A Local Search Framework for Experimental Design*

Lap Chi Lau[†] Hong Zhou[‡]

Abstract

We present a local search framework to design and analyze both combinatorial algorithms and rounding algorithms for experimental design problems. This framework provides a unifying approach to match and improve all known results in D/A/E-design and to obtain new results in previously unknown settings.

- For combinatorial algorithms, we provide a new analysis of the classical Fedorov's exchange method. We prove that this simple local search algorithm works well as long as there exists an almost optimal solution with good condition number. Moreover, we design a new combinatorial local search algorithm for E-design using the regret minimization framework.
- For rounding algorithms, we provide a unified randomized exchange algorithm to match and improve previous results for D/A/E-design. Furthermore, the algorithm works in the more general setting to approximately satisfy multiple knapsack constraints, which can be used for weighted experimental design and for incorporating fairness constraints into experimental design.

1 Introduction

In experimental design problems, we are given vectors $v_1, \ldots, v_n \in \mathbb{R}^d$ and a parameter $b \geq d$, and the goal is to choose a (multi-)subset S of b vectors so that $\sum_{i \in S} v_i v_i^T$ optimizes some objective function. The most popular and well-studied objective functions are:

- D-design: Maximizing $\left(\det\left(\sum_{i \in S} v_i v_i^T\right)\right)^{\frac{1}{d}}$.
- A-design: Minimizing $\operatorname{tr}\left(\left(\sum_{i \in S} v_i v_i^T\right)^{-1}\right)$.
- E-design: Maximizing $\lambda_{\min}\left(\sum_{i \in S} v_i v_i^T\right)$.

Two settings are studied in the literature. One is the "with repetition" setting where each vector is allowed to be chosen multiple times, and the other is the "without repetition" setting where each vector is allowed to be chosen at most once. There is a simple reduction from the with repetition setting to the without repetition setting. All the results in this paper apply in the more general without repetition setting.

These problems of choosing a representative subset of vectors have a wide range of applications.

- Experimental design is a classical topic in statistics with extensive literature [19, 5, 34, 23], where the goal is to choose b (noisy) linear measurements from $v_1, \ldots, v_n \in \mathbb{R}^d$ so as to maximize the statistical efficiency of estimating an unknown vector in $\mathbb{R}^d$.
- In machine learning, they are used in active learning [4], feature selection [10], and data summarization [30, 11].
- In numerical linear algebra, they are used in column subset selection [7], sparse least square regression [9], and matrix approximation [17, 18].
- In signal processing, they are used in sensor placement problems [24], and optimal subsampling in graph signal processing [12, 13, 14].
- In network design, the problem of choosing a subgraph with at most b edges to minimize the total effective resistance [22, 26] is an A-design problem, and the problem of choosing a subgraph with at most b edges to maximize the algebraic connectivity [21, 25, 26] is an E-design problem.

We refer the interested reader to [38, 36, 28, 33, 3] for more discussions of these applications and further references on related work.

1.1 Our Results

We present both combinatorial algorithms and rounding algorithms for experimental design problems. A main contribution in this paper is to show that these two types of algorithms can be analyzed using the same local search framework. Using this framework, we match and improve all known results and also obtain some new results.

1.1.1 Combinatorial Algorithms

The Fedorov's exchange method [19] starts with an arbitrary initial set S_0 of b vectors, and in each step $t \geq 1$ it aims to exchange one of the vectors, $S_t \leftarrow S_{t-1} - v_i + v_j$ where $v_i \in S_{t-1}$ and $v_j \notin S_{t-1}$, to improve the objective value, and stops if such an improving exchange is not possible. The simplicity of this algorithm and its good empirical performance [16, 29, 31] make the method widely used [6]. The approximation guarantee of this method is only analyzed rigorously in a recent work [28], and we extend their analysis in multiple directions.

For D-design, it was proved in [28] that Fedorov's

*Supported by NSERC Discovery Grant 2950-120715. Full paper is avaliable at arXiv preprint arXiv:2010.15805.

[†]University of Waterloo. lapchi@uwaterloo.ca

[‡]University of Waterloo. hong.zhou@uwaterloo.ca

exchange method gives a polynomial time approximation algorithm for all inputs in the with repetition setting, and we extend their result to the without repetition setting.

THEOREM 1.1. *The Fedorov's exchange method is a polynomial time $\frac{b-d-1}{b}$-approximation algorithm for D-design in the without repetition setting. In particular, this is a $(1-\varepsilon)$-approximation algorithm whenever $b \geq d+1+\frac{d}{\varepsilon}$ for any $\varepsilon > 0$.*

For A-design, it was shown in [28] that there are arbitrarily bad local optimal solutions for the Fedorov's exchange method. Interestingly, we prove that Fedorov's exchange method works well as long as there exists an almost optimal solution with good condition number. This provides a new insight about when the local search method works well, and this condition may hold in practical instances. As a corollary, this also extends the analysis of Fedorov's exchange method in [28] when all the vectors are short to the without repetition setting (see Section 3.2).

THEOREM 1.2. *Let $X := \sum_{i=1}^n x(i) \cdot v_i v_i^T$ with $\sum_{i=1}^n x(i) = b$ and $x_i \in [0,1]$ for $1 \leq i \leq n$ be a fractional solution to A-design. For any $\varepsilon \in (0,1)$, the Fedorov's exchange method returns an integral solution $Z = \sum_{i=1}^n z(i) \cdot v_i v_i^T$ with $\sum_{i=1}^n z(i) \leq b$ and $z(i) \in \{0,1\}$ for $1 \leq i \leq n$ such that*

$$\operatorname{tr}\left(Z^{-1}\right) \leq (1+\varepsilon) \cdot \operatorname{tr}(X^{-1}) \quad \textit{whenever}$$
$$b \geq \Omega\Big(\frac{d+\sqrt{\operatorname{tr}(X)\operatorname{tr}\left(X^{-1}\right)}}{\varepsilon}\Big).$$

In particular, let $\kappa = \frac{\lambda_{\max}(X^*)}{\lambda_{\min}(X^*)}$ be the condition number of an optimal solution X^*, then the Fedorov's exchange method gives a $(1+\varepsilon)$-approximation algorithm for A-design whenever $b \geq \Omega\Big(\frac{(1+\sqrt{\kappa})\cdot d}{\varepsilon}\Big)$, and the time complexity is polynomial in $n, d, \frac{1}{\varepsilon}, \kappa$.

For E-design, there are no known combinatorial local search algorithms, and there are examples showing that Fedorov's exchange method does not work even if there exists a well-conditioned optimal solution (see full version of the paper for examples). Using the regret minimization framework in [1, 3], however, we prove that a modified local search algorithm using a "smoothed" objective function for E-design works as long as there exists an almost optimal solution with good condition number.

THEOREM 1.3. *Let $X := \sum_{i=1}^n x(i) \cdot v_i v_i^T$ with $\sum_{i=1}^n x(i) = b$ and $x(i) \in [0,1]$ for $1 \leq i \leq n$ be a fractional solution to E-design. For any $\varepsilon \in (0,1)$, there is a combinatorial local search algorithm which returns an integral solution $Z = \sum_{i=1}^n z(i) \cdot v_i v_i^T$ with $\sum_{i=1}^n z(i) \leq b$ and $z(i) \in \{0,1\}$ for $1 \leq i \leq n$ such that*

$$\lambda_{\min}(Z) \geq (1-\varepsilon) \cdot \lambda_{\min}(X) \quad \textit{whenever}$$
$$b \geq \Omega\Big(\frac{d}{\varepsilon^2}\sqrt{\frac{\lambda_{\mathrm{avg}}(X)}{\lambda_{\min}(X)}}\Big),$$

where $\lambda_{avg}(X) = \frac{\operatorname{tr}(X)}{d}$ is the average eigenvalue of X.

In particular, let $\kappa = \frac{\lambda_{\max}(X^*)}{\lambda_{\min}(X^*)}$ be the condition number of an optimal solution X^*, then the combinatorial local search method gives a polynomial time $(1-\varepsilon)$-approximation algorithm for E-design whenever $b \geq \Omega\left(\frac{d\sqrt{\kappa}}{\varepsilon^2}\right)$, and the time complexity is polynomial in $n, d, \frac{1}{\varepsilon}, \kappa$.

A combinatorial "capping" procedure was used in [28] to reduce the A-design problem to the case when every vector is "short", for which Fedorov's exchange method works. This capping procedure, however, crucially leveraged that a vector can be chosen multiple times. We do not have a preprocessing procedure to reduce A-design and E-design in the without repetition setting to the case when Theorem 1.2 and Theorem 1.3 apply. We leave it as an open problem to design a fully combinatorial algorithm for A-design and E-design in the general case.

1.1.2 Rounding Algorithms for Convex Programming Relaxations There are natural convex programming relaxations for the D/A/E-design problems. The best known rounding algorithms for these three problems are all quite different, i.e. approximate positively correlated distributions for D-design [36], proportional volume sampling for A-design [33], and regret minimization for E-design [3, 26]. Although the one-sided spectral rounding result in [3, 26] (see Section 2.3) provides a general solution for a large class of experimental design problems including D/A/E-design, this only works under the stronger assumption that $b \geq \Omega\left(\frac{d}{\varepsilon^2}\right)$ and it was unclear how to unify the best known algorithmic results.

Surprisingly, we prove that the iterative randomized rounding algorithm for E-design in [26] can be modified just slightly to match and improve the previous results for D/A-design, as well as to extend them to handle multiple knapsack constraints. To this end, we bypass the one-sided spectral rounding problem. Instead, we perform a refined analysis for the iterative randomized rounding algorithm, in which the minimum eigenvalue of the current solution plays an unexpectedly crucial role for D/A-design as well. This provides a unified rounding algorithm to achieve the optimal results for the natural convex programming relaxations for these experimental design problems.

In D/A/E-design with knapsack constraints, we are given vectors $v_1, \ldots, v_n \in \mathbb{R}^d$, knapsack constraints $c_1, \ldots, c_m \in \mathbb{R}^n_+$ and budgets $b_1, \ldots, b_m \geq 0$, and the goal is to find a solution $z \in \{0,1\}^n$ with $\langle c_i, z \rangle \leq b_i$ for $1 \leq i \leq m$ to optimize the objective value. Consider the following natural convex programming relaxations for D/A-design.

$$(1.1)\quad \begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{maximize}} && \log\det\left(\sum\nolimits_{i=1}^n x(i)\cdot v_i v_i^T\right) \quad \text{or} \\ &\underset{x \in \mathbb{R}^n}{\text{minimize}} && \operatorname{tr}\left(\left(\sum\nolimits_{i=1}^n x(i)\cdot v_i v_i^T\right)^{-1}\right) \\ &\text{subject to} && \langle c_j, x\rangle \leq b_j, \qquad \text{for } 0 \leq j \leq m, \\ & && 0 \leq x(i) \leq 1, \qquad \text{for } 1 \leq i \leq n. \end{aligned}$$

THEOREM 1.4. *Let $x \in [0,1]^n$ be an optimal fractional solution to D/A-design with knapsack constraints. For any $\varepsilon \leq \frac{1}{200}$, if each knapsack constraint budget satisfies $b_j \geq \frac{2d\|c_j\|_\infty}{\varepsilon}$, then there is a randomized exchange algorithm which returns in polynomial time an integral solution $Z = \sum_{i=1}^n z(i)\cdot v_i v_i^T$ with $z(i) \in \{0,1\}$ for $1 \leq i \leq n$ such that*

$$\det\left(\sum_{i=1}^n z(i)\cdot v_i v_i^T\right)^{\frac{1}{d}} \geq \big(1-O(\varepsilon)\big)\cdot\det\left(\sum_{i=1}^n x(i)\cdot v_i v_i^T\right)^{\frac{1}{d}} \text{ or}$$

$$\operatorname{tr}\left(\left(\sum_{i=1}^n z(i)\cdot v_i v_i^T\right)^{-1}\right) \leq (1+\varepsilon)\cdot\operatorname{tr}\left(\left(\sum_{i=1}^n x(i)\cdot v_i v_i^T\right)^{-1}\right)$$

for D/A-design respectively with probability at least $1 - O\left(\frac{k^2}{\varepsilon^2}\cdot e^{-\Omega(\sqrt{d})}\right)$ where $k = O(\|x\|_1 + d)$. Furthermore, each knapsack constraint $\langle c_j, z\rangle \leq b_j$ is satisfied with probability at least $1 - e^{-\Omega(\varepsilon d)}$.

Note that D/A-design with a cardinality constraint is the special case when there is only one cost constraint and $c = \vec{1}$. In this special case, Theorem 1.4 improves the previous results in [36, 33] by removing the term $O\left(\frac{1}{\varepsilon^2}\log\left(\frac{1}{\varepsilon}\right)\right)$ from their assumption $b \geq \Omega\left(\frac{d}{\varepsilon} + \frac{1}{\varepsilon^2}\log\left(\frac{1}{\varepsilon}\right)\right)$, and this achieves the optimal integrality gap result for D-design [36] and A-design [33] (see the full version of the paper for a proof). In the general case with knapsack constraints, Theorem 1.4 improves the previous result in [26], which requires a stronger assumption that $b_j \geq \Omega\left(\frac{d\|c_j\|_\infty}{\varepsilon^2}\right)$ to obtain the same approximation guarantee. The knapsack constraints can be used for weighted experimental design and for incorporating fairness constraints in experimental design, which we will discuss in the next subsection.

1.1.3 Some Applications

We discuss some applications of our results in specific instances of experimental design problems.

Fair and Diverse Data Summarization: In the data summarization problem, we are given n data points $v_1, \ldots, v_n \in \mathbb{R}^d$, and the objective is to choose a subset of b data points that provides a "fair" and "diverse" summary of the data. For diversity, the D-design objective of maximizing determinant is a popular measure used in previous work [30, 11]. For fairness, the partition constraints [32, 11] for D-design are used to partition the set X of data points into p disjoint groups $X_1 \cup \cdots \cup X_p$ and to ensure that b_i data points are chosen in X_i where $\sum_{i=1}^p b_i = b$.

We believe that Theorem 1.4 for D-design with knapsack constraints provides an alternative solution for this problem. The main advantage is that the knapsack constraints are more flexible in that they do not require the groups to be disjoint. For instance, we can have knapsack constraints on arbitrary subsets $X_1, \ldots, X_p \subseteq X$ of the form $\sum_{j \in X_i} x(j) \leq b_i$ to ensure that at most b_i data points are chosen in group X_i, so that we can handle constraints of overlapping groups such as race, age, gender (e.g. at most 50% of the chosen vectors correspond to men/women), etc. Also, the approximation guarantee in Theorem 1.4 is stronger than the constant factor approximation for D-design with partition constraint [32], and the convex programming relaxation used in Theorem 1.4 is simpler and easier to be solved than the more sophisticated one used in [32].

Minimizing Total Effective Resistance: Ghosh, Boyd and Saberi [22] studied the problem of choosing a subgraph with at most b edges to minimize the total effective resistance, and showed that this is a special case of A-design. The proportional volume sampling algorithm by Nikolov, Singh and Tantipongpipat [33] achieves a $(1+\varepsilon)$-approximation for this problem when $b \geq \Omega(\frac{n}{\varepsilon} + \frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon})$ where n is the number of vertices in the graph. Lau and Zhou [26] considered the weighted problem of choosing a subgraph with total edge cost at most b to minimize the total effective resistance, and gave a $(1+\varepsilon)$-approximation algorithm when $b \geq \Omega\left(\frac{n\|c\|_\infty}{\varepsilon^2}\right)$ where c is the cost vector of the edges. Theorem 1.4 improves these two results (see the full version of the paper for a proof of the following corollary).

