Unconstrained MAP Inference, Exponentiated Determinantal Point Processes, and Exponential Inapproximability

Naoto Ohsaka
NEC Corporation

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 prove the following complexity-theoretic hardness results that explain the difficulty in approximating unconstrained MAP inference and the normalizing constant for E-DPPs.

- Unconstrained MAP inference for an $n \times n$ matrix is NP-hard to approximate within a $2^{\beta n}$-factor, where $\beta = 10^{-10^{13}}$. This result improves upon a $(\frac{9}{8} - \epsilon)$-factor inapproximability given by Kulesza and Taskar (2012).
- 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 $2^{\beta p n}$-factor. This gives a(nother) negative answer to open questions posed by Kulesza and Taskar (2012); Ohsaka and Matsuoka (2020).

1 Introduction

Selecting a small set of "diverse" items from large data is an essential task in machine learning. 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 and Rains, 2005). Suppose we are given n items (e.g., images or documents) associated with feature vectors $\{\boldsymbol{\phi}_i\}_{i \in [n]}$ and an $n \times n$ Gram matrix $\mathbf{A}$ such that $A_{i,j} = \langle \boldsymbol{\phi}_i, \boldsymbol{\phi}_j \rangle$ 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 $\{\boldsymbol{\phi}_i\}_{i \in S}$, dissimilar items are likely to appear in the selected subsets, which ensures set diversity. DPPs exhibit fascinating properties that make them suitable for 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 and Taskar (2012) for details).

The present study aims at analyzing two exceptionally hard problems on DPPs through the lens of complexity theory—unconstrained 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 and Taskar, 2011; Chao et al., 2015), YouTube video recommendation (Wilhelm et al., 2018), active learning (Bıyık 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 has been challenging despite its simplicity. Typically, the Greedy algorithm is used as a heuristic, whereas the current best approximation factor is e^n (Nikolov, 2015). Indeed, Kulesza and Taskar (2012) have shown that unconstrained MAP inference is NP-hard to approximate within a $(\frac{9}{8} - \epsilon)$-factor,[1] which is the current best lower bound. On the other hand, *size-constrained* MAP inference (i.e., $|S| = k$) is NP-hard to approximate within an exponential factor of 2^{ck} for some $c > 0$ (Çivril and Magdon-Ismail, 2013), which does *not*, however, directly apply to the unconstrained case. Closing the

Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

[1] We define approximation factor ρ so that $\rho \geq 1$; see §2.

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 (lower bound)	approximability (upper bound)
unconstrained $S \subseteq [n]$	$2^{\beta n}$ for $\beta = 10^{-10^{13}}$ (**this paper, Theorem 2.2**) $\frac{9}{8} - \epsilon$ (Kulesza and Taskar, 2012)	e^n (Nikolov, 2015) $n!^2$ (Çivril and Magdon-Ismail, 2009)
size-constrained $S \subseteq [n], \lvert S \rvert = k$	2^{ck} $(c > 0)$ (Koutis, 2006; Çivril and Magdon-Ismail, 2013) $(2^{\frac{1}{506}} - \epsilon)^k$ (Di Summa et al., 2014)	e^k (Nikolov, 2015) $k!^2$ (Çivril and Magdon-Ismail, 2009)

gap between the lower bound ($\approx \frac{9}{8}$) and upper bound ($= \mathrm{e}^n$) is the first question addressed in this study.

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 $\propto \det(\mathbf{A}_S)^p$ (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, while setting $p = 0$ results in a uniform distribution. 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 only if p is an even integer (Gurvits, 2005; Ohsaka and Matsuoka, 2020), and the case of $p \leq 1$ admits a fully polynomial-time randomized approximation scheme (FPRAS)[2] based on an approximate sampler (Anari et al., 2019; Robinson et al., 2019). The second question in this study is to find the value of p such that the normalizing constant is hard to approximate.

Our research questions can be summarized as follows:

> **Q1.** *Is unconstrained MAP inference on DPPs exponentially inapproximable?*
> **Q2.** *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 two complexity-theoretic hardness results.

(§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 $2^{\beta n}$-factor for $\beta = 10^{-10^{13}}$.*

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.

(§3) Exponential Inapproximability of Exponentiated DPPs.

Our second result is the following, which is derived by applying Theorem 2.2 (cf. Figure 1).

> **Corollary 3.2** (informal)**.** *For every fixed number $p \geq \beta^{-1} = 10^{10{,}000{,}000{,}000{,}000}$, 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, §7.2) and Ohsaka and Matsuoka (2020, §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 3.4). 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 3.3).

1.2 Related Work

MAP Inference on DPPs. In Theoretical Computer Science, unconstrained MAP inference on DPPs is known as *determinant maximization* (DetMax for short). The size-constrained version (k-DetMax for short), which restricts the output to size k for parameter k, finds applications in computational geometry

[2] An FPRAS is a randomized algorithm that outputs an e^{ϵ}-approximation with probability at least 3/4 and runs in polynomial time in the input size and ϵ^{-1}.

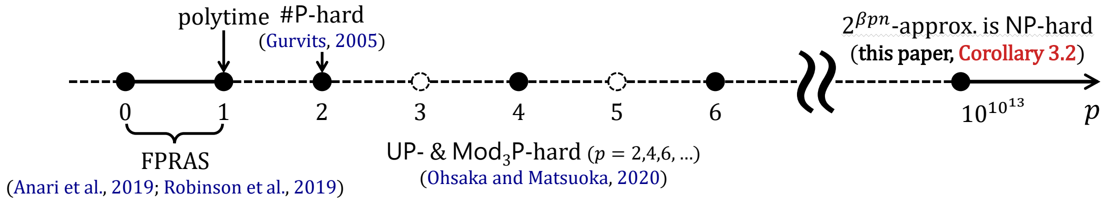

Figure 1: Computational complexity of the normalizing constant for exponentiated DPPs. Our result is $2^{\beta p n}$-factor inapproximability for $p \geq 10^{10^{13}}$. Tractability for p's on dashed lines and circles remains open.

and discrepancy theory (Nikolov, 2015).

On the inapproximability side, Ko, Lee, and Queyranne (1995) prove that DetMax and k-DetMax 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-DetMax is also exponentially inapproximable in k. Çivril and Magdon-Ismail (2013) directly prove a similar result for k-DetMax. Di Summa, Eisenbrand, Faenza, and Moldenhauer (2014) study the special case of k-DetMax, 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-DetMax 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-DetMax; this is the current best approximation factor. Invoking Nikolov's algorithm for all k immediately yields an e^n-approximation for DetMax.

In Machine Learning, 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 and Ou, 2016), and others (Gillenwater et al., