Approximate Positive Correlated Distributions and Approximation Algorithms for D-optimal Design

Mohit Singh* Weijun Xie†

Abstract

Experimental design is a classical area in statistics [21] and has also found new applications in machine learning[2]. In the combinatorial experimental design problem, the aim is to estimate an unknown m-dimensional vector x from linear measurements where a Gaussian noise is introduced in each measurement. The goal is to pick k out of the given n experiments so as to make the most accurate estimate of the unknown parameter x. Given a set S of chosen experiments, the most likelihood estimate x' can be obtained by a least squares computation. One of the robust measures of error estimation is the D-optimality criterion [27] which aims to minimize the generalized variance of the estimator. This corresponds to minimizing the volume of the standard confidence ellipsoid for the estimation error $x - x'$. The problem gives rise to two natural variants depending on whether repetitions of experiments is allowed or not. The latter variant, while being more general, has also found applications in geographical location of sensors [19].

We show a close connection between approximation algorithms for the D-optimal design problem and constructions of *approximately m-wise positively correlated distributions.* This connection allows us to obtain a $\frac{1}{e}$-approximation for the D-optimal design problem with and without repetitions giving the first constant factor approximation for the problem. We then consider the case when the number of experiments chosen is much larger than the dimension m and show one can obtain $(1-\epsilon)$-approximation if $k \geq \frac{2m}{\epsilon}$ when repetitions are allowed and if $k = O\left(\frac{m}{\epsilon} + \frac{1}{\epsilon^2} \cdot \log \frac{1}{\epsilon}\right)$ when no repetitions are allowed improving on previous work.

1 Introduction

Experimental design is a classical area in statistics [5, 16, 17, 21, 27] and has also found new applications in machine learning [2, 32]. In the combinatorial experimental design problem, our aim is to estimate a vector $x \in \mathbb{R}^m$ from linear measurements of the form $y_i = a_i^T x + \eta_i$ where $a_i \in \mathbb{R}^m$ characterizes the i^{th} experiment and η_i is an independent Gaussian noise with zero mean and fixed variance σ^2 and $1 \leq i \leq n$ indexes all candidate experiments. We are also given an integer $k \geq m$ and the goal is to pick k out of the given n experiments so as to make the *most accurate estimate* of the parameter x. If a set of experiments S is chosen, the most likelihood estimate is given by

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

One of the most robust measures of error estimation is the D-optimality criterion where the goal is to maximize $\det\left(\sum_{i\in S} a_i a_i^T\right)^{\frac{1}{m}}$. This corresponds to minimizing the volume of the standard confidence ellipsoid for the estimation error $x - \widehat{x}$. In the problem description, the same experiment may or may not be allowed to chosen multiple times. We refer the problem as *D-optimal design with repetitions* if we are allowed to pick an experiment more than once and *D-optimal design* otherwise. The latter problem has also been studied as the sensor selection problem [19] where the goal is to find the set of sensor locations to so obtain the most accurate estimate of certain parameter to be measured. It is easy to see that D-optimal design with repetitions is a special case of the D-optimal design problem[1].

1.1 Our Results and Contributions Formally, in the D-optimal design problem, we are given a collection of vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ and integer $k \geq m$ and the goal is to pick a subset $S \subseteq \{1, \ldots, n\}$ of size k to maximize

$$det\left(\sum_{i \in S} a_i a_i^T\right)^{\frac{1}{m}}.$$

In the D-optimal design problem with replacement, the solution S can be a multi-set. The D-optimal design problem is known to be NP-hard [33] and our aim is to design approximation algorithms for the problem.

Our first contribution is to reduce the design of approximation algorithms for the D-optimal design problem to finding distributions that are approximately positively correlated. To make this connection formal,

*School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332. Email: mohitsinghr@gmail.com.

†Department of Industrial & Systems Engineering, Virginia Tech, Blacksburg, VA 24060. Email: wxie@vt.edu.

[1]For the reduction, introduce k copies of each vector in the instance to construct an instance with nk vectors.

we give the following definition. Also, we let $[N] = \{1, \ldots, N\}$ for any positive integer N.

DEFINITION 1. *Given $x \in [0,1]^n$ such that $1^T x = k$ 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 of $1, \ldots, n$ respectively, thus $X_i = 1$ if $i \in \mathcal{S}$ and 0 otherwise for each $1 \leq i \leq n$ where random set $\mathcal{S}$ is sampled from μ. Then $X_1, \ldots, X_n$ are m-wise α-positively correlated for some $0 \leq \alpha \leq 1$ if for each $T \subseteq [n]$ such that $|T| \leq m$, we have*

$$\mathbb{P}[T \subseteq \mathcal{S}] \geq \alpha^{|T|} \prod_{i \in T} x_i.$$

With a slight abuse of notation, we call the distribution μ to be m-wise α-positively correlated if the above condition is satisfied. Observe that if $\alpha = 1$, then the above definition implies that the random variables $X_1, \ldots, X_n$ are m-wise positively correlated. We insist on an approximate version where we relax the positive correlation condition by a multiplicative factor. The following lemma shows the crucial role played by m-wise approximate positively correlated distributions in design of algorithms for D-optimal design.

LEMMA 1.1. *For any $\alpha \leq 1$, the D-optimal design problem has a randomized α-approximation algorithm if for every $x \in [0,1]^n$ and integers $m \leq k \leq n$, such that $\sum_{i \in [n]} x_i = k$ there exists an efficiently computable distribution that is m-wise α-positively correlated.*

The proof of Lemma 1.1 relies on the polynomial formulation of a natural convex relaxation of the D-optimal design problem. We show that a m-wise α-positively correlated distribution leads to an randomized algorithm for the D-optimal design problem that approximates each of the coefficients in the polynomial formulation. The convex relaxation and the proof of the Lemma is given in Section 2.

We then utilize the connection in Lemma 1.1 to design an $\frac{1}{e}$-approximation algorithm giving the first constant factor approximation for D-optimal design problem. Previously, constant factor approximations were known only for restricted range of parameters [2, 6, 32] (see related work for details).

THEOREM 1.1. *For any integers $m \leq k \leq n$ and $x \in [0,1]^n$ such that $1^T x = k$, there exists an efficiently computable distribution μ on subsets of $[n]$ of size k such that the indicator random variables of $1, \ldots, n$ are m-wise $1/e$-positively correlated. Thus, there is a $\frac{1}{e}$-approximation algorithm for the D-optimal design problem.*

The distribution μ in Theorem 1.1 is the product distribution where $\mu(\mathcal{S}) \propto \prod_{i \in \mathcal{S}} x_i$ for each $\mathcal{S}$ of size k. We show that it is approximately positively correlated with the claimed parameters. The technical ingredient in the proof are various inequalities on symmetric polynomials [34]. We also show how to derandomize the above algorithm and obtain a deterministic algorithm achieving the same guarantee. We also remark that the bound $\frac{1}{e}$ is best possible for the parameter α in m-wise α-positively correlated distributions. The proof of Theorem 1.1 appears in Section 3.

While in general, the bounds are tight, we show that they can be improved when k is larger than m, a case that has been considered in previous work as well [2, 32]. Indeed, in this case we obtain substantial improvements.

THEOREM 1.2. *For any integers $m \leq k \leq n$ such that and $x \in [0,1]^n$ such that $\sum_{i \in [n]} x_i = k$, there exists an efficiently computable distribution μ on subsets of $[n]$ of size k such that the indicator random variables of $1, \ldots, n$ are m-wise $(1-\epsilon)$-positively correlated if $k = O\left(\frac{1}{\epsilon^2} \log\left(\frac{1}{\epsilon}\right) + \frac{m}{\epsilon}\right)$. Thus there exists a $(1-\epsilon)$-approximation for the D-optimal design problem when $k = O\left(\frac{1}{\epsilon^2} \log\left(\frac{1}{\epsilon}\right) + \frac{m}{\epsilon}\right)$.*

To obtain the claimed distribution in Theorem 1.2, we start with an independent randomized rounding with marginals given by vector x. If the random set thus obtained has size more than k, we apply a simple contention resolution method by selecting a random set of size k of the selected elements. The proof of Theorem 1.2 appears in Section 4.

