Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Approximation Algorithms for D-optimal Design

Mohit Singh, Weijun Xie

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Approximation Algorithms for *D*-optimal Design

Mohit Singh,[a] Weijun Xie[b]

[a] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332; [b] Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061
Contact: mohitsinghr@gmail.com (MS); wxie@vt.edu, https://orcid.org/0000-0001-5157-1194 (WX)

Abstract. Experimental design is a classical statistics problem, and its aim is to estimate an unknown vector from linear measurements where a Gaussian noise is introduced in each measurement. For the combinatorial experimental design problem, the goal is to pick a subset of experiments so as to make the most accurate estimate of the unknown parameters. In this paper, we will study one of the most robust measures of error estimation—the *D*-optimality criterion, which corresponds to minimizing the volume of the confidence ellipsoid for the estimation error. The problem gives rise to two natural variants depending on whether repetitions of experiments are allowed or not. We first propose an approximation algorithm with a $1/e$-approximation for the *D*-optimal design problem with and without repetitions, giving the first constant-factor approximation for the problem. We then analyze another sampling approximation algorithm and prove that it is asymptotically optimal. Finally, for *D*-optimal design with repetitions, we study a different algorithm proposed by the literature and show that it can improve this asymptotic approximation ratio. All the sampling algorithms studied in this paper are shown to admit polynomial-time deterministic implementations.

Keywords: *D*-optimal design • approximation algorithm • determinant • derandomization

1. Introduction

Experimental design is a classical problem in statistics (Atkinson et al. [5], Federer [18], Fedorov [19], Kirk [24], Pukelsheim [32]), and recently, it has also been applied to machine learning (Allen-Zhu et al. [3], Wang et al. [39]). In an experimental design problem, its aim is to estimate an m-dimensional vector $\beta \in \mathbb{R}^m$ from n linear measurements of the form $y_i = a_i^T \beta + \tilde{\epsilon}_i$ for each $i \in [n] := \{1, 2, \ldots, n\}$, where vector $a_i \in \mathbb{R}^m$ characterizes the i experiment and $\{\tilde{\epsilon}_i\}_{i \in [n]}$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance σ^2 (i.e., $\tilde{\epsilon}_i \sim \mathcal{N}(0, \sigma^2)$ for all $i \in [n]$). Because of limited resources, it might be quite expensive to conduct all of the n experiments. Therefore, as a compromise, in the combinatorial experimental design problem, we are given an integer $k \in [m, n]$, and our goal is to pick k out of the n experiments so as to make the *most accurate estimate* of the parameters denoted as $\widehat{\beta}$. Suppose that a size k subset of experiments $S \subseteq [n]$ is chosen; then, the most likelihood estimation of β (compare with Joshi and Boyd [22]) is given by

$$\widehat{\beta} = \left(\sum_{i \in S} a_i a_i^\top \right)^{-1} \sum_{i \in S} y_i a_i.$$

There are many criteria on how to choose the best estimation among all of the possible size k subsets of n experiments (see Atkinson et al. [5] for a review). One of the most robust measures of error estimation is known as *D*-optimality criterion, where the goal is to choose the best size k subset S to maximize $[\det(\sum_{i \in S} a_i a_i^\top)]^{\frac{1}{m}}$: that is, the combinatorial optimization problem

$$\max_{S} \left\{ f(S) := \left[\det\left(\sum_{i \in S} a_i a_i^\top \right) \right]^{\frac{1}{m}} : \mathrm{Supp}(S) \subseteq [n], |S| = k \right\}, \tag{1}$$

where $\mathrm{Supp}(S)$ denotes the support of set S and $|S|$ denotes the cardinality of set S. Note that optimization model (1) corresponds to minimizing the volume of the confidence ellipsoid for the estimation error $\beta - \widehat{\beta}$. Equivalently, the objective function of (1) can be chosen as determinant itself (i.e., $\det(\sum_{i \in S} a_i a_i^\top)$) or log determinant ($\log \det(\sum_{i \in S} a_i a_i^\top)$) (Joshi and Boyd [22]). However, both objective functions are problematic for

the following reasons: (1) the determinant function is nonconvex and has numerical issue, especially when k, m, n are large; and (2) although log-determinant function is concave, it can also be numerically unstable, in particular when the determinant is close to zero. Therefore, in this paper, we follow the work in Sagnol and Harman [33] and consider mth root of determinant function, which is concave and numerically stable.

In the problem description, the same experiment may or may not be allowed to choose multiple times. We refer to the problem as *D-optimal design with repetitions* if we are allowed to pick an experiment more than once and *D-optimal design without repetitions* otherwise. Correspondingly, for D-optimal design with repetitions and without repetitions, in (1), the subset S denotes a *multiset*, where elements of S can be duplicated, and a *conventional* set, where elements of S must be different from each other, respectively. The former problem has been studied very extensively in statistics (Kirk [24], Pukelsheim [32], Sagnol and Harman [33]). The latter problem has also been studied as the sensor selection problem (Joshi and Boyd [22]), where the goal is to find the best subset of sensor locations to obtain the most accurate estimate of unknown parameters. It is easy to see that D-optimal design with repetitions is a special case of the D-optimal design problem without repetitions. To do so, for the D-optimal design with repetitions, we can create k copies of each vector, which reduces to the D-optimal design without repetitions with nk vectors.

The remainder of the paper is organized as follows. Section 2 details the problem setting, reviews related literature, and summarizes our contributions. Section 3 develops and analyzes a randomized algorithm, its approximation results, and deterministic implementation. Section 4 proposes another randomized algorithm and its deterministic implementation as well as asymptotic behavior. Section 5 analyzes a randomized algorithm for D-optimal design with repetitions proposed by literature, and investigates its deterministic implementation as well as approximation ratios. Finally, Section 6 concludes the paper.

1.1. Notations

The following notation is used throughout the paper. We use bold letters (e.g., x, A) to denote vectors or matrices, and we use corresponding nonbold letters to denote their components. We let $\mathbf{e}$ be the all-ones vector. We let $\mathbb{R}_+, \mathbb{Q}_+, \mathbb{Z}_+$ denote the sets of positive real numbers, rational numbers, and integers, respectively. Given a positive integer N and a set Q, we let $[N] := \{1, \dots, N\}$, $N! = \prod_{i \in [N]} i$, $|Q|$ denote its cardinality, $\text{Supp}(Q)$ denote the support of Q, and $\binom{Q}{N}$ denote all of the possible subsets of Q with cardinality that equals N. Given a multiset Q, for each $i \in \text{Supp}(Q) \subseteq [n]$, we let function $M_Q(i)$ denote the number of occurrences of i in Q, and for any $i \in [n]$, $Q(i)$ denotes its ith component. Given a matrix A and two sets R, T, we let $\det(A)$ denote its determinant if A is a square matrix, let $A_{R,T}$ denote a submatrix of A with rows and columns from sets R, T, let A_i denote ith column of matrix A, and let A_R denote submatrix of A with columns from set R. For any positive integer r, we let I_r denote $r \times r$ identity matrix. We use $\tilde{\mathcal{S}}$ to denote a random set. For notational convenience, give a set $S \in [n]$ and vector $x \in \mathbb{R}_+^n$, we define $f(S) = [\det(\sum_{i \in S} a_i a_i^\top)]^{\frac{1}{m}}$ and $f(x) = [\det(\sum_{i \in [n]} x_i a_i a_i^\top)]^{\frac{1}{m}}$. Additional notation will be introduced as needed.

2. Model Formulation, Related Literature, and Contributions
2.1. Model Formulation

To formulate D-optimal design problem (1) as an equivalent mathematical program, we first let set B denote the set of all nonnegative numbers (i.e., $B = \mathbb{R}_+$) if repetitions are allowed; otherwise, $B = [0,1]$. Next, we introduce integer decision variable $x_i \in B \cap \mathbb{Z}_+$ to represent how many times ith experiment will be chosen for each $i \in [n]$. With the notation introduced, the D-optimal design problem can be formulated as a mixed integer convex program:

$$w^* := \max_{x,w} \left\{ w : w \le f(x) = \left[\det\left(\sum_{i \in [n]} x_i a_i a_i^\top \right) \right]^{\frac{1}{m}}, \sum_{i \in [n]} x_i = k, x \in B^n \cap \mathbb{Z}_+^n \right\}, \quad (2)$$

where for notational convenience, we let $f(x)$ denote the objective function. Note that, if $B = \mathbb{R}_+$, (2) is equivalent to D-optimal design with repetitions, and if $B = [0,1]$, then (2) corresponds to D-optimal design without repetitions.

It can be easily shown that $f(x) = [\det(\sum_{i \in [n]} x_i a_i a_i^\top)]^{\frac{1}{m}}$ is concave in x (compare with Ben-Tal and Nemirovski [8]). Therefore, a straightforward convex relaxation of (2) is to relax the binary vector x to continuous (i.e., $x \in [0,1]^n$), which is formulated as below:

$$\widehat{w} := \max_{x,w} \left\{ w : w \le \left[\det\left(\sum_{i \in [n]} x_i a_i a_i^\top \right) \right]^{\frac{1}{m}}, \sum_{i \in [n]} x_i = k, x \in B^n \right\}. \tag{3}$$

Note that (3) is a tractable convex program (compare with Joshi and Boyd [22]); thus, it is efficiently solvable. Recently, Sagnol and Harman [33] proposed an alternative second-order conic program formulation for (3), which can be solved by a more effective interior point method or even off-the-shelf solvers (e.g., CPLEX, Gurobi, and MOSEK). We remark that, according to Ben-Tal and Nemirovski [8], the time complexity of solving (3) is $O(n^5)$.

2.2. Related Literature

As remarked earlier, experimental design is a classical area in statistics. We refer the reader to Pukelsheim [32, chapter 9] on details about D-optimality criterion as well as other related (A, E, T) criteria. The combinatorial version, where the number of each experiment that needs to be chosen is an integer as in (2), is also called exact experimental design. It turns out that the D-optimality criterion is proven to be NP-hard (Welch [40]). Thus, there has been plenty of works on heuristic methods, including local search and its variants, to obtain good solutions (Fedorov [19], Joshi and Boyd [22]).

From an approximation algorithm viewpoint, D-optimal design has received a lot of attention recently. For example, Bouhtou et al. [9] gave a $(n/k)^{\frac{1}{m}}$-approximation algorithm, and Wang et al. [39], building on Avron et al. [6], gave a $(1 + \epsilon)$ approximation if $k \ge m^2/\epsilon$. Recently, Allen-Zhu et al. [3] realized the connection between this problem and matrix sparsification (Batson et al. [7], Spielman and Srivastava [37]) and used regret minimization methods (Allen-Zhu et al. [2]) to obtain $O(1)$-approximation algorithm if $k \ge 2m$ and $(1 + \epsilon)$ approximation when $k \ge O(m/\epsilon^2)$. We also remark that their results are general and also, applicable to other optimality criteria.

Another closely related problem is the largest j-simplex problem, the problem description of which is as follows: (Largest j-simplex problem) Given a set of n vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ and integer $k \le m$, pick a set of S of k vectors to maximize the k root of the pseudodeterminant of $X = \sum_{i \in S} a_i a_i^\top$ (i.e., the geometric mean of the nonzero eigenvalues of X).

The problem has also received much attention recently (Di Summa et al. [17], Khachiyan, [23], Nikolov [28]), and Nikolov [28] gave a $1/e$-approximation algorithm. Observe that the special case of $k = m$ of this problem coincides with the special case of $k = m$ for the D-optimal design problem. Indeed, Nikolov's algorithm, although applicable, results in a $e^{-\frac{k}{m}}$ approximation for the D-optimal design problem. Recently, matroid constrained versions of the largest j-simplex problem have also been studied (Anari and Gharan [4], Nikolov and Singh [29], Straszak and Vishnoi [38]).

The D-optimality criterion is also closely related to constrained submodular maximization (Nemhauser et al. [27]), a classical combinatorial problem, for which there has been much progress recently (Calinescu et al. [14], Krause and Golovin [25]). Indeed, the set function $m \log f(S) := \log \det(\sum_{i \in S} a_i a_i^\top)$ is known to be submodular (Shamaiah et al. [34]). Unfortunately, this submodular function is not necessarily nonnegative, a prerequisite for all of the results on constrained submodular maximization, and thus, these results are not directly applicable. We also remark that, for a multiplicative guarantee for the det objective, we would aim for an additive guarantee for log det objective.

2.3. Contributions

In this paper, we make several contributions to the approximations of D-optimal design problems both with and without repetitions. Our approximation algorithms are randomized, where we sample k experiments out of n with given marginals. This type of sampling procedure has been studied intensively in the approximation algorithms, and many different schemes have been proposed (Brewer and Hanif [11]), where most of them exhibit a negative correlation of various degrees (Brändén and Jonasson [10]). In this work, we will study sampling schemes, which exhibit approximate positive correlation and prove the approximation ratios of these algorithms.

All of the proposed approximations come with approximation guarantees. Given an *approximation ratio* $\gamma \in (0,1]$, a γ-approximation algorithm for D-optimal design returns a solution $\bar{x} \in B^n \cap \mathbb{Z}_+^n$ such that $\sum_{i \in [n]} \bar{x}_i = k$

and $f(\bar{x}) \geq \gamma w^*$ (i.e., the solution is feasible and has an objective value at least γ times the optimal value). The approximation ratios of our randomized algorithms only hold in the sense of expectation. To improve them, we further propose polynomial time deterministic implementations for all of the randomized algorithms with iterated conditional expectation method, which also achieve the same approximation guarantees. The following is a summary of our contributions:

1. We develop a $1/e$-approximation algorithm and its polynomial time deterministic implementation, giving the first constant factor approximation for D-optimal design problem for both with and without repetitions. Previously, constant factor approximations were known only for a restricted range of parameters (Allen-Zhu et al. [3], Avron et al. [6], Wang et al. [39]) (see related work for details).