COROLLARY 1.1. *For any $0 < \varepsilon < 1$, there is a polynomial time randomized $(1+\varepsilon)$-approximation algorithm for minimizing total effective resistance in an edge weighted graph whenever $b \geq \Omega\left(\frac{n\|c\|_\infty}{\varepsilon}\right)$.*

Maximizing Algebraic Connectivity: Ghosh and Boyd [21] studied the problem of choosing a subgraph with total cost at most b that maximizes the algebraic connectivity, *i.e.* the second smallest eigenvalue of its Laplacian matrix. Kolla, Makarychev, Saberi and Teng [25] provided the first algorithm with non-trivial approximation guarantee in the zero-one cost setting. Lau and Zhou [26] observed that this is a special case of E-design and gave a $(1-\varepsilon)$-approximation algorithm when $b \geq \Omega\left(\frac{n\|c\|_\infty}{\varepsilon^2}\right)$ where c is the cost vector of the edges. All previous results are based on convex programming. Theorem 1.3 provides a combinatorial algorithm for the unweighted problem, where the goal is to choose b edges to maximize the algebraic connectivity, and shows that it has a good performance as long as the optimal value is large (see the full version of the paper for a proof).

Corollary 1.2. *For any $0 < \varepsilon < 1$, there is a polynomial time combinatorial $(1-\varepsilon)$-approximation algorithm for maximizing algebraic connectivity in an unweighted graph whenever $b \geq \Omega\left(\frac{n}{\varepsilon^4 \lambda_2^*}\right)$, where λ_2^* is the optimal value for the problem.*

1.2 Techniques We extend the randomized approach in [26] to analyze both combinatorial local search algorithms and to design improved approximation algorithms for D/A-design with knapsack constraints. The approach in [26] is based on the regret minimization framework developed in [3] for the one-sided spectral rounding problem. In the following, we will first present the techniques used in these two previous results, and then present the new ideas in this paper.

Previous Techniques: In [3], Allen-Zhu, Li, Singh, and Wang first solved the natural convex programming relaxation for experimental design and obtained a solution $x \in \mathbb{R}^n$, and performed a linear transformation so that $\sum_{i=1}^n x(i) \cdot v_i v_i^T = I$. They showed that the experimental design problem is reduced to the following one-sided spectral rounding problem, where the goal to find a subset $S \subseteq [n]$ so that $\sum_{i\in S} v_i v_i^T \succcurlyeq (1-\varepsilon)I$ and $|S| \leq \sum_{i=1}^n x(i) = b$. To solve this problem, they started from an arbitrary initial set S_0 of b vectors, and in each step $t \geq 1$ they sought to find a pair $i_t \in S_{t-1}$ and $j_t \notin S_{t-1}$ so that the new solution $S_t \leftarrow S_{t-1} - v_{i_t} + v_{j_t}$ improves the current solution S_{t-1} in terms of a potential function related to the minimum eigenvalue. Using the regret minimization framework that maintains a density matrix A_t at each step t (see Section 2.2), they proved that choosing

$$i_t := \text{argmin}_{i\in S_{t-1}} \frac{\langle v_i v_i^T, A_t\rangle}{1-2\alpha\langle v_i v_i^T, A_t^{\frac{1}{2}}\rangle} \quad \text{and}$$

$$j_t := \text{argmax}_{j\notin S_{t-1}} \frac{\langle v_j v_j^T, A_t\rangle}{1+2\alpha\langle v_j v_j^T, A_t^{\frac{1}{2}}\rangle}$$

would improve the current solution S_{t-1} as long as $\sum_{i\in S_t} v_i v_i^T \not\succcurlyeq (1-\varepsilon)I$.

In [26], the idea is to use the following probability distributions to sample i_t and j_t:

$$\begin{aligned} &\Pr(i_t = i) \propto \big(1 - x(i)\big) \cdot (1 - 2\alpha\langle v_i v_i^T, A_t^{\frac{1}{2}}\rangle) \text{ and} \\ &\Pr(i_j = j) \propto x(j) \cdot (1 + 2\alpha\langle v_j v_j^T, A_t^{\frac{1}{2}}\rangle). \end{aligned} \tag{1.2}$$

The advantage of doing random sampling is that the knapsack constraints will be approximately preserved, while the potential function related to the minimum eigenvalue is expected to improve. Freedman's martingale inequality and a new concentration inequality for non-martingales are used to prove that all these quantities are close to their expected values with high probability.

Analysis of Combinatorial Algorithms: In this paper, we use this randomized approach in [26] to analyze both combinatorial algorithms and rounding algorithms. For combinatorial local search algorithms, one difference from the previous analysis in [28] is that we compare the objective of the current integral solution to that of an optimal *fractional* solution. When the objective value of the fractional solution is considerably better than that of the current integral solution, we use the fractional solution to define appropriate probability distributions similar to that in (1.2) to sample i_t and j_t so that the expected objective value of $S_t \leftarrow S_{t-1} - v_{i_t} + v_{j_t}$ improves, and this would imply the existence of an improving pair in Fedorov's exchange method. One advantage of this approach is that this allows us the flexibility to compare the current integral solution to a fractional solution with smaller budget which still has its objective value close to the optimal one.

Our analysis is arguably simpler than that in [28] which uses a dual fitting method while we only do a primal analysis. More importantly, our analysis shows that if the optimal fractional solution is well-conditioned (e.g. $\sum_{i=1}^n x(i) \cdot v_i v_i = I_d$), then the Fedorov's exchange method indeed performs as well as the best known rounding algorithms. This gives us a new insight that the only important step in rounding algorithms for the unweighted experimental design problems is the ability to first transform the optimal fractional solution to the identity matrix. For E-design, simply doing Fedorov's exchange method on the objective function $\lambda_{\min}\left(\sum_{i\in S_t} v_i v_i^T\right)$ would not work (see full versino of the paper for examples), and instead we apply the Fedorov's exchange method to the potential function in the regret minimization framework, which is morally the same as the potential function

$\operatorname{tr}\left(\left(\sum_{i\in S_t} v_i v_i^T - lI_d\right)^{-1}\right)$ used by Batson, Spielman and Srivastava for spectral sparsification [8].

Analysis of Rounding Algorithms: For the rounding algorithm for experimental design with knapsack constraints, surprisingly we prove that a minor modification of the algorithm for E-design in [26] would work for D/A-design with improved approximation guarantees! Essentially, we just use the algorithm for E-design but only require that the solution to have minimum eigenvalue $\frac{3}{4}$ rather than $1-\varepsilon$. Our analysis has two phases. In the first phase, using the results in [26], we show that the randomized exchange algorithm will find a solution with minimum eigenvalue at least $\frac{3}{4}$ in polynomial time with high probability whenever $b \geq \Omega\left(\frac{d}{\varepsilon}\right)$ (rather than $b \geq \Omega\left(\frac{d}{\varepsilon^2}\right)$ in order to achieve minimum eigenvalue at least $1-\varepsilon$). In the second phase, we prove that the minimum eigenvalue will maintain to be at least $\frac{1}{4}$ with high probability when ε is not too tiny, and then the objective value for D-design and A-design will improve to $(1 \pm \varepsilon)$ times the optimal objective value in polynomial time with high probability. The condition that the minimum eigenvalue is at least $\frac{1}{4}$ is used crucially in multiple places for the analysis of the second phase. Interestingly, it is used in showing that the probability distributions in (1.2) for E-design are also good for improving the objective value for D-design and A-design. Moreover, it is crucially used in the martingale concentration argument, e.g. to show that the martingale is bounded and to prove upper bounds on the variance of the changes. For the martingale concentration argument, we also use the optimality conditions for convex programs to prove that the vectors with fractional value are "short" in order to bound the quantities involved. Overall, the analysis for the rounding algorithm is quite involved, but it provides a unifying algorithm to achieve the optimal results for the natural convex relaxations for D/A/E-design. Please refer to Section 4 for a more detailed outline of the analysis.

1.3 Previous Work

All three experimental design problems are NP-hard [15, 39] and also APX-hard [35, 33, 15]. Despite the long history and the wide interest, strong approximation algorithms for these problems are only obtained recently.

D-design: Singh and Xie [36] designed an $(1-\varepsilon)$-approximation algorithm for D-design in the with repetition setting when $b \geq \frac{2d}{\varepsilon}$, and in the without repetition setting when $b = \Omega\left(\frac{d}{\varepsilon} + \frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$. Their algorithm is by rounding an optimal solution to a natural convex program relaxation using approximate positively correlated distributions.

Madan, Singh, Tantipongpipat and Xie [28] analyzed the Fedorov's exchange method and proved that it gives an $(1-\varepsilon)$-approximation algorithm for D-design as long as $b \geq d + \frac{d}{\varepsilon}$, which improves upon the above result. However, they only provide a polynomial time implementation of the local search algorithm to achieve this guarantee in the less general with repetition setting.

A-design: Nikolov, Singh and Tantipongpipat [33] designed an $(1+\varepsilon)$-approximation algorithm for A-design in the with repetition setting when $b \geq d + \frac{d}{\varepsilon}$, and in the without repetition setting when $b = \Omega\left(\frac{d}{\varepsilon} + \frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$. Their algorithm is by rounding an optimal solution to a natural convex program relaxation using proportional volume sampling.

Madan, Singh, Tantipongpipat and Xie [28] also analyzed the Fedorov's exchange method for A-design, and showed that there are arbitrarily bad local optimal solutions. On the other hand, they proved that Fedorov's exchange method works when all the input vectors are "short", and they designed a "capping procedure" to reduce the general case to the case when all vectors are short. As a result, they obtained a combinatorial $(1+\varepsilon)$-approximation algorithm, without solving convex programs, for A-design when $b \geq \Omega\left(\frac{d}{\varepsilon^4}\right)$ in the with repetition setting.

E-design: Allen-Zhu, Li, Singh and Wang [2, 3] designed an $(1-\varepsilon)$-approximation algorithm for E-design in the with and without repetition settings when $b \geq \Omega\left(\frac{d}{\varepsilon^2}\right)$. Their algorithm is by rounding an optimal solution to a natural convex program relaxation using the regret minimization framework, which was initially developed for the spectral sparsification problem [1]. They formulated and solved a "one-sided spectral rounding problem" (see Section 2.3), and showed that experimental design with any objective function satisfying some mild regularity assumptions, including D/A/E-design, can be reduced to the one-sided spectral rounding problem. Their algorithm for one-sided spectral rounding can be viewed as a local search algorithm, and this was the starting point of the current work.

Nikolov, Singh and Tantipongpipat [33] showed that the assumption $b \geq \Omega\left(\frac{d}{\varepsilon^2}\right)$ is necessary to achieve $(1-\varepsilon)$-approximation for E-design using the natural convex program, and Lau and Zhou [26] showed that the assumption $b \geq \Omega\left(\frac{d}{\varepsilon^2}\right)$ is necessary for the one-sided spectral rounding problem. These suggest that the regret minimization framework may not be used to match the results for D/A-design, but we bypass the one-sided spectral rounding problem to prove Theorem 1.4.

Experimental design with additional constraints: Lau and Zhou [26] considered the generalization of the experimental design problem with additional knapsack constraints as in (1.1). In particular, it generalizes the experimental design problems to the *weighted* problems, where each vector v_i has a weight $c(i)$ and the goal is to choose a subset S of vectors with $\sum_{i\in S} c(i) \leq b$ to optimize the objective value. Using a randomized iterative rounding algorithm, they obtained a $(1\pm\varepsilon)$-approximation algorithm for weighted D/A/E-design when $b \geq \Omega\left(\frac{d\|c\|_\infty}{\varepsilon^2}\right)$.

Using more sophisticated convex programming relaxations, Nikolov and Singh [32] designed an approximation algorithm for D-design under partition constraints. Recently, Madan, Nikolov, Singh and Tantipongpipat [27] designed an approximation algorithm for D-design under general matroid constraints.

2 Preliminaries

We recall some basic linear algebra in Section 2.1. Then, we review the regret minimization framework in Section 2.2, and the iterative randomized rounding algorithm for one-sided spectral rounding in Section 2.3. Finally, we state some useful inequalities for the analysis of martingales and the analysis of the objective functions in Section 2.4 and Section 2.5 respectively.

2.1 Linear Algebra We write $\mathbb{R}$ and $\mathbb{R}_+$ as the sets of real numbers and non-negative real numbers. Throughout the paper, we use italic sans-serif font for vectors and matrices, e.g. x, A.

All the vectors in this paper only have real entries. Let $\mathbb{R}^d$ denote the d-dimensional Euclidean space. We write $\vec{1}_d$ as the d-dimensional all-one vector. Given a vector $x \in \mathbb{R}^d$, we write $x(i)$ as the i-th entry of vector x, and write $x(S) := \sum_{i\in S} x(i)$ for any subset $S \subseteq [d]$. We denote $\|x\|_2$ the ℓ_2-norm, $\|x\|_1$ the ℓ_1-norm, and $\|x\|_\infty$ the ℓ_∞-norm of x. A vector $v \in \mathbb{R}^d$ is a column vector, and its transpose is denoted by v^T. Given two vectors $x, y \in \mathbb{R}^d$, the inner product is defined as $\langle x, y\rangle := \sum_{i=1}^n x(i)\cdot y(i)$. The Cauchy-Schwarz inequality says that $\langle x, y\rangle \leq \|x\|\,\|y\|$.

All matrices considered in this paper are real symmetric matrices. We denote the $d \times d$ identity matrix by I_d or simply I when the dimension is clear from the context. We write $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ as the maximum and the minimum eigenvalue of a matrix M. The trace of a matrix M, denoted by $\mathrm{tr}(M)$, is defined as the sum of the diagonal entries of M. It is well-known that $\mathrm{tr}(M) = \sum_{i=1}^n \lambda_i(M)$ where $\lambda_i(M)$ denotes the i-th eigenvalue of M.

A matrix M is a positive semidefinite (PSD) matrix, denoted as $M \succcurlyeq 0$, if M is symmetric and all the eigenvalues are nonnegative, or equivalently, the quadratic form $x^T M x \geq 0$ for any vector x. We use $A \succcurlyeq B$ to denote $A - B \succcurlyeq 0$ for matrices A and B. We write $\mathbb{S}^n_+$ as the set of all n-dimensional PSD matrices. Let $M \succcurlyeq 0$ be a PSD matrix with eigendecomposition $M = \sum_i \lambda_i v_i v_i^T$, where $\lambda_i \geq 0$ is the i-th eigenvalue and v_i is the corresponding eigenvector. The square root of M is $M^{1/2} := \sum_i \sqrt{\lambda_i} v_i v_i^T$. Given two matrices A and B of the same size, the Frobenius inner product of A, B is denoted as $\langle A, B\rangle := \sum_{i,j} A(i,j)\cdot B(i,j) = \mathrm{tr}(A^T B)$. The following are two standard facts: $A, B \succcurlyeq 0 \implies \langle A, B\rangle \geq 0$ and $A \succcurlyeq 0, B \succcurlyeq C \succcurlyeq 0 \implies \langle A, B\rangle \geq \langle A, C\rangle$.

2.2 Regret Minimization We will use the regret minimization framework developed in [1, 3] for spectral sparsification and one-sided spectral rounding. The framework is for an online optimization setting. In each iteration t, the player chooses an action matrix A_t from the set of density matrices $\Delta_d = \{A \in \mathbb{R}^{d\times d} \mid A \succcurlyeq 0, \mathrm{tr}(A) = 1\}$, which can be understood as a probability distribution over the set of unit vectors. The player then observes a feedback matrix F_t and incurs a loss of $\langle A_t, F_t\rangle$. After τ iterations, the regret of the player is defined as

$$R_\tau := \sum_{t=1}^{\tau}\langle A_t, F_t\rangle - \inf_{B\in\Delta_d}\sum_{t=1}^{\tau}\langle B, F_t\rangle = \sum_{t=1}^{\tau}\langle A_t, F_t\rangle - \lambda_{\min}\left(\sum_{t=1}^{\tau} F_t\right),$$

which is the difference between the loss of the player actions and the loss of the best fixed action B, that can be assumed to be a rank one matrix vv^T. The objective of the player is to minimize the regret. A well-known algorithm for regret minimization is follow-the-regularized-leader which plays the action

$$A_t = \mathrm{argmin}_{A\in\Delta_d}\left\{w(A) + \alpha\cdot\sum_{l=0}^{t-1}\langle A, F_l\rangle\right\},$$

where $w(A)$ is a regularization term and α is a parameter called the learning rate that balances the loss and the regularization. Here F_0 is an initial feedback which is given before the game started. Different choices of regularization give different algorithms for regret minimization. The choice that we will use is the $\ell_{\frac{1}{2}}$-regularizer $w(A) = -2\,\mathrm{tr}(A^{\frac{1}{2}})$ introduced in [1], which plays the action

$$A_t = \left(l_t I + \alpha\sum_{l=0}^{t-1} F_l\right)^{-2}, \tag{2.3}$$

where l_t is the unique constant that ensures $A_t \in \Delta_d$. Allen-Zhu, Li, Singh and Wang [3] proves the following regret bound for rank-two feedback matrices with $\ell_{\frac{1}{2}}$ regularizer.