As remarked earlier, the D-optimal design with repetition is a special case of the D-optimal design problem. Therefore, all the above results apply to this case as well. But, for the case when k is larger than m, we can obtain improved bounds.

THEOREM 1.3. *Thus for any integers $m \leq k \leq n$, there exists a $(1-\epsilon)$-approximation for the D-optimal design problem with repetition when $k \geq \frac{2m}{\epsilon}$.*

The randomized algorithm for Theorem 1.3 relies on a simple randomized algorithm similar to that in Nikolov [24]. The proof differs significantly from Nikolov [24] and relies on relationships between conditional Poisson sampling and multinomial distributions as well as Poisson limit theorems. The proof of Theorem 1.3 appears in Section 5.

1.2 Related Work

As remarked earlier, experimental design is a classical area in Statistics; we refer the reader to Pukelsheim[27], Chapter 9 on details about D-optimality criterion as well as other related (A, E, T)-criteria. The combinatorial version, where each experiment needs to chosen integrally as in this work, is also called exact experimental design and the D-optimality criterion is NP-hard [33]. In contrast, if the experiments are allowed to be picked fractionally, then D-optimality criterion reduces to a convex program(see for example [10]) and [28] give methods to solve the convex program fast. There has also been work on heuristic methods, such as local search and its variants, to obtain good solutions [17].

From an approximation algorithm viewpoint, the problem has received attention lately. Bouhou et al. [9] give a $\left(\frac{n}{k}\right)^{\frac{1}{m}}$-approximation algorithm. Wang et al. [32] building on [6] give a $(1+\epsilon)$-approximation if $k \geq \frac{m^2}{\epsilon}$. Recently, Allen-Zhu et al [2] use the connection of the problem to sparsification[7, 29] and use regret minimization methods [3] and gave $O(1)$-approximation algorithm if $k \geq 2m$ and $(1+\epsilon)$-approximation when $k \geq O\left(\frac{m}{\epsilon^2}\right)$. We also remark that their results also apply to other optimality criteria.

A closely related problem is the largest j-simplex problem. This problem is equivalent to the following: Given a set of n vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ and integer $k \leq m$, pick a set of S of k vectors to maximize the k^{th}-root of the pseudo-determinant of $X = \sum_{i \in S} a_i a_i^T$, i.e., the geometric mean of the non-zero eigenvalues of X. The problem has received much attention recently [20, 24, 31] and Nikolov[24] gave a $\frac{1}{e}$-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, while applicable, results in a $\frac{1}{e^{\frac{k}{m}}}$-approximation for the D-optimal design problem. Recently, matroid constrained versions of the largest j-simplex problem have also been studied [4, 25, 30].

The D-optimality criteria is also closely related to constrained submodular maximization, a classical problem [23], for which there has been much progress recently[14]. Indeed the set function $f(S) := \log\det\left(\sum_{i \in S} a_i a_i^T\right)$ is known to be submodular. Unfortunately, the submodular function f is not necessarily non-negative, a prerequisite for all the results on constrained submodular maximization and thus these results are not applicable. Moreover, for a multiplicative guarantee for the det objective, we would aim for an additive guarantee for log det objective.

Sampling k objects out of n with given marginals has been studied intensely and many different schemes have been proposed [12]. Indeed most of them exhibit negative correlation of various degrees [11]. Here we are interested in schemes which exhibit approximate positive correlation and thus allow the marginals to be satisfied approximately. We expect that the concept of approximate positive correlation as well the rounding methods proposed in the paper will be of independent interest.

Notations: 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 $[N] := \{1, \ldots, N\}$, $N! = \prod_{i \in [N]} i$, $|Q|$ denote its cardinality and $\binom{Q}{N}$ denote all the possible subsets of Q whose cardinality equals to N. 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. And we use $\mathcal{S}$ to denote a random set.

2 Convex Relaxation and Positively Correlated Distributions

In this section, we introduce a convex relaxation for the problem that has been extensively studied [19, 28] and prove Lemma 1.1.

2.1 Convex Relaxation

We first note that the D-optimal design problem can be formulated as a mixed integer convex program below:

(2.1)
$$\max_{x \in \{0,1\}^n, w} \left\{ w : w \leq \left[\det\left(\sum_{i \in [n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}, \sum_{i \in [n]} x_i = k \right\},$$

where $\left[\det\left(\sum_{i \in [n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$ is concave in x (c.f., [8]). A straightforward convex relaxation of (2.1) is to relax the binary vector x as continuous one, i.e., $x \in [0,1]^n$, which is formulated as below

(2.2)
$$\max_{x \in [0,1]^n, w} \left\{ w : w \leq \left[\det\left(\sum_{i \in [n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}, \sum_{i \in [n]} x_i = k \right\}.$$

(2.2) is a convex program, which can be solved efficiently. Recently, in [28], the authors proposed a second order conic program (SOCP) formulation for (2.2) for which a more effective interior point method can be used to solve it.

We also remark that in Lemma 1.1, we can replace x by the optimal solution to the continuous relaxation of D-optimal design problem (2.2). This allows us to establish approximation bounds.

2.2 Proof of Lemma 1.1 Before proving Lemma 1.1, we would like to introduce some useful results below. The following lemmas follows from Cauchy-Binet formula [13] and use properties of determinants as a polynomial in entries of the matrix. The proofs appear in the appendix. For any set X and integer k, we let $\binom{X}{k} = \{Y \subseteq X : |Y| = k\}$ denote the set of subsets of X of cardinality k.

LEMMA 2.1. *Suppose $a_i \in \mathbb{R}^m$ for $i \in T$ with $|T| \geq m$, then*

$$\det\left(\sum_{i\in T} a_i a_i^\top\right) = \sum_{S\in\binom{T}{m}} \det\left(\sum_{i\in S} a_i a_i^\top\right). \tag{2.3}$$

LEMMA 2.2. *For any $x \in [0,1]^n$, then*

$$\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right) = \sum_{S\in\binom{[n]}{m}} \prod_{i\in S} x_i \det\left(\sum_{i\in S} a_i a_i^\top\right). \tag{2.4}$$

Now we are ready to prove the main Lemma 1.1.

Proof. (Proof of Lemma 1.1) Suppose that x is the optimal solution to (2.2). We consider the distribution μ given by Lemma 1.1 for this x which satisfies the conditions of Definition 1. We now consider the randomized algorithm that samples a random set $\mathcal{S}$ from this distribution μ and returns this as the solution. We show this randomized algorithm satisfies the guarantee claimed in the Lemma. All expectation and probabilities of events are under the probability measure μ and for ease of notation we drop it from the notation.

Since $\left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$ is at least as large as the optimal value to D-optimal design problem (2.2), we only need to show that

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

or equivalently

$$\mathbb{E}\left[\det\left(\sum_{i\in\mathcal{S}} a_i a_i^\top\right)\right] \geq \alpha^m \det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right). \tag{2.5}$$

This indeed holds since

$$\begin{aligned}
\mathbb{E}\left[\det\left(\sum_{i\in\mathcal{S}} a_i a_i^\top\right)\right] &= \sum_{S\in\binom{[n]}{k}} \mathbb{P}[\mathcal{S}=S] \det\left(\sum_{i\in S} a_i a_i^\top\right)\\
&= \sum_{S\in\binom{[n]}{k}} \mathbb{P}[\mathcal{S}=S] \sum_{T\in\binom{S}{m}} \det\left(\sum_{i\in T} a_i a_i^\top\right)\\
&= \sum_{T\in\binom{[n]}{m}} \mathbb{P}[T\subseteq\mathcal{S}] \det\left(\sum_{i\in T} a_i a_i^\top\right)\\
&\geq \alpha^m \sum_{T\in\binom{[n]}{m}} \prod_{i\in T} x_i \det\left(\sum_{i\in T} a_i a_i^\top\right)\\
&= \alpha^m \det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)
\end{aligned}$$

where the first equality is because of definition of probability measure μ, the second equality is due to Lemma 2.2, 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 2.2.