2. We study a different sampling algorithm and its polynomial time deterministic implementation, showing that its solution is $(1 - \epsilon)$ optimal if $k \geq 4m/\epsilon + 12/\epsilon^2 \log(1/\epsilon)$ for any given $\epsilon \in (0, 1)$. These results substantially improve the previous work (Allen-Zhu et al. [3], Wang et al. [39]).

3. For D-optimal design with repetitions, we investigate a simple randomized algorithm similar to that in Nikolov [28], study its polynomial time deterministic implementation, and provide a significant different analysis of the approximation guarantee. We show that the proposed algorithm yields $(1 - \epsilon)$ approximation for the D-optimal design problem with repetitions when $k \geq (m - 1)/\epsilon$.

Note that the preliminary version of the paper appeared in the *ACM-SIAM Symposium on Discrete Algorithms* (Singh and Xie [35]). Compared with [35], this paper has the following major improvement: (1) For the constant factor approximation algorithm presented in Section 3, we simplify its polynomial time deterministic implementation and improve its analysis. (2) For the asymptotically optimal algorithm in Section 4, we simplify its sampling procedure and derive its polynomial time deterministic implementation. (3) For the approximation algorithm of D-optimal design with repetitions, we improve its approximation ratio analysis and propose its polynomial time deterministic implementation.

3. Approximation Algorithm for *D*-optimal Design Problem

In this section, we will propose a sampling procedure and prove its approximation ratio. We also develop an efficient way to implement this algorithm, and finally, we will show its polynomial time deterministic implementation with the same performance guarantees.

First of all, we note that, for D-optimal design problem with repetitions (i.e., $B = \mathbb{R}_+$ in (2)), it can be equivalently reformulated as D-optimal design problem without repetitions by creating k copies of vectors $\{a_i\}_{i \in [n]}$. Therefore, the approximation for D-optimal design problem without repetitions directly applies to that with repetitions. Hence, in this section and the next sections, we will only focus on D-optimal design problem without repetitions (i.e., in (2) and (3), we only consider $B = [0, 1]$).

3.1. Sampling Algorithm and Its Efficient Implementation

In this subsection, we will introduce a sampling procedure and explain its efficient implementation.

In this sampling algorithm, we first suppose $(\widehat{x}, \widehat{w})$ to be an optimal solution to the convex relaxation (3), where $\widehat{x} \in [0, 1]^n$ with $\sum_{i \in [n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$. Then, we randomly choose a size k subset $\tilde{\mathcal{S}}$ according to the following probability:

$$\mathbb{P}\left[\tilde{\mathcal{S}} = S\right] = \frac{\prod_{j \in S} \widehat{x}_j}{\sum_{\bar{S} \in \binom{[n]}{k}} \prod_{i \in \bar{S}} \widehat{x}_i} \tag{4}$$

for every $S \in \binom{[n]}{k}$, where $\binom{[n]}{k}$ denotes all of the possible size k subsets of $[n]$.

Note that the sampling procedure in (4) is not an efficient implementable description, because there are $\binom{n}{k}$ candidate subsets to be sampled from. Therefore, we propose an efficient way (i.e., Algorithm 1) to obtain a size k random subset $\tilde{\mathcal{S}}$ with probability distribution in (4). We first observe a useful way to compute the probabilities in Algorithm 1 efficiently.

Observation 1. Suppose $x \in \mathbb{R}^t$ and integer $0 \leq r \leq t$; then, $\sum_{S \in \binom{[t]}{r}} \prod_{i \in S} x_i$ is the coefficient of y^r of the polynomial $\prod_{i \in [t]} (1 + x_i y)$.

In fact, it has been shown that the product of two polynomials with degree at most t can be done in $O(t \log t)$ amount of time by the fast Fourier transform (Cooley and Tukey [16]). Thus, by the divide-and-conquer approach, it takes $O(t \log^2 t)$ time to expand the polynomial $\prod_{i \in [t]} (1 + x_i y)$ (i.e., it takes $O(t \log^2 t)$ time to compute the coefficient of y^r of the polynomial $\prod_{i \in [t]} (1 + x_i y)$).

Note that we need to compute the probability $\mathbb{P}[\tilde{\mathcal{G}} = S]$ in (4) efficiently. The main idea of the efficient implementation is to sample elements one by one based on conditioning on the chosen elements and unchosen elements and update the conditional probability distribution sequentially. Indeed, suppose that we are given a subset of chosen elements S with $|S| < k$ and a subset of unchosen elements T with $|T| < n - k$, which can be empty; then, probability that the jth experiment with $j \in [n] \setminus (S \cup T)$ will be chosen is equal to

$$\mathbb{P}[j \text{ will be chosen}|S, T] = \frac{\widehat{x}_j \left(\sum_{\bar{S} \in \binom{[n]\setminus(S\cup T)}{k-1-|S|}} \prod_{\tau \in \bar{S}} \widehat{x}_\tau \right)}{\left(\sum_{\bar{S} \in \binom{[n]\setminus(S\cup T)}{k-|S|}} \prod_{\tau \in \bar{S}} \widehat{x}_\tau \right)}.$$

In the formula, the denominator and numerator can be computed efficiently based on Observation 1. Thus, we flip a coin with success rate equal to the probability, which clearly has the following two outcomes: if j is chosen, then update $S := S \cup \{j\}$; otherwise, update $T := T \cup \{j\}$. Then, go to the next iteration, and repeat this procedure until $|S| = k$. By applying iterated conditional probability, the probability that S is chosen equal to $\prod_{j \in S} \widehat{x}_j / (\sum_{\bar{S} \in \binom{[n]}{k}} \prod_{j \in \bar{S}} \widehat{x}_j)$ (i.e., (4)) holds. The detailed implementation is shown in Algorithm 1. Note that the time complexity of Algorithm 1 is $O(n^2)$.

Algorithm 1 Efficient Implementation of Sampling Procedure (4) with Constant Factor Approximation

1. Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = [0, 1]$, where $\widehat{x} \in [0, 1]^n$ with $\sum_{i \in [n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$
2. Initialize chosen set $\tilde{\mathcal{G}} = \emptyset$ and unchosen set $T = \emptyset$
3. Two factors: $A_1 = \sum_{\bar{S} \in \binom{[n]}{k}} \prod_{i \in \bar{S}} \widehat{x}_i, A_2 = 0$
4. **for** $j = 1, \ldots, n$ **do**
5.   **if** $|\tilde{\mathcal{G}}| == k$ **then**
6.   break
7.   **else if** $|T| = n - k$ **then**
8. $\tilde{\mathcal{G}} = [n] \setminus T$
9.   break
10.   **end if**
11.   Let $A_2 = (\sum_{\bar{S} \in \binom{[n]\setminus(\tilde{\mathcal{G}}\cup T)}{k-1-|\tilde{\mathcal{G}}|}} \prod_{\tau \in \bar{S}} \widehat{x}_\tau)$

12.   Sample a $(0, 1)$ uniform random variable U
13.   **if** $\widehat{x}_j A_2 / A_1 \geq U$ **then**
14.   Add j to set $\tilde{\mathcal{G}}$
15. $A_1 = A_2$
16.   **else**
17.   Add j to set T
18. $A_1 = A_1 - \widehat{x}_j A_2$
19.   **end if**
20. **end for**
21. Output $\tilde{\mathcal{G}}$

3.2. m-wise α-positively Correlated Probability Distributions

In this subsection, we will introduce the main proof idea of the approximation guarantees, which is to analyze the probability distribution (4) and show that it is approximately positively correlated (Byrka et al. [13]). The formal derivation will be in the next subsection.

Recall that a set of random variables $X_1, \ldots, X_n$ is pairwise positively correlated if, for each $i, j \in [n]$, we have $\mathbb{E}[X_i X_j] \geq \mathbb{E}[X_i] \cdot \mathbb{E}[X_j]$, which for $\{0, 1\}$-valued random variables, translates to $\mathbb{P}[X_i = 1, X_j = 1] \geq \mathbb{P}[X_i = 1] \cdot \mathbb{P}[X_j = 1]$. This definition aims to capture settings where random variables are more likely to take similar values than independent random variables with the same marginals. More generally, given an integer m, $\{0, 1\}$-valued random variables $X_1, \ldots, X_n$ are called m-wise positively correlated random variables if $\mathbb{P}[X_i = 1 \; \forall i \in T] \geq \prod_{i \in T} Pr[X_i = 1]$ for all $T \subseteq [n]$ where $|T| = m$. We provide an generalized definition of positive correlation that is crucial to our analysis.

Definition 1. Given $x \in [0,1]^n$ such that $\sum_{i \in [n]} x_i = k$ for integer $k \geq 1$, let μ be a probability distribution on subsets of $[n]$ of size k. Let $X_1, \ldots, X_n$ denote the indicator random variables; thus, $X_i = 1$ if $i \in \tilde{\mathcal{S}}$ and 0 otherwise for each $i \in [n]$ where random set $\tilde{\mathcal{S}}$ of size k is sampled from μ. Then, $X_1, \ldots, X_n$ are m-wise α-positively correlated for some $\alpha \in [0,1]$ if, for each $T \subseteq [n]$ such that $|T| = m$, we have

$$\mathbb{P}[X_i = 1 \ \forall i \in T] = \mathbb{P}\left[T \subseteq \tilde{\mathcal{S}}\right] \geq \alpha^m \prod_{i \in T} x_i.$$

With a slight abuse of notation, we call the distribution μ to be *m-wise α-positively correlated* with respect to x if the condition is satisfied. Observe that, if $\alpha = 1$, then the definition implies that the random variables $X_1, \ldots, X_n$ are m-wise positively correlated.

The following lemma shows the crucial role played by m-wise approximate positively correlated distributions in the design of algorithms for D-optimal design.

Lemma 1. *Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation* (3). *Then, for any $\alpha \in (0,1]$, if there exists an efficiently computable distribution that is m-wise α-positively correlated with respect to $\widehat{x}$, then the D-optimal design problem has a randomized α-approximation algorithm: that is,*

$$\left\{ \mathbb{E}\left[\det\left(\sum_{i \in \tilde{\mathcal{S}}} a_i a_i^\top \right) \right] \right\}^{\frac{1}{m}} \geq \alpha w^*,$$

where random set $\tilde{\mathcal{S}}$ with size k is the output of the approximation algorithm.

Proof of Lemma 1 relies on the polynomial formulation of matrix determinant and convex relaxation of the D-optimal design problem. We show that a m-wise α-positively correlated distribution leads to a randomized algorithm for the D-optimal design problem that approximates each of the coefficients in the polynomial formulation. Note that the approximation ratio α in Lemma 1 only holds in the sense of expectation. Therefore, one might need to derandomize the algorithm to achieve the approximation ratio.

Before proving Lemma 1, we would like to introduce some useful results. The following lemmas follow from the Cauchy–Binet equation (Broida and Williamson [12]) and use the fact that a matrix's determinant is polynomial in entries of the matrix. Interested readers can find the proofs in Appendix A.

Lemma 2. *Suppose that $a_i \in \mathbb{R}^m$ for $i \in T$ with $|T| \geq m$; then the following identity holds:*

$$\det\left(\sum_{i \in T} a_i a_i^\top \right) = \sum_{S \in \binom{T}{m}} \det\left(\sum_{i \in S} a_i a_i^\top \right). \tag{5}$$

Proof. See Appendix A.1. □

Lemma 3. *For any $x \in [0,1]^n$, then the following identity holds:*

$$\det\left(\sum_{i \in [n]} x_i a_i a_i^\top \right) = \sum_{S \in \binom{[n]}{m}} \prod_{i \in S} x_i \det\left(\sum_{i \in S} a_i a_i^\top \right). \tag{6}$$

Proof. See Appendix A.2. □

Now, we are ready to prove Lemma 1.

Proof of Lemma 1. Note that $(\widehat{x}, \widehat{w})$ is an optimal solution to (3) and the distribution μ given by Lemma 1 for this $\widehat{x}$, which satisfies the conditions of Definition 1. We now consider the randomized algorithm that samples a random set $\tilde{\mathcal{S}}$ from this distribution μ. We show that this randomized algorithm satisfies the guarantee claimed in the lemma. All expectations and probabilities of events are under the probability measure μ, and for simplicity, we drop it from the notation.