THEOREM 2.1. (LEMMA 2.5 IN [3]) *Suppose the action matrix $A_t \in \mathbb{R}^{d\times d}$ is of the form of (2.3) for some $\alpha > 0$. Suppose the initial feedback matrix $F_0 \in \mathbb{S}^d$ is a symmetric matrix, and for all $t \geq 1$ each feedback matrix F_t is of the form $v_{j_t} v_{j_t}^T - v_{i_t} v_{i_t}^T$ for some $v_{j_t}, v_{i_t} \in \mathbb{R}^d$ such that $\alpha\langle v_{i_t} v_{i_t}^T, A_t^{\frac{1}{2}}\rangle < \frac{1}{2}$. Then for any given density* *matrix U,*

$$\sum_{t=1}^{\tau}\langle F_t, U\rangle \geq \sum_{t=1}^{\tau}\left(\frac{\langle v_{j_t}v_{j_t}^T, A_t\rangle}{1+2\alpha\langle v_{j_t}v_{j_t}^T, A_t^{\frac{1}{2}}\rangle} - \frac{\langle v_{i_t}v_{i_t}^T, A_t\rangle}{1-2\alpha\langle v_{i_t}v_{i_t}^T, A_t^{\frac{1}{2}}\rangle}\right) - \frac{D(A_1, U)}{\alpha},$$

where $D(A_1, U) := \operatorname{tr}(A_1^{\frac{1}{2}}) - 2\operatorname{tr}(U^{\frac{1}{2}}) + \langle A_1^{-\frac{1}{2}}, U\rangle$ is the Bregman divergence of the $\ell_{\frac{1}{2}}$-regularizer.

The above theorem is used in [3] to give a lower bound on the minimum eigenvalue of $\lambda_{\min}(\sum_{t=0}^{\tau} F_t)$, by bounding $D(A_1, U) \leq \alpha\langle F_0, U\rangle + 2\sqrt{d}$. We will use the following more refined version where there is an extra $\lambda_{\min}(F_0)$ term. (see full version of the paper for a proof)

Corollary 2.1. *Under the same assumptions as in Theorem 2.1,*

$$\lambda_{\min}\left(\sum_{t=0}^{\tau} F_t\right) \geq \sum_{t=1}^{\tau}\left(\frac{\langle v_{j_t}v_{j_t}^T, A_t\rangle}{1+2\alpha\langle v_{j_t}v_{j_t}^T, A_t^{\frac{1}{2}}\rangle} - \frac{\langle v_{i_t}v_{i_t}^T, A_t\rangle}{1-2\alpha\langle v_{i_t}v_{i_t}^T, A_t^{\frac{1}{2}}\rangle}\right) - \frac{2\sqrt{d}}{\alpha} + \lambda_{\min}(F_0).$$

One technical point used in [2, 3] is that the partial solution $Z_t := \sum_{l=0}^{t-1} F_l$ at time t and the action matrix A_t at time t have the same eigenbasis due to (2.3). This allows one to bound $\langle Z_t, A_t\rangle$ and $\langle Z_t, A_t^{\frac{1}{2}}\rangle$ as follows.

Lemma 2.1. (Claim 2.11 in [3]) *Let $Z \succcurlyeq 0$ be an $d \times d$ positive semidefinite matrix and $A = (\alpha Z + lI)^{-2}$ for some $\alpha > 0$ where l is the unique constant such that A is a density matrix. Then*

$$\langle Z, A\rangle \leq \frac{\sqrt{d}}{\alpha} + \lambda_{\min}(Z) \quad \text{and} \quad \alpha\langle Z, A^{\frac{1}{2}}\rangle \leq d + \alpha\sqrt{d}\cdot\lambda_{\min}(Z).$$

We will use the above lemma in the analysis of the combinatorial algorithm for E-design. We also use the following simple fact about the action matrix.

Lemma 2.2. *For any $d \times d$ matrix $A \succcurlyeq 0$ satisfying $\operatorname{tr}(A) = 1$, $\operatorname{tr}\left(A^{\frac{1}{2}}\right) \leq \sqrt{d}$.*

2.3 One-Sided Spectral Rounding and Iterative Randomized Rounding The following one-sided spectral rounding result was formulated and proved by Allen-Zhu, Li, Singh and Wang [3]. It is the main theorem that implies a $(1\pm\varepsilon)$-approximation algorithm for a large class of experimental design problems with one cardinality constraint, whenever the budget $b \geq \Omega\left(\frac{d}{\varepsilon^2}\right)$.

Theorem 2.2. ([3]) *Let $v_1, v_2, \ldots, v_n \in \mathbb{R}^d$, $x \in [0,1]^m$ and $b = \sum_{i=1}^m x(i)$. Suppose $\sum_{i=1}^m x(i)\cdot v_i v_i^T = I_d$ and $b \geq \frac{5d}{\varepsilon^2}$ for some $\varepsilon \in (0, \frac{1}{3}]$. Then there is a polynomial time algorithm to return a subset $S \subseteq [m]$ with*

$$|S| \leq b \quad \text{and} \quad \sum_{i\in S} v_i v_i^T \succcurlyeq (1-3\varepsilon)\cdot I_d.$$

Theorem 2.2 was extended in [26] to incorporate non-negative linear constraints, where the goal is to output a subset S that approximately satisfies the spectral lower bound and also $c(S) \approx \langle c, x\rangle$ for any non-negative linear constraint $c \in \mathbb{R}_+^n$. The randomized exchange algorithm in this paper is almost the same as the iterative randomized rounding algorithm in [26]. In the following, we describe the algorithm in [26] and state several results that we will use in our analysis.

In the iterative randomized rounding algorithm, we start with an initial solution set $S_0 \subseteq [n]$, which is constructed by sampling each vector i with probability $x(i)$ independently for all $i \in [n]$. In each iteration t, with the current solution set S_{t-1}, we randomly choose a vector $i_t \in S_{t-1}$ and a vector $j_t \in [n] \setminus S_{t-1}$ and set $S_t \leftarrow S_{t-1} - i_t + j_t$. The sampling distributions in each iteration depend on the action matrix A_t as defined in Eq. (2.3). The probability of choosing a vector $i_t \in S_{t-1}$ is $\frac{1}{k}\cdot\left(1 - x(i)\right)\cdot\left(1 - 2\alpha\langle v_i v_i^T, A_t^{\frac{1}{2}}\rangle\right)$, and the probability of choosing a vector $j_t \in [n] \setminus S_{t-1}$ is $\frac{1}{k}\cdot x(j)\cdot\left(1 + 2\alpha\langle v_j v_j^T, A_t^{\frac{1}{2}}\rangle\right)$, where k is a large enough denominator so that the two distributions are well-defined. Informally, the terms $1 - x(i)$ and $x(j)$ in the sampling probability help us maintain the cost $c(S_t) \approx \langle c, x\rangle$, and the term depending on A_t help us improve the spectral lower bound. The following are the precise statements that we will use in this paper.

Theorem 2.3. (Theorem 3.8 of [26]) *Let $\alpha = \frac{\sqrt{d}}{\gamma}$ be the learning rate used in computing the action matrix as defined in (2.3). Let τ be the first time such that the solution S_τ of the iterative randomized rounding algorithm satisfies $\sum_{i\in S_\tau} v_i v_i^T \succcurlyeq (1-2\gamma)\cdot I$. The probability that $\tau \leq \frac{2k}{\gamma}$ is at least $1 - e^{-\Omega(\sqrt{d})}$.*

Theorem 2.4. (Theorem 3.12 of [26]) *Let $\alpha = \frac{\sqrt{d}}{\gamma}$ be the learning rate used in computing the action matrix as defined in (2.3). Suppose that the solution S_t of the iterative randomized rounding algorithm satisfies $\lambda_{\min}\left(\sum_{i\in S_t} v_i v_i^T\right) < 1$ for all $1 \leq t < \tau$. Then, for any given $c \in \mathbb{R}_+^n$ and any $\delta \in [0,1]$,*

$$\Pr\left[c(S_\tau) \leq (1+\delta)\cdot\langle c, x\rangle + \frac{15d\,\|c\|_\infty}{\gamma}\right] \geq 1 - e^{-\Omega\left(\frac{\delta d}{\gamma}\right)}.$$

The iterative randomized rounding algorithm can be applied to the weighted D/A/E-design problems with the following guarantees (we rephrased Theorem 4.11 in [26] for multiple knapsack constraints setting). Note that the dependence in ε is optimal for E-design, but not optimal for A/E-design, which is exactly the motivation of proposing a new randomized exchange algorithm in Section 4.

We will also use the following two lemmas in [26] that were used in proving Theorem 2.3. Define

(2.4)
$$\Delta_t^+ := \frac{\langle v_{j_t} v_{j_t}^T, A_t\rangle}{1+2\alpha\langle v_{j_t} v_{j_t}^T, A_t^{\frac{1}{2}}\rangle}, \quad \Delta_t^- := \frac{\langle v_{i_t} v_{i_t}^T, A_t\rangle}{1-2\alpha\langle v_{i_t} v_{i_t}^T, A_t^{\frac{1}{2}}\rangle},$$

and $\Delta_t := \Delta_t^+ - \Delta_t^-$. Throughout the paper, we denote $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid S_{t-1}]$.

Lemma 2.3. (Lemma 3.5 and Lemma 3.7 of [26]) *Let $\tau \geq \tau' \geq 1$ be two time steps in the iterative randomized rounding algorithm, and let $\lambda := \max_{\tau' \leq t \leq \tau} \lambda_{\min}\left(\sum_{i\in S_t} v_i v_i^T\right)$. Let Δ_t be defined as in (2.4) and k be the denominator used in the sampling distributions. Then*

$$\sum_{t=\tau'+1}^{\tau} \mathbb{E}_t\left[\Delta_t\right] \geq \sum_{t=\tau'+1}^{\tau} \frac{1}{k}\left(1-\gamma-\lambda_{\min}(Z_{t-1})\right) \geq \frac{\tau-\tau'}{k}\cdot(1-\gamma-\lambda).$$

Furthermore, for any $\eta > 0$,

$$\Pr\left[\sum_{t=\tau'+1}^{\tau}\Delta_t \leq \sum_{t=\tau'+1}^{\tau}\mathbb{E}_t[\Delta_t]-\eta\right] \leq \exp\left(-\frac{\eta^2 k\sqrt{d}/2}{(\tau-\tau')\gamma(1+\lambda+\gamma)+\eta k\gamma/3}\right).$$

2.4 Martingales A sequence of random variables $Y_1,\ldots,Y_\tau$ is a martingale with respect to a sequence of random variables $Z_1,\ldots,Z_\tau$ if for all $t>0$, it holds that

1. Y_t is a function of $Z_1,\ldots,Z_{t-1}$;
2. $\mathbb{E}[|Y_t|] < \infty$;
3. $\mathbb{E}[Y_{t+1}|Z_1,\ldots,Z_t] = Y_t$.

We will use the following theorem by Freedman to bound the probability that Y_τ is large.

Theorem 2.5. ([20, 37]) *Let $\{Y_t\}_t$ be a real-valued martingale with respect to $\{Z_t\}_t$, and $\{X_t = Y_t - Y_{t-1}\}_t$ be the difference sequence. Assume that $X_t \leq R$ deterministically for $1 \leq t \leq \tau$. Let $W_t := \sum_{j=1}^t \mathbb{E}[X_j^2|Z_1,...,Z_{j-1}]$ for $1 \leq t \leq \tau$. Then, for all $\delta \geq 0$ and $\sigma^2 > 0$,*

$$\Pr\left(\exists t\in[\tau] : Y_t \geq \delta \text{ and } W_t \leq \sigma^2\right) \leq \exp\left(\frac{-\delta^2/2}{\sigma^2+R\delta/3}\right).$$

2.5 Inequalities for Objective Functions To analyze the change of the D-design objective value under exchange operations, we will derive a bound for determinant under rank-two update based on the following well-known matrix determinant lemma.

Lemma 2.4. *For any invertible matrix X and any vector $v \in \mathbb{R}^d$,*

$$\det(X \pm vv^T) = \det(X)(1 \pm \langle vv^T, X^{-1}\rangle).$$

The following determinant lower bound under a rank-two update is a simple consequence and was implicitly contained in [28].

Lemma 2.5. *Given a matrix $A \succ 0$ and two vectors $u, v \in \mathbb{R}^d$, if $\langle uu^T, A^{-1}\rangle \leq 1$, then*

$$\det(A - uu^T + vv^T) \geq \det(A)(1-\langle uu^T, A^{-1}\rangle)(1+\langle vv^T, A^{-1}\rangle).$$

Similarly, to analyze the change of A-design objective value under exchange operations, we need an upper bound on the trace of the inverse of a matrix under rank-two update. We use an observation in [3] to derive the inequality below (see the full version of the paper for a proof).

Lemma 2.6. *Let $A \in \mathbb{R}^{d\times d} \succ 0$ and $v, u \in \mathbb{R}^d$. If $2\langle vv^T, A^{-1}\rangle < 1$, then it holds for any $X \succcurlyeq 0$ that*

$$\langle X, \left(A - vv^T + uu^T\right)^{-1}\rangle \leq \langle X, A^{-1}\rangle + \frac{\langle X, A^{-1}vv^TA^{-1}\rangle}{1-2\langle vv^T, A^{-1}\rangle} - \frac{\langle X, A^{-1}uu^TA^{-1}\rangle}{1+2\langle uu^T, A^{-1}\rangle}.$$

We will also use the following simple claim in the analysis of combinatorial algorithms, whose proof is by checking the derivatives of $f(x)$ and $g(x)$.

Claim 2.6. *The functions $f(x) = \frac{x-c_1}{c_2+c_3\sqrt{x}}$ and $g(x) = \frac{x-c_1}{c_2+c_3 x}$ with $c_1, c_2, c_3 \geq 0$ are monotone increasing for $x \geq 0$.*

3 Combinatorial Algorithms

In this section, we present combinatorial local search algorithms for D/A/E-design problems. In Section 3.1, we show that Fedorov's exchange method is a polynomial time algorithm to achieve $\frac{b-d-1}{b}$-approximation for D-design, which extends the result in [28] to the without repetition setting. In Section 3.2, we analyze Fedorov's exchange method for A-design, and prove that it works well as long as there is a well-conditioned optimal solution. As a corollary, this extends the result in [28] for A-design to the without repetition setting, with an arguably simpler proof.

We can also show that Fedorov's exchange method does not work with the minimum eigenvalue objective, and we propose a modified local search algorithm and prove that it works well as long as there is a well-conditioned optimal solution. See the full version of the paper for more details.

A common theme in the analysis of all these algorithms is that we compare the current integral solution S to an optimal fractional solution x. As long as the objective value of x is significantly better than that of S, we use x to define two probability distributions to sample a pair of vectors v_i, v_j so that the expected objective value of $S - i + j$ improves that of S considerably, and so we can conclude that the combinatorial algorithms will find such an improving exchange pair. One advantage of this approach is that this allows us the flexibility to compare with a fractional solution with smaller budget (which still has objective value close to the optimal one), and this makes the analysis easier and simpler.