3 Approximation Algorithm for D-optimal Design Problem

In this section, we will propose a sampling procedure and show the approximation ratio based upon Lemma 1.1 to prove Theorem 1.1. We also develop an efficient way to implementing this algorithm and finally we will remark that this sampling algorithm can be derandomized.

3.1 Analysis of Sampling Scheme In this subsection, we will develop and analyze a sampling procedure. In this sampling procedure, we are given $x \in [0,1]^n$ with $\sum_{i\in[n]} x_i = k$. Then we randomly choose a size-k subset $\mathcal{S}$ such that

$$\mathbb{P}[\mathcal{S}=S] = \frac{\prod_{j\in S} x_j}{\sum_{S\in\binom{[n]}{k}} \prod_{i\in S} x_i} \tag{3.6}$$

for every $S \in \binom{[n]}{k}$.

We are going to derive approximation bounds on positive correlation for the measure induced by this sampling procedure. The key idea is to derive lower bound for $\mathbb{P}[T\subseteq\mathcal{S}]$ for any $T \in \binom{[n]}{m}$ in comparison to $\prod_{i\in T} x_i$. Observe that we have

$$\mathbb{P}[T\subseteq\mathcal{S}] = \frac{\sum_{S\in\binom{[n]}{k}: T\subseteq S} \prod_{j\in S} x_j}{\sum_{S\in\binom{[n]}{k}} \prod_{i\in S} x_i}.$$

Observe that the denominator 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 $[n] \setminus T$. This observation will allow us to use inequalities on symmetric polynomials and reduce the worst case ratio of $\mathbb{P}[T \subseteq \mathcal{S}]$ with $\prod_{i\in T} x_i$ 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.

Before proving the main proposition, we first introduce two well-known results for sum of homogeneous and symmetric polynomials.

LEMMA 3.1. *(Maclaurin's inequality [22]) Given a set S, an integer $s \in \{0, 1, \cdots, |S|\}$ and nonnegative vector $x \in \mathbb{R}_+^{|S|}$, we must have*

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

And

LEMMA 3.2. *(Generalized Newton's inequality [34]) 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*

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

Now we are ready to prove the main proposition.

PROPOSITION 3.1. *Suppose $\mathcal{S}$ is the random variable as defined in (3.6). Then for any $T \subseteq [n]$ such that $|T| = m$, we have*

$$\mathbb{P}[T \subseteq \mathcal{S}] \geq \frac{1}{g(m,n,k)}\prod_{i\in T} x_i,$$

where

(3.7)

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

Proof. According to Lemma 1.1 and sampling procedure in (3.6), we have

$$\mathbb{P}[T \subseteq \mathcal{S}] = \prod_{j\in T} x_j \frac{\sum_{R\in\binom{[n]\setminus T}{k-m}}\prod_{j\in R} x_j}{\sum_{\bar{S}\in\binom{[n]}{k}}\prod_{i\in\bar{S}} x_i}$$
$$= \prod_{j\in T} x_j \frac{\sum_{R\in\binom{[n]\setminus T}{k-m}}\prod_{j\in R} x_j}{\sum_{\tau=0}^{m}\sum_{W\in\binom{T}{\tau}}\prod_{i\in W} x_i\left(\sum_{Q\in\binom{[n]\setminus T}{k-\tau}}\prod_{i\in Q} x_i\right)}$$

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

(3.8a)

$$A_T(x) = \frac{\sum_{\tau=0}^{m}\sum_{W\in\binom{T}{\tau}}\prod_{i\in W} x_i\left(\sum_{Q\in\binom{[n]\setminus T}{k-\tau}}\prod_{i\in Q} x_i\right)}{\sum_{R\in\binom{[n]\setminus T}{k-m}}\prod_{j\in R} x_j}.$$

According to Definition 1, it is sufficient to find a lower bound to $\frac{1}{\prod_{i\in T} x_i}\mathbb{P}[T \subseteq \mathcal{S}]$, i.e.,

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

Or equivalently, we would like to find an upper bound of $A_T(x)$ for any x which satisfies $\sum_{i\in[n]} x_i = k, x \in [0,1]^n$, i.e., show that

$$g(m,n,k) \geq \max_{x\in[0,1]^n}\left\{A_T(x) : \sum_{i\in[n]} x_i = k\right\}. \tag{3.8b}$$

In the following steps, we first observe that in (3.8a), 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 $\{x_i\}_{i\in[n]\setminus T}$ are also equal. We also show that $x_j \leq x_i$ for each $i \in T$ and $j \in [n] \setminus T$. This allows us to reduce the optimization problem in R.H.S. of (3.8b) to a single variable optimization problem, i.e., (3.7). The proof is now separated into following three claims.

(i) First, we claim that

CLAIM 1. *The optimal solution to (3.8b) must satisfy the following condition - for each $i \in T$ and $j \in [n] \setminus T$, $x_j \leq x_i$.*

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

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

where b_1, b_2, b_3, c_1, c_2 are all non-negative numbers with

$$b_1 = \sum_{\bar{S} \in \binom{[n]\backslash\{i',j'\}}{k}} \prod_{i \in \bar{S}} x_i,$$

$$b_2 = \sum_{\bar{S} \in \binom{[n]\backslash\{i',j'\}}{k-1}} \prod_{i \in \bar{S}} x_i,$$

$$b_3 = \sum_{\bar{S} \in \binom{[n]\backslash\{i',j'\}}{k-2}} \prod_{i \in \bar{S}} x_i$$

$$c_1 = \sum_{R \in \binom{[n]\backslash(T\cup\{j'\})}{k-m}} \prod_{j \in R} x_j,$$

$$c_2 = \sum_{R \in \binom{[n]\backslash(T\cup\{j'\})}{k-m-1}} \prod_{j \in R} x_j.$$

Note that

$$x_{i'} x_{j'} \leq \frac{1}{4}\left(x_{i'} + x_{j'}\right)^2.$$

Therefore, $A_T(x)$ has a larger value if we replace $x_{i'}, x_{j'}$ by their average, i.e. $x_{i'} := \frac{1}{2}\left(x_{i'} + x_{j'}\right), x_{j'} := \frac{1}{2}\left(x_{i'} + x_{j'}\right)$.

(ii) Next, we claim that

CLAIM 2. *for any feasible x to* (3.8b), *and for each $S \subseteq [n]$ and $s \in \{0, 1, \cdots, |S|\}$, we must have*

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

Proof. This directly follows from Lemma 3.1.

And also

CLAIM 3. *for each $T \subseteq [n]$ with $|T| = k$ and $\tau \in \{0, 1, \cdots, m\}$,*

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

Proof. This can be shown by Claim 2 and Lemma 3.2

$$\frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \left(\sum_{R \in \binom{[n]\backslash T}{k-m}} \prod_{j \in R} x_j\right)\left(\sum_{i \in [n]\backslash T} x_i\right)^{m-\tau}$$
$$\geq \frac{\binom{n-m}{k-\tau}}{\binom{n-m}{m-\tau}\binom{n-m}{k-m}} \left(\sum_{R \in \binom{[n]\backslash T}{k-m}} \prod_{j \in R} x_j\right)\left(\sum_{S \in \binom{[n]\backslash T}{m-\tau}} \prod_{i \in S} x_i\right)$$
$$\geq \sum_{Q \in \binom{[n]\backslash T}{k-\tau}} \prod_{i \in Q} x_i$$

where the first inequality is due to Claim 2, and the last inequality is because of Lemma 3.2.

(iii) Thus, by Claim 3, for any feasible x to (3.8b), $A_T(x)$ in (3.8a) can be upper bounded as

$$A_T(x) \leq \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \cdot \left(\sum_{i \in [n]\backslash T} x_i\right)^{m-\tau} \sum_{W \in \binom{T}{\tau}} \prod_{i \in W} x_i$$
$$\leq \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} \cdot \left(\sum_{i \in [n]\backslash T} x_i\right)^{m-\tau} \left(\sum_{i \in T} x_i\right)^{\tau}$$
$$\leq \max_y \left\{ \sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau} \cdot (k-y)^{m-\tau}(y)^\tau : \frac{mk}{n} \leq y \leq m \right\} := g(m,n,k)$$

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

Next, we derive the upper bound of $g(m,n,k)$ in (3.7), which is a single-variable optimization problem. In the proof below, we first observe that for any given (m,k) with $m \leq k$, $g(m,n,k)$ is monotone non-decreasing in n. This motivates us to find an upper bound on $\lim_{n\to\infty} g(m,n,k)$, which leads to the proposition below.