Because $[\det(\sum_{i \in [n]} \widehat{x}_i a_i a_i^\top)]^{\frac{1}{m}}$ is at least as large as the optimal value to D-optimal design problem (3) with $B = [0,1]$, we only need to show that

$$\left\{ \mathbb{E}\left[\det\left(\sum_{i \in \tilde{\mathcal{S}}} a_i a_i^\top \right) \right] \right\}^{\frac{1}{m}} \geq \alpha \left[\det\left(\sum_{i \in [n]} \widehat{x}_i a_i a_i^\top \right) \right]^{\frac{1}{m}},$$

or equivalently,

$$\mathbb{E}\left[\det\left(\sum_{i\in\tilde{\mathcal{G}}}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right)\right]\geq\alpha^m\det\left(\sum_{i\in[n]}\widehat{x}_i\boldsymbol{a}_i\boldsymbol{a}_i^\top\right). \tag{7}$$

This indeed holds, because

$$\mathbb{E}\left[\det\left(\sum_{i\in\tilde{\mathcal{G}}}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right)\right]=\sum_{S\in\binom{[n]}{k}}\mathbb{P}\left[\tilde{\mathcal{G}}=S\right]\det\left(\sum_{i\in S}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right)=\sum_{S\in\binom{[n]}{k}}\mathbb{P}\left[\tilde{\mathcal{G}}=S\right]\sum_{T\in\binom{S}{m}}\det\left(\sum_{i\in T}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right)$$

$$=\sum_{T\in\binom{[n]}{m}}\mathbb{P}\left[T\subseteq\tilde{\mathcal{G}}\right]\det\left(\sum_{i\in T}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right)\geq\alpha^m\sum_{T\in\binom{[n]}{m}}\prod_{i\in T}\widehat{x}_i\det\left(\sum_{i\in T}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right)=\alpha^m\det\left(\sum_{i\in[n]}\widehat{x}_i\boldsymbol{a}_i\boldsymbol{a}_i^\top\right),$$

where the first equality is because of the definition of probability measure μ, the second equality is due to Lemma 3, the third equality is due to the interchange of summation, the first inequality is due to Definition 1, and the fourth equality is because of Lemma 3. $\square$

3.3. Analysis of Sampling Scheme

In this subsection, we will analyze the proposed sampling procedure (i.e., deriving the approximation ratio α of sampling procedure (4) in Lemma 1). The key idea is to derive lower bound for the ratio

$$\frac{\mathbb{P}\left[T\subseteq\tilde{\mathcal{G}}\right]}{\prod_{i\in T}\widehat{x}_i}$$

for any $T\in\binom{[n]}{m}$, where $(\widehat{x},\widehat{w})$ is an optimal solution to (3) with $B=[0,1]$.

By the definition of random set $\tilde{\mathcal{G}}$ in (4), the probability $\mathbb{P}[T\subseteq\tilde{\mathcal{G}}]$ is equal to

$$\mathbb{P}\left[T\subseteq\tilde{\mathcal{G}}\right]=\frac{\sum_{S\in\binom{[n]}{k}:T\subseteq S}\prod_{j\in S}\widehat{x}_j}{\sum_{\bar{S}\in\binom{[n]}{k}}\prod_{i\in\bar{S}}\widehat{x}_i}. \tag{8}$$

Observe that the denominator in (8) is a degree k polynomial that is invariant under any permutation of $[n]$. Moreover, the numerator is also invariant under any permutation of T as well as any permutation of $[n]\setminus T$. These observations allow us to use inequalities on symmetric polynomials and reduce the worst-case ratio of $\mathbb{P}[T\subseteq\tilde{\mathcal{G}}]$ with $\prod_{i\in T}\widehat{x}_i$ to a single-variable optimization problem as shown in the following proposition. We then analyze the single-variable optimization problem to prove the desired bound.

Proposition 1. *Let $\tilde{\mathcal{G}}$ be the random set defined in (4). Then, for any $T\subseteq[n]$ such that $|T|=m$, we have*

$$\mathbb{P}\left[T\subseteq\tilde{\mathcal{G}}\right]\geq\frac{1}{g(m,n,k)}\prod_{i\in T}\widehat{x}_i:=\alpha^m\prod_{i\in T}\widehat{x}_i,$$

where

$$g(m,n,k)=\max_y\left\{\sum_{\tau=0}^m\frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}}\frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}(y)^\tau:\frac{mk}{n}\leq y\leq m\right\}. \tag{9}$$

Proof. See Appendix A.3. $\square$

Next, we derive the upper bound of $g(m,n,k)$ in (9), which is a single-variable optimization problem. To derive the desired results, we first observe that, for any given (m,k) with $m\leq k$, $g(m,n,k)$ is monotone nondecreasing in n. This motivates us to find an upper bound on $\lim_{n\to\infty}g(m,n,k)$, which leads to Proposition 2.

Proposition 2. *For any $n\geq k\geq m$, we have*

$$\alpha^{-1}=\left[g(m,n,k)\right]^{\frac{1}{m}}\leq\lim_{\tau\to\infty}\left[g(m,\tau,k)\right]^{\frac{1}{m}}\leq\min\left\{e,1+\frac{k}{k-m+1}\right\}. \tag{10}$$

Proof. See Appendix A.4. □

Finally, we present our first approximation result below.

Theorem 1. *For any positive integers $m \le k \le n$, Algorithm 1 yields $1/e$ approximation for the D-optimal design problem.*

Proof. The result directly follows from Lemma 1 and Proposition 2 given that $n \ge k \ge m$. □

We also note that, when k is large enough, Algorithm 1 is a near-0.5 approximation.

Corollary 1. *Given $\epsilon \in (0,1)$ and positive integers $m \le k \le n$, if $k \ge (m-1)/(2\epsilon)$, then Algorithm 1 yields $(0.5 - \epsilon)$ approximation for the D-optimal design problem.*

Proof. For any $\epsilon \in (0,1)$, from Proposition 2, let the lower-bound approximation ratio $\alpha \ge [\min\{e, 1 + k/(k-m+1)\}]^{-1} \ge 0.5 - \epsilon$, or equivalently, let

$$1 + \frac{k}{k-m+1} \le \frac{1}{0.5 - \epsilon}.$$

Then, the conclusion follows. □

3.4. Deterministic Implementation

The approximation ratios presented in the previous subsection only hold in the sense of expectation. In this subsection, we will overcome this issue and present a deterministic Algorithm 2 with the same approximation guarantees. The key idea is to derandomize Algorithm 2 using the method of conditional expectation (compare with Spencer and Spencer [36]), and the main challenge is how to compute the conditional expectation efficiently. We next will show that it can be done by evaluating a determinant of a $n \times n$ matrix in which entries are linear polynomials in three indeterminates.

In this deterministic Algorithm 2, suppose that we are given a subset $S \subseteq [n]$ such that $|S| = s \le k$. Then, the expectation of mth power of objective function given S is

$$
\begin{aligned}
H(S) &:= \mathbb{E}\left[\det\left(\sum_{i \in \tilde{\mathcal{S}}} a_i a_i^\top\right) \middle| S \subseteq \tilde{\mathcal{S}}\right] \\
&= \sum_{U \in \binom{[n]\backslash S}{k-s}} \frac{\prod_{j \in U} \widehat{x}_j}{\sum_{\bar{U} \in \binom{[n]\backslash S}{k-s}} \prod_{i \in \bar{U}} \widehat{x}_i} \det\left(\sum_{i \in U} a_i a_i^\top + \sum_{i \in S} a_i a_i^\top\right) \\
&= \left(\sum_{\bar{U} \in \binom{[n]\backslash S}{k-s}} \prod_{i \in \bar{U}} \widehat{x}_i\right)^{-1} \sum_{U \in \binom{[n]\backslash S}{k-s}} \prod_{j \in U} \widehat{x}_j \sum_{R \in \binom{U \cup S}{m}} \det\left(\sum_{i \in R} a_i a_i^\top\right) \\
&= \frac{\sum_{R \in \binom{[n]}{m}, r := |R \backslash S| \le k-s} \prod_{j \in R \backslash S} \widehat{x}_j \det\left(\sum_{i \in R} a_i a_i^\top\right) \sum_{W \in \binom{[n]\backslash(S \cup R)}{k-s-r}} \prod_{j \in W} \widehat{x}_j}{\sum_{\bar{U} \in \binom{[n]\backslash S}{k-s}} \prod_{i \in \bar{U}} \widehat{x}_i},
\end{aligned}
\tag{11}
$$

where the second equality is a direct computing of the conditional probability, the third equality is due to Lemma 2, and the last one is because of interchange of summation.

Algorithm 2 Derandomization of Algorithm 1

1. Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = [0,1]$, where $\widehat{x} \in [0,1]^n$ with $\sum_{i \in [n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$
2. Initialize chosen set $S = \emptyset$
3. **do**
4.   Let $j^* \in \arg\max_{j \in [n]\backslash \tilde{\mathcal{S}}} H(S \cup j)$, where $H(S \cup j)$ defines (11), and its denominator and numerator can be computed by Observation 1 and Proposition 3, respectively
5.   Add j^* to set $\tilde{\mathcal{S}}$
6. **while** $|S| < k$
7. Output S

Note that the denominator in (11) can be computed efficiently according to Observation 1. Next, we show that the numerator in (11) can be also computed efficiently.

Proposition 3. *Let matrix $A = [a_1, \ldots, a_n]$. Consider the following function:*

$$F(t_1, t_2, t_3) = \det\left(I_n + t_1 \operatorname{diag}(y)^{\frac{1}{2}} A^\top A \operatorname{diag}(y)^{\frac{1}{2}} + \operatorname{diag}(y)\right), \tag{12}$$

where $t_1, t_2, t_3 \in \mathbb{R}, y \in \mathbb{R}^n$ are indeterminate and

$$y_i = \begin{cases} t_3, & \text{if } i \in S \\ \widehat{x}_i t_2, & \text{otherwise.} \end{cases}$$

Then, the coefficient of $t_1^m t_2^{k-s} t_3^s$ in $F(t_1, t_2, t_3)$ equals to

$$\sum_{R \in \binom{[n]}{m}, r:=|R \setminus S| \leq k-s} \prod_{j \in R \setminus S} \widehat{x}_j \det\left(\sum_{i \in R} a_i a_i^\top\right) \sum_{W \in \binom{[n] \setminus (S \cup R)}{k-s-r}} \prod_{j \in W} \widehat{x}_j. \tag{13}$$

Proof. First of all, we can rewrite $F(t_1, t_2, t_3)$ as

$$F(t_1, t_2, t_3) = \det(I_n + \operatorname{diag}(y)) \det\left(I_n + t_1 \operatorname{diag}(e + y)^{-\frac{1}{2}} \operatorname{diag}(y)^{\frac{1}{2}} A^\top A \operatorname{diag}(y)^{\frac{1}{2}} \operatorname{diag}(e + y)^{-\frac{1}{2}}\right)$$

$$= \prod_{i \in S}(1 + t_3) \prod_{i \in [n] \setminus S}\left(1 + \widehat{x}_i t_2\right) \det\left(I_n + t_1 W^\top W\right),$$

where the ith column of matrix W is

$$W_i = \begin{cases} \sqrt{\dfrac{t_3}{1 + t_3}} a_i, & \text{if } i \in S \\[2ex] \sqrt{\dfrac{\widehat{x}_i t_2}{1 + \widehat{x}_i t_2}} a_i, & \text{otherwise.} \end{cases}$$

Note that the coefficient of t_1^m in $\det(I_n + t_1 W^\top W)$ is equal to the one of $\prod_{i \in [n]}(1 + t_1 \Lambda_i)$, where $\{\Lambda_i\}_{i \in [n]}$ are the eigenvalues of $W^\top W$. Thus, the coefficient of t_1^m is

$$\sum_{R \in \binom{[n]}{m}} \prod_{i \in R} \Lambda_i = \sum_{R \in \binom{[n]}{m}} \det((W^\top W)_{R,R}) = \sum_{R \in \binom{[n]}{m}} \det\left(\sum_{i \in R} W_i W_i^\top\right) = \sum_{R \in \binom{[n]}{m}} \det\left(\sum_{i \in R} W_i W_i^\top\right),$$

where P_{R_1, R_2} denotes a submatrix of P with rows and columns from sets R_1, R_2, the first equality is due to the property of the eigenvalues (Horn and Johnson [20, theorem 1.2.12]), the second inequality is because the length of each column of W is m, and the third equality is because the determinant of singular matrix is zero.

Therefore, the coefficient of $t_1^m t_2^{k-s} t_3^s$ in $F(t_1, t_2, t_3)$ is equivalent to the one of

$$\prod_{i \in S}(1 + t_3) \prod_{i \in [n] \setminus S}\left(1 + \widehat{x}_i t_2\right) \sum_{R \in \binom{[n]}{m}} \det\left(\sum_{i \in R} W_i W_i^\top\right).$$

By Lemma 3 with $n = m$ and the definition of matrix W, the coefficient of $t_1^m t_2^{k-s} t_3^s$ in $F(t_1, t_2, t_3)$ is further equivalent to the one of

$$\prod_{i \in S}(1 + t_3) \prod_{i \in [n] \setminus S}\left(1 + \widehat{x}_i t_2\right) \sum_{R \in \binom{[n]}{m}} t_1^m \prod_{j \in R \setminus S}\frac{\widehat{x}_j}{1 + t_2 \widehat{x}_j} \prod_{j \in R \cap S}\frac{t_3}{1 + t_3} \det\left(\sum_{i \in R} a_i a_i^\top\right)$$

$$= t_1^m \sum_{R \in \binom{[n]}{m}} t_2^{|R \setminus S|} t_3^{|R \cap S|}(1 + t_3)^{|S \setminus R|} \prod_{i \in [n] \setminus (S \cup R)}\left(1 + \widehat{x}_i t_2\right) \prod_{j \in R \setminus S} \widehat{x}_j \det\left(\sum_{i \in R} a_i a_i^\top\right),$$

which is equal to (13) by collecting coefficients of $t_1^m t_2^{k-s} t_3^s$. $\square$

Note that the characteristic function of a matrix can be computed efficiently by using the Faddeev–LeVerrier algorithm (Hou [21]). Thus, the polynomial $F(t_1, t_2, t_3)$ is efficiently computable with time complexity of $O(n^4)$.