The following notations will be used throughout this section. Given a fractional solution $x \in [0,1]^n$ and an integral solution $S \subseteq [n]$, we denote

$$X := \sum\nolimits_{i=1}^{n} x(i) \cdot v_i v_i^T, \quad X_S := \sum\nolimits_{i\in S} x(i) \cdot v_i v_i^T,$$
$$x(S) := \sum\nolimits_{i \in S} x(i), \qquad Z := \sum\nolimits_{i\in S} v_i v_i^T.$$

3.1 Combinatorial Local Search Algorithm for D-Design We analyze the following version of Fedorov's exchange method for D-design, where we always choose a pair that maximizes the improvement of the objective value and we stop as soon as the improvement is not large enough.

Fedorov's Exchange Method for D-Design
Input: n vectors $v_1, ..., v_n \in \mathbb{R}^d$, a budget $b \geq d$.

1. Let $S_0 \subseteq [n]$ be an arbitrary set of full-rank vectors with $|S_0| = b$.
2. Let $t \leftarrow 1$ and $Z_1 := \sum_{i\in S_{t-1}} v_i v_i^T$.
3. **Repeat**
   (a) Find $i_t \in S_{t-1}$ and $j_t \in [n] \setminus S_{t-1}$ such that
   $$(i_t, j_t) = \underset{\substack{(i,j):\\ i\in S_{t-1},\\ j\in[n]\setminus S_{t-1}}}{\arg\max} \det\left(Z_t - v_i v_i^T + v_j v_j^T\right).$$
   (b) Set $S_t \leftarrow S_{t-1} \cup \{j_t\} \setminus \{i_t\}$ and $Z_{t+1} \leftarrow Z_t - v_{i_t} v_{i_t}^T + v_{j_t} v_{j_t}^T$ and $t \leftarrow t + 1$.

   Until $\det(Z_t) < \left(1 + \frac{d}{4b^3}\right) \det(Z_{t-1})$.
4. Return S_{t-2} as the solution set.

To analyze the change of the objective value in each iteration, note that $\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle \leq 1$ for any t as $i_t \in S_{t-1}$, and so it follows from Lemma 2.5 that

$$\det(Z_{t+1}) = \det(Z_t - v_{i_t} v_{i_t}^T + v_{j_t} v_{j_t}^T)$$
$$\geq \det(Z_t) \cdot (1 - \underbrace{\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle}_{\text{loss}}) \cdot (1 + \underbrace{\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle}_{\text{gain}}).$$

Therefore, in order to lower bound the determinant of the solution, we lower bound the "gain" term and upper bound the "loss" term to quantify the progress in each iteration. First, we prove the existence of i_t with small loss, with respect to a fractional solution x with $\|x\|_1 = q < b$.

LEMMA 3.1. (LOSS) *For any $x \in [0,1]^n$ with $\sum_{i=1}^n x(i) = q < b$ and any $S \subseteq [n]$ with $|S| = b$, there exists $i \in S$ with*

$$\langle v_i v_i^T, Z^{-1}\rangle \leq \frac{d - \langle X_S, Z^{-1}\rangle}{b - x(S)}.$$

Proof. Consider the probability distribution of removing a vector v_i where each $i \in S$ is sampled with probability $\big(1 - x(i)\big) / \sum_{j\in S} \big(1 - x(j)\big)$, so that the "staying" probability is proportional to the value $x(i)$. Note that the denominator is positive as $x(S) = q < b$, and thus the probability distribution is well-defined. Then, the expected loss using this probability distribution is

$$\mathbb{E}\big[\langle v_{i_t} v_{i_t}^T, Z^{-1}\rangle\big] = \frac{\sum_{i\in S} \big(1 - x(i)\big) \cdot \langle v_i v_i^T, Z^{-1}\rangle}{\sum_{j\in S} \big(1 - x(j)\big)}$$
$$= \frac{d - \langle X_S, Z^{-1}\rangle}{b - x(S)},$$

where the last equality follows as $\sum_{i\in S} v_i v_i^T = Z$ and $|S| = b$. Therefore, there must exist one vector i with $\langle v_i v_i^T, Z^{-1}\rangle$ at most the expected value. □

Next, we prove the existence of j_t with large gain, again with respect to a fractional solution x with $\|x\|_1 = q \leq b$.

LEMMA 3.2. (GAIN) *For any $x \in [0,1]^n$ with $\sum_{i=1}^n x(i) = q < b$ and any $S \subseteq [n]$ with $|S| = b$ and $x(S) < q$, there exists $j \in [n] \setminus S$ with*

$$\langle v_j v_j^T, Z^{-1}\rangle \geq \frac{\langle X, Z^{-1}\rangle - \langle X_S, Z^{-1}\rangle}{q - x(S)}.$$

Proof. Consider the probability distribution of adding a vector v_j where each $j \in [n] \setminus S$ is sampled with probability $x(j) / \sum_{i\in[n]\setminus S} x(i)$, so that the "adding" probability is proportional to the value $x(i)$. Note that the denominator is positive by our assumption that $x(S) < q$, and so the probability distribution is well-defined. Then, the expected gain using this probability distribution is

$$\mathbb{E}[\langle v_j v_j^T, Z^{-1}\rangle] = \frac{\sum_{j\in[n]\setminus S} x(j)\cdot\langle v_j v_j^T, Z^{-1}\rangle}{\sum_{i\in[n]\setminus S} x(i)} = \frac{\langle X, Z^{-1}\rangle - \langle X_S, Z^{-1}\rangle}{q - x(S)}.$$

Therefore, there must exist one vector j with $\langle v_j v_j^T, Z^{-1}\rangle$ at least the expected value. □

We are about ready to analyze when the objective value would increase. The following lemma will be used to relate the numerator of the gain term to the current objective value $\det(Z)$ (see the full version of the paper for a proof).

Lemma 3.3. *For any given $d \times d$ positive definite matrices $A, B \succ 0$,*

$$\langle A, B\rangle \geq d\cdot\det(A)^{\frac{1}{d}}\cdot\det(B)^{\frac{1}{d}}.$$

The following is the main technical result for D-design, which lower bounds the improvement of the objective value in each iteration. In the proof, we compare our current integral solution S with size b to a fractional solution y with size $q = b - d - \frac{1}{2}$.

Proposition 3.1. (Progress) *Let $x \in [0,1]^n$ be a feasible solution to the convex programming relaxation (1.1) for D-design with $\sum_{i=1}^n x(i) = b$ for $b \geq d+1$. Let Z_t be the current solution in the t-th iteration of Fedorov's exchange method. Then*

$$\det(Z_t)^{\frac{1}{d}} \leq \frac{b-d-1}{b}\cdot\det(X)^{\frac{1}{d}} \implies \det(Z_{t+1}) \geq \left(1 + \frac{d}{4b^3}\right)\cdot\det(Z_t).$$

Proof. We consider the following scaled-down version y, Y of the fractional solution x, X. Define

$$q := b - d - \frac{1}{2}, \quad y := \frac{q}{b}x, \quad Y := \sum_{i=1}^n y(i)\cdot v_i v_i^T = \frac{q}{b}X.$$

Note that $\det(Y)^{\frac{1}{d}} = \frac{q}{b}\cdot\det(X)^{\frac{1}{d}}$ and $\frac{1}{2} \leq q < b$. Let $S := S_{t-1}$ be the current solution set at time t. Note that we can assume $x(S) < b$ and hence $y(S) < q$, as otherwise $\det(Z_t) \geq \det(X)$ and there is nothing to prove. Hence, we can apply Lemma 3.1 and Lemma 3.2 on Y and S to ensure the existence of $i_t \in S$ and $j_t \in [n]\setminus S$ such that

$$\det(Z_{t+1}) \geq \det(Z_t)\cdot\left(1 - \frac{d - \langle Y_S, Z_t^{-1}\rangle}{b - y(S)}\right)\underbrace{\left(1 + \frac{\langle Y, Z_t^{-1}\rangle - \langle Y_S, Z_t^{-1}\rangle}{q - y(S)}\right)}_{(*)}.$$

The $(*)$ term can be bounded by

$$\begin{aligned}(*) &\geq 1 + \frac{d\cdot\det(Y)^{\frac{1}{d}}\cdot\det(Z_t^{-1})^{\frac{1}{d}} - \langle Y_S, Z_t^{-1}\rangle}{q - y(S)}\\ &\geq 1 + \frac{d\cdot\frac{q}{b}\det(X)^{\frac{1}{d}}\cdot\frac{b}{b-d-1}\det(X)^{-\frac{1}{d}} - \langle Y_S, Z_t^{-1}\rangle}{q - y(S)}\\ &= 1 + \frac{\left(1+\frac{1}{2q-1}\right)d - \langle Y_S, Z_t^{-1}\rangle}{q - y(S)},\end{aligned}$$

where the first inequality follows from Lemma 3.3, the second inequality follows from $Y = \frac{q}{b}X$ and the assumption $\det(Z)^{\frac{1}{d}} \leq \frac{b-d-1}{b}\det(X)^{\frac{1}{d}}$, and the last equality is by $q = b - d - \frac{1}{2}$.

To lower bound the improvement, we write $a := d - \langle Y_S, Z_t^{-1}\rangle$ as a shorthand, and then the multiplicative change factor is

$$\begin{aligned}&\left(1 - \frac{d - \langle Y_S, Z_t^{-1}\rangle}{b - y(S)}\right)\cdot\left(1 + \frac{\left(1+\frac{1}{2q-1}\right)d - \langle Y_S, Z_t^{-1}\rangle}{q - y(S)}\right)\\ &= \left(1 - \frac{a}{b - y(S)}\right)\cdot\left(1 + \frac{a + \frac{d}{2q-1}}{q - y(S)}\right)\\ &= 1 + \frac{(b-q)a - a^2 + \frac{(b-y(S))d}{2q-1} - \frac{ad}{2q-1}}{\big(b - y(S)\big)\cdot\big(q - y(S)\big)}\\ &\geq 1 + \frac{(b-q)a - a^2 + \frac{(b-q)d}{2q-1} - \frac{ad}{2q-1}}{\big(b - y(S)\big)\cdot\big(q - y(S)\big)},\end{aligned}$$

where the last inequality follows as $y(S) \leq q$. Let $f(x) = -x^2 + (b-q)x - \frac{dx}{2q-1} + \frac{(b-q)d}{2q-1}$ be a univariate quadratic function in x. Note that $f''(x) < 0$, and thus $\min_{x\in[x_1,x_2]} f(x)$ is attained at one of the two ends $x = x_1$ or $x = x_2$. Since $a = d - \langle Y_S, Z_t^{-1}\rangle \in [0,d]$, the numerator of the second term above is lower bounded by

$$\begin{aligned}f(a) &\geq \min_{x\in[0,d]} f(x) \geq \min\{f(0), f(d)\}\\ &= \min\left\{\frac{(b-q)d}{2q-1}, \frac{2qd(b-q-d)}{2q-1}\right\}\\ &= \min\left\{\frac{(d+\frac{1}{2})d}{2(b-d-1)}, \frac{(b-d-\frac{1}{2})d}{2(b-d-1)}\right\} \geq \frac{d}{4(b-d-1)},\end{aligned}$$

where the equality in the second line is by plugging in $q = b - d - \frac{1}{2}$, and the last inequality follows the assumption $b \geq d + 1$. Therefore, we conclude that

$$\begin{aligned}\det(Z_{t+1}) &\geq \det(Z_t)\left(1 + \frac{d}{4(b-d-1)\big(b-y(S)\big)\big(q-y(S)\big)}\right)\\ &\geq \det(Z_t)\left(1 + \frac{d}{4b^3}\right).\end{aligned}$$

□

The main result in this subsection follows immediately from Proposition 3.1.

Proof. (Proof of Theorem 1.1) Let $X^* = \sum_{i=1}^n x^*(i) \cdot v_i v_i^T$ be an optimal fractional solution for D-design with budget b for $x^* \in [0,1]^n$. Let $Z_1 \succ 0$ be an arbitrary initial solution.

When the combinatorial local search algorithm terminates at the τ-th iteration, the termination condition implies that $\det(Z_{\tau+1}) < \left(1 + \frac{d}{4b^3}\right) \det(Z_\tau)$. It follows from Proposition 3.1 with $X = X^*$ that

$$\det(Z_\tau)^{\frac{1}{d}} \geq \frac{b-d-1}{b} \cdot \det(X^*)^{\frac{1}{d}},$$

and thus the returned solution of the Fedorov's exchange method is an $\frac{b-d-1}{b}$-approximate solution.

Finally, we bound the time complexity of the algorithm. If the algorithm runs for $\tau > \frac{8b^3}{d} \ln \frac{\det(X^*)}{\det(Z_1)}$ iterations, then the termination condition implies that the determinant of $Z_{\tau+1}$ is at least

$$\det(Z_{\tau+1}) \geq \left(1 + \frac{d}{4b^3}\right)^\tau \det(Z_1) \geq e^{\frac{d\tau}{8b^3}} \det(Z_1) > \det(X^*),$$

where the second inequality follows as $(1 + \frac{d}{4b^3}) \geq e^{\frac{d}{8b^3}}$ for $\frac{d}{4b^3} \leq \frac{1}{4}$. It was proved in Appendix C of [28] that $\ln \frac{\det(X^*)}{\det(Z_1)}$ is polynomial in d, b and ℓ, which is the maximum number of bits to represent the numbers in the entries of the vectors. Specifically, they proved that $\det(Z_1) \geq 2^{-4(2b\ell+1)d^2}$ and $\det(X^*) \leq 2^{4(2n\ell+1)d^2}$, and so $\tau = O(db^3 n\ell)$ iterations of the algorithm is enough. □

3.2 Combinatorial Local Search Algorithm for A-Design

We analyze the following version of Fedorov's exchange method for A-design, where we always choose a pair that maximizes the improvement of the objective value and we stop as soon as the improvement is not large enough.

Fedorov's Exchange Method for A-Optimal Design

Input: n vectors $v_1, ..., v_n \in \mathbb{R}^d$, a budget $b \geq d$, and an accuracy parameter $\varepsilon \in (0,1)$.

1. Let $S_0 \subseteq [n]$ be an arbitrary set of full-rank vectors with $|S_0| = b$.
2. Let $t \leftarrow 1$ and $Z_1 \leftarrow \sum_{i \in S_0} v_i v_i^T$.
3. **Repeat**
   (a) Find $i_t \in S_{t-1}$ and $j_t \in [n] \setminus S_{t-1}$ such that
   $$(i_t, j_t) = \underset{\substack{(i,j):\\ i \in S_{t-1},\\ j \in [n]\setminus S_{t-1}}}{\arg\min} \operatorname{tr}\left(\left(Z_t - v_i v_i^T + v_j v_j^T\right)^{-1}\right).$$
   (b) Set $S_t \leftarrow S_{t-1} \cup \{j_t\} \setminus \{i_t\}$ and $Z_{t+1} \leftarrow Z_t - v_{i_t} v_{i_t}^T + v_{j_t} v_{j_t}^T$ and $t \leftarrow t+1$.

   Until $\operatorname{tr}(Z_t^{-1}) > \left(1 - \frac{\varepsilon}{b}\right) \operatorname{tr}(Z_{t-1}^{-1})$.
4. Return S_{t-2} as the solution set.

To analyze the change of the objective value in each iteration, we apply Lemma 2.6 which states that if $2\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle < 1$ then

(3.5)
$$\operatorname{tr}(Z_{t+1}^{-1}) - \operatorname{tr}(Z_t^{-1}) \leq \underbrace{\frac{\langle v_{i_t} v_{i_t}^T, Z_t^{-2}\rangle}{1 - 2\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle}}_{\text{loss}} - \underbrace{\frac{\langle v_{j_t} v_{j_t}^T, Z_t^{-2}\rangle}{1 + 2\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle}}_{\text{gain}}.$$

Therefore, to upper bound the A-design objective of the solution, we upper bound the loss term and lower bound the gain term to quantify the progress in each iteration.

In the following lemma, we first prove the existence of i_t with small loss term, with respect to a fractional solution x with $\|x\|_1 = q < b - 2d$. Note that we only restrict our choice of i_t to those vectors that satisfy $2\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle < 1$ so that (3.5) applies, clearly Fedorov's exchange method could only do better by considering all possible vectors in the current solution.