PROPOSITION 3.2. *For any $n \geq k \geq m$, we have*

$$[g(m,n,k)]^{\frac{1}{m}} \leq \min\left\{e, 1+\frac{k}{k-m+1}\right\}. \tag{3.9}$$

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

CLAIM 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 (3.7) for any given $m \leq k \leq n$, i.e.,

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

Clearly, y^* is feasible to (3.7) with pair $(m, n+1, k)$. We only need to show that

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

In other words, it is sufficient to show for any $0 \leq \tau \leq m$,

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

which is equivalent to prove

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

The above inequality holds since for any positive integers p, q with $p < q$, we must have $\frac{p}{q} \leq \frac{p+1}{q+1}$.

(ii) By Claim 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 $n \geq k \geq m$. Therefore, from now on, we only consider the case when $n \to \infty$ for any fixed $k \geq m$.

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

$$\begin{aligned} R_2(t) &:= \frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}} \left(1+\frac{k-y}{n-m}t+\frac{1}{2!}\left(\frac{k-y}{n-m}t\right)^2+\ldots\right)^{n-m} \left(1+\frac{y}{m}t+\frac{1}{2!}\left(\frac{y}{m}t\right)^2+\ldots\right)^m \\ &= \frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}}\left(e^{\frac{k-y}{n-m}t}\right)^{n-m}\left(e^{\frac{y}{m}t}\right)^m \end{aligned}$$

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

$$\begin{aligned} &\lim_{n\to\infty}\frac{1}{k!}\frac{d^k R_1(t)}{dt^k}\bigg|_{t=0} \\ &= \lim_{n\to\infty}\sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau \\ &\leq \lim_{n\to\infty}\sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}} \frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau \\ &+\sum_{\tau=0}^{m}\sum_{\substack{i_j\in\mathbb{Z}_+,\forall j\in[n]\\ \sum_{j\in[n-m]} i_j = k-\tau \\ \sum_{j\in[n]\setminus[n-m]} i_j=\tau \\ \max_{j\in[n]} i_j\geq 2}} \frac{1}{\prod_{j\in[n]} i_j!} \frac{1}{(n-m)^{m-\tau}\binom{n-m}{k-m}}\frac{1}{m^\tau}(k-y)^{m-\tau}y^\tau \\ &:= \lim_{n\to\infty}\frac{1}{k!}\frac{d^k R_2(t)}{dt^k}\bigg|_{t=0} \\ &= \lim_{n\to\infty}\frac{k^k}{k!}\frac{(n-m)^{k-m}}{(k-y)^{k-m}\binom{n-m}{k-m}} \\ &\leq \lim_{n\to\infty}\frac{k^k}{k!}\frac{(n-m)^{k-m}}{(k-m)^{k-m}\binom{n-m}{k-m}} \\ &= \frac{k^k}{k!}\frac{(k-m)!}{(k-m)^{k-m}} := R_3(m,k) \end{aligned}$$

where the first inequality is due to the non-negativity of the second term of $\frac{1}{k!}\frac{d^k R_2(t)}{dt^k}\Big|_{t=0}$, 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,

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

whose first derivative over k is equal to

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

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

$$\begin{aligned}\log \frac{R_3(m,k+1)}{R_3(m,k)} &= k \log\left(1+\frac{1}{k}\right) \\ &- (k-m)\log\left(1+\frac{1}{(k-m)}\right) \\ &\ge \lim_{k\to\infty} \log \frac{R_3(m,k+1)}{R_3(m,k)} = 0\end{aligned}$$

Thus, $R_3(m,k)$ is upper bounded when $k \to \infty$, i.e.,

$$\begin{aligned}R_3(m,k) &\le \lim_{k'\to\infty} R_3(m,k') \\ &= \lim_{k'\to\infty} \left[\left(1-\frac{m}{k'}\right)^{-\frac{k'}{m}}\right]^m \\ &\cdot \frac{(k'-m)^m}{k'(k'-1)\cdots(k'-m+1)} = e^m,\end{aligned}$$

where the last equality is due to the fact that $\lim_{k'\to\infty}\left(1-\frac{m}{k'}\right)^{-\frac{k'}{m}} = e$ and $\lim_{k'\to\infty} \frac{(k'-m)^m}{k'(k'-1)\cdots(k'-m+1)} = 1$. Therefore,

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

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

$$\begin{aligned}g(m,n,k) &\le \lim_{n'\to\infty} g(m,n',k) \\ &= \max_{0\le y\le m}\left\{\sum_{\tau=0}^{m} \frac{(k-m)!}{(k-\tau)!}\frac{\binom{m}{\tau}}{m^\tau}(k-y)^{m-\tau}y^\tau\right\}.\end{aligned}$$

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

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

where the last inequality is due to $\frac{(k-m)!}{(k-\tau)!} = \frac{1}{(k-\tau)\cdots(k-m+1)} \le \left(\frac{1}{k-m+1}\right)^{m-\tau}$.

Therefore, we have

$$\begin{aligned}&[g(m,n,k)]^{\frac{1}{m}} \\ &= \left[\max_y\left\{\sum_{\tau=0}^{m} \frac{\binom{n-m}{k-\tau}}{(n-m)^{m-\tau}\binom{n-m}{k-m}}\frac{\binom{m}{\tau}}{m^\tau}\right.\right. \\ &\left.\left.\cdot(k-y)^{m-\tau}y^\tau : \frac{mk}{n} \le y \le m\right\}\right]^{\frac{1}{m}} \\ &\le 1+\frac{k}{k-m+1}\end{aligned}$$

for any $m \le k \le n$.

Theorem 1.1 directly follows from Proposition 3.2 when $n \ge k \ge m$.

We also note that when k is large enough, the sampling algorithm is a near 0.5-approximation.

Corollary 3.1. *For any integers $m \le k \le n$ such that and $x \in [0,1]^n$ such that $\sum_{i\in[n]} x_i = k$, there exists an efficiently computable distribution μ on subsets of $[n]$ of size k such that the indicator random variables of $1,\ldots,n$ are m-wise $0.5-\epsilon$-positively correlated if $k \ge \frac{m}{\epsilon}$. Thus there exists a $(0.5-\epsilon)$-approximation for the D-optimal design problem when $k \ge \frac{m}{\epsilon}$.*

Proof. As $0 < \epsilon < 1$, from Proposition 3.2, let

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

Then the conclusion follows.

3.2 Efficient Implementation In this subsection, we propose an efficient implementation (i.e., Algorithm 1) for the sampling procedure proposed in the previous subsection. We first observe a useful way to computing the probabilities in Algorithm 1 efficiently.

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

The main idea of the efficient implementation is to sample elements one by one and update the conditional probability distribution. We show how to sample from these updated distributions efficiently. For any given $x \in [0,1]^n$ with $\sum_{i\in[n]} x_i = k$ and a size-k subset $\mathcal{S} = S$ such that $|S| = k$, we need to compute the probability $\mathbb{P}[\mathcal{S} = S]$ in (3.6) efficiently. Indeed, this probability can be computed sequentially: given that a subset of chosen elements S with $|S| < k$ and a subset of unchosen elements T with $|T| < n-k$, then probability that $j \notin (S\cup T)$ will be chosen is equal to

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

The denominator and numerator can be computed efficiently based on Observation 1. If j is chosen, then update $S := S\cup\{j\}$; otherwise, update $T := T\cup\{j\}$. Then go to next iteration until $|S| = k$. By applying sequential conditional probability, we have

$$\mathbb{P}[\mathcal{S} = S] = \frac{\prod_{j\in S} x_j}{\sum_{\bar S\in\binom{[n]}{k}} \prod_{j\in\bar S} x_j}.$$

The detailed implementation is shown in Algorithm 1.

