

Some Inapproximability Results of MAP Inference and Exponentiated Determinantal Point Processes

Naoto Ohsaka

CyberAgent, Inc.

40-1 Udagawacho Shibuya-ku, Tokyo, Japan

OHSAKA_NAOTO@CYBERAGENT.CO.JP

Abstract

We study the computational complexity of two hard problems on determinantal point processes (DPPs). One is *maximum a posteriori (MAP) inference*, i.e., to find a principal submatrix having the maximum determinant. The other is probabilistic inference on *exponentiated DPPs (E-DPPs)*, which can sharpen or weaken the diversity preference of DPPs with an exponent parameter p . We present several complexity-theoretic hardness results that explain the difficulty in approximating MAP inference and the normalizing constant for E-DPPs. We first prove that unconstrained MAP inference for an $n \times n$ matrix is NP-hard to approximate within a factor of $2^{\beta n}$, where $\beta = 10^{-10^{13}}$. This result improves upon the best-known inapproximability factor of $(\frac{9}{8} - \epsilon)$, and rules out the existence of any polynomial-factor approximation algorithm assuming $P \neq NP$. We then show that log-determinant maximization is NP-hard to approximate within a factor of $\frac{5}{4}$ for the unconstrained case and within a factor of $1 + 10^{-10^{13}}$ for the size-constrained monotone case. In particular, log-determinant maximization does not admit a polynomial-time approximation scheme unless $P = NP$. As a corollary of the first result, we demonstrate that the normalizing constant for E-DPPs of any (fixed) constant exponent $p \geq \beta^{-1} = 10^{10^{13}}$ is NP-hard to approximate within a factor of $2^{\beta p n}$, which is in contrast to the case of $p \leq 1$ admitting a fully polynomial-time randomized approximation scheme.

1. Introduction

Selecting a small set of “diverse” items from large data is an essential task in artificial intelligence. *Determinantal point processes (DPPs)* provide a probabilistic model on a discrete set that captures the notion of diversity using the matrix determinant (Macchi, 1975; Borodin & Rains, 2005). Suppose we are given n items (e.g., images or documents) associated with feature vectors $\{\phi_i\}_{i \in [n]}$ and an $n \times n$ Gram matrix \mathbf{A} such that $A_{i,j} = \langle \phi_i, \phi_j \rangle$ for all $i, j \in [n]$. The DPP defined by \mathbf{A} is a distribution over the power set $2^{[n]}$ such that the probability of drawing a subset $S \subseteq [n]$ is proportional to $\det(\mathbf{A}_S)$. Since $\det(\mathbf{A}_S)$ is equal to the squared volume of the parallelepiped spanned by $\{\phi_i\}_{i \in S}$, dissimilar items are likely to appear in the selected subsets, which ensures set diversity. DPPs exhibit fascinating properties that make them suitable for artificial intelligence and machine learning applications; e.g., many inference tasks are computationally tractable, including normalization, marginalization, and sampling, and efficient learning algorithms have been developed (see, e.g., the survey of Kulesza & Taskar, 2012 for details).

The present study aims at analyzing two exceptionally hard problems on DPPs through the lens of complexity theory—MAP inference and probabilistic inference on exponentiated DPPs.

Unconstrained MAP Inference. Seeking the most diverse subset that has the highest probability, i.e., *maximum a posteriori (MAP) inference*, is motivated by numerous applications, e.g., document summarization (Kulesza & Taskar, 2011; Chao et al., 2015; Perez-Beltrachini & Lapata, 2021), tweet timeline generation (Yao et al., 2016), YouTube video recommendation (Wilhelm et al., 2018), active learning (Biyik et al., 2019), and video summarization (Gong et al., 2014; Han et al., 2017). In particular, we focus on *unconstrained* MAP inference, which is equivalent to finding a principal submatrix with the maximum determinant, i.e., $\max_{S \subseteq [n]} \det(\mathbf{A}_S)$, and it has been computationally challenging despite its simplicity. Typically, the GREEDY algorithm is used as a heuristic, whereas the current best approximation algorithm achieves a factor of e^n (Nikolov, 2015). Kulesza and Taskar (2012) have shown that unconstrained MAP inference is NP-hard to approximate within a factor of $(\frac{9}{8} - \epsilon)$,¹ which is the best-known lower bound. On the other hand, *size-constrained* MAP inference (i.e., it must hold that $|S| = k$ for an input k) is NP-hard to approximate within an exponential factor of 2^{ck} for some $c > 0$ (Çivril & Magdon-Ismail, 2013), which does *not*, however, directly apply to the unconstrained case. Closing the gap between the lower bound ($\approx \frac{9}{8}$) and upper bound ($= e^n$) on unconstrained MAP inference is the first question addressed in this study.

Log-Determinant Maximization. In the artificial intelligence and machine learning communities, the performance of algorithms for MAP inference on DPPs is often evaluated in terms of the *logarithm* of the determinant (Gillenwater et al., 2012; Han & Gillenwater, 2020). So, we examine the approximability of the *log-determinant maximization* problem, i.e., $\max_{S \subseteq [n]} \log \det(\mathbf{A}_S)$. The unconstrained case is known to admit a 2-factor approximation algorithm (Buchbinder et al., 2015; Buchbinder & Feldman, 2018). If the minimum eigenvalue of \mathbf{A} is at least 1, the size-constrained case (i.e., $|S| = k$) is a special case of *monotone submodular maximization*, for which the GREEDY algorithm has an approximation factor of $\frac{e}{e-1} \approx 1.58$ (Han & Gillenwater, 2020; Sharma et al., 2015). On the other hand, explicit inapproximability results have not been known (to the best of our knowledge), which is the second question. Note that such hardness results for $\log \det(\mathbf{A}_S)$ and $\det(\mathbf{A}_S)$ are “incomparable” in the sense that a multiplicative approximation to one does *not* imply that to the other.

Exponentiated DPPs. Given an $n \times n$ positive semi-definite matrix \mathbf{A} , an *exponentiated DPP (E-DPP)* of exponent $p > 0$ defines a distribution whose probability mass for $S \subseteq [n]$ is proportional to $\det(\mathbf{A}_S)^p$ (Mariet et al., 2018). We can sharpen or weaken the diversity preference by tuning the value of exponent parameter p : increasing p prefers more diverse subsets than DPPs, setting $p = 0$ results in a uniform distribution, and E-DPPs coincide with DPPs if $p = 1$. Though computing the normalizing constant, i.e., $\sum_{S \subseteq [n]} \det(\mathbf{A}_S)^p$, lies at the core of efficient probabilistic inference on E-DPPs, it seems not to have a closed-form expression. Currently, some hardness results on *exact* computation are known if p is an even integer (Gurvits, 2005; Ohsaka & Matsuoka, 2020), and the case of $p \leq 1$ admits a fully polynomial-time randomized approximation scheme (FPRAS)² based on an approximate

1. We define approximation factor ρ so that $\rho \geq 1$. See Section 2 for the formal definition.
 2. An FPRAS is a randomized algorithm that outputs an e^ϵ -approximation with probability at least $\frac{3}{4}$ and runs in polynomial time in the input size and ϵ^{-1} .

Table 1: Computational complexity of MAP inference on DPPs, i.e., $\max_S \det(\mathbf{A}_S)$. Our result is $2^{\beta n}$ -factor inapproximability, improving the known lower bound of $\approx \frac{9}{8}$ and matching the best upper bound of e^n .

constrained?	inapproximability	approximability
unconstrained $S \subseteq [n]$	$2^{\beta n}$ ($\beta = 10^{-10^{13}}$) (this paper, Theorem 2.2) $\frac{9}{8} - \epsilon$ (Kulesza & Taskar, 2012)	e^n (Nikolov, 2015) $n!^2$ (Çivril & Magdon-Ismail, 2009)
size-constrained $S \subseteq [n], S = k$	2^{ck} (Koutis, 2006; Çivril & Magdon-Ismail, 2013) $(2^{\frac{1}{506}} - \epsilon)^k$ (Di Summa et al., 2014)	e^k (Nikolov, 2015) $k!^2$ (Çivril & Magdon-Ismail, 2009)

sampler (Anari et al., 2019; Robinson et al., 2019). The third question in this study is to find the value of p such that the normalizing constant is hard to approximate.

Our research questions in the present study can be summarized as follows:

- Q1.** *Is unconstrained MAP inference on DPPs exponentially inapproximable?*
- Q2.** *Is log-determinant maximization constant-factor inapproximable?*
- Q3.** *For what value of p is the normalizing constant for E-DPPs inapproximable?*

1.1 Our Contributions

We answer the above questions affirmatively by presenting three complexity-theoretic hardness results.

(Section 2) Exponential Inapproximability of Unconstrained MAP Inference on DPPs. Our first result is the following (cf. Table 1):

Theorem 2.2 (informal). *Unconstrained MAP inference on DPPs for an $n \times n$ matrix is NP-hard to approximate within a factor of $2^{\beta n}$, where $\beta = 10^{-10^{13}}$.*

This result significantly improves upon the best-known lower bound of Kulesza and Taskar (2012). Though the universal constant $\beta = 10^{-10^{13}}$ is extremely small, Theorem 2.2 justifies why any polynomial-factor approximation algorithm for unconstrained MAP inference has not been found. Our lower bound $2^{\beta n}$ matches the best upper bound e^n (Nikolov, 2015), up to a constant in the exponent. The proof is obtained by carefully extending the proof technique of Çivril and Magdon-Ismail (2013) to the unconstrained case.

(Section 3) Constant-Factor Inapproximability of Log-Determinant Maximization. We then obtain the following answer to our second question:

Theorems 3.1 and 3.2 (informal). *Log-determinant maximization is NP-hard to approximate within a constant factor. In particular,*

- *the unconstrained case is NP-hard to approximate within a factor of $\frac{5}{4}$, and*
- *the size-constrained case is NP-hard to approximate within a factor of $1 + 10^{-10^{13}}$, even when the minimum eigenvalue of an input matrix is at least 1.*

The above results imply that the existing algorithms’ approximation factors are tight, up to a constant. In particular, log-determinant maximization does not admit a polynomial-time approximation scheme (PTAS),³ unless $P = NP$.

(Section 4) Exponential Inapproximability of Exponentiated DPPs. Our third result is the following, which is derived by applying Theorem 2.2 (cf. Figure 1):

Corollary 4.2 (informal). *For every fixed number $p \geq \beta^{-1} = 10^{10^{13}}$, it is NP-hard to approximate the normalizing constant for E-DPPs of exponent p for an $n \times n$ matrix within a factor of $2^{\beta^{pn}}$. Moreover, we cannot generate a sample from E-DPPs of exponent p .*

This is the first inapproximability result regarding E-DPPs of constant exponent p and gives a new negative answer to open questions posed by Kulesza and Taskar (2012, Section 7.2) and Ohsaka and Matsuoka (2020, Section 6). The factor $2^{\beta^{pn}}$ is tight up to a constant in the exponent because $2^{\mathcal{O}(pn)}$ -factor approximation is possible in polynomial time (Observation 4.4). The latter statement means that in contrast to the case of $p \leq 1$, an efficient approximate sampler does not exist whenever $p \geq 10^{10^{13}}$. We stress that when applying a $(\frac{9}{8} - \epsilon)$ -factor inapproximability of Kulesza and Taskar (2012) instead of Theorem 2.2, we would be able to derive inapproximability only if $p = \Omega(n)$ (see Remark 4.3).

This article is an extended version of our conference paper presented at the 24th International Conference on Artificial Intelligence and Statistics (Ohsaka, 2021), which includes the following new results on log-determinant maximization in Section 3:

- We improve an inapproximability factor for the unconstrained case from $\frac{8}{7} \approx 1.143$ (Ohsaka, 2021, Remark 2.10) to $\frac{5}{4} = 1.25$ (Theorem 3.1).
- We prove a constant-factor inapproximability for the size-constrained monotone case (Theorem 3.2).

1.2 Related Work

MAP Inference on DPPs. In the theoretical computer science community, unconstrained MAP inference on DPPs is known as *determinant maximization* (DETMAX for

3. A PTAS is an approximation algorithm that takes a precision parameter $\epsilon > 0$ and outputs an e^ϵ -approximation in polynomial time in the input size when ϵ is fixed.

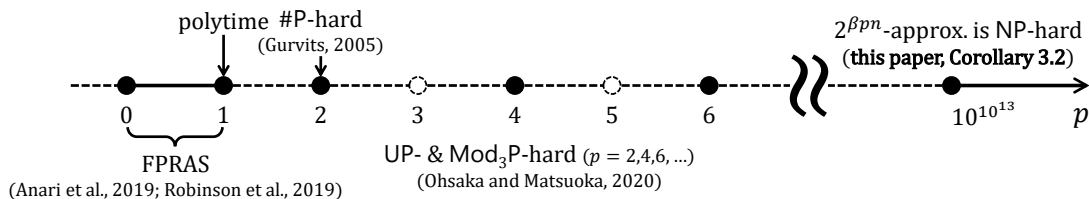


Figure 1: Computational complexity of the normalizing constant for exponentiated DPPs. Our result is $2^{\beta p n}$ -factor inapproximability for $p \geq 10^{10^{13}}$. Tractability for p 's on dashed lines and circles remains open.

short). The size-constrained version (k -DETMEX for short), which restricts the output to size k for parameter k , finds applications in computational geometry and discrepancy theory (Nikolov, 2015).