LEMMA 3.4. (LOSS) *For any $x \in [0,1]^n$ with $\sum_{i=1}^n x(i) = q < b - 2d$ and any $S \subseteq [n]$ with $|S| = b$, there exists $i \in S' := \{j \in S : 2\langle v_j v_j^T, Z^{-1}\rangle < 1\}$ with*

$$\frac{\langle v_i v_i^T, Z^{-2}\rangle}{1 - 2\langle v_i v_i^T, Z^{-1}\rangle} \leq \frac{\operatorname{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle}{b - x(S) - 2d}.$$

Proof. Consider the probability distribution of removing a vector v_i with probability

$$\Pr[i_t = i] = \frac{\left(1 - x(i)\right) \cdot \left(1 - 2\langle v_i v_i^T, Z^{-1}\rangle\right)}{\sum_{j \in S'} \left(1 - x(j)\right) \cdot \left(1 - 2\langle v_j v_j^T, Z^{-1}\rangle\right)}$$

for each $i \in S'$. We first check that the probability distribution is well-defined. Note that the numerator is non-negative as $1 - 2\langle v_i v_i^T, Z^{-1}\rangle > 0$ for each $i \in S'$. The denominator is

$$\begin{aligned}
&\sum_{j \in S'} \left(1 - x(j)\right)\left(1 - 2\langle v_j v_j^T, Z^{-1}\rangle\right) \\
&\geq \sum_{j \in S} \left(1 - x(j)\right) \cdot \left(1 - 2\langle v_j v_j^T, Z^{-1}\rangle\right) \\
&\geq \sum_{j \in S} \left(1 - x(j)\right) - 2\sum_{j \in S} \langle v_j v_j^T, Z^{-1}\rangle \\
&= b - x(S) - 2d \ > \ 0,
\end{aligned}$$

where the first inequality holds as $1 - 2\langle v_j v_j^T, Z^{-1}\rangle \leq 0$ for $j \in S \setminus S'$, the second inequality follows from

$1 - x(j) \le 1$ for each $j \in [n]$, and the equality is by $|S| = b$ and $\langle \sum_{j \in S} v_j v_j^T, Z^{-1} \rangle = \langle Z, Z^{-1} \rangle = d$, and the strict inequality is by the assumption $b > q + 2d \ge x(S) + 2d$. Thus, $\Pr[i_t = i] \ge 0$ for each $i \in S'$, and clearly $\sum_{i \in S'} \Pr[i_t = i] = 1$. The expected loss using this probability distribution is

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\langle v_{i_t} v_{i_t}^T, Z^{-2} \rangle}{1 - 2\langle v_{i_t} v_{i_t}^T, Z^{-1} \rangle}\right] \\
=&\sum_{i \in S'} \frac{\big(1-x(i)\big)\big(1-2\langle v_i v_i^T, Z^{-1}\rangle\big)}{\sum_{j \in S'} \big(1-x(j)\big)\big(1-2\langle v_j v_j^T, Z^{-1}\rangle\big)} \cdot \frac{\langle v_i v_i^T, Z^{-2}\rangle}{1-2\langle v_i v_i^T, Z^{-1}\rangle} \\
=&\frac{\sum_{i \in S'} \big(1 - x(i)\big)\langle v_i v_i^T, Z^{-2}\rangle}{\sum_{j \in S'} \big(1 - x(j)\big)\big(1 - 2\langle v_j v_j^T, Z^{-1}\rangle\big)} \\
\le&\frac{\operatorname{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle}{b - x(S) - 2d},
\end{aligned}
$$

where the last inequality follows from the inequality above for the denominator and

$$
\begin{aligned}
&\sum_{i \in S'} \big(1 - x(i)\big)\langle v_i v_i^T, Z^{-2}\rangle \le \sum_{i \in S} \big(1 - x(i)\big)\langle v_i v_i^T, Z^{-2}\rangle \\
=&\langle Z, Z^{-2}\rangle - \langle X_S, Z^{-2}\rangle = \operatorname{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle
\end{aligned}
$$

for the numerator. Therefore, there exists an $i \in S'$ with loss at most the expected value. ◻

Next we show the existence of j_t with large gain term, again with respect to a fractional solution x.

LEMMA 3.5. (GAIN) *For any $x \in [0,1]^n$ with $\sum_{i=1}^n x(i) = q < b$ and any $S \subseteq [n]$ with $|S| = b$ and $x(S) < q$, there exists $j \in [n] \setminus S$ with*

$$
\frac{\langle v_j v_j^T, Z^{-2}\rangle}{1 + 2\langle v_j v_j^T, Z^{-1}\rangle} \ge \frac{\langle X, Z^{-2}\rangle - \langle X_S, Z^{-2}\rangle}{q - x(S) + 2\langle X, Z^{-1}\rangle}.
$$

Proof. Consider the probability distribution of adding a vector v_j where each $j \in [n] \setminus S$ is sampled with probability

$$
\Pr[j_t = j] = \frac{x(j) \cdot \big(1 + 2\langle v_j v_j^T, Z^{-1}\rangle\big)}{\sum_{i \in [n]\setminus S} x(i) \cdot \big(1 + 2\langle v_i v_i^T, Z^{-1}\rangle\big)}
$$

for each $j \in [n]\setminus S$. Note that the denominator is positive by the assumption $x(S) < q$ which implies that $x([n] \setminus S) > 0$, and so the probability distribution is well-defined. The expected gain using this probability distribution is

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\langle v_{j_t} v_{j_t}^T, Z^{-2}\rangle}{1 + 2\langle v_{j_t} v_{j_t}^T, Z^{-1}\rangle}\right] \\
=&\sum_{j \in [n]\setminus S} \frac{x(j)\big(1 + 2\langle v_j v_j^T, Z^{-1}\rangle\big)}{\sum_{i \in [n]\setminus S} x(i)\big(1 + 2\langle v_i v_i^T, Z^{-1}\rangle\big)} \cdot \frac{\langle v_j v_j^T, Z^{-2}\rangle}{1 + 2\langle v_j v_j^T, Z^{-1}\rangle} \\
=&\frac{\sum_{j \in [n]\setminus S} x(j)\langle v_j v_j^T, Z^{-2}\rangle}{\sum_{i \in [n]\setminus S} x(i)\big(1 + 2\langle v_i v_i^T, Z^{-1}\rangle\big)} \\
=&\frac{\langle X, Z^{-2}\rangle - \langle X_S, Z^{-2}\rangle}{q - x(S) + \sum_{i \in [n]\setminus S} 2x(i)\langle v_i v_i^T, Z^{-1}\rangle} \\
\ge&\frac{\langle X, Z^{-2}\rangle - \langle X_S, Z^{-2}\rangle}{q - x(S) + 2\langle X, Z^{-1}\rangle},
\end{aligned}
$$

where the third equality is by $\sum_{i=1}^n x(i) = q$, and the last inequality holds as $\sum_{i \in [n]\setminus S} x(i) \cdot v_i v_i^T \preccurlyeq X$. Therefore, there exists $j \in [n]\setminus S$ with gain at least the expected value. ◻

We are about ready to analyze when the objective value would decrease. The following lemma will be used to bound the denominator of the gain term, and also to relate the numerator of the gain term to the current objective value $\operatorname{tr}(Z^{-1})$ (see the full version of the paper for a proof).

LEMMA 3.6. *For any given $d \times d$ positive definite matrices $A, B \succ 0$,*

$$
\langle A, B^2\rangle \ge \frac{(\operatorname{tr}(B))^2}{\operatorname{tr}\left(A^{-1}\right)} \quad \text{and} \tag{3.6}
$$

$$
\langle A, B\rangle \le \sqrt{\operatorname{tr}(A) \cdot \langle A, B^2\rangle}. \tag{3.7}
$$

The following is the main technical result for A-design, which lower bounds the improvement of the objective value in each iteration. Note that the result depends on $\operatorname{tr}(X) \cdot \operatorname{tr}(X^{-1})$.

PROPOSITION 3.2. (PROGRESS) *Let $x \in [0,1]^n$ be a fractional solution with $\sum_{i=1}^n x(i) = q$. Let Z_t be the current solution in the t-th iteration of Fedorov's exchange method. For any $\varepsilon > 0$, if*

$$
\begin{aligned}
&\operatorname{tr}(Z_t^{-1}) \ge (1+\varepsilon)\operatorname{tr}(X^{-1}) \quad \textit{and} \\
&b \ge q + 2d + 2(1+\varepsilon)\sqrt{\operatorname{tr}(X) \cdot \operatorname{tr}(X^{-1})},
\end{aligned}
$$

then $\operatorname{tr}\left(Z_{t+1}^{-1}\right) \le \left(1 - \frac{\varepsilon}{b}\right) \cdot \operatorname{tr}\left(Z_t^{-1}\right)$.

Proof. Let $S := S_{t-1}$ be the current solution set at time t. Note that $x(S) < q$, as otherwise $\operatorname{tr}(Z^{-1}) \le \operatorname{tr}(X^{-1})$ and the assumption does not hold. Hence, we can apply Lemma 3.5 to prove the existence of a $j_t \in [n]\setminus S$ such that the gain term is

$$\begin{aligned}\frac{\langle v_{j_t} v_{j_t}^T, Z^{-2}\rangle}{1+2\langle v_{j_t} v_{j_t}^T, Z^{-1}\rangle} &\geq \frac{\langle X, Z^{-2}\rangle - \langle X_S, Z^{-2}\rangle}{q - x(S) + 2\langle X, Z^{-1}\rangle} \\ &\geq \frac{\langle X, Z^{-2}\rangle - \langle X_S, Z^{-2}\rangle}{q - x(S) + 2\sqrt{\mathrm{tr}(X)\cdot\langle X, Z^{-2}\rangle}} \\ &\geq \frac{\frac{\mathrm{tr}(Z^{-1})^2}{\mathrm{tr}(X^{-1})} - \langle X_S, Z^{-2}\rangle}{q - x(S) + 2\sqrt{\mathrm{tr}(X)\cdot\frac{\mathrm{tr}(Z^{-1})^2}{\mathrm{tr}(X^{-1})}}} \\ &= \frac{\frac{\mathrm{tr}(Z^{-1})}{\mathrm{tr}(X^{-1})}\cdot \mathrm{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle}{q - x(S) + \frac{2\,\mathrm{tr}(Z^{-1})}{\mathrm{tr}(X^{-1})}\cdot\sqrt{\mathrm{tr}(X)\cdot\mathrm{tr}(X^{-1})}} \\ &\geq \frac{(1+\varepsilon)\,\mathrm{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle}{q - x(S) + 2(1+\varepsilon)\sqrt{\mathrm{tr}(X)\cdot\mathrm{tr}(X^{-1})}} \\ &\geq \frac{(1+\varepsilon)\,\mathrm{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle}{b - x(S) - 2d},\end{aligned}$$

where the second inequality is by (3.7), the third inequality follows from $\langle X, Z^{-2}\rangle \geq \frac{(\mathrm{tr}(Z^{-1}))^2}{\mathrm{tr}(X^{-1})}$ by (3.6) and an application of Claim 2.6 with $f(x) = \frac{x-c_1}{c_2+c_3\sqrt{x}}$ to establish monotonicity, the fourth inequality follows from the first assumption that $\mathrm{tr}(Z^{-1}) \geq (1+\varepsilon)\,\mathrm{tr}(X^{-1})$ and another application of Claim 2.6 with $g(x) = \frac{x-c_1}{c_2+c_3 x}$ to establish monotonicity, and the last inequality follows from the second assumption that $b \geq q + 2d + 2(1+\varepsilon)\sqrt{\mathrm{tr}(X)\cdot\mathrm{tr}(X^{-1})}$.

For the loss term, note that $q < b - 2d$ by the assumption on b, and so we can apply Lemma 3.4 to prove the existence of an $i_t \in S' \subseteq S$ such that the loss term is

$$\frac{\langle v_{i_t} v_{i_t}^T, Z^{-2}\rangle}{1-2\langle v_{i_t} v_{i_t}^T, Z^{-1}\rangle} \leq \frac{\mathrm{tr}(Z^{-1}) - \langle X_S, Z^{-2}\rangle}{b - x(S) - 2d}.$$

Since $i_t \in S'$ satisfies $2\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle < 1$, we can apply (3.5) to conclude that

$$\begin{aligned}\mathrm{tr}(Z_{t+1}^{-1}) - \mathrm{tr}(Z_t^{-1}) &= \mathrm{tr}\left((Z_t - v_{i_t} v_{i_t}^T + v_{j_t} v_{j_t}^T)^{-1}\right) - \mathrm{tr}\left(Z_t^{-1}\right) \\ &\leq \frac{\langle v_{i_t} v_{i_t}^T, Z^{-2}\rangle}{1-2\langle v_{i_t} v_{i_t}^T, Z^{-1}\rangle} - \frac{\langle v_{j_t} v_{j_t}^T, Z^{-2}\rangle}{1+2\langle v_{j_t} v_{j_t}^T, Z^{-1}\rangle} \\ &\leq \frac{-\varepsilon\,\mathrm{tr}(Z_t^{-1})}{b - x(S) - 2d} \leq -\frac{\varepsilon}{b}\,\mathrm{tr}(Z_t^{-1}).\end{aligned}$$

∎

The main result in this subsection follows from Proposition 3.2 by a simple scaling argument.

Proof. (Proof of Theorem 1.2) We consider the following scaled-down version y, Y of the fractional solution x, X.

$$q := b - 2d - 2(1+\varepsilon)\sqrt{\mathrm{tr}(X)\cdot\mathrm{tr}((X)^{-1})},$$
$$y := \frac{q}{b}\cdot x,\ \ Y := \sum_{i=1}^n y(i)\cdot v_i v_i^T = \frac{q}{b}\cdot X.$$

Note that $\mathrm{tr}(Y)\cdot\mathrm{tr}(Y^{-1}) = \mathrm{tr}(X)\cdot\mathrm{tr}(X^{-1})$ and so it holds that $b \geq q + 2d + 2(1+\varepsilon)\sqrt{\mathrm{tr}(Y)\cdot\mathrm{tr}(Y^{-1})}$. Thus, we can apply Proposition 3.2 on y to conclude that if the algorithm terminates at the τ-th iteration such that $\mathrm{tr}(Z_{\tau+1}^{-1}) > \left(1-\frac{\varepsilon}{b}\right)\mathrm{tr}\left(Z_\tau^{-1}\right)$ then

$$\begin{aligned}\mathrm{tr}\left(Z_\tau^{-1}\right) < (1+\varepsilon)\cdot\mathrm{tr}\left(Y^{-1}\right) &= \frac{(1+\varepsilon)b}{q}\cdot\mathrm{tr}\left(X^{-1}\right) \\ &\leq \left(1+O(\varepsilon)\right)\cdot\mathrm{tr}\left(X^{-1}\right),\end{aligned}$$

where the last inequality follows from the assumption $b = \Omega\left(\frac{1}{\varepsilon}\left(d + \sqrt{\mathrm{tr}(X)\,\mathrm{tr}(X^{-1})}\right)\right)$ which implies that $q \geq (1-O(\varepsilon))b$. This proves the approximation guarantee of the returned solution.

Finally, we bound the time complexity of the algorithm. If the algorithm runs for $\tau > \frac{b}{\varepsilon}\ln\frac{\mathrm{tr}(Z_1^{-1})}{\mathrm{tr}(X^{-1})}$ iterations, then the termination condition implies that the objective value of $Z_{\tau+1}$ is at most

$$\mathrm{tr}(Z_{\tau+1}^{-1}) \leq \left(1-\frac{\varepsilon}{b}\right)^\tau \mathrm{tr}\left(Z_1^{-1}\right) \leq e^{-\frac{\varepsilon\tau}{b}}\,\mathrm{tr}\left(Z_1^{-1}\right) \leq \mathrm{tr}\left(X^{-1}\right).$$

Note that $\ln\frac{\mathrm{tr}(Z_1^{-1})}{\mathrm{tr}(X^{-1})}$ is upper bounded by a polynomial in d, n and the input size as proved in [28] (and the corresponding bound for D-design is discussed in the proof of Theorem 1.1 in Section 3.1). ∎

As a corollary, we extend the analysis of Fedorov's exchange method in [28] to the more general without repetition setting (see the full version of the paper).

