Proportional Volume Sampling and Approximation Algorithms for A-Optimal Design

Aleksandar Nikolov* Mohit Singh† Uthaipon Tao Tantipongpipat‡

October 31, 2018

Abstract

We study the A-optimal design problem where we are given vectors $v_1, \ldots, v_n \in \mathbb{R}^d$, an integer $k \geq d$, and the goal is to select a set S of k vectors that minimizes the trace of $\left(\sum_{i \in S} v_i v_i^\top\right)^{-1}$. Traditionally, the problem is an instance of optimal design of experiments in statistics [35] where each vector corresponds to a linear measurement of an unknown vector and the goal is to pick k of them that minimize the average variance of the error in the maximum likelihood estimate of the vector being measured. The problem also finds applications in sensor placement in wireless networks [22], sparse least squares regression [8], feature selection for k-means clustering [9], and matrix approximation [13, 14, 5]. In this paper, we introduce *proportional volume sampling* to obtain improved approximation algorithms for A-optimal design.

Given a matrix, proportional volume sampling involves picking a set of columns S of size k with probability proportional to $\mu(S)$ times $\det(\sum_{i \in S} v_i v_i^\top)$ for some measure μ. Our main result is to show the approximability of the A-optimal design problem can be reduced to *approximate* independence properties of the measure μ. We appeal to hard-core distributions as candidate distributions μ that allow us to obtain improved approximation algorithms for the A-optimal design. Our results include a d-approximation when $k = d$, an $(1+\epsilon)$-approximation when $k = \Omega\left(\frac{d}{\epsilon} + \frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right)$ and $\frac{k}{k-d+1}$-approximation when repetitions of vectors are allowed in the solution. We also consider generalization of the problem for $k \leq d$ and obtain a k-approximation.

We also show that the proportional volume sampling algorithm gives approximation algorithms for other optimal design objectives (such as D-optimal design [36] and generalized ratio objective [27]) matching or improving previous best known results. Interestingly, we show that a similar guarantee cannot be obtained for the E-optimal design problem. We also show that the A-optimal design problem is NP-hard to approximate within a fixed constant when $k = d$.

*University of Toronto. Supported by an NSERC Discovery Grant (RGPIN-2016-06333), and part of this work was done while he was visiting the Simons Institute for the Theory of Computing semester on Bridging Continuous and Discrete Optimization, partially supported through NSF grant CCF-1740425.

†Georgia Institute of Technology. Supported in part by National Science Foundation grant CCF-1717947.

‡Georgia Institute of Technology. National Science Foundation grants CCF-1717947 and CCF-1740776 (Georgia Institute of Technology TRIAD), and by a Georgia Institute of Technology ARC fellowship.

1 Introduction

Given a collection of vectors, a common problem is to select a subset of size $k \leq n$ that *represents* the given vectors. To quantify the representability of the chosen set, typically one considers spectral properties of certain natural matrices defined by the vectors. Such problems arise as experimental design [18, 35] in statistics; feature selection [9] and sensor placement problems [22] in machine learning; matrix sparsification [6, 38] and column subset selection [5] in numerical linear algebra. In this work, we consider the optimization problem of choosing the representative subset that aims to optimize the A-*optimality criterion* in experimental design.

Experimental design is a classical problem in statistics [35] with recent applications in machine learning [22, 43]. Here the goal is to estimate an unknown vector $w \in \mathbb{R}^d$ via linear measurements of the form $y_i = v_i^\top w + \eta_i$ where v_i are possible experiments and η_i is assumed to be small i.i.d. unbiased Gaussian error introduced in the measurement. Given a set S of linear measurements, the maximum likelihood estimate $\hat{w}$ of w can be obtained via a least squares computation. The error vector $w - \hat{w}$ has a Gaussian distribution with mean 0 and covariance matrix $\left(\sum_{i \in S} v_i v_i^\top\right)^{-1}$. In the optimal experimental design problem the goal is to pick a cardinality k set S out of the n vectors such that the measurement error is minimized. Minimality is measured according to different criteria, which quantify the "size" of the covariance matrix. In this paper, we study the classical A-optimality criterion, which aims to minimize the average variance over directions, or equivalently the trace of the covariance matrix, which is also the expectation of the squared Euclidean norm of the error vector $w - \hat{w}$.

We let V denote the $d \times n$ matrix whose columns are the vectors $v_1, \ldots, v_n$ and $[n] = \{1, \ldots, n\}$. For any set $S \subseteq [n]$, we let V_S denote the $d \times |S|$ submatrix of V whose columns correspond to vectors indexed by S. Formally, in the A-optimal design problem our aim is to find a subset S of cardinality k that minimizes the trace of $(V_S V_S^\top)^{-1} = \left(\sum_{i \in S} v_i v_i^\top\right)^{-1}$. We also consider the A-optimal design problem with repetitions, where the chosen S can be a multi-set, thus allowing a vector to chosen more than once.

Apart from experimental design, the above formulation finds application in other areas such as sensor placement in wireless networks [22], sparse least squares regression [8], feature selection for k-means clustering [9], and matrix approximation [5]. For example, in matrix approximation [13, 14, 5] given a $d \times n$ matrix V, one aims to select a set S of k such that the Frobenius norm of the Moore-Penrose pseudoinverse of the selected matrix V_S is minimized. It is easy to observe that this objective equals the A-optimality criterion for the vectors given by the columns of V.

1.1 Our Contributions and Results Our main contribution is to introduce the *proportional volume sampling* class of probability measures to obtain improved approximation algorithms for the A-optimal design problem. We obtain improved algorithms for the problem with and without repetitions in regimes where k is close to d as well as in the asymptotic regime where $k \geq d$. Let $\mathcal{U}_k$ denote the collection of subsets of $[n]$ of size exactly k and $\mathcal{U}_{\leq k}$ denote the subsets of $[n]$ of size at most k. We will consider distributions on sets in $\mathcal{U}_k$ as well as $\mathcal{U}_{\leq k}$ and state the following definition more generally.

Definition 1.1. *Let μ be probability measure on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$). Then the proportional volume sampling with measure μ picks a set $S \in \mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$) with probability proportional to $\mu(S)\det(V_S V_S^\top)$.*

Observe that when μ is the uniform distribution and $k \leq d$ then we obtain the standard volume sampling [17] where one picks a set S proportional to $\det(V_S V_S^\top)$, or, equivalently, to the volume of the parallelopiped spanned by the vectors indexed by S. The volume sampling measure has received much attention and efficient algorithms are known for sampling from it [17, 21]. More recently, efficient algorithms were obtained even when $k \geq d$ [26, 36]. We discuss the computational issues of sampling from proportional volume sampling in Lemma 1.1 and Section 6.2.

Our first result shows that approximating the A-optimal design problem can be reduced to finding distributions on $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$) that are *approximately independent*. First, we define the exact formulation of approximate independence needed in our setting.

Definition 1.2. *Given integers $d \leq k \leq n$ and a vector $x \in [0,1]^n$ such that $1^\top x = k$, we call a measure μ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$), α-approximate $(d-1,d)$-wise independent with respect to x if for any subsets $T, R \subseteq [n]$ with $|T| = d-1$ and $|R| = d$, we have*

$$\frac{\mathsf{Pr}_{\mathcal{S}\sim\mu}[T \subseteq \mathcal{S}]}{\mathsf{Pr}_{\mathcal{S}\sim\mu}[R \subseteq \mathcal{S}]} \leq \alpha \frac{x^T}{x^R}$$

where $x^L := \prod_{i \in L} x_i$ for any $L \subseteq [n]$. We omit "with respect to x" when the context is clear.

Observe that if the measure μ corresponds to picking each element i independently with probability x_i, then $\frac{\mathsf{Pr}_{\mathcal{S}\sim\mu}[T\subseteq\mathcal{S}]}{\mathsf{Pr}_{\mathcal{S}\sim\mu}[R\subseteq\mathcal{S}]} = \frac{x^T}{x^R}$. However, this distribution has support on all sets and not just sets in $\mathcal{U}_k$ or $\mathcal{U}_{\leq k}$, so it is not allowed by the definition above.

Our first result reduces the search for approximation algorithms for A-optimal design to construction of approximate $(d-1,d)$-wise independent distributions. This result generalizes the connection between volume sampling and A-optimal design established in [5] to proportional volume sampling, which allows us to exploit the power of the convex relaxation and get a significantly improved approximation.

Theorem 1.1. *Given integers $d \leq k \leq n$, suppose that for any a vector $x \in [0,1]^n$ such that $1^\top x = k$ there exists a distribution μ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$) that is α-approximate $(d-1,d)$-wise independent. Then the proportional volume sampling with measure μ gives an α-approximation algorithm for the A-optimal design problem.*

In the above theorem, we in fact only need an approximately independent distribution μ for the optimal solution x of the natural convex relaxation for the problem, which is given in (2.1)–(2.3). The result also bounds the integrality gap of the convex relaxation by α. Theorem 1.1 is proved in Section 2.

Theorem 1.1 reduces our aim to constructing distributions that have approximate $(d-1,d)$-independence. We focus our attention on the general class of *hard-core distributions*. We call μ a *hard-core* distribution with parameter $\lambda \in \mathbb{R}^n_+$ if $\mu(S) \propto \lambda^S := \prod_{i\in S} \lambda_i$ for each set in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$). Convex duality implies that hard-core distributions have the maximum entropy among all distributions which match the marginals of μ [10]. Observe that, while μ places non-zero probability on exponentially many sets, it is enough to specify μ succinctly by describing λ. Hard-core distributions over various structures including spanning trees [20] or matchings [23, 24] in a graph display *approximate independence* and this has found use in combinatorics as well as algorithm design. Following this theme, we show that certain hard core distributions on $\mathcal{U}_k$ and $\mathcal{U}_{\leq k}$ exhibit approximate $(d-1,d)$-independence when $k = d$ and in the asymptotic regime when $k >> d$.

THEOREM 1.2. *Given integers $d \le k \le n$ and a vector $x \in [0,1]^n$ such that $1^\top x = k$, there exists a hard-core distribution μ on sets in $\mathcal{U}_k$ that is d-approximate $(d-1, d)$-wise independent when $k = d$. Moreover, for any $\epsilon > 0$, if $k = \Omega\left(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$, then there is a hard-core distribution μ on $\mathcal{U}_{\le k}$ that is $(1+\epsilon)$-approximate $(d-1, d)$-wise independent. Thus we obtain a d-approximation algorithm for the A-optimal design problem when $k = d$ and $(1+\epsilon)$-approximation algorithm when $k = \Omega\left(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$.*

The above theorem relies on two natural hard-core distributions. In the first one, we consider the hard-core distribution with parameter $\lambda = x$ on sets in $\mathcal{U}_k$ and in the second we consider the hard-core distribution with parameter $\lambda = \frac{(1-\epsilon)x}{1-(1-\epsilon)x}$ (defined co-ordinate wise) on sets in $\mathcal{U}_{\le k}$. We prove the theorem in Section 3.

Our techniques also apply to the A-optimal design problem with repetitions where we obtain an even stronger result, described below. The main idea is to introduce multiple, possibly exponentially many, copies of each vector, depending on the fractional solution, and then apply proportional volume sampling to obtain the following result.

THEOREM 1.3. *For all $k \ge d$ and $0 < \epsilon \le 1$, there is a $(\frac{k}{k-d+1} + \epsilon)$-approximation algorithm for the A-optimal design problem with repetitions. In particular, there is a $(1+\epsilon)$-approximation when $k \ge d + \frac{d}{\epsilon}$.*

We remark that the integrality gap of the natural convex relaxation is at least $\frac{k}{k-d+1}$ (see Section 7.2) and thus the above theorem results in an exact characterization of the integrality gap of the convex program (2.1)–(2.3), stated in the following corollary. The proof of Theorem 1.3 appears in Section 6.3.

