
**Mathematics of Operations Research**

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

# Proportional Volume Sampling and Approximation Algorithms for A-Optimal Design

Aleksandar Nikolov, Mohit Singh, Uthaipon (Tao) Tantipongpipat

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 47, No. 2, May 2022, pp. 847–877
ISSN 0364-765X (print), ISSN 1526-5471 (online)
http://pubsonline.informs.org/journal/moor



# Proportional Volume Sampling and Approximation Algorithms for *A*-Optimal Design

**Aleksandar Nikolov,[a] Mohit Singh,[b] Uthaipon (Tao) Tantipongpipat[b,*]**

[a] Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada; [b] H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332
*Corresponding author
**Contact:** anikolov@cs.toronto.edu (AN); mohitsinghr@gmail.com (MS); uthaipon@gmail.com, https://orcid.org/0000-0002-5573-5165 (U(T)T)

**Received:** April 11, 2019
**Revised:** June 12, 2020
**Accepted:** October 11, 2020
**Published Online in Articles in Advance:** February 15, 2022

**MSC2020 Subject Classification:** Primary: 90C27; secondary: 62J05, 62K05; 62K15; 68W20; 68W25

**https://doi.org/10.1287/moor.2021.1129**



**Abstract.** We study optimal design problems in which the goal is to choose a set of linear measurements to obtain the most accurate estimate of an unknown vector. We study the $A$-optimal design variant where the objective is to minimize the average variance of the error in the maximum likelihood estimate of the vector being measured. We introduce the *proportional volume sampling* algorithm to obtain nearly optimal bounds in the asymptotic regime when the number $k$ of measurements made is significantly larger than the dimension $d$ and obtain the first approximation algorithms whose approximation factor does not degrade with the number of possible measurements when $k$ is small. The algorithm also gives approximation guarantees for other optimal design objectives such as $D$-optimality and the generalized ratio objective, matching or improving the previously best-known results. We further show that bounds similar to ours cannot be obtained for $E$-optimal design and that $A$-optimal design is NP-hard to approximate within a fixed constant when $k = d$.

**Funding:** A. Nikolov was supported by Natural Sciences and Engineering Research Council of Canada 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 the National Science Foundation [Grant CCF-1740425]. M. Singh and U. (T.) Tantipongpipat are supported in part by the National Science Foundation [Grant CCF-24067E5]. U. (T.) Tantipongpipat is supported in part by the National Science Foundation [Grant CCF-1740776 (Georgia Institute of Technology TRIAD)] and a Georgia Institute of Technology Algorithm and Randomness Center fellowship.

**Keywords:** optimal design • design of experiments • A-optimal design • D-optimal design • generalized ratio objective • linear regression • volume sampling • proportional volume sampling • approximation algorithms • NP-hardness

## 1. Introduction

Given a collection of vectors, a common problem is to select a subset of size $k \le n$ 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 (Fedorov [23], Pukelsheim [43]) in statistics; feature selection (Boutsidis and Magdon-Ismail [8]) and sensor placement problems (Joshi and Boyd [27]) in machine learning; matrix sparsification (Batson et al. [6], Spielman and Srivastava [45]); and column subset selection (Avron and Boutsidis [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 (Pukelsheim [43]) with recent applications in machine learning (Joshi and Boyd [27], Wang et al. [50]). Here the goal is to estimate an unknown vector $\boldsymbol{w} \in \mathbb{R}^d$ via linear measurements of the form $y_i = \boldsymbol{v}_i^\top \boldsymbol{w} + \eta_i$, where $\boldsymbol{v}_i$ are possible experiments and $\eta_i$ is assumed to be small independently and identically distributed unbiased Gaussian error introduced in the measurement. Given a set $S$ of linear measurements, the maximum likelihood estimate $\hat{\boldsymbol{w}}$ of $\boldsymbol{w}$ can be obtained via a least-squares computation. The error vector $\boldsymbol{w} - \hat{\boldsymbol{w}}$ has a Gaussian distribution with mean zero and covariance matrix $\left(\sum_{i\in S} \boldsymbol{v}_i \boldsymbol{v}_i^\top\right)^{-1}$. 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 $\boldsymbol{w} - \hat{\boldsymbol{w}}$.

We let $V$ denote the $d \times n$ matrix whose columns are the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 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 $\left(V_S V_S^\top\right)^{-1} = \left(\sum_{i\in S} \boldsymbol{v}_i \boldsymbol{v}_i^\top\right)^{-1}$.

We also consider the $A$-optimal design problem with repetitions, where the chosen $S$ can be a multiset, thus allowing a vector to chosen more than once.

Apart from experimental design, this formulation finds application in other areas such as sensor placement in wireless networks (Joshi and Boyd [27]), sparse least-squares regression (Boutsidis et al. [9]), feature selection for $k$-means clustering (Boutsidis and Magdon-Ismail [8]), and matrix approximation (Avron and Boutsidis [5]). For example, in matrix approximation (Avron and Boutsidis [5], de Hoog and Mattheij [15, 16]) 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 $d$ and in the asymptotic regime where $k \geq d$. The improvement is summarized in Table 1. 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 $\mathcal{U}_k$ and $\mathcal{U}_{\leq k}$ and state the following definition more generally.

**Definition 1.** Let $\mu$ be probability measure on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$). Then the proportional volume sampling with measure $\mu$ 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 $\mu$ is the uniform distribution and $k \leq d$ then we obtain the standard volume sampling (Deshpande et al. [21]) 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 (Deshpande and Rademacher [19], Deshpande and Vempala [20], Guruswami and Sinop [26]). More recently, efficient algorithms were obtained even when $k \geq d$ (Li et al. [31], Singh and Xie [44]). We discuss the computational issues of sampling from proportional volume sampling in Lemma 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 2.** Given integers $d \leq k \leq n$ and a vector $\boldsymbol{x} \in [0,1]^n$ such that $1^\top \boldsymbol{x} = k$, we call a measure $\mu$ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$), $\alpha$-approximate $(d-1, d)$-wise independent with respect to $\boldsymbol{x}$ if for any subsets $T, R \subseteq [n]$ with $|T| = d-1$ and $|R| = d$, we have

$$\frac{\Pr_{\mathcal{S}\sim\mu}[T \subseteq \mathcal{S}]}{\Pr_{\mathcal{S}\sim\mu}[R \subseteq \mathcal{S}]} \leq \alpha \frac{x^T}{x^R},$$

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 $\mu$ corresponds to picking each element $i$ independently with probability $x_i$, then

$$\frac{\Pr_{\mathcal{S}\sim\mu}[T \subseteq \mathcal{S}]}{\Pr_{\mathcal{S}\sim\mu}[R \subseteq \mathcal{S}]} = \frac{x^T}{x^R}$$

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 previous definition.

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 Avron and Boutsidis [5] to proportional volume sampling, which allows us to exploit the power of the convex relaxation and get a significantly improved approximation.

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

**Table 1.** Summary of approximation ratios of $A$-optimal results. We list the best applicable previous work for comparison.

| Problem | Our result | Previous work |
|---|---|---|
| Case $k = d$ | $d$[a] | $n - d + 1$ (Avron and Boutsidis [5]) |
| Asymptotic $k \gg d$ without repetitions | $1 + \epsilon$, for $k \geq \Omega\left(d/\epsilon + \log(1/\epsilon)/\epsilon^2\right)$ | $1 + \epsilon$, for $k \geq \Omega(d/\epsilon^2)$ (Allen-Zhu et al. [2]) |
| Arbitrary $k$ and $d$ with repetitions | $k/k - d + 1$[a] | $n - d + 1$ (Avron and Boutsidis [5]) |
| Asymptotic $k \gg d$ with repetitions | $1 + \epsilon$, for $k \geq d + (d/\epsilon)$[a] | $1 + \epsilon$, for $k \geq \Omega(d/\epsilon^2)$ (Allen-Zhu et al. [2]) |

[a]Ratios are tight with matching integrality gap of the convex relaxation (1)–(3).

In this theorem, we in fact only need an approximately independent distribution $\mu$ for the optimal solution $\boldsymbol{x}$ of the natural convex relaxation for the problem, which is given in (1)–(3). The result also bounds the integrality gap of the convex relaxation by $\alpha$. Theorem 1 is proved in Section 2.

Theorem 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 $\mu$ 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 $\mu$ (Boyd and Vandenberghe [10]). Observe that, although $\mu$ places nonzero probability on exponentially many sets, it is enough to specify $\mu$ succinctly by describing $\lambda$. Hard-core distributions over various structures including spanning trees (Gharan et al. [25]) or matchings (Kahn [28, 29]) in a graph display *approximate independence*, and this has found use in combinatorics and 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 $k \gg d$.

**Theorem 2.** *Given integers $d \leq k \leq n$ and a vector $x \in [0, 1]^n$ such that $1^\top \boldsymbol{x} = k$, there exists a hard-core distribution $\mu$ on sets in $\mathcal{U}_k$ that is d-approximate $(d-1, d)$-wise independent when $k = d$. Moreover, for any $\epsilon > 0$, if $k = \Omega(d/\epsilon + (1/\epsilon^2)\log(1/\epsilon))$, then there is a hard-core distribution $\mu$ on $\mathcal{U}_{\leq k}$ 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 $k = \Omega(d/\epsilon + (1/\epsilon^2)\log(1/\epsilon))$.*

This theorem relies on two natural hard-core distributions. In the first one, we consider the hard-core distribution with parameter $\lambda = \boldsymbol{x}$ on sets in $\mathcal{U}_k$ and in the second we consider the hard-core distribution with parameter $\lambda = (1-\epsilon)\boldsymbol{x}/1 - (1-\epsilon)\boldsymbol{x}$ (defined coordinate-wise) on sets in $\mathcal{U}_{\leq k}$. 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 later. 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 3.** *For all $k \geq d$ and $0 < \epsilon \leq 1$, there is a $((k/(k-d+1)) + \epsilon)$-approximation algorithm for the A-optimal design problem with repetitions. In particular, there is a $(1+\epsilon)$-approximation when $k \geq d + d/\epsilon$.*

We remark that the integrality gap of the natural convex relaxation is at least $k/(k-d+1)$ (see Section 7.2), and thus the previous theorem results in an exact characterization of the integrality gap of the convex program (1)–(3), stated in the following corollary. The proof of Theorem 3 appears in Section 6.3.

**Corollary 1.** *For any integers $k \geq d$, the integrality gap of the convex program (1)–(3) for the A-optimal design with repetitions is exactly $k/(k-d+1)$.*

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

**Theorem 4.** *There exists a constant $c > 1$ such that the A-optimal design problem is* NP-*hard to c-approximate when $k = d$.*

The $k \leq d$ 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 $\Sigma_{i=1}^k \lambda_i^{-1}$, where $\lambda_1, \ldots, \lambda_k$ are the $k$ largest eigenvalues of the matrix $V_S V_S^\top$. Although 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 that are as *spread out* as possible. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are the points in a data set, we can see an optimal solution as a maximally diverse representative sample of the data set. 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 Nikolov [36] for references to prior work). Çivril and Magdon-Ismail [13] also studied an analogous problem with the sum in the objective replaced by a maximum (which extends $E$-optimal design).

Although our rounding extends easily to the $k \leq d$ 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 5.** *There exists a* poly$(d, n)$*-time k-approximation algorithm for the A-optimal design problem when $k \leq d$.*

**1.1.1. General Objectives.** Experimental design problems come with many different objectives including $A$, $D$, $E$, $G$, $T$, and $V$, each corresponding to a different function of the covariance matrix of the error $\boldsymbol{w} - \hat{\boldsymbol{w}}$. Any algorithm that solves $A$-design can solve $V$-optimal design 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, Singh and Xie [44]; and generalized ratio objective, Mariet and Sra [34]) matching or improving previous best-known results. We refer the reader to Section 5.3 for details.

**1.1.2. Integrality Gap and *E*-Optimality.** Given the results mentioned previously, a natural question is whether all objectives for optimal design behave similarly in terms of approximation algorithms. Indeed, recent results of Allen-Zhu et al. [1, 2] and Wang et al. [50] give the $(1+\epsilon)$-approximation algorithm in the asymptotic regime, $k \geq \Omega(d/\epsilon^2)$ and $k \geq \Omega(d^2/\epsilon)$, 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 $k = \Omega(d/\epsilon^2)$, 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 (Alon [3], Nilli [40]). We prove an Alon-Boppana style bound for the unnormalized Laplacian of not necessarily regular graphs with a given average degree.

**1.1.3. Restricted Invertibility Principle for Harmonic Mean.** As an application of Theorem 5, we prove a restricted invertibility principle (RIP) (Bourgain and Tzafriri [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 $\|\cdot\|_{HS} = \sqrt{\operatorname{tr}(VV^\top)}$ is the Hilbert-Schmidt, or Frobenius, norm of $V$, and $\|\cdot\|$ is the operator norm. The classical restricted invertibility principle (Bourgain and Tzafriri [7], Spielman and Srivastava [46], Vershynin [49]) 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 $\Omega(\|V\|_{HS}/\sqrt{m})$. Nikolov [36] showed there exists a submatrix $V_S$ of $k$ columns of $V$ so that the geometric mean of 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 $r$. Although 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 $\Omega(\epsilon\|V\|_{HS}^2/m)$, whereas the tight restricted invertibility principle of Spielman and Srivastava [45] would only give $\epsilon^2$ in the place of $\epsilon$. This restricted invertibility principle can also be derived from the results of Naor et al. [35], 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.

**1.1.4. Computational Issues.** Although it is not clear whether sampling from proportional volume sampling is possible under general assumptions (e.g., given a sampling oracle for $\mu$), we obtain an efficient sampling algorithm when $\mu$ is a hard-core distribution.

**Lemma 1.** *There exists a* $\operatorname{poly}(d,n)$*-time algorithm that, given a matrix $d \times n$ matrix $V$, integer $k \leq n$, and a hard-core distribution $\mu$ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\leq k}$) with parameter $\lambda$, efficiently samples a set from the proportional volume measure defined by $\mu$.*

When $k \leq d$ and $\mu$ 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 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 Theorems 2 and 3 in Sections 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 $(\sum_{i\in S} v_i v_i^\top)^{-1}$. 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 Purkelsheim [43] for a broad survey on experimental design.

Avron and Boutsidis [5] studied the $A$- and $E$-optimal design problems and analyzed various combinatorial algorithms and algorithms based on volume sampling, and achieved approximation ratio $(n-d+1)/(k-d+1)$. Wang et al. [50] found connections between optimal design and matrix sparsification and used these connections to obtain a $(1+\epsilon)$-approximation when $k \geq d^2/\epsilon$, and also approximation algorithms under certain technical assumptions. More recently, Allen-Zhu et al. [1, 2] obtained a $(1+\epsilon)$-approximation when $k = \Omega(d/\epsilon^2)$ 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 $\Sigma_{i\in S} v_i v_i^\top$, respectively. More generally, their result applies to any objective function that satisfies certain regularity criteria.

Improved bounds for $D$-optimality were obtained by Singh and Xie [44], who give an $e$-approximation for all $k$ and $d$, and $(1+\epsilon)$-approximation algorithm when $k = \Omega(d/\epsilon + (1/\epsilon^2)\log(1/\epsilon))$, with a weaker condition of $k \geq (2d)/\epsilon$ if repetitions are allowed. The $D$-optimality criterion when $k \leq d$ has also been extensively studied. It captures maximum a posteriori inference in constrained determinantal point process models (Kulesza et al. [30]) and the maximum volume simplex problem. Nikolov [36], 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 (Anari and Gharan [5], Nikolov and Singh [37], Straszak and Vishnoi [48]).

Çivril and Magdon-Ismail [13] also studied several related problems in the $k \leq d$ 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 $\Sigma_{i\in S} v_i v_i^\top$, is closely related to the problem of matrix sparsification (Batson et al. [6], Spielman and Srivastava [45]) 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 Bourgain and Tzafriri [7] and was later strengthened by Vershynin [49], Spielman and Srivastava [46], and Naor and Youssef [35]. Spielman and Srivastava [46] gave a deterministic algorithm to find the well-invertible submatrix whose existence is guaranteed by the theorem. Besides its numerous applications in geometry (Vershynin [49], Youssef [51]), the principle has also found applications to differential privacy (Nikolov et al. [39]) and to approximation algorithms for discrepancy (Nikolov and Talwar [38]).