On the inapproximability side, Ko, Lee, and Queyranne (1995) prove that DETMAX and k -DETMEX are both NP-hard, and Koutis (2006) shows NP-hardness of approximating the largest k -simplex in a \mathcal{V} -polytope within a factor of 2^{ck} for some $c > 0$, implying that k -DETMEX is also exponentially inapproximable in k . Çivril and Magdon-Ismail (2013) directly prove a similar result for k -DETMEX. Di Summa, Eisenbrand, Faenza, and Moldenhauer (2014) study the special case of k -DETMEX, where k is fixed to the rank of an input matrix, which is still NP-hard to approximate within $(2^{\frac{1}{506}} - \epsilon)^k$ for any $\epsilon > 0$. Kulesza and Taskar (2012) use the reduction technique developed by Çivril and Magdon-Ismail (2009) to show an inapproximability factor of $(\frac{9}{8} - \epsilon)$ for DETMAX for any $\epsilon > 0$, which is the current best lower bound.

On the algorithmic side, Çivril and Magdon-Ismail (2009) prove that the GREEDY algorithm for k -DETMEX achieves an approximation factor of $k!^2 = 2^{\mathcal{O}(k \log k)}$. In their celebrated work, Nikolov (2015) gives an e^k -approximation algorithm for k -DETMEX; this is the current best approximation factor. Invoking the algorithm of Nikolov (2015) for all k immediately yields an e^n -approximation for DETMAX. The recent work focuses on DETMAX under complex constraints such as partition constraints (Nikolov & Singh, 2016) and matroid constraints (Madan et al., 2020).

In artificial intelligence and machine learning applications, unconstrained MAP inference is preferable if we do not (or cannot) prespecify the desired size of output, e.g., tweet timeline generation (Yao et al., 2016), object detection (Lee et al., 2016), change-point detection (Zhang & Ou, 2016), and others (Gillenwater et al., 2012; Han et al., 2017; Chen et al., 2018; Chao et al., 2015). Since the log-determinant as a set function $f(S) \triangleq \log \det(\mathbf{A}_S)$ for a positive semi-definite matrix \mathbf{A} is submodular,⁴ the GREEDY algorithm for monotone submodular maximization (Nemhauser et al., 1978) is widely used, which guarantees an $(\frac{e}{e-1})$ -factor approximation (with respect to f) under a size constraint if the minimum eigenvalue of \mathbf{A} is at least 1; i.e., f is a nonnegative, monotone, and submodular function (Han & Gillenwater, 2020; Sharma et al., 2015). Though the log-determinant f is *not* necessarily monotone, for which GREEDY has, in fact, a poor approximation guarantee, it works pretty

4. We say that a set function $f : 2^{[n]} \rightarrow \mathbb{R}$ is *submodular* if $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ for all $S, T \subseteq [n]$.

well in practice (Yao et al., 2016; Zhang & Ou, 2016). Several attempts have been made to scale up GREEDY (Han et al., 2017; Chen et al., 2018; Han & Gillenwater, 2020; Gartrell et al., 2021), whose naive implementation requires quartic time in n . Other than GREEDY, Gillenwater, Kulesza, and Taskar (2012) propose a gradient-based efficient algorithm with an approximation factor of 4 (with respect to f). Subsequently, a 2-factor approximation algorithm for unconstrained nonmonotone submodular maximization (including $\max_{S \subseteq [n]} f(S)$ as a special case) is developed (Buchbinder et al., 2015; Buchbinder & Feldman, 2018). Our results give constant-factor inapproximability for log-determinant maximization.

Exponentiated DPPs. We review known results on the computational complexity of the normalizing constant of E-DPPs, i.e., $\sum_{S \subseteq [n]} \det(\mathbf{A}_S)^p$ for $\mathbf{A} \in \mathbb{Q}^{n \times n}$, briefly appearing in Zou and Adams (2012); Gillenwater (2014). The case of $p = 1$ enjoys a simple closed-form expression that $\sum_{S \subseteq [n]} \det(\mathbf{A}_S) = \det(\mathbf{A} + \mathbf{I})$ (Kulesza & Taskar, 2012). Such a closed-form is unknown if $p < 1$, but a Markov chain Monte Carlo algorithm mixes in polynomially many steps thanks to their log-concavity (Anari et al., 2019; Robinson et al., 2019), implying an FPRAS.

On the other hand, the case of $p > 1$ seems a little more difficult. Kulesza and Taskar (2012) posed efficient computation of the normalizing constant for E-DPPs as an open question. Surprisingly, Gurvits (2005, 2009) has proven that computing $\sum_{S \subseteq [n]} \det(\mathbf{A}_S)^2$ for a P-matrix \mathbf{A} is #P-hard, but it is approximable within an e^n -factor (Anari & Gharan, 2017). Mariet, Sra, and Jegelka (2018) derive an upper bound on the mixing time of sampling algorithms parameterized by p and eigenvalues of \mathbf{A} . Ohsaka and Matsuoka (2020) derive UP-hardness and Mod₃P-hardness for every positive even integer $p = 2, 4, 6, \dots$. On the positive side, Ohsaka and Matsuoka (2020) develop $r^{\mathcal{O}(pr)} n^{\mathcal{O}(1)}$ -time algorithms for integer exponent p , where r is the rank or the treewidth of \mathbf{A} . Our study strengthens previous work by giving the first inapproximability result for every (fixed) constant exponent $p \geq 10^{10^{13}}$.

1.3 Notations and Definitions

For a positive integer n , let $[n] \triangleq \{1, 2, \dots, n\}$. For a finite set S and an integer k , we write $\binom{S}{k}$ for the family of all size- k subsets of S . The Euclidean norm is denoted $\|\cdot\|$; i.e., $\|\mathbf{v}\| \triangleq \sqrt{\sum_{i \in [d]} (v(i))^2}$ for a vector \mathbf{v} in \mathbb{R}^d . We use $\langle \cdot, \cdot \rangle$ for the standard inner product; i.e., $\langle \mathbf{v}, \mathbf{w} \rangle \triangleq \sum_{i \in [d]} v(i) \cdot w(i)$ for two vectors \mathbf{v}, \mathbf{w} in \mathbb{R}^d . For an $n \times n$ matrix \mathbf{A} and an index set $S \subseteq [n]$, we use \mathbf{A}_S to denote the principal submatrix of \mathbf{A} whose rows and columns are indexed by S . For a matrix \mathbf{A} in $\mathbb{R}^{n \times n}$, its *determinant* is defined as

$$\det(\mathbf{A}) \triangleq \sum_{\sigma \in \mathfrak{S}_n} \text{sgn}(\sigma) \prod_{i \in [n]} A_{i, \sigma(i)},$$

where \mathfrak{S}_n is the symmetric group on $[n]$, and $\text{sgn}(\sigma)$ is the sign of a permutation σ . We define $\det(\mathbf{A}_\emptyset) \triangleq 1$. For a set $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of n vectors in \mathbb{R}^d , the *volume* of the parallelepiped spanned by \mathbf{V} is defined as

$$\text{vol}(\mathbf{V}) \triangleq \|\mathbf{v}_1\| \cdot \prod_{2 \leq i \leq n} \left\| \mathbf{v}_i - \text{proj}_{\{\mathbf{v}_1, \dots, \mathbf{v}_{i-1}\}}(\mathbf{v}_i) \right\|, \quad (1)$$

where $\text{proj}_{\mathbf{P}}(\cdot)$ is an operator of orthogonal projection onto the subspace spanned by vectors in \mathbf{P} . We finally introduce the *AM-GM inequality*: for any n nonnegative real numbers x_1, \dots, x_n , we have

$$\frac{1}{n} \sum_{i \in [n]} x_i \geq \left(\prod_{i \in [n]} x_i \right)^{\frac{1}{n}}.$$

2. Exponential Inapproximability of Unconstrained MAP Inference

We prove an exponential-factor inapproximability result for unconstrained MAP inference on DPPs, which is identical to the following determinant maximization problem:

Definition 2.1. Given a positive semi-definite matrix \mathbf{A} in $\mathbb{Q}^{n \times n}$, *determinant maximization* (DETMAX) asks to find a subset S of $[n]$ such that the determinant $\det(\mathbf{A}_S)$ of a principal submatrix is maximized. The optimal value of DETMAX is denoted $\text{maxdet}(\mathbf{A}) \triangleq \max_{S \subseteq [n]} \det(\mathbf{A}_S)$.

We say that a polynomial-time algorithm ALG is a ρ -*approximation algorithm* for $\rho \geq 1$ if for all $\mathbf{A} \in \mathbb{Q}^{n \times n}$, it holds that

$$\det(\text{ALG}(\mathbf{A})) \geq \left(\frac{1}{\rho} \right) \cdot \text{maxdet}(\mathbf{A}),$$

where $\text{ALG}(\mathbf{A})$ is the output of ALG on \mathbf{A} . The factor ρ can be a function in the input size n , e.g., $\rho(n) = 2^n$, and (asymptotically) smaller ρ is a better approximation factor. We also define $[s(n), c(n)]$ -GAP-DETMAX for two functions $c(n)$ and $s(n)$ as a problem of deciding whether $\text{maxdet}(\mathbf{A}) \geq c(n)$ or $\text{maxdet}(\mathbf{A}) < s(n)$.⁵ If $[s(n), c(n)]$ -GAP-DETMAX is NP-hard, then so is approximating DETMAX within a factor of $\frac{c(n)}{s(n)}$.

We are now ready to state our result formally.

Theorem 2.2. *There exist universal constants λ_c and λ_s such that $\lambda_c - \lambda_s > 10^{-10^{13}}$ and $[2^{\lambda_s n}, 2^{\lambda_c n}]$ -GAP-DETMAX is NP-hard, where n is the order of an input matrix. In particular, it is NP-hard to approximate DETMAX within a factor of $2^{\beta n}$, where $\beta = 10^{-10^{13}}$.*

Remark 2.3. *The universal constant $\beta = 10^{-10^{13}}$ is so extremely small that $2^{\beta n} \approx 1$ for real-world matrices, whose possible size n is limited inherently. The significance of Theorem 2.2 is that it can rule out the existence of any polynomial-factor approximation algorithm (unless $\mathbf{P} = \mathbf{NP}$). As a corollary, we also show inapproximability for E-DPPs of constant exponent p .*

The input for DETMAX is often given by a *Gram matrix* $\mathbf{A} \in \mathbb{Q}^{n \times n}$, where $A_{i,j} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ for all $i, j \in [n]$ for n vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in \mathbb{Q}^d . In such a case, we have a simple relation between the principal minor and the volume of the parallelepiped that $\det(\mathbf{A}_S) = \text{vol}(\{\mathbf{v}_i\}_{i \in S})^2$ for every subset $S \subseteq [n]$. DETMAX is thus essentially equivalent to the following optimization problem:

5. Precisely, \mathbf{A} is “promised” to satisfy either $\text{maxdet}(\mathbf{A}) \geq c(n)$ (*yes instance*) or $\text{maxdet}(\mathbf{A}) < s(n)$ (*no instance*). Such a problem is called a *promise* problem.

Definition 2.4. Given a set $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of n vectors in \mathbb{Q}^d , *volume maximization* (VOLMAX) asks to find a subset \mathbf{S} of \mathbf{V} such that the volume $\text{vol}(\mathbf{S})$ is maximized. The optimal value of VOLMAX is denoted $\text{maxvol}(\mathbf{V}) \triangleq \max_{S \subseteq [n]} \text{vol}(\{\mathbf{v}_i\}_{i \in S})$.

Observe that there exists a $\rho(n)$ -approximation algorithm for DETMAX *if and only if* there exists a $\sqrt{\rho(n)}$ -approximation algorithm for VOLMAX.

Outline of the Remainder of Section 2. Section 2.1 introduces projection games to be reduced to VOLMAX and Raz's parallel repetition theorem. Section 2.2 reviews the indistinguishability of projection games. Section 2.3 describes Main Lemma, which is crucial in proving Theorem 2.2, and Section 2.4 is devoted to the proof of Main Lemma.

2.1 Projection Game and Parallel Repetition Theorem

We introduce projection games followed by the parallel repetition theorem.

Definition 2.5. A *2-player 1-round projection game* is specified by a quintuple $\mathfrak{G} = (X, Y, E, \Sigma, \Pi)$ such that the following conditions are satisfied:

- (X, Y, E) is a bipartite graph with vertex sets X and Y and an edge set E between X and Y ,
- Σ is an alphabet, and
- $\Pi = \{\pi_e\}_{e \in E}$ is a constraint set, where π_e for each edge $e \in E$ is a function $\Sigma \rightarrow \Sigma$.

A *labeling* σ is defined as a label assignment of each vertex of the bipartite graph, i.e., $\sigma : (X \uplus Y) \rightarrow \Sigma$. An edge $e = (x, y) \in E$ is said to be *satisfied* by σ if $\pi_e(\sigma(x)) = \sigma(y)$. The *value* of a projection game \mathfrak{G} , denoted $\text{val}(\mathfrak{G})$, is defined as the maximum fraction of edges satisfied over all possible labelings σ , i.e.,

$$\text{val}(\mathfrak{G}) \triangleq \max_{\sigma : (X \uplus Y) \rightarrow \Sigma} \frac{1}{|E|} \sum_{e=(x,y) \in E} \mathbb{1}[\pi_e(\sigma(x)) = \sigma(y)].$$

The LABELCOVER problem is defined as finding a labeling that satisfies the maximum fraction of edges in the bipartite graph of a projection game.