COROLLARY 1.1. *For any integers $k \ge d$, the integrality gap of the convex program (2.1)–(2.3) for the A-optimal design with repetitions is exactly $\frac{k}{k-d+1}$.*

We also show that A-optimal design is NP-hard for $k = d$ and moreover, hard to approximate within a constant factor.

THEOREM 1.4. *There exists a constant $c > 1$ such that the A-optimal design problem is NP-hard to c-approximate when $k = d$.*

The $k \le d$ case. The A-optimal design problem has a natural extension to choosing fewer than d vectors: our objective in this case is to select a set $S \subseteq [n]$ of size k so that we minimize $\sum_{i=1}^{k} \lambda_i^{-1}$, where $\lambda_1, \ldots, \lambda_k$ are the k largest eigenvalues of the matrix $V_S V_S^\top$. While this problem no longer corresponds to minimizing the variance in an experimental design setting, we will abuse terminology and still call it the A-optimal design problem. This is a natural formulation of the geometric problem of picking a set of vectors which are as "spread out" as possible. If $v_1, \ldots, v_n$ are the points in a dataset, we can see an optimal solution as a maximally diverse representative sample of the dataset. Similar problems, but with a determinant objective, have been widely studied in computational geometry, linear algebra, and machine learning: for example the largest volume simplex problem, and the maximum subdeterminant problem (see [29] for references to prior work). [12] also studied an analogous problem with the sum in the objective replaced by a maximum (which extends E-optimal design).

While our rounding extends easily to the $k \le d$ regime, coming up with a convex relaxation becomes less trivial. We do find such a relaxation and obtain the following result whose proof appears in Section 5.1.

THEOREM 1.5. *There exists a $\mathrm{poly}(d,n)$-time k-approximation algorithm for the A-optimal design problem when $k \le d$.*

General Objectives. Experimental design problems come with many different objectives including A, D, E, G, T, V, each corresponding to a different function of the covariance matrix of the error $w - \hat{w}$. We note that any algorithm that solves A-optimal objective can solve V-optimal objective by prepossessing vectors with a linear transformation. In addition, we show that the proportional volume sampling algorithm gives approximation algorithms for other optimal design objectives (such as D-optimal design [36] and generalized ratio objective [27]) matching or improving previous best known results. We refer the reader to Section 5.3 for details.

Integrality Gap and E-optimality. Given the results mentioned above, a natural question is whether all objectives for optimal design behave similarly in terms of approximation algorithms. Indeed, recent results of [2, 1] and [43] give the $(1+\epsilon)$-approximation algorithm in the asymptotic regime, $k \ge \Omega\left(\frac{d}{\epsilon^2}\right)$ and $k \ge \Omega\left(\frac{d^2}{\epsilon}\right)$, for many of these variants. In contrast, we show the *optimal bounds* that can be obtained via the standard convex relaxation are different for different objectives. We show that for the E-optimality criterion (in which we minimize the largest eigenvalue of the covariance matrix) getting a $(1+\epsilon)$-approximation with the natural convex relaxation requires $k = \Omega(\frac{d}{\epsilon^2})$, both with and without repetitions. This is in sharp contrast to results we obtain here for A, D-optimality and other generalized ratio objectives. Thus, different criteria behave differently in terms of approximability. Our proof of the integrality gap (in Section 7.1) builds on a connection to spectral graph theory and in particular on the Alon-Boppana bound [3, 34]. We prove an Alon-Boppana style bound for the unnormalized Laplacian of not necessarily regular graphs with a given average degree.

Restricted Invertibility Principle for Harmonic Mean. As an application of Theorem 1.5, we prove a restricted invertibility principle (RIP) [7] for the harmonic mean of singular values. The RIP is a robust version of the elementary fact in linear algebra that if V is a $d \times n$ rank r matrix, then it has an invertible submatrix V_S for some $S \subseteq [n]$ of size r. The RIP shows that if V has stable rank r, then it has a well-invertible submatrix consisting of $\Omega(r)$ columns. Here the stable rank of V is the ratio $(\|V\|_{HS}^2/\|V\|^2)$, where $\|\cdot\|_{HS} = \sqrt{\mathrm{tr}(VV^\top)}$ is the Hilbert-Schmidt, or Frobenius, norm of V, and $\|\cdot\|$ is the operator norm. The classical restricted invertibility principle [7, 42, 37] shows that when the stable rank of V is r, then there exists a subset of its columns S of size $k = \Omega(r)$ so that the k-th singular value of V_S is $\Omega\left(\|V\|_{HS}/\sqrt{m}\right)$. [29] showed there exists a submatrix V_S of k columns of V so that the geometric mean its top k singular values is on the same order, even when k equals the stable rank. We show an analogous result for the harmonic mean when k is slightly less than r. While this is implied by the classical restricted invertibility principle, the dependence on parameters is better in our result for the harmonic mean. For example, when $k = (1-\epsilon)r$, the harmonic mean of squared singular values of V_S can be made at least $\Omega\left(\epsilon\|V\|_{HS}^2/m\right)$, while the tight restricted invertibility principle of Spielman and Srivastava [38] would only give ϵ^2 in the place of ϵ. This restricted invertibility principle can also be derived from the results of [28], but their arguments, unlike ours, do not give an efficient algorithm to compute the submatrix V_S. See Section 5.2 for the precise formulation of our restricted invertibility principle.

Computational Issues. While it is not clear whether sampling from proportional volume sampling is possible under general assumptions (for example given a sampling oracle for μ), we obtain an efficient sampling algorithm when μ is a hard-core distribution.

LEMMA 1.1. *There exists a* $\mathrm{poly}(d, n)$*-time algorithm that, given a matrix $d \times n$ matrix V, integer $k \le n$, and a hard-core distribution μ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\le k}$) with parameter λ, efficiently samples a set from the proportional volume measure defined by μ.*

When $k \le d$ and μ is a hard-core distribution, the proportional volume sampling can be implemented by the standard volume sampling after scaling the vectors appropriately. When $k > d$, such a method does not suffice and we appeal to properties of hard-core distributions to obtain the result. We also present an efficient implementation of Theorem 1.3 which runs in time polynomial in $\log(1/\epsilon)$. This requires more work since the basic description of the algorithm involves implementing proportional volume sampling on an exponentially-sized ground set. This is done in Section 6.3.

We also outline efficient deterministic implementation of algorithms in Theorem 1.2 and 1.3 in Section 6.2 and 6.4.

1.2 Related Work

Experimental design is the problem of maximizing information obtained from selecting subsets of experiments to perform, which is equivalent to minimizing the covariance matrix $\left(\sum_{i\in S} v_i v_i^\top\right)^{-1}$. We focus on A-optimality, one of the criteria that has been studied intensely. We restrict our attention to approximation algorithms for these problems and refer the reader to [35] for a broad survey on experimental design.

[5] studied the A- and E-optimal design problems and analyzed various combinatorial algorithms and algorithms based on volume sampling, and achieved approximation ratio $\frac{n-d+1}{k-d+1}$. [43] found connections between optimal design and matrix sparsification, and used these connections to obtain a $(1+\epsilon)$-approximation when $k \ge \frac{d^2}{\epsilon}$, and also approximation algorithms under certain technical assumptions. More recently, [2, 1] obtained a $(1+\epsilon)$-approximation when $k = \Omega\left(\frac{d}{\epsilon^2}\right)$ both with and without repetitions. We remark that their result also applies to other criteria such as E and D-optimality that aim to maximize the minimum eigenvalue, and the geometric mean of the eigenvalues of $\sum_{i\in S} v_i v_i^\top$, respectively. More generally, their result applies to any objective function that satisfies certain regularity criteria.

Improved bounds for D-optimality were obtained by [36] who give an e-approximation for all k and d, and $(1+\epsilon)$-approximation algorithm when $k = \Omega(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$, with a weaker condition of $k \ge \frac{2d}{\epsilon}$ if repetitions are allowed. The D-optimality criterion when $k \le d$ has also been extensively studied. It captures maximum a-posteriori inference in constrained determinantal point process models [25], and also the maximum volume simplex problem. [29], improving on a long line of work, gave a e-approximation. The problem has also been studied under more general matroid constraints rather than cardinality constraints [30, 4, 40].

[12] also studied several related problems in the $k \le d$ regime, including D- and E-optimality. We are not aware of any prior work on A-optimality in this regime.

The criterion of E-optimality, whose objective is to maximize the minimum eigenvalue of $\sum_{i\in S} v_i v_i^\top$, is closely related to the problem of matrix sparsification [6, 38] but incomparable. In matrix sparsification, we are allowed to weigh the selected vectors, but need to bound both the largest and the smallest eigenvalue of the matrix we output.

The restricted invertibility principle was first proved in the work of [7], and was later strengthened by [42], [37], and [28]. Spielman and Srivastava gave a deterministic algorithm to find the well-invertible submatrix whose existence is guaranteed by the theorem. Besides its numerous applications in geometry (see [42] and [44]), the principle has also found applications to differential privacy [33], and to approximation algorithms for discrepancy [32].

Volume sampling where a set S is sampled with probability proportional to $\det(V_S V_S^\top)$ has been studied extensively and efficient algorithms were given by [17] and improved by [21]. The probability distribution is also called a *determinantal point process* (DPP) and finds many applications in machine learning [25]. Recently, fast algorithms for volume sampling have been considered in [15, 16].

While NP-hardness is known for the D- and E-optimality criteria [12], to the best of our knowledge no NP-hardness for A-optimality was known prior to our work. Proving such a hardness result was stated as an open problem in [5].

2 Approximation via Near Independent Distributions

In this section, we prove Theorem 1.1 and give an α-approximation algorithm for the A-optimal design problem given an α-approximate $(d-1, d)$-independent distribution μ.

We first consider the convex relaxation for the problem given below for the settings without and with repetitions. This relaxation is classical, and already appears in, e.g. [11]. It is easy to see that the objective $\operatorname{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$ is convex ([10], section 7.5). For this section, we focus on the case when repetitions are not allowed. The relaxation when repetitions are allowed is by replacing constraints $0 \le x_i \le 1$ by $0 \le x_i$.

$$\text{(2.1)} \qquad \min \operatorname{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$$

$$\text{(2.2)} \qquad \text{s.t.} \quad \sum_{i=1}^n x_i = k$$

$$\text{(2.3)} \qquad 0 \le x_i \le 1 \quad \forall i \in [n]$$

Let us denote the optimal value of (2.1)–(2.3) by CP. By plugging in the indicator vector of an optimal integral solution for x, we see that $\mathsf{CP} \le \mathsf{OPT}$, where OPT denotes the value of the optimal solution.

2.1 Approximately Independent Distributions

Let us use the notation $x^S = \prod_{i\in S} x_i$, V_S a matrix of column vectors $v_i \in \mathbb{R}^d$ for $i \in S$, and $V_S(x)$ a matrix of column vectors $\sqrt{x_i} v_i \in \mathbb{R}^d$ for $i \in S$. Let $e_k(x_1, \ldots, x_n)$ be the degree k elementary symmetric polynomial in the variables $x_1, \ldots, x_n$, i.e. $e_k(x_1, \ldots, x_n) = \sum_{S\in\mathcal{U}_k} x^S$. By convention, $e_0(x) = 1$ for any x. For any positive semidefinite $n \times n$ matrix M, we define $E_k(M)$ to be $e_k(\lambda_1, \ldots, \lambda_n)$, where $\lambda(M) = (\lambda_1, \ldots, \lambda_n)$ is the vector of eigenvalues of M. Notice that $E_1(M) = \operatorname{tr}(M)$ and $E_n(M) = \det(M)$.

To prove Theorem 1.1, we give Algorithm 1 which is a general framework to sample S to solve the A-optimal design problem.