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 Deshpande and Rademacher [19] and improved by Guruswami and Sinop [26]. The probability distribution is also called a *determinantal point process* (DPP) and finds many applications in machine learning (Kulesza et al. [30]). Recently, fast algorithms for volume sampling have been considered in Dereziński and Warmuth [17, 18].

Although NP-hardness is known for the $D$- and $E$-optimality criteria (Çivril and Magdon-Ismail [13]), 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 Avron and Boutsidis [5].

## 2. Approximation via Near Independent Distributions

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

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 Chernoff [12]. It is easy to see that the objective $\mathrm{tr}\left(\Sigma_{i=1}^n x_i v_i v_i^\top\right)^{-1}$ is convex (Boyd and Vandenberghe [10], section 7.5). For this section, we focus on the case when repetitions are no allowed.

| With repetitions | Without repetitions | |
|---|---|---|
| $\min \mathrm{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$ | $\min \mathrm{tr}\left(\sum_{i=1}^n x_i v_i v_i^\top\right)^{-1}$ | (1) |
| s.t. $\sum_{i=1}^n x_i = k$ | s.t. $\sum_{i=1}^n x_i = k$ | (2) |
| $0 \leq x_i \quad \forall i \in [n]$ | $0 \leq x_i \leq 1 \quad \forall i \in [n]$ | (3) |

Let us denote the optimal value of (1)–(3) by CP (convex program). By plugging in the indicator vector of an optimal integral solution for $x$, we see that $\mathsf{CP} \leq \mathsf{OPT}$, where OPT (optimal) 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 $x_1, \ldots, x_n$, that is, $e_k(x_1, \ldots, x_n) = \Sigma_{S\in\mathcal{U}_k} x^S$. 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 $E_1(M)=\mathrm{tr}(M)$ and $E_n(M)=\det(M)$.

To prove Theorem 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 $\mu$ 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 $\Sigma_{i=1}^n x_i=k$.
3: Sample set $\mathcal{S}$ (from $\mathcal{U}_{\le k}$ or $\mathcal{U}_k$) where $\Pr[\mathcal{S}=S]\propto\mu(S)\det(V_SV_S^\top)$ for any $S\in\mathcal{U}_k$ (or $\mathcal{U}_{\le k}$). ▷ $\mu(S)$ 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 that is needed for proving Theorem 1.

**Lemma 2.** *Let $T\subseteq[n]$ be of size no more than $d$. Then,*

$$\det\left(V_T(x)^\top V_T(\boldsymbol{x})\right)=x^T\det\left(V_T^\top V_T\right).$$

**Proof.** The statement is true by multilinearity of the determinant and the exact formula for $V_T(x)^\top V_T(x)$ as follows. The matrix $V_T(x)^\top V_T(x)$ has $(i,j)$ entry

$$\left(V_T(x)^\top V_T(x)\right)_{i,j}=\sqrt{x_i}\boldsymbol{v}_i\cdot\sqrt{x_j}v_j=\sqrt{x_ix_j}\boldsymbol{v}_i\cdot v_j$$

for each pair $i,j\in[|T|]$. By the multilinearity of the determinant, we can take the factor $\sqrt{x_i}$ out from each row $i$ of $V_T(\boldsymbol{x})^\top V_T(\boldsymbol{x})$ and the factor $\sqrt{x_j}$ out from each column $j$ of $V_T(x)^\top V_T(x)$. This gives

$$\det\left(V_T(\boldsymbol{x})^\top V_T(\boldsymbol{x})\right)=\prod_{i\in[|T|]}\sqrt{x_i}\prod_{j\in[|T|]}\sqrt{x_j}\det\left(V_T^\top V_T\right)=x^T\det\left(V_T^\top V_T\right).\quad\square$$

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

$$E_k\left(VV^\top\right)=E_k\left(V^\top V\right)=\sum_{S\in\mathcal{U}_k}\det\left(V_S^\top V_S\right).\tag{4}$$

Identity (4) appeared in Mariet and Sra [34] and, specifically for $k=d-1$, as lemma 3.8 in Avron and Boutsidis [5]. Now we are ready to prove Theorem 1.

**Proof of Theorem 1.** Let $\mu'$ 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 $\mu(S)\det(V_SV_S^\top)$. Because $\mathrm{tr}\left(\Sigma_{i\in[n]}x_iv_iv_i^\top\right)^{-1}=\mathsf{CP}\le\mathsf{OPT}$, it is enough to show that

$$\mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\mathrm{tr}\left(\sum_{i\in\mathcal{S}}\boldsymbol{v}_iv_i^\top\right)^{-1}\right]\le\alpha\,\mathrm{tr}\left(\sum_{i\in[n]}x_i\boldsymbol{v}_iv_i^\top\right)^{-1}.\tag{5}$$

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 (Avron and Boutsidis [5]): for any $d\times d$ matrix $M$ of rank $d$, we have

$$\mathrm{tr}\,M^{-1}=\sum_{i=1}^d\frac{1}{\lambda_i(M)}=\frac{e_{d-1}(\lambda(M))}{e_d(\lambda(M))}=\frac{E_{d-1}(M)}{\det M}.\tag{6}$$

Therefore, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\mathrm{tr}\left(\sum_{i\in\mathcal{S}}\boldsymbol{v}_iv_i^\top\right)^{-1}\right]&=\sum_{S\in\mathcal{U}}\Pr_{\mu'}[\mathcal{S}=S]\,\mathrm{tr}\left(V_SV_S^\top\right)^{-1}\\&=\sum_{S\in\mathcal{U}}\frac{\mu(S)\det\left(V_SV_S^\top\right)}{\sum_{S'\in\mathcal{U}}\mu(S')\det\left(V_{S'}V_{S'}^\top\right)}\frac{E_{d-1}\left(V_SV_S^\top\right)}{\det\left(V_SV_S^\top\right)}\\&=\frac{\sum_{S\in\mathcal{U}}\mu(S)E_{d-1}\left(V_SV_S^\top\right)}{\sum_{S\in\mathcal{U}}\mu(S)\det\left(V_SV_S^\top\right)}.\end{aligned}$$

We can now apply the Cauchy-Binet Equation (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

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

where we change the order of summation at the second to last equality. Next, we apply (6) and the Cauchy-Binet Equation (4) in a similar way to the matrix $V(\boldsymbol{x})V(\boldsymbol{x})^\top$:

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

where we use the fact that $\det\left(V_R(\boldsymbol{x})^\top V_R(\boldsymbol{x})\right) = x^R\det\left(V_R^\top V_R\right)$ and $\det\left(V_T(\boldsymbol{x})^\top V_T(\boldsymbol{x})\right) = x^T\det\left(V_T^\top V_T\right)$ in the last equality by Lemma 2.

Hence, Inequality (5), which we want to show is equivalent to

$$\frac{\sum_{|T|=d-1,T\subseteq[n]}\det\left(V_T^\top V_T\right)\Pr_\mu[\mathcal{S}\supseteq T]}{\sum_{|R|=d,R\subseteq[n]}\det\left(V_R^\top V_R\right)\Pr_\mu[\mathcal{S}\supseteq R]} \le \alpha\frac{\sum_{|T|=d-1,T\subseteq[n]}\det\left(V_T^\top V_T\right)x^T}{\sum_{|R|=d,R\subseteq[n]}\det\left(V_R^\top V_R\right)x^R}, \tag{7}$$

which is equivalent to

$$\begin{aligned}&\sum_{|T|=d-1,\,|R|=d}\det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right)\cdot x^R\cdot\Pr_\mu[\mathcal{S}\supseteq T]\\ &\le \alpha\sum_{|T|=d-1,\,|R|=d}\det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right)\cdot x^T\cdot\Pr_\mu[\mathcal{S}\supseteq R].\end{aligned} \tag{8}$$

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 $|R| = d$,

$$\det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right)\cdot x^R\cdot\Pr_\mu[\mathcal{S}\supseteq T] \le \alpha\det\left(V_T^\top V_T\right)\det\left(V_R^\top V_R\right)\cdot x^T\cdot\Pr_\mu[\mathcal{S}\supseteq R]. \tag{9}$$

Summing (9) over all $T, R$ proves (8). □

## 3. Approximating Optimal Design Without Repetitions

In this section, we prove Theorem 2 by constructing $\alpha$-approximate $(d-1, d)$-independent distributions for appropriate values of $\alpha$. We first consider the case when $k = d$ and then the asymptotic case when $k = \Omega(\frac{d}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$. 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. The *d*-Approximation for *k*=*d*

We prove the following lemma that, together with Theorem 1, implies the $d$-approximation for $A$-optimal design when $k = d$.

**Lemma 3.** *Let $k = d$. The hard-core distribution $\mu$ 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 $\mu(S) = x^S/Z$, where $Z = \Sigma_{S'\in\mathcal{U}_k}x^{S'}$ is the normalization factor. For any $T\subseteq[n]$ such that $|T| = d-1$, we have

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

We use $k=d$ and $\Sigma_{i\in[n]\setminus T} x_i \le k=d$. 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

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

### 3.2. The $(1+\epsilon)$-Approximation

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

**Lemma 4.** *Fix some* $0<\epsilon\le 2$, *and let* $k=\Omega(d/\epsilon+\log(1/\epsilon)/\epsilon^2)$. *The hard-core distribution* $\mu$ *on* $\mathcal{U}_{\le k}$ *with parameter* $\lambda$, *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 $\beta=1+\epsilon/4$, and $\xi_i=x_i/\beta$. Observe that the probability mass under $\mu$ of any set $S$ of size at most $k$ is proportional to $(\prod_{i\in S}\xi_i)(\prod_{i\notin S}(1-\xi_i))$. Thus, $\mu$ 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 $\xi_i$; then we have $\mu(S)=\Pr[\mathcal{B}=S\,|\,|\mathcal{B}|\le k]$ 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 previous observation, for $\mathcal{S}$ sampled according to $\mu$, and $\mathcal{B}$ as earlier, 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 $\Pr[\mathcal{B}\supseteq T]=\xi^T=x^T/\beta^{d-1}$. To simplify the probability in the denominator, let us introduce, for each $i\in[n]$, the indicator random variable $Y_i$, defined to be one if $i\in\mathcal{B}$ and zero otherwise. By the choice of $\mathcal{B}$, the $Y_i$ s are independent Bernoulli random variables with mean $\xi_i$, 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 $\xi^R=x^R/\beta^d$, and plugging into the previous inequality, we get

$$\frac{\Pr[\mathcal{S}\supseteq T]}{\Pr[\mathcal{S}\supseteq R]}\le\beta\frac{x^T}{x^R\Pr\left[\sum_{i\notin R}Y_i\le k-d\right]}.\tag{10}$$

We claim that

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

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

$$\Pr[Y>k-d]<e^{-\frac{\delta^2}{3\beta}(k-x(R))},\tag{11}$$

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 (11) simplifies to

$$\frac{\delta^2(k-x(R))}{3\beta}=\frac{((\beta-1)\,k+x\,(R)-\beta d)^2}{3\beta\,(k-x\,(R))}\geq\frac{((\beta-1)\,k-\beta d)^2}{3\beta k}.$$

For the bound $\Pr[Y>k-d]\leq\epsilon/4$, it suffices to have

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

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

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

$$\frac{\Pr[\mathcal{S}\supseteq T]}{\Pr[\mathcal{S}\supseteq R]}\leq\beta\frac{x^T}{x^R\Pr\left[\sum_{i\notin R}Y_i\leq k-d\right]}\leq\frac{1+\frac{\epsilon}{4}x^T}{1-\frac{\epsilon}{4}x^R},\tag{12}$$

and

$$\frac{1+\frac{\epsilon}{4}}{1-\frac{\epsilon}{4}}\leq1+\epsilon.\quad\square$$

The $(1+\epsilon)$-approximation for large enough $k$ in Theorem 2 now follows directly from Lemma 4 and Theorem 1.

## 4. Approximately Optimal Design with Repetitions

In this section, we consider the $A$-optimal design without the bound $x_i\leq1$ and prove Theorem 3. That is, we allow the sample set $S$ to be a multiset. 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 obtain an improved approximation.

We first describe a sampling Algorithm 2 that achieves a $(k(1+\epsilon))/(k-d+1)$-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 $\mu$ is defined on sets of the new larger ground set $U$ over copies of the original input vectors. The distribution $\mu$ 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 that 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 $\mu$ is uniform and that $U$ has only at most $n$ distinct vectors.

**Algorithm 2** (Approximation Algorithm for A-Optimal Design with Repetitions)

1: Given $x\in\mathbb{R}^n_+$ with $\Sigma_{i=1}^n x_i=k$, $\epsilon>0$, and vectors $v_1,\ldots,v_n$.

2: Let $q=(2n)/(\epsilon k)$. Set $x_i':=((k-n/q)/k)x_i$ for each $i$, and round each $x_i'$ up to a multiple of $1/q$.

3: If $\Sigma_{i=1}^n x_i'<k$, add $1/q$ to any $x_i'$ until $\Sigma_{i=1}^n x_i'=k$.

4: Create $qx_i'$ copies of vector $\boldsymbol{v}_i$ for each $i\in[n]$. Denote $W$ the set of size $\Sigma_{i=1}^n qx_i'=qk$ 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 $\Pr[\mathcal{S}=S]\propto\det\left(W_SW_S^\top\right)$ for each $S\subseteq U$ of size $k$.

6: Set $X_i=\Sigma_{w\in W_S}\mathbb{1}(\boldsymbol{w}\text{ is a copy of }\boldsymbol{v}_i)$ for all $i\in[n]$ ▷ Get an integral solution $X$ by counting numbers of copies of $\boldsymbol{v}_i$ in $\mathcal{S}$.

7: Output $X$.

**Lemma 5.** *Algorithm 2, when given as input $x\in\mathbb{R}^n_+$ s.t. (such that) $\Sigma_{i=1}^n x_i=k$, $1\geq\epsilon>0$, and $v_1,\ldots,v_n$, outputs a random $X\in\mathbb{Z}^n_+$ with $\Sigma_{i=1}^n X_i=k$ such that*

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

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

$$\mathbb{E}\left[\operatorname{tr}\left(\sum_{i=1}^{n} X_i \boldsymbol{v}_i v_i^{\top}\right)^{-1}\right] \le \frac{k}{k-d+1} \operatorname{tr}\left(\sum_{i=1}^{n} x_i' \boldsymbol{v}_i v_i^{\top}\right)^{-1} \le \frac{k(1+\epsilon)}{k-d+1} \operatorname{tr}\left(\sum_{i=1}^{n} x_i \boldsymbol{v}_i v_i^{\top}\right)^{-1}.$$

The second inequality is by observing that the scaling $x_i' := ((k-n/q)/k)x_i$ multiplies the objective $\operatorname{tr}(\Sigma_{i=1}^n x_i \boldsymbol{v}_i v_i^{\top})^{-1}$ by a factor of

$$\left(\frac{k-n/q}{k}\right)^{-1} = \left(1-\frac{\epsilon}{2}\right)^{-1} \le 1+\epsilon$$

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 $\operatorname{tr}(\Sigma_{i=1}^n x_i' \boldsymbol{v}_i v_i^{\top})^{-1}$ and $\operatorname{tr}(\Sigma_{i=1}^n X_i \boldsymbol{v}_i v_i^{\top})^{-1}$ from the with-repetition setting over $V$ and $[n]$ to the without-repetition setting over $W$ and $U$. First, $\operatorname{tr}(\Sigma_{i=1}^n x_i' \boldsymbol{v}_i v_i^{\top})^{-1} = \operatorname{tr}(\Sigma_{i\in U} y_i w_i w_i^{\top})^{-1}$, where $y_i = 1/q$ are all equal over all $i \in U$, and $w_i$ is the copied vector in $W$ at index $i \in U$. Second, $\operatorname{tr}(\Sigma_{i=1}^n X_i \boldsymbol{v}_i v_i^{\top})^{-1} = \operatorname{tr}(\Sigma_{i\in\mathcal{S}\subseteq U} w_i w_i^{\top})^{-1}$.