We then define the *product* of two games followed by the parallel repetition of a game.

Definition 2.6. Let $\mathfrak{G}_1 = (X_1, Y_1, E_1, \Sigma_1, \{\pi_{1,e}\}_{e \in E_1})$ and $\mathfrak{G}_2 = (X_2, Y_2, E_2, \Sigma_2, \{\pi_{2,e}\}_{e \in E_2})$ be two projection games. The *product* of \mathfrak{G}_1 and \mathfrak{G}_2 , denoted $\mathfrak{G}_1 \otimes \mathfrak{G}_2$, is defined as a new game $(X_1 \times X_2, Y_1 \times Y_2, E, \Sigma_1 \times \Sigma_2, \Pi = \{\pi_e\}_{e \in E})$, where $E \triangleq \{((x_1, x_2), (y_1, y_2)) \mid (x_1, y_1) \in E_1, (x_2, y_2) \in E_2\}$, and for each edge $e = ((x_1, x_2), (y_1, y_2)) \in E$, $\pi_e : \Sigma_1 \times \Sigma_2 \rightarrow \Sigma_1 \times \Sigma_2$ is defined as $\pi_e((i_1, i_2)) \triangleq (\pi_{1,(x_1,y_1)}(i_1), \pi_{2,(x_2,y_2)}(i_2))$ for labels $i_1 \in \Sigma_1$ and $i_2 \in \Sigma_2$.

The ℓ -*hold parallel repetition* of \mathfrak{G} for any positive integer ℓ is defined as

$$\mathfrak{G}^{\otimes \ell} \triangleq \underbrace{\mathfrak{G} \otimes \dots \otimes \mathfrak{G}}_{\ell \text{ times}}.$$

Raz (1998) proved the *parallel repetition theorem*, which states that for every (not necessarily projection) game \mathfrak{G} with $\text{val}(\mathfrak{G}) = 1 - \epsilon$, it holds that $\text{val}(\mathfrak{G}^{\otimes \ell}) \leq (1 - \bar{\epsilon})^{\frac{\ell}{\log |\Sigma|}}$, where

$\bar{\epsilon}$ is a constant depending only on ϵ . Once establishing an inapproximability factor for LABELCOVER, we can amplify it by taking parallel repetition. The use of the parallel repetition theorem has led to inapproximability results for many (NP-hard) optimization problems, such as SETCOVER (Feige, 1998) and MAXCLIQUE (Håstad, 1999). We refer to a tighter, explicit bound derived by Dinur and Steurer (2014).

Theorem 2.7 (Dinur & Steurer, 2014, Corollary 1). *For any projection game \mathfrak{G} with $\text{val}(\mathfrak{G}) \leq 1 - \epsilon$ for some $\epsilon > 0$, the ℓ -fold parallel repetition of \mathfrak{G} satisfies that*

$$\text{val}(\mathfrak{G}^{\otimes \ell}) \leq \left(1 - \frac{\epsilon^2}{16}\right)^\ell.$$

2.2 Indistinguishability of Projection Games

We review the indistinguishability of (the value of) projection games, in other words, inapproximability of LABELCOVER. The following theorem shows that we cannot decide in polynomial time whether a projection game has a value 1 or has a value less than $1 - \epsilon$ for some $\epsilon > 0$. Though its proof is widely known, we include it in Appendix A to explicitly describe the value of such ϵ .

Theorem 2.8 (See, e.g., Feige, 1998; Håstad, 2001; Trevisan, 2004; Vazirani, 2013; Tamaki, 2015). *Let $\mathfrak{G} = (X, Y, E, \Sigma, \Pi)$ be a projection game such that (X, Y, E) is a 15-regular bipartite graph (i.e., each vertex of $X \uplus Y$ is incident to exactly 15 edges), where $|X| = |Y| = 5n$ and $|E| = 75n$ for some positive integer n divisible by 3, and $|\Sigma| = 7$. Then, it is NP-hard to distinguish between $\text{val}(\mathfrak{G}) = 1$ and $\text{val}(\mathfrak{G}) < 1 - \frac{1}{206,401}$.*

A projection game satisfying the conditions in Theorem 2.8 is said to be *special* in this paper. Owing to Theorems 2.7 and 2.8, for any ℓ , it is NP-hard to decide whether the ℓ -fold parallel repetition $\mathfrak{G}^{\otimes \ell}$ of a special projection game satisfies $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$ or

$$\text{val}(\mathfrak{G}^{\otimes \ell}) < \left(1 - \frac{1}{(206,401)^2 \cdot 16}\right)^\ell < 2^{-2 \cdot 10^{-12} \ell}.$$

Hereafter, we let $\alpha \triangleq 2 \cdot 10^{-12}$.

2.3 Main Lemma and Proof of Theorem 2.2

The proof of Theorem 2.2 relies on a reduction from the ℓ -fold parallel repetition of a special projection game to VOLMAX. Throughout the remainder of this section, we fix the value of ℓ as $\ell \triangleq \left\lceil \frac{4}{\alpha} \right\rceil = 2 \cdot 10^{12}$. Our main lemma in the following can be thought of as an extension of Çivril and Magdon-Ismaïl (2013). The proof is deferred to the next subsection.

Lemma 2.9 (Main Lemma). *There is a polynomial-time reduction from the ℓ -fold parallel repetition $\mathfrak{G}^{\otimes \ell}$ of a special projection game to an instance $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ of VOLMAX such that $N = 2 \cdot (35n)^\ell$ for some integer n , each vector of \mathbf{V} is normalized (i.e., $\|\mathbf{v}_i\| = 1$ for all $i \in [N]$), and the following conditions are satisfied:*

- (Completeness) *If $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there exists a set \mathbf{S} of K vectors from \mathbf{V} with volume $\text{vol}(\mathbf{S}) = 1$, where $K = \frac{N}{7^\ell}$.*

- (Soundness) If $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$, then any set \mathbf{S} of k vectors from \mathbf{V} satisfies the following properties:
 1. $0 \leq k < \frac{4}{5}K$: $\text{vol}(\mathbf{S}) \leq 1$.
 2. $\frac{4}{5}K \leq k \leq N$: $\text{vol}(\mathbf{S}) < 2^{-\beta^\circ k}$, where $\beta^\circ = 10^{-10^{12.4}}$.

By Lemma 2.9, we can prove Theorem 2.2 as follows.

Proof of Theorem 2.2. Let $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ be an instance of VOLMAX reduced from the ℓ -fold parallel repetition $\mathfrak{G}^{\otimes \ell}$ by Lemma 2.9. Create a new instance of VOLMAX $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$, where $\mathbf{w}_i = 2^{\beta^\circ} \cdot \mathbf{v}_i$ for each $i \in [N]$ (which is a polynomial-time reduction as 2^{β° is constant). If $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there is a set \mathbf{S} of $K = N/7^\ell$ vectors from \mathbf{W} such that $\text{vol}(\mathbf{S}) = 2^{\beta^\circ K}$. On the other hand, if $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$, then $\text{maxvol}(\mathbf{W})$ is (strictly) bounded from above by $2^{\frac{4}{5}\beta^\circ K}$ through the following case analysis on the size of $\mathbf{S} \subseteq \mathbf{W}$:

- 1. $0 \leq |\mathbf{S}| < \frac{4}{5}K$: $\text{vol}(\mathbf{S}) < 2^{\frac{4}{5}\beta^\circ K}$.
- 2. $\frac{4}{5}K \leq |\mathbf{S}| \leq N$: $\text{vol}(\mathbf{S}) < 2^{(\beta^\circ - \beta^\circ)|\mathbf{S}|} \leq 1$.

It is thus NP-hard to decide whether $\text{maxvol}(\mathbf{W}) \geq 2^{\frac{\beta^\circ}{7^\ell} N}$ or $\text{maxvol}(\mathbf{W}) < 2^{\frac{4\beta^\circ}{5 \cdot 7^\ell} N}$ by Theorems 2.7 and 2.8. Owing to the relation between VOLMAX and DETMAX, $[2^{\frac{8\beta^\circ}{5 \cdot 7^\ell} n}, 2^{\frac{2\beta^\circ}{7^\ell} n}]$ -GAP-DETMAX is also NP-hard, where n is the order of an input matrix. In particular, it is NP-hard to approximate DETMAX within a factor of $2^{(\lambda_c - \lambda_s)n}$, where $\lambda_c = \frac{2\beta^\circ}{7^\ell}$ and $\lambda_s = \frac{8\beta^\circ}{5 \cdot 7^\ell}$. Observing that $\beta = 10^{-10^{13}} < 10^{-10^{12.7}} < \lambda_c - \lambda_s$ suffices to complete the proof. \square

2.4 Proof of Main Lemma

We now prove Main Lemma. We first introduce tools from Çivril and Magdon-Ismail (2013).

Lemma 2.10 (Çivril & Magdon-Ismail, 2013, Union Lemma). *Let \mathbf{P} and \mathbf{Q} be two (finite) sets of vectors in \mathbb{R}^d . Then, we have the following:*

$$\text{vol}(\mathbf{P} \cup \mathbf{Q}) \leq \text{vol}(\mathbf{Q}) \cdot \prod_{\mathbf{v} \in \mathbf{P}} d(\mathbf{v}, \mathbf{Q}),$$

where $d(\mathbf{v}, \mathbf{Q})$ denotes the distance of \mathbf{v} to the subspace spanned by \mathbf{Q} ; i.e.,

$$d(\mathbf{v}, \mathbf{Q}) \triangleq \|\mathbf{v} - \text{proj}_{\mathbf{Q}}(\mathbf{v})\|.$$

Lemma 2.11 (Çivril & Magdon-Ismail, 2013, Lemma 13). *For any positive integer ℓ , there exists a set of 2^ℓ vectors $\mathbf{B}^{(\ell)} = \{\mathbf{b}_1, \dots, \mathbf{b}_{2^\ell}\}$ of dimension $2^{\ell+1}$ such that the following conditions are satisfied:*

- Each element of vectors is either 0 or $2^{-\frac{\ell}{2}}$.
- $\|\mathbf{b}_i\| = 1$ for all $i \in [2^\ell]$.
- $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = \frac{1}{2}$ for all $i, j \in [2^\ell]$ with $i \neq j$.
- $\langle \mathbf{b}_i, \overline{\mathbf{b}_j} \rangle = \frac{1}{2}$ for all $i, j \in [2^\ell]$ with $i \neq j$, where $\overline{\mathbf{b}_j} \triangleq 2^{-\frac{\ell}{2}} \cdot \mathbf{1} - \mathbf{b}_j$. Note that $\langle \mathbf{b}_i, \overline{\mathbf{b}_i} \rangle = 0$.

Moreover, $\mathbf{B}^{(\ell)}$ can be constructed in time $\mathcal{O}(4^\ell)$.

Our Reduction. We explain how to reduce from special projection games to VOLMAX. Let $\mathfrak{G}^{\otimes \ell} = (X, Y, E, \Sigma, \Pi)$ be the ℓ -fold parallel repetition of a special projection game. By definition, (X, Y, E) is a 15^ℓ -regular bipartite graph, where $|X| = |Y| = (5n)^\ell$, $|E| = (75n)^\ell$, and $|\Sigma| = 7^\ell$ for some integer n . Assume that $\Sigma = [7^\ell]$ for notational convenience.

For each pair of a vertex of $X \uplus Y$ and a label of Σ , we define a vector as follows. Each vector consists of $|E|$ blocks, each of which is $2^{3\ell+1}$ -dimensional and is either a vector in the set $\mathbf{B}^{(3\ell)} = \{\mathbf{b}_1, \dots, \mathbf{b}_{2^{3\ell}}\}$ or the zero vector $\mathbf{0}$. Let $\mathbf{v}_{x,i}$ (resp. $\mathbf{v}_{y,i}$) denote the vector for a pair $(x, i) \in X \times \Sigma$ (resp. $(y, i) \in Y \times \Sigma$), and let $\mathbf{v}_{x,i}(e)$ (resp. $\mathbf{v}_{y,i}(e)$) denote the block of $\mathbf{v}_{x,i}$ (resp. $\mathbf{v}_{y,i}$) corresponding to edge $e \in E$. Each block is defined as follows:

$$\mathbf{v}_{x,i}(e) \triangleq \begin{cases} \frac{\overline{\mathbf{b}_{\pi_e(i)}}}{15^{\ell/2}} & \text{if } e \text{ is incident to } x, \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

$$\mathbf{v}_{y,i}(e) \triangleq \begin{cases} \frac{\mathbf{b}_i}{15^{\ell/2}} & \text{if } e \text{ is incident to } y, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Since each vector contains exactly 15^ℓ blocks chosen from $\mathbf{B}^{(3\ell)}$, it is normalized; i.e., $\|\mathbf{v}_{x,i}\| = \|\mathbf{v}_{y,i}\| = 1$ for all $x \in X, y \in Y, i \in \Sigma$. Note that \mathbf{v}_{x_1, i_1} and \mathbf{v}_{x_2, i_2} (resp. \mathbf{v}_{y_1, i_1} and \mathbf{v}_{y_2, i_2}) are orthogonal for any $x_1, x_2 \in X$ (resp. $y_1, y_2 \in Y$) and $i_1, i_2 \in \Sigma$ if $x_1 \neq x_2$ (resp. $y_1 \neq y_2$), and $\mathbf{v}_{x,i}$ and $\mathbf{v}_{y,j}$ for $x \in X, y \in Y, i, j \in \Sigma$ are orthogonal if $(x, y) \in E$ and $\pi_{(x,y)}(i) = j$ as $\langle \mathbf{v}_{x,i}, \mathbf{v}_{y,j} \rangle = \frac{1}{15^\ell} \langle \overline{\mathbf{b}_{\pi_{(x,y)}(i)}}, \mathbf{b}_j \rangle = 0$, or if $(x, y) \notin E$.