Algorithm 1 The proportional volume sampling algorithm

1: Given an input $V = [v_1, \ldots, v_n]$ where $v_i \in \mathbb{R}^d$, k a positive integer, and measure μ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\le k}$).
2: Solve convex relaxation CP to get a fractional solution $x \in \mathbb{R}^n_+$ with $\sum_{i=1}^n x_i = k$.
3: Sample set $\mathcal{S}$ (from $\mathcal{U}_{\le k}$ or $\mathcal{U}_k$) where $\Pr[\mathcal{S} = S] \propto \mu(S)\det(V_S V_S^\top)$ for any $S \in \mathcal{U}_k$ (or $\mathcal{U}_{\le k}$). $\triangleright$ $\mu(S)$ may be defined using the solution x
4: Output $\mathcal{S}$ (If $|\mathcal{S}| < k$, add $k - |\mathcal{S}|$ arbitrary vectors to $\mathcal{S}$ first).

We first prove the following lemma which is needed for proving Theorem 1.1.

Lemma 2.1. *Let $T \subseteq [n]$ be of size no more than d. Then*

$$\det(V_T(x)^\top V_T(x)) = x^T \det(V_T^\top V_T)$$

The proof is presented in the full version of the paper [31].

We also need the following identity, which is well-known and extends the Cauchy-Binet formula for the determinant to the functions E_k.

$$\text{(2.4)} \qquad E_k(VV^\top) = E_k(V^\top V) = \sum_{S\in\mathcal{U}_k} \det(V_S^\top V_S).$$

The identity (2.4) appeared in [27] and, specifically for $k = d-1$, as Lemma 3.8 in [5]. Now we are ready to prove Theorem 1.1.

Proof of Theorem 1.1: Let μ' denote the sampling distribution over $\mathcal{U}$, where $\mathcal{U} = \mathcal{U}_k$ or $\mathcal{U}_{\le k}$, with probability of sampling $S \in \mathcal{U}$ proportional to $\mu(S)\det(V_S V_S^\top)$. Because $\operatorname{tr}\left(\sum_{i\in[n]} x_i v_i v_i^\top\right)^{-1} = \mathsf{CP} \le \mathsf{OPT}$, it is enough to show that

$$\mathbb{E}_{\mathcal{S}\sim\mu'}\left[\operatorname{tr}\left(\sum_{i\in\mathcal{S}} v_i v_i^\top\right)^{-1}\right] \le \alpha \operatorname{tr}\left(\sum_{i\in[n]} x_i v_i v_i^\top\right)^{-1}. \tag{2.5}$$

Note that in case $|\mathcal{S}| < k$, algorithm $\mathcal{A}$ adds $k - |\mathcal{S}|$ arbitrary vector to $\mathcal{S}$, which can only decrease the objective value of the solution.

First, a simple but important observation ([5]): for any $d \times d$ matrix M of rank d, we have

$$\operatorname{tr} M^{-1} = \sum_{i=1}^{d} \frac{1}{\lambda_i(M)} = \frac{e_{d-1}(\lambda(M))}{e_d(\lambda(M))} = \frac{E_{d-1}(M)}{\det M}. \tag{2.6}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{S}\sim\mu'}\left[\operatorname{tr}\left(\sum_{i\in\mathcal{S}} v_i v_i^\top\right)^{-1}\right] \\
&= \sum_{S\in\mathcal{U}} \Pr_{\mu'}[\mathcal{S} = S] \operatorname{tr}\left(V_S V_S^\top\right)^{-1} \\
&= \sum_{S\in\mathcal{U}} \frac{\mu(S)\det\left(V_S V_S^\top\right)}{\sum_{S'\in\mathcal{U}} \mu(S')\det(V_{S'} V_{S'}^\top)} \frac{E_{d-1}(V_S V_S^\top)}{\det\left(V_S V_S^\top\right)} \\
&= \frac{\sum_{S\in\mathcal{U}} \mu(S) E_{d-1}(V_S V_S^\top)}{\sum_{S\in\mathcal{U}} \mu(S)\det(V_S V_S^\top)}.
\end{aligned}$$

We can now apply the Cauchy-Binet formula (2.4) for E_{d-1}, $E_d = \det$, and the matrix $V_S V_S^\top$ to the numerator and denominator on the right hand side, and we get

$$\begin{aligned}
&\mathbb{E}_{\mathcal{S}\sim\mu'}\left[\operatorname{tr}\left(\sum_{i\in\mathcal{S}} v_i v_i^\top\right)^{-1}\right] \\
&= \frac{\sum_{S\in\mathcal{U}} \sum_{|T|=d-1, T\subseteq S} \mu(S)\det(V_T^\top V_T)}{\sum_{S\in\mathcal{U}} \mu(S) \sum_{|R|=d, R\subseteq S} \det(V_R^\top V_R)} \\
&= \frac{\sum_{|T|=d-1, T\subseteq[n]} \det\left(V_T^\top V_T\right) \sum_{S\in\mathcal{U}, S\supseteq T} \mu(S)}{\sum_{|R|=d, R\subseteq[n]} \det\left(V_R^\top V_R\right) \sum_{S\in\mathcal{U}, S\supseteq R} \mu(S)} \\
&= \frac{\sum_{|T|=d-1, T\subseteq[n]} \det\left(V_T^\top V_T\right) \Pr_{\mu}[\mathcal{S} \supseteq T]}{\sum_{|R|=d, R\subseteq[n]} \det\left(V_R^\top V_R\right) \Pr_{\mu}[\mathcal{S} \supseteq R]}
\end{aligned}$$

where we change the order of summation at the second to last equality. Next, we apply (2.6) and the Cauchy-Binet formula (2.4) in a similar way to the matrix $V(x)V(x)^\top$:

$$\begin{aligned}
\operatorname{tr}\left(V(x)V(x)^\top\right)^{-1} &= \frac{E_{d-1}(V(x)V(x)^\top)}{\det(V(x)V(x)^\top)} \\
&= \frac{\sum_{|T|=d-1, T\subseteq[n]} \det(V_T(x)^\top V_T(x))}{\sum_{|R|=d, R\subseteq[n]} \det(V_R(x)^\top V_R(x))} \\
&= \frac{\sum_{|T|=d-1, T\subseteq[n]} \det\left(V_T^\top V_T\right) x^T}{\sum_{|R|=d, R\subseteq[n]} \det\left(V_R^\top V_R\right) x^R}
\end{aligned}$$

where we use the fact that $\det(V_R(x)^\top V_R(x)) = x^R \det(V_R^\top V_R)$ and $\det(V_T(x)^\top V_T(x)) = x^T \det(V_T^\top V_T)$ in the last equality by Lemma 2.1.

Hence, the inequality (2.5) which we want to show is equivalent to

$$\frac{\sum_{|T|=d-1, T\subseteq[n]} \det\left(V_T^\top V_T\right) \Pr_{\mu}[\mathcal{S} \supseteq T]}{\sum_{|R|=d, R\subseteq[n]} \det\left(V_R^\top V_R\right) \Pr_{\mu}[\mathcal{S} \supseteq R]} \le \alpha \frac{\sum_{|T|=d-1, T\subseteq[n]} \det\left(V_T^\top V_T\right) x^T}{\sum_{|R|=d, R\subseteq[n]} \det\left(V_R^\top V_R\right) x^R} \tag{2.7}$$

which is equivalent to

$$\begin{aligned}
&\sum_{|T|=d-1, |R|=d} \det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right) \cdot x^R \cdot \Pr_{\mu}[\mathcal{S} \supseteq T] \\
&\le \alpha \sum_{|T|=d-1, |R|=d} \det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right) \cdot x^T \cdot \Pr_{\mu}[\mathcal{S} \supseteq R].
\end{aligned} \tag{2.8}$$

By the assumption that $\frac{\Pr_{\mu}[\mathcal{S}\supseteq T]}{\Pr_{\mu}[\mathcal{S}\supseteq R]} \le \alpha \frac{x^T}{x^R}$ for each subset $T, R \subseteq [n]$ with $|T| = d-1$ and $|R| = d$,

$$\begin{aligned}
&\det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right) \cdot x^R \cdot \Pr_{\mu}[\mathcal{S} \supseteq T] \\
&\le \alpha \det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right) \cdot x^T \cdot \Pr_{\mu}[\mathcal{S} \supseteq R]
\end{aligned} \tag{2.9}$$

Summing (2.9) over all T, R proves (2.8). $\square$

3 Approximating Optimal Design without Repetitions

In this section, we prove Theorem 1.2 by constructing α-approximate $(d-1, d)$-independent distributions for appropriate values of α. We first consider the case when $k = d$ and then the asymptotic case when $k = \Omega\left(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$. We also remark that the argument for $k = d$ can be generalized for all $k \le d$, and we discuss this generalization in Section 5.1.

3.1 d-approximation for $k = d$ We prove the following lemma which, together with Theorem 1.1, implies the d-approximation for A-optimal design when $k = d$.

LEMMA 3.1. *Let $k = d$. The hard-core distribution μ on $\mathcal{U}_k$ with parameter x is d-approximate $(d-1, d)$-independent.*

Proof. Observe that for any $S \in \mathcal{U}_k$, we have $\mu(S) = \frac{x^S}{Z}$ where $Z = \sum_{S' \in \mathcal{U}_k} x^{S'}$ is the normalization factor. For any $T \subseteq [n]$ such that $|T| = d-1$, we have

$$\Pr_{\mathcal{S}\sim\mu}[\mathcal{S} \supseteq T] = \sum_{S \in \mathcal{U}_k : S \supseteq T} \frac{x^S}{Z} = \frac{x^T}{Z} \cdot \left(\sum_{i \in [n]\setminus T} x_i \right) \le d\frac{x^T}{Z}.$$

where we use $k = d$ and $\sum_{i\in[n]\setminus T} x_i \le k = d$. For any $R \subseteq [n]$ such that $|R| = d$, we have

$$\Pr_{\mathcal{S}\sim\mu}[\mathcal{S} \supseteq R] = \sum_{S \in \mathcal{U}_k : S \supseteq R} \frac{x^S}{Z} = \frac{x^R}{Z}.$$

Thus for any $T, R \subseteq [n]$ such that $|T| = d-1$ and $|R| = d$, we have

$$\frac{\Pr_{\mathcal{S}\sim\mu}[\mathcal{S} \supseteq T]}{\Pr_{\mathcal{S}\sim\mu}[\mathcal{S} \supseteq R]} \le d\frac{x^T}{x^R}.$$

3.2 $(1+\epsilon)$-approximation Now, we show that there is a hard-core distribution μ on $\mathcal{U}_{\le k}$ that is $(1+\epsilon)$-approximate $(d-1, d)$-independent when $k = \Omega\left(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$.

LEMMA 3.2. *Fix some $0 < \epsilon \le 2$, and let $k = \Omega\left(\frac{d}{\epsilon} + \frac{\log(1/\epsilon)}{\epsilon^2}\right)$. The hard-core distribution μ on $\mathcal{U}_{\le k}$ with parameter λ, defined by*

$$\lambda_i = \frac{x_i}{1 + \frac{\epsilon}{4} - x_i},$$

is $(1+\epsilon)$-approximate $(d-1, d)$-wise independent.