Let $\mu'$ 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 $\mu'$ satisfies

$$\mathbb{E}_{\mathcal{S}\sim\mu'}\left[\operatorname{tr}\left(\sum_{i\in\mathcal{S}\subseteq U} w_i w_i^{\top}\right)^{-1}\right] \le \frac{k}{k-d+1} \operatorname{tr}\left(\sum_{i\in U} y_i w_i w_i^{\top}\right)^{-1}. \tag{13}$$

Observe that $\mu'$ 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 $\mu$ is uniform. Hence, by Theorem 1, it is enough to show that for all $T, R \subseteq U$ with $|T| = d-1, |R| = d$,

$$\frac{\Pr_{\mu}[\mathcal{S}\supseteq T]}{\Pr_{\mu}[\mathcal{S}\supseteq R]} \le \left(\frac{k}{k-d+1}\right)\frac{y^T}{y^R}, \tag{14}$$

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

$$\frac{\Pr_{\mu}[\mathcal{S}\supseteq T]}{\Pr_{\mu}[\mathcal{S}\supseteq R]} = \frac{\binom{qk-d+1}{k-d+1}\Big/\binom{qk}{k}}{\binom{qk-d}{k-d}\Big/\binom{qk}{k}} = \frac{qk-d+1}{k-d+1} \quad \text{and} \quad \frac{y^T}{y^R} = \frac{1}{y_i} = q. \tag{15}$$

Therefore, (14) holds because

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

**Remark 1.** The approximation ratio for A-optimality with repetitions for $k \ge d$ is tight because it matches the integrality gap lower bound stated in Theorem 20.

## 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.

### 5.1. The *k*-Approximation Algorithm for *k* ≤ *d*

Recall that our aim is to select a set $S\subseteq[n]$ of size $k \le d$ that minimizes $\Sigma_{i=1}^k \lambda_i^{-1}$, 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 because when $k < d$, the inverse of $M(S) = \Sigma_{i\in S}\, v_i v_i^\top$ for $|S| = k$ is no longer well defined. We write a new convex program:

$$\min \frac{E_{k-1}\left(\sum_{i=1}^{n} x_i v_i v_i^\top\right)}{E_k\left(\sum_{i=1}^{n} x_i v_i v_i^\top\right)}, \tag{16}$$

s.t.

$$\sum_{i=1}^{n} x_i = k, \tag{17}$$

$$0 \le x_i \le 1 \quad \forall i \in [n]. \tag{18}$$

Once again we denote the optimal value of (16)–(18) by CP. Although the proof that this relaxes the original problem is easy, the convexity is nontrivial. Fortunately, ratios of symmetric polynomials are known to be convex.

**Lemma 6.** *The optimization problem* (16)–(18) *is a convex relaxation of the A-optimal design problem when* $k \le d$.

**Proof.** To prove convexity, we first note that the function $f(M) = (E_k(M))/(E_{k-1}(M))$ is concave on positive semidefinite matrices $M$ of rank at least $k$. This was proved by Bullen and Marcus [11, theorem 4] for positive definite $M$, and can be extended to $M$ of rank at least $k$ by a limiting argument. Alternatively, we can use the theorem of Marcus and Lopes [33] that the function $g(\lambda) = (e_k(\lambda))/(e_{k-1}(\lambda))$ is concave on vectors $\lambda \in \mathbb{R}^d$ with nonnegative entries and at least $k$ positive entries. Because $g$ is symmetric under permutations of its arguments and concave, and $f(M) = g(\lambda(M))$, where $\lambda(M)$ is the vector of eigenvalues of $M$, by a classical result of Davis [14], the function $f$ is concave on positive semidefinite matrices of rank at least $k$.

Notice that Objective (16) equals $1/(f(M(x)))$ for the linear matrix-valued function $M(x) = \Sigma_{i=1}^{n} x_i v_i v_i^\top$. Therefore, to prove that (16) is convex in $x$ for nonnegative $x$, it suffices to show that $1/f(M)$ is convex in $M$ for positive semidefinite $M$. Because the function $1/z$ is convex and monotone decreasing over positive reals $z$, and $f$ is concave and nonnegative over positive semidefinite matrices of rank at least $k$, we have that $1/(f(M))$ is convex in $M$, as desired. Then (16)–(18) is an optimization problem with a convex objective and affine constraints, so we have a convex optimization problem.

Let OPT be the optimal value of the $A$-optimal design problem and let $S$ be an optimal solution. We need to show that $\mathrm{CP} \le \mathrm{OPT}$. To this end, let $x$ be the indicator vector of $S$, that is, $x_i = 1$ if and only if $i \in S$, and $x_i = 0$ otherwise. Then,

$$\mathrm{CP} \le \frac{E_{k-1}(M(S))}{E_k(M(S))} = \frac{\sum_{i=1}^{k}\prod_{j\ne i}\lambda_j(M(S))}{\prod_i \lambda_i(M(S))} = \sum_{i=1}^{k}\frac{1}{\lambda_i(M(S))} = \mathrm{OPT}.$$

Previously, $\lambda_1(M(S)), \ldots, \lambda_k(M(S))$ are, again, the nonzero eigenvalues of $M(S) = \sum_{i\in S} v_i v_i^\top$. □

We shall use the natural analog of proportional volume sampling: given a measure $\mu$ 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, for example, the algorithm of Deshpande and Rademacher [19].

The following version of Theorem 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 6.** *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* $\mu$ *on* $\mathcal{U}_k$ *that is* $\alpha$*-approximate* $(k-1,k)$*-wise independent. Then for* $x$ *the optimal solution of* (16)–(18), *proportional volume sampling with measure* $\mu$ *gives a polynomial time* $\alpha$*-approximation algorithm for the A-optimal design problem.*

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

**Theorem 7.** *For any* $k \le d$, *proportional volume sampling with the hard-core measure* $\mu$ *on* $\mathcal{U}_k$ *with parameter* $x$ *equal to the optimal solution of* (16)–(18) *gives a* $k$*-approximation to the A-optimal design problem.*

**Proof.** In view of Theorem 6, we only need to show that $\mu$ is $k$-approximate $(k-1,k)$-wise independent. This is a straightforward calculation: for $\mathcal{S} \sim \mu$, and any $T \subseteq [n]$ of size $k-1$ and $R \subseteq [n]$ of size $k$,

$$\frac{\Pr[\mathcal{S} \supseteq T]}{\Pr[\mathcal{S} \supseteq R]} = \frac{x^T \sum_{i\notin T} x_i}{x^R} \le k\frac{x^T}{x^R}.$$

This completes the proof. □

The algorithm can be derandomized using the method of conditional expectations analogously to the case of $k = d$ that we will show in Theorem 14.

The $k$-approximation also matches the integrality gap of (16)–(18). 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 zeros. On such an instance, the convex program (16)–(18) is equivalent to the convex program (1)–(3). Thus, the integrality gap that we will show in Theorem 20 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. In this section we use the notation $\|M\|_p = (\Sigma_{i=1}^d |\lambda_i|^p)^{1/p}$ for the Schatten-$p$ norm of a symmetric $d$ by $d$ matrix $M$ with eigenvalues $\lambda_1, \ldots, \lambda_d$. When $M$ is positive semidefinite, this is simply $\|M\|_p = \mathrm{tr}(M^p)^{1/p}$. The Schatten-infinity norm $\|M\|_\infty$ equals the largest absolute value of the eigenvalues of $M$.

**Theorem 8.** *Let $v_1, \ldots, v_n \in \mathbb{R}^d$, and $c_1, \ldots, c_n \in \mathbb{R}_+$, and define $M = \Sigma_{i=1}^n c_i v_i v_i^\top$. For any $p \in (1, \infty]$ and $q \in [1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, and any integer $k \leq r_p = \|M\|_1^q / \|M\|_p^q$, there exists a subset $S \subseteq [n]$ of size $k$ such that the $k$ largest eigenvalues $\lambda_1, \ldots, \lambda_k$ of the matrix $\Sigma_{i \in S} v_i v_i^\top$ satisfy*

$$\left(\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\lambda_i}\right)^{-1} \geq \left(1 - \left(\frac{k-1}{r_p}\right)^{\frac{1}{q}}\right) \cdot \frac{\mathrm{tr}\,(M)}{\sum_{i=1}^n c_i}.$$

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

The proof of Theorem 8 relies on the following lemma.

**Lemma 7.** *Let $v_1, \ldots, v_n \in \mathbb{R}^d$, and $c_1, \ldots, c_n \in \mathbb{R}_+$, and define $M = \Sigma_{i=1}^n c_i v_i v_i^\top$. For any $k \leq d$, there exists a set $S$ of size $k$ such that the $k$ largest eigenvalues $\lambda_1, \ldots, \lambda_k$ of $\Sigma_{i \in S} v_i v_i^\top$ satisfy*

$$\left(\frac{1}{k}\Sigma_{i=1}^{k}\frac{1}{\lambda_i}\right)^{-1} \geq \frac{E_k(M)}{E_{k-1}(M)} \cdot \frac{k}{\Sigma_{i=1}^n c_i}.$$

*Moreover, such a set can be found in deterministic polynomial time.*

**Proof.** Without loss of generality we can assume that $\Sigma_{i=1}^n c_i = k$. Then, by Theorem 7, proportional volume sampling with the hard-core measure $\mu$ on $\mathcal{U}_k$ with parameter $c = (c_1, \ldots, c_n)$ gives a random set $\mathcal{S}$ of size $k$ such that

$$\mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\lambda_i(M(\mathcal{S}))}\right] \leq \frac{E_{k-1}(M)}{E_k(M)},$$

where $\lambda_i(M(\mathcal{S}))$ is the $i$th largest eigenvalues of $M(\mathcal{S}) = \sum_{i \in \mathcal{S}} v_i v_i^\top$. This implies the existence of the claimed set. The fact that the set can be found in deterministic polynomial time follows by Theorem 14. □

A similar result is implicit in Naor and Youssef [35]. In particular, combining lemma 18 and equality (12) in Naor and Youssef [35] shows that, in the setting of Lemma 7, for any $k \leq \mathrm{rank}\, M$, there exists a set $S$ 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

$$\left(\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\lambda_i}\right)^{-1} \geq \frac{\sum_{i=k}^d \lambda_i(M)}{\sum_{i=1}^n c_i}, \tag{19}$$

where $\lambda_1(M) \geq \ldots \geq \lambda_d(M)$ are the eigenavalues of $M$. Our Lemma 7 is stronger in a couple of ways. First, it gives an efficient algorithm to compute the set $S$. Furthermore, it gives a bound which dominates the one in (18). Namely, for any vector $\lambda \in \mathbb{R}_+^d$, such that $\lambda_1 \geq \ldots \geq \lambda_d$, we have

$$k \cdot e_k(\lambda) = \sum_{T \subseteq [d]: |T| = k-1} \lambda^T \sum_{i \notin T} \lambda_i \geq \sum_{T \subseteq [d]: |T| = k-1} \lambda^T \sum_{i=k}^{d} \lambda_i = e_{k-1}(\lambda) \cdot \left(\sum_{i=k}^{d} \lambda_i\right). \tag{20}$$

**Proof of Theorem 8.** Equation (5) in Naor and Youssef [35] shows that, for any $k \leq d$,

$$\sum_{i=k}^{d} \lambda_i(M) \geq \left(1 - \left(\frac{k-1}{r_p}\right)^{\frac{1}{q}}\right) \cdot \mathrm{tr}\,(M).$$

This inequality, together with Lemma 7 and (20) imply the theorem. □

### 5.3. Generalized Ratio Objective

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

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

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

$$\min_{S\subseteq[n],|S|=k} \left(\frac{E_{l'}\left(V_S V_S^\top\right)}{E_l\left(V_S V_S^\top\right)}\right)^{\frac{1}{l-l'}}. \tag{21}$$

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

$$\min_{S\subseteq[n],|S|=k} \left(\frac{1}{\det\left(V_S V_S^\top\right)}\right)^{\frac{1}{d}}. \tag{22}$$

A closely related generalization between $A$- and $D$-criteria was considered in Mariet and Sra [34]. Indeed, their generalization corresponds to the case when $l = d$ and $l'$ takes any value from 0 and $d-1$.

| With repetitions | Without repetitions | |
|---|---|---|
| $\min \left(\frac{E_{l'}\left(V(x)V(x)^\top\right)}{E_l\left(V(x)V(x)^\top\right)}\right)^{\frac{1}{l-l'}}$ | $\min \left(\frac{E_{l'}\left(V(x)V(x)^\top\right)}{E_l\left(V(x)V(x)^\top\right)}\right)^{\frac{1}{l-l'}}$ | (1) |
| s.t. $\sum_{i=1}^n x_i = k$ | s.t. $\sum_{i=1}^n x_i = k$ | (2) |
| $0 \le x_i \quad \forall i \in [n]$ | $0 \le x_i \le 1 \quad \forall i \in [n]$ | (3) |

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, summarized in Table 2, 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 relaxations for with and without repetitions as (23)–(25).

We now show that $\left(\frac{E_{l'}\left(V(x)V(x)^\top\right)}{E_l\left(V(x)V(x)^\top\right)}\right)^{\frac{1}{l-l'}}$ is convex in $x$.

**Lemma 8.** *Let $d$ be a positive integer. For any given pair $0 \le l' < l \le d$, the function*

$$f_{l',l}(M) = \left(\frac{E_{l'}(M)}{E_l(M)}\right)^{\frac{1}{l-l'}} \tag{26}$$

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

**Table 2.** Summary of approximation ratio obtained by our work on generalized ratio problem.

| Problem | $A$-optimal ($l' = d-1, l = d$) | $\min_{\lvert S\rvert=k}\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}$ | $D$-optimal ($l' = 0, l = d$) |
|---|---|---|---|
| Case $k = d$ | $d$ | $l \cdot [(l-l')!]^{-\frac{1}{l-l'}} \le \frac{el}{l-l'}$ | $e$ |
| Asymptotic $k \gg d$ without Repetitions | $1+\epsilon$, for | $1+\epsilon$, for | $1+\epsilon$, for |
| | $k \ge \Omega\left(\frac{d}{\epsilon} + \frac{\log 1/\epsilon}{\epsilon^2}\right)$ | $k \ge \Omega\left(\frac{l}{\epsilon} + \frac{\log 1/\epsilon}{\epsilon^2}\right)$ | $k \ge \Omega\left(\frac{d}{\epsilon} + \frac{\log 1/\epsilon}{\epsilon^2}\right)$ |
| Arbitrary $k$ and $d$ with Repetitions | $\frac{k}{k-d+1}$ | $\frac{k}{k-l+1}$ | $\frac{k}{k-d+1}$ |
| Asymptotic $k \gg d$ with Repetitions | $1+\epsilon$, for $k \ge d + \frac{d}{\epsilon}$ | $1+\epsilon$, for $k \ge l + \frac{l}{\epsilon}$ | $1+\epsilon$, for $k \ge d + \frac{d}{\epsilon}$ |

**Proof.** By theorem 3 in Bullen and Marcus [11], $(f_{l',l}(M))^{-1} = ((E_l(M))/(E_{l'}(M)))^{1/l-l'}$ is concave on positive semidefinite matrices $M$ for each $0 \le l' < l \le d$. The function $\frac{1}{z}$ is convex and monotone decreasing over the positive reals $z$, and this, together with the concavity of $(f_{l',l}(M))^{-1}$ and that $(f_{l',l}(M))^{-1} > 0$, implies that $f_{l',l}(M)$ is convex in $M$. □

**5.3.2. Approximation via $(l', l)$-Wise Independent Distribution.** Let $0 \le l' < l \le d$ and $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\le k}\}$. We first show connection of approximation guarantees on objectives $((E_{l'}(V_S V_S^\top))/(E_l(V_S V_S^\top)))^{1/l-l'}$ and $(E_{l'}(V_S V_S^\top))/(E_l(V_S V_S^\top))$. Suppose we already solve the convex relaxation of generalized ratio problem (23)–(25) and get a fractional solution $x \in \mathbb{R}^n$. Suppose that a randomized algorithm $\mathcal{A}$, on receiving input $V \in \mathbb{R}^{d\times n}$ and $x \in \mathbb{R}^n$, outputs $S \in \mathcal{U}$ such that

$$\mathop{\mathbb{E}}_{S\sim\mathcal{A}}\left[\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right] \le \alpha' \frac{E_{l'}(V(x)V(x)^\top)}{E_l(V(x)V(x)^\top)} \tag{27}$$

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

$$\mathbb{E}\left[\frac{E_{l'}(M)}{E_l(M)}\right] \ge \mathbb{E}\left[\left(\frac{E_{l'}(M)}{E_l(M)}\right)^{\frac{1}{l-l'}}\right]^{l-l'} \tag{28}$$

for any semipositive definite matrix $M$. Combining (27) and (28) gives

$$\mathop{\mathbb{E}}_{S\sim\mathcal{A}}\left[\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}\right] \le \alpha\left(\frac{E_{l'}(V(x)V(x)^\top)}{E_l(V(x)V(x)^\top)}\right)^{\frac{1}{l-l'}}, \tag{29}$$

where $\alpha = (\alpha')^{1/(l-l')}$. Therefore, it is sufficient for an algorithm to satisfy (27) and give a bound on $\alpha'$ in order to solve the generalized ratio problem up to factor $\alpha$.