COROLLARY 3.1. *Let $x \in [0,1]^n$ be a fractional solution to the convex programming relaxation* (1.1) *for A-design with $\sum_{i=1}^n x(i) = b$. If $\|v_i\|^2 \leq \frac{\varepsilon^2 b}{2\,\mathrm{tr}(X^{-1})}$ for each $1 \leq i \leq n$ and $b \geq \Omega\left(\frac{d}{\varepsilon}\right)$ for some $\varepsilon \in (0,1)$, then Fedorov's exchange method for A-design returns a solution with at most b vectors with objective value at most $\left(1+O(\varepsilon)\right)\cdot\mathrm{tr}\left(X^{-1}\right)$ in polynomial time.*

4 D/A-Optimal Design with Knapsack Constraints

In this section, we propose the following randomized exchange algorithm to solve the D/A-optimal design problems with knapsack constraints.

Randomized Exchange Algorithm

Input: n vectors $u_1, ..., u_n \in \mathbb{R}^d$, an accuracy parameter $\varepsilon \in (0,1)$, and m knapsack constraints $c_j \in \mathbb{R}_+^n$ with budgets $b_j \geq \frac{d\|c_j\|_\infty}{\varepsilon}$ for all $j \in [m]$.

1. Solve the convex programming relaxation (1.1) for D-design or A-design and obtain an optimal solution $x \in [0,1]^n$. Let $X = \sum_{i=1}^n x(i)\cdot u_i u_i^T$.

2. Preprocessing: Let $v_i \leftarrow X^{-1/2} u_i$ for all $i \in [n]$, so that $\sum_{i=1}^n x(i) \cdot v_i v_i^T = I$.

3. Initialization: $t \leftarrow 1$, $S_0 \leftarrow \emptyset$, $\alpha \leftarrow 8\sqrt{d}$, and $k \leftarrow 2\|x\|_1 + 120d$.

4. Add i into S_0 independently with probability $x(i)$ for each $i \in [n]$. Let $Z_1 \leftarrow \sum_{i \in S_0} v_i v_i^T$.

5. While the termination condition is not satisfied and $t = O\left(\frac{k}{\varepsilon}\right)$ do the following, where the termination conditions for D-design and A-design are respectively

   $\det(Z_t)^{1/d} \geq 1 - 10\varepsilon$ and $\langle X^{-1}, Z_t^{-1} \rangle \leq (1+\varepsilon)\operatorname{tr}(X^{-1})$.

   (a) $S_t \leftarrow \text{Exchange}(S_{t-1})$.

   (b) Set $Z_{t+1} \leftarrow \sum_{i \in S_t} v_i v_i^T$ and $t \leftarrow t+1$.

6. Return S_{t-1} as the solution.

The exchange subroutine is described as follows.

Exchange Subroutine

1. Compute the action matrix $A_t := (\alpha Z_t - l_t I)^{-2}$, where $Z_t = \sum_{i \in S_{t-1}} v_i v_i^T$ and l_t is the unique scalar such that $A_t \succ 0$ and $\operatorname{tr}(A_t) = 1$.

2. Define $S_t' := \{i \in S_{t-1} \mid 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle \leq \frac{1}{2}\}$.

3. Sample $i_t \in S_{t-1}'$ from the following probability distribution

$$\Pr(i_t = i) = \frac{1 - x(i)}{k}\left(1 - 2\alpha\langle v_i v_i^T, A_t^{1/2}\rangle\right), \text{ for } i \in S_{t-1}',$$

$$\Pr(i_t = \emptyset) = 1 - \sum_{i \in S_{t-1}'} \frac{1 - x(i)}{k}\left(1 - 2\alpha\langle v_i v_i^T, A_t^{1/2}\rangle\right).$$

4. Sample $j_t \in [n] \backslash S_{t-1}$ from the following probability distribution

$$\Pr(j_t = j) = \frac{x(j)}{k}\left(1 + 2\alpha\langle v_j v_j^T, A_t^{1/2}\rangle\right), \text{ for } j \in [n]\backslash S_{t-1},$$

$$\Pr(j_t = \emptyset) = 1 - \sum_{j \in [n]\backslash S_{t-1}} \frac{x(j)}{k}\left(1 + 2\alpha\langle v_j v_j^T, A_t^{1/2}\rangle\right).$$

5. Return $S_t := S_{t-1} \cup \{j_t\} \backslash \{i_t\}$.

REMARK 4.1. *The randomized exchange algorithm is almost the same as the iterative randomized rounding algorithm in [26]. There are only two differences. One is that $\alpha \leftarrow 8\sqrt{d}$ instead of $\alpha \leftarrow \frac{\sqrt{d}}{\varepsilon}$ in [26]. The other is that the termination condition for E-design, which is $\lambda_{\min}(Z_t) \geq 1 - 2\varepsilon$, is replaced by the termination condition for D-design or the termination condition for A-design.*

The parameter α is used to control the approximation guarantee of the algorithm for E-design. If the termination condition is $\lambda_{\min}(Z_t) \geq \frac{3}{4}$, then it was proved in Theorem 2.3 (Theorem 3.8 of [26]) that the algorithm will terminate successfully in $O(k)$ steps with high probability.

Intuition and Proof Ideas Given the analysis of Fedorov's exchange method for D-design and A-design in Section 3, the most natural algorithm is to use the same distributions there for the rounding algorithm as well. We use D-design to illustrate the difficulty of analyzing this natural algorithm and to motivate the modifications made in the randomized exchange algorithm. By applying Lemma 2.5 repeatedly, for any $\tau \geq 1$,

$$\det(Z_{\tau+1}) \geq \det(Z_1) \prod_{t=1}^{\tau} (1 - v_{i_t}^T Z_t^{-1} v_{i_t})(1 + v_{j_t}^T Z_t^{-1} v_{j_t}).$$

Using the distributions $\Pr(i_t = i) \propto 1 - x_i$ and $\Pr(j_t = j) \propto x_j$ as in Section 3.1, Lemma 3.2 and Lemma 3.1 shows that there exist $i_t \in S_{t-1}$ and $j_t \notin S_{t-1}$ such that setting $S_t \leftarrow S_{t-1} - i_t + j_t$ will improve the D-design objective. However, if we randomly sample i_t and j_t from these distributions, we cannot prove that the objective value is consistently improving with good probability. For D-design, we are analyzing a product of random variables where each random variable could have a large variance, and existing martingale inequalities are not applicable to establish concentration of the product.

To bound the variance, one important observation is that when x is an optimal fractional solution, it follows from the optimality condition of the convex programming relaxation that any vector v_i with $x(i) \in (0,1)$ satisfies $\|v_i\|_2^2 \leq \varepsilon$. The current algorithm is motivated by the observation that if we can also lower bound the minimum eigenvalue of Z_t, then we can upper bound $v^T Z_t^{-1} v$ and this would allow us to establish concentration of the objective value. So our idea is to use the same algorithm in [26] for E-design to ensure that the minimum eigenvalue of Z_t is at least $\Omega(1)$ as mentioned in Remark 4.1. Surprisingly, we prove that sampling from the distributions for E-design can also improve the objective values for D-design and A-design, and this is particularly interesting for A-design where the minimum eigenvalue condition is needed to prove so. Having these in place, we can use Freedman's martingale inequality to prove that the objective values for D-design and A-design will be improving consistently after the minimum eigenvalue of the current solution is at least $\Omega(1)$.

Proof Outline and Organization In the analysis of the randomized exchange algorithm, we conceptually divide the algorithm into two phases. In the first phase, we show that the minimum eigenvalue of the current solution will reach $\frac{3}{4}$ in $O(k)$ iterations with high probability. In the second phase, we prove that the objective value for D/A-design will be $(1 \pm \varepsilon)$-approximation of the optimal in $O\left(\frac{k}{\varepsilon}\right)$ iterations with high probability. The following is a proof outline.

1. In Section 4.1.1, we first prove that the probability distributions in the exchange subroutine are well defined for $k = O(\|x\|_1 + d)$. This will be used to upper bound the number of iterations and the failure probability.
2. In Section 4.1.2, we prove that the minimum eigenvalue will reach $\frac{3}{4}$ in $O(k)$ iterations with high probability. Furthermore, the minimum eigenvalue will be at least $\frac{1}{4}$ during the next $\Theta\left(\frac{k}{\varepsilon}\right)$ iterations with good probability, for which we require the assumption that ε is not too small. The proofs are based on the regret minimization framework and the iterative randomized rounding algorithm developed in [3, 26].
3. In Section 4.2, we prove that the objective value of D-design will improve consistently with high probability. This is the more technical parts of the proof. We use the minimum eigenvalue condition in multiple places in the martingale concentration arguments. We also need the optimality conditions for the martingale concentration arguments. The analysis of the randomized exchange algorithm for A-design has a similar flavor, we refer the readers to the full version of the paper for more details.
4. In Section 4.1.3, we prove the main approximation results including Theorem 1.4 for experimental design, by combining the previous steps and using the concentration inequality for the knapsack constraints proved in [26]. We also prove Corollary 1.1 as an application of the main result.
5. Finally, in Section 4.1.4, we improve the success probability of the randomized exchange algorithm in some special cases, including the basic setting of having only one cardinality constraint, by combining with the algorithm in [26].

4.1 Analysis of the Common Algorithm The algorithm is identical for D-design and A-design except the termination condition. In this subsection, we will present the proofs of the common parts and the main results, and then present the specific proofs for D-design in Section 4.2.

4.1.1 Probability Distributions in the Exchange Subroutine In this subsection, we show that the probability distributions in the exchange subroutine are well defined for $k = O(\|x\|_1 + d)$, which will be used to upper bound the number of iterations and the failure probability of the algorithm (see the full version of the paper for a proof).

LEMMA 4.1. *The probability distributions in the randomized exchange algorithm are well-defined within* $O\left(\frac{k}{\varepsilon}\right)$ *iterations with probability at least* $1 - O\left(\frac{k}{\varepsilon} \cdot e^{-\Omega(d)}\right)$.

Henceforth, we assume that the randomized exchange algorithm is well-defined.

4.1.2 Lower Bounding Minimum Eigenvalue As discussed above, the minimum eigenvalue of Z_t plays a key role in our analysis of the algorithm. We conceptually divide the execution of the randomized exchange algorithm into two phases. In the first phase, we show that the minimum eigenvalue of the current solution will reach $\frac{3}{4}$ in $O(k)$ iterations with high probability.

PROPOSITION 4.1. *The probability that the randomized exchange algorithm has terminated successfully within* $16k$ *iterations or three exists* $\tau_1 \le 16k$ *with* $\lambda_{\min}(Z_{\tau_1}) \ge \frac{3}{4}$ *is at least* $1 - \exp(-\Omega(\sqrt{d}))$.

Proof. As noted in Remark 4.1, except for the termination condition, the randomized exchange algorithm is exactly the same as the algorithm in [26] with $\alpha = 8\sqrt{d}$. So, the proposition follows from Theorem 2.3 with $\gamma = \frac{1}{8}$ and $q = 2$. □

In the second phase, we prove that the minimum eigenvalue of Z_t is at least $\frac{1}{4}$ in the next $\Theta\left(\frac{k}{\varepsilon}\right)$ iterations with good probability.

PROPOSITION 4.2. *Suppose* $\lambda_{\min}(Z_{\tau_1}) \ge \frac{3}{4}$ *for some* τ_1. *In the randomized exchange algorithm, the probability that* $\lambda_{\min}(Z_t) \ge \frac{1}{4}$ *for all* $\tau_1 \le t \le \tau_1 + \frac{2k}{\varepsilon}$ *is at least* $1 - \frac{4k^2}{\varepsilon^2} \cdot e^{-\Omega(\sqrt{d})}$.

Proof. Consider the bad event that there exists a time $t \in [\tau_1, \tau_1 + \frac{2k}{\varepsilon}]$ with $\lambda_{\min}(Z_t) < \frac{1}{4}$. As the initial solution Z_{τ_1} satisfies $\lambda_{\min}(Z_{\tau_1}) \ge \frac{3}{4}$, there must exist a time period $[t_0, t_1] \subseteq [\tau_1, \tau_1 + \frac{2k}{\varepsilon})$ such that $\lambda_{\min}(Z_{t_0}) \ge \frac{3}{4}$, $\lambda_{\min}(Z_{t_1+1}) < \frac{1}{4}$, and $\lambda_{\min}(Z_t) \in [\frac{1}{4}, \frac{3}{4})$ for all $t \in [t_0 + 1, t_1]$.

We show that the decrease of the minimum eigenvalue from t_0 to t_1 implies that the sum of Δ_t defined in (2.4) has decreased significantly. Let $F_{t_0} = Z_{t_0}$ and $F_t = v_{j_t} v_{j_t}^T - v_{i_t} v_{i_t}^T$ for all $t \in [t_0 + 1, t_1]$. Note that the exchange subroutine ensures that $\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle \le \frac{1}{4}$ for any t. So, it follows from Corollary 2.1 with $\alpha = 8\sqrt{d}$ that

$$\frac{1}{4} > \lambda_{\min}(Z_{t_1+1}) \geq \sum_{t=t_0+1}^{t_1} \Delta_t - \frac{2\sqrt{d}}{\alpha} + \lambda_{\min}(Z_{t_0}) \geq \sum_{t=t_0+1}^{t_1} \Delta_t + \frac{1}{2}$$
$$\implies \quad \sum_{t=t_0+1}^{t_1} \Delta_t < -\frac{1}{4}.$$

On the other hand, Δ_t is expected to be positive when $\lambda_{\min}(Z_t) < \frac{3}{4}$. The expectation bound in Lemma 2.3 with $\tau' = t_0$, $\tau = t_1$, $\lambda \leq \frac{3}{4}$ and $\gamma = \frac{1}{8}$ implies that $\sum_{t=t_0+1}^{t_1} \mathbb{E}_t[\Delta_t] \geq \frac{t_1-t_0}{8k}$. So, the sum has a large deviation from the expectation,

$$\sum_{t=t_0+1}^{t_1} \Delta_t \leq \sum_{t=t_0+1}^{t_1} \mathbb{E}_t[\Delta_t] - \left(\frac{1}{4} + \frac{t_1 - t_0}{8k}\right).$$

We can apply the concentration bound in Lemma 2.3 with $\gamma = \frac{1}{8}$, $\lambda \leq \frac{3}{4}$ and $\eta = \frac{1}{4} + \frac{t_1-t_0}{8k}$ to upper bound this probability by

$$\Pr\left[\sum_{t=t_0+1}^{t_1} \Delta_t \leq \sum_{t=t_0+1}^{t_1} \mathbb{E}_t[\Delta_t] - \left(\frac{1}{4} + \frac{t_1 - t_0}{8k}\right)\right]$$
$$\leq \exp\left(-\frac{4\left(\frac{1}{4} + \frac{t_1-t_0}{8k}\right)^2 k\sqrt{d}}{(t_1 - t_0)(1 + \frac{3}{4} + \frac{1}{8}) + \left(\frac{1}{4} + \frac{t_1-t_0}{8k}\right)k/3}\right)$$
$$\leq \exp\left(-\Omega(\sqrt{d})\right),$$

where the last inequality follows as the denominator is in the order of $\Theta(k+t_1-t_0)$, and the numerator is in the order of $\Omega\left(1+\frac{t_1-t_0}{k}+\frac{(t_1-t_0)^2}{k^2}\right)\cdot k\sqrt{d} = \Omega(k+t_1-t_0)\cdot\sqrt{d}$. The proposition follows by applying the union bound over the at most $\frac{4k^2}{\varepsilon^2}$ possible pairs of t_0 and t_1 from time τ_1 to $\tau_1 + \frac{2k}{\varepsilon}$. ∎

4.1.3 Main Approximation Results

In this subsection, we prove the main approximation results for experimental design, including Theorem 1.4. We will do so by first assuming the following theorem about the improvement of the objective value in the second phase, which will be proved in Section 4.2 for D-design and in the full version of the paper for A-design.

