w}^*)}{\varphi_0(S^*)}$ take the value $\frac{n}{s-1}$, and we show that this approximation factor of $\frac{n}{s-1}$ is optimal for $n = 1$.

References

- [1] A. A. Ageev and M. I. Sviridenko. Pipage rounding: a new method of constructing algorithms with proven performance guarantee. *Journal of*

- Combinatorial Optimization*, 8:2004, 2004.
- [2] C.L. Atwood. Sequences converging to D-optimal designs of experiments. *Annal of statistics*, 1:342–352, 1973.
 - [3] M. Branderhorst, I. Walmsley, R. Kosut, and H. Rabitz. Optimal experiment design for quantum state tomography of a molecular vibrational mode. *Journal of Physics B: Atomic Molecular and Optical Physics*, 41:074004, 2008.
 - [4] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák. Maximizing a submodular set function subject to a matroid constraint. In *Proceedings of the 12th international conference on Integer Programming and Combinatorial Optimization, IPCO*, volume 4513, pages 182–196, 2007.
 - [5] H. Dette, A. Pepelyshev, and A. Zhigljavsky. Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Computational Statistics & Data Analysis*, 53(2):312 – 320, 2008.
 - [6] V.V. Fedorov. *Theory of optimal experiments*. New York : Academic Press, 1972. Translated and edited by W. J. Studden and E. M. Klimko.
 - [7] J. Kiefer. Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, 62(2):277–288, 1975.
 - [8] A. Kulik, H. Shachnai, and T. Tamir. Maximizing submodular set functions subject to multiple linear constraints. In *SODA '09: Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms*, pages 545–554, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.
 - [9] G.L. Nemhauser, L.A. Wolsey, and M.L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 1978.
 - [10] F. Pukelsheim. On linear regression designs which maximize information. *Journal of statistical planning and inference*, 4:339–364, 1980.
 - [11] F. Pukelsheim. *Optimal Design of Experiments*. Wiley, 1993.
 - [12] G. Sagnol. Polynomial-time approximability results for combinatorial problems arising in optimal experimental design. Draft.
 - [13] G. Sagnol, M. Bouhtou, and S. Gaubert. Optimization of network traffic measurement: a semidefinite programming approach. In *Proceedings of the International Conference on Engineering Optimization (ENGOPT)*, Rio De Janeiro, Brazil, 2008. ISBN 978-85-7650-152-7.
 - [14] H.H. Song, L. Qiu, and Y. Zhang. Netquest: A flexible framework for largescale network measurement. In *ACM SIGMETRICS'06*, St Malo, France, 2006.
 - [15] D. Uciński and M. Patan. D-optimal design of a monitoring network for parameter estimation of distributed systems. *Journal of Global Optimization*, 39(2):291–322, 2007.

- [16] J. Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *ACM Symposium on Theory of Computing, STOC'08*, pages 67–74, 2008.
- [17] Y. Yu. Monotonic convergence of a general algorithm for computing optimal designs, 2009. [arXiv.org:0905.2646v2](https://arxiv.org/abs/0905.2646v2).