Algorithm 2 proceeds as follows. We start with an empty subset S of chosen elements, and for each $j \notin S$, we compute the expected mth power of objective function that j will be chosen $H(S \cup j)$. We add j^* to S, where

$j^* \in \arg\max_{j \in [n] \setminus S} H(S \cup j)$. Then, go to next iteration. This procedure will terminate if $|S| = k$. Note that Algorithm 2 requires $O(nk)$ evaluations of function $H(S \cup j)$; thus, the time complexity is $O(n^5 k)$. Hence, in practice, we recommend Algorithm 1 because of its shorter running time.

The approximation results for Algorithm 2 are identical to Theorem 1 and Corollary 1, which are summarized as follows.

Theorem 2. *For any positive integers $m \le k \le n$,*

  i. *deterministic Algorithm 2 is efficiently computable and yields $1/e$ approximation for the D-optimal design problem; and*

  ii. *given $\epsilon \in (0,1)$, if $k \ge (m-1)/(2\epsilon)$, then deterministic Algorithm 2 yields $(0.5 - \epsilon)$ approximation for the D-optimal design problem.*

4. Improving Approximation Bound in Asymptotic Regime

In this section, we propose another sampling Algorithm 3, which achieves asymptotic optimality (i.e., the output of Algorithm 3 is close to optimal when $k/m \to \infty$). We also show the derandomization of Algorithm 3. Similar to the previous section, in this section, we consider the D-optimal design problem without repetitions (i.e., in (2) and (3), we let $B = [0,1]$).

In Algorithm 3, suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = [0,1]$, $\epsilon \in (0,1)$ is a positive threshold, and $\mathcal{N}$ is a random permutation of $[n]$. Then, for each $j \in \mathcal{N}$, we select j with probability $x_j/(1+\epsilon)$, and let $\tilde{\mathcal{S}}$ be the set of selected elements. If $|\tilde{\mathcal{S}}| < k$, then we can add $k - |\tilde{\mathcal{S}}|$ more elements from $[n] \setminus \tilde{\mathcal{S}}$. However, if $|\tilde{\mathcal{S}}| > k$, then we repeat the sampling procedure. Algorithm 3 has time complexity $O(n)$. In addition, note that the difference between Algorithms 1 and 3 is that, in Algorithm 3, we inflate the probability of choosing jth experiment by $1/(1+\epsilon)$. This condition guarantees that, when $k \gg m$, according to concentration inequality, the probability of size m subset T to be chosen will be nearly equal to $\prod_{j \in T}[x_j/(1+\epsilon)]$.

Algorithm 3 Asymptotic Sampling Algorithm
1. Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = [0,1]$, where $\widehat{x} \in [0,1]^n$ with
    $\sum_{i \in [n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$
2. Initialize $\tilde{\mathcal{S}} = \emptyset$ and a positive number $\epsilon > 0$
3. **do**
4.   Let set $\mathcal{N}$ be a random permutation set of $\{1, \dots, n\}$
5.   **for** $j \in \mathcal{N}$ **do**
6.   Sample a $(0,1)$ uniform random variable U
7.   **if** $U \le \widehat{x}/(1+\epsilon)$ **then**
8.   Add j to set $\tilde{\mathcal{S}}$
9.   **end if**
10.   **end for**
11. **while** $|\tilde{\mathcal{S}}| > k$
12. **if** $|\tilde{\mathcal{S}}| < k$ **then**   ▷ Greedy step to enforce $|\tilde{\mathcal{S}}| = k$
13.   Let $j^* \in \arg\max_{j \in [n] \setminus \tilde{\mathcal{S}}} [\det(\sum_{i \in \tilde{\mathcal{S}}} a_i a_i^\top + a_j a_j^\top)]^{\frac{1}{m}}$
14.   Add j^* to set $\tilde{\mathcal{S}}$
15. **end if**
16. Output $\tilde{\mathcal{S}}$

4.1. Analysis of Sampling Algorithm 3

To analyze sampling Algorithm 3, we first show the following probability bound. The key idea is to prove the lower bound $1/\prod_{i \in T} \widehat{x}_i \mathbb{P}\{T \subseteq \tilde{\mathcal{S}} \mid |\tilde{\mathcal{S}}| \le k\}$ by using the Chernoff bound [15].

Lemma 4. *Let $\epsilon > 0$ and $\tilde{\mathcal{S}} \subseteq [n]$ be a random set output from Algorithm 3. Given $T \subseteq [n]$ with $|T| = m \le n$, then we have*

$$\alpha^m := \frac{1}{\prod_{i \in T} \widehat{x}_i} \mathbb{P}\left\{ T \subseteq \tilde{\mathcal{S}} \,\middle|\, |\tilde{\mathcal{S}}| \le k \right\} \ge (1+\epsilon)^{-m}\left(1 - e^{-\frac{(\epsilon k - (1+\epsilon)m)^2}{k(2+\epsilon)(1+\epsilon)}}\right), \tag{14}$$

where α is in Definition 1. In addition, when $k \ge 4m/\epsilon + 12/\epsilon^2 \log(1/\epsilon)$, then,

$$\alpha^m \ge (1-\epsilon)^m. \tag{15}$$

Proof. We note that $\tilde{\mathcal{G}} \subseteq [n]$ is a random set, where each $i \in [n]$ is independently sampled according to Bernoulli random variable X_i with the probability of success $\hat{x}_i/(1 + \epsilon)$. According to Definition 1 and by ignoring the greedy procedure Algorithm 3, it is sufficient to derive the lower bound of $1/\prod_{i \in T} \hat{x}_i \mathbb{P}\{T \subseteq \tilde{\mathcal{G}} || \tilde{\mathcal{G}}| \leq k\}$: that is,

$$\frac{\mathbb{P}\left\{T \subseteq \tilde{\mathcal{G}} \middle| |\tilde{\mathcal{G}}| \leq k\right\}}{\prod_{i \in T} \hat{x}_i} = \frac{\mathbb{P}\left\{T \subseteq \tilde{\mathcal{G}}, |\tilde{\mathcal{G}}| \leq k\right\}}{\prod_{i \in T} \hat{x}_i \mathbb{P}\left\{|\tilde{\mathcal{G}}| \leq k\right\}} \geq (1 + \epsilon)^{-m} \mathbb{P}\left\{\sum_{i \in [n] \setminus T} X_i \leq k - m\right\},$$

where the first inequality is due to $\mathbb{P}\{|\tilde{\mathcal{G}}| \leq k\} \leq 1$.

Therefore, it is sufficient to bound the following probability

$$\mathbb{P}\left\{\sum_{i \in [n] \setminus T} X_i \leq k - m\right\}.$$

Because $X_i \in \{0, 1\}$ for each $i \in [n]$, $\mathbb{E}[\sum_{i \in [n] \setminus T} X_i] = 1/(1 + \epsilon) \sum_{i \in [n] \setminus T} \hat{x}_i$. According to the Chernoff bound [15], we have

$$\mathbb{P}\left\{\sum_{i \in [n] \setminus T} X_i > (1 + \bar{\epsilon}) \mathbb{E}\left[\sum_{i \in [n] \setminus T} X_i\right]\right\} \leq e^{-\frac{\bar{\epsilon}^2}{2 + \bar{\epsilon}} \mathbb{E}\left[\sum_{i \in [n] \setminus T} X_i\right]}.$$

Here, $\bar{\epsilon}$ is a positive constant. Therefore, by choosing $\bar{\epsilon} = (1 + \epsilon)(k - m)/(\sum_{i \in [n] \setminus T} \hat{x}_i) - 1$, we have

$$(1 + \epsilon)^{-m} \mathbb{P}\left\{\sum_{i \in [n] \setminus T} X_i \leq k - m\right\} \geq (1 + \epsilon)^{-m}\left(1 - e^{-\frac{\bar{\epsilon}^2 \sum_{i \in [n] \setminus T} \hat{x}_i}{(2 + \bar{\epsilon})(1 + \epsilon)}}\right). \tag{16}$$

Note that $k - m \leq \sum_{i \in [n] \setminus T} X_i \leq k$, $\epsilon - (1 + \epsilon)m/k \leq \bar{\epsilon} \leq \epsilon$, and $\epsilon k - (1 + \epsilon)m \leq \bar{\epsilon} \sum_{i \in [n] \setminus T} \hat{x}_i \leq \epsilon(k - m)$. Suppose that $k \geq (1/\epsilon + 1)m$; then, the left-hand side of (16) can be further lower bounded as

$$(1 + \epsilon)^{-m}\left(1 - e^{-\frac{\bar{\epsilon}^2 \sum_{i \in [n] \setminus T} \hat{x}_i}{(2 + \bar{\epsilon})(1 + \epsilon)}}\right) \geq (1 + \epsilon)^{-m}\left(1 - e^{-\frac{(\epsilon k - (1 + \epsilon)m)^2}{k(2 + \epsilon)(1 + \epsilon)}}\right).$$

To prove (15), it remains to show

$$1 - e^{-\frac{(\epsilon k - (1 + \epsilon)m)^2}{k(2 + \epsilon)(1 + \epsilon)}} \geq ((1 - \epsilon)(1 + \epsilon))^m,$$

or equivalently,

$$\log\left[1 - \left(1 - \epsilon^2\right)^m\right] \geq -\frac{(\epsilon k - (1 + \epsilon)m)^2}{k(2 + \epsilon)(1 + \epsilon)}, \tag{17}$$

which holds if

$$k \geq \left(1 + \frac{1}{\epsilon}\right)\left(2m - \left(1 + \frac{2}{\epsilon}\right)\log\left[1 - \left(1 - \epsilon^2\right)^m\right]\right).$$

We note that $-\log[1 - (1 - \epsilon^2)^m]$ is nonincreasing over $m \geq 1$; therefore, it is upper bounded by $2\log(1/\epsilon)$. Hence, (17) holds if $k \geq 4m/\epsilon + 12/\epsilon^2 \log(1/\epsilon)$. □

Finally, we state our main approximation result.

Theorem 3. *For any integers $m \leq k \leq n$ and $\epsilon \in (0, 1)$, if $k \geq 4m/\epsilon + 12/\epsilon^2 \log(1/\epsilon)$, then Algorithm 3 is a $(1 - \epsilon)$ approximation for the D-optimal design problem.*

Proof. The result directly follows from Lemmas 1 and 4. □

4.2. Deterministic Implementation
Similar to Section 3.4, the approximation ratios presented in the previous subsection only hold in the sense of expectation. In this subsection, we will overcome this issue and present a deterministic Algorithm 3 with the same approximation guarantees.

In this deterministic Algorithm 4, S denotes a subset such that $|S| = s \le k$. Then, the expectation of mth power of objective function given S is

$$
\begin{aligned}
H(S) &:= \mathbb{E}\left[\det\left(\sum_{i\in\tilde{\mathcal{S}}} a_i a_i^\top\right)\middle|\, S \subseteq \tilde{\mathcal{S}}, |\tilde{\mathcal{S}}| \le k\right] \\
&= \frac{\sum_{\kappa=s}^{k} \sum_{U\in\binom{[n]\backslash S}{\kappa-s}} \prod_{j\in U} \frac{\hat{x}_i}{1+\epsilon} \prod_{j\in[n]\backslash(S\cup U)}\left(1 - \frac{\hat{x}_i}{1+\epsilon}\right) \det\left(\sum_{i\in U} a_i a_i^\top + \sum_{i\in S} a_i a_i^\top\right)}{\sum_{\kappa=1}^{k} \sum_{\bar{U}\in\binom{[n]\backslash S}{\kappa-s}} \prod_{i\in\bar{U}} \frac{\hat{x}_i}{1+\epsilon} \prod_{j\in[n]\backslash(S\cup\bar{U})}\left(1 - \frac{\hat{x}_i}{1+\epsilon}\right)} \\
&= \left(\sum_{\kappa=s}^{k} \sum_{\bar{U}\in\binom{[n]\backslash S}{\kappa-s}} \prod_{i\in\bar{U}} \frac{\hat{x}_i}{1+\epsilon-\hat{x}_i}\right)^{-1} \sum_{\kappa=s}^{k} \sum_{U\in\binom{[n]\backslash S}{\kappa-s}} \prod_{j\in U} \frac{\hat{x}_i}{1+\epsilon-\hat{x}_i} \sum_{R\in\binom{U\cup S}{m}} \det\left(\sum_{i\in R} a_i a_i^\top\right) \\
&= \frac{\sum_{\kappa=s}^{k} \sum_{R\in\binom{[n]}{m}, r:=|R\backslash S|\le\kappa-s} \prod_{j\in R\backslash S} \frac{\hat{x}_j}{1+\epsilon-\hat{x}_j} \det\left(\sum_{i\in R} a_i a_i^\top\right) \sum_{W\in\binom{[n]\backslash(S\cup R)}{\kappa-s-r}} \prod_{j\in W} \frac{\hat{x}_j}{1+\epsilon-\hat{x}_j}}{\sum_{\kappa=s}^{k} \sum_{\bar{U}\in\binom{[n]\backslash S}{\kappa-s}} \prod_{i\in\bar{U}} \frac{\hat{x}_i}{1+\epsilon-\hat{x}_i}},
\end{aligned}
\tag{18}
$$

where the second equality is a direct computing of the conditional probability, the third equality is due to Lemma 2 (dividing both denominator and numerator by $\prod_{j\in[n]\backslash S}(1 - \hat{x}_i/(1+\epsilon))$) and the convention, set $\binom{S}{\tau} = \emptyset$ if $\tau > |S|$ or $\tau < 0$), and the fourth one is because of interchange of summation.