THEOREM 4.2. *Suppose that* $\lambda_{\min}(Z_{\tau_1}) \geq \frac{3}{4}$ *and* $\lambda_{\min}(Z_t) \geq \frac{1}{4}$ *for* $t \geq \tau_1$. *For both D-design and A-design, if* $b_j \geq \frac{d\|c_j\|_\infty}{\varepsilon}$ *for all* $j \in [m]$ *for some* $\varepsilon \leq \frac{1}{100}$, *then the probability that the randomized exchange algorithm has not terminated by time* $\tau_1 + \frac{2k}{\varepsilon}$ *is at most* $e^{-\Omega(\sqrt{d})}$.

First, we prove the following bicriteria approximation result for D/A-design with knapsack constraints, by combining the previous steps and using the concentration inequality for the knapsack constraints proved in [26].

THEOREM 4.3. *Given* $\varepsilon \leq \frac{1}{100}$, *if* $b_j \geq \frac{d\|c_j\|_\infty}{\varepsilon}$ *for all* $j \in [m]$, *the randomized exchange algorithm returns a solution set* S *within* $16k + \frac{2k}{\varepsilon}$ *iterations such that*

$$\det\left(\sum_{i\in S} u_i u_i^T\right)^{\frac{1}{d}} \geq (1 - 10\varepsilon) \cdot \det(X)^{\frac{1}{d}} \quad \text{or}$$
$$\operatorname{tr}\left(\left(\sum_{i\in S} u_i u_i^T\right)^{-1}\right) \leq (1 + \varepsilon) \cdot \operatorname{tr}\left(X^{-1}\right)$$

for D-design and A-design respectively with probability at least $1 - O\left(\frac{k^2}{\varepsilon^2} \cdot e^{-\Omega(\sqrt{d})}\right)$, *where* X *is an optimal fractional solution. Moreover, for each* $j \in [m]$, *the solution set* S *satisfies*

$$c_j(S) \leq (1 + \varepsilon)b_j + 120d\,\|c_j\|_\infty \leq \left(1 + O(\varepsilon)\right)b_j$$

with probability at least $1 - e^{-\Omega(\varepsilon d)}$.

Proof. We start with defining some bad events for the randomized exchange algorithm.

- B_1: there exists some $t \leq 16k + \frac{2k}{\varepsilon}$ such that the sampling distributions are not well-defined.
- B_2: the algorithm has not terminated successfully within $16k$ iterations and $\tau_1 > 16k$ where τ_1 is the first time such that $\lambda_{\min}(Z_{\tau_1}) \geq \frac{3}{4}$.
- B_3: there exists some $\tau_1 \leq t \leq \tau_1 + \frac{2k}{\varepsilon}$ such that $\lambda_{\min}(Z_t) < 1/4$.
- B_4: the termination condition for D/A-design is not satisfied for all $\tau_1 \leq t \leq \tau_1 + \frac{2k}{\varepsilon}$.

If none of the bad events happens, then either the algorithm has terminated successfully within $16k$ iterations or the termination condition for D/A-design will be satisfied at some time $t \leq \tau_1 + \frac{2k}{\varepsilon} \leq 16k + \frac{2k}{\varepsilon}$. So, the probability that the randomized exchange algorithm has not satisfied the termination condition within $16k + \frac{2k}{\varepsilon}$ iterations is upper bounded by

$$\Pr[B_1 \cup B_2 \cup B_3 \cup B_4]$$
$$\leq \Pr[B_1] + \Pr[B_2 \cap \neg B_1] + \Pr[B_3 \cap \neg B_2 \cap \neg B_1] + \Pr[B_4 \cap \neg B_1 \cap \neg B_2 \cap \neg B_3]$$
$$\leq O\left(\frac{k}{\varepsilon}e^{-\Omega(d)}\right) + O\left(e^{-\Omega(\sqrt{d})}\right) + O\left(\frac{k^2}{\varepsilon^2}e^{-\Omega(\sqrt{d})}\right) + O\left(e^{-\Omega(\sqrt{d})}\right)$$
$$\leq O\left(\frac{k^2}{\varepsilon^2} \cdot e^{-\Omega(\sqrt{d})}\right),$$

where $\Pr[B_1]$ is bounded in Lemma 4.1, $\Pr[B_2 \cap \neg B_1]$ is bounded in Proposition 4.1, $\Pr[B_3 \cap \neg B_2 \cap \neg B_1]$ is bounded in Proposition 4.2, and $\Pr[B_4 \cap \neg B_1 \cap \neg B_2 \cap \neg B_3]$ is bounded in Theorem 4.2.

For D-design, since $v_i = X^{-\frac{1}{2}} u_i$, the termination condition implies the approximation guarantee as

$$\det\left(\sum_{i\in S} v_i v_i^T\right)^{\frac{1}{d}} > 1 - 10\varepsilon$$
$$\implies \quad \det\left(\sum_{i\in S} u_i u_i^T\right)^{\frac{1}{d}} \geq (1 - 10\varepsilon) \cdot \det(X)^{\frac{1}{d}}.$$

For A-design, note that

(4.8)
$$\left\langle X^{-1}, \Big(\sum_{i\in S} v_i v_i^T\Big)^{-1}\right\rangle = \left\langle X^{-1}, \Big(\sum_{i\in S} X^{-\frac{1}{2}} u_i u_i^T X^{-\frac{1}{2}}\Big)^{-1}\right\rangle$$
$$= \left\langle I, \Big(\sum_{i\in S} u_i u_i^T\Big)^{-1}\right\rangle = \operatorname{tr}\left(\Big(\sum_{i\in S} u_i u_i^T\Big)^{-1}\right),$$

and so the termination condition also implies the approximation guarantee as

$$\langle X^{-1}, (\textstyle\sum_{i\in S} v_i v_i^T)^{-1}\rangle \le (1+\varepsilon)\operatorname{tr}(X^{-1})$$
$$\implies \quad \operatorname{tr}\big((\textstyle\sum_{i\in S} u_i u_i^T)^{-1}\big) \le (1+\varepsilon)\operatorname{tr}(X^{-1}).$$

Finally, we consider the knapsack constraints. Note that the termination conditions of both D/A-design imply $\lambda_{\min}(Z_t) < 1$ before the algorithm terminates. So, we can apply Theorem 2.4 with $\gamma = \frac{1}{8}$ to conclude that the returned solution S satisfies

$$c_j(S) \le (1+\varepsilon)\langle c_j, x\rangle + 120d\,\|c_j\|_\infty$$
$$\le (1+\varepsilon)b_j + 120d\,\|c_j\|_\infty \le (1+O(\varepsilon))b_j$$

with probability at least $1-\exp(-\Omega(\varepsilon d))$, where the last inequality follows from $b_j \ge \frac{d\|c_j\|_\infty}{\varepsilon}$. □

We can prove the main theorem of this section Theorem 1.4 by turning the above bicriteria approximation result to a true approximation result with a simple scaling argument (see the full version of the paper).

4.1.4 Improving Success Probability for Unweighted Experimental Design

In Theorem 1.4, the success probability becomes zero when $k \ge \varepsilon \cdot e^{\Omega(\sqrt{d})}$. We can remove the dependency on k in some special cases by combining Theorem 1.4 with the result in [26]. This further leads to an improvement over the previous result on A-design with a single cardinality constraint by replacing the assumption $b \ge \Omega(\frac{d}{\varepsilon} + \frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon})$ in [33] with $b \ge \frac{2d}{\varepsilon}$, although there is a mild assumption on the range of ε. We refer the interested readers to the full version of the paper for more details.

4.2 Analysis of the D-Design Objective

We will prove Theorem 4.2 for D-design in this subsection. Let τ_1 be the start time of the second phase. For ease of notation, we simply reset $\tau_1 = 1$, as the first time step in the second phase. By assumption, $\lambda_{\min}(Z_1) \ge \frac{3}{4}$ and $\lambda_{\min}(Z_t) \ge \frac{1}{4}$ for all $t \ge 1$, which will be crucial in the analysis.

To analyze the objective value for D-design, our plan is to transform the product of random variables in Lemma 2.5 into a sum of random variables in the exponent as follows,

$$\det(Z_{\tau+1}) \ge \det(Z_1)\underbrace{\prod_{t=1}^{\tau}(1-\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle)(1+\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle)}_{(*)}.$$

Ideally, we would like to bound the $(*)$ term by

(4.9)
$$(*) \ge \exp\left(\sum_{t=1}^{\tau}(1-4\varepsilon)\underbrace{\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle}_{\text{gain } g_t} - (1+5\varepsilon)\underbrace{\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle}_{\text{loss } l_t}\right),$$

where the inequalities $1-x \ge e^{(1-4\varepsilon)x}$ and $1-x \ge e^{-(1+5\varepsilon)x}$ only hold when $x \in [0, 4\varepsilon]$ and ε is small enough such as $\varepsilon \le \frac{1}{50}$.

So, for our plan to work, we need to bound the gain term $\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle$ and the loss term $\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle$. To do so, we prove in Lemma 4.4 that in an optimal fractional solution x, every vector v_i with $0 < x(i) < 1$ satisfies the condition that $\|v_i\|_2^2 \le \varepsilon$. Note that, in the randomized exchange algorithm, all the vectors v_i with $x(i) = 1$ are added into the initial solution S_0 with probability one and will never be removed from the current solution, and all the vectors v_i with $x(i) = 0$ will never be added into the current solution. Therefore, Lemma 4.4 implies that $\|v_{i_t}\|_2^2 \le \varepsilon$ and $\|v_{j_t}\|_2^2 \le \varepsilon$ for all $t \ge 1$. Together with the assumption that $Z_t \succcurlyeq \frac{1}{4}I$ for all $t \ge 1$, we can ensure that $\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle \le 4\varepsilon$ and $\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle \le 4\varepsilon$ for all $t \ge 1$, and hence (4.9) holds.

Once this transformation is done and (4.9) is established, we can apply Freedman's martingale inequality to prove concentration of the exponent. In the following, we define the gain g_t, loss l_t and progress Γ_t in the t-th iteration as

$$g_t := \langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle, \quad l_t := \langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle, \quad \text{and}$$
$$\Gamma_t := (1-4\varepsilon)g_t - (1+5\varepsilon)l_t.$$

In Section 4.2.1, we will prove that the expected progress is large if the current solution is far from optimal. Then, in Section 4.2.2, we will prove that the total progress is concentrated around its expectation, where the minimum eigenvalue assumption is crucial in the martingale concentration argument. Finally, we finish the proof of Theorem 4.2 for D-design in Section 4.2.3, and present the proof of Lemma 4.4 in Section 4.2.4.

4.2.1 Expected Improvement of the D-Design Objective

The following Lemma bounds the conditional expectation of progress Γ_t, and shows that $\mathbb{E}_t[\Gamma_t]$ is large if the current objective value $\det(Z_t)^{\frac{1}{d}}$ is small. The proof follows by similar calculation as in the proofs of Lemma 3.1 and Lemma 3.2 (see the full version of the paper for more details).

LEMMA 4.2. *Let $\gamma \geq 1$. Let S_{t-1} be the solution set at time t and $Z_t = \sum_{i \in S_{t-1}} v_i v_i^T$ for $1 \leq t \leq \tau$. Suppose $\det(Z_t)^{\frac{1}{d}} \leq \lambda$ for $1 \leq t \leq \tau$. Then*

$$\sum_{t=1}^{\tau} \mathbb{E}_t[\Gamma_t] \geq \Big(\frac{1-4\varepsilon}{\lambda} - (1+5\varepsilon)\Big) \cdot \frac{d\tau}{k}.$$

4.2.2 Martingale Concentration Argument Here we show that the total progress is concentrated around the expectation. The proof uses the minimum eigenvalue assumption and the short vector condition from Lemma 4.4.

LEMMA 4.3. *Suppose $Z_t \succcurlyeq \frac{1}{4} I$ and $\|v_{i_t}\|_2^2 \leq \varepsilon$ and $\|v_{j_t}\|_2^2 \leq \varepsilon$ for $\varepsilon \leq \frac{1}{100}$ for all $1 \leq t \leq \tau$. Then, for any $\eta > 0$,*

$$\Pr\left[\sum_{t=1}^{\tau} \Gamma_t \leq \sum_{t=1}^{\tau} \mathbb{E}_t[\Gamma_t] - \eta\right] \leq \exp\Big(-\Omega\Big(\frac{\eta^2 k}{\varepsilon\tau d^{1.5} + \varepsilon\eta k}\Big)\Big).$$

Proof. We define two sequences of random variables $\{X_t\}_t$ and $\{Y_t\}_t$, where $X_t := \mathbb{E}_t[\Gamma_t] - \Gamma_t$ and $Y_t := \sum_{l=1}^{t} X_l$. It is easy to check that $\{Y_t\}_t$ is a martingale with respect to $\{S_t\}_t$. We will use Freedman's inequality to bound $\Pr(Y_\tau \geq \eta)$. To apply Freedman's inequality, we need to upper bound X_t and $E_t[X_t^2]$. Note that

$$0 \leq g_t = \langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle \leq 4\varepsilon \text{ and } 0 \leq l_t = \langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle \leq 4\varepsilon$$

by our assumptions that $Z_t \succcurlyeq \frac{1}{4} I$ and $\|v_{i_t}\|_2^2 \leq \varepsilon$ and $\|v_{j_t}\|_2^2 \leq \varepsilon$ for $1 \leq t \leq \tau$. These imply that

$$\begin{aligned} X_t = \mathbb{E}_t[\Gamma_t] - \Gamma_t &\leq (1-4\varepsilon)\mathbb{E}_t[g_t] + (1+5\varepsilon) l_t \\ &\leq (2+\varepsilon)\cdot 4\varepsilon \leq 10\varepsilon, \end{aligned}$$

where the last inequality holds for $\varepsilon \leq \frac{1}{2}$. To upper bound $\mathbb{E}_t[X_t^2]$, we first upper bound $\mathbb{E}_t[g_t]$ and $\mathbb{E}_t[l_t]$.

$$\begin{aligned} \mathbb{E}_t[g_t] &= \sum_{j \in [n]\setminus S_{t-1}} \frac{x(j)}{k}\big(1 + 2\alpha\langle v_j v_j^T, A_t^{\frac{1}{2}}\rangle\big)\langle v_j v_j^T, Z_t^{-1}\rangle \\ &\leq \frac{1+16\varepsilon\sqrt{d}}{k} \sum_{j \in [n]\setminus S_{t-1}} x(j)\langle v_j v_j^T, Z_t^{-1}\rangle \\ &\leq \frac{1+16\varepsilon\sqrt{d}}{k} \cdot \mathrm{tr}(Z_t^{-1}) \leq \frac{4d + 64\varepsilon d^{1.5}}{k}, \end{aligned}$$

where the first inequality holds as $\alpha = 8\sqrt{d}$, $A_t \preccurlyeq I$ and $\|v_j\|_2^2 \leq \varepsilon$ for $j \in [n] \setminus S_{t-1}$ with $x(j) > 0$, the second inequality follows as $\sum_{i=1}^{n} x(i) \cdot v_i v_i^T = I$, and the last inequality follows from the assumption that $Z_t \succcurlyeq \frac{1}{4} I$. We note that $\mathbb{E}_t[l_t] \leq \frac{d}{k}$, which follows by a similar proof of Lemma 3.1. So, we can upper bound $\mathbb{E}_t[X_t^2]$ by

$$\mathbb{E}_t[X_t^2] \leq 10\varepsilon\mathbb{E}_t[|X_t|] \leq O(\varepsilon)\cdot(\mathbb{E}_t[g_t] + \mathbb{E}_t[l_t]) \leq O\Big(\frac{\varepsilon d^{1.5}}{k}\Big),$$

where the first inequality is by the upper bound on X_t, and the last inequality is by the loose bound that $\mathbb{E}_t[g_t] \leq O\big(\frac{d^{1.5}}{k}\big)$. Therefore, $\sum_{t=1}^{\tau} \mathbb{E}_t[X_t^2] \leq O\big(\frac{\varepsilon\tau d^{1.5}}{k}\big)$.