Proof. For simplicity of notation, let us denote $\beta = 1 + \frac{\epsilon}{4}$, and $\xi_i = \frac{x_i}{\beta}$. Observe that the probability mass under μ of any set S of size at most k is proportional to $\left(\prod_{i\in S}\xi_i\right)\left(\prod_{i\notin S}(1-\xi_i)\right)$. Thus, μ is equivalent to the following distribution: sample a set $\mathcal{B} \subseteq [n]$ by including every $i \in [n]$ in $\mathcal{B}$ independently with probability ξ_i; then we have $\mu(S) = \Pr[\mathcal{B} = S \mid |\mathcal{B}| \le k]$ for every S of size at most k. Let us fix for the rest of the proof arbitrary sets $T, R \subseteq [n]$ of size $d-1$ and d, respectively. By the observation above, for $\mathcal{S}$ sampled according to μ, and $\mathcal{B}$ as above, we have

$$\frac{\Pr[\mathcal{S} \supseteq T]}{\Pr[\mathcal{S} \supseteq R]} = \frac{\Pr[\mathcal{B} \supseteq T \text{ and } |\mathcal{B}| \le k]}{\Pr[\mathcal{B} \supseteq R \text{ and } |\mathcal{B}| \le k]} \le \frac{\Pr[\mathcal{B} \supseteq T]}{\Pr[\mathcal{B} \supseteq R \text{ and } |\mathcal{B}| \le k]}$$

We have $\Pr[\mathcal{B} \supseteq T] = \xi^T = \frac{x^T}{\beta^{d-1}}$. To simplify the probability in the denominator, let us introduce, for each $i \in [n]$, the indicator random variable Y_i, defined to be 1 if $i \in \mathcal{B}$ and 0 otherwise. By the choice of $\mathcal{B}$, the Y_i's are independent Bernoulli random variables with mean ξ_i, respectively. We can write

$$\begin{aligned}
&\Pr[\mathcal{B} \supseteq R \text{ and } |\mathcal{B}| \le k] \\
&= \Pr\left[\forall i \in R : Y_i = 1 \text{ and } \sum_{i\notin R} Y_i \le k-d\right] \\
&= \Pr[\forall i \in R : Y_i = 1] \Pr\left[\sum_{i\notin R} Y_i \le k-d\right],
\end{aligned}$$

where the last equality follows by the independence of the Y_i. The first probability on the right hand side is just $\xi^R = \frac{x^R}{\beta^d}$, and plugging into the inequality above, we get

$$\frac{\Pr[\mathcal{S} \supseteq T]}{\Pr[\mathcal{S} \supseteq R]} \le \beta \frac{x^T}{x^R \Pr[\sum_{i\notin R} Y_i \le k-d]}. \tag{3.10}$$

We claim that

$$\Pr[\sum_{i\notin R} Y_i \le k-d] \ge 1 - \frac{\epsilon}{4}$$

as long as $k = \Omega\left(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$. The proof follows from standard concentration of measure arguments. Let $Y = \sum_{i\notin R} Y_i$, and observe that $\mathbb{E}[Y] = \frac{1}{\beta}(k - x(R))$, where $x(R)$ is shorthand for $\sum_{i\in R} x_i$. By Chernoff's bound,

$$\Pr[Y > k-d] < e^{-\frac{\delta^2}{3\beta}(k - x(R))} \tag{3.11}$$

where

$$\delta = \frac{\beta(k-d)}{k - x(R)} - 1 = \frac{(\beta-1)k + x(R) - \beta d}{k - x(R)}.$$

The exponent on the right hand side of (3.11) simplifies to

$$\begin{aligned}
\frac{\delta^2(k - x(R))}{3\beta} &= \frac{((\beta-1)k + x(R) - \beta d)^2}{3\beta(k - x(R))} \\
&\ge \frac{((\beta-1)k - \beta d)^2}{3\beta k}
\end{aligned}$$

For the bound $\Pr[Y > k-d] \le \frac{\epsilon}{4}$, it suffices to have

$$(\beta-1)k - \beta d \ge \sqrt{3\beta\log(4/\epsilon)k}.$$

Assuming that $k \ge \frac{C\log(4/\epsilon)}{\epsilon^2}$ for a sufficiently big constant C, the right hand side is at most $\frac{\epsilon k}{8}$. So, as long as $k \geq \frac{\beta d}{\beta - 1 - \frac{\epsilon}{8}}$, the inequality is satisfied and $\Pr[Y > k-d] < \frac{\epsilon}{4}$, as we claimed.

The proof of the lemma now follows since for any $|T| = d-1$ and $|R| = d$, we have

(3.12)
$$\frac{\Pr[\mathcal{S} \supseteq T]}{\Pr[\mathcal{S} \supseteq R]} \leq \beta \frac{x^T}{x^R \Pr[\sum_{i \notin R} Y_i \leq k-d]} \leq \frac{1+\frac{\epsilon}{4}}{1-\frac{\epsilon}{4}} \frac{x^T}{x^R},$$

and $\frac{1+\frac{\epsilon}{4}}{1-\frac{\epsilon}{4}} \leq 1+\epsilon$.

The $(1+\epsilon)$-approximation for large enough k in Theorem 1.2 now follows directly from Lemma 3.2 and Theorem 1.1.

4 Approximately Optimal Design with Repetitions

In this section, we consider the A-optimal design without the bound $x_i \leq 1$ and prove Theorem 1.3. That is, we allow the sample set S to be a multi-set. We obtain a tight bound on the integrality gap in this case. Interestingly, we reduce the problem to a special case of A-optimal design without repetitions that allows us to obtained an improved approximation.

We first describe a sampling Algorithm 2 that achieves a $\frac{k(1+\epsilon)}{k-d+1}$-approximation for any $\epsilon > 0$. In the algorithm, we introduce $\mathrm{poly}(n, 1/\epsilon)$ number of copies of each vector to ensure that the fractional solution assigns equal fractional value for each copy of each vector. Then we use the proportional volume sampling where the measure distribution μ is defined on sets of the new larger ground set U over copies of the original input vectors. The distribution μ is just the uniform distribution over subsets of size k of U, and we are effectively using traditional volume sampling over U. Notice, however, that the distribution over multisets of the original set of vectors is different. The proportional volume sampling used in the algorithm can be implemented in the same way as the one used for without repetition setting, as described in Section 6.1, which runs in $\mathrm{poly}(n, d, k, 1/\epsilon)$ time.

In Section 6.3, we describe a new implementation of proportional volume sampling procedure which improves the running time to $\mathrm{poly}(n, d, k, \log(1/\epsilon))$. The new algorithm is still efficient even when U has exponential size by exploiting the facts that μ is uniform and that U has only at most n distinct vectors.

Lemma 4.1. *Algorithm 2, when given as input $x \in \mathbb{R}^n_+$ s.t. $\sum_{i=1}^n x_i = k$, $1 \geq \epsilon > 0$, and $v_1, \ldots, v_n$, outputs a random $X \in \mathbb{Z}^n_+$ with $\sum_{i=1}^n X_i = k$ such that*

$$\mathbb{E}\left[\mathrm{tr}\left(\sum_{i=1}^n X_i v_i v_i^\top\right)^{-1}\right] \leq \frac{k(1+\epsilon)}{k-d+1} \mathrm{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$$

Algorithm 2 Approximation Algorithm for A-optimal design with repetitions

1: Given $x \in \mathbb{R}^n_+$ with $\sum_{i=1}^n x_i = k$, $\epsilon > 0$, and vectors $v_1, \ldots, v_n$.
2: Let $q = \frac{2n}{\epsilon k}$. Set $x'_i := \frac{k - n/q}{k} x_i$ for each i, and round each x'_i up to a multiple of $1/q$.
3: If $\sum_{i=1}^n x'_i < k$, add $1/q$ to any x'_i until $\sum_{i=1}^n x'_i = k$.
4: Create qx'_i copies of vector v_i for each $i \in [n]$. Denote W the set of size $\sum_{i=1}^n qx'_i = qk$ of all those copies of vectors. Denote U the new index set of $W$of size qk. ▷ This implies that we can assume that our new fractional solution $y_i = 1/q$ is equal over all $i \in U$
5: Sample a subset $\mathcal{S}$ of U of size k where $\Pr[\mathcal{S} = S] \propto \det(W_S W_S^\top)$ for each $S \subseteq U$ of size k.
6: Set $X_i = \sum_{w \in W_{\mathcal{S}}} \mathbb{1}(w \text{ is a copy of } v_i)$ for all $i \in [n]$ ▷ Get an integral solution X by counting numbers of copies of v_i in $\mathcal{S}$.
7: Output X.

Proof. Define $x'_i, y, W, U, \mathcal{S}, X$ as in the algorithm. We will show that

$$\mathbb{E}\left[\mathrm{tr}\left(\sum_{i=1}^n X_i v_i v_i^\top\right)^{-1}\right] \leq \frac{k}{k-d+1} \mathrm{tr}\left(\sum_{i=1}^n x'_i v_i v_i^\top\right)^{-1}$$
$$\leq \frac{k(1+\epsilon)}{k-d+1} \mathrm{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$$

The second inequality is by observing that the scaling $x'_i := \frac{k-n/q}{k} x_i$ multiplies the objective $\mathrm{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$ by a factor of $\left(\frac{k-n/q}{k}\right)^{-1} = (1 - \epsilon/2)^{-1} \leq 1+\epsilon$, and that rounding x_i up and adding $1/q$ to any x_i can only decrease the objective.

To show the first inequality, we first translate the two key quantities $\mathrm{tr}\left(\sum_{i=1}^n x'_i v_i v_i^\top\right)^{-1}$ and $\mathrm{tr}\left(\sum_{i=1}^n X_i v_i v_i^\top\right)^{-1}$ from the with-repetition setting over V and $[n]$ to the without-repetition setting over W and U. First, $\mathrm{tr}\left(\sum_{i=1}^n x'_i v_i v_i^\top\right)^{-1} = \mathrm{tr}\left(\sum_{i \in U} y_i w_i w_i{}^\top\right)^{-1}$, where $y_i = \frac{1}{q}$ are all equal over all $i \in U$, and w_i is the copied vector in W at index $i \in U$. Second, $\mathrm{tr}\left(\sum_{i=1}^n X_i v_i v_i^\top\right)^{-1} = \mathrm{tr}\left(\sum_{i \in \mathcal{S} \subseteq U} w_i w_i{}^\top\right)^{-1}$.

Let μ' be the distribution over subsets S of U of size k defined by $\mu'(S) \propto \det(W_S W_S^\top)$. It is, therefore, sufficient to show that the sampling distribution μ'

satisfies

$$\mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\operatorname{tr}\left(\sum_{i\in\mathcal{S}\subseteq U} w_i {w_i}^\top\right)^{-1}\right]$$

$$\le \frac{k}{k-d+1}\operatorname{tr}\left(\sum_{i\in U} y_i w_i {w_i}^\top\right)^{-1} \tag{4.13}$$

Observe that μ' is the same as sampling a set $S\subseteq U$ of size k with probability proportional to $\mu(S)\det(W_S W_S^\top)$ where μ is uniform. Hence, by Theorem 1.1, it is enough to show that for all $T,R\subseteq U$ with $|T|=d-1, |R|=d$,

$$\frac{\Pr_\mu[\mathcal{S}\supseteq T]}{\Pr_\mu[\mathcal{S}\supseteq R]} \le \left(\frac{k}{k-d+1}\right)\frac{y^T}{y^R} \tag{4.14}$$

With μ being uniform and y_i being all equal to $1/q$, the calculation is straightforward:

$$\frac{\Pr_\mu[\mathcal{S}\supseteq T]}{\Pr_\mu[\mathcal{S}\supseteq R]} = \frac{\binom{qk-d+1}{k-d+1}/\binom{qk}{k}}{\binom{qk-d}{k-d}/\binom{qk}{k}} = \frac{qk-d+1}{k-d+1}, \text{and} \tag{4.15}$$

$$\frac{y^T}{y^R} = \frac{1}{y_i} = q$$

Therefore, (4.14) holds because

$$\frac{\Pr_\mu[\mathcal{S}\supseteq T]}{\Pr_\mu[\mathcal{S}\supseteq R]}\cdot\left(\frac{y^T}{y^R}\right)^{-1} = \frac{qk-d+1}{k-d+1}\cdot\frac{1}{q}$$

$$\le \frac{qk}{k-d+1}\cdot\frac{1}{q} = \frac{k}{k-d+1} \tag{4.16}$$

Remark 4.1. *The approximation ratio for A-optimality with repetitions for $k\ge d$ is tight, since it matches the integrality gap lower bound stated in Theorem 7.3.*

5 Generalizations

In this section we show that our arguments extend to the regime $k\le d$ and give a k-approximation (without repetitions), which matches the integrality gap of our convex relaxation. We also derive a restricted invertibility principle for the harmonic mean of eigenvalues. The proofs in this section are in the full version of the paper [31].