To show (27), we first define the proportional $l$-volume sampling and $\alpha$-approximate $(l', l)$-wise independent distribution.

**Definition 3.** Let $\mu$ be probability measure on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\le k}$). Then the proportional $l$-volume sampling with measure $\mu$ picks a set of vectors indexed by $S \in \mathcal{U}_k$ (or $\mathcal{U}_{\le k}$) with probability proportional to $\mu(S)E_l(V_S V_S^\top)$.

**Definition 4.** Given integers $d, k, n$, a pair of integers $0 \le l' \le l \le d$, and a vector $x \in [0,1]^n$ such that $1^\top x = k$, we call a measure $\mu$ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\le k}$), $\alpha$-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

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

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 (27) to $\alpha$-approximate $(l', l)$-wise independence properties of a certain distribution $\mu$ by using proportional $l$-volume sampling.

**Theorem 9.** *Given integers $d, k, n$, $V = [v_1 \dots v_n] \in \mathbb{R}^{d\times n}$, and a vector $\boldsymbol{x} \in [0, 1]^n$ such that $1^\top \boldsymbol{x} = k$, suppose there exists a distribution $\mu$ on sets in $\mathcal{U}_k$ (or $\mathcal{U}_{\le k}$) and is $\alpha$-approximate $(l', l)$-wise independent for some $0 \le l' < l \le d$. Then the proportional $l$-volume sampling with measure $\mu$ gives an $\alpha$-approximation algorithm for minimizing $((E_{l'}(V_S V_S^\top))/(E_l(V_S V_S^\top)))^{1/(l-l')}$ over subsets $S \subseteq [n]$ of size $k$.*

**Proof.** Let $\mu'$ 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 $\mu(S)E_l(V_S V_S^\top)$. We mechanically replace $T, R, d-1, d,$ and det in the proof of Theorem 1 with $T', T, l', l,$ and $E_l$ to obtain

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

We finish the proof by observing that (27) implies (29), as discussed earlier. □

The following subsections generalize algorithms and proofs for with and without repetitions. The algorithm for generalized ratio problem can be summarized in Algorithm 3. 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 $0 \le l' < l \le d$.

2: Solve the convex relaxation

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

where $J = [0,1]$ if without repetitions or $\mathbb{R}^+$ if with repetitions.

3: **if** $k = l$ **then**

4: Sample $\mu'(S) \propto x^S E_l(V_S V_S^\top)$ for each $S \in \mathcal{U}_k$

5: **else if** without repetition setting and $k \ge \Omega\left(d/\epsilon + (\log(1/\epsilon))/\epsilon^2\right)$ **then**

6: Sample $\mu'(S) \propto \lambda^S E_l(V_S V_S^\top)$ for each $S \in \mathcal{U}_{\le k}$ where $\lambda_i := x_i/(1 + \epsilon/4 - x_i)$

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: **end if**

10: 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 that generalizes Lemmas 3 and 4. The $\alpha$-approximate $(l', l)$-wise independence property, together with Theorem 9, implies an approximation guarantee for generalized ratio problem without repetitions for $k = l$ and asymptotically for $k = \Omega(l/\epsilon + (1/\epsilon^2)\log(1/\epsilon))$.

**Theorem 10.** *Given integers $d, k, n$, a pair of integers $0 \le l' \le l \le d$, and a vector $x \in [0,1]^n$ such that $1^\top x = k$, the hard-core distribution $\mu$ on sets in $\mathcal{U}_k$ with parameter $x$ is $\alpha$-approximate $(l', l)$-wise independent when $k = l$ for*

$$\alpha = l \cdot [(l-l')!]^{-\frac{1}{l-l'}} \le \frac{el}{l-l'}. \tag{30}$$

*Moreover, for any $0 < \epsilon \le 2$ when $k = \Omega(\frac{l}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$, the hard-core distribution $\mu$ on $\mathcal{U}_{\le k}$ with parameter $\lambda$, 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 $\left((E_{l'}(V_S V_S^\top))/(E_l(V_S V_S^\top))\right)^{1/(l-l')}$ over subsets $S \subseteq [n]$ of size $k$, we obtain*

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

**Proof.** We first prove the result for $k = 1$. For all $T', T \subseteq [n]$ such that $|T'| = l', |T| = l$,

$$\frac{\Pr_{\mathcal{S}\sim\mu}[S \supseteq T']}{\Pr_{\mathcal{S}\sim\mu}[S \supseteq T]} = \frac{\sum_{|S|=k,\, S \supseteq T'} x^S}{\sum_{|S|=k,\, S \supseteq T} x^S} = \frac{x^{T'} \sum_{L \in \binom{[n]\setminus T'}{k-l'}} x^L}{x^T} \le \frac{x^{T'} \sum_{L \in \binom{[n]}{k-l'}} x^L}{x^T}.$$

We now use Maclaurin's inequality (Lin and Trudinger [31]) to bound the quantity on the right-hand side

$$\sum_{L \in \binom{[n]}{k-l'}} x^L = e_{l-l'}(x) \le \binom{n}{l-l'} (e_1(x)/n)^{l-l'} \le \frac{n^{l-l'}}{(l-l')!}(l/n)^{l-l'} = \frac{l^{l-l'}}{(l-l')!}. \tag{31}$$

Therefore,

$$\frac{\Pr_{\mathcal{S}\sim\mu}[S \supseteq T']}{\Pr_{\mathcal{S}\sim\mu}[S \supseteq T]} \le \frac{l^{l-l'}}{(l-l')!}\frac{x^{T'}}{x^T}, \tag{32}$$

which proves the $(l', l)$-wise independent property of $\mu$ with required approximation ratio from (30).

We now prove the result for $k = \Omega(l/\epsilon + (l/\epsilon^2)\log(l/\epsilon))$. The proof follows similarly from Lemma 4 by replacing $T, R$ with $T', T$ of sizes $l', l$ instead of sizes $d-1, d$. In particular, Equation (10) becomes

$$\frac{\Pr[\mathcal{S} \supseteq T']}{\Pr[\mathcal{S} \supseteq T]} \le \left(1 + \frac{\epsilon}{4}\right)^{l-l'} \frac{x^{T'}}{x^T \Pr\left[\sum_{i \notin T} Y_i \le k - l\right]}. \tag{33}$$

The Chernoff's bound (11) still holds by mechanically replacing $d, R$ with $l, T$ respectively. The resulting approximation ratio $\alpha$ satisfies

$$\alpha^{l-l'} = \frac{\left(1 + \frac{\epsilon}{4}\right)^{l-l'}}{1 - \frac{\epsilon}{4}} \le (1+\epsilon)^{l-l'},$$

where the inequality holds because $\epsilon \le 2$. □

**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 5.

**Theorem 11.** *Given $V = [v_1 v_2 \ldots v_n]$, where $\boldsymbol{v}_i \in \mathbb{R}^d$, a pair of integers $0 \le l' \le l \le d$, an integer $k \ge l$, and $1 \ge \epsilon > 0$, there is an $\alpha$-approximation algorithm for minimizing $\left(\left(E_{l'}(V_s V_s^\top)\right)/\left(E_l(V_s V_s^\top)\right)\right)^{1/(l-l')}$ over subsets $S \subseteq [n]$ of size $k$ with repetitions for*

$$\alpha \le \frac{k(1+\epsilon)}{k-l+1}. \tag{34}$$

**Proof.** We use the algorithm similar to Algorithm 2 except that in Step 5, we sample $S \subseteq U$ of size $k$ where $\Pr[\mathcal{S} = S] \propto E_1(W_s {W_s}^T)$ in place of $\Pr[\mathcal{S} = S] \propto E_1(W_s W_s^\top)$. The analysis follows on the same line as in Lemma 5. In Lemma 5, it is sufficient to show that the uniform distribution $\mu$ over subsets $S \subseteq U$ of size $k$ is $k/(k-d+1)$-approximate $(d-1, d)$-wise independent (as in (13)). Here, it is sufficient to show that the uniform distribution $\mu$ is $k/(k-l+1)$-approximate $(l', l)$-wise independent. For $T, T' \subseteq [n]$ of size $(l', l)$, the calculation of $\left(\Pr_\mu[\mathcal{S} \supseteq T']\right)/\left(\Pr_\mu[\mathcal{S} \supseteq T]\right)$ and $\frac{y^{T'}}{y^T}$ is straightforward

$$\frac{\Pr_\mu[\mathcal{S} \supseteq T']}{\Pr_\mu[\mathcal{S} \supseteq T]} = \frac{\binom{qk-l'}{k-l'} \Big/ \binom{qk}{k}}{\binom{qk-l}{k-l} \Big/ \binom{qk}{k}} \le \frac{(qk)^{l-l'}(k-l)!}{(k-l')!} \quad \text{and} \quad \frac{y^{T'}}{y^T} = q^{l-l'}. \tag{35}$$

Therefore, $\mu$ is $\alpha$-approximate $(l', l)$-wise independent for

$$\alpha = \left(\frac{\Pr_\mu[\mathcal{S} \supseteq T']}{\Pr_\mu[\mathcal{S} \supseteq T]} \cdot \frac{y^T}{y^{T'}}\right)^{\frac{1}{l-l'}} \le \left(\frac{(qk)^{l-l'}(k-l)!}{(k-l')!} q^{l'-l}\right)^{\frac{1}{l-l'}}$$
$$= \frac{k}{[(k-l')(k-l'-1)\cdots(k-l+1)]^{\frac{1}{l-l'}}} \le \frac{k}{k-l+1}$$

as we wanted to show. □

We note that the $l$-proportional volume sampling in the proof of Theorem 11 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 $((E_{l'}(V_s V_s^\top))/(E_l(V_s V_s^\top)))^{1/(l-l')}$ 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 12.** *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 $V = [v_1, \ldots, v_n] \in \mathbb{R}^{d \times n}$ to the problem of minimizing $((E_{l'}(V_s V_s^\top))/(E_l(V_s V_s^\top)))^{1/(l-l')}$ 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 $\left((E_{l'}(V_s V_s^\top))/(E_l(V_s V_s^\top))\right)^{1/(l-l')}$ over subsets $S \subseteq [n]$ of size $k$. The theorem applies to both with and without repetitions.

**Proof.** The instance $V = [v_1, \ldots, v_n]$ will be the same for with and without repetitions. For each $1 \le i \le d$, let $e_i$ denote the unit vector in the direction of axis $i$. Choose

$$v_i = \begin{cases} \sqrt{N} \cdot e_i & \text{for } i = 1, \ldots, l' \\ e_i & \text{for } i = 1, \ldots, l' \end{cases},$$

where $N > 0$ is a constant to be chosen later. Set $v_i, i > l$ to be at least $k$ copies of each of these $v_i$ for $i \le l$, as we can make $n$ as big as needed. Hence, we may assume that we are allowed to pick only $v_i, i \le l$, but with repetitions.

Let $S^*$ represent the set of vectors in OPT and $y_i$ be the number of copies of $v_i$ in $S^*$ for $1 \le i \le l$. Clearly $y_i \ge 1$ for all $i = 1, \ldots, l$ (else the objective is unbounded). The eigenvalues of $V_{S^*} V_{S^*}^\top$ are

$$\lambda(V_{S^*} V_{S^*}^\top) = (y_1 N, y_2 N, \ldots, y_{l'} N, y_{l'+1}, y_{l'+2}, \ldots, y_l, 0, \ldots, 0).$$

Hence, both $E_{l'}(V_{S^*} V_{S^*}^\top) = e_{l'}(\lambda)$ and $E_l(V_{S^*} V_{S^*}^\top) = e_l(\lambda)$ are polynomials in variables $N$ of degree $l'$.

Now let $N \to \infty$. To compute $(\mathsf{OPT})^{l-l'} = (E_{l'}(V_{S^*} V_{S^*}^\top))/(E_l(V_{S^*} V_{S^*}^\top))$, we only need to compute the coefficient of the highest degree monomial $N^{l'}$. The coefficient of $N^{l'}$ in $e_{l'}(\lambda), e_l(\lambda)$ are exactly $\prod_{i=1}^{l'} y_i, \prod_{i=1}^{l'} y_i$, and therefore

$$(\mathsf{OPT})^{l-l'} = \frac{E_{l'}(V_{S^*} V_{S^*}^\top)}{E_l(V_{S^*} V_{S^*}^\top)} \to \frac{\prod_{i=1}^{l'} y_i}{\prod_{i=1}^{l} y_i} = \left(\prod_{i=l'+1}^{l} y_i\right)^{-1}.$$

Observe that $\prod_{i=l'+1}^{l} y_i$ is maximized under the budget constraint $\Sigma_{i=1}^{l} y_i = |S^*| = k$ when $y_j = 1$ for $j = 1, \ldots, l'$. Therefore,

$$\prod_{i=l'+1}^{l} y_i \le \left(\frac{1}{l-l'} \sum_{i=l'+1}^{l} y_i\right)^{l-l'} = \left(\frac{k-l'}{l-l'}\right)^{l-l'},$$

where the inequality is by AM-GM (arithmetic mean and geometric mean). Hence, $\mathsf{OPT}$ is lower bounded by a quantity that converges to $(l-l')/(k-l')$ as $N \to \infty$.

We now give a valid fractional solution $x$ to upper bound CP for each $N > 0$. Choose

$$x_i = \begin{cases} \dfrac{k}{\sqrt{N}} & \text{for } i = 1, \ldots, l' \\ \dfrac{k - \dfrac{kl'}{\sqrt{N}}}{l-l'} & \text{for } i = l'+1, \ldots, l \\ 0 & \text{for } i > l \end{cases}.$$

Then, eigenvalues of $V(x)V(x)^\top$ are

$$\begin{aligned} \lambda' := \lambda(V(x)V(x)^\top) &= (x_1 N, x_2 N, \ldots, x_{l'} N, x_{l'+1}, x_{l'+2}, \ldots, x_l, 0, \ldots, 0) \\ &= (k\sqrt{N}, k\sqrt{N}, \ldots, k\sqrt{N}, x_{l'+1}, x_{l'+2}, \ldots, x_l, 0, \ldots, 0). \end{aligned}$$

Now as $N \to \infty$, the dominating terms of $E_{l'}(V(x)V(x)^\top) = e_{l'}(\lambda')$ is $\prod_{i=1}^{l'}(k\sqrt{N}) = k^{l'}(\sqrt{N})^{l'}$. Also, we have

$$\begin{aligned} E_l(V(x)V(x)^\top) = e_l(\lambda') &= \prod_{i=1}^{l'}(k\sqrt{N}) \prod_{i=l'+1}^{l} x_i \\ &= k^{l'}\left(\frac{k - \frac{kl'}{\sqrt{N}}}{l-l'}\right)^{l-l'} (\sqrt{N})^{l'} \to k^{l'}\left(\frac{k}{l-l'}\right)^{l-l'} (\sqrt{N})^{l'}. \end{aligned}$$

Hence,

$$\mathsf{CP} \le \left(\frac{E_{l'}(V(x)V(x)^\top)}{E_l(V(x)V(x)^\top)}\right)^{l-l'} \to \frac{l-l'}{k}.$$

Therefore, $\frac{\mathrm{OPT}}{\mathrm{CP}}$ is lower bounded by a ratio that converges to

$$\frac{l-l'}{k-l'}\cdot\frac{k}{l-l'}=\frac{k}{k-l'}.\quad\square$$

## 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.

### 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 9.** *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:*

$$F(t_1,t_2,t_3)=\det\left(I_n+t_1\mathrm{diag}(y)+t_1t_2\mathrm{diag}(y)^{1/2}VV^\top\mathrm{diag}(y)^{1/2}\right)$$

*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*

$$\sum_{|S|=k,\,I\subseteq S,\,J\cap S=\emptyset}\lambda^S\sum_{|T|=d_0,\,T\subseteq S}\det(V_T^\top V_T)\tag{36}$$

*is the coefficient of the monomial $t_1^kt_2^{d_0}t_3^{|I|}$. Moreover, this quantity can be computed in $O(n^3d_0k|I|\cdot\log(d_0k|I|))$ number of arithmetic operations.*

**Proof.** Let us first fix some $S\subseteq[n]$. Then we have

$$\sum_{|T|=d_0,\,T\subseteq S}\det(V_T^\top V_T)=E_{d_0}(V_S^\top V_S)=[t_2^{d_0}]\det(I_S+t_2V_SV_S^\top),$$

where the notation $[t_2^{d_0}]p(t_2)$ denotes the coefficient of $t^{d_0}$ in the polynomial $p(t_2)=\det(I_S+t_2V_SV_S^\top)$. The first equality is just Cauchy-Binet, and the second one is standard and follows from the Leibniz formula for the determinant. Therefore, (36) equals

$$\left[t_2^{d_0}\right]\sum_{|S|=k,\,I\subseteq S,\,J\cap S=\emptyset}\lambda^S\det(I_S+t_2V_SV_S^\top).$$

To complete the proof, we establish the following claim.