Finally, we can apply Freedman's martingale inequality (Theorem 2.5) with $R = 10\varepsilon$ and $\sigma^2 = O\big(\frac{\varepsilon\tau d^{1.5}}{k}\big)$ to conclude that

$$\Pr(Y_\tau \geq \eta) \leq \exp\Big(-\frac{\eta^2/2}{\sigma^2 + R\eta/3}\Big) = \exp\Big(-\Omega\Big(\frac{\eta^2 k}{\varepsilon\tau d^{1.5} + \varepsilon\eta k}\Big)\Big).$$

The lemma follows by noting that $Y_\tau \geq \eta$ is equivalent to $\sum_{t=1}^{\tau} \Gamma_t \leq \sum_{t=1}^{\tau} \mathbb{E}_t[\Gamma_t] - \eta$. ∎

4.2.3 Proof of Theorem 4.2 for D-design We are ready to prove Theorem 4.2 for D-design. Let $\tau = \frac{2k}{\varepsilon}$. Suppose the second phase of the algorithm has not terminated by time τ. Then $\lambda = \max_{1 \leq t \leq \tau+1} \det(Z_t)^{\frac{1}{d}} < 1 - 10\varepsilon$. Thus, Lemma 4.2 implies that

$$\sum_{t=1}^{\tau} \mathbb{E}_t\left[\Gamma_t\right] \geq \Big(\frac{1-4\varepsilon}{\lambda} - (1+5\varepsilon)\Big) \cdot \frac{d\tau}{k} \geq \frac{\varepsilon d\tau}{k} = 2d.$$

On the other hand, the initial solution of the second phase satisfies $Z_1 \succcurlyeq \frac{3}{4} I$, which implies that $\det(Z_1) \geq \big(\frac{3}{4}\big)^d$. As the knapsack constraints satisfy $b_j \geq \frac{d\|c_j\|_\infty}{\varepsilon}$ for $j \in [m]$, we know from Lemma 4.4 that $\|v_i\|_2^2 \leq \varepsilon$ for each i with $0 < x(i) < 1$. Note that, in the randomized exchange algorithm, all i_t and j_t satisfy $0 < x(i_t) < 1$ and $0 < x(j_t) < 1$. Together with the assumption that $Z_t \succcurlyeq \frac{1}{4} I$ for all $1 \leq t \leq \tau$, we have $\langle v_{j_t} v_{j_t}^T, Z_t^{-1}\rangle \leq 4\varepsilon$ and $\langle v_{i_t} v_{i_t}^T, Z_t^{-1}\rangle \leq 4\varepsilon$ for all $1 \leq t \leq \tau$. Hence, we can apply (4.9) to deduce that

$$\begin{aligned} &1 > \det(Z_{\tau+1}) \geq \det(Z_1)\exp\left(\sum_{t=1}^{\tau}\Gamma_t\right) \geq \Big(\frac{3}{4}\Big)^d \exp\left(\sum_{t=1}^{\tau}\Gamma_t\right) \\ &\implies \quad \sum_{t=1}^{\tau} \Gamma_t \leq d\ln\frac{4}{3} \leq d. \end{aligned}$$

Therefore, we can apply Lemma 4.3 with $\eta = d$ and $\tau = \frac{2k}{\varepsilon}$ to conclude that

$$\begin{aligned} &\Pr\left[\max_{1 \leq t \leq \tau+1} \det(Z_t)^{\frac{1}{d}} < 1 - 10\varepsilon\right] \leq \Pr\left[\sum_{t=1}^{\tau}\Gamma_t < \sum_{t=1}^{\tau}\mathbb{E}_t[\Gamma_t] - d\right] \\ &\leq \exp\left(-\Omega\left(\frac{d^2 k}{\varepsilon\big(\frac{2k}{\varepsilon}\big)d^{1.5} + \varepsilon dk}\right)\right) \leq \exp(-\Omega(\sqrt{d})). \end{aligned}$$

4.2.4 Optimality Condition of the Convex Program for D-Design The following lemma uses the assumption about the budgets to prove that all vectors with fractional value are short.

LEMMA 4.4. *Let $x \in [0,1]^n$ be an optimal fractional solution of the convex programming relaxation* (1.1) *for D-design. Let $X = \sum_{i=1}^{n} x(i) \cdot u_i u_i^T$, and $v_i = X^{-\frac{1}{2}} u_i$ for $1 \leq i \leq n$. Suppose $b_j \geq \frac{d\|c_j\|_\infty}{\varepsilon}$ for $1 \leq j \leq m$. Then $\|v_i\|_2^2 \leq \varepsilon$ for each $1 \leq i \leq n$ with $0 < x(i) < 1$.*

Proof. We rewrite the convex programming relaxation (1.1) for D-design as follows:

$$\begin{aligned} \underset{x\in\mathbb{R}^d, X\in\mathbb{R}^{d\times d}}{\text{maximize}} \quad & \log\det(X) \\ \text{subject to} \quad & X = \sum_{i=1}^n x(i)\cdot u_i u_i^T, \\ & \langle c_j, x\rangle \le b_j, \quad \forall j\in[m], \\ & 0\le x(i)\le 1, \quad \forall i\in[n]. \end{aligned}$$

We will use a dual characterization to investigate the length of the vectors. We introduce a dual variable Y for the first equality constraint, a dual variable $\mu_j \ge 0$ for each of the budget constraint $b_j - \langle c_j, x\rangle \ge 0$, a dual variable $\beta_i^- \ge 0$ for each non-negative constraint $x(i) \ge 0$, and a dual variable $\beta_i^+ \ge 0$ for each capacity constraint $1-x(i)\ge 0$. The Lagrange function $L(x, X, Y, \mu, \beta^+, \beta^-)$ is defined as

$$\begin{aligned} &\log\det(X) + \left\langle Y, \sum_{i=1}^n x(i) u_i u_i^T - X\right\rangle + \sum_{j=1}^m \mu_j\left(b_j - \langle c_j, x\rangle\right) \\ &\qquad + \sum_{i=1}^n \beta_i^- x(i) + \sum_{i=1}^n \beta_i^+(1-x(i)) \\ &= \log\det(X) - \langle Y, X\rangle + \sum_{j=1}^m \mu_j b_j + \sum_{i=1}^n \beta_i^+ \\ &\qquad + \sum_{i=1}^n x(i)\left(\langle Y, u_i u_i^T\rangle - \sum_{j=1}^m \mu_j c_j(i) + \beta_i^- - \beta_i^+\right). \end{aligned}$$

The Lagrangian dual program is $\min_{Y,\mu\ge 0,\beta^+\ge 0,\beta^-\ge 0} \max_{x,X} L(x, X, Y, \mu, \beta^+, \beta^-)$. Note that we can assume that $Y \succcurlyeq 0$, as otherwise the inner maximization problem is unbounded above. Given $Y \succcurlyeq 0, \mu \ge 0, \beta^+ \ge 0, \beta^- \ge 0$, the maximizers x, X of the Lagrange function satisfy the optimality conditions that

$$\begin{aligned} &\nabla_X L = X^{-1} - Y = 0 \quad \text{and} \\ &\nabla_{x(i)} L = \langle Y, u_i u_i^T\rangle - \sum_{j=1}^m \mu_j c_j(i) + \beta_i^- - \beta_i^+ = 0. \end{aligned}$$

Thus, the Lagrangian dual program can be written as

$$\begin{aligned} \min_{\substack{Y\succcurlyeq 0,\mu\ge 0,\\ \beta^+\ge 0,\beta^-\ge 0}} \quad & \log\det\left(Y^{-1}\right) - d + \sum_{j=1}^m \mu_j b_j + \sum_{i=1}^n \beta_i^+ \\ \text{subject to} \quad & \langle Y, u_i u_i^T\rangle = \sum_{j=1}^m \mu_j c_j(i) - \beta_i^- + \beta_i^+, \ \forall i\in[n]. \end{aligned}$$

It is easy to verify that $x = \delta\vec{1}$ is a strictly feasible solution of the primal program for a small enough δ. So, Slater's condition implies that strong duality holds. Let x, X be an optimal solution for the primal program, and Y, μ, β^+, β^- be an optimal solution for the dual program. The Lagrangian optimality condition implies that $Y = X^{-1}$, and it follows that

$$\begin{aligned} &\log\det(X) = \log\det(X) - d + \sum_{j=1}^m \mu_j b_j + \sum_{i=1}^n \beta_i^+ \\ &\implies \sum_{j=1}^m \mu_j b_j \le d \implies \sum_{j=1}^m \mu_j \|c_j\|_\infty \le \varepsilon, \end{aligned}$$

where the last implication follows by the assumption $b_j \ge \frac{d\|c_j\|_\infty}{\varepsilon}$ for each $j\in[m]$.

Finally, by the complementary slackness conditions, we have $\beta_i^- \cdot x(i) = 0$ and $\beta_i^+\cdot(1-x(i)) = 0$ for each $i\in[n]$. Therefore, for each i with $0 < x(i) < 1$, we must have $\beta_i^+ = \beta_i^- = 0$, which implies that

$$\begin{aligned} &\sum_{j=1}^m \mu_j c_j(i) = \langle Y, u_i u_i^T\rangle = \langle X^{-1}, u_i u_i^T\rangle = \|v_i\|_2^2 \\ &\implies \qquad \|v_i\|_2^2 \le \sum_{j=1}^m \mu_j \|c_j\|_\infty \le \varepsilon. \end{aligned}$$

□

Acknowledgment

We would like to thank Tsz Chiu Kwok and Mohit Singh for many useful discussions and helpful comments on the writing.

References

[1] Zeyuan Allen-Zhu, Zhenyu Liao, Lorenzo Orecchia. *Spectral sparsification and regret minimization beyond matrix multiplicative updates*. In Proceedings of the forty-seventh annual ACM Symposium on Theory of Computing (STOC), 237–245, 2015.

[2] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. *Near-optimal design of experiments via regret minimization*. In Proceedings of the 34th International Conference on Machine Learning (ICML), 70:126–135, 2017.

[3] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. *Near-optimal discrete optimization for experimental design: A regret minimization approach*. Mathematical Programming, Jan 10:1–40. Springer, 2020.

[4] Dana Angluin. *Queries and concept learning*. Machine learning, 2(4):319–342, 1988.

[5] Anthony C Atkinson and Alexander N Donev. *Optimum experimental designs*. Clarendon Press, 1992.

[6] Anthony C Atkinson, Alexander N Donev, and Randall D Tobias. *Optimum experimental designs, with SAS*, Oxford Statistical Science Series Vol 34. Oxford University Press, 2007.

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

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

[9] Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail. *Sparse Features for PCA-like Linear Regression* In Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS), pages 2285–2293, 2011.

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

[11] Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, and Nisheeth Vishnoi.

Fair and Diverse DPP-Based Data Summarization. In Proceedings of the 34th International Conference on Machine Learning (ICML), pages 80:716–725, 2018.

[12] Luiz FO Chamon and Alejandro Ribeiro. *Greedy sampling of graph signals.* IEEE Transactions on Signal Processing, 66(1):34–47, IEEE, 2018.

[13] Siheng Chen, Aliaksei Sandryhaila, Jose MF Moura, and Jelena Kovacevic. *Signal recovery on graphs: Variation minimization.* IEEE Transactions on Signal Processing, 63(17):4609–4624, IEEE, 2015.

[14] Siheng Chen, Aliaksei Sandryhaila, Jose MF Moura, and Jelena Kovacevic. *Signal recovery on graphs: Fundamental limits of sampling strategies.* IEEE Transactions on Signal Processing, 2(4):539–554, IEEE, 2016.

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

[16] R Dennis Cook and Christopher J Nachtrheim. *A Comparison of Algorithms for Constructing Exact D-Optimal Designs.* Technometrics, 22(3):315–324. Taylor & Francis, 1980.

[17] Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang. *Matrix approximation and projective clustering via volume sampling.* Theory of Computing, 2(1):225–247, 2006.

[18] Amit Deshpande and Santosh Vempala. *Adaptive sampling and fast low-rank matrix approximation.* In Proceedings of the 9th international conference on Approximation Algorithms for Combinatorial Optimization Problems (APPROX), and 10th international conference on Randomization and Computation (RANDOM), pages 292–303, 2006.

[19] Valerii Vadimovich Fedorov. *Theory of optimal experiments.* Academic Press, 1972.

[20] David A. Freedman. *On tail probabilities for martingales.* Annals of Probability, 3(1):100–118, 1975.

[21] Arpita Ghosh and Stephen Boyd. *Growing well-connected graphs.* In Proceedings of the 45th IEEE Conference on Decision and Control (CDC), pages 6605–6611. IEEE, 2006.

[22] Arpita Ghosh, Stephen Boyd, and Amin Saberi. *Minimizing effective resistance of a graph.* SIAM review, 50(1):37–66, 2008.

[23] Peter Goos and Bradley Jones. *Optimal Design of Experiments: A Case Study Approach.* Wiley & Sons, 2011.

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

[25] Alexandra Kolla, Yury Makarychev, Amin Saberi, and Shang-Hua Teng. *Subgraph sparsification and nearly optimal ultrasparsifiers.* In Proceedings of the 42nd ACM symposium on Theory of computing (STOC), pages 57–66. ACM, 2010.

[26] Lap Chi Lau and Hong Zhou. *A Spectral Approach to Network Design.* arXiv preprint arXiv:2003.07810, 2020.

[27] Vivek Madan, Aleksandar Nikolov, Mohit Singh, and Uthaipon Tantipongpipat. *Maximizing Determinants under Matroid Constraints.* arXiv preprint arXiv:2004.07886, 2020.

[28] Vivek Madan, Mohit Singh, Uthaipon Tantipongpipat, and Weijun Xie. *Combinatorial Algorithms for Optimal Design.* In Proceedings of Conference on Learning Theory (COLT), pages 2210–2258, 2019.

[29] Alan J Miller and Nam-Ky Nguyen. *A Fedorov exchange algorithm for D-optimal design.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 43(4):669–677. Wiley-Blackwell, 1994.

[30] Baharan Mirzasoleiman. Big Data Summarization Using Submodular Functions. PhD Thesis, ETH Zurich, 2017.

[31] Nam-Ky Nguyen and Alan J Miller. *A review of some exchange algorithms for constructing discrete D-optimal designs.* Computational Statistics & Data Analysis, 14(4):489–498. Elsevier, 1992.

[32] Aleksandar Nikolov, and Mohit Singh. *Maximizing determinants under partition constraints.* In Proceedings of the forty-eighth annual ACM Symposium on Theory of Computing (STOC), 192–201, 2016.

[33] Aleksandar Nikolov, Mohit Singh, and Uthaipon Tao Tantipongpipat. *Proportional volume sampling and approximation algorithms for A-optimal design.* In Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 1369–1386. ACM-SIAM, 2019.

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

[35] Marco Di Summa, Friedrich Eisenbrand, Yuri Faenza, and Carsten Moldenhauer. *On Largest Volume Simplices and Sub-Determinants.* In Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 315–323. ACM-SIAM, 2015.

[36] Mohit Singh and Weijun Xie. *Approximate Positively Correlated Distributions and Approximation Algorithms for D-optimal Design.* In Proceedings of Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 2240-2255. ACM-SIAM, 2018.

[37] Joel A. Tropp. *Freedman's inequality for matrix martingales.* Electronic Communications in Probability, 16:262–270, 2011.

[38] Yining Wang, Adams Wei Yu, and Aarti Singh. *On computationally tractable selection of experiments in measurement-constrained regression models.* The Journal of Machine Learning Research, 18(1):5238–5278, 2017.

[39] William J Welch. *Algorithmic complexity: three NP-hard problems in computational statistics.* Journal of Statistical Computation and Simulation, 15(1):17–25, Taylor & Francis, 1982.