We then define an instance \mathbf{V} of VOLMAX as follows:

$$\mathbf{V} \triangleq \{\mathbf{v}_{x,i} \mid x \in X, i \in \Sigma\} \uplus \{\mathbf{v}_{y,i} \mid y \in Y, i \in \Sigma\}.$$

Here, \mathbf{V} contains $N \triangleq 2 \cdot (35n)^\ell$ vectors. Define $K \triangleq |X| + |Y| = 2 \cdot (5n)^\ell$; it holds that $K = N/7^\ell$. Construction of \mathbf{V} from $\mathfrak{G}^{\otimes \ell}$ can be done in polynomial time in n . In what follows, we show that \mathbf{V} satisfies the conditions listed in Main Lemma.

Completeness.

Lemma 2.12. *If $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there exists a set \mathbf{S} of K vectors from \mathbf{V} such that $\text{vol}(\mathbf{S}) = 1$.*

Proof. Let $\sigma : (X \uplus Y) \rightarrow \Sigma$ be an (optimal) labeling satisfying all the edges of E . For each edge $e = (x, y) \in E$, we have $\langle \mathbf{v}_{x, \sigma(x)}, \mathbf{v}_{y, \sigma(y)} \rangle = 0$ since $\pi_e(\sigma(x)) = \sigma(y)$. We further have $\langle \mathbf{v}_{x_1, \sigma(x_1)}, \mathbf{v}_{x_2, \sigma(x_2)} \rangle = 0$ for $x_1, x_2 \in X$ whenever $x_1 \neq x_2$, $\langle \mathbf{v}_{y_1, \sigma(y_1)}, \mathbf{v}_{y_2, \sigma(y_2)} \rangle = 0$ for $y_1, y_2 \in Y$ whenever $y_1 \neq y_2$, and $\langle \mathbf{v}_{x, \sigma(x)}, \mathbf{v}_{y, \sigma(y)} \rangle = 0$ for $x \in X, y \in Y$ whenever $(x, y) \notin E$. Hence, K vectors in the set defined as $\mathbf{S} \triangleq \{\mathbf{v}_{x, \sigma(x)} \mid x \in X\} \uplus \{\mathbf{v}_{y, \sigma(y)} \mid y \in Y\}$ are orthogonal to each other, implying $\text{vol}(\mathbf{S}) = 1$. \square

Soundness. Different from Çivril and Magdon-Ismail (2013), we need to bound the volume of every subset $\mathbf{S} \subseteq \mathbf{V}$. We consider two cases: **1.** $0 \leq |\mathbf{S}| < \frac{4}{5}K$ and **2.** $\frac{4}{5}K \leq |\mathbf{S}| \leq N$.

Soundness 1. $0 \leq |\mathbf{S}| < \frac{4}{5}K$.

Lemma 2.13. *Suppose $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$. For any set \mathbf{S} of less than $\frac{4}{5}K$ vectors from \mathbf{V} , it holds that $\text{vol}(\mathbf{S}) \leq 1$.*

Proof. The proof is a direct consequence of the fact that every vector of \mathbf{V} is normalized. \square

Soundness 2. $\frac{4}{5}K \leq |\mathbf{S}| \leq N$. For a set \mathbf{S} of vectors from \mathbf{V} , we use the following notations:

$$\begin{aligned} \mathbf{S}_X &\triangleq \{\mathbf{v}_{x,i} \in \mathbf{S} \mid x \in X, i \in \Sigma\}, & \mathbf{S}_Y &\triangleq \{\mathbf{v}_{y,i} \in \mathbf{S} \mid y \in Y, i \in \Sigma\}, \\ X(\mathbf{S}) &\triangleq \{x \in X \mid \exists i \in \Sigma, \mathbf{v}_{x,i} \in \mathbf{S}\}, & Y(\mathbf{S}) &\triangleq \{y \in Y \mid \exists i \in \Sigma, \mathbf{v}_{y,i} \in \mathbf{S}\}, \\ \text{rep}(\mathbf{S}_X) &\triangleq |\mathbf{S}_X| - |X(\mathbf{S})|, & \text{rep}(\mathbf{S}_Y) &\triangleq |\mathbf{S}_Y| - |Y(\mathbf{S})|. \end{aligned}$$

Here, $\text{rep}(\mathbf{S}_X)$ and $\text{rep}(\mathbf{S}_Y)$ mean how many times the same vertex appears (i.e., the number of repetitions) in the vectors of \mathbf{S}_X and \mathbf{S}_Y , respectively. The following lemma given by Çivril and Magdon-Ismail (2013) bounds the volume of \mathbf{S}_X and \mathbf{S}_Y in terms of $\text{rep}(\mathbf{S}_X)$ and $\text{rep}(\mathbf{S}_Y)$, respectively, which is proved in Appendix B for the sake of completeness.

Lemma 2.14 (Çivril & Magdon-Ismail, 2013, Lemma 16). *For any set \mathbf{S} of vectors from \mathbf{V} , it holds that*

$$\text{vol}(\mathbf{S}_X) \leq \left(\frac{\sqrt{3}}{2}\right)^{\text{rep}(\mathbf{S}_X)} \quad \text{and} \quad \text{vol}(\mathbf{S}_Y) \leq \left(\frac{\sqrt{3}}{2}\right)^{\text{rep}(\mathbf{S}_Y)}.$$

We first show that both $X(\mathbf{S})$ and $Y(\mathbf{S})$ contain $\Omega(k)$ vertices if their volume is sufficiently large.

Claim 2.15. *For any set \mathbf{S} of $k \geq \frac{4}{5}K$ vectors from \mathbf{V} , if $\text{vol}(\mathbf{S}) \geq 2^{-ck}$ for some number $c > 0$, then it holds that*

$$|X(\mathbf{S})| > \left(\frac{3}{8} - 10c\right)k \quad \text{and} \quad |Y(\mathbf{S})| > \left(\frac{3}{8} - 10c\right)k.$$

Proof. Observe first that $\text{vol}(\mathbf{S}_X) \geq \text{vol}(\mathbf{S}) \geq 2^{-ck}$. By Lemma 2.14, we have $(\sqrt{3}/2)^{\text{rep}(\mathbf{S}_X)} \geq \text{vol}(\mathbf{S}_X) \geq 2^{-ck}$, implying $\text{rep}(\mathbf{S}_X) \leq ck / \log_2(2/\sqrt{3}) < 5ck$. Similarly, we have $\text{rep}(\mathbf{S}_Y) < 5ck$. Using the facts that $|\mathbf{S}_X| = |X(\mathbf{S})| + \text{rep}(\mathbf{S}_X)$, $|\mathbf{S}_Y| = |Y(\mathbf{S})| + \text{rep}(\mathbf{S}_Y)$, and $k = |\mathbf{S}_X| + |\mathbf{S}_Y|$, we bound $|X(\mathbf{S})|$ from below as follows:

$$\begin{aligned} |X(\mathbf{S})| &= k - |\mathbf{S}_Y| - \text{rep}(\mathbf{S}_X) = k - |Y(\mathbf{S})| - \text{rep}(\mathbf{S}_Y) - \text{rep}(\mathbf{S}_X) \\ &> k - |Y| - 10ck = \left(1 - \frac{(5n)^\ell}{k} - 10c\right)k \\ &\geq \left(\frac{3}{8} - 10c\right)k, \end{aligned}$$

where the last inequality follows from the fact that $k \geq \frac{4}{5}K$ and $K = 2 \cdot (5n)^\ell$. Similarly, we have $|Y(\mathbf{S})| > (\frac{3}{8} - 10c)k$. \square

We now show that no vector set has a volume close to 1 if $\text{val}(\mathfrak{G}^{\otimes \ell})$ is small.

Lemma 2.16. *Suppose $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$. For any set \mathbf{S} of k vectors from \mathbf{V} with $k \geq \frac{4}{5}K$, it holds that $\text{vol}(\mathbf{S}) < 2^{-\beta^\circ k}$, where $\beta^\circ = 10^{-10^{12.4}}$.*

Proof. The proof is by contradiction. Suppose there exists a set \mathbf{S} of $k \geq \frac{4}{5}K$ vectors from \mathbf{V} such that $\text{vol}(\mathbf{S}) \geq 2^{-\beta^\circ k}$.

Consider a labeling $\sigma : (X \uplus Y) \rightarrow \Sigma$ defined as follows:

$$\sigma(z) \triangleq \begin{cases} \text{any } i \text{ such that } \mathbf{v}_{z,i} \in \mathbf{S}_X & \text{if } z \in X(\mathbf{S}), \\ \text{any } i \text{ such that } \mathbf{v}_{z,i} \in \mathbf{S}_Y & \text{if } z \in Y(\mathbf{S}), \\ \text{any element of } \Sigma & \text{otherwise.} \end{cases} \quad (2)$$

The choice of i 's can be arbitrary. Define $\mathbf{P} \triangleq \{\mathbf{v}_{x,\sigma(x)} \mid x \in X(\mathbf{S})\}$ and $\mathbf{Q} \triangleq \{\mathbf{v}_{y,\sigma(y)} \mid y \in Y(\mathbf{S})\}$. Our aim is to show that the volume of $\mathbf{P} \uplus \mathbf{Q} \subseteq \mathbf{S}$ is sufficiently small. To use Lemma 2.10, we bound the distance of the vectors of \mathbf{P} to \mathbf{Q} . Since \mathbf{Q} forms an orthonormal basis by construction and it holds that

$$\|\mathbf{v}_{x,\sigma(x)}\|^2 = \|\text{proj}_{\mathbf{Q}}(\mathbf{v}_{x,\sigma(x)})\|^2 + d(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q})^2$$

for each $x \in X(\mathbf{S})$, we have

$$d(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q}) = \sqrt{1 - \sum_{\mathbf{v}_{y,\sigma(y)} \in \mathbf{Q}} \langle \mathbf{v}_{x,\sigma(x)}, \mathbf{v}_{y,\sigma(y)} \rangle^2}.$$

If an edge $(x, y) \in E$ between $X(\mathbf{S})$ and $Y(\mathbf{S})$ is not satisfied by σ ; i.e., $\pi_{(x,y)}(\sigma(x)) \neq \sigma(y)$, then we have

$$\langle \mathbf{v}_{x,\sigma(x)}, \mathbf{v}_{y,\sigma(y)} \rangle = \left\langle \frac{\mathbf{b}_{\pi_{(x,y)}(\sigma(x))}}{15^{\ell/2}}, \frac{\mathbf{b}_{\sigma(y)}}{15^{\ell/2}} \right\rangle = \frac{1}{2 \cdot 15^\ell}.$$

Consequently, we obtain

$$d(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q}) = \left(1 - \frac{U(x)}{4 \cdot 15^{2\ell}}\right)^{\frac{1}{2}},$$

where $U(x)$ is defined as the number of unsatisfied edges between x and $Y(\mathbf{S})$. Using Lemma 2.10 and the fact that $\text{vol}(\mathbf{Q}) \leq 1$, we have

$$\begin{aligned} \text{vol}(\mathbf{P} \uplus \mathbf{Q}) &\leq \text{vol}(\mathbf{Q}) \cdot \prod_{x \in X(\mathbf{S})} d(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q}) \\ &\leq \left(\prod_{x \in X(\mathbf{S})} \left(1 - \frac{U(x)}{4 \cdot 15^{2\ell}}\right) \right)^{\frac{1}{2}} \\ &\leq \left(\frac{1}{|X(\mathbf{S})|} \sum_{x \in X(\mathbf{S})} \left(1 - \frac{U(x)}{4 \cdot 15^{2\ell}}\right) \right)^{\frac{|X(\mathbf{S})|}{2}}, \end{aligned} \quad (3)$$

where the last inequality is by the AM–GM inequality.