**Claim 1.** Let $L$ be an $n\times n$ matrix, and let $\lambda,I,J,k,y$ be as in the statement of the lemma. Then,

$$\begin{aligned}\sum_{|S|=k,\,I\subseteq S,\,J\cap S=\emptyset}\lambda^S\det(L_{S,S})&=\left[t_3^{|I|}\right]E_k\left(\mathrm{diag}(y)^{1/2}L\,\mathrm{diag}(y)^{1/2}\right)\\&=\left[t_1^kt_3^{|I|}\right]\det\left(I_n+t_1\mathrm{diag}(y)^{1/2}L\,\mathrm{diag}(y)^{1/2}\right).\end{aligned}$$

**Proof.** By Cauchy-Binet,

$$\begin{aligned}E_k\left(\mathrm{diag}(y)^{1/2}L\,\mathrm{diag}(y)^{1/2}\right)&=\sum_{|S|=k}y^S\det(L_{S,S})\\&=\sum_{|S|=k,\,J\cap S=\emptyset}t_3^{|S\cap I|}\lambda^S\det(L_{S,S}).\end{aligned}$$

The first equality follows. The second is, again, a consequence of the Leibniz formula for the determinant. □

Plugging in $L = I_n + t_2 VV^\top$ in Claim 1 gives that (36) equals

$$\left[t_1^k t_2^{d_0} t_3^{|I|}\right]\det\left(I_n + t_1 \mathrm{diag}(y)^{1/2}\left(I_n + t_2 VV^\top\right)\mathrm{diag}(y)^{1/2}\right)$$
$$= \left[t_1^k t_2^{d_0} t_3^{|I|}\right]\det\left(I_n + t_1 \mathrm{diag}(y) + t_1 t_2 \mathrm{diag}(y)^{1/2} VV^\top \mathrm{diag}(y)^{1/2}\right).$$

This completes the proof. For the running time, the standard computation time of matrix multiplication and determinant of $n \times n$ matrices is $O(n^3)$ entry-wise arithmetic operations. We need to keep all monomials in the form $t_1^a t_2^b t_3^c$, where $a \le k, b \le d_0, c \le |I|$, of which there are $O(d_0 k|I|)$. By representing multivariate monomials in single variable (Pan [41]), we may use fast Fourier transform to do one polynomial multiplication of entries of the matrix in $O(d_0 k|I| \cdot \log(d_0 k|I|))$ number of arithmetic operations. This gives the total running time of $O(n^3 d_0 k|I| \cdot \log(d_0 k|I|))$. □

Using the previous lemma, we now prove the following theorem that will directly imply Lemma 1.

**Theorem 13.** *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 randomized algorithm $\mathcal{A}$ that 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 $\mu'$ with hard-core measure $\mu$ on $\mathcal{U}$ with parameter $\lambda$. Moreover, the algorithm runs in $O(n^4 dk^2 \log(dk))$ number of arithmetic operations.*

**Observation 1.** Wang et al. [50] show 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 $O((k + d^2)^4 dk^2 \log(dk))$. Although the degrees in $d, k$ are not small, this runtime is independent of $n$.

**Observation 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 Allen-Zhu et al. [2]. In particular, solving the continuous relaxation of $A$-optimal design takes $O(n^{2+\omega} \log n)$ number of iterations by standard ellipsoid method and $O((n + d^2)^{3.5})$ number of iterations by SDP (semi-definite program), where $O(n^\omega)$ denotes the runtime of $n \times n$ matrix multiplication. In most applications where $n \gg k$, these running times dominates one of proportional volume sampling.

**Proof.** We can sample by starting with an empty set $S = \emptyset$. Then, in each step $i = 1, 2, \ldots, n$, the algorithm decides with the correct probability

$$\Pr_{\mathcal{S}\sim\mu'}[i \in \mathcal{S} | I \subseteq \mathcal{S}, J \cap \mathcal{S} = \emptyset],$$

whether to include $i$ in $S$ or not, given that we already know that we have included $I$ in $S$ and excluded $J$ from $S$ from previous steps $1, 2, \ldots, i-1$. Let $I' = I \cup \{i\}$. This probability equals

$$\Pr_{\mathcal{S}\sim\mu'}[i \in \mathcal{S} | I \subseteq \mathcal{S}, J \cap \mathcal{S} = \emptyset] = \frac{\Pr_{\mathcal{S}\sim\mu'}[I' \subseteq \mathcal{S}, J \cap \mathcal{S} = \emptyset]}{\Pr_{\mathcal{S}\sim\mu'}[I \subseteq \mathcal{S}, J \cap \mathcal{S} = \emptyset]}$$
$$= \frac{\sum_{S \in \mathcal{U}, I' \subseteq S, J \cap S = \emptyset} \lambda^S \det(V_S V_S^\top)}{\sum_{S \in \mathcal{U}, I \subseteq S, J \cap S = \emptyset} \lambda^S \det(V_S V_S^\top)}$$
$$= \frac{\sum_{S \in \mathcal{U}, I' \subseteq S, J \cap S = \emptyset} \lambda^S \sum_{|R| = d, R \subset S} \det(V_R V_R^\top)}{\sum_{S \in \mathcal{U}, I \subseteq S, J \cap S = \emptyset} \lambda^S \sum_{|R| = d, R \subset S} \det(V_R V_R^\top)},$$

where we apply the Cauchy-Binet formula in the last equality. For $\mathcal{U} = \mathcal{U}_k$, both the numerator and denominator are summations over $S$ restricted to $|S| = k$, which can be computed in $O(n^3 dk^2 \log(dk))$ number of arithmetic operations by Lemma 9. For the case $\mathcal{U} = \mathcal{U}_{\le k}$, we can evaluate summations in the numerator and denominator restricted to $|S| = k_0$ for each $k_0 = 1, 2, \ldots k$ by computing polynomial $F(t_1, t_2, t_3)$ in Lemma 9 only once and then sum those quantities over $k_0$. □

### 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 14.** *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 $\mathcal{A}'$ that outputs $S^* \subseteq [n]$ of size $k$ such that*

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

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

Again, with the assumption that $n \le k + d^2$ (Observation 1), the runtime for deterministic proportional volume sampling is $O((k+d^2)^4 dk^2 \log(dk))$.

**Proof.** To prove the theorem, we derandomize the sampling algorithm in Theorem 13 by the method of conditional expectations. The deterministic algorithm starts with $S^* = \emptyset$, and then chooses, at each step $i = 1, 2, \ldots, n$, whether to pick $i$ to be in $S^*$ or not, given that we know from previous steps to include or exclude each element $1, 2, \ldots, i-1$ from $S^*$. The main challenge is to calculate exactly the quantity of the form

$$X(I,J) := \mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\mathrm{tr}\left(V_{\mathcal{S}}V_{\mathcal{S}}^\top\right)^{-1} \middle| I \subset \mathcal{S}, J \cap \mathcal{S} = \emptyset\right]$$

where $I, J \subseteq [n]$ are disjoint. If we can efficiently calculate the quantity of such form, the algorithm can, at each step $i = 1, 2, \ldots, n$, calculate $X(I' \cup \{i\}, J')$ and $X(I', J' \cup \{i\})$, where $I', J' \subseteq [i-1]$ denote elements we have decided to pick and not to pick, respectively, and then include $i$ to $S^*$ if and only if $X(I' \cup \{i\}, J') \ge X(I', J' \cup \{i\})$.
The quantity $X(I,J)$ equals

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{S}\sim\mu'}\left[\mathrm{tr}\left(V_{\mathcal{S}}V_{\mathcal{S}}^\top\right)^{-1} | I \subset \mathcal{S}, J \cap \mathcal{S} = \emptyset\right] &= \sum_{\substack{S\in\mathcal{U},\\ I\subseteq S, J\cap S=\emptyset}} \Pr_{\mu'}[\mathcal{S} = S | I \subseteq \mathcal{S}, \mathcal{S}\cap J = \emptyset]\,\mathrm{tr}\left[\left(V_S V_S^\top\right)^{-1}\right] \\
&= \sum_{\substack{S\in\mathcal{U},\\ I\subseteq S, J\cap S=\emptyset}} \frac{\lambda^S \det\left(V_S V_S^\top\right)}{\sum_{S'\in\mathcal{U}, I\subseteq S, J\cap S=\emptyset}\lambda^{S'}\det\left(V_{S'}V_{S'}^\top\right)}\,\mathrm{tr}\left[\left(V_S V_S^\top\right)^{-1}\right] \\
&= \frac{\sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\emptyset} \lambda^S E_{d-1}\left(V_S V_S^\top\right)}{\sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\emptyset} \lambda^S \sum_{|R|=d, R\subset S}\det\left(V_R V_R^\top\right)} \\
&= \frac{\sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\emptyset} \lambda^S \sum_{|T|=d-1, T\subset S}\det\left(V_T^\top V_T\right)}{\sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\emptyset} \lambda^S \sum_{|R|=d, R\subset S}\det\left(V_R V_R^\top\right)},
\end{aligned}$$

where we write inverse of trace as ratio of symmetric polynomials of eigenvalues in the third equality and use the Cauchy-Binet formula for the third and the fourth equality. The rest of the proof is now identical to the proof of Theorem 13, except with different parameters $d_0 = d-1, d$ in $f(t_1, t_2, t_3)$ when applying Lemma 9. □

### 6.3. Efficient Randomized Implementation of $k/(k-d+1)$-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 Sections 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 9 not useful, as the matrix needed to be computed has dimension $|U| \times |U|$. The main difference, however, is that the parameter $\lambda$ is now a constant, allowing us to obtain sums by computing a more compact $d \times d$ matrix.

**Lemma 10.** *Let $V = [v_1, \ldots, v_m]$ be a matrix of vectors $\boldsymbol{v}_i \in \mathbb{R}^d$ 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 $\Sigma_{|T|=d_0, |F\cap R|=r} \det\left(V_T^\top V_T\right)$ is the coefficient of $t_1^{d-d_0} t_2^{d_0-r} t_3^r$ in*

$$f(t_1, t_2, t_3) = \det\left(t_1 I_d + \sum_{i\in F} t_3 \boldsymbol{v}_i v_i^\top + \sum_{i\notin F} t_2 \boldsymbol{v}_i v_i^\top\right), \tag{37}$$

*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 $O(n(d-d_0+1)d_0^2 d^2 \log d)$ number of arithmetic operations.*

**Proof.** First, note that $\det\left(t_1 I + \Sigma_{i\in F} t_3 v_i v_i^\top + \Sigma_{i\notin F} t_2 v_i v_i^\top\right) = \prod_{i=1}^d (t_1 + \nu_i)$, where $\nu(M) = \{\nu_1, \ldots, \nu_d\}$ is the vector of eigenvalues of the matrix $M = \Sigma_{i\in F} t_3 v_i v_i^\top + \Sigma_{i\notin F} t_2 v_i v_i^\top$. Hence, the coefficient of $t_1^{d-d_0}$ in $\det\left(t_1 I + \Sigma_{i\in F} t_3 v_i v_i^\top + \Sigma_{i\notin F} t_2 v_i v_i^\top\right)$ is $e_{d_0}(\nu(M))$.

Next, observe that $M$ is in the form $V'V'^\top$ where $V'$ is the matrix where columns are $\sqrt{t_3} v_i$, $i \in F$ and $\sqrt{t_2} v_i, i \notin F$. Applying Cauchy-Binet to $E_{d_0}(V'V'^\top)$, we get

$$
\begin{aligned}
E_{d_0}\left(\sum_{i\in F} t_3 v_i v_i^\top + \sum_{i\notin F} t_2 v_i v_i^\top\right) &= E_{d_0}(V'V'\top) = \sum_{|T|=d_0} \det\left(V_T'^\top V_T'\right) \\
&= \sum_{l=0}^{|F|} \sum_{|T|=d_0, |T\cap F|=l} \det\left(V_T'^\top V_T'\right) \\
&= \sum_{l=0}^{|F|} \sum_{|T|=d_0, |T\cap F|=l} t_3^l t_2^{d_0-l} \det\left(V_T^\top V_T\right),
\end{aligned}
$$

where we use Lemma 2 for the last equality. The desired quantity $\sum_{|T|=d_0, |F\cap R|=r} \det\left(V_T^\top V_T\right)$ is then exactly the coefficient at $l = r$ in the sum on the righ- hand side.

To compute the running time, because there are only $n$ distinct vectors, we may represent sets $V, F$ compactly with distinct $v_i$'s and number of copies of each distinct $v_i$'s. Therefore, computing the matrix sum takes $O(nd^2)$ entry-wise operations. Next, the standard computation time of determinant of $d \times d$ matrix is $O(d^3)$ entry-wise arithmetic operations. This gives a total of $O(nd^2 + d^3) = O(nd^2)$ entry-wise operations.

For each entry-wise operation, we keep all monomials in the form $t_1^a t_2^b t_3^c$, where $a \le d - d_0, b \le d_0 - r, c \le r$, of which there are $O((d - d_0 + 1)d_0^2)$. By representing multivariate monomials in a single variable (Pan [41]) of degree $O((d - d_0 + 1)d_0^2)$, we may use fast Fourier transform to do one polynomial multiplication of entries of the matrix in $O((d - d_0 + 1)d_0^2 \log d)$ number of arithmetic operations. This gives the total runtime of $O(n(d - d_0 + 1)d_0^2 d^2 \log d)$ arithmetic operations. □

**Lemma 11.** *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 $\Sigma_{|S|=k, S\supseteq F} E_{d_0}(V_S V_S^\top)$ with $O(n(d - d_0 + 1)d_0^2 d^2 \log d)$ number of arithmetic operations.*

**Proof.** We apply Cauchy-Binet:

$$
\begin{aligned}
\sum_{|S|=k, S\supseteq F} E_{d_0}(V_S V_S^\top) &= \sum_{|S|=k, S\supseteq F} \sum_{|T|=d_0, T\subset S} \det\left(V_T^\top V_T\right) \\
&= \sum_{|T|=d_0} \det\left(V_T^\top V_T\right) \binom{m - |F| - d_0 + |F\cap T|}{k - |F| - d_0 + |F\cap T|} \\
&= \sum_{r=0}^{d} \binom{m - |F| - d_0 + r}{k - |F| - d_0 + r} \sum_{|T|=d_0, |F\cap T|=r} \det\left(V_T^\top V_T\right),
\end{aligned}
$$

where we change the order of summations for the second equality and enumerate over possible sizes of $F \cap T$ to get the third equality. We compute $f(t_1, t_2, t_3)$ in Lemma 10 once with $O(n(d - d_0 + 1)d_0^2 d^2 \log d)$ number of arithmetic operations, so we obtain values of $\Sigma_{|T|=d_0, |F\cap T|=r} \det\left(V_T^\top V_T\right)$ for all $r = 0, \ldots, d_0$. The rest is a straightforward calculation. □

We now present an efficient sampling procedure for Algorithm 2. We want to sample $S$ proportional to $\det\left(W_S W_S^\top\right)$. 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 (37) is still efficiently computable because exponentially many of these copies of vectors are the same.