Note that the κth entry with $\kappa \in \{s,\ldots,k\}$ of the denominator in (18) can be computed efficiently according to Observation 1 by letting $x_i := \hat{x}_i/(1+\epsilon-\hat{x}_i)$. Meanwhile, κth entry with $\kappa \in \{s,\ldots,k\}$ of the numerator in (11) can be also computed efficiently by Proposition 3 by letting $\hat{x}_i := \hat{x}_i/(1+\epsilon-\hat{x}_i)$. Therefore, the conditional expectation in (11) is efficiently computable.

In summary, Algorithm 4 proceeds as follows. We start with an empty subset S of chosen elements, and for each $j \notin S$, we compute the expected mth power of objective function that j will be chosen (i.e., $H(S \cup j)$). We update $S := S \cup \{j^*\}$, where $j^* \in \arg\max_{j\in[n]\backslash S} H(S \cup j)$. Then, go to the next iteration. This procedure will terminate if $|S| = k$. Similar to Algorithm 2, the time complexity of Algorithm 4 is $O(n^5 k^2)$. Hence, in practice, we recommend the more efficient Algorithm 3.

The approximation result for Algorithm 4 is identical to Theorem 3, which is summarized as follows.

Theorem 4. *For any positive integers $m \le k \le n$ and $\epsilon \in (0,1)$, deterministic Algorithm 4 is efficiently computable and yields $1 - \epsilon$ approximation for the D-optimal design problem if $k \ge 4m/\epsilon + 12/\epsilon^2 \log(1/\epsilon)$.*

Algorithm 4 Derandomization of Algorithm 3

1. Suppose that $(\hat{x}, \hat{w})$ is an optimal solution to the convex relaxation (3) with $B = [0,1]$, where $\hat{x} \in [0,1]^n$ with $\sum_{i\in[n]} \hat{x}_i = k$ and $\hat{w} = f(\hat{x})$
2. Initialize chosen set $S = \emptyset$
3. **do**
4.   Let $j^* \in \arg\max_{j\in[n]\backslash\tilde{\mathcal{S}}} H(S \cup j)$, where $H(S \cup j)$ defines (18), and all of the entries of its denominator and numerator can be computed by Observation 1 by letting $x_i := \hat{x}_i/(1+\epsilon-\hat{x}_i)$ and by Proposition 3 by letting $\hat{x}_i := \hat{x}_i/1+\epsilon-\hat{x}_i$, respectively
5.   Add j^* to set S
6. **while** $|S| < k$
7. Output S

5. Approximation Algorithm for D-optimal Design Problem with Repetitions

In this section, we consider the D-optimal design problem with repetitions: that is, we let $B = \mathbb{R}_+$ in (2) and (3). We will propose a new analysis of the algorithm proposed by Nikolov [28], derive its approximation ratio, and show its deterministic implementation. Again, in this section, we also let $(\hat{x}, \hat{w})$ be an optimal solution to the convex relaxation (3), where $\hat{x} \in \mathbb{R}_+^n$ with $\sum_{i\in[n]} \hat{x}_i = k$ and $\hat{w} = f(\hat{x})$. Because the set of all nonnegative rational vectors is dense in the set of all nonnegative real vectors, thus without loss of generality, we assume that $\hat{x}$ is a nonnegative rational vector (i.e., $\hat{x} \in \mathbb{Q}_+^n$).

In Nikolov [28], the author suggested obtaining the k sample set $\tilde{\mathcal{S}}$ with replacement (i.e., $\tilde{\mathcal{S}}$ can be a multiset). The sampling procedure can be separated into k steps. At each step, a sample s is selected with probability $\mathbb{P}\{s = i\} = \hat{x}_i/k$ (note that $\hat{x} \in \mathbb{R}_+^n$ with $\sum_{i\in[n]} \hat{x}_i = k$). The detailed description is in Algorithm 5.

This sampling procedure can be interpreted as follows: let $\{X_i\}_{i\in[n]}$ be independent Poisson random variables where X_i has arrival rate $\widehat{x}_i$. We note that, conditioning on total number of arrivals equal to k (i.e., $\sum_{i\in[n]} X_i = k$), the distribution of $\{X_i\}_{i\in[n]}$ is multinomial (compare with Albert and Denis [1]), where there are k trials and the probability of ith entry to be chosen is $\widehat{x}_i/k$. We terminate this sampling procedure if the total number of arrivals equals to k. Note that the time complexity of Algorithm 5 is $O(n)$.

Algorithm 5 Sampling Algorithm for *D*-optimal Design with Repetitions
1. Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = \mathbb{R}_+$, where $\widehat{x} \in \mathbb{Q}_+^n$ with
    $\sum_{i\in[n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$
2. Initialize chosen multiset $\tilde{\mathcal{S}} = \emptyset$
3. **for** $j = 1, \dots, k$ **do**
4.   Sample s from $[n]$ with probability $\mathbb{P}\{s = i\} = \widehat{x}_i/k$
5.   Let $\tilde{\mathcal{S}} = \tilde{\mathcal{S}} \cup \{s\}$
6. **end for**
7. Output $\tilde{\mathcal{S}}$

To analyze Algorithm 5, let us consider another Algorithm 6, which turns out to be arbitrarily close to Algorithm 5. Because $\widehat{x}$ is a nonnegative rational vector (i.e., $\widehat{x} \in \mathbb{Q}_+^n$), we let q be a common multiple of the denominators of rational numbers $\widehat{x}_1, \dots, \widehat{x}_n$ (i.e., $q\widehat{x}_1, \dots, q\widehat{x}_n \in \mathbb{Z}_+$). Next, we create a multiset $\mathcal{A}_q$, which contains $q\widehat{x}_i$ copies of index i for each $i \in [n]$ (i.e., $|\mathcal{A}_q| = qk$). Finally, we sample a subset $\tilde{\mathcal{S}}_q$ of k items from set $\mathcal{A}_q$ uniformly (i.e., with probability $\binom{qk}{k}^{-1}$). The detailed description is in Algorithm 6. In this case, the sampling procedure has the following interpretation. Because sum of i.i.d. Bernoulli random variables is binomial, we let $\{X_i'\}_{i\in[n]}$ be independent binomial random variables where X_i' has number of trials $q\widehat{x}_i$ and probability of success $1/q$ for each $i \in [n]$. We terminate the sampling procedure if the total number of succeeded trials equals to k.

The following lemma shows that the probability distributions of outputs of Algorithms 5 and 6 can be arbitrarily close.

Lemma 5. *Let $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}_q$ be outputs of Algorithms 5 and 6, respectively. Then, $\tilde{\mathcal{S}}_q \xrightarrow{\mu} \tilde{\mathcal{S}}$ (i.e., the probability distribution of $\tilde{\mathcal{S}}_q$ converges to $\tilde{\mathcal{S}}$ as $q \to \infty$).*

Proof. Consider two classes of independent random variables $\{X_i\}_{i\in[n]}$, $\{X_i'\}_{i\in[n]}$, where X_i is Poisson random variable with arrival rate $\widehat{x}_i$ for each $i \in [n]$ and X_i' is binomial random variable with number of trials $q\widehat{x}_i$ and probability of success $1/q$ for each $i \in [n]$, respectively.

Given a size k multiset $\mathcal{R}$ with support $\text{Supp}\mathcal{R} \subseteq [n]$ and $M_{\mathcal{R}}(i)$ denoting the number of occurrences of i in $\mathcal{R}$, according to the description of Algorithm 6, we have

$$\mathbb{P}\{\tilde{\mathcal{S}} = \mathcal{R}\} = \mathbb{P}\left\{X_i = M_{\mathcal{R}}(i), \forall i \in [n] \,\middle|\, \sum_{i\in[n]} X_i = k\right\} = \frac{\mathbb{P}\{X_i = M_{\mathcal{R}}(i), \forall i \in [n], \sum_{i\in[n]} X_i = k\}}{\mathbb{P}\{\sum_{i\in[n]} X_i = k\}}$$

$$= \mathbb{I}(|\mathcal{R}| = k)\frac{\prod_{i\in[n]} \mathbb{P}\{X_i = M_{\mathcal{R}}(i)\}}{\mathbb{P}\{\sum_{i\in[n]} X_i = k\}},$$

where the first equality is from the description of Algorithm 5, the second equality is by the definition of conditional probability, the third equality is because $\{X_i\}_{i\in[n]}$ are independent from each other, and $\mathbb{I}(\cdot)$ denotes indicator function. Similarly, we also have

$$\mathbb{P}\{\tilde{\mathcal{S}}_q = \mathcal{R}\} = \mathbb{I}(|\mathcal{R}| = k)\frac{\prod_{i\in[n]} \mathbb{P}\{X_i' = M_{\mathcal{R}}(i)\}}{\mathbb{P}\{\sum_{i\in[n]} X_i' = k\}}.$$

Followed by the well-known Poisson limit theorem (compare with Papoulis and Unnikrishna Pillai [31]), X_i and X_i' have the same distribution as $q \to \infty$ for any $i \in [n]$. Therefore,

$$\mathbb{P}\{\tilde{\mathcal{S}}_q = \mathcal{R}\} \to \mathbb{P}\{\tilde{\mathcal{S}} = \mathcal{R}\},$$

when $q \to \infty$ (that is, the outputs of Algorithms 5 and 6 have the same distribution when $q \to \infty$). $\square$

Algorithm 6 Approximation of Algorithm 5
1. Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = \mathbb{R}^+$, where $\widehat{x} \in \mathbb{Q}_+^n$ with $\sum_{i\in[n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$
2. Let q be a common multiple of the denominators of rational numbers $\widehat{x}_1, \ldots, \widehat{x}_n$ (i.e., $q\widehat{x}_1, \ldots, q\widehat{x}_n \in \mathbb{Z}_+$)
3. Duplicate $q\widehat{x}_i$ copies of index i for each $i \in [n]$ as set $\mathcal{A}_q$ (i.e., $|\mathcal{A}_q| = qk$)
4. Sample a subset $\tilde{\mathcal{S}}_q$ of k items from set $\mathcal{A}_q$ with probability $\binom{qk}{k}^{-1}$
5. Output $\tilde{\mathcal{S}}_q$

Now, we are ready to present our approximation results for Algorithm 5. The proof idea is based on Lemma 5 (that is, we first analyze Algorithm 6 and apply its result to Algorithm 5 by letting $q \to \infty$).

Proposition 4. *Let $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}_q$ be outputs of Algorithms 5 and 6, respectively. Then, we have*

$$\left(\mathbb{E}\left[\left(f\left(\tilde{\mathcal{S}}_q\right)\right)^m\right]\right)^{\frac{1}{m}} \geq \left(\mathbb{E}\left[\left(f\left(\tilde{\mathcal{S}}\right)\right)^m\right]\right)^{\frac{1}{m}} \geq g(m,k)^{-1} w^*,$$

where

$$g(m,k) = \left[\frac{(k-m)!k^m}{k!}\right]^{\frac{1}{m}} \leq \min\left\{e, \frac{k}{k-m+1}\right\}. \tag{19}$$

Proof. We will first show the approximation ratio of Algorithm 6 and then, apply it to Algorithm 5 by Lemma 5 when $q \to \infty$.

i. Let $(\bar{x}_q', \bar{w}_q')$ be output of Algorithm 6. Similar to Proof of Theorem 1, we have

$$\mathbb{E}\left[\left(\bar{w}_q'\right)^m\right] = \sum_{S \in \binom{\mathcal{A}_q}{k}} \frac{1}{\binom{qk}{k}} \det\left(\sum_{i\in S} a_i a_i^\top\right) = \frac{q^m}{\binom{qk}{k}} \sum_{S \in \binom{\mathcal{A}_q}{k}} \frac{1}{q^m} \det\left(\sum_{i\in S} a_i a_i^\top\right)$$

$$= \frac{q^m}{\binom{qk}{k}} \sum_{S \in \binom{\mathcal{A}_q}{k}} \frac{1}{q^m} \sum_{T \in \binom{S}{m}} \det\left(\sum_{i\in T} a_i a_i^\top\right) = \frac{q^m \binom{qk-m}{k-m}}{\binom{qk}{k}} \sum_{T \in \binom{\mathcal{A}_q}{m}} \frac{1}{q^m} \det\left(\sum_{i\in T} a_i a_i^\top\right)$$

$$= \frac{q^m \binom{qk-m}{k-m}}{\binom{qk}{k}} \det\left(\sum_{i\in[n]} \widehat{x}_i a_i a_i^\top\right) \geq \frac{k!}{(k-m)!k^m}\left[f(\widehat{x})\right]^m \geq \frac{k!}{(k-m)!k^m}\left(w^*\right)^m,$$

where the first and second equalities are due to Algorithm 6, the third equality is because of Lemma 2 and $k \geq m$, the fourth equality is due to interchange of summation, the fifth equality is because of the identity $\sum_{i\in[n]} \widehat{x}_i a_i a_i^\top = \sum_{i\in\mathcal{A}_q} \frac{1}{q} a_i a_i^\top$, the first inequality holds because $(qk)^m(qk-m)!/(qk)! \geq 1$, and the last inequality is because $\widehat{x}$ is an optimal solution of the continuous relaxation.

ii. From Lemma 5, we know that the output of Algorithm 6 has the same probability distribution as the output of Algorithm 5 when $q \to \infty$. Thus, we have

$$\mathbb{E}[(\bar{w})^m] = \lim_{q\to\infty} \mathbb{E}\left[\left(\bar{w}_q'\right)^m\right] = \lim_{q\to\infty} \frac{q^m \binom{qk-m}{k-m}}{\binom{qk}{k}} \det\left(\sum_{i\in[n]} \widehat{x}_i a_i a_i^\top\right)$$

$$= \frac{k!}{(k-m)!k^m} \det\left(\sum_{i\in[n]} \widehat{x}_i a_i a_i^\top\right) = \frac{k!}{(k-m)!k^m}\left[f(\widehat{x})\right]^m.$$

iii. Next, let

$$g(m,k) = \left[\frac{(k-m)!k^m}{k!}\right]^{\frac{1}{m}},$$

and we would like to investigate its bound.
First note that the function

$$\log\left(\frac{g(m,k+1)}{g(m,k)}\right) = m\log\left(1 + \frac{1}{k}\right) + \log\left(1 - \frac{m}{k+1}\right),$$

which is nondecreasing over $k \in [m, \infty)$. Thus, we have

$$\log\left(\frac{g(m, k+1)}{g(m, k)}\right) \leq \lim_{k' \to \infty} \log\left(\frac{g(m, k'+1)}{g(m, k')}\right) = 0,$$

that is, $g(m, k) \leq g(m, m) = [m^m/m!]^{\frac{1}{m}} \leq e$.