5.1 k-Approximation Algorithm for $k\le d$ Recall that our aim is to select a set $S\subseteq[n]$ of size $k\le d$ that minimizes $\sum_{i=1}^k \lambda_i^{-1}$, where $\lambda_1,\ldots,\lambda_k$ are the k largest eigenvalues of the matrix $V_S V_S^\top$. We need to reformulate our convex relaxation since when $k<d$, the inverse of $M(S)=\sum_{i\in S} v_i v_i^\top$ for $|S|=k$ is no longer well-defined. We write a new convex program:

$$\min \frac{E_{k-1}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)}{E_k\left(\sum_{i=1}^n x_i v_i v_i^\top\right)} \tag{5.17}$$

s.t.

$$\sum_{i=1}^n x_i = k \tag{5.18}$$

$$0\le x_i\le 1 \quad \forall i\in[n] \tag{5.19}$$

Once again we denote the optimal value of (5.17)–(5.19) by CP. While the proof that this relaxes the original problem is easy, the convexity is non-trivial. Fortunately, ratios of symmetric polynomials are known to be convex.

Lemma 5.1. *The optimization problem* (5.17)–(5.19) *is a convex relaxation of the A-optimal design problem when $k\le d$.*

We shall use the natural analog of proportional volume sampling: given a measure μ on subsets of size k, we sample a set S with probability proportional to $\mu(S)E_k(M(S))$. In fact, we will only take $\mu(S)$ proportional to x^S, so this reduces to sampling S with probability proportional to $E_k(\sum_{i\in S} x_i v_i v_i^\top)$, which is the standard volume sampling with vectors scaled by $\sqrt{x_i}$, and can be implemented efficiently using, e.g. the algorithm of [17].

The following version of Theorem 1.1 still holds with this modified proportional volume sampling. The proof is exactly the same, except for mechanically replacing every instance of determinant by E_k, of E_{d-1} by E_{k-1}, and in general of d by k.

Theorem 5.1. *Given integers $k\le d\le n$ and a vector $x\in[0,1]^n$ such that $1^\top x=k$, suppose there exists a measure μ on $\mathcal{U}_k$ that is α-approximate $(k-1,k)$-wise independent. Then for x the optimal solution of* (5.17)–(5.19), *proportional volume sampling with measure μ gives a polynomial time α-approximation algorithm for the A-optimal design problem.*

We can now give the main approximation guarantee we have for $k\le d$.

Theorem 5.2. *For any $k\le d$, proportional volume sampling with the hard-core measure μ on $\mathcal{U}_k$ with parameter x equal to the optimal solution of* (5.17)–(5.19) *gives a k-approximation to the A-optimal design problem.*

The algorithm can be derandomized using the method of conditional expectations analogously to the case of $k = d$ that we will show in Theorem 6.2.

The k-approximation also matches the integrality gap of (5.17)–(5.19). Indeed, we can take a k-dimensional integrality gap instance $v_1, \ldots, v_n$, and embed it in $\mathbb{R}^d$ for any $d > k$ by padding each vector with 0's. On such an instance, the convex program (5.17)–(5.19) is equivalent to the convex program (2.1)–(2.3). Thus the integrality gap that we will show in Theorem 7.3 implies an integrality gap of k for all $d \geq k$.

5.2 Restricted Invertibility Principle for Harmonic Mean Next we state and prove our restricted invertibility principle for harmonic mean in a general form.

THEOREM 5.3. *Let $v_1, \ldots, v_n \in \mathbb{R}^d$, and $c_1, \ldots, c_n \in \mathbb{R}_+$, and define $M = \sum_{i=1}^n c_i v_i v_i^\top$. For any $k \leq r = \frac{\operatorname{tr}(M)}{\|M\|}$, there exists a subset $S \subseteq [n]$ of size k such that the k largest eigenvalues $\lambda_1, \ldots, \lambda_k$ of the matrix $\sum_{i \in S} v_i v_i^\top$ satisfy*

$$\left(\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\lambda_i}\right)^{-1} \geq \frac{r-k+1}{r} \cdot \frac{\operatorname{tr}(M)}{\sum_{i=1}^n c_i}.$$

Moreover, such a set S can be computed in deterministic polynomial time.

We note that Theorem 5.3 also follows from Lemma 18 and and equation (12) of [28]. However, their proof of their Lemma 18 does not yield an efficient algorithm to compute the set S, as it relies on a volume maximization argument.

5.3 The Generalized Ratio Objective In A-optimal design, given $V = [v_1 \ldots v_n] \in \mathbb{R}^{d \times n}$, we state the objective as minimizing

$$\operatorname{tr}\left(\sum_{i \in S} v_i v_i^\top\right)^{-1} = \frac{E_{d-1}(V_S V_S^\top)}{E_d(V_S V_S^\top)}.$$

over subsets $S \subseteq [n]$ of size k. In this section, for any given pair of integers $0 \leq l' < l \leq d$, we consider the following *generalized ratio problem*:

$$\min_{S \subseteq [n], |S|=k} \left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}} \tag{5.20}$$

The above problem naturally interpolates between A-optimality and D-optimality. This follows since for $l = d$ and $l' = 0$, the objective reduces to

$$\min_{S \subseteq [n], |S|=k} \left(\frac{1}{\det(V_S V_S^\top)}\right)^{\frac{1}{d}}. \tag{5.21}$$

A closely related generalization between A- and D-criteria was considered in [27]. Indeed, their generalization corresponds to the case when $l = d$ and l' takes any value from 0 and $d-1$.

In this section, we show that our results extend to solving generalized ratio problem. We begin by describing a convex program for the generalized ratio problem. We then generalize the proportional volume sampling algorithm to *proportional l-volume sampling*. Following the same plan as in the proof of A-optimality, we then reduce the approximation guarantee to near-independence properties of certain distribution. Here again, we appeal to the same product measure and obtain identical bounds on the performance of the algorithm. The efficient implementations of approximation algorithms for generalized ratio problem are described in Section 6.5.

5.3.1 Convex Relaxation As in solving A-optimality, we may define a relaxation for without repetitions as (5.22)-(5.24). We replace $0 \leq x_i \leq 1$ by $0 \leq x_i$ when repetitions are allowed.

$$\min \left(\frac{E_{l'}\left(V(x)V(x)^\top\right)}{E_l\left(V(x)V(x)^\top\right)}\right)^{\frac{1}{l-l'}} \tag{5.22}$$

$$\text{s.t.} \quad \sum_{i=1}^n x_i = k \tag{5.23}$$

$$0 \leq x_i \leq 1 \quad \forall i \in [n] \tag{5.24}$$

We have that the objective $\left(\frac{E_{l'}\left(V(x)V(x)^\top\right)}{E_l\left(V(x)V(x)^\top\right)}\right)^{\frac{1}{l-l'}}$ remains convex in x.

LEMMA 5.2. *Let d be a positive integer. For any given pair $0 \leq l' < l \leq d$, the function*

$$f_{l',l}(M) = \left(\frac{E_{l'}(M)}{E_l(M)}\right)^{\frac{1}{l-l'}} \tag{5.25}$$

is convex over $d \times d$ positive semidefinite matrix M.

5.3.2 Approximation via (l', l)-Wise Independent Distribution Let $0 \leq l' < l \leq d$ and $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\leq k}\}$. We first show connection of approximation guarantees on objectives $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ and $\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}$. Suppose we already solve the convex relaxation of generalized ratio problem (5.22)-(5.24) and get a fractional solution $x \in \mathbb{R}^n$. Suppose that a randomized algorithm $\mathcal{A}$, upon receiving input $V \in \mathbb{R}^{d \times n}$ and $x \in \mathbb{R}^n$, outputs $S \in \mathcal{U}$ such that

$$\mathop{\mathbb{E}}_{S \sim \mathcal{A}}\left[\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right] \leq \alpha' \frac{E_{l'}(V(x)V(x)^\top)}{E_l(V(x)V(x)^\top)} \tag{5.26}$$

for some constant $\alpha' > 0$. By the convexity of the function $f(z) = z^{l-l'}$ over positive reals z, we have

$$\mathbb{E}\left[\frac{E_{l'}(M)}{E_l(M)}\right] \geq \mathbb{E}\left[\left(\frac{E_{l'}(M)}{E_l(M)}\right)^{\frac{1}{l-l'}}\right]^{l-l'} \tag{5.27}$$

for any semi-positive definite matrix M. Combining (5.26) and (5.27) gives

$$\mathop{\mathbb{E}}_{\mathcal{S}\sim\mathcal{A}}\left[\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}\right] \leq \alpha\left(\frac{E_{l'}(V(x)V(x)^\top)}{E_l(V(x)V(x)^\top)}\right)^{\frac{1}{l-l'}} \tag{5.28}$$

where $\alpha = (\alpha')^{\frac{1}{l-l'}}$. Therefore, it is sufficient for an algorithm to satisfy (5.26) and give a bound on α' in order to solve the generalized ratio problem up to factor α.

To show (5.26), we first define the proportional l-volume sampling and α-approximate (l', l)-wise independent distribution.

DEFINITION 5.1. *Let μ be probability measure on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$). Then the proportional l-volume sampling with measure μ picks a set of vectors indexed by $S \in \mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$) with probability proportional to $\mu(S)E_l(V_S V_S^\top)$.*

DEFINITION 5.2. *Given integers d, k, n, a pair of integers $0 \leq l' \leq l \leq d$, and a vector $x \in [0,1]^n$ such that $1^\top x = k$, we call a measure μ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$), α-approximate (l', l)-wise independent with respect to x if for any subsets $T', T \subseteq [n]$ with $|T'| = l'$ and $|T| = l$, we have*

$$\frac{\Pr_{\mathcal{S}\sim\mu}[T' \subseteq \mathcal{S}]}{\Pr_{\mathcal{S}\sim\mu}[T \subseteq \mathcal{S}]} \leq \alpha^{l-l'} \cdot \frac{x^{T'}}{x^T}$$

where $x^L := \prod_{i\in L} x_i$ for any $L \subseteq [n]$. We omit "with respect to x" when the context is clear.

The following theorem reduces the approximation guarantee in (5.26) to α-approximate (l', l)-wise independence properties of a certain distribution μ by utilizing proportional l-volume sampling.

THEOREM 5.4. *Given integers d, k, n, $V = [v_1 \dots v_n] \in \mathbb{R}^{d\times n}$, and a vector $x \in [0,1]^n$ such that $1^\top x = k$, suppose there exists a distribution μ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$) and is α-approximate (l', l)-wise independent for some $0 \leq l' < l \leq d$. Then the proportional l-volume sampling with measure μ gives an α-approximation algorithm for minimizing $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ over subsets $S \subseteq [n]$ of size k.*

The following subsections generalize algorithms and proofs for with and without repetitions. The algorithm for generalized ratio problem can be summarized in Algorithm 3. Note that efficient implementation of the sampling is described in Section 6.5.

Algorithm 3 Generalized ratio approximation algorithm

1: Given an input $V = [v_1, \dots, v_n]$ where $v_i \in \mathbb{R}^d$, k a positive integer, and a pair of integers $0 \leq l' < l \leq d$.
2: Solve the convex relaxation

$$x = \mathrm{argmin}_{x\in J^n : 1^\top x = k}\left(\frac{E_{l'}\left(V(x)V(x)^\top\right)}{E_l\left(V(x)V(x)^\top\right)}\right)^{\frac{1}{l-l'}}$$

where $J = [0,1]$ if without repetitions or $\mathbb{R}^+$ if with repetitions.
3: **if** $k = l$ **then**
4: Sample $\mu'(S) \propto x^S E_l\left(V_S V_S^\top\right)$ for each $S \in \mathcal{U}_k$
5: **else if** without repetition setting and $k \geq \Omega\left(\frac{d}{\epsilon} + \frac{\log(1/\epsilon)}{\epsilon^2}\right)$ **then**
6: Sample $\mu'(S) \propto \lambda^S E_l\left(V_S V_S^\top\right)$ for each $S \in \mathcal{U}_{\leq k}$ where $\lambda_i := \frac{x_i}{1+\epsilon/4-x_i}$
7: **else if** with repetition setting **then**
8: Run Algorithm 2, except modifying the sampling step to sample a subset $\mathcal{S}$ of U of size k with $\Pr[\mathcal{S} = S] \propto E_l(W_S W_S^\top)$.
9: Output $\mathcal{S}$ (If $|\mathcal{S}| < k$, add $k - |\mathcal{S}|$ arbitrary vectors to $\mathcal{S}$ first).