**Theorem 15.** *Given inputs $n, d, k, \epsilon, x \in \mathbb{R}_+^n$ with $\Sigma_{i=1}^n x_i = k$, 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 $\mu'$ over all subsets $S \subseteq U$ of size $k$, where $\mu'$ is defined by $\Pr_{\mathcal{S}\sim\mu'}[\mathcal{S} = S] \propto \det\left(W_S W_S^\top\right)$ for each $S \subseteq U, |S| = k$. Moreover, $\mathcal{A}$ runs in $O(n^2 d^4 k \log d)$ number of arithmetic operations.*

Theorem 15 says that Steps 4 and 5 in Algorithm 2 can be efficiently implemented. Other steps except 4 and 5 obviously use $O(n^2 d^4 k \log d)$ number of arithmetic operations, so the previous statement implies that Algorithm

2 runs in $O(n^2d^4k\log d)$ number of arithmetic operations. Again, by Observation 1, the number of arithmetic operations is in fact $O((k+d^2)^2d^4k\log d)$.

**Proof.** Let $m_i = qx_i'$ be the number of copies of vector $v_i$ (recall that $q=(2n)/(\epsilon k)$). Let $w_{i,j}$ denote the $j$th copy of vector $v_i$. Write $U=\{(i,j): i\in[n], j\in[m_i]\}$ be the new set of indices after the copying procedure. Denote $\mathcal{S}$ a random subset (not multiset) of $U$ that we want to sample. Write $W$ as the matrix with columns $w_{i,j}$ for all $(i,j)\in U$. Let $E_i=\{w_{i,j}: j=1,\dots,m_i\}$ be the set of copies of vector $v_i$. For any $A\subseteq U$, we say that $A$ *has* $k_i$ *copies of* $v_i$ to mean that $|A\cap E_i|=k_i$.

We can define the sampling algorithm $\mathcal{A}$ by sampling, at each step $t=1,\dots,n$, how many copies of $v_i$ are to be included in $\mathcal{S}\subseteq U$. Denote $\mu'$ the volume sampling on $W$ we want to sample. The problem then reduces to efficiently computing

$$\Pr_{\mu'}[\mathcal{S} \text{ has } k_t \text{ copies of } v_t | \mathcal{S} \text{ has } k_i \text{ copies of } v_i,\ \forall i=1,\dots,t-1] = \frac{\Pr_{\mu'}[\mathcal{S} \text{ has } k_i \text{ copies of } v_i,\ \forall i=1,\dots,t]}{\Pr_{\mu'}[\mathcal{S} \text{ has } k_i \text{ copies of } v_i,\ \forall i=1,\dots,t-1]} \tag{38}$$

for each $k_t=0,1,\dots,k-\Sigma_{i=1}^{t-1}k_i$. Thus, it suffices to efficiently compute quantity (38) for any given $1\le t\le n$ and $k_1,\dots,k_t$ such that $\Sigma_{i=1}^{t}k_i\le k$.

We now fix $t,k_1,\dots,k_t$. For any $i\in[n]$, getting any set of $k_i$ copies of $v_i$ is the same, that is, events $\mathcal{S}\cap E_i=F_i$ and $\mathcal{S}\cap E_i=F_i'$ under $\mathcal{S}\sim\mu'$ have the same probability for any subsets $F_i,F_i'\subseteq E_i$ of the same size. Therefore, we fix one set of $k_i$ copies of $v_i$ to be $F_i=\{w_{i,j}: j=1,\dots,k_i\}$ for all $i\in[n]$ and obtain

$$\Pr[\mathcal{S} \text{ has } k_i \text{ copies of } v_i,\ \forall i=1,\dots,t] = \prod_{i=1}^{t}\binom{m_i}{k_i}\Pr[\mathcal{S}\cap E_i=F_i,\ \forall i=1,\dots t].$$

Therefore, (38) equals

$$\begin{aligned}&\prod\nolimits_{i=1}^{t}\binom{m_i}{k_i}\Pr[\mathcal{S}\cap E_i=F_i,\ \forall i=1,\dots t]\Big/\prod\nolimits_{i=1}^{t-1}\binom{m_i}{k_i}\Pr[\mathcal{S}\cap E_i=F_i,\ \forall i=1,\dots t-1]\\ &=\binom{m_t}{k_t}\frac{\sum_{|S|=k,S\cap E_i=F_i,\forall=1,\dots t}\det(W_SW_S^\top)}{\sum_{|S|=k,S\cap E_i=F_i,\forall=1,\dots t-1}\det(W_SW_S^\top)}.\end{aligned} \tag{39}$$

To compute the numerator, define $W'$ a matrix of vectors in $W$ restricted to indices $U\backslash(\bigcup_{i=1}^{t}E_i\backslash F_i)$, and $F:=\bigcup_{i=1}^{t}F_i$, then we have

$$\sum_{|S|=k,\,S\subseteq W,\,S\cap E_i=F_i,\,\forall i=1,\dots t}\det(W_SW_S^\top)=\sum_{|S|=k,\,S\subseteq W',\,S\supseteq F}\det(W_S'W_S'^\top). \tag{40}$$

By Lemma 11, the number of arithmetic operations to compute (40) is $O(n(d-d_0+1)d_0^2d^2\log d)=O(nd^4\log d)$ (by applying $d_0=d$). Therefore, because in each step $t=1,2,\dots,n$, we compute (38) at most $k$ times for different values of $k_t$, the total number of arithmetic operations for sampling algorithm $\mathcal{A}$ is $O(n^2d^4k\log d)$. □

**Remark 2.** Although Theorem 15 and Observation 1 imply that randomized rounding for $A$-optimal design with repetition takes $O((k+d^2)^2d^4k\log d)$ number of arithmetic operations, this does not take into account the size of numbers used in the computation which may scale with input $\epsilon$. It is not hard to see that the sizes of coefficients $f(t_1,t_2,t_3)$ in Lemma 10, of the number $\binom{m-|F|-d_0+r}{k-|F|-d_0+r}$ in the proof of Lemma 11 and of $\binom{m_t}{k_t}$ in (39) scale linearly with $O(k\log(m))$, where $m=\sum_{i=1}^{n}m_i$. As we apply $m\le qk=(2n)/\epsilon$ in the proof of Theorem 15, the runtime of randomized rounding for $A$-optimal design with repetition, after taking into account the size of numbers in the computation, has an extra factor of $k\log(n/\epsilon)$ and becomes $O((k+d^2)^2d^4k^2\log d\log(k+d^2/\epsilon)))$.

### 6.4. Efficient Deterministic Implementation of $k/(k-d+1)$-Approximation Algorithm with Repetitions

We show a *deterministic* implementation of proportional volume sampling used for the $k/(k-d+1)$-approximation algorithm with repetitions. In particular, we derandomized the efficient implementation of Steps 4 and 5 of Algorithm 2 and show that the running time of deterministic version is the same as that of the randomized one.

**Theorem 16.** *Given inputs $n, d, k, \epsilon, \mathbf{x} \in \mathbb{R}^n_+$ with $\Sigma_{i=1}^n x_i = k$, and vectors $v_1, \ldots, v_n$ to Algorithm 2, we define $q, U, W$ as in Algorithm 2. Then, there exists 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 $\mu'$ is a distribution over all subsets $S \subseteq U$ of size $k$ defined by $\mu'(S) \propto \det\left(W_S W_S^\top\right)$ for each set $S \subseteq U$ of size $k$. Moreover, $\mathcal{A}'$ runs in $O(n^2 d^4 k \log d)$ number of arithmetic operations.*

Again, together with Observation 1 and Remark 2, Theorem 16 implies that the $k/(k-d+1)$-approximation algorithm for $A$-optimal design with repetitions can be implemented deterministically in $O((k+d^2)^2 d^4 k \log d)$ number of arithmetic operations and, after taking into account the size of numbers in the computation, in $O\left((k+d^2)^2 d^4 k^2 \log d \log\left((k+d^2)/\epsilon\right)\right)$ time.

**Proof.** We can define the deterministic algorithm $\mathcal{A}'$ by deciding, at each step $t = 1, \ldots, n$, how many copies of $\boldsymbol{v}_i$ are to be included in $S^* \subseteq U$. The problem then reduces to efficiently computing

$$X(k_1, \ldots, k_t) := \mathop{\mathbb{E}}_{\mu'}\left[\operatorname{tr}\left(W_{\mathcal{S}} W_{\mathcal{S}}^\top\right)^{-1} \,\middle|\, \mathcal{S} \text{ has } k_i \text{ copies of } \boldsymbol{v}_i,\ \forall i = 1, \ldots, t-1, t\right], \tag{41}$$

where $k_1, \ldots, k_{t-1}$ is already decided by previously steps of the algorithm, and now we compute (41) for each $k_t = 0, 1, \ldots, k - \Sigma_{i=1}^{t-1} k_i$. $\mathcal{A}'$ then chooses value of $k_t$ that maximizes (41) to complete step $t$.

Recall the definitions from proof of Theorem 15 that $F_i, E_i$ are the sets of fixed $k_i$ copies and all copies of $\boldsymbol{v}_i$, respectively, $W'$ is the matrix of vectors in $W$ restricted to indices $U \backslash (\bigcup_{i-1}^t E_i \backslash F_i)$, and $F := \bigcup_{i=1}^t F_i$. Consider that

$$\begin{aligned}
X(k_1, \ldots, k_t) &= \sum_{\substack{S \subseteq U;|S|=k;\\ |S\cap E_i|=k_i, \forall i=1,\ldots,t}} \Pr_{\mu'}[\mathcal{S} = S | \mathcal{S} \text{ has } k_i \text{ copies of } \boldsymbol{v}_i,\ \forall i = 1, \ldots, t] \operatorname{tr}\left[\left(W_S W_S^\top\right)^{-1}\right] \\
&= \sum_{\substack{S \subseteq U;|S|=k;\\ |S\cap E_i|=k_i, \forall i=1,\ldots,t}} \frac{\det\left(W_S W_S^\top\right)}{\sum_{S' \subseteq U;|S'|=k;|S'\cap E_i|=k_i,\forall i=1,\ldots,t} \det\left(W_{S'} W_{S'}^\top\right)} \operatorname{tr}\left[\left(W_S W_S^\top\right)^{-1}\right] \\
&= \frac{\sum_{S \subseteq U;|S|=k;|S\cap E_i|=k_i,\forall i=1,\ldots,t} E_{d-1}\left(W_S W_S^\top\right)}{\sum_{S \subseteq U;|S|=k;|S\cap E_i|=k_i,\forall i=1,\ldots,t} \det\left(W_S W_S^\top\right)} \\
&= \frac{\prod_{i=1}^t \binom{m_i}{k_i} \sum_{S \subseteq U;|S|=k;S\supseteq F} E_{d-1}\left(W'_S W_S'^\top\right)}{\prod_{i=1}^t \binom{m_i}{k_i} \sum_{S \subseteq U;|S|=k;S\supseteq F} \det\left(W'_S W_S'^\top\right)} \\
&= \frac{\sum_{S \subseteq U;|S|=k;S\supseteq F} E_{d-1}\left(W'_S W_S'^\top\right)}{\sum_{S \subseteq U;|S|=k;S\supseteq F} \det\left(W'_S W_S'^\top\right)}.
\end{aligned}$$

By Lemma 11, we can compute the numerator and denominator in $O\left(n(d-d_0+1)d_0^2 d^2 \log d\right) = O\left(nd^4 \log d\right)$ (by applying $d_0 = d-1, d$) number of arithmetic operations. Therefore, because in each step $t = 1, 2, \ldots, n$, we compute (41) at most $k$ times for different values of $k_t$, the total number of arithmetic operations for sampling algorithm $\mathcal{A}$ is $O\left(n^2 d^4 k \log d\right)$.

### 6.5. Efficient Implementations for the Generalized Ratio Objective

In Sections 6.1 and 6.2, we obtain efficient randomized and deterministic implementations of proportional volume sampling with measure $\mu$ when $\mu$ 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 Sections 6.3 and 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 $O(n^2 d^4 k \log d)$ number of arithmetic operations. In this section, we show that the results from Sections 6.1 to 6.4 generalize to proportional $l$-volume sampling for generalized ratio problem.

**Theorem 17.** *Let $n, d, k$ be positive integers, $\lambda \in \mathbb{R}^n_+$, $\mathcal{U} \in \{\mathcal{U}_k, \mathcal{U}_{\leq k}\}$, $V = [v_1, \ldots, v_n] \in \mathbb{R}^{d\times n}$, and $0 \leq l' < l \leq d$ be a pair of integers. Let $\mu'$ be the $l$-proportional volume sampling distribution over $\mathcal{U}$ with hard-core measure $\mu$ of parameter $\lambda$, that is $\mu'(S) \propto \lambda^S E_l\left(V_S V_S^\top\right)$ for all $S \in \mathcal{U}$. There are*

- *an implementation to sample from* $\mu'$ *that runs in* $O(n^4 lk^2 \log(lk))$ *number of arithmetic operations, and*
- *a deterministic algorithm that outputs a set* $S^* \in \mathcal{U}$ *of size k such that*

$$\left(\frac{E_{l'}(V_{S^*}V_{S^*}^\top)}{E_l(V_{S^*}V_{S^*}^\top)}\right)^{\frac{1}{l-l'}} \geq \mathop{\mathbb{E}}_{S\sim\mu'}\left[\left(\frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}\right)^{\frac{1}{l-l'}}\right] \tag{42}$$

*that runs in* $O(n^4 lk^2 \log(lk))$ *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* $\mu'$ *be the l-proportional volume sampling over all subsets* $S \subseteq U$ *of size k with measure* $\mu$ *that is uniform, that is,* $\mu'(S) \propto E_l(W_S W_S^\top)$ *for all* $S \subseteq U, |S| = k$*. There are*

- *an implementation to sample from* $\mu'$ *that runs in* $O(n^2(d-l+1)l^2 d^2 k \log d)$ *number of arithmetic operations, and*
- *a deterministic algorithm that outputs a set* $S^* \in \mathcal{U}$ *of size k such that*

$$\left(\frac{E_{l'}(W_{S^*}W_{S^*}^\top)}{E_l(W_{S^*}W_{S^*}^\top)}\right)^{\frac{1}{l-l'}} \geq \mathop{\mathbb{E}}_{S\sim\mu'}\left[\left(\frac{E_{l'}(W_S W_S^\top)}{E_l(W_S W_S^\top)}\right)^{\frac{1}{l-l'}}\right] \tag{43}$$

*that runs in* $O(n^2((d-l'+1)l'^2 + (d-l+1)l^2)d^2 k \log d)$ *number of arithmetic operations.*

As in Observation 1, we can replace $n = k + d^2$ in all running times in Theorem 17 so that running times of all variants of proportional volume sampling are independent of $n$. We also note, as in Remark 2, that running times of $l$-proportional volume sampling over $m$ vectors with $n$ distinct vectors have 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 $[m]$.

**Proof.** By the convexity of $f(z) = z^{l-l'}$ over positive reals $z$, we have $\mathbb{E}[X] \geq (\mathbb{E}[X^{1/(l-l')}])^{l-l'}$ for a nonnegative random variable $X$. Therefore, to show (42), it is sufficient to show that

$$\frac{E_{l'}(V_{S^*}V_{S^*}^\top)}{E_l(V_{S^*}V_{S^*}^\top)} \geq \mathop{\mathbb{E}}_{S\sim\mu'} \frac{E_{l'}(V_S V_S^\top)}{E_l(V_S V_S^\top)}. \tag{44}$$

That is, it is enough to derandomized with respect to the objective $(E_{l'}(V_S V_S^\top))/(E_l(V_S V_S^\top))$, and the same is true for showing (43). Hence, we choose to calculate the conditional expectations with respect to this objective.

We follow the exact same calculation for $l$-proportional volume sampling for generalized ratio objective as original proofs of efficient implementations of all four algorithms in $A$-optimal objective. We observe that those proofs in $A$-optimal objective ultimately rely on the ability to, given disjoint $I, J \subseteq [n]$ (or in the other case, $[m]$), efficiently compute

$$\sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\phi} \lambda^S \sum_{|R|=d, R\subseteq S} \det(V_R V_R^\top) \text{ and } \sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\phi} \lambda^S \sum_{|T|=d-1, T\subseteq S} \det(V_T^\top V_T)$$

(or in the other case, replace $V$ with $W$ and $\lambda^S = 1$ for all $S$). The proofs for generalized ratio objective follow the same line as those proofs of four algorithms, except that we instead need to efficiently compute

$$\sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\phi} \lambda^S \sum_{|T|=l, R\subseteq S} \det(V_T^\top V_T) \text{ and } \sum_{S\in\mathcal{U}, I\subseteq S, J\cap S=\phi} \lambda^S \sum_{|T'|=l', T'\subseteq S} \det(V_{T'}^\top V_{T'})$$

(note the change of $R, T$ of size $d, d-1$ to $T, T'$ of size $l, l'$, respectively). However, the computations can indeed be done efficiently by using different $d_0 = l', l$ instead of $d_0 = d-1, d$ when applying Lemmas 9–11 in the proofs and then following a similar calculation. The proofs for running times are identical. □

## 7. Integrality Gaps

In this section, we show the integrality gap of the natural convex relaxations of $A$- and $E$-optimal problems.