Algorithm 1 Sampling Algorithm with Constant Factor Approximation

Given $x \in [0,1]^n$ with $\sum_{i\in[n]} x_i = k$ and $w = \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
Initialize chosen set $\mathcal{S} = \emptyset$ and unchosen set $T = \emptyset$
Two factors: $A_1 = \sum_{\bar S\in\binom{[n]}{k}} \prod_{i\in\bar S} x_i, A_2 = 0$
for $j = 1,\ldots,n$ **do**
  if $|\mathcal{S}| == k$ **then**
    break
  else if $|T| = n-k$ **then**
    $\mathcal{S} = [n]\setminus T$
    break
  end if
  Let $A_2 = \left(\sum_{\bar S\in\binom{[n]\setminus(\mathcal{S}\cup T)}{k-1-|\mathcal{S}|}} \prod_{\tau\in\bar S} x_\tau\right)$
  Sample a $(0,1)$ uniform random variable U
  if $x_j A_2/A_1 \le U$ **then**
    Add j to set $\mathcal{S}$
    $A_1 = A_2$
  else
    Add j to set T
    $A_1 = A_1 - x_j A_2$
  end if
end for
Output $\mathcal{S}$

3.3 Deterministic Implementation In this subsection, we will present a deterministic Algorithm 2, which is a derandomization of Algorithm 1 using the method of conditional expectation. The main challenge is to show that we can compute the conditional expectation efficiently. We show how to do this by evaluating a determinant of a $n\times n$ matrix whose entries are linear polynomials in three indeterminates.

In this deterministic Algorithm 2, suppose we have two disjoint subsets S, T, where S is a chosen subset and T is a unchosen set such that $|S| = s \le k, |T| = t \le n-k, S\cap T = \emptyset$. Then the expectation of mth power of objective function given these two subsets is

(3.11)
$$\begin{aligned}
H(S,T) &:= \mathbb{E}\left[\det\left(\sum_{i\in\mathcal{S}} a_i a_i^\top\right)\middle| S\subseteq\mathcal{S}, T\cap\mathcal{S} = \emptyset\right] \\
&= \sum_{U\in\binom{[n]\setminus(S\cup T)}{k-s}} \frac{\prod_{j\in U} x_j}{\sum_{\bar U\in\binom{[n]\setminus(S\cup T)}{k-s}} \prod_{i\in\bar U} x_i} \\
&\quad\cdot \det\left(\sum_{i\in U} a_i a_i^\top + \sum_{i\in S} a_i a_i^\top\right) \\
&= \left(\sum_{\bar U\in\binom{[n]\setminus(S\cup T)}{k-s}} \prod_{i\in\bar U} x_i\right)^{-1} \sum_{U\in\binom{[n]\setminus(S\cup T)}{k-s}} \prod_{j\in U} x_j \\
&\quad\cdot \sum_{R\in\binom{U\cup S}{m}} \det\left(\sum_{i\in R} a_i a_i^\top\right) \\
&= \left(\sum_{U\in\binom{[n]\setminus(S\cup T)}{k-s}} \prod_{i\in U} x_i\right)^{-1} \sum_{R\in\binom{[n]\setminus T}{m}, |R\setminus S| = r\le k-s} \\
&\quad\cdot \prod_{j\in R\setminus S} x_j \det\left(\sum_{i\in R} a_i a_i^\top\right) \sum_{W\in\binom{[n]\setminus(S\cup T\cup R)}{k-s-r}} \prod_{j\in W} x_j,
\end{aligned}$$

where the first equality is a direct computing of the conditional probability, the second equality is due to Lemma 2.1 and the third one is because of interchange of summation.

Note that denominator term in (3.11) can be computed efficiently according to Observation 1. Next, we show that the numerator in (3.11) can be also computed efficiently below.

PROPOSITION 3.3. *Suppose $x \in \mathbb{R}_+^n, a_i \in \mathbb{R}^m$ for each $i \in [n]$ with $m \le n$ and $A = [a_1, \dots, a_n]$. Consider the following function*

$$F(t_1,t_2,t_3) = \det\Big(I_n + t_1 \operatorname{diag}(y)^{\frac{1}{2}} A^\top A \operatorname{diag}(y)^{\frac{1}{2}} + \operatorname{diag}(y)\Big), \tag{3.12}$$

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

$$y_i = \begin{cases} t_3, & \text{if } i \in S \\ 0, & \text{if } i \in T \\ x_i t_2, & \text{otherwise} \end{cases}.$$

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

$$\sum_{R \in \binom{[n]\setminus T}{m}, |R\setminus S| = r \le k-s} \prod_{j \in R\setminus S} x_j \det\left(\sum_{i\in R} a_i a_i^\top\right) \cdot \sum_{W \in \binom{[n]\setminus(S\cup T\cup R)}{k-s-r}} \prod_{j\in W} x_j. \tag{3.13}$$

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

$$\begin{aligned} F(t_1,t_2,t_3) &= \det(I_n + \operatorname{diag}(y)) \det\Big(I_n + t_1 \operatorname{diag}(\mathrm{e}+y)^{-\frac{1}{2}} \operatorname{diag}(y)^{\frac{1}{2}} A^\top A \operatorname{diag}(y)^{\frac{1}{2}} \operatorname{diag}(\mathrm{e}+y)^{-\frac{1}{2}}\Big) \\ &= \prod_{i\in S}(1+t_3) \prod_{i\in[n]\setminus(S\cup T)} (1+x_i t_2) \det(I_n + t_1 B^\top B) \end{aligned}$$

where the ith column of matrix B is

$$B_i = \begin{cases} \sqrt{\frac{t_3}{1+t_3}} a_i, & \text{if } i \in S \\ 0, & \text{if } i \in T \\ \sqrt{\frac{x_i t_2}{1+x_i t_2}} a_i, & \text{otherwise} \end{cases}.$$

Note that the coefficient of t_1^m in $\det(I_n + t_1 B^\top B)$ 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 $B^\top B$. Thus, the coefficient of t_1^m is

$$\sum_{R\in\binom{[n]}{m}} \prod_{i\in R} \lambda_i = \sum_{R\in\binom{[n]}{m}} \det\left((B^\top B)_{R,R}\right) = \sum_{R\in\binom{[n]}{m}} \det\left(\sum_{i\in R} B_i B_i^\top\right) = \sum_{R\in\binom{[n]\setminus T}{m}} \det\left(\sum_{i\in R} B_i B_i^\top\right)$$

where $P_{R,T}$ denotes a submatrix of P with rows and columns from sets R, T, the first equality is due to the property of the eigenvalues (Theorem 1.2.12 in [18]), and the second inequality is because the length of each column of B is m, and the third equality is because the determinant of singular matrix is 0.

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

$$\prod_{i\in S}(1+t_3) \prod_{i\in[n]\setminus(S\cup T)} (1+x_i t_2) \sum_{R\in\binom{[n]\setminus T}{m}} \det\left(\sum_{i\in R} B_i B_i^\top\right).$$

By Lemma 2.2 with $n = m$ and the definition of matrix B, 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

$$\begin{aligned} &\prod_{i\in S}(1+t_3) \prod_{i\in[n]\setminus(S\cup T)} (1+x_i t_2) \\ &\cdot \sum_{R\in\binom{[n]\setminus T}{m}} t_1^m \prod_{j\in R\setminus S} \frac{x_j}{1+t_2 x_j} \prod_{j\in R\cap S} \frac{t_3}{1+t_3} \\ &\cdot \det\left(\sum_{i\in R} a_i a_i^\top\right) \\ &= t_1^m \sum_{R\in\binom{[n]\setminus T}{m}} t_2^{|R\setminus S|} t_3^{|R\cap S|} (1+t_3)^{|S\setminus R|} \\ &\cdot \prod_{i\in[n]\setminus(S\cup T\cup R)} (1+x_i t_2) \prod_{j\in R\setminus S} x_j \det\left(\sum_{i\in R} a_i a_i^\top\right) \end{aligned}$$

which is equal to (3.13) by collecting coefficients of $t_1^m t_2^{k-s} t_3^s$.