5.3.3 Approximation Guarantee for Generalized Ratio Problem without Repetitions

We prove the following theorem which generalizes Lemmas 3.1 and 3.2. The α-approximate (l', l)-wise independence property, together with Theorem 5.4, implies an approximation guarantee for generalized ratio problem without repetitions for $k = l$ and asymptotically for $k = \Omega\left(\frac{l}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$.

THEOREM 5.5. *Given integers d, k, n, a pair of integers $0 \leq l' \leq l \leq d$, and a vector $x \in [0,1]^n$ such that $1^\top x = k$, the hard-core distribution μ on sets in $\mathcal{U}_k$ with parameter x is α-approximate (l', l)-wise independent when $k = l$ for*

$$\alpha = l \cdot [(l-l')!]^{-\frac{1}{l-l'}} \leq \frac{el}{l-l'} \tag{5.29}$$

Moreover, for any $0 < \epsilon \leq 2$ when $k = \Omega\left(\frac{l}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$, the hard-core distribution μ on $\mathcal{U}_{\leq k}$ with parameter λ, defined by

$$\lambda_i = \frac{x_i}{1 + \frac{\epsilon}{4} - x_i},$$

is $(1+\epsilon)$-approximate (l', l)-wise independent.

Thus for minimizing the generalized ratio problem $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ over subsets $S \subseteq [n]$ of size k, we obtain

- *$(\frac{el}{l-l'})$-approximation algorithm when $k = l$, and*
- *$(1+\epsilon)$-approximation algorithm when $k = \Omega\left(\frac{l}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\right)$.*

5.3.4 Approximation Guarantee for Generalized Ratio Problem with Repetitions

We now consider the generalized ratio problem *with repetitions*. The following statement is a generalization of Lemma 4.1.

THEOREM 5.6. *Given $V = [v_1\ v_2 \dots v_n]$ where $v_i \in \mathbb{R}^d$, a pair of integers $0 \le l' \le l \le d$, an integer $k \ge l$, and $1 \ge \epsilon > 0$, there is an α-approximation algorithm for minimizing $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ over subsets $S \subseteq [n]$ of size k with repetitions for*

$$\alpha \le \frac{k(1+\epsilon)}{k-l+1} \tag{5.30}$$

We note that the l-proportional volume sampling in the proof of Theorem 5.6 can be implemented efficiently, and the proof is outlined in Section 6.5.

5.3.5 Integrality Gap

Finally, we state an integrality gap for minimizing generalized ratio objective $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ over subsets $S \subseteq [n]$ of size k. The integrality gap matches our approximation ratio of our algorithm with repetitions when k is large.

THEOREM 5.7. *For any given positive integers k, d and a pair of integers $0 \le l' \le l \le d$ with $k > l'$, there exists an instance $V = [v_1, \dots, v_n] \in \mathbb{R}^{d\times n}$ to the problem of minimizing $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ over subsets $S \subseteq [n]$ of size k such that*

$$\mathsf{OPT} \ge \left(\frac{k}{k-l'} - \delta\right) \cdot \mathsf{CP}$$

for all $\delta > 0$, where OPT denotes the value of the optimal integral solution and CP denotes the value of the convex program.

This implies that the integrality gap is at least $\frac{k}{k-l'}$ for minimizing $\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ over subsets $S \subseteq [n]$ of size k. The theorem applies to both with and without repetitions.

6 Efficient Algorithms

In this section, we outline efficient sampling algorithms, as well as deterministic implementations of our rounding algorithms, both for with and without repetition settings. The proofs in this section are in the full version of the paper [31].

6.1 Efficient Randomized Proportional Volume

Given a vector $\lambda \in \mathbb{R}^n_+$, we show that proportional volume sampling with $\mu(S) \propto \lambda^S$ for $S \in \mathcal{U}$, where $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\le k}\}$ can be done in time polynomial in the size n of the ground set. We start by stating a lemma which is very useful both for the sampling algorithms and the deterministic implementations.

LEMMA 6.1. *Let $\lambda \in \mathbb{R}^n_+, v_1, \dots, v_n \in \mathbb{R}^d$, and $V = [v_1, \dots, v_n]$. Let $I, J \subseteq [n]$ be disjoint. Let $1 \le k \le n, 0 \le d_0 \le d$. Consider the following function*

$$F(t_1,t_2,t_3) := \det\left(I_n + t_1 \operatorname{diag}(y) + t_1 t_2 \operatorname{diag}(y)^{1/2} V V^\top \operatorname{diag}(y)^{1/2}\right)$$

where $t_1, t_2, t_3 \in \mathbb{R}$ are indeterminate, I_n is the $n \times n$ identity matrix, and $y \in \mathbb{R}^n$ with

$$y_i = \begin{cases} \lambda_i t_3, & \text{if } i \in I \\ 0, & \text{if } i \in J \\ \lambda_i, & \text{otherwise} \end{cases}.$$

Then $F(t_1,t_2,t_3)$ is a polynomial and the quantity

$$\sum_{|S|=k, I\subseteq S, J\cap S=\emptyset} \lambda^S \sum_{|T|=d_0, T\subseteq S} \det(V_T^\top V_T) \tag{6.31}$$

is the coefficient of the monomial $t_1^k t_2^{d_0} t_3^{|I|}$. Moreover, this quantity can be computed in $O\left(n^3 d_0 k|I| \cdot \log(d_0 k|I|)\right)$ number of arithmetic operations.

Using the above lemma, we now prove the following theorem that will directly imply Lemma 1.1.

THEOREM 6.1. *Let $\lambda \in \mathbb{R}^n_+, v_1, \dots, v_n \in \mathbb{R}^d, 1 \le k \le n$, $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\le k}\}$, and $V = [v_1, \dots, v_n]$. Then there is a randomized algorithm $\mathcal{A}$ which outputs $\mathcal{S} \in \mathcal{U}$ such that*

$$\Pr_{\mathcal{S}\sim\mathcal{A}}[\mathcal{S} = S] = \frac{\lambda^S \det(V_S V_S^\top)}{\sum_{S'\in\mathcal{U}} \lambda^{S'} \det(V_{S'} V_{S'}^\top)} =: \mu'(S)$$

That is, the algorithm correctly implements proportional volume sampling μ' with hard-core measure μ on $\mathcal{U}$ with parameter λ. Moreover, the algorithm runs in $O\left(n^4 dk^2 \log(dk)\right)$ number of arithmetic operations.

Observation 6.1. *[43] shows that we may assume that the support of an extreme fractional solution of convex relaxation has size at most $k + d^2$. Thus, the runtime of proportional volume sampling is $O\left((k+d^2)^4 dk^2 \log(dk)\right)$. While the degrees in d, k are not small, this runtime is independent of n.*

Observation 6.2. *It is true in theory and observed in practice that solving the continuous relaxation rather than the rounding algorithm is a bottleneck in computation time, as discussed in [2]. In particular, solving the continuous relaxation of A-optimal design takes $O\left(n^{2+\omega}\log n\right)$ number of iterations by standard ellipsoid method and $O\left((n+d^2)^{3.5}\right)$ number of iterations by SDP, where $O(n^\omega)$ denotes the runtime of $n\times n$ matrix multiplication. In most applications where $n >> k$, these running times dominates one of proportional volume sampling.*

6.2 Efficient Deterministic Proportional Volume We show that for hard-core measures there is a deterministic algorithm that achieves the same objective value as the expected objective value achieved by proportional volume sampling. The basic idea is to use the method of conditional expectations.

Theorem 6.2. *Let $\lambda \in \mathbb{R}^n_+, v_1, \ldots, v_n \in \mathbb{R}^d, 1 \le k \le n$, $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\le k}\}$, and $V = [v_1, \ldots, v_n]$. Then there is a deterministic algorithm $\mathcal{A}'$ which outputs $S^* \subseteq [n]$ of size k such that*

$$\operatorname{tr}\left(V_{S^*}V_{S^*}^\top\right)^{-1} \ge \mathop{\mathbb{E}}_{\mu'}\left[\operatorname{tr}\left(V_\mathcal{S}V_\mathcal{S}^\top\right)^{-1}\right]$$

where μ' is the probability distribution defined by $\mu'(S) \propto \lambda^S \det(V_S V_S^\top)$ for all $S \in \mathcal{U}$. Moreover, the algorithm runs in $O\left(n^4 dk^2 \log(dk)\right)$ number of arithmetic operations.

Again, with the assumption that $n \le k + d^2$ (Observation 6.1), the runtime for deterministic proportional volume sampling is $O\left((k+d^2)^4 dk^2 \log(dk)\right)$.

6.3 Efficient Randomized Implementation of $\frac{k}{k-d+1}$-Approximation Algorithm With Repetitions First, we need to state several Lemmas needed to compute particular sums. The main motivation that we need a different method from Section 6.1 and 6.2 to compute a similar sum is that we want to allow the ground set U of indices of all copies of vectors to have an exponential size. This makes Lemma 6.1 not useful, as the matrix needed to be computed has dimension $|U| \times |U|$. The main difference, however, is that the parameter λ is now a constant, allowing us to obtain sums by computing a more compact $d \times d$ matrix.

Lemma 6.2. *Let $V = [v_1, \ldots, v_m]$ be a matrix of vectors $v_i \in \mathbb{R}^d$ with $n \ge d$ distinct vectors. Let $F \subseteq [m]$ and let $0 \le r \le d$ and $0 \le d_0 \le d$ be integers. Then the quantity $\sum_{|T|=d_0, |F\cap R|=r} \det(V_T^\top V_T)$ is the coefficient of $t_1^{d-d_0} t_2^{d_0-r} t_3^r$ in*

$$f(t_1,t_2,t_3) = \det\left(t_1 I_d + \sum_{i\in F} t_3 v_i v_i^\top + \sum_{i\notin F} t_2 v_i v_i^\top\right) \tag{6.32}$$

where $t_1, t_2, t_3 \in \mathbb{R}$ are indeterminate and I_d is the $d\times d$ identity matrix. Furthermore, this quantity can be computed in $O\left(n(d-d_0+1)d_0^2 d^2 \log d\right)$ number of arithmetic operations.

Lemma 6.3. *Let $V = [v_1, \ldots, v_m]$ be a matrix of vectors $v_i \in \mathbb{R}^d$ with $n \ge d$ distinct vectors. Let $F \subseteq [m]$ and let $0 \le r \le d$ and $0 \le d_0 \le d$ be integers. There is an algorithm to compute $\sum_{|S|=k, S\supseteq F} E_{d_0}(V_S V_S^\top)$ with $O\left(n(d-d_0+1)d_0^2 d^2 \log d\right)$ number of arithmetic operations.*

We now present an efficient sampling procedure for Algorithm 2. We want to sample S proportional to $\det(W_S W_S^\top)$. The set S is a subset of all copies of at most n distinct vectors, and there can be exponentially many copies. However, the key is that the quantity $f(t_1, t_2, t_3)$ in (6.32) is still efficiently computable because exponentially many of these copies of vectors are the same.