### 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 $v_1, \ldots, v_n \in \mathbb{R}^d$, and our goal is to select a set $S \subseteq [n]$ of size $k$, but this time we minimize the objective $\|(\sum_{i\in S} v_i v_i^\top)^{-1}\|$, where $\|\cdot\|$ is the operator norm, that is, 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 $i$th smallest eigenvalue of $M$.

This problem also has a natural convex relaxation, analogous to the one we use for the $A$ objective:

$$\max \lambda_1\left(\sum_{i=1}^{n} x_i v_i v_i^\top\right), \tag{45}$$

s.t.

$$\sum_{i=1}^{n} x_i = k, \tag{46}$$

$$0 \le x_i \le 1 \quad \forall i \in [n]. \tag{47}$$

We prove the following integrality gap result for (45)–(47).

**Theorem 18.** *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* (45)–(47) *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 $k = \Omega(d/\epsilon + (\log(1/\epsilon))/\epsilon^2)$. Theorem 18 shows that such a result is impossible for the $E$-objective, for which the results in Allen-Zhu et al. [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 $x = (k/\binom{d+1}{2})1$, that is, each coordinate of $x$ is $k/\binom{d+1}{2}$. Then $M(x) = \sum_{i=1}^{d+1}\sum_{j=i+1}^{d+1} x_{ij}v_{ij}v_{ij}^\top = (2k)/(dI_d)$, and the objective value of the solution is $(2k)/d$.

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 $(1+\epsilon)$CP, then the smallest eigenvalue of $M(S) = \sum_{(i,j)\in S} v_{ij}v_{ij}^\top$ is at least $(2k)/(d(1+\epsilon)) \ge (1-\epsilon)(2k)/(d)$. Because $M(S) = B^\top L_G B$, this means that the second smallest eigenvalue of $L_G$ is at least $(1-\epsilon)(2k)/(d)$. The average degree $\Delta$ of $G$ is $(2k)/(d+1)$. Therefore, we have a graph $G$ on $d+1$ vertices with average degree $\Delta$ for which the second smallest eigenvalue of its Laplacian is at least $(1-\epsilon)(1-(1)/(d+1))\Delta \ge (1-2\epsilon)\Delta$, where the inequality holds for $d$ large enough. The classical Alon-Boppana bound (Alon [3], Nilli [40]) shows that, up to lower order terms, the second smallest eigenvalue of the Laplacian of a $\Delta$-regular graph is at most $\Delta - 2\sqrt{\Delta}$. If our graph $G$ were regular, this would imply that $(2k)/(d+1) = \Delta \ge 1/\epsilon^2$. In order to prove Theorem 18, 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 nonregular graphs: see the recent preprint (Srivastava and Trevisan [47]) 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 19.** *Let $G = (V,E)$ be a graph with average degree $\Delta = (2|E|)/|V|$, and let $L_G$ be its unnormalized Laplacian matrix. Then, as long as $\Delta$ is large enough, and $|V|$ is large enough with respect to $\Delta$,*

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

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

**Proof.** By the variational characterization of eigenvalues, we need to find a unit vector $x$, orthogonal to one, such that $x^\top L_G x \le \Delta - c\sqrt{\Delta}$. Our goal is to use a vector $x$ similar to the one used in the lower bound on the number of edges of a spectral sparsifier in Batson et al. [6]. However, to apply this strategy, we need to make sure that $G$ has a low degree vertex most of whose neighbors have low degree. This requires most of the work in the proof.

So that we do not have to worry about making our "test vector" orthogonal to one, observe that

$$\lambda_2(L_G) = \min_{x\in\mathbb{R}^V} \frac{x^\top L_G x}{x^\top x - (1^\top x)^2/|V|}. \tag{48}$$

Indeed, the denominator equals $\boldsymbol{y}^\top \boldsymbol{y}$ for the projection $\boldsymbol{y}$ of $x$ orthogonal to one, and the numerator is equal to $\boldsymbol{y}^\top L_G \boldsymbol{y}$. Here, and in the remainder of the proof, we work in $\mathbb{R}^V$, the space of $|V|$-dimensional real vectors indexed by $V$, and think of $L_G$ as being indexed by $V$ as well.

Observe that if $G$ has a vertex $u$ of degree $\Delta(u)$ at most $\Delta - 1/10\sqrt{\Delta}$, we are done. In that case we can pick $x \in \mathbb{R}^V$ such that $x_u = 1$ and $x_v = 0$ for all $v \neq u$. Then

$$\frac{x^\top L_G x}{x^\top x - (1^\top x)^2/n} = \frac{\sum_{(u,v)\in E}(x_u - x_v)^2}{1 - \frac{1}{|V|}} \leq \frac{\Delta - \frac{1}{10}\sqrt{\Delta}}{1 - \frac{1}{|V|}},$$

which, by (48), implies the theorem for all large enough $|V|$. Therefore, for the rest of the proof, we will assume that $\Delta(u) \geq \Delta - 1/10\sqrt{\Delta}$ for all $u \in V$.

Define $T = \{u \in V : \Delta(u) \geq \Delta + 1/2\sqrt{\Delta}\}$ to be the set of large-degree vertices and let $S = V\backslash T$. Observe that

$$\begin{aligned} |V|\Delta &\geq |T|\left(\Delta + \frac{1}{2}\sqrt{\Delta}\right) + |S|\left(\delta - \frac{1}{10}\sqrt{\Delta}\right) \\ &= |V|\Delta + \left(\frac{1}{2}|T| - \frac{1}{10}|S|\right)\sqrt{\Delta}. \end{aligned}$$

Therefore, $|S| \geq 5|T|$, and, because $T$ and $S$ partition $V$, we have $|S| \geq 5/6|V|$.

Define $\alpha = \min\,\{|\{v \sim u : v \in T\}|/(\Delta - 1/10\sqrt{\Delta}) : u \in S\}$, where $v \sim u$ means that $v$ is a neighbor of $u$. We need to find a vertex in $S$ such that only a small fraction of its neighbors is in $T$; that is, we need an upper bound on $\alpha$. To show such an upper bound, let us define $E(S,T)$ to be the set of edges between $S$ and $T$; then

$$\frac{1}{2}\Delta|V| = |E| \geq |E(S,T)| \geq |S|\alpha\left(\Delta - \frac{1}{10}\sqrt{\Delta}\right) \geq \frac{5}{6}|V|\alpha\Delta\left(1 - \frac{1}{10\sqrt{\Delta}}\right).$$

Therefore, $\alpha \leq 3/5(1 - 1/10\sqrt{\Delta})^{-1}$.

Let $u \in S$ be a vertex with at most $\alpha\Delta - \alpha/10\sqrt{\Delta}$ neighbors in $T$, and let $\delta = |\{v \sim u : v \in S\}|$. By the choice of $u$,

$$\delta \geq \Delta(u) - \alpha\Delta + \frac{\alpha}{10}\sqrt{\Delta} \geq (1-\alpha)\,\Delta\left(1 - \frac{1}{10\sqrt{\Delta}}\right).$$

Assume that $\Delta$ is large enough so that $\left(1 - 1/(10\sqrt{\Delta})\right) \geq 16/25$. Then, $\delta \geq 16/25(1-\alpha)\Delta$.

We are now ready to define our test vector $x$ and complete the proof. Let $x_u = 1$, $x_v = 1/\sqrt{\delta}$ for any neighbor $v$ of $u$, which is in $S$, and $x_w = 0$ for any $\boldsymbol{w}$ which is in $T$ or is not a neighbor of $u$. We calculate

$$\begin{aligned} x^\top L_G x &= |\{v \sim u : v \in S\}|\left(1 - \tfrac{1}{\sqrt{\delta}}\right)^2 + |\{v \sim u : v \in T\}| + \sum_{v\sim u,\, v\in S}\ \sum_{\boldsymbol{w}\sim v,\, \boldsymbol{w}\neq u} \frac{1}{\delta} \\ &\leq \delta\left(1 - \tfrac{1}{\sqrt{\delta}}\right)^2 + \Delta(u) - \delta + \Delta + \frac{1}{2}\sqrt{\Delta} - 1, \end{aligned}$$

where we used the fact for any $v \in S$, $\Delta(v) \leq \Delta + 1/2\sqrt{\Delta}$ by definition of $S$. The right-hand side simplifies to

$$\Delta(u) - 2\sqrt{\delta} + \Delta + \frac{1}{2}\sqrt{\Delta} \leq 2\Delta - \left(\frac{8}{5}\sqrt{(1-\alpha)} - \frac{1}{2}\right)\sqrt{\Delta}.$$

Because $\alpha \leq 3/5(1 - 1/10\sqrt{\Delta})^{-1}$, $8/5\sqrt{(1-\alpha)} - 1/2 \geq 1/2$ for all large enough $\Delta$, and by (48), we have

$$\lambda_2(G) \leq \frac{x^\top L_G x}{x^\top x - (1^\top x)^2} \leq \frac{2\Delta - \frac{1}{2}\sqrt{\Delta}}{2\left(1 - \frac{1+\sqrt{\Delta}}{2|V|}\right)} = \left(\Delta - \frac{1}{4}\sqrt{\Delta}\right)\left(1 - \frac{1+\sqrt{\Delta}}{2|V|}\right)^{-1}.$$

The theorem now follows as long as $|V| \geq C\Delta$ for a sufficiently large constant $C$. □

To finish the proof of Theorem 18, 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 $\Delta = (2k)/(d+1)$ satisfies

$\lambda_2(L_G) \geq (1-2\epsilon)\Delta$. By Theorem 19, $\lambda_2(L_G) \leq \Delta - c\sqrt{\Delta}$ for large enough $d$ with respect to $\Delta$. We have $\Delta \geq c^2/(4\epsilon^2)$, and rearranging the terms proves the theorem.

The proof of Theorem 19 does not require the graph $G$ to be simple; that is, parallel edges are allowed. This means that the integrality gap in Theorem 18 holds for the $E$-optimal design problem with repetitions as well.

### 7.2. Integrality Gap for *A*-Optimality

The integrality gap of $A$-optimal design problem can be stated as follows.

**Theorem 20.** *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 $k/(k-d+1)$. The theorem statement applies to both with and without repetitions.

**Proof.** The instance $V = [v_1, \ldots, v_n]$ will be the same with or without repetitions. For each $1 \leq i \leq d$, let $e_i$ denote the unit vector in direction of axis $i$. Let $v_i = N \cdot e_i$ for each $i = 1, \ldots, d-1$, where $N > 0$ is a constant to be chosen later and $v_d = e_d$. Set the rest $v_i, i > d$ to be at least $k$ copies of each of these $v_i$ for $i \leq d$, as we can make $n$ as big as needed. Hence, we may assume that we are allowed to pick only $v_i, i \leq d$, but with repetitions.

The fractional optimal solution which can be calculated by Lagrange's multiplier technique is $y^* = (\delta_0, \delta_0, \ldots, \delta_0, k-(d-1)\delta_0)$ for small $\delta_0 = k/(\sqrt{N}+d-1)$. The optimal integral solution is $x^* = (1, 1, \ldots, 1, k-d+1)$. Therefore, as $N \to \infty$, we have $\mathsf{CP} = (d-1)/(\delta_0 N) + 1/(k-(d-1)\delta_0) \to 1/k$, and $\mathsf{OPT} = (d-1)/N + 1/(k-d) + 1 \to 1/(k-d+1)$. Hence,

$$\frac{\mathsf{OPT}}{\mathsf{CP}} \to \frac{k}{k-d+1},$$

proving the theorem. □

## 8. Hardness of Approximation

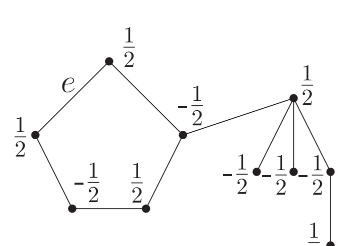
In this section, we show that the $A$-optimal design problem is NP-hard to approximate within a fixed constant when $k = d$. To the best of our knowledge, no hardness results for this problem were previously known. Our reduction is inspired by the hardness of approximation for $D$-optimal design proved in Di Summa et al. [22]. The hard problem we reduce from is an approximation version of Partition into Triangles.

Before we prove our main hardness result, Theorem 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 one, and all its other coordinates are equal to zero. 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 12.** *Let $V$ be the vertex by edge incidence matrix of a graph $G = ([d], E)$, as described previously. 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 $t = d/3$, $\mathrm{tr}((V_S V_S^\top)^{-1}) = 3d/4$; and*
- *for any $t$, $\mathrm{tr}((V_S V_S^\top)^{-1}) \geq d - 3t/4$.*

**Figure 1.** Values of the coordinates of $u_e$ for $e \in C_\ell$.



**Proof.** Let $H_1, \dots, H_c$ be the connected components of $H$. First we claim that the invertibility of $V_S$ implies that none of the $H_\ell$ is bipartite. Indeed, if some $H_\ell$ were bipartite, with bipartition $L \cup R$, then the nonzero vector $x$ defined by

$$x_i = \begin{cases} 1 & i \in L \\ -1 & i \in R \\ 0 & \text{otherwise,} \end{cases}$$

is in the kernel of $V_S$. In particular, each $H_\ell$ must have at least as many edges as vertices. Because the number of edges of $H$ equals the number of vertices, it follows that *every* connected component $H_\ell$ must have exactly as many edges as vertices, too. In particular, this means that every $H_\ell$ is the disjoint union of a spanning tree and an edge, and the edge creates an odd-length cycle.

Let us explicitly describe the inverse $V_S^{-1}$. For each $e \in S$ we need to give a vector $u_e \in \mathbb{R}^d$ so that $u_e^\top v_e = 1$ and $u_e^\top v_f = 0$ for every $f \in S, f \neq e$. Then $U^\top = V_S^{-1}$, where $U = (u_e)_{e \in S}$ is the matrix whose columns are the $u_e$ vectors. Let $H_\ell$ be, as above, one of the connected components of $H$. We will define the vectors $u_e$ for all edges $e$ in $H_\ell$; the vectors for edges in the other connected components are defined analogously. Let $C_\ell$ be the unique cycle of $H_\ell$. Recall that $C_\ell$ must be an odd cycle. For any $e = (i,j)$ in $C_\ell$, we set the $i$th and the $j$th coordinate of $u_e$ to $\frac{1}{2}$. Let $T$ be the spanning tree of $H_\ell$ derived from removing the edge $e$. We set the coordinates of $u_e$ corresponding to vertices of $H_\ell$ other than $i$ and $j$ to either $-1/2$ or $+1/2$, so that the vertices of any edge of $T$ receive values with opposite signs. This can be done by setting the coordinate of $u_e$ corresponding to vertex $k$ in $H_\ell$ to $1/2(-1)^{\delta_T(i,k)}$, where $\delta_T(i,k)$ is the distance in $T$ between $i$ and $k$. Because $C_\ell$ is an odd cycle, $\delta_T(i,j)$ is even, and this assignment is consistent with the values we already determined for $i$ and $j$. Finally, the coordinates of $u_e$ that do not correspond to vertices of $H_\ell$ are set to zsero. Figure 1 provides an example. It is easy to verify that $u_e^\top v_e = 1$ and $u_e^\top v_f = 0$ for any edge $f \neq e$. Notice that $\|u_e\|_2^2 = d_\ell/4$, where $d_\ell$ is the number of vertices (and also the number of edges) of $H_\ell$.