Now consider bounding $\sum_{x \in X(\mathbf{S})} U(x)$ from below, which is equal to the total number of unsatisfied edges between $X(\mathbf{S})$ and $Y(\mathbf{S})$ by σ . Substituting β° for c in Claim 2.15

derives $|X(\mathbf{S})| > (\frac{3}{8} - 10\beta^\circ)k$ and $|Y(\mathbf{S})| > (\frac{3}{8} - 10\beta^\circ)k$. Because less than $2^{-\alpha\ell}$ -fraction of edges in E can be satisfied by any labeling (including σ) by assumption, and more than $(\frac{3}{8} - 10\beta^\circ)k \cdot 15^\ell$ edges are incident to $X(\mathbf{S})$ (resp. $Y(\mathbf{S})$), the number of unsatisfied edges *incident to $X(\mathbf{S})$* (resp. $Y(\mathbf{S})$) is at least

$$\left(\frac{3}{8} - 10\beta^\circ\right) k \cdot 15^\ell - 2^{-\alpha\ell} \cdot (75n)^\ell = \left[\left(\frac{3}{8} - 10\beta^\circ\right) \frac{k}{(5n)^\ell} - 2^{-\alpha\ell}\right] (75n)^\ell. \quad (4)$$

Consequently, the number of unsatisfied edges *between $X(\mathbf{S})$ and $Y(\mathbf{S})$* is at least twice Eq. (4) minus “the number of unsatisfied edges incident to $X(\mathbf{S})$ or $Y(\mathbf{S})$ ” (which is at most $(75n)^\ell$); namely, we have

$$\sum_{x \in X(\mathbf{S})} U(x) \geq 2 \cdot (\text{value of Eq. (4)}) - (75n)^\ell \quad (5)$$

$$= \left[\left(\frac{3}{8} - 10\beta^\circ\right) \frac{2k}{(5n)^\ell} - 1 - 2^{-\alpha\ell+1}\right] (75n)^\ell \quad (6)$$

$$\geq \left[\left(\frac{1}{16} - 10\beta^\circ\right) \frac{2k}{(5n)^\ell} - 2^{-\alpha\ell+1}\right] (75n)^\ell, \quad (7)$$

where we have used the fact that $k \geq \frac{4}{5}K$ and $K = 2 \cdot (5n)^\ell$. With this inequality, we further expand Eq. (3) as

$$\begin{aligned} \text{vol}(\mathbf{P} \uplus \mathbf{Q}) &\leq \left(1 - \frac{\sum_{x \in X(\mathbf{S})} U(x)}{|X(\mathbf{S})|} \frac{1}{4 \cdot 15^{2\ell}}\right)^{\frac{|X(\mathbf{S})|}{2}} \\ &\leq \exp\left(-\frac{\left[\left(\frac{1}{16} - 10\beta^\circ\right) \frac{2k}{(5n)^\ell} - 2^{-\alpha\ell+1}\right] (75n)^\ell |X(\mathbf{S})|}{|X(\mathbf{S})| \cdot 8 \cdot 15^{2\ell}}\right) \\ &\leq \exp\left(-\left[\left(\frac{1}{16} - 10\beta^\circ\right) \frac{1}{4 \cdot 15^\ell} - 2^{-\alpha\ell+1} \frac{5}{64 \cdot 15^\ell}\right] k\right) \\ &= \exp\left(-\frac{1 - 160\beta^\circ - 5 \cdot 2^{-\alpha\ell+1}}{64 \cdot 15^\ell} k\right). \end{aligned}$$

Since $\beta^\circ = 10^{-10^{12.4}} < \frac{1-5 \cdot 2^{-\alpha\ell+1}}{64 \cdot 15^\ell \cdot \log_e(2) + 160}$ (recall that $\ell = \lceil \frac{4}{\alpha} \rceil$) and $\beta^\circ > 0$, we finally have $\text{vol}(\mathbf{S}) \leq \text{vol}(\mathbf{P} \uplus \mathbf{Q}) < 2^{-\beta^\circ k}$, a contradiction. \square

Proof of Main Lemma. Let $\mathfrak{G}^{\otimes \ell}$ be the ℓ -fold parallel repetition of a special projection game, and let \mathbf{V} be an instance of VOLMAX reduced from $\mathfrak{G}^{\otimes \ell}$ according to the procedure described in the beginning of this subsection. Then, the completeness follows from Lemma 2.12 and the soundness follows from Lemmas 2.13 and 2.16. \square

3. Constant-Factor Inapproximability for Log-Determinant Maximization

Here, we present constant-factor inapproximability results for log-determinant maximization. Let us begin with the definition of two optimization problems: Given a positive semi-definite matrix \mathbf{A} in $\mathbb{Q}^{n \times n}$, *log-determinant maximization* (LOGDETMAX) requests to find a

subset $S \subseteq [n]$ that maximizes $\log \det(\mathbf{A}_S)$, and *size-constrained log-determinant maximization* (k -LOGDETMAX) requests to find a size- k subset $S \in \binom{[n]}{k}$ that maximizes $\log \det(\mathbf{A}_S)$. Denote $\max\det_k(\mathbf{A}) \triangleq \max_{S \in \binom{[n]}{k}} \det(\mathbf{A}_S)$ and $\max\text{vol}_k(\mathbf{V}) \triangleq \max_{\mathbf{S} \in \binom{\mathbf{V}}{k}} \text{vol}(\mathbf{S})$ for a vector set \mathbf{V} .

Our first result is a $\frac{5}{4}$ -factor inapproximability for LOGDETMAX, which is an immediate consequence of the proof of Theorem 2.2.

Theorem 3.1. *It is NP-hard to approximate LOGDETMAX within a factor of $\frac{5}{4}$.*

Proof. The proof of Theorem 2.2 implies that determining whether $\log_2 \max\det(\mathbf{A}) \geq \lambda_c n$ or $\log_2 \max\det(\mathbf{A}) < \lambda_s n$ is NP-hard. Observing that $\frac{\lambda_c}{\lambda_s} = \frac{5}{4}$ is sufficient to complete the proof. \square

Our next result states that k -LOGDETMAX is inapproximable within a constant factor even if the log-determinant function is nonnegative, monotone, and submodular.

Theorem 3.2. *It is NP-hard to approximate k -LOGDETMAX within a factor of $1 + 10^{-10^{13}}$, even if the minimum eigenvalue of an input matrix \mathbf{A} is at least 1; namely, the set function $f(S) \triangleq \log \det(\mathbf{A}_S)$ is nonnegative, monotone, and submodular.*

The following lemma is essential for proving Theorem 3.2.

Lemma 3.3. *There is a polynomial-time reduction from the ℓ -fold parallel repetition $\mathfrak{G}^{\otimes \ell}$ of a special projection game to a set $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ of N vectors such that $N = 2 \cdot (35n)^\ell$ for some integer n , each vector of \mathbf{V} is normalized, and the following conditions are satisfied:*

- (Completeness) *If $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there exists a set \mathbf{S} of K vectors from \mathbf{V} such that $\text{vol}(\mathbf{S}) = 1$, where $K = \frac{N}{7^\ell}$.*
- (Soundness) *If $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$, then any set \mathbf{S} of K vectors from \mathbf{V} satisfies that $\text{vol}(\mathbf{S}) < 2^{-\beta^\circ K}$, where $\beta^\circ = 10^{-10^{12.4}}$.*
- (Minimum eigenvalue) *The minimum eigenvalue of the Gram matrix defined from \mathbf{V} is at least $\frac{1}{15^{\ell+1}}$.*

By Lemma 3.3, we can prove Theorem 3.2 as follows:

Proof of Theorem 3.2. Let \mathbf{V} be a set of N vectors reduced from the ℓ -fold parallel repetition $\mathfrak{G}^{\otimes \ell}$ by Lemma 3.3. Create a new vector set $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$, where $\mathbf{w}_i = (15^\ell + 1)^{1/2} \cdot \mathbf{v}_i$ for each $i \in [N]$. If $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there exists a set \mathbf{S} of $K = N/7^\ell$ vectors from \mathbf{W} such that $\text{vol}(\mathbf{S}) = (15^\ell + 1)^{K/2}$. On the other hand, if $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$, then $\max\text{vol}_k(\mathbf{W})$ is (strictly) bounded from above by $2^{-\beta^\circ K} \cdot (15^\ell + 1)^{K/2}$. By Theorems 2.7 and 2.8, it is NP-hard to decide whether $\max\text{vol}_k(\mathbf{W}) \geq (15^\ell + 1)^{K/2}$ or $\max\text{vol}_k(\mathbf{W}) < 2^{-\beta^\circ K} \cdot (15^\ell + 1)^{K/2}$.

Consider a Gram matrix \mathbf{A} derived from \mathbf{W} and a submodular set function $f : 2^{[N]} \rightarrow \mathbb{R}$ such that $f(S) \triangleq \log_2 \det(\mathbf{A}_S)$ for each $S \subseteq [N]$. Since \mathbf{A} has a minimum eigenvalue at least 1 by assumption, f is monotone (Sharma et al., 2015, Proposition 2), and thus is nonnegative. Since it is NP-hard to decide whether $\log_2 \max\det_k(\mathbf{A}) \geq \log_2((15^\ell + 1)^K)$

or $\log_2 \max \det_k(\mathbf{A}) < \log_2(2^{-2\beta^\circ K} \cdot (15^\ell + 1)^K)$, it is also NP-hard to approximate k -LOGDETMAX within a factor of

$$\frac{\log_2((15^\ell + 1)^K)}{\log_2(2^{-2\beta^\circ K} \cdot (15^\ell + 1)^K)} = \frac{\log_2(15^\ell + 1)}{\log_2(15^\ell + 1) - 2\beta^\circ} > 1 + 10^{-10^{13}}.$$

This completes the proof. \square

The remainder of this section is devoted to the proof of Lemma 3.3.

3.1 Proof of Lemma 3.3

Again, the proof of Lemma 3.3 relies on a reduction from the ℓ -fold parallel repetition of a special projection game. One might think that the vector set \mathbf{V} introduced in Section 2.3 can be used as it is; however, we have the following issue: Suppose there exist a vertex $x \in X$ and two labels $i_1, i_2 \in \Sigma$ such that $\pi_e(i_1) = \pi_e(i_2)$ for every edge e incident to x . Then, we have $\mathbf{v}_{x,i_1} = \mathbf{v}_{x,i_2}$; i.e., \mathbf{V} is *linearly dependent*. It then turns out that for a Gram matrix \mathbf{A} defined from \mathbf{V} , $\det(\mathbf{A}_\emptyset) = 1$ and $\det(\mathbf{A}) = 0$, implying that the set function $f_c(S) \triangleq c \log \det(\mathbf{A}_S)$ is never nonnegative nor monotone for any $c > 0$.

To circumvent this issue, we “modify” the reduction as follows. Let $\mathfrak{G}^{\otimes \ell} = (X, Y, E, \Sigma, \Pi)$ be the ℓ -fold parallel repetition of a special projection game. Recall that (X, Y, E) is a 15^ℓ -regular bipartite graph, where $|X| = |Y| = (5n)^\ell$, $|E| = (75n)^\ell$, and $|\Sigma| = 7^\ell$ for some integer n . Assume that $\Sigma = [7^\ell]$.

For each pair of a vertex of $X \uplus Y$ and a label of Σ , we define a vector as follows. Each vector consists of $|E|$ blocks, each of which is $2^{3\ell+1}$ -dimensional and is either a vector in the set $\mathbf{B}^{(3\ell)} = \{\mathbf{b}_1, \dots, \mathbf{b}_{2^{3\ell}}\}$ or the zero vector $\mathbf{0}$, and consists of $|X \uplus Y| \cdot |\Sigma|$ entries, each of which is a scalar. Let $\mathbf{v}_{x,i}$ (resp. $\mathbf{v}_{y,i}$) denote the vector for a pair $(x, i) \in X \times \Sigma$ (resp. $(y, i) \in Y \times \Sigma$), and let $\mathbf{v}_{x,i}(e)$ (resp. $\mathbf{v}_{y,i}(e)$) denote the block of $\mathbf{v}_{x,i}$ (resp. $\mathbf{v}_{y,i}$) corresponding to edge $e \in E$, let $v_{x,i}(z, j)$ (resp. $v_{y,i}(z, j)$) denote the entry of $\mathbf{v}_{x,i}$ (resp. $\mathbf{v}_{y,i}$) corresponding to pair $(z, j) \in (X \uplus Y) \times \Sigma$. Hence, each vector is $((75n)^\ell \cdot 2^{3\ell+1} + 2 \cdot (35n)^\ell)$ -dimensional. Each block and entry is defined as follows:

$$\mathbf{v}_{x,i}(e) \triangleq \begin{cases} \frac{\overline{\mathbf{b}_{\pi_e(i)}}}{(15^\ell + 1)^{1/2}} & \text{if } e \text{ is incident to } x, \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad v_{x,i}(z, j) \triangleq \begin{cases} \frac{1}{(15^\ell + 1)^{1/2}} & \text{if } (z, j) = (x, i), \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{v}_{y,i}(e) \triangleq \begin{cases} \frac{\mathbf{b}_i}{(15^\ell + 1)^{1/2}} & \text{if } e \text{ is incident to } y, \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad v_{y,i}(z, j) \triangleq \begin{cases} \frac{1}{(15^\ell + 1)^{1/2}} & \text{if } (z, j) = (x, i), \\ 0 & \text{otherwise.} \end{cases}$$

Since each vector contains exactly 15^ℓ blocks chosen from $\mathbf{B}^{(3\ell)}$ and exactly one element whose value is $\frac{1}{(15^\ell + 1)^{1/2}}$, it is normalized; i.e., $\|\mathbf{v}_{x,i}\| = \|\mathbf{v}_{y,i}\| = 1$ for all $x \in X, y \in Y, i \in \Sigma$. Note that \mathbf{v}_{x_1, i_1} and \mathbf{v}_{x_2, i_2} (resp. \mathbf{v}_{y_1, i_1} and \mathbf{v}_{y_2, i_2}) are orthogonal for any $x_1, x_2 \in X$ (resp. $y_1, y_2 \in Y$) and $i_1, i_2 \in \Sigma$ if $x_1 \neq x_2$ (resp. $y_1 \neq y_2$), and $\mathbf{v}_{x,i}$ and $\mathbf{v}_{y,j}$ for $x \in X, y \in Y, i, j \in \Sigma$ are orthogonal if $(x, y) \in E$ and $\pi_{(x,y)}(i) = j$, or if $(x, y) \notin E$.