Theorem 6.3. *Given inputs $n, d, k, \epsilon, x \in \mathbb{R}^n_+$ with $\sum_{i=1}^n x_i = k$, and vectors $v_1, \ldots, v_n$ to Algorithm 2 we define q, U, W as in Algorithm 2. Then, there exists an implementation $\mathcal{A}$ that samples $\mathcal{S}$ from the distribution μ' over all subsets $S \subseteq U$ of size k, where μ' is defined by $\Pr_{\mathcal{S}\sim\mu'}[\mathcal{S} = S] \propto \det(W_S W_S^\top)$ for each $S \subseteq U, |S| = k$. Moreover, $\mathcal{A}$ runs in $O\left(n^2 d^4 k \log d\right)$ number of arithmetic operations.*

Theorem 6.3 says that steps (4)-(5) in Algorithm 2 can be efficiently implemented. Other steps except (4)-(5) obviously use $O\left(n^2 d^4 k \log d\right)$ number of arithmetic operations, so the above statement implies that Algorithm 2 runs in $O\left(n^2 d^4 k \log d\right)$ number of arithmetic operations. Again, by Observation 6.1, the number of arithmetic operations is in fact $O\left((k+d^2)^2 d^4 k \log d\right)$.

Remark 6.1. *Although Theorem 6.3 and Observation 6.1 imply that randomized rounding for A-optimal design with repetition takes $O\left((k+d^2)^2 d^4 k \log d\right)$ number of arithmetic operations, this does not take into account the size of numbers used in the computation which may scale with input ϵ. After taking this into account in the proofs in Lemma 6.2 and Lemma 6.3, the runtime* *of randomized rounding for A-optimal design with repetition becomes $O\left((k+d^2)^2 d^4 k^2 \log d \log(\frac{k+d^2}{\epsilon})\right)$. The proof is presented in the full version of the paper [31].*

6.4 Efficient Deterministic Implementation of $\frac{k}{k-d+1}$-Approximation Algorithm With Repetitions We show a *deterministic* implementation of proportional volume sampling used for the $\frac{k}{k-d+1}$-approximation algorithm with repetitions. In particular, we derandomized the efficient implementation of steps (4)-(5) of Algorithm 2, and show that the running time of deterministic version is the same as that of the randomized one.

THEOREM 6.4. *Given inputs $n, d, k, \epsilon, x \in \mathbb{R}^n_+$ with $\sum_{i=1}^n x_i = k$, and vectors $v_1, \dots, v_n$ to Algorithm 2, we define q, U, W as in Algorithm 2. Then, there exists a deterministic algorithm $\mathcal{A}'$ that outputs $S^* \subseteq U$ of size k such that*

$$\operatorname{tr}\left(W_{S^*}W_{S^*}^\top\right)^{-1} \geq \mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\operatorname{tr}\left(W_{\mathcal{S}}W_{\mathcal{S}}^\top\right)^{-1}\right]$$

where μ' is a distribution over all subsets $S \subseteq U$ of size k defined by $\mu'(S) \propto \det(W_S W_S^\top)$ for each set $S \subseteq U$ of size k. Moreover, $\mathcal{A}'$ runs in $O\left(n^2 d^4 k \log d\right)$ number of arithmetic operations.

Again, together with Observation 6.1 and Remark 6.1, Theorem 6.4 implies that the $\frac{k}{k-d+1}$-approximation algorithm for A-optimal design with repetitions can be implemented deterministically in $O\left((k+d^2)^2 d^4 k \log d\right)$ number of arithmetic operations and, after taking into account the size of numbers in the computation, in $O\left((k+d^2)^2 d^4 k^2 \log d \log(\frac{k+d^2}{\epsilon})\right)$ time.

6.5 Efficient Implementations for the Generalized Ratio Objective In Section 6.1-6.2 we obtain efficient randomized and deterministic implementations of proportional volume sampling with measure μ when μ is a hard-core distribution over all subsets $S \in \mathcal{U}$ (where $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\leq k}\}$) with any given parameter $\lambda \in \mathbb{R}^n_+$. Both implementations run in $O\left(n^4 dk^2 \log(dk)\right)$ number of arithmetic operations. In Section 6.3-6.4, we obtain efficient randomized and deterministic implementations of proportional volume sampling over exponentially-sized matrix $W = [w_{i,j}]$ of m vectors containing n *distinct* vectors in $O\left(n^2 d^4 k \log d\right)$ number of arithmetic operations. In this section, we show that the results from Section 6.1-6.4 generalize to proportional l-volume sampling for generalized ratio problem.

THEOREM 6.5. *Let n, d, k be positive integers, $\lambda \in \mathbb{R}^n_+$, $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\leq k}\}$, $V = [v_1, \dots, v_n] \in \mathbb{R}^{d\times n}$, and $0 \leq l' < l \leq d$ be a pair of integers. Let μ' be the l-proportional volume sampling distribution over $\mathcal{U}$ with hard-core measure μ of parameter λ, i.e. $\mu'(S) \propto \lambda^S E_l\left(V_S V_S^\top\right)$ for all $S \in \mathcal{U}$. There are*

- *an implementation to sample from μ' that runs in $O\left(n^4 lk^2 \log(lk)\right)$ number of arithmetic operations, and*
- *a deterministic algorithm that outputs a set $S^* \in \mathcal{U}$ of size k such that*
  (6.33)
  $$\left(\frac{E_{l'}(V_{S^*}V_{S^*}^\top)}{E_l(V_{S^*}V_{S^*}^\top)}\right)^{\frac{1}{l-l'}} \geq \mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\left(\frac{E_{l'}(V_{\mathcal{S}}V_{\mathcal{S}}^\top)}{E_l(V_{\mathcal{S}}V_{\mathcal{S}}^\top)}\right)^{\frac{1}{l-l'}}\right]$$
  that runs in $O\left(n^4 lk^2 \log(lk)\right)$ number of arithmetic operations.

Moreover, let $W = [w_{i,j}]$ be a matrix of m vectors where $w_{i,j} = v_i$ for all $i \in [n]$ and j. Denote U the index set of W. Let μ' be the l-proportional volume sampling over all subsets $S \subseteq U$ of size k with measure μ that is uniform, i.e. $\mu'(S) \propto E_l\left(W_S W_S^\top\right)$ for all $S \subseteq U, |S| = k$. There are

- *an implementation to sample from μ' that runs in $O\left(n^2(d-l+1)l^2 d^2 k \log d\right)$ number of arithmetic operations, and*
- *a deterministic algorithm that outputs a set $S^* \in \mathcal{U}$ of size k such that*
  (6.34)
  $$\left(\frac{E_{l'}(W_{S^*}W_{S^*}^\top)}{E_l(W_{S^*}W_{S^*}^\top)}\right)^{\frac{1}{l-l'}} \geq \mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\left(\frac{E_{l'}(W_{\mathcal{S}}W_{\mathcal{S}}^\top)}{E_l(W_{\mathcal{S}}W_{\mathcal{S}}^\top)}\right)^{\frac{1}{l-l'}}\right]$$
  that runs in
  $$O\left(n^2\left((d-l'+1)l'^2 + (d-l+1)l^2\right)d^2 k \log d\right)$$
  number of arithmetic operations.

As in Observation 6.1, note that we can replace $n = k + d^2$ in all running times in Theorem 6.5 so that running times of all variants of proportional volume sampling are independent of n. We also note, as in Remark 6.1, that running times of l-proportional volume sampling over m vectors with n distinct vectors has an extra factor of $k \log m$ after taking into account the size of numbers in computation, allowing us to do sampling over exponential-sized ground set $[m]$.

7 Integrality Gaps

7.1 Integrality Gap for E-Optimality Here we consider another objective for optimal design of experiments, the E-optimal design objective, and show that our results in the asymptotic regime do not extend to it. Once again, the input is a set of vectors $v_1, \dots, v_n \in \mathbb{R}^d$, and our goal is to select a set $S \subseteq [n]$ of size k, but this time we minimize the objective $\left\|(\sum_{i\in S} v_i v_i^\top)^{-1}\right\|$, where $\|\cdot\|$ is the operator norm, i.e. the largest singular value. By taking the inverse of the objective, this is equivalent to maximizing $\lambda_1(\sum_{i\in S} v_i v_i^\top)$, where $\lambda_i(M)$ denotes the ith smallest eigenvalue of M. This problem also has a natural convex relaxation, analogous to the one we use for the A objective:

$$\max \lambda_1\left(\sum_{i=1}^n x_i v_i v_i^\top\right) \tag{7.35}$$

s.t.

$$\sum_{i=1}^n x_i = k \tag{7.36}$$

$$0 \le x_i \le 1 \quad \forall i \in [n] \tag{7.37}$$

We prove the following integrality gap result for (7.35)–(7.37).

Theorem 7.1. *There exists a constant $c > 0$ such that the following holds. For any small enough $\epsilon > 0$, and all integers $d \ge d_0(\epsilon)$, if $k < \frac{cd}{\epsilon^2}$, then there exists an instance $v_1, \dots v_n \in \mathbb{R}^d$ of the E-optimal design problem, for which the value* CP *of (7.35)–(7.37) satisfies*

$$\mathsf{CP} > (1+\epsilon)\mathsf{OPT} = (1+\epsilon) \max_{S\subseteq[n]:|S|=k} \lambda_1\left(\sum_{i\in S} v_i v_i^\top\right)$$

Recall that for the A-objective we achieve a $(1+\epsilon)$-approximation for $k = \Omega(\frac{d}{\epsilon} + \frac{\log(1/\epsilon)}{\epsilon^2})$. Theorem 7.1 shows that such a result is impossible for the E-objective, for which the results in [1] cannot be improved.

Our integrality gap instance comes from a natural connection to spectral graph theory. Let us first describe the instance for any given d. We first define $n = \binom{d+1}{2}$ vectors in $\mathbb{R}^{d+1}$, one for each unordered pair $(i,j) \in \binom{[d+1]}{2}$. The vector corresponding to (i,j), $i<j$, is u_{ij} and has value 1 in the i-th coordinate, -1 in the j-th coordinate, and 0 everywhere else. In other words, the u_{ij} vectors are the columns of the vertex by edge incidence matrix U of the complete graph K_{d+1}, and $UU^\top = (d+1)I_{d+1} - J_{d+1}$ is the (unnormalized) Laplacian of K_{d+1}. (We use I_m for the $m\times m$ identity matrix, and J_m for the $m\times m$ all-ones matrix.) All the u_{ij} are orthogonal to the all-ones vector 1; we define our instance by writing u_{ij} in an orthonormal basis of this subspace: pick any orthonormal basis $b_1, \dots, b_d$ of the subspace of $\mathbb{R}^{d+1}$ orthogonal to 1, and define $v_{ij} = B^\top u_{ij}$ for $B = (b_i)_{i=1}^d$. Thus

$$M = \sum_{i=1}^{d+1}\sum_{j=i+1}^{d+1} v_{ij}v_{ij}^\top = (d+1)I_d.$$

We consider the fractional solution $x = \frac{k}{\binom{d+1}{2}}1$, i.e. each coordinate of x is $k/\binom{d+1}{2}$. Then $M(x) = \sum_{i=1}^{d+1}\sum_{j=i+1}^{d+1} x_{ij}v_{ij}v_{ij}^\top = \frac{2k}{d}I_d$, and the objective value of the solution is $\frac{2k}{d}$.

Consider now any integral solution $S \subseteq \binom{[d+1]}{2}$ of the E-optimal design problem. We can treat S as the edges of a graph $G = ([d+1], S)$, and the Laplacian L_G of this graph is $L_G = \sum_{(i,j)\in S} u_{ij}u_{ij}^\top$. If the objective value of S is at most $(1+\epsilon)\mathsf{CP}$, then the smallest eigenvalue of $M(S) = \sum_{(i,j)\in S} v_{ij}v_{ij}^\top$ is at least $\frac{2k}{d(1+\epsilon)} \ge (1-\epsilon)\frac{2k}{d}$. Since $M(S) = B^\top L_G B$, this means that the second smallest eigenvalue of L_G is at least $(1-\epsilon)\frac{2k}{d}$. The average degree Δ of G is $\frac{2k}{d+1}$. So, we have a graph G on $d+1$ vertices with average degree Δ for which the second smallest eigenvalue of its Laplacian is at least $(1-\epsilon)(1-\frac{1}{d+1})\Delta \ge (1-2\epsilon)\Delta$, where the inequality holds for d large enough. The classical Alon-Boppana bound ([3, 34]) shows that, up to lower order terms, the second smallest eigenvalue of the Laplacian of a Δ-regular graph is at most $\Delta - 2\sqrt{\Delta}$. If our graph G were regular, this would imply that $\frac{2k}{d+1} = \Delta \ge \frac{1}{\epsilon^2}$. In order to prove Theorem 7.1, we extend the Alon-Boppana bound to not necessarily regular graphs, but with worse constants. There is an extensive body of work on extending the Alon-Boppana bound to non-regular graphs: see the recent preprint [39] for an overview of prior work on this subject. However, most of the work focuses either on the normalized Laplacian or the adjacency matrix of G, and we were unable to find the statement below in the literature.