However, because $(k - m)!/k! \leq 1/(k - m + 1)^m$, thus, we have

$$g(m, k) \leq \left[\frac{k^m}{(k - m + 1)^m}\right]^{\frac{1}{m}} = \frac{k}{k - m + 1}.$$

Hence, $g(m, k) \leq \min\{e, k/(k - m + 1)\}$. □

From Proposition 4, we note that, when k is large enough, the output of Algorithm 5 is almost optimal. Finally, we present our approximation results blow.

Theorem 5. *For any positive integers $m \leq k \leq n$,*
   i. *both Algorithms 5 and 6 yield $1/e$ approximation for the D-optimal design problem with repetitions and*
   ii. *given any $\epsilon \in (0, 1)$, if $k \geq (m - 1)/\epsilon$, then both Algorithms 5 and 6 yield $(1 - \epsilon)$ approximation for the D-optimal design problem with repetitions.*

Proof. The first result directly follows from Proposition 4. For the second one, given $\epsilon \in (0, 1)$, by Proposition 4, let

$$\frac{k - m + 1}{k} \geq 1 - \epsilon.$$

Then, the conclusion follows by letting $k \geq (m - 1)/\epsilon$. □

To conclude this part, we remark that, although the results from previous sections hold for *D*-optimal design with repetitions as well, Algorithm 5 has tighter approximation ratios. Therefore, investigating the convex relaxation solution and approximation algorithms of *D*-optimal design with repetitions alone does help us improve the approximation bounds.

5.1. Deterministic Implementation

Similar to Section 3.4, the approximation ratios presented in the previous subsection hold in the sense of expectation. Recall that $(\widehat{x}, \widehat{w})$ is an optimal solution of (3) with $\widehat{x} \in \mathbb{Q}_+^n$, q is a common multiple of the denominators $\widehat{x}_1, \ldots, \widehat{x}_n$, and multiset $\mathcal{A}_q$ of size qk contains $q\widehat{x}_i$ copies of index i for each $i \in [n]$. In this subsection, we will show that the deterministic Algorithm 7 applies to Algorithm 5, which achieves the same approximation ratios.

In this deterministic Algorithm 7, let S be a subset such that $|S| = s \leq k$. Let $\tilde{\mathcal{S}}_q, \tilde{\mathcal{S}}$ be outputs of Algorithms 5 and 6, respectively. We know that, from Lemma 5, $\tilde{\mathcal{S}}_q \xrightarrow{\mu} \tilde{\mathcal{S}}$ when $q \to \infty$. Thus, the expectation of mth power of objective function given S is

$$
\begin{aligned}
H(S) &:= \mathbb{E}\left[\det\left(\sum_{i \in \tilde{\mathcal{S}}} a_i a_i^\top\right) \middle| S \subseteq \tilde{\mathcal{S}}\right] = \lim_{q \to \infty} \mathbb{E}\left[\det\left(\sum_{i \in \tilde{\mathcal{S}}_q} a_i a_i^\top\right) \middle| S \subseteq \tilde{\mathcal{S}}_q\right] \\
&= \lim_{q \to \infty} \sum_{U \in \binom{\mathcal{A}_q \setminus S}{k-s}} \frac{\frac{1}{q^k}}{\sum_{\bar{U} \in \binom{\mathcal{A}_q \setminus S}{k-s}} \frac{1}{q^k}} \det\left(\sum_{i \in U} a_i a_i^\top + \sum_{i \in S} a_i a_i^\top\right) \\
&= \lim_{q \to \infty} \binom{qk-s}{k-s}^{-1} \sum_{U \in \binom{\mathcal{A}_q \setminus S}{k-s}} \sum_{R \in \binom{U \cup S}{m}} \det\left(\sum_{i \in R} a_i a_i^\top\right) \\
&= \lim_{q \to \infty} \sum_{r=1}^{\min\{k-s, m\}} \frac{q^r \binom{qk-s-r}{k-s-r}}{\binom{qk-s}{k-s}} \sum_{R \in \binom{\mathcal{A}_q}{m}, |R \setminus S| = r} q^{-r} \det\left(\sum_{i \in R} a_i a_i^\top\right)
\end{aligned}
\tag{20}
$$

$$
= \sum_{r=1}^{\min\{k-s,m\}} \lim_{q \to \infty} \frac{q^r \binom{qk-s-r}{k-s-r}}{\binom{qk-s}{k-s}} \lim_{q \to \infty} \sum_{R \in \binom{\mathscr{A}_q}{m}, |R \setminus S|=r} q^{-r} \det\left(\sum_{i \in R} a_i a_i^\top\right),
$$

$$
= \sum_{r=1}^{\min\{k-s,m\}} \frac{(k-s)!}{k^r (k-s-r)!} \lim_{q \to \infty} \sum_{R \in \binom{\mathscr{A}_q}{m}, |R \setminus S|=r} q^{-r} \det\left(\sum_{i \in R} a_i a_i^\top\right),
$$

where the second equality is due to $\tilde{\mathscr{S}}_q \xrightarrow{\mu} \tilde{\mathscr{S}}$ when $q \to \infty$, the third equality is a direct computing of the conditional probability, the fourth equality is due to Lemma 2, and the fifth and last equalities are because both $\lim_{q \to \infty} \frac{q^r \binom{qk-s-r}{k-s-r}}{\binom{qk-s}{k-s}}$ and $\lim_{q \to \infty} \sum_{R \in \binom{\mathscr{A}_q}{m}, |R \setminus S|=r} q^{-r} \det(\sum_{i \in R} a_i a_i^\top)$ exist and are finite for each $r \in \{1, \ldots, \min\{k-s, m\}\}$.

Algorithm 7 Derandomization of Algorithm 6
1. Suppose that $(\widehat{x}, \widehat{w})$ is an optimal solution to the convex relaxation (3) with $B = [0,1]$, where $\widehat{x} \in [0,1]^n$ with $\sum_{i \in [n]} \widehat{x}_i = k$ and $\widehat{w} = f(\widehat{x})$
2. Initialize chosen set $S = \emptyset$
3. **do**
4.   Let $j^* \in \arg\max_{j \in [n] \setminus \mathscr{S}} H(S \cup j)$, where $H(S \cup j)$ defines (20), and the limit can be computed by Lemma 6
5.   Add j^* to set S
6. **while** $|S| < k$
7. Output S

We note that, from Lemma 3, for each $r = 1, \ldots, \min\{k-s, m\}$, the limit in (20) can be computed efficiently according to the following lemma.

Lemma 6. *For each $r = 1, \ldots, \min\{k-s, m\}$,*
  i. *the term $\sum_{R \in \binom{\mathscr{A}_q}{m}, |R \setminus S|=r} q^{-r} \det(\sum_{i \in R} a_i a_i^\top)$ is equal to the coefficient t^r of the following determinant function*

$$
\det\left(\frac{t}{q} \sum_{i \in \mathscr{A}_q \setminus S} a_i a_i^\top + \sum_{i \in S} a_i a_i^\top\right); \text{ and}
$$

  ii. *the term $\lim_{q \to \infty} \sum_{R \in \binom{\mathscr{A}_q}{m}, |R \setminus S|=r} q^{-r} \det(\sum_{i \in R} a_i a_i^\top)$ is equal to the coefficient t^r of the following determinant function*

$$
\det\left(t \sum_{i \in [n]} \widehat{x}_i a_i a_i^\top + \sum_{i \in S} a_i a_i^\top\right).
$$

Proof.
  i. The results follow directly by Lemma 3.
  ii. By part (i), the term $\lim_{q \to \infty} \sum_{R \in \binom{\mathscr{A}_q}{m}, |R \setminus S|=r} q^{-r} \det(\sum_{i \in R} a_i a_i^\top)$ is equal to the coefficient t^r of the following determinant function:

$$
\lim_{q \to \infty} \det\left(\frac{t}{q} \sum_{i \in \mathscr{A}_q \setminus S} a_i a_i^\top + \sum_{i \in S} a_i a_i^\top\right).
$$

Because $\lim_{q \to \infty} 1/q \sum_{i \in \mathscr{A}_q \setminus S} a_i a_i^\top = \sum_{i \in [n]} \widehat{x}_i a_i a_i^\top$ and $\det(\cdot)$ is a continuous function, thus we arrive at the conclusion. $\square$

Algorithm 7 proceeds as follows. We start with an empty subset S of chosen elements, and for each $j \notin S$, we compute the expected mth power of objective function that j will be chosen $H(S \cup j)$. We update $S := S \cup \{j^*\}$, where $j^* \in \arg\max_{j \in [n] \setminus S} H(S \cup j)$. Then, go to next iteration. This procedure will terminate if $|S| = k$. Similar to Algorithms 2 and 4, the time complexity of Algorithm 7 is $O(m^4 n k^2)$. Thus, in practice, we recommend Algorithm 5 for D-optimal design problem with repetitions.

The approximation results for Algorithm 7 are identical to those in Theorem 5.

Theorem 6. *For any positive integers $m \leq k \leq n$ and $\epsilon \in (0, 1)$,*

   i. *deterministic Algorithm 7 is efficiently computable and yields $1/e$ approximation for the D-optimal design problem with repetitions; and*

   ii. *given $\epsilon \in (0, 1)$, if $k \geq (m - 1)/\epsilon$, then deterministic Algorithm 7 yields $(1 - \epsilon)$ approximation for the D-optimal design problem with repetitions.*

6. Closing Remarks and Conclusion

In this section, we make our final remarks about the proposed algorithms and present a conclusion of this paper.

6.1. Closing Remarks

We first remark that the proposed methods also work for A-optimality design, which was studied in Nikolov et al. [30]. In their paper, the authors also showed that the proposed methods might not work for other criteria. For D-optimal design problem without repetitions, if $k \approx m$, then we recommend sampling Algorithm 1 because of its efficiency and accuracy; if $k \gg m$, then we recommend sampling Algorithm 3 because of its efficiency and asymptotic optimality. For D-optimal design problem with repetitions, we recommend sampling Algorithm 5, because it is much more efficient than its deterministic counterpart.

6.2. Conclusion

In this paper, we show that D-optimal design problem admits $\frac{1}{e}$-approximation guarantee. That is, we propose a sampling algorithm and its deterministic implementation, the solution of which is at most $1/e$ of the true optimal objective value, giving the first constant approximation ratio for this problem. We also analyze a different sampling algorithm, which achieves the asymptotic optimality: that is, the output of the algorithm is $(1 - \epsilon)$ approximation if $k \geq 4m/\epsilon + 12/\epsilon^2 \log(1/\epsilon)$ for any $\epsilon \in (0, 1)$. For D-optimal design problem with repetitions (i.e., each experiment can be picked multiple times), our sampling algorithm and its derandomization improve asymptotic approximation ratio (i.e., the output of the algorithm is $(1 - \epsilon)$ approximation if $k \geq (m - 1)/\epsilon$ for any $\epsilon \in (0, 1)$). For future research, we would like to investigate if more sophisticated relaxation schemes can be used to improve the approximation analyses. Another direction is to prove the tightness of the approximation bounds. In particular, we conjecture that, for D-optimal design problem with or without repetitions, to achieve $(1 - \epsilon)$ approximation, one must have $k = \Omega(m/\epsilon)$.