We then define an instance \mathbf{V} of VOLMAX as follows:

$$\mathbf{V} \triangleq \{\mathbf{v}_{x,i} \mid x \in X, i \in \Sigma\} \uplus \{\mathbf{v}_{y,i} \mid y \in Y, i \in \Sigma\}.$$

Here, \mathbf{V} contains $N \triangleq 2 \cdot (35n)^\ell$ vectors. Define $K \triangleq |X| + |Y| = 2 \cdot (5n)^\ell$; it holds that $K = N/7^\ell$. Construction of \mathbf{V} from $\mathfrak{G}^{\otimes \ell}$ can be done in polynomial time in n .

Completeness.

Lemma 3.4. *If $\text{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there exists a set \mathbf{S} of K vectors from \mathbf{V} such that $\text{vol}(\mathbf{S}) = 1$.*

Proof. The proof can be done in a similar manner to that for Lemma 2.12. □

Soundness. For a vector set $\mathbf{S} \subseteq \mathbf{V}$, we use the same notation \mathbf{S}_X , \mathbf{S}_Y , $X(\mathbf{S})$, $Y(\mathbf{S})$, $\text{rep}(\mathbf{S}_X)$, and $\text{rep}(\mathbf{S}_Y)$ as those introduced in Section 2.4. We first prove the following, whose proof is similar to that for Lemma 2.14 and deferred to Appendix B.

Lemma 3.5. *For any set \mathbf{S} of vectors from \mathbf{V} , it holds that*

$$\text{vol}(\mathbf{S}_X) \leq \left(\frac{\sqrt{3.01}}{2} \right)^{\text{rep}(\mathbf{S}_X)} \quad \text{and} \quad \text{vol}(\mathbf{S}_Y) \leq \left(\frac{\sqrt{3.01}}{2} \right)^{\text{rep}(\mathbf{S}_Y)}.$$

By using Lemma 3.5, we show that both $X(\mathbf{S})$ and $Y(\mathbf{S})$ contain $\Omega(K)$ vertices if its volume is sufficiently large, whose proof is similar to that for Claim 2.15 and deferred to Appendix B.

Claim 3.6. *For any set \mathbf{S} of K vectors from \mathbf{V} , if $\text{vol}(\mathbf{S}) \geq 2^{-cK}$ for some number $c > 0$, then it holds that*

$$|X(\mathbf{S})| > \left(\frac{1}{2} - 10c \right) K \quad \text{and} \quad |Y(\mathbf{S})| > \left(\frac{1}{2} - 10c \right) K.$$

We now show that no vector set of size K has a volume close to 1 if $\text{val}(\mathfrak{G}^{\otimes \ell})$ is small, whose proof is similar to that for Lemma 2.16 and deferred to Appendix B.

Lemma 3.7. *Suppose $\text{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$. For any set \mathbf{S} of K vectors from V , it holds that $\text{vol}(\mathbf{S}) < 2^{-\beta^\circ K}$, where $\beta^\circ = 10^{-10^{12.4}}$.*

Eigenvalue. We finally bound the minimum eigenvalue of the Gram matrix from below.

Lemma 3.8. *The minimum eigenvalue of a Gram matrix defined from \mathbf{V} is at least $\frac{1}{15^\ell + 1}$.*

Proof. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a Gram matrix defined from \mathbf{V} ; i.e., it holds that $A_{i,j} \triangleq \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ for all $i, j \in [N]$. By the Courant–Fischer min-max theorem, the minimum eigenvalue of \mathbf{A} , denoted $\lambda_1(\mathbf{A})$, is equal to the minimum possible value of the Rayleigh quotient; namely,

$$\lambda_1(\mathbf{A}) = \min_{\|\mathbf{b}\|=1} \mathbf{b}^\top \mathbf{A} \mathbf{b}.$$

Recall that each vector \mathbf{v} of \mathbf{V} consists of blocks $\mathbf{v}(e)$ corresponding to edge $e \in E$ and entries $v(z, j)$ corresponding to pair $(z, j) \in (X \uplus Y) \times \Sigma$. Suppose an N -dimensional vector

\mathbf{b} is indexed by pairs of $(X \uplus Y) \times \Sigma$, that is, let $b(z, i)$ denote an entry of \mathbf{b} corresponding to pair $(z, i) \in (X \uplus Y) \times \Sigma$. By definition of vectors of \mathbf{V} , we have

$$\begin{aligned} \mathbf{b}^\top \mathbf{A} \mathbf{b} &\geq \sum_{(z,j) \in (X \uplus Y) \times \Sigma} \left(\sum_{(x,i) \in X \times \Sigma} b(x, i) \cdot v_{x,i}(z, j) + \sum_{(y,i) \in Y \times \Sigma} b(y, i) \cdot v_{y,i}(z, j) \right)^2 \\ &\geq \sum_{(z,j) \in (X \uplus Y) \times \Sigma} (b(z, j) \cdot v_{z,j}(z, j))^2 \\ &= \frac{1}{15^\ell + 1} \sum_{(z,j) \in (X \uplus Y) \times \Sigma} b(z, j)^2 = \frac{1}{15^\ell + 1} \|\mathbf{b}\|^2. \end{aligned}$$

Consequently, we have $\lambda_1(\mathbf{A}) \geq \min_{\|\mathbf{b}\|=1} \frac{1}{15^\ell + 1} \|\mathbf{b}\|^2 = \frac{1}{15^\ell + 1}$, which finishes the proof. \square

Proof of Lemma 3.3. Let $\mathfrak{G}^{\otimes \ell}$ be the ℓ -fold parallel repetition of a special projection game, and let \mathbf{V} be a vector set reduced by the procedure described in the beginning of this section. Then, the completeness follows from Lemma 3.4, the soundness follows from Lemma 3.7, and the minimum eigenvalue bound follows from Lemma 3.8. \square

4. Exponential Inapproximability for Exponentiated DPPs

We derive the exponential inapproximability of exponentiated DPPs. Given an $n \times n$ positive semi-definite matrix \mathbf{A} , the *exponentiated DPP (E-DPP)* of exponent $p > 0$ defines a distribution over the power set $2^{[n]}$, whose probability mass for each subset $S \subseteq [n]$ is proportional to $\det(\mathbf{A}_S)^p$. We use $\mathcal{Z}^p(\mathbf{A})$ to denote the normalizing constant; namely,

$$\mathcal{Z}^p(\mathbf{A}) \triangleq \sum_{S \subseteq [n]} \det(\mathbf{A}_S)^p.$$

$\mathcal{Z}^p(\mathbf{A})$ must be at least 1 by the positive semi-definiteness of \mathbf{A} . We say that an estimate $\widehat{\mathcal{Z}}^p$ is a ρ -approximation to \mathcal{Z}^p for $\rho \geq 1$ if it holds that

$$\mathcal{Z}^p \leq \widehat{\mathcal{Z}}^p \leq \rho \cdot \mathcal{Z}^p.$$

For two probability distributions $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ on Ω , the *total variation distance* is defined as

$$\frac{1}{2} \sum_{x \in \Omega} |\mu_x - \eta_x|$$

We first present the following theorem stating that assuming exponential inapproximability of DETMAX, we can neither estimate \mathcal{Z}^p and thus the probability mass for any subset accurately nor generate a random sample from E-DPPs in polynomial time for a sufficiently large p .

Theorem 4.1. *Suppose there exist universal constants λ_c and λ_s such that $[2^{\lambda_s n}, 2^{\lambda_c n}]$ -GAP-DETMAX is NP-hard. Then, for every fixed number $p > \frac{1}{\lambda_c - \lambda_s}$, it is NP-hard to approximate*

$\mathcal{Z}^p(\mathbf{A})$ for a positive semi-definite matrix \mathbf{A} in $\mathbb{Q}^{n \times n}$ within a factor of $2^{((\lambda_c - \lambda_s)p - 1)n}$. Moreover, unless $\text{RP} = \text{NP}$,⁶ no polynomial-time algorithm can draw a sample from a distribution whose total variation distance from E-DPPs of exponent $p > \frac{1}{\lambda_c - \lambda_s}$ is at most $\frac{1}{3}$.

As a corollary of Theorems 2.2 and 4.1, we have the following inapproximability result on E-DPPs.

Corollary 4.2. *For every fixed number $p \geq \beta^{-1} = 10^{10^{13}}$, it is NP-hard to approximate $\mathcal{Z}^p(\mathbf{A})$ for an $n \times n$ positive semi-definite matrix \mathbf{A} within a factor of $2^{\beta p n}$. Moreover, no polynomial-time algorithm can draw a sample whose total variation distance from the E-DPP of exponent p is at most $\frac{1}{3}$, unless $\text{RP} = \text{NP}$.*

Proof. The proof of Theorem 2.2 implies that $\lambda_c = \frac{2\beta^\circ}{7^\ell}$ and $\lambda_s = \frac{8\beta^\circ}{5 \cdot 7^\ell}$ satisfy the conditions in Theorem 4.1 and that $10^{10^{12.7}} > \frac{1}{\lambda_c - \lambda_s}$. Observe then that $(\lambda_c - \lambda_s)p - 1 > \beta p$ for $\beta = 10^{-10^{13}}$ and $p \geq 10^{10^{13}}$, which suffices to complete the proof. \square

Proof of Theorem 4.1. Consider the E-DPP of exponent $p > \frac{1}{\lambda_c - \lambda_s}$ defined by a positive semi-definite matrix $\mathbf{A} \in \mathbb{Q}^{n \times n}$. Suppose $p = \frac{q}{\lambda_c - \lambda_s}$ for some $q > 1$. We prove the first argument. If there exists a set $S \subseteq [n]$ such that $\det(\mathbf{A}_S) \geq 2^{\lambda_c n}$, then $\mathcal{Z}^p(\mathbf{A})$ is at least $2^{\lambda_c p n}$. On the other hand, if every set $S \subseteq [n]$ satisfies $\det(\mathbf{A}_S) < 2^{\lambda_s n}$, then $\mathcal{Z}^p(\mathbf{A})$ is less than $2^{\lambda_s p n + n}$. If a $2^{(q-1)n}$ -approximation to $\mathcal{Z}^p(\mathbf{A})$ is given, we can distinguish the two cases (i.e., we can solve $[2^{\lambda_s n}, 2^{\lambda_c n}]$ -GAP-DETMAX) because $2^{(q-1)n} = 2^{\lambda_c p n} / 2^{\lambda_s p n + n}$. We then prove the second argument. Assume that $\max \det(\mathbf{A}) \geq 2^{\lambda_c n}$. Sampling S from the E-DPP, we have “ $\det(\mathbf{A}_S) > 2^{\lambda_s n}$ ” (which is a certificate of the case) with probability at least $\frac{2^{\lambda_c p n}}{\mathcal{Z}^p(\mathbf{A})}$, and we have “ $\det(\mathbf{A}_S) \leq 2^{\lambda_s n}$ ” with probability at most $\frac{2^{\lambda_s p n + n}}{\mathcal{Z}^p(\mathbf{A})}$. Hence, provided a polynomial-time algorithm to generate a random sample whose total variation distance from the E-DPP is at most $\frac{1}{3}$, we can use it to find the certificate with probability at least

$$\left(1 - \frac{2^{\lambda_s p n + n}}{2^{\lambda_s p n + n} + 2^{\lambda_c p n}}\right) - \frac{1}{3} \geq \frac{2}{3} - \frac{1}{1 + 2^{q n - n}} \geq \frac{1}{2} \quad (\text{as long as } n \geq \frac{3}{q - 1}),$$

implying $\text{RP} = \text{NP}$. This completes the proof. \square

Remark 4.3. *Theorem 4.1 holds even when λ_c and λ_s are functions in n . If we apply a $(\frac{9}{8} - \epsilon)$ -factor inapproximability due to Kulesza and Taskar (2012), then we would obtain $\frac{1}{\lambda_c - \lambda_s} = \Theta(n)$; thus, p must be $\Omega(n)$, which is weaker than Corollary 4.2. Theorem 2.2 is crucial for ruling out approximability for constant p .*

We finally observe that a $2^{\mathcal{O}(pn)}$ -approximation to \mathcal{Z}^p can be derived using a $2^{\mathcal{O}(n)}$ -approximation algorithm for DETMAX (Nikolov, 2015). This means that Corollary 4.2 is tight up to a constant in the exponent (when $p \geq 10^{10^{13}}$).

Observation 4.4. *There exists a polynomial-time algorithm that approximates $\mathcal{Z}^p(\mathbf{A})$ for an $n \times n$ positive semi-definite matrix \mathbf{A} within a factor of $(2 \cdot e^p)^n$.*

6. RP is the class of decision problems for which there exists a probabilistic polynomial-time Turing machine that accepts a *yes* instance with probability $\geq \frac{1}{2}$ and always rejects a *no* instance. It is believed that $\text{RP} \neq \text{NP}$.