The Algorithm 2 proceeds as follows. Given a subset of chosen elements S with $|S| < k$ and a subset of unchosen elements T with $|T| < n-k$ is not chosen, and $j \notin (S \cup T)$, we compute the expected mth power of objective function that j will be chosen or not $H(S\cup j, T), H(S, T\cup j)$. If $H(S\cup j, T) \ge H(S, T\cup j)$, then j is chosen, then update $S := S \cup \{j\}$; otherwise, update $T := T \cup \{j\}$. Then go to next iteration.

The approximation result for Algorithm 2 is identical to Theorem 1.1 and Corollary 3.1.

4 Improving Approximation Bound in Asymptotic Regime

In this section, we propose another sampling Algorithm 3 achieves asymptotic optimality, i.e. the output of Algorithm 3 is close to optimal when $k/m \to \infty$. In Algorithm 3, we are given $x \in [0,1]^n$ with $\sum_{i\in[n]} x_i = k$, a positive threshold $\epsilon > 0$ and a random permutation $\mathcal{N}$ of $[n]$. Then for each $j \in \mathcal{N}$, we select j with probability $\frac{x_j}{1+\epsilon}$, and let $\mathcal{S}$ be the set of selected elements. If $|\mathcal{S}| < k$,

Algorithm 2 Deterministic Algorithm

Given $x \in [0,1]^n$ with $\sum_{i\in[n]} x_i = k$ and $w = \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
Initialize chosen set $\mathcal{S} = \emptyset$ and unchosen set $T = \emptyset$
for $j = 1,\dots,n$ **do**
  if $|\mathcal{S}| == k$ **then**
    break
  else if $|T| = n-k$ **then**
    $\mathcal{S} = [n] \setminus T$
  end if
  if $H(\mathcal{S}\cup j, T) \geq H(\mathcal{S}, T\cup j)$ **then**
    Add j to set $\mathcal{S}$
  else
    Add j to set T
  end if
end for
Output $\mathcal{S}$

then we can add $k - |\mathcal{S}|$ more elements from $[n] \setminus \mathcal{S}$. On the other hand, if $|\mathcal{S}| > k$, then we apply a simple contention resolution method, i.e., choose k elements uniformly from set $\mathcal{S}$.

To analyze sampling Algorithm 3, we first show the following probability bound. The key idea is to prove the lower bound $\frac{1}{\prod_{i\in T} x_i}\mathbb{P}\{T \subseteq \mathcal{S}\}$ by using Chernoff Bound [15].

Lemma 4.1. *Let $\epsilon > 0$ and $\mathcal{S} \subseteq [n]$ be a random set output from Algorithm 3. Given $T \subseteq [n]$ with $|T| = m \leq n$, then we have*

$$\alpha^m \geq (1+\epsilon)^{-m}\left(1 - e^{-\frac{(\epsilon k-(1+\epsilon)m)^2}{k(2+\epsilon)(1+\epsilon)}}\right), \tag{4.14}$$

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

$$\alpha^m \geq (1-\epsilon)^m. \tag{4.15}$$

Proof. We note that $\mathcal{S} \subseteq [n]$ is a random set, where each $i \in [n]$ is independently sampled according to Bernoulli random variable X_i with the probability of success $\frac{x_i}{1+\epsilon}$. According to Definition 1, it is sufficient to lower bound $\frac{1}{\prod_{i\in T} x_i}\mathbb{P}\{T \subseteq \mathcal{S}\}$. Then from Algorithm 3, $\frac{1}{\prod_{i\in T} x_i}\mathbb{P}\{T \subseteq \mathcal{S}\}$ is lower bounded by

$$\begin{aligned}
&\frac{1}{\prod_{i\in T} x_i}\mathbb{P}\{T \subseteq \mathcal{S}\} \geq \frac{1}{\prod_{i\in T} x_i}\mathbb{P}\{T \subseteq \mathcal{S}', |\mathcal{S}'| \leq k\}\\
&+ \frac{1}{\prod_{i\in T} x_i}\mathbb{P}\{T \subseteq \mathcal{S} \subseteq \mathcal{S}', |\mathcal{S}'| \geq k+1\}\\
&= (1+\epsilon)^{-m}\mathbb{P}\left\{\sum_{i\in[n]\setminus T} X_i \leq k-m\right\}\\
&+ \frac{1}{\prod_{i\in T} x_i}\sum_{j=k+1}^{n}\mathbb{P}\{T \subseteq \mathcal{S} \subseteq \mathcal{S}' | T \subseteq \mathcal{S}', |\mathcal{S}'| = j\}\\
&\cdot \mathbb{P}\{T \subseteq \mathcal{S}', \mathcal{S}' = j\}\\
&= (1+\epsilon)^{-m}\mathbb{P}\left\{\sum_{i\in[n]\setminus T} X_i \leq k-m\right\}\\
&+ (1+\epsilon)^{-m}\sum_{j=k+1}^{n}\frac{\binom{j-m}{k-m}}{\binom{j}{k}}\mathbb{P}\left\{\sum_{i\in[n]\setminus T} X_i = j-m\right\}\\
&\geq (1+\epsilon)^{-m}\mathbb{P}\left\{\sum_{i\in[n]\setminus T} X_i \leq k-m\right\}
\end{aligned}$$

where the first inequality is because we are ignoring the greedy step in Algorithm 3 when $|S| < k$, and the second inequality is due to $\mathbb{P}\left\{\sum_{i\in[n]\setminus T} X_i = j-m\right\} \geq 0$.

Therefore, we would like to bound the following probability

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

Since $X_i \in [0,1]$ for each $i \in [n]$, and $\mathbb{E}[\sum_{i\in[n]\setminus T} X_i] = \frac{1}{1+\epsilon}\sum_{i\in[n]\setminus T} x_i$. According to Chernoff bound [15], we have

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

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

$$\begin{aligned}
&(1+\epsilon)^{-m}\mathbb{P}\left\{\sum_{i\in[n]\setminus T} X_i \leq k-m\right\}\\
&\geq (1+\epsilon)^{-m}\left(1 - e^{-\frac{\bar{\epsilon}^2\sum_{i\in[n]\setminus T} x_i}{(2+\bar{\epsilon})(1+\epsilon)}}\right).
\end{aligned} \tag{4.16}$$

Note that $k - m \le \sum_{i\in[n]\setminus T} X_i \le k$, $\epsilon - (1+\epsilon)\frac{m}{k} \le \bar{\epsilon} \le \epsilon$ and $\epsilon k - (1+\epsilon)m \le \bar{\epsilon}\sum_{i\in[n]\setminus T} x_i \le \epsilon(k-m)$. Suppose that $k \ge \frac{1+\epsilon}{\epsilon}m$, then the left-hand side of (4.16) can be further lower bounded as

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

To prove(4.15), we only need to show

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

or equivalently,

$$\text{(4.17)} \qquad \log\left[1 - \left(1-\epsilon^2\right)^m\right] \ge -\frac{(\epsilon k - (1+\epsilon)m)^2}{k(2+\epsilon)(1+\epsilon)},$$

which holds if

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

We note that $-\log\left[1-\left(1-\epsilon^2\right)^m\right]$ is non-increasing over $m \ge 1$, therefore is upper bounded by $2\log(\frac{1}{\epsilon})$. Hence, (4.17) holds if $k \ge \frac{4m}{\epsilon} + \frac{12}{\epsilon^2}\log(\frac{1}{\epsilon})$.