Acknowledgments

Appendix A. Proofs
A.1. Proof of Lemma 2

Lemma 2. *Suppose that $a_i \in \mathbb{R}^m$ for $i \in T$ with $|T| \geq m$; then, we have the following identity*

$$\det\left(\sum_{i \in T} a_i a_i^\top\right) = \sum_{S \in \binom{T}{m}} \det\left(\sum_{i \in S} a_i a_i^\top\right). \tag{5}$$

Proof. Suppose that $T = \{i_1, \ldots, i_{|T|}\}$. Let matrix $A = [a_{i_1}, \ldots, a_{i_{|T|}}]$; then, we have

$$\det\left(\sum_{i \in T} a_i a_i^\top\right) = \det(AA^\top). \tag{A.1}$$

Next, the right-hand side of (A.1) is equivalent to

$$\det(AA^\top) = \sum_{S \in \binom{T}{m}} \det(A_S)^2 = \sum_{S \in \binom{T}{m}} \det(A_S A_S^\top) = \sum_{S \in \binom{T}{m}} \det\left(\sum_{i \in S} a_i a_i^\top\right),$$

where A_S is the submatrix of A with columns from subset S, the first equality is due to the Cauchy–Binet formula (Broida and Williamson [12]), the second equality is because A_S is a square matrix, and the last inequality is the definition of $A_S A_S^\top$. $\square$

A.2. Proof of Lemma 3

Lemma 3. *For any $x \in [0,1]^n$, then the following identify holds*

$$\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right) = \sum_{S\in\binom{[n]}{m}} \prod_{i\in S} x_i \det\left(\sum_{i\in S} a_i a_i^\top\right). \tag{6}$$

Proof. Let $P = \mathrm{diag}(x) \in \mathbb{R}^{n\times n}$ be the diagonal matrix with diagonal vector equal to x and matrix $A = [a_1, \ldots, a_n]$. By Lemma 2, we have

$$\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right) = \det\left(\sum_{i\in[n]} (\sqrt{x_i}a_i)(\sqrt{x_i}a_i)^\top\right) = \sum_{S\in\binom{[n]}{m}} \det\left(\sum_{i\in S} x_i a_i a_i^\top\right). \tag{A.2a}$$

Note that $\sum_{i\in S} x_i a_i a_i^\top = A_S P_S A_S^\top$, where A_S is the submatrix of A with columns from subset $\tilde{\mathcal{S}}$, and P_S is the square submatrix of P with rows and columns from S. Thus, (A.2a) further yields

$$\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right) = \sum_{S\in\binom{[n]}{m}} \det\left(\sum_{i\in S} x_i a_i a_i^\top\right) = \sum_{S\in\binom{[n]}{m}} \det(A_S P_S A_S^\top) = \sum_{S\in\binom{[n]}{m}} \det(A_S)^2 \det(P_S)$$
$$= \sum_{S\in\binom{[n]}{m}} \prod_{i\in S} x_i \det\left(\sum_{i\in S} a_i a_i^\top\right), \tag{A.2b}$$

where the third and fourth equalities are because the determinant of products of square matrices is equal to the products of individual determinants. $\square$

A.3. Proof of Proposition 1

Before proving Proposition 1, we first introduce two well-known results for sum of homogeneous and symmetric polynomials.

Lemma A.1. *(Maclaurin's Inequality (Lin and Trudinger [26])). Given a set S, an integer $s \in \{0, 1, \cdots, |S|\}$, and nonnegative vector $x \in \mathbb{R}_+^{|\tilde{\mathcal{S}}|}$, we must have*

$$\frac{1}{|S|}\left(\sum_{i\in S} x_i\right) \geq \sqrt[s]{\frac{1}{\binom{|S|}{s}}\left(\sum_{Q\in\binom{S}{s}} \prod_{i\in Q} x_i\right)}.$$

Lemma A.2. *(Generalized Newton's Inequality (Xu [41])). Given a set S, two nonnegative positive integers $s, \tau \in \mathbb{Z}_+$ such that $s, \tau \leq |S|$, and nonnegative vector $x \in \mathbb{R}_+^{|S|}$, then we have*

$$\frac{\left(\sum_{R\in\binom{S}{s}} \prod_{j\in R} x_j\right)}{\binom{|S|}{s}} \frac{\left(\sum_{R\in\binom{S}{\tau}} \prod_{i\in R} x_i\right)}{\binom{|S|}{\tau}} \geq \frac{\sum_{Q\in\binom{S}{s+\tau}} \prod_{i\in Q} x_i}{\binom{|\tilde{\mathcal{S}}|}{s+\tau}}.$$

Now, we are ready to prove the main proposition.

Proposition 1. *Let $\tilde{\mathcal{S}}$ be the random set defined in (4). Then, for any $T \subseteq [n]$ such that $|T| = m$, we have*

$$\mathbb{P}\left[T \subseteq \tilde{\mathcal{S}}\right] \geq \frac{1}{g(m,n,k)} \prod_{i\in T} \widehat{x}_i := \alpha^m \prod_{i\in T} \widehat{x}_i,$$

where

$$g(m,n,k) = \max_y \left\{ \sum_{\tau=0}^m \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} (k-y)^{m-\tau}(y)^\tau : \frac{mk}{n} \leq y \leq m \right\}. \tag{9}$$

Proof. According to Lemma 1 and sampling procedure in (4), we have

$$\mathbb{P}\left[T \subseteq \tilde{\mathcal{S}}\right] = \prod_{j\in T} \widehat{x}_j \frac{\sum_{R\in\binom{[n]\setminus T}{k-m}} \prod_{j\in R} \widehat{x}_j}{\sum_{\tilde{S}\in\binom{[n]}{k}} \prod_{i\in\tilde{S}} \widehat{x}_i} = \prod_{j\in T} \widehat{x}_j \frac{\sum_{R\in\binom{[n]\setminus T}{k-m}} \prod_{j\in R} \widehat{x}_j}{\sum_{\tau=0}^m \sum_{W\in\binom{T}{\tau}} \prod_{i\in W} \widehat{x}_i \left(\sum_{Q\in\binom{[n]\setminus T}{k-\tau}} \prod_{i\in Q} \widehat{x}_i\right)},$$

where the second equality uses the following identity

$$\binom{[n]}{k} = \bigcup_{\tau=0}^{m} \left\{ W \cup Q : W \in \binom{T}{\tau}, Q \in \binom{[n] \setminus T}{k - \tau} \right\}.$$

We now let

$$A_T(x) = \frac{\sum_{\tau=0}^{m} \sum_{W \in \binom{T}{\tau}} \prod_{i \in W} \widehat{x}_i \left(\sum_{Q \in \binom{[n] \setminus T}{k-\tau}} \prod_{i \in Q} \widehat{x}_i \right)}{\sum_{R \in \binom{[n] \setminus T}{k-m}} \prod_{j \in R} \widehat{x}_j}. \tag{A.3a}$$

According to Definition 1, it is sufficient to find a lower bound to $\frac{1}{\prod_{i \in T} \widehat{x}_i} \mathbb{P}[T \subseteq \tilde{\mathcal{G}}]$: that is,

$$\frac{1}{g(m,n,k)} \leq \min_x \left\{ \frac{1}{\prod_{i \in T} \widehat{x}_i} \mathbb{P}\left[T \subseteq \tilde{\mathcal{G}}\right] = \frac{1}{A_T(x)} : \sum_{i \in [n]} \widehat{x}_i = k, x \in [0,1]^n \right\}.$$

Equivalently, we would like to find an upper bound of $A_T(x)$ for any x that satisfies $\sum_{i \in [n]} \widehat{x}_i = k, x \in [0,1]^n$: that is, show that

$$g(m,n,k) \geq \max_x \left\{ A_T(x) : \sum_{i \in [n]} \widehat{x}_i = k, x \in [0,1]^n \right\}. \tag{A.3b}$$

In the following steps, we first observe that, in (A.3a), the components of $\{x_i\}_{i \in T}$ and $\{x_i\}_{i \in [n] \setminus T}$ are both symmetric in the expression of $A_T(x)$. We will show that, for the worst case, $\{x_i\}_{i \in T}$ are all equal and that $\{x_i\}_{i \in [n] \setminus T}$ are also equal. We also show that $\widehat{x}_j \leq \widehat{x}_i$ for each $i \in T$ and $j \in [n] \setminus T$. This allows us to reduce the optimization problem in right-hand side (R.H.S.) of (A.3b) to a single-variable optimization problem (i.e., (9)). The proof is now separated into the following three claims.

First, we will prove the following claim.

Claim A.1. *The optimal solution to* (A.3b) *must satisfy the following condition: for each $i \in T$ and $j \in [n] \setminus T$, $\widehat{x}_j \leq \widehat{x}_i$.*

Proof. We prove it by contradiction. Suppose that there exists $i' \in T$ and $j' \in [n] \setminus T$, where $\widehat{x}_{i'} < \widehat{x}_{j'}$. By collecting the coefficients of $1, \widehat{x}_{i'}, \widehat{x}_{j'}, \widehat{x}_{i'} \widehat{x}_{j'}$, we have

$$A_T(x) = \frac{b_1 + b_2 \widehat{x}_{i'} + b_2 \widehat{x}_{j'} + b_3 \widehat{x}_{i'} \widehat{x}_{j'}}{c_1 + c_2 \widehat{x}_{j'}},$$

where b_1, b_2, b_3, c_1, c_2 are all nonnegative numbers with

$$b_1 = \sum_{\bar{S} \in \binom{[n] \setminus \{i',j'\}}{k}} \prod_{i \in \bar{S}} \widehat{x}_i, b_2 = \sum_{\bar{S} \in \binom{[n] \setminus \{i',j'\}}{k-1}} \prod_{i \in \bar{S}} \widehat{x}_i, b_3 = \sum_{\bar{S} \in \binom{[n] \setminus \{i',j'\}}{k-2}} \prod_{i \in \bar{S}} \widehat{x}_i$$

$$c_1 = \sum_{R \in \binom{[n] \setminus (T \cup \{j'\})}{k-m}} \prod_{j \in R} \widehat{x}_j, c_2 = \sum_{R \in \binom{[n] \setminus (T \cup \{j'\})}{k-m-1}} \prod_{j \in R} \widehat{x}_j.$$

According to the following inequality,

$$\widehat{x}_{i'} \widehat{x}_{j'} \leq \frac{1}{4} \left(\widehat{x}_{i'} + \widehat{x}_{j'} \right)^2,$$

therefore, $A_T(x)$ has a larger value if we replace $\widehat{x}_{i'}, \widehat{x}_{j'}$ by their average (i.e., $\widehat{x}_{i'} := \frac{1}{2}(\widehat{x}_{i'} + \widehat{x}_{j'}), \widehat{x}_{j'} := \frac{1}{2}(\widehat{x}_{i'} + \widehat{x}_{j'})$). $\square$

Second, we will prove the following claim.

Claim A.2. *For any feasible x to* (A.6b) *and for each $S \subseteq [n]$ and $s \in \{0, 1, \ldots, |S|\}$, we must have*

$$\sum_{Q \in \binom{S}{s}} \prod_{i \in Q} \widehat{x}_i \leq \frac{\binom{|S|}{s}}{|S|^s} \left(\sum_{i \in S} \widehat{x}_i \right)^s.$$

Proof. This directly follows from Lemma A.1. $\square$

Also, we will prove the following claim.

Claim A.3. *For each $T \subseteq [n]$ with $|T| = k$ and $\tau \in \{0, 1, \cdots, m\}$, we must have*

$$\sum_{Q \in \binom{[n]\backslash T}{k-\tau}} \prod_{i \in Q} \widehat{x}_i \leq \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \left(\sum_{R \in \binom{[n]\backslash T}{k-m}} \prod_{j \in R} \widehat{x}_j \right) \left(\sum_{i \in [n]\backslash T} \widehat{x}_i \right)^{m-\tau}.$$

Proof. This can be shown by Claim A.2 and Lemma A.2, that is

$$\frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \left(\sum_{R \in \binom{[n]\backslash T}{k-m}} \prod_{j \in R} \widehat{x}_j \right) \left(\sum_{i \in [n]\backslash T} \widehat{x}_i \right)^{m-\tau}$$

$$\geq \frac{\binom{n-m}{k-\tau}}{\binom{n-m}{m-\tau}\binom{n-m}{k-m}} \left(\sum_{R \in \binom{[n]\backslash T}{k-m}} \prod_{j \in R} \widehat{x}_j \right) \left(\sum_{S \in \binom{[n]\backslash T}{m-\tau}} \prod_{i \in S} \widehat{x}_i \right) \geq \sum_{Q \in \binom{[n]\backslash T}{k-\tau}} \prod_{i \in Q} \widehat{x}_i,$$

where the first inequality is due to Claim A.2 and the last inequality is because of Lemma A.2. □

Thus, by Claim A.3, for any feasible x to (A.3b), the function $A_T(x)$ in (A.3a) can be upper bounded by

$$A_T(x) \leq \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \left(\sum_{i \in [n]\backslash T} \widehat{x}_i \right)^{m-\tau} \sum_{W \in \binom{T}{\tau}} \prod_{i \in W} \widehat{x}_i$$

$$\leq \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} \left(\sum_{i \in [n]\backslash T} \widehat{x}_i \right)^{m-\tau} \left(\sum_{i \in T} \widehat{x}_i \right)^\tau \qquad (A.3c)$$

$$\leq \max_y \left\{ \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} (k-y)^{m-\tau}(y)^\tau : \frac{mk}{n} \leq y \leq m \right\} := g(m,n,k),$$

where the second inequality is because of Claim A.2, and the last inequality is because we let $y = \sum_{i \in T} \widehat{x}_i$, which is no larger than m, maximize over it and Claim A.1 yields that $y/m \geq (k-y)/(n-m)$ (i.e., $mk/n \leq y \leq m$). This completes the proof. □

A.4. Proof of Proposition 2

Proposition 2. *For any $n \geq k \geq m$, we have*

$$\alpha^{-1} = [g(m,n,k)]^{\frac{1}{m}} \leq \lim_{\tau \to \infty} [g(m,\tau,k)]^{\frac{1}{m}} \leq \min\left\{ e, 1 + \frac{k}{k-m+1} \right\}. \qquad (10)$$

Proof.
i. First of all, we prove the following claim.