Proof. Fix an $n \times n$ positive semi-definite matrix \mathbf{A} in $\mathbb{Q}^{n \times n}$. Let $S \subseteq [n]$ be an e^n -approximation to DETMAX obtained by running the algorithm of Nikolov (2015) for every $k \in [n]$; i.e., it holds that

$$e^{-n} \cdot \maxdet(\mathbf{A}) \leq \det(\mathbf{A}_S) \leq \maxdet(\mathbf{A}).$$

Thanks to the positive semi-definiteness of \mathbf{A} , it is easily verified that

$$\frac{1}{2^n} \cdot \mathcal{Z}^p(\mathbf{A}) \leq \maxdet(\mathbf{A})^p \leq \mathcal{Z}^p(\mathbf{A}).$$

Combining two inequalities, we have

$$\frac{1}{\rho} \cdot \mathcal{Z}^p(\mathbf{A}) \leq \det(\mathbf{A}_S)^p \leq \mathcal{Z}^p(\mathbf{A}),$$

where $\rho \triangleq 2^n e^{pn}$. Hence, $\rho \cdot \det(\mathbf{A}_S)^p$ is a ρ -approximation of $\mathcal{Z}^p(\mathbf{A})$, completing the proof. \square

5. Concluding Remarks and Open Questions

We have established three inapproximability results for MAP inference on DPPs and the normalizing constant for exponentiated DPPs. We conclude this paper with three open questions.

- **Optimal bound for unconstrained MAP inference.** The universal constant $\beta = 10^{-10^{13}}$ in Theorem 2.2 seems extremely small despite e^n -factor approximability (Nikolov, 2015); improving the value of β is a potential research direction.
- **Complexity class of log-determinant maximization.** Theorems 3.1 and 3.2 state that LOGDETMAX is $\mathcal{O}(1)$ -factor inapproximable, while LOGDETMAX is known to be in the class APX.⁷ Is LOGDETMAX APX-complete?
- **Smallest exponent p for which \mathcal{Z}^p is inapproximable.** Our upper bound $10^{10^{13}}$ on the exponent p in Corollary 4.2 is surprisingly large. Can we find a (smaller) “threshold” p_c such that \mathcal{Z}^p is approximable if $p \leq p_c$ and inapproximable otherwise?

Acknowledgments

This work was mostly done while the author was at NEC Corporation. The author would like to thank Leonid Gurvits for pointing out to us the reference (Gurvits, 2009), and thank the referees for helpful suggestions on the presentation of the paper.

7. APX is the class of NP optimization problems that admit constant-factor approximation algorithms.

Appendix A. Proof of Theorem 2.8

The proof of Theorem 2.8 is based on a series of gap-preserving reductions from MAX-3SAT to LABELCOVER. In MAX-3SAT, given a 3-conjunctive normal form (3-CNF) Boolean formula ϕ , of which each clause contains at most three variables (e.g., $\phi = (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee \overline{x_3} \vee x_4) \wedge (\overline{x_1} \vee x_3 \vee \overline{x_4})$), we are asked to find a truth assignment that satisfies the maximum fraction of the clauses of ϕ . The decision version of MAX-3SAT is known to be NP-hard; indeed, Håstad (2001) established the following indistinguishability:

Theorem A.1 (Håstad, 2001, Theorem 6.5). *Given a 3-CNF Boolean formula ϕ , it is NP-hard to distinguish between the following two cases for any constant $\epsilon > 0$:*

- (Completeness) ϕ is satisfiable.
- (Soundness) No truth assignment can satisfy at least $(\frac{7}{8} + \epsilon)$ -fraction of the clauses of ϕ .

We first reduce from MAX-3SAT to MAX-3SAT(29), which is a special case of MAX-3SAT where every variable appears in at most 29 clauses of a 3-CNF Boolean formula.

Theorem A.2 (Trevisan, 2004, Theorem 7; Vazirani, 2013, Theorem 29.11; Tamaki, 2015, Theorem 4.7). *There is a polynomial-time gap-preserving reduction that transforms an instance ϕ of MAX-3SAT to an instance ψ of MAX-3SAT(29) such that the following is satisfied:*

- (Completeness) If ϕ is satisfiable, then so is ψ .
- (Soundness) If no truth assignment satisfies at least $(1 - \epsilon)$ fraction of the clauses of ϕ , then no truth assignment satisfies at least $(1 - \frac{\epsilon}{43})$ -fraction of the clauses of ψ .

We next reduce from MAX-3SAT(29) to MAX-E3SAT(5), which is a special case of MAX-3SAT where an input 3-CNF formula contains n variables and $5n/3$ clauses for some positive integer divisible by 3, each clause contains exactly 3 literals, and each variable appears in exactly 5 clauses but never appears twice in the same clause.

Theorem A.3 (Feige, 1998, Proposition 2.1.2; Tamaki, 2015, Theorem 4.11). *There is a polynomial-time gap-preserving reduction that transforms an instance ϕ of MAX-3SAT(29) to an instance ψ of MAX-E3SAT(5) such that the following is satisfied:*

- (Completeness) If ϕ is satisfiable, then so is ψ .
- (Soundness) If no truth assignment satisfies at least $(1 - \epsilon)$ -fraction of the clauses of ϕ , then no truth assignment satisfies at least $(1 - \frac{\epsilon}{200})$ -fraction of the clauses of ψ .

We further reduce from MAX-E3SAT(5) to LABELCOVER (i.e., a projection game) as follows.

Theorem A.4 (Tamaki, 2015, Proof of Theorem 4.2). *There is a polynomial-time reduction that transforms an instance ϕ of MAX-E3SAT(5) with n variables and $5n/3$ clauses to a projection game $\mathfrak{G} = (X, Y, E, \Sigma, \Pi)$ such that the following is satisfied:*

- $|X| = n, |Y| = 5n/3, |E| = 5n, |\Sigma| = 7$, and each vertex of X and Y has exactly degree 5 and 3, respectively.
- (Completeness) If ϕ is satisfiable, then $\text{val}(\mathfrak{G}) = 1$.
- (Soundness) If no truth assignment satisfies at least $(1 - \epsilon)$ -fraction of the clauses of ϕ , then $\text{val}(\mathfrak{G}) \leq 1 - \frac{\epsilon}{3}$.

We finally reduce from the above-mentioned projection game to a special projection game (see Section 2.2) as follows.

Lemma A.5. *Let $\mathfrak{G}' = (X', Y', E', \Sigma', \Pi' = \{\pi'_e\}_{e \in E'})$ be a projection game satisfying the conditions in Theorem A.4, i.e., $|X'| = n, |Y'| = 5n/3, |E'| = 5n$, and $|\Sigma'| = 7$ for some positive integer n divisible by 3. There is a polynomial-time gap-preserving reduction that transforms \mathfrak{G}' to a new projection game $\mathfrak{G} = (X, Y, E, \Sigma, \Pi = \{\pi_e\}_{e \in E})$ such that the following is satisfied:*

- (X, Y, E) is a 15-regular bipartite graph (i.e., each vertex of $X \uplus Y$ is incident to exactly 15 edges), where $|X| = |Y| = 5n$ and $|E| = 75n$,
- $|\Sigma| = 7$, and
- $\text{val}(\mathfrak{G}) = \text{val}(\mathfrak{G}')$.

Proof. We construct \mathfrak{G} as follows. We first create $X \triangleq \bigcup_{x \in X'} \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}\}$ and $Y \triangleq \bigcup_{y \in Y'} \{y^{(1)}, y^{(2)}, y^{(3)}\}$, where each $x^{(i)}$ for $i \in [5]$ and $y^{(j)}$ for $j \in [3]$ is a copy of $x \in X$ and $y \in Y$, respectively. We then build an edge set as $E \triangleq \bigcup_{i \in [5], j \in [3]} \{(x^{(i)}, y^{(j)}) \in X \times Y \mid (x, y) \in E'\}$, and $\pi_e \triangleq \pi'_{(x,y)}$ for each $e = (x^{(i)}, y^{(j)}) \in E$. Then, $\text{val}(\mathfrak{G})$ is easily verified to be $\text{val}(\mathfrak{G}')$. \square

Proof of Theorem 2.8. By Theorems A.1 to A.4 and Lemma A.5, we can reduce a 3-CNF Boolean formula ϕ to a projection game \mathfrak{G} such that the following is satisfied:

- \mathfrak{G} is special (i.e., \mathfrak{G} satisfies the conditions in Theorem 2.8).
- If ϕ is satisfiable, then $\text{val}(\mathfrak{G}) = 1$.
- If no truth assignment satisfy at most $(\frac{7}{8} + \epsilon)$ -fraction of the clauses of ϕ , then $\text{val}(\mathfrak{G}) \leq 1 - \frac{1}{206,400} + \frac{\epsilon}{25,800}$, for any $\epsilon > 0$.

We can specify $\epsilon > 0$ so that $1 - \frac{1}{206,400} + \frac{\epsilon}{25,800} < 1 - \frac{1}{206,401}$, which completes the proof. \square

Appendix B. Missing Proofs in Section 2 and Section 3

Proof of Lemma 2.14. Fix $\mathbf{S} \subseteq \mathbf{V}$. Let \mathbf{Q} be a set of $|X(\mathbf{S})|$ vectors from \mathbf{S}_X with no repetitions, i.e., $\mathbf{Q} \subseteq \mathbf{S}_X$ such that $|\mathbf{Q}| = |X(\mathbf{S})|$ and $\text{rep}(\mathbf{Q}) = 0$, and let $\mathbf{P} = \mathbf{S}_X \setminus \mathbf{Q}$. For each vector $\mathbf{v}_{x,i}$ in \mathbf{P} , there is exactly one vector $\mathbf{v}_{x,i'}$ in \mathbf{Q} with $i \neq i'$; hence, we have

$$d(\mathbf{v}_{x,i}, \mathbf{Q}) \leq d(\mathbf{v}_{x,i}, \{\mathbf{v}_{x,i'}\}) = \|\mathbf{v}_{x,i} - \langle \mathbf{v}_{x,i}, \mathbf{v}_{x,i'} \rangle \mathbf{v}_{x,i'}\| \leq \frac{\sqrt{3}}{2}.$$

By Lemma 2.10, we have

$$\text{vol}(\mathbf{S}_X) \leq \text{vol}(\mathbf{Q}) \cdot \prod_{\mathbf{v}_{x,i} \in \mathbf{P}} d(\mathbf{v}_{x,i}, \mathbf{Q}) \leq \left(\frac{\sqrt{3}}{2}\right)^{|\mathbf{P}|} = \left(\frac{\sqrt{3}}{2}\right)^{\text{rep}(\mathbf{S}_X)}.$$

The proof for $\text{vol}(\mathbf{S}_Y)$ is similar. \square

Proof of Lemma 3.5. Fix $\mathbf{S} \subseteq \mathbf{V}$. Let \mathbf{Q} be a set of $|X(\mathbf{S})|$ vectors from \mathbf{S}_X with no repetitions, i.e., $\mathbf{Q} \subseteq \mathbf{S}_X$ such that $|\mathbf{Q}| = |X(\mathbf{S})|$ and $\text{rep}(\mathbf{Q}) = 0$, and let $\mathbf{P} = \mathbf{S}_X \setminus \mathbf{Q}$. For each vector $\mathbf{v}_{x,i}$ in \mathbf{P} , there is exactly one vector $\mathbf{v}_{x,i'}$ in \mathbf{Q} with $i \neq i'$; hence, we have

$$\begin{aligned} d(\mathbf{v}_{x,i}, \mathbf{Q}) &\leq d(\mathbf{v}_{x,i}, \{\mathbf{v}_{x,i'}\}) = \|\mathbf{v}_{x,i} - \langle \mathbf{v}_{x,i}, \mathbf{v}_{x,i'} \rangle \mathbf{v}_{x,i'}\| \\ &= \sqrt{1 - \langle \mathbf{v}_{x,i}, \mathbf{v}_{x,i'} \rangle^2} \leq \sqrt{1 - \left(\frac{15^\ell}{15^\ell + 1} \cdot \frac{1}{2}\right)^2} \\ &< \frac{\sqrt{3.01}}{2}. \end{aligned}$$

By Lemma 2.10, we have

$$\text{vol}(\mathbf{S}_X) \leq \text{vol}(\mathbf{Q}) \cdot \prod_{\mathbf{v}_{x,i} \in \mathbf{P}} d(\mathbf{v}_{x,i}, \mathbf{Q}) \leq \left(\frac{\sqrt{3.01}}{2}\right)^{|\mathbf{P}|} = \left(\frac{\sqrt{3.01}}{2}\right)^{\text{rep}(\mathbf{S}_X)}.$$

The proof for $\text{vol}(\mathbf{S}_Y)$ is similar. \square

Proof of Claim 3.6. Observe first that $\text{vol}(\mathbf{S}_X) \geq \text{vol}(\mathbf{S}) \geq 2^{-cK}$. By Lemma 3.5, we have $(\sqrt{3.01}/2)^{\text{rep}(\mathbf{S}_X)} \geq \text{vol}(\mathbf{S}_X) \geq 2^{-cK}$, implying $\text{rep}(\mathbf{S}_X) \leq cK / \log_2(2/\sqrt{3.01}) < 5cK$. Similarly, we have $\text{rep}(\mathbf{S}_Y) < 5cK$. Using the fact that $|\mathbf{S}_X| = |X(\mathbf{S})| + \text{rep}(\mathbf{S}_X)$, $|\mathbf{S}_Y| = |Y(\mathbf{S})| + \text{rep}(\mathbf{S}_Y)$, and $K = |\mathbf{S}_X| + |\mathbf{S}_Y|$, we bound $|X(\mathbf{S})|$ from below as follows:

$$\begin{aligned} |X(\mathbf{S})| &= K - |\mathbf{S}_Y| - \text{rep}(\mathbf{S}_X) > K - |Y| - 10cK \\ &= \left(1 - \frac{(5n)^\ell}{K} - 10c\right) K = \left(\frac{1}{2} - 10c\right) K. \end{aligned}$$

Similarly, we have $|Y(\mathbf{S})| > (\frac{1}{2} - 10c)K$. \square

Proof of Lemma 3.7. The proof is by contradiction. Suppose there exists a set \mathbf{S} of K vectors from V such that $\text{vol}(\mathbf{S}) \geq 2^{-\beta^\circ K}$.