Algorithm 3 Asymptotic Sampling Algorithm

Given $x \in [0,1]^n$ with $\sum_{i\in[n]} x_i = k$ and $w = \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
Initialize $\mathcal{S} = \emptyset$ and a positive number $\epsilon > 0$
Let set $\mathcal{N}$ be a random permutation set of $\{1,\dots,n\}$
for $j \in \mathcal{N}$ **do**
    Sample a $(0,1)$ uniform random variable U
    if $U \le \frac{x_j}{1+\epsilon}$ **then**
        Add j to set $\mathcal{S}$
    end if
end for
if $|\mathcal{S}| > k$ **then** ▷ Fix the case when $|\mathcal{S}| > k$ by uniform sampling
    Select a subset T uniformly from set $\mathcal{S}$
    Let $\mathcal{S} := T$
end if
while $|\mathcal{S}| < k$ **do** ▷ Greedy step to enforce $|\mathcal{S}| = k$
    Let $j^* \in \arg\max_{j\in[n]\setminus\mathcal{S}} \left[\det\left(\sum_{i\in\mathcal{S}} a_i a_i^\top + a_j a_j^\top\right)\right]^{\frac{1}{m}}$
    Add j^* to set $\mathcal{S}$
end while
Output $\mathcal{S}$

Now Theorem 1.2 follows directly from Lemmas 1.1 and 4.1.

5 Approximation Algorithm for *D*-optimal Design Problem with Repetition

For the *D*-optimal design problem with repetition, it can be reformulated as

(5.18)
$$\max_{x\in\mathbb{Z}_+^n, w}\left\{w : w \le \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}, \sum_{i\in[n]} x_i = k\right\},$$

where the decision variable x is general integer rather than binary. Hence, similar to (2.2) except that $x \ge 0$, its convex relaxation is

(5.19)
$$\max_{x\in\mathbb{R}_+^n, w}\left\{w : w \le \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}, \sum_{i\in[n]} x_i = k\right\},$$

In [24], the author suggested to obtain k-sample set $\mathcal{S}$ with replacement, i.e. $\mathcal{S}$ can be a multi-set. The sampling procedure can be separated into k steps. At each step, a sample s is selected with probability $\mathbb{P}\{s = i\} = \frac{x_i}{k}$, given that $x \in \mathbb{R}_+^n$ with $\sum_{i\in[n]} x_i = k$. The detailed description is in Algorithm 4. 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 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 [1], where there are k trials and the probability of ith entry to be chosen is $\frac{x_i}{k}$. We terminate this sampling procedure if the total number of arrivals equals to k.

Algorithm 4 Sampling Algorithm for *D*-optimal Design with Repetition

Given $x \in \mathbb{R}_+^n$ with $\sum_{i\in[n]} x_i = k$ and $w = \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
Initialize chosen multi-set $\mathcal{S} = \emptyset$ and vector $\bar{x} = 0 \in \mathbb{R}^n$
for $j = 1,\dots,k$ **do**
    Sample s from $[n]$ with probability distribution $\mathbb{P}\{s = i\} = \frac{x_i}{k}$
    Let $\mathcal{S} = \mathcal{S} \cup \{s\}$ and $\bar{x}_s = \bar{x}_s + 1$
end for
Let $\bar{w} = \left[\det\left(\sum_{i\in[n]} \bar{x}_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
Output $(\bar{x}, \bar{w})$

To analyze Algorithm 4, we propose another Algorithm 5 which will be arbitrarily close to it. In Algorithms 5, we first assume that without loss of generality,

x is a nonnegative rational vector (i.e., $x \in \mathbb{Q}_+^n$) since set of all nonnegative rational vectors is dense in the set of all nonnegative real vectors. Then we let q be a common multiple of the denominators of rational numbers $x_1, \ldots, x_n$, i.e. $qx_1, \ldots, qx_n \in \mathbb{Z}_+$. Next, we create a multi-set $\mathcal{A}$, which contains qx_i copies of vector a_i for each $i \in [n]$, i.e. $|\mathcal{A}| = qk$. Finally, we sample a subset $\mathcal{A}_S$ of k items from set $\mathcal{A}$ uniformly, i.e. with probability $\binom{qk}{k}^{-1}$. The detailed description is in Algorithm 5. In this case, the sampling procedure can be interpreted as below. As sum of i.i.d. Bernoulli random variables is Binomial, hence we let $\{X_i\}_{i\in[n]}$ be independent binomial random variables where X_i has number of trials qx_i and probability of success $\frac{1}{q}$ for each $i \in [n]$. We terminate the sampling procedure if the total number of succeeded trials equals to k.

LEMMA 5.1. *Let $(\bar{x}, \bar{w})$ and $(\bar{x}'_q, \bar{w}'_q)$ be outputs of Algorithms 4 and 5, respectively. Then $\bar{x}'_q \xrightarrow{\mu} \bar{x}$, i.e. the probability distribution of $\bar{x}_q$ converges to $\bar{x}$ as $q \to \infty$.*

Proof. Consider independent random variables $\{X_i\}_{i\in[n]}$, $\{X'_i\}_{i\in[n]}$, where X_i is Poisson random variable with arrival rate x_i for each $i \in [n]$ and X'_i is binomial random variable with number of trials qx_i and probability of success $\frac{1}{q}$ for each $i \in [n]$.

Given an integer vector $s \in \mathbb{Z}_+^n$, clearly, and 5, we have

$$\begin{aligned}
&\mathbb{P}\{\bar{x}_i = s_i, \forall i \in [n]\}\\
&= \mathbb{P}\left\{X_i = s_i, \forall i \in [n] \middle| \sum_{i\in[n]} X_i = k\right\}\\
&= \frac{\mathbb{P}\left\{X_i = s_i, \forall i \in [n], \sum_{i\in[n]} X_i = k\right\}}{\mathbb{P}\left\{\sum_{i\in[n]} X_i = k\right\}}\\
&= \mathbb{I}(\sum_{i\in[n]} s_i = k)\frac{\prod_{i\in[n]}\mathbb{P}\{X_i = s_i\}}{\mathbb{P}\left\{\sum_{i\in[n]} X_i = k\right\}},
\end{aligned}$$

where the first equality is from the description of Algorithm 4, 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

$$\mathbb{P}\{\bar{x}'_i = s_i, \forall i \in [n]\} = \mathbb{I}(\sum_{i\in[n]} s_i = k)\frac{\prod_{i\in[n]}\mathbb{P}\{X'_i = s_i\}}{\mathbb{P}\left\{\sum_{i\in[n]} X'_i = k\right\}}.$$

Followed by the well-known Poisson limit theorem (c.f. [26]), X_i and X'_i have the same distribution as $q \to \infty$ for any $i \in [n]$. Therefore,

$$\mathbb{P}\{\bar{x}'_i = s_i, \forall i \in [n]\} \to \mathbb{P}\{\bar{x}_i = s_i, \forall i \in [n]\},$$

when $q \to \infty$, i.e., the outputs of Algorithm 4 and 5 have the same distribution when $q \to \infty$.

Algorithm 5 Approximation of Algorithm 4

1: Given $x \in \mathbb{Q}_+^n$ with $\sum_{i\in[n]} x_i = k$ and $w = \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
2: Let q be a common multiple of the denominators of rational numbers $x_1, \ldots, x_n$, i.e. $qx_1, \ldots, qx_n \in \mathbb{Z}_+$
3: Duplicate qx_i copies of vector a_i for each $i \in [n]$ as set $\mathcal{A}$, i.e. $|\mathcal{A}| = qk$
4: Sample a subset $\mathcal{A}_S$ of k items from set $\mathcal{A}$ with probability $\binom{qk}{k}^{-1}$
5: Initialize $\bar{x}'_q = 0$ and set $(\bar{x}'_q)_i = \sum_{b\in\mathcal{A}_S} \mathbb{I}(b = a_i)$ for each $i \in [n]$
6: Let $\bar{w}'_q = \left[\det\left(\sum_{i\in[n]} (\bar{x}'_q)_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$
7: Output $(\bar{x}'_q, \bar{w}'_q)$

Now we are ready to present our approximation results for Algorithm 4. The proof idea is base on Lemma 5.1, i.e., we first analyze Algorithm 5 and the result holds for Algorithm 4 by letting $q \to \infty$.

PROPOSITION 5.1. *Given $x \in \mathbb{R}_+^n$ with $\sum_{i\in[n]} x_i = k$ and $w = \left[\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)\right]^{\frac{1}{m}}$, $(\bar{x}, \bar{w})$ be the output of Algorithm 4. Then $(\mathbb{E}[\bar{w}^m])^{\frac{1}{m}} \geq g(m,k)^{-1} w$, where*

$$g(m,k) = \left[\frac{(k-m)!k^m}{k!}\right]^{\frac{1}{m}} \leq \min\left\{e, \frac{k}{k-m+1}\right\}. \tag{5.20}$$

Proof. We will first show the approximation ratio of Algorithm 5 and then apply it to Algorithm 4 by Lemma 5.1 when $q \to \infty$.