It remains to describe $u_e$ when $e = (i,j) \notin C_\ell$. Let $T$ be the tree derived from $H_\ell$ by contracting $C_\ell$ to a vertex $r$, and set $r$ as the root of $T$. Without loss of generality, assume that $j$ is the endpoint of $e$, which is further from $r$ in $T$. We set the $j$th coordinate of $u_e$ equal to one. We set the coordinates of $u_e$ corresponding to vertices in the subtree of $T$ below $j$ to either $-1$ or $+1$ so that the signs alternate down each path from $j$ to a leaf of $T$ below $j$. This can be achieved by setting the coordinate of $u_e$ corresponding to vertex $k$ to $(-1)^{\delta_T(j,k)}$, where $\delta_T(j,k)$ is the distance between $j$ and $k$ in $T$. All other coordinates of $u_e$ are set equal to zero. Figure 2 provides an example. Notice that $\|u_e\|_2^2 \geq 1$ (and in fact equals the number of nodes in the subtree of $T$ below the node $j$).

We are now ready to finish the proof. Clearly if $[d]$ can be partitioned into $t = d/3$ disjoint triangles, and the union of their edges is $S$, then

$$\operatorname{tr}\left(\left(V_S {V_S}^T\right)^{-1}\right) = \operatorname{tr}\left(U\,U^\top\right) = \sum_{e \in S} \|u_e\|_2^2 = \frac{3|S|}{4} = \frac{3d}{4}.$$

In the general case, we have

$$\begin{aligned} \operatorname{tr}\left(\left(V_S {V_S}^T\right)^{-1}\right) = \operatorname{tr}\left(U\,U^\top\right) &= \sum_{e \in S} \|u_e\|_2^2 \\ &\geq \sum_{\ell=1}^{c} \frac{|C_\ell| \,.\, d_\ell}{4} + d_\ell - |C_\ell| \\ &\geq \frac{9t}{4} + d - 3t = d - \frac{3t}{4}, \end{aligned}$$

where $|C_\ell|$ is the length of $C_\ell$, and $d_\ell$ is the number of edges (and also the number of vertices) in $H_\ell$. The final inequality follows because any connected component $H_\ell$ that is not a triangle contributes at least $d_\ell$ to the sum. □

Recall that in the Partition into Triangles problem, we are given a graph $G = (W, E)$ and need to decide whether $W$ can be partitioned into $|W|/3$ vertex-disjoint triangles. This problem is NP-complete (Garey and Johnson [24] present a proof in chapter 3 and cite personal communication with Schaeffer), and this, together with Lemma 12, suffices to show that the $A$-optimal design problem is NP-hard when $k = d$. To prove hardness of approximation,

**Figure 2.** Values of the coordinates of $u_e$ for $e \notin C_\ell$.

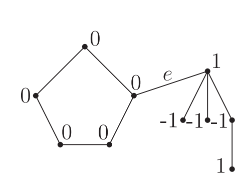
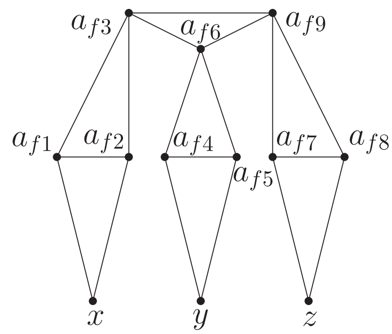


**Figure 3.** Subgraph with edges $E_f$ for the triple $f = \{x, y, z\}$ (adapted from Garey and Johnson [24]).




we prove hardness of a gap version of Partition into Triangles. In fact, we just observe that the reduction from three-dimensional matching to Partition into Triangles in Garey and Johnson [24] and known hardness of approximation of three-dimensional matching give the result we need.

**Lemma 13.** *Given a graph $G = (W, E)$, it is* NP*-hard to distinguish the two cases:*
1. *$W$ can be partitioned into $|W|/3$ vertex-disjoint triangles;and*
2. *every set of vertex-disjoint triangles in $G$ has cardinality at most $\alpha|W|/3$,*
*where $\alpha \in (0, 1)$ is an absolute constant.*

To prove Lemma 13, we use a theorem of Petrank.

**Theorem 21** (Petrank [42]). *Given a collection of triples $F \subseteq X \times Y \times Z$, where $X$, $Y$, and $Z$ are three disjoint sets of size $m$ each, and each element of $X \cup Y \cup Z$ appears in at most three triples of $F$, it is* NP*-hard to distinguish the two cases:*
1. *there is a set of disjoint triples $M \subseteq F$ of cardinality $m$; and*
2. *every set of disjoint triples $M \subseteq F$ has cardinality at most $\beta m$,*
*where $\beta \in (0, 1)$ is an absolute constant.*

We note that Petrank gives a slightly different version of the problem, in which the set $M$ is allowed to have intersecting triples, and the goal is to maximize the number of elements $X \cup Y \cup Z$ that are covered exactly once. Petrank shows that it is hard to distinguish between the cases when every element is covered exactly once, and the case when at most $3\beta m$ elements are covered exactly once. It is immediate that this also implies Theorem 21.

**Proof of Lemma 13.** We will show that the reduction in Garey and Johnson [24] from three-dimensional matching to Partition into Triangles is approximation preserving. This follows in a straightforward way from the argument in Garey and Johnson [24], but we repeat the reduction and its analysis for the sake of completeness.

Given $F \subseteq X \cup Y \cup Z$ such that each element of $X \cup Y \cup Z$ appears in at most three triples of $F$, we construct a graph $G = (W, E)$ on the vertices $X \cup Y \cup Z$ and $9|F|$ additional vertices: $a_{f1}, \ldots a_{f9}$ for each $f \in F$. For each triple $f \in F$, we include in $E$ the edges $E_f$ shown in Figure 3. The subgraphs spanned by the sets $E_f$, $E_g$ for two different triples $f$ and $g$ are edge-disjoint, and the only vertices they share are in $X \cup Y \cup Z$.

First, we show that, if $F$ has a matching $M$ covering all elements of $X \cup Y \cup Z$, then $G$ can be partitioned into vertex-disjoint triangles. Indeed, for each $f = \{x, y, z\} \in M$ we can take the triangles $\{x, a_{f1}, a_{f2}\}$, $\{y, a_{f4}, a_{f5}\}$, $\{z, a_{f7}, a_{f8}\}$, and $\{a_{f3}, a_{f6}, a_{f9}\}$. For each $f \notin M$, we can take the triangles $\{a_{f1}, a_{f2}, a_{f3}\}$, $\{a_{f4}, a_{f5}, a_{f6}\}$, and $\{a_{f7}, a_{f8}, a_{f9}\}$.

In the other direction, assume there exists a set $T$ of at least $\alpha(|W|/3)$ vertex disjoint triangles in $G$, for a value of $\alpha$ to be chosen shortly. We need to show that $F$ contains a matching of at least $\beta m$ triples. To this end, we construct a set $M$ that contains all triples $f$, for each $E_f$ that contains at least four triangles of $T$. Notice that the only way to pick three vertex disjoint triangles from $E_f$ is to include the lower three triangles (see Figure 3), so any two triples $f$ and $g$ in $M$ must be disjoint. The cardinality of $T$ is at most $4|M| + 3(|F| - |M|) = |M| + 3|F|$.

Therefore,

$$|M| + 3|F| \geq \alpha \frac{|W|}{3} = \alpha(m + 3|F|),$$

and we have $|M| \geq \alpha m - (1 - \alpha)3|F| \geq (10\alpha - 9)m$, where we used the fact that $|F| \leq 3m$ because each element of $X$ appears in at most three triples of $F$. Then, if $\alpha \geq (9 + \beta)/10$ we have $|M| \geq \beta m$. This finishes the proof of the lemma. □

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

**Proof of Theorem 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 12, $\operatorname{tr}((V_S V_S^\top)^{-1}) = 3d/3$.

Next, consider the case in which every set of vertex-disjoint triangles in $G$ has cardinality at most $\alpha(d/3)$. 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 $\alpha(d/3)$ connected components that are triangles, because any two triangles in distinct connected components are necessarily vertex-disjoint. Therefore, by Lemma 12, $\operatorname{tr}((V_S V_S^\top)^{-1}) \ge ((4 - \alpha)d)/4$.

It follows that a $c$-approximation algorithm for the $A$-optimal design problem, for any $c < (4 - \alpha)/3$, can be used to distinguish between the two cases of Lemma 13, and therefore, the $A$-optimal design problem is NP-hard to $c$-approximate. □

## Acknowledgments

The authors thank Bobby Kleinberg for a useful discussion.

## References

[1] Allen-Zhu Z, Li Y, Singh A, Wang Y (2017) Near-optimal design of experiments via regret minimization. Precup D, Teh YW, eds. *Proc. 34th Internat. Conf. on Machine Learning*, vol. 70, 126–135.
[2] Allen-Zhu Z, Li Y, Singh A, Wang Y (2017) Near-optimal discrete optimization for experimental design: A regret minimization approach. Preprint, submitted November 14, https://arxiv.org/abs/1711.05174.
[3] Alon N (1986) Eigenvalues and expanders. *Combinatorica* 6(2):83–96.
[4] Anari N, Gharan SO (2017) A generalization of permanent inequalities and applications in counting and optimization. Hatami H, McKenzie P, King V, eds. *Proc. 49th Annual ACM SIGACT Sympos. Theory Comput.*, 384–396
[5] Avron H, Boutsidis C (2013) Faster subset selection for matrices and applications. *SIAM J. Matrix Analysis Appl.* 34(4):1464–1499.
[6] Batson J, Spielman DA, Srivastava N (2012) Twice-ramanujan sparsifiers. *SIAM J. Comput.* 41(6):1704–1721.
[7] Bourgain J, Tzafriri L (1987) Invertibility of large submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.* 57(2):137–224.
[8] Boutsidis C, Magdon-Ismail M (2013) Deterministic feature selection for k-means clustering. *IEEE Trans. Inform. Theory* 59(9):6099–6110.
[9] Boutsidis C, Drineas P, Magdon-Ismail M (2011) Sparse features for pca-like linear regression. Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira FCN, Weinberger KQ, eds. *Advances in Neural Information Processing Systems*, 2285–2293.
[10] Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).
[11] Bullen P, Marcus M (1961) Symmetric means and matrix inequalities. *Proc. Amer. Math. Soc.*, 285–290.
[12] Chernoff H (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* 23(4):493–507.
[13] Çivril A, Magdon-Ismail M (2009) On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Comput. Sci.* 410(47-49):4801–4811.
[14] Davis C (1957) All convex invariant functions of hermitian matrices. *Arch. Math. (Basel)* 8:276–278.
[15] de Hoog FR, Mattheij RMM (2007) Subset selection for matrices. *Linear Algebra Its Appl.* 422(2-3):349–359.
[16] de Hoog FR, Mattheij RMM (2011) A note on subset selection for matrices. *Linear Algebra Its Appl.* 434(8):1845–1850.
[17] Dereziński M, Warmuth MK (2017) Subsampling for ridge regression via regularized volume sampling. Preprint, submitted February 23, https://arxiv.org/abs/1710.05110.
[18] Dereziński M, Warmuth MK (2017) Unbiased estimates for linear regression via volume sampling. *NIPS'17: Proc. 31st Internat. Conf. Neural Inform. Processing Systems*, 3087–3096.
[19] Deshpande A, Rademacher L (2010) Efficient volume sampling for row/column subset selection. *51th Annual IEEE Sympos. Foundations Comput. Sci.* (IEEE Computer Society, New York), 329–338.
[20] Deshpande A, Vempala S (2006) Adaptive sampling and fast low-rank matrix approximation. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (Springer, Berlin), 292–303.
[21] Deshpande A, Rademacher L, Vempala S, Wang G (2006) Matrix approximation and projective clustering via volume sampling. *Proc. 17th Annual ACM-SIAM Sympos. Discrete Algorithm* (Society for Industrial and Applied Mathematics, Philadelphia), 1117–1126.
[22] Di Summa M, Eisenbrand F, Faenza Y, Moldenhauer C (2015) On largest volume simplices and sub-determinants. *Proc. 26th Annual ACM-SIAM Sympos. Discrete Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia), 315–323.
[23] Fedorov VV (1972) *Theory of Optimal Experiments* (Elsevier, New York).
[24] Garey MR, Johnson DS (1979) *Computers and Intractability* (W. H. Freeman and Co., San Francisco).
[25] Gharan SO, Saberi A, Singh M (2011) A randomized rounding approach to the traveling salesman problem. *IEEE 52nd Annual Sympos. on Foundations of Comput. Sci.* (IEEE, New York), 550–559.
[26] Guruswami V, Sinop AK (2012) Optimal column-based low-rank matrix reconstruction. *Proc. 23rd Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 1207–1214.
[27] Joshi S, Boyd S (2009) Sensor selection via convex optimization. *IEEE Trans. Signal Processing* 57(2):451–462.
[28] Kahn J (1996) Asymptotics of the chromatic index for multigraphs. *J. Combinatorial Theory Ser. B* 68(2):233–254.
[29] Kahn J (2000) Asymptotics of the list-chromatic index for multigraphs. *Random Structures Algorithms* 17(2):117–156.

[30] Kulesza A, Taskar B (2012) Determinantal point processes for machine learning. *Foundations Trends Machine Learning* 5(2–3):123–286.
[31] Li C, Jegelka S, Sra S (2017) Column subset selection via polynomial time dual volume sampling. Preprint, submitted November 16, https://arxiv.org/abs/1703.02674.
[32] Lin M, Trudinger NS (1993) On some inequalities for elementary symmetric function. *Bull. Australian Math. Soc.* 50(2):317–326.
[33] Marcus M, Lopes L (1957) Inequalities for symmetric functions and Hermitian matrices. *Canadian J. Math.* 9:305–312.
[34] Mariet ZE, Sra S (2017) Elementary symmetric polynomials for optimal experimental design. *31st Conf. Neural Inform. Processing Systems (NIPS 2017), Long Beach, CA,* 2136–2145.
[35] Naor A, Youssef P (2017) *Revisited RI* (Springer International Publishing, Cham, Switzerland).
[36] Nikolov A (2015) Randomized rounding for the largest simplex problem. *Proc. 47th Annual ACM Sympos. Theory of Comput.* (ACM), 861–870.
[37] Nikolov A, Singh M (2016) Maximizing determinants under partition constraints. *Proc. Forty-Eighth Annual ACM Symp. Theory Comput. (STOC)* (Association for Computing Machinery, New York).
[38] Nikolov A, Talwar K (2015) Approximating hereditary discrepancy via small width ellipsoids. Indyk P, ed. *Proc. 26th Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 324–336.
[39] Nikolov A, Talwar K, Zhang L (2016) The geometry of differential privacy: The small database and approximate cases. *SIAM J. Comput.* 45(2):575–616.
[40] Nilli A (1991) On the second eigenvalue of a graph. *Discrete Math.* 91(2): 207–210.
[41] Pan VY (1994) Simple multivariate polynomial multiplication. *J. Symbolic Comput*. 18(3):183–186.
[42] Petrank E (1994) The hardness of approximation: Gap location. *Comput. Complexity* 4(2):133–157.
[43] Pukelsheim F (2006) *Optimal Design of Experiments* (SIAM, Philadelphia).
[44] Singh M, Xie W (2018) Approximate positive correlated distributions and approximation algorithms for d-optimal design. *Proc. Twenty-Ninth Annual ACM-SIAM Sympos. Discrete Algorithms* (Society for Industrial and Applied Mathematics), 2240–2255.
[45] Spielman DA, Srivastava N (2011) Graph sparsification by effective resistances. *SIAM J. Comput.* 40(6):1913–1926.
[46] Spielman DA, Srivastava N (2012) An elementary proof of the restricted invertibility theorem. *Israel J. Math.* 190:83–91.
[47] Srivastava N, Trevisan L (2017) An alon-boppana type bound for weighted graphs and lowerbounds for spectral sparsification. *Proc. Twenty-Ninth Annual ACM-SIAM Sympos. Discrete Algorithms* (Society for Industrial and Applied Mathematics), 1306–1315.
[48] Straszak D, Vishnoi NK (2017) Real stable polynomials and matroids: Optimization and counting. Hatami H, McKenzie P, King V, eds. *Proc. 49th Annual ACM SIGACT Sympos. Theory of Comput*. (ACM), 370–383.
[49] Vershynin R (2001) John's decompositions: Selecting a large part. *Israel J. Math.* 122(1):253–277.
[50] Wang Y, Yu AW, Singh A (2016) On computationally tractable selection of experiments in regression models. Preprint, submitted December 20, https://arxiv.org/abs/1601.02068.
[51] Youssef P (2014) Restricted invertibility and the Banach-Mazur distance to the cube. *Mathematika* 60(1):201–218.