Claim A.4. *For any $m \leq k \leq n$, we have*

$$g(m,n,k) \leq g(m,n+1,k).$$

Proof. Let y^* be the maximizer to (9) for any given $m \leq k \leq n$: that is,

$$g(m,n,k) = \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} (k-y^*)^{m-\tau}(y^*)^\tau.$$

Clearly, y^* is feasible to (9) with pair $(m, n+1, k)$. We only need to show that

$$g(m,n,k) \leq \sum_{\tau=0}^{m} \frac{\binom{n+1-m}{k-\tau}}{(n+1-m)^{m-\tau}\binom{n+1-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} (k-y^*)^{m-\tau}(y^*)^\tau.$$

In other words, it is sufficient to show, for any $0 \le \tau \le m$, *we must have*

$$\frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \le \frac{\binom{n+1-m}{k-\tau}}{(n+1-m)^{m-\tau}\binom{n+1-m}{k-m}},$$

which is equivalent to prove

$$\frac{n-k}{n-m} \cdot \frac{n-k-1}{n-m} \cdots \frac{n-k-m+\tau+1}{n-m} \le \frac{n+1-k}{n+1-m} \cdot \frac{n+1-k-1}{n+1-m} \cdots \frac{n+1-k-m+\tau+1}{n+1-m}.$$

The inequality holds, because for any positive integers p, q with $p < q$, we must have $\frac{p}{q} \le \frac{p+1}{q+1}$. □

ii. By Claim A.4, it is sufficient to investigate the bound $\lim_{n' \to \infty} g(m, n', k)$, which provides an upper bound to $g(m, n, k)$ for any integers $n \ge k \ge m$. Therefore, from now on, we only consider the case when $n \to \infty$ for any fixed $k \ge m$.

Note that, for any given y, $\sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} (k-y)^{m-\tau} y^\tau$ is the coefficient of t^k in the following polynomial:

$$R_1(t) := \frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}} \left(1 + \frac{k-y}{n-m}t\right)^{n-m} \left(1 + \frac{y}{m}t\right)^m,$$

which is upper bounded by

$$R_2(t) := \frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}} \left(1 + \frac{k-y}{n-m}t + \frac{1}{2!}\left(\frac{k-y}{n-m}t\right)^2 + \dots\right)^{n-m} \left(1 + \frac{y}{m}t + \frac{1}{2!}\left(\frac{y}{m}t\right)^2 + \dots\right)^m$$

$$= \frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}} \left(e^{\frac{k-y}{n-m}t}\right)^{n-m} \left(e^{\frac{y}{m}t}\right)^m$$

because of the inequality $e^r = 1 + r + \frac{1}{2}r^2 + \dots$ for any r and $t \ge 0$. Therefore, we also have

$$\lim_{n \to \infty} \frac{1}{k!}\frac{d^k R_1(t)}{dt^k}\Big|_{t=0} = \lim_{n \to \infty} \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau$$

$$\le \lim_{n \to \infty} \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau$$

$$+ \sum_{\tau=0}^{m} \sum_{\substack{i_j \in \mathbb{Z}_+, \forall j \in [n] \\ \sum_{j \in [n-m]} i_j = k-\tau \\ \sum_{j \in [n] \setminus [n-m]} i_j = \tau \\ \max_{j \in [n]} i_j \ge 2}} \frac{1}{\prod_{j \in [n]} i_j!} \frac{1}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{1}{m^\tau}(k-y)^{m-\tau}y^\tau := \lim_{n \to \infty} \frac{1}{k!}\frac{d^k R_2(t)}{dt^k}\Big|_{t=0}$$

$$= \lim_{n \to \infty} \frac{k^k}{k!} \frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}} \le \lim_{n \to \infty} \frac{k^k}{k!} \frac{(n-m)^{k-m}}{(k-m)^{k-m}\binom{n-m}{k-m}}$$

$$= \frac{k^k}{k!} \frac{(k-m)!}{(k-m)^{k-m}} := R_3(m, k),$$

where the first inequality is because of the nonnegativity of the second term of $\cdot \frac{1}{k!}\frac{d^k R_2(t)}{dt^k}\big|_{t=0}$, the second and third equalities are because of two equivalent definitions of $R_2(t)$, the last inequality is due to $y \le m$, and the fourth equality holds because of $n \to \infty$.

Note that $R_3(m, k)$ is nondecreasing over $k \in [m, \infty)$. Indeed, for any given m,

$$\log \frac{R_3(m, k+1)}{R_3(m, k)} = k \log\left(1 + \frac{1}{k}\right) - (k-m)\log\left(1 + \frac{1}{(k-m)}\right),$$

the first derivative over k of which is equal to

$$\log\left(1+\frac{1}{k}\right)-\frac{1}{k+1}-\log\left(1+\frac{1}{(k-m)}\right)+\frac{1}{k-m+1}\leq 0,$$

and that is, $\log\frac{R_3(m,k+1)}{R_3(m,k)}$ is nonincreasing over k. Therefore,

$$\log\frac{R_3(m,k+1)}{R_3(m,k)}=k\log\left(1+\frac{1}{k}\right)-(k-m)\log\left(1+\frac{1}{(k-m)}\right)\geq \lim_{k\to\infty}\log\frac{R_3(m,k+1)}{R_3(m,k)}=0.$$

Thus, $R_3(m,k)$ is upper bounded when $k\to\infty$: that is,

$$R_3(m,k)\leq \lim_{k'\to\infty}R_3(m,k')=\lim_{k'\to\infty}\left[\left(1-\frac{m}{k'}\right)^{-\frac{k'}{m}}\right]^m\frac{(k'-m)^m}{k'(k'-1)\cdots(k'-m+1)}=e^m,$$

where the last equality is because $\lim_{k'\to\infty}(1-\frac{m}{k'})^{-\frac{k'}{m}}=e$ and $\lim_{k'\to\infty}\frac{(k'-m)^m}{k'(k'-1)\cdots(k'-m+1)}=1$. Therefore, the following inequalities hold:

$$\lim_{n\to\infty}[g(m,n,k)]^{\frac{1}{m}}=\lim_{n\to\infty}\left[\max_y\left\{\sum_{\tau=0}^{m}\frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}}\frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau:\frac{mk}{n}\leq y\leq m\right\}\right]^{\frac{1}{m}}.$$
$$\leq [R_3(m,k)]^{\frac{1}{m}}\leq e$$

iii. We now compute another bound $1+k/(k-m+1)$ for $[g(m,n,k)]^{\frac{1}{m}}$, which can be smaller than e when k is large. By Claim A.4, we have

$$g(m,n,k)\leq \lim_{n'\to\infty}g(m,n',k)=\max_y\left\{\sum_{\tau=0}^{m}\frac{(k-m)!\binom{m}{\tau}}{(k-\tau)!m^\tau}(k-y)^{m-\tau}y^\tau:0\leq y\leq m\right\}.$$

Note that $0\leq y\leq m$; thus, $k-y\leq k$. Therefore, we have

$$\lim_{n\to\infty}g(m,n,k)\leq \sum_{\tau=0}^{m}\frac{(k-m)!\binom{m}{\tau}}{(k-\tau)!}k^{m-\tau}\leq \sum_{\tau=0}^{m}\left(\frac{k}{k-m+1}\right)^{m-\tau}=\left(1+\frac{k}{k-m+1}\right)^m,$$

where the last inequality is because

$$\frac{(k-m)!}{(k-\tau)!}=\frac{1}{(k-\tau)\cdots(k-m+1)}\leq\left(\frac{1}{k-m+1}\right)^{m-\tau}.$$

Therefore, we have

$$[g(m,n,k)]^{\frac{1}{m}}=\left[\max_y\left\{\sum_{\tau=0}^{m}\frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}}\frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau:\frac{mk}{n}\leq y\leq m\right\}\right]^{\frac{1}{m}}$$
$$\leq 1+\frac{k}{k-m+1}$$

for any $m\leq k\leq n$. □

References

[1] Albert I, Denis J-B (2011) Dirichlet and multinomial distributions: Properties and uses in Jags. *Anal. (Oxford)* 31:1141–1155.
[2] Allen-Zhu Z, Liao Z, Orecchia L (2015) Spectral sparsification and regret minimization beyond matrix multiplicative updates. *Proc. 47th Annual ACM Sympos. Theory Comput.* (ACM, New York), 237–245.
[3] Allen-Zhu Z, Li Y, Singh A, Wang Y (2017) Near-optimal design of experiments via regret minimization. *Proc. 34th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 70 (PMLR), 126–135.
[4] Anari N, Gharan SO (2017) A generalization of permanent inequalities and applications in counting and optimization. Preprint, submitted February 9, https://arxiv.org/abs/1702.02937.
[5] Atkinson AC, Donev AN, Tobias RD (2007) *Optimum Experimental Designs, with SAS*, vol. 34 (Oxford University Press, Oxford, UK).
[6] Avron H, Boutsidis C, Toledo S, Zouzias A (2013) Efficient dimensionality reduction for canonical correlation analysis. *Proc. 30th Internat. Conf. Machine Learn. (ICML-13)*, Proceedings Machine Learning Research, vol. 28(PMLR), 347–355.
[7] Batson J, Spielman DA, Srivastava N (2012) Twice-ramanujan sparsifiers. *SIAM J. Comput.* 41(6):1704–1721.
[8] Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, vol. 2 (SIAM, Philadelphia).

[9] Bouhtou M, Gaubert S, Sagnol G (2010) Submodularity and randomized rounding techniques for optimal experimental design. *Electronic Notes Discrete Math.* 36:679–686.

[10] Brändén P, Jonasson J (2012) Negative dependence in sampling. *Scandinavian J. Statist.* 39(4):830–838.

[11] Brewer KRW, Hanif M (2013) *Sampling with Unequal Probabilities*, vol. 15 (Springer Science & Business Media, New York).

[12] Broida JG, Williamson SG (1989) *A Comprehensive Introduction to Linear Algebra* (Addison-Wesley, Reading, MA).

[13] Byrka J, Pensyl T, Rybicki B, Srinivasan A, Trinh K (2014) An improved approximation for *k*-median, and positive correlation in budgeted optimization. *Proc. 26th Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 737–756.

[14] Calinescu G, Chekuri C, Pál M, Vondrák J (2011) Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.* 40(6):1740–1766.

[15] Chernoff H (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* 493–507.

[16] Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* 19(90):297–301.

[17] Di Summa M, Eisenbrand F, Faenza Y, Moldenhauer C (2015) On largest volume simplices and sub-determinants. *Proc. 26th Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 315–323.

[18] Federer WT (1955) *Experimental Design* (Macmillan Co., New York).

[19] Fedorov VV (1972) *Theory of Optimal Experiments* (Academic Press, New York).

[20] Horn RA, Johnson CR (1985) *Matrix Analysis* (Cambridge University Press, Cambridge, UK).

[21] Hou S-H (1998) Classroom note: A simple proof of the Leverrier–Faddeev characteristic polynomial algorithm. *SIAM Rev.* 40(3):706–709.

[22] Joshi S, Boyd S (2009) Sensor selection via convex optimization. *IEEE Trans. Signal Processing* 57(2):451–462.

[23] Khachiyan LG (1996) Rounding of polytopes in the real number model of computation. *Math. Oper. Res.* 21(2):307–320.

[24] Kirk RE (1982) *Experimental Design* (John Wiley & Sons, New York).

[25] Krause A, Golovin D (2014) Submodular function maximization. Accessed November 25, 2019, http://www.cs.cmu.edu/afs/.cs.cmu.edu/Web/People/dgolovin/papers/submodular_survey12.pdf.

[26] Lin M, Trudinger NS (1994) On some inequalities for elementary symmetric function. *Bull. Australian Math. Soc.* 50(2):317–326.

[27] Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions—I. *Math. Programming* 14(1):265–294.

[28] Nikolov A (2015) Randomized rounding for the largest simplex problem. *Proc. 47th Annual ACM Sympos. Theory Comput.* (ACM, New York), 861–870.

[29] Nikolov A, Singh M (2016) Maximizing determinants under partition constraints. *Proc. 48th Annual ACM Sympos. Theory Comput.* (ACM, New York), 192–201.

[30] Nikolov A, Singh M, Tantipongpipat UT (2018) Proportional volume sampling and approximation algorithms for *A*-optimal design. Preprint, submitted February 22, https://arxiv.org/abs/1802.08318.

[31] Papoulis A, Unnikrishna Pillai S (2002) *Probability, Random Variables, and Stochastic Processes* (Tata McGraw-Hill Education, New York).

[32] Pukelsheim F (2006) *Optimal Design of Experiments* (SIAM, Philadelphia).

[33] Sagnol G, Harman R (2015) Computing exact *D*-optimal designs by mixed integer second-order cone programming. *Ann. Statist.* 43(5):2198–2224.

[34] Shamaiah M, Banerjee S, Vikalo H (2010) Greedy sensor selection: Leveraging submodularity. *Proc. 49th IEEE Conf. Decision Control (CDC)* (IEEE, New York), 2572–2577.

[35] Singh M, Xie W (2018) Approximate positively correlated distributions and approximation algorithms for *D*-optimal design. *Proc. 29th Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 2240–2255.

[36] Spencer JH, Spencer JH (1987) *Ten Lectures on the Probabilistic Method*, vol. 52 (SIAM, Philadelphia).

[37] Spielman DA, Srivastava N (2011) Graph sparsification by effective resistances. *SIAM J. Comput.* 40(6):1913–1926.

[38] Straszak D, Vishnoi NK (2016) Real stable polynomials and matroids: Optimization and counting. Preprint, submitted November 14, https://arxiv.org/abs/1611.04548.

[39] Wang Y, Yu AW, Singh A (2016) On computationally tractable selection of experiments in regression models. Preprint, submitted January 9, https://arxiv.org/abs/1601.02068.

[40] Welch WJ (1982) Algorithmic complexity: Three NP-hard problems in computational statistics. *J. Statist. Comput. Simulation* 15(1):17–25.

[41] Xu J (2008) Generalized Newton-like inequalities. *J. Inequalities Pure Appl. Math.* 9(3), Paper No. 85.