(i) Let $(\bar{x}'_q, \bar{w}'_q)$ be output of Algorithm 5. Similar to the proof of Theorem 3.1, we have

$$\mathbb{E}[(\bar{w}'_q)^m] = \sum_{S\in\binom{[qk]}{k}} \frac{1}{\binom{qk}{k}} \det\left(\sum_{i\in S} a_i a_i^\top\right)$$
$$= \frac{q^m}{\binom{qk}{k}} \sum_{S\in\binom{[qk]}{k}} \frac{1}{q^m} \det\left(\sum_{i\in S} a_i a_i^\top\right)$$
$$= \frac{q^m}{\binom{qk}{k}} \sum_{S\in\binom{[qk]}{k}} \frac{1}{q^m} \sum_{T\in\binom{S}{m}} \det\left(\sum_{i\in T} a_i a_i^\top\right)$$
$$= \frac{q^m\binom{qk-m}{k-m}}{\binom{qk}{k}} \sum_{T\in\binom{[qk]}{m}} \frac{1}{q^m} \det\left(\sum_{i\in T} a_i a_i^\top\right)$$
$$= \frac{q^m\binom{qk-m}{k-m}}{\binom{qk}{k}} \det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right)$$

where the first and second equalities are due to Algorithm 5, the third equality is because of Lemma 2.1 and $k \geq m$, the fourth equality is due to interchange of summation and the last equality is because of the identity $\sum_{i\in[n]} x_i a_i a_i^\top = \sum_{i\in[qk]} \frac{1}{q} a_i a_i^\top$.

(ii) From Lemma 5.1, we know that the output of Algorithm 5 has the same distribution of the output of Algorithm 4 when $q \rightarrow$. Thus, we have

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

(iii) Hence, 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

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

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

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

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

On the other hand, since $\frac{(k-m)!}{k!} \leq \frac{1}{(k-m+1)^m}$, thus

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

Hence, $g(m,k) \leq \min\left\{e, \frac{k}{k-m+1}\right\}$.

From Proposition 5.1, we note that when k is large enough, the solution of Algorithm 4 is almost optimal. This proves Theorem 1.3.

Proof. (Proof of Theorem 1.3) Let $(\bar{x}, \bar{w})$ be the output of Algorithm 4.For any $\epsilon \in (0,1)$, from Proposition 5.1, let

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

Then the conclusion follows by letting $k \geq \frac{2m}{\epsilon}$.

References

[1] I. Albert and J.-B. Denis. Dirichlet and multinomial distributions: properties and uses in jags. *Analysis*, 31:1141–1155, 2011.

[2] Z. Allen-Zhu, Y. Li, A. Singh, and Y. Wang. Near-optimal design of experiments via regret minimization. In *Proceedings of the ICML*.

[3] Z. Allen-Zhu, Z. Liao, and L. Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 237–245. ACM, 2015.

[4] N. Anari and S. O. Gharan. A generalization of permanent inequalities and applications in counting and optimization. *arXiv preprint arXiv:1702.02937*, 2017.

[5] A. Atkinson, A. Donev, and R. Tobias. *Optimum experimental designs, with SAS*, volume 34. Oxford University Press, 2007.

[6] H. Avron, C. Boutsidis, S. Toledo, and A. Zouzias. Efficient dimensionality reduction for canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 347–355, 2013.

[7] J. Batson, D. A. Spielman, and N. Srivastava. Twice-ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721, 2012.

[8] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. SIAM, 2001.

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

[10] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[11] P. Brändén and J. Jonasson. Negative dependence in sampling. *Scandinavian Journal of Statistics*, 39(4):830–838, 2012.

[12] K. R. Brewer and M. Hanif. *Sampling with unequal probabilities*, volume 15. Springer Science & Business Media, 2013.

[13] J. G. Broida and S. G. Williamson. *A comprehensive introduction to linear algebra*. Addison-Wesley Reading, Mass., 1989.

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

[15] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[16] W. T. Federer et al. Macmillan Co., New York and London, 1955.

[17] V. V. Fedorov. *Theory of optimal experiments*. Elsevier, 1972.

[18] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1985.

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

[20] L. G. Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.

[21] R. E. Kirk. *Experimental design*. Wiley Online Library, 1982.

[22] M. Lin and N. S. Trudinger. *On Some Inequalities for Elementary Symmetric Function*. Australian National University, School of Mathematical Sciences, Centre for Mathematics and its Applications, 1993.

[23] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functionsi. *Mathematical Programming*, 14(1):265–294, 1978.

[24] A. Nikolov. Randomized rounding for the largest simplex problem. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 861–870. ACM, 2015.

[25] A. Nikolov and M. Singh. Maximizing determinants under partition constraints. In *STOC*, pages 192–201, 2016.

[26] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[27] F. Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

[28] G. Sagnol and R. Harman. Computing exact d-optimal designs by mixed integer second-order cone programming. *The Annals of Statistics*, 43(5):2198–2224, 2015.

[29] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.

[30] D. Straszak and N. K. Vishnoi. Real stable polynomials and matroids: optimization and counting. *arXiv preprint arXiv:1611.04548*, 2016.

[31] M. D. Summa, F. Eisenbrand, Y. Faenza, and C. Moldenhauer. On largest volume simplices and subdeterminants. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 315–323. Society for Industrial and Applied Mathematics, 2015.

[32] Y. Wang, A. W. Yu, and A. Singh. On computationally tractable selection of experiments in regression models. *arXiv preprint arXiv:1601.02068*, 2016.

[33] W. J. Welch. Algorithmic complexity: three np-hard problems in computational statistics. *Journal of Statistical Computation and Simulation*, 15(1):17–25, 1982.

[34] J. Xu. Generalized newton–like inequalities. *J. Inequal. Pure and Appl. Math*, 9(3), 2008.

6 Omitted Proofs

Proof. [Proof of Lemma 2.1]Suppose that $T = \{i_1, \ldots, i_{|T|}\}$. Let matrix $A = [a_{i_1}, \ldots, a_{i_{|T|}}]$, then

$$\det\left(\sum_{i\in T} a_i a_i^\top\right) = \det\left(AA^\top\right). \tag{6.22a}$$

Next the right-hand side of (6.22a) is equivalent to

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

where A_S is the submatrix of A with columns from subset S, the first equality is due to Cauchy-Binet Formula [13], the second equality is because A_S is a square matrix, and the last inequality is the definition of $A_S A_S^\top$.

Proof. [Proof of Lemma 2.2]Let $P = \operatorname{diag}(x) \in \mathbb{R}^{n\times n}$ be the diagonal matrix with diagonal vector equal to x

and matrix $A = [a_1, \ldots, a_n]$. By Lemma 2.1, we have

$$\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right) = \det\left(\sum_{i\in[n]} (\sqrt{x_i}a_i)(\sqrt{x_i}a_i)^\top\right) = \sum_{S\in\binom{[n]}{m}} \det\left(\sum_{i\in S} x_i a_i a_i^\top\right). \tag{6.23a}$$

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 S, and P_S is the square submatrix of P with rows and columns from S. Thus, (6.23a) further yields

$$\begin{aligned}\det\left(\sum_{i\in[n]} x_i a_i a_i^\top\right) &= \sum_{S\in\binom{[n]}{m}} \det\left(\sum_{i\in S} x_i a_i a_i^\top\right) \\ &= \sum_{S\in\binom{[n]}{m}} \det\left(A_S P_S A_S^\top\right) \\ &= \sum_{S\in\binom{[n]}{m}} \det\left(A_S\right)^2 \det\left(P_S\right) \\ &= \sum_{S\in\binom{[n]}{m}} \prod_{i\in S} x_i \det\left(\sum_{i\in S} a_i a_i^\top\right)\end{aligned} \tag{6.23b}$$

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