Consider a labeling $\sigma : (X \uplus Y) \rightarrow \Sigma$ defined according to Eq. (2) in the proof of Lemma 2.16. Define $\mathbf{P} \triangleq \{\mathbf{v}_{x,\sigma(x)} \mid x \in X(\mathbf{S})\}$ and $\mathbf{Q} \triangleq \{\mathbf{v}_{y,\sigma(y)} \mid y \in Y(\mathbf{S})\}$. Here, to use Lemma 2.10, we bound the distance of the vectors of \mathbf{P} to \mathbf{Q} . Similar to the proof of Lemma 2.16, we have for each $\mathbf{v}_{x,\sigma(x)} \in \mathbf{P}$,

$$d(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q}) = \sqrt{1 - \sum_{\mathbf{v}_{y,\sigma(y)} \in \mathbf{Q}} \langle \mathbf{v}_{x,\sigma(x)}, \mathbf{v}_{y,\sigma(y)} \rangle^2} = \left(1 - \frac{U(x)}{4 \cdot (15^\ell + 1)^2}\right)^{\frac{1}{2}},$$

where $U(x)$ is defined as the number of unsatisfied edges between x and $Y(\mathbf{S})$. Using Lemma 2.10 and the fact that $\text{vol}(\mathbf{Q}) \leq 1$, we have

$$\begin{aligned} \text{vol}(\mathbf{P} \uplus \mathbf{Q}) &\leq \text{vol}(\mathbf{Q}) \cdot \prod_{x \in X(\mathbf{S})} d(\mathbf{v}_{x, \sigma(x)}, \mathbf{Q}) \\ &\leq \left(\prod_{x \in X(\mathbf{S})} \left(1 - \frac{U(x)}{4 \cdot (15^\ell + 1)^2} \right) \right)^{\frac{1}{2}} \\ &\leq \left(\frac{1}{|X(\mathbf{S})|} \sum_{x \in X(\mathbf{S})} \left(1 - \frac{U(x)}{4 \cdot (15^\ell + 1)^2} \right) \right)^{\frac{|X(\mathbf{S})|}{2}}, \end{aligned} \quad (8)$$

where the last inequality is by the AM–GM inequality.

Now consider bounding $\sum_{x \in X(\mathbf{S})} U(x)$ from below, which is equal to the total number of unsatisfied edges between $X(\mathbf{S})$ and $Y(\mathbf{S})$ by σ . Substituting β° for c in Claim 3.6 derives $|X(\mathbf{S})| > (\frac{1}{2} - 10\beta^\circ)K$ and $|Y(\mathbf{S})| > (\frac{1}{2} - 10\beta^\circ)K$. Because less than $2^{-\alpha\ell}$ -fraction of edges in E can be satisfied by any labeling (including σ) by assumption, and more than $(\frac{1}{2} - 10\beta^\circ)K$ edges are incident to $X(\mathbf{S})$ (resp. $Y(\mathbf{S})$), the number of unsatisfied edges *incident to* $X(\mathbf{S})$ (resp. $Y(\mathbf{S})$) is at least

$$\left[\left(\frac{1}{2} - 10\beta^\circ \right) \frac{K}{(5n)^\ell} - 2^{-\alpha\ell} \right] (75n)^\ell. \quad (9)$$

Consequently, the number of unsatisfied edges *between* $X(\mathbf{S})$ and $Y(\mathbf{S})$ is at least twice Eq. (9) minus “the number of unsatisfied edges incident to $X(\mathbf{S})$ or $Y(\mathbf{S})$ ” (which is at most $(75n)^\ell$); namely, we have

$$\sum_{x \in X(\mathbf{S})} U(x) \geq \left[\left(\frac{1}{4} - 10\beta^\circ \right) \frac{2K}{(5n)^\ell} - 2^{-\alpha\ell+1} \right] (75n)^\ell.$$

We further expand Eq. (8) as follows:

$$\begin{aligned} \text{vol}(\mathbf{P} \uplus \mathbf{Q}) &\leq \left(1 - \frac{\sum_{x \in X(\mathbf{S})} U(x)}{|X(\mathbf{S})|} \frac{1}{4 \cdot (15^\ell + 1)^2} \right)^{\frac{|X(\mathbf{S})|}{2}} \\ &\leq \exp \left(- \frac{\left[\left(\frac{1}{4} - 10\beta^\circ \right) \frac{2K}{(5n)^\ell} - 2^{-\alpha\ell+1} \right] (75n)^\ell}{|X(\mathbf{S})|} \frac{|X(\mathbf{S})|}{8 \cdot (15^\ell + 1)^2} \right) \\ &\leq \exp \left(- \left[\left(\frac{1}{4} - 10\beta^\circ \right) \frac{15^\ell}{4 \cdot (15^\ell + 1)^2} - 2^{-\alpha\ell+1} \frac{15^\ell}{16 \cdot (15^\ell + 1)^2} \right] K \right) \\ &= \exp \left(- \frac{1 - 40\beta^\circ - 2^{-\alpha\ell+1}}{16 \cdot (15^\ell + 1)^2} 15^\ell \cdot K \right) \\ &\leq \exp \left(- \frac{1 - 40\beta^\circ - 2^{-\alpha\ell+1}}{32 \cdot 15^\ell} K \right). \end{aligned}$$

Since $\beta^\circ = 10^{-10^{12.4}} < \frac{1 - 2^{-\alpha\ell+1}}{32 \cdot 15^\ell \cdot \log_e(2) + 40}$ and $\beta^\circ > 0$, we finally have $\text{vol}(\mathbf{S}) \leq \text{vol}(\mathbf{P} \uplus \mathbf{Q}) < 2^{-\beta^\circ K}$, a contradiction. \square

References

- Anari, N., & Gharan, S. O. (2017). A generalization of permanent inequalities and applications in counting and optimization. In *STOC*, pp. 384–396.
- Anari, N., Liu, K., Gharan, S. O., & Vinzant, C. (2019). Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC*, pp. 1–12.
- Bıyık, E., Wang, K., Anari, N., & Sadigh, D. (2019). Batch active learning using determinantal point processes. *CoRR*, *abs/1906.07975*.
- Borodin, A., & Rains, E. M. (2005). Eynard-Mehta theorem, Schur process, and their Pfaffian analogs. *J. Stat. Phys.*, *121*(3–4), 291–317.
- Buchbinder, N., & Feldman, M. (2018). Deterministic algorithms for submodular maximization problems. *ACM Trans. Algorithms*, *14*(3), 1–20.
- Buchbinder, N., Feldman, M., Naor, J. S., & Schwartz, R. (2015). A tight linear time (1/2)-approximation for unconstrained submodular maximization. *SIAM J. Comput.*, *44*(5), 1384–1402.
- Chao, W.-L., Gong, B., Grauman, K., & Sha, F. (2015). Large-margin determinantal point processes. In *UAI*, pp. 191–200.
- Chen, L., Zhang, G., & Zhou, E. (2018). Fast greedy MAP inference for determinantal point process to improve recommendation diversity. In *NeurIPS*, pp. 5622–5633.
- Çivril, A., & Magdon-Ismail, M. (2009). On selecting a maximum volume sub-matrix of a matrix and related problems. *Theor. Comput. Sci.*, *410*(47–49), 4801–4811.
- Çivril, A., & Magdon-Ismail, M. (2013). Exponential inapproximability of selecting a maximum volume sub-matrix. *Algorithmica*, *65*(1), 159–176.
- Di Summa, M., Eisenbrand, F., Faenza, Y., & Moldenhauer, C. (2014). On largest volume simplices and sub-determinants. In *SODA*, pp. 315–323.
- Dinur, I., & Steurer, D. (2014). Analytical approach to parallel repetition. In *STOC*, pp. 624–633.
- Feige, U. (1998). A threshold of $\ln n$ for approximating set cover. *J. ACM*, *45*(4), 634–652.
- Gartrell, M., Han, I., Dohmatob, E., Gillenwater, J., & Brunel, V.-E. (2021). Scalable learning and MAP inference for nonsymmetric determinantal point processes. In *ICLR*.
- Gillenwater, J., Kulesza, A., & Taskar, B. (2012). Near-optimal MAP inference for determinantal point processes. In *NIPS*, pp. 2735–2743.
- Gillenwater, J. A. (2014). *Approximate Inference for Determinantal Point Processes*. Ph.D. thesis, University of Pennsylvania.
- Gong, B., Chao, W., Grauman, K., & Sha, F. (2014). Diverse sequential subset selection for supervised video summarization. In *NIPS*, pp. 2069–2077.
- Gurvits, L. (2005). On the complexity of mixed discriminants and related problems. In *MFCS*, pp. 447–458.

- Gurvits, L. (2009). On complexity of the mixed volume of parallelograms. In *EuroCG*, pp. 337–341.
- Han, I., & Gillenwater, J. (2020). MAP inference for customized determinantal point processes via maximum inner product search. In *AISTATS*, pp. 2797–2807.
- Han, I., Kambadur, P., Park, K., & Shin, J. (2017). Faster greedy MAP inference for determinantal point processes. In *ICML*, pp. 1384–1393.
- Håstad, J. (1999). Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Math.*, 182(1), 105–142.
- Håstad, J. (2001). Some optimal inapproximability results. *J. ACM*, 48(4), 798–859.
- Ko, C.-W., Lee, J., & Queyranne, M. (1995). An exact algorithm for maximum entropy sampling. *Oper. Res.*, 43(4), 684–691.
- Koutis, I. (2006). Parameterized complexity and improved inapproximability for computing the largest j -simplex in a V -polytope. *Inf. Process. Lett.*, 100(1), 8–13.
- Kulesza, A., & Taskar, B. (2011). Learning determinantal point processes. In *UAI*, pp. 419–427.
- Kulesza, A., & Taskar, B. (2012). Determinantal point processes for machine learning. *Found. Trends Mach. Learn.*, 5(2–3), 123–286.
- Lee, D., Cha, G., Yang, M.-H., & Oh, S. (2016). Individualness and determinantal point processes for pedestrian detection. In *ECCV*, pp. 330–346.
- Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. Appl. Probab.*, 7(1), 83–122.
- Madan, V., Nikolov, A., Singh, M., & Tantipongpipat, U. (2020). Maximizing determinants under matroid constraints. In *FOCS*, pp. 565–576.
- Mariet, Z. E., Sra, S., & Jegelka, S. (2018). Exponentiated strongly Rayleigh distributions. In *NeurIPS*, pp. 4464–4474.
- Nemhauser, G. L., Wolsey, L. A., & Fisher, M. L. (1978). An analysis of the approximations for maximizing submodular set functions—I. *Math. Program.*, 14, 265–294.
- Nikolov, A. (2015). Randomized rounding for the largest simplex problem. In *STOC*, pp. 861–870.
- Nikolov, A., & Singh, M. (2016). Maximizing determinants under partition constraints. In *STOC*, pp. 192–201.
- Ohsaka, N. (2021). Unconstrained MAP inference, exponentiated determinantal point processes, and exponential inapproximability. In *AISTATS*, pp. 154–162.
- Ohsaka, N., & Matsuoka, T. (2020). On the (in)tractability of computing normalizing constants for the product of determinantal point processes. In *ICML*, pp. 6600–6609.
- Perez-Beltrachini, L., & Lapata, M. (2021). Multi-document summarization with determinantal point process attention. *J. Artif. Intell. Res.*, 71, 371–399.
- Raz, R. (1998). A parallel repetition theorem. *SIAM J. Comput.*, 27(3), 763–803.
- Robinson, J., Sra, S., & Jegelka, S. (2019). Flexible modeling of diversity with strongly log-concave distributions. In *NeurIPS*, pp. 15199–15209.

- Sharma, D., Kapoor, A., & Deshpande, A. (2015). On greedy maximization of entropy. In *ICML*, pp. 1330–1338.
- Tamaki, S. (2015). Parallel repetition of two-prover one-round games: An exposition. *Interdiscip. Inf. Sci.*, 21(4), 289–306.
- Trevisan, L. (2004). Inapproximability of combinatorial optimization problems. *CoRR*, *abs/cs/0409043*.
- Vazirani, V. V. (2013). *Approximation Algorithms*. Springer Science & Business Media.
- Wilhelm, M., Ramanathan, A., Bonomo, A., Jain, S., Chi, E. H., & Gillenwater, J. (2018). Practical diversified recommendations on YouTube with determinantal point processes. In *CIKM*, pp. 2165–2173.
- Yao, J., Fan, F., Zhao, W. X., Wan, X., Chang, E. Y., & Xiao, J. (2016). Tweet timeline generation with determinantal point processes. In *AAAI*, pp. 3080–3086.
- Zhang, M. J., & Ou, Z. (2016). Block-wise MAP inference for determinantal point processes with application to change-point detection. In *SSP*, pp. 1–5.
- Zou, J. Y., & Adams, R. P. (2012). Priors for diversity in generative latent variable models. In *NIPS*, pp. 2996–3004.