Theorem 7.2. *Let $G = (V,E)$ be a graph with average degree $\Delta = \frac{2|E|}{|V|}$, and let L_G be its unnormalized Laplacian matrix. Then, as long as Δ is large enough, and $|V|$ is large enough with respect to Δ,*

$$\lambda_2(L_G) \le \Delta - c\sqrt{\Delta},$$

where $\lambda_2(L_G)$ is the second smallest eigenvalue of L_G, and $c > 0$ is an absolute constant.

The proof is presented in the full version of the paper [31].

To finish the proof of Theorem 7.1, recall that the existence of a $(1+\epsilon)$-approximate solution S to our instance implies that, for all large enough d, the graph

$G = ([d+1], S)$ with average degree $\Delta = \frac{2k}{d+1}$ satisfies $\lambda_2(L_G) \geq (1-2\epsilon)\Delta$. By Theorem 7.2, $\lambda_2(L_G) \leq \Delta - c\sqrt{\Delta}$ for large enough d with respect to Δ. We have $\Delta \geq \frac{c^2}{4\epsilon^2}$, and re-arranging the terms proves the theorem.

Note that the proof of Theorem 7.2 does not require the graph G to be simple, i.e. parallel edges are allowed. This means that the integrality gap in Theorem 7.1 holds for the E-optimal design problem with repetitions as well.

7.2 Integrality Gap for A-optimality

THEOREM 7.3. *For any given positive integers k, d, there exists an instance $V = [v_1, \ldots, v_n] \in \mathbb{R}^{d\times n}$ to the A-optimal design problem such that*

$$\mathsf{OPT} \geq \left(\frac{k}{k-d+1} - \delta\right) \cdot \mathsf{CP}$$

for all $\delta > 0$, where OPT denotes the value of the optimal integral solution and CP denotes the value of the convex program.

This implies that the gap is at least $\frac{k}{k-d+1}$. The theorem statement applies to both with and without repetitions.

The proof is presented in the full version of the paper [31].

8 Hardness of Approximation

In this section we show that the A-optimal design problem is NP-hard to approximate within a fixed constant when $k = d$. To the best of our knowledge, no hardness results for this problem were previously known. Our reduction is inspired by the hardness of approximation for D-optimal design proved in [41]. The hard problem we reduce from is an approximation version of Partition into Triangles.

Before we prove our main hardness result, Theorem 1.4, we describe the class of instances we consider, and prove some basic properties. Given a graph $G = ([d], E)$, we define a vector v_e for each edge $e = (i,j)$ so that its i-th and j-th coordinates are equal to 1, and all its other coordinates are equal to 0. Then the matrix $V = (v_e)_{e\in E}$ is the undirected vertex by edge incidence matrix of G. The main technical lemma needed for our reduction follows.

LEMMA 8.1. *Let V be the vertex by edge incidence matrix of a graph $G = ([d], E)$, as described above. Let $S \subseteq E$ be a set of d edges of G so that the submatrix V_S is invertible. Then each connected component of the subgraph $H = ([d], S)$ is the disjoint union of a spanning tree and an edge. Moreover, if t of the connected components of H are triangles, then*

- *for $t = \frac{d}{3}$, $\operatorname{tr}((V_S V_S^\top)^{-1}) = \frac{3d}{4}$;*
- *for any t, $\operatorname{tr}((V_S V_S^\top)^{-1}) \geq d - \frac{3t}{4}$.*

The proof is presented in the full version of the paper [31].

Recall that in the Partition into Triangles problem we are given a graph $G = (W, E)$, and need to decide if W can be partitioned into $\frac{|W|}{3}$ vertex-disjoint triangles. This problem is NP-complete ([19] present a proof in Chapter 3 and cite personal communication with Schaeffer), and this, together with Lemma 8.1, suffice to show that the A-optimal design problem is NP-hard when $k = d$. To prove hardness of approximation, we prove hardness of a gap version of Partition into Triangles. In fact, we just observe that the reduction from 3-Dimensional Matching to Partition into Triangles in [19] and known hardness of approximation of 3-Dimensional Matching give the result we need.

LEMMA 8.2. *Given a graph $G = (W, E)$, it is NP-hard to distinguish the two cases:*

1. *W can be partitioned into $\frac{|W|}{3}$ vertex-disjoint triangles;*
2. *every set of vertex-disjoint triangles in G has cardinality at most $\alpha\frac{|W|}{3}$,*

where $\alpha \in (0,1)$ is an absolute constant.

The proof is presented in the full version of the paper [31].

We now have everything in place to finish the proof of our main hardness result.

Proof of Theorem 1.4: We use a reduction from (the gap version of) Partition into Triangles to the A-optimal design problem. In fact the reduction was already described in the beginning of the section: given a graph $G = ([d], E)$, it outputs the columns v_e of the vertex by edge incidence matrix V of G.

Consider the case in which the vertices of G can be partitioned into vertex-disjoint triangles. Let S be the union of the edges of the triangles. Then, by Lemma 8.1, $\operatorname{tr}((V_S V_S^\top)^{-1}) = \frac{3d}{4}$.

Next, consider the case in which every set of vertex-disjoint triangles in G has cardinality at most $\alpha\frac{d}{3}$. Let S be any set of d edges in E such that V_S is invertible. The subgraph $H = ([d], S)$ of G can have at most $\alpha\frac{d}{3}$ connected components that are triangles, because any two triangles in distinct connected components are necessarily vertex-disjoint. Therefore, by Lemma 8.1, $\operatorname{tr}((V_S V_S^\top)^{-1}) \geq \frac{(4-\alpha)d}{4}$.

It follows that a c-approximation algorithm for the A-optimal design problem, for any $c < \frac{4-\alpha}{3}$, can be used to distinguish between the two cases of Lemma 8.2, and, therefore, the A-optimal design problem is NP-hard to c-approximate. $\square$

Acknowledgement The authors thank Bobby Kleinberg for a useful discussion.

References

[1] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal design of experiments via regret minimization. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 126–135. PMLR, 2017.

[2] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: A regret minimization approach. *arXiv preprint arXiv:1711.05174*, 2017.

[3] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of computing (Singer Island, Fla., 1984).

[4] Nima Anari and Shayan Oveis Gharan. A generalization of permanent inequalities and applications in counting and optimization. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 384–396. ACM, 2017.

[5] Haim Avron and Christos Boutsidis. Faster subset selection for matrices and applications. *SIAM Journal on Matrix Analysis and Applications*, 34(4):1464–1499, 2013.

[6] Joshua Batson, Daniel A Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721, 2012.

[7] J. Bourgain and L. Tzafriri. Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987.

[8] Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail. Sparse features for pca-like linear regression. In John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain.*, pages 2285–2293, 2011.

[9] Christos Boutsidis and Malik Magdon-Ismail. Deterministic feature selection for k-means clustering. *IEEE Transactions on Information Theory*, 59(9):6099–6110, 2013.

[10] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[11] Herman 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.

[12] Ali Çivril and Malik Magdon-Ismail. On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Computer Science*, 410(47-49):4801–4811, 2009.

[13] FR de Hoog and RMM Mattheij. Subset selection for matrices. *Linear Algebra and its Applications*, 422(2-3):349–359, 2007.

[14] FR de Hoog and RMM Mattheij. A note on subset selection for matrices. *Linear Algebra and its Applications*, 434(8):1845–1850, 2011.

[15] Michał Dereziński and Manfred K Warmuth. Subsampling for ridge regression via regularized volume sampling. *arXiv preprint arXiv:1710.05110*, 2017.

[16] Michał Dereziński and Manfred K Warmuth. Unbiased estimates for linear regression via volume sampling. In *Advances in Neural Information Processing Systems*, pages 3087–3096, 2017.

[17] Amit Deshpande and Luis Rademacher. Efficient volume sampling for row/column subset selection. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 329–338. IEEE Computer Society, 2010.

[18] Valerii Vadimovich Fedorov. *Theory of optimal experiments.* Elsevier, 1972.

[19] Michael R. Garey and David S. Johnson. *Computers and intractability.* W. H. Freeman and Co., San Francisco, Calif., 1979. A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences.

[20] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 550–559. IEEE, 2011.

[21] Venkatesan Guruswami and Ali Kemal Sinop. Optimal column-based low-rank matrix reconstruction. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 1207–1214. SIAM, 2012.

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

[23] Jeff Kahn. Asymptotics of the chromatic index for multigraphs. *journal of combinatorial theory, Series B*, 68(2):233–254, 1996.

[24] Jeff Kahn. Asymptotics of the list-chromatic index for multigraphs. *Random Structures & Algorithms*, 17(2):117–156, 2000.

[25] Alex Kulesza, Ben Taskar, et al. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012.

[26] Chengtao Li, Stefanie Jegelka, and Suvrit Sra. Column subset selection via polynomial time dual volume

sampling. *arXiv preprint arXiv:1703.02674*, 2017.

[27] Zelda E Mariet and Suvrit Sra. Elementary symmetric polynomials for optimal experimental design. In *Advances in Neural Information Processing Systems*, pages 2136–2145, 2017.

[28] Assaf Naor and Pierre Youssef. *Restricted Invertibility Revisited*, pages 657–691. Springer International Publishing, Cham, 2017.

[29] Aleksandar 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.

[30] Aleksandar Nikolov and Mohit Singh. Maximizing determinants under partition constraints. In *STOC*, pages 192–201, 2016.

[31] Aleksandar Nikolov, Mohit Singh, and Uthaipon Tao Tantipongpipat. Proportional volume sampling and approximation algorithms for a-optimal design. *arXiv preprint arXiv:1802.08318*, 2018.

[32] Aleksandar Nikolov and Kunal Talwar. Approximating hereditary discrepancy via small width ellipsoids. In Piotr Indyk, editor, *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 324–336. SIAM, 2015.

[33] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: The small database and approximate cases. *SIAM J. Comput.*, 45(2):575–616, 2016.

[34] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.

[35] Friedrich Pukelsheim. *Optimal design of experiments.* SIAM, 2006.

[36] Mohit Singh and Weijun Xie. Approximate positive correlated distributions and approximation algorithms for d-optimal design. *To appear in SODA*, 2018.

[37] D.A. Spielman and N. Srivastava. An elementary proof of the restricted invertibility theorem. *Israel Journal of Mathematics*, pages 1–9, 2010.

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

[39] Nikhil Srivastava and Luca Trevisan. An alon-boppana type bound for weighted graphs and lowerbounds for spectral sparsification. *CoRR*, abs/1707.06364, 2017.

[40] Damian Straszak and Nisheeth K. Vishnoi. Real stable polynomials and matroids: optimization and counting. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 370–383. ACM, 2017.

[41] Marco Di Summa, Friedrich Eisenbrand, Yuri Faenza, and Carsten Moldenhauer. On largest volume simplices and sub-determinants. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 315–323. Society for Industrial and Applied Mathematics, 2015.

[42] R. Vershynin. John's decompositions: Selecting a large part. *Israel Journal of Mathematics*, 122(1):253–277, 2001.

[43] Yining Wang, Adams Wei Yu, and Aarti Singh. On computationally tractable selection of experiments in regression models. *arXiv preprint arXiv:1601.02068*, 2016.

[44] Pierre Youssef. Restricted invertibility and the Banach-Mazur distance to the cube. *Mathematika*, 60(1):201–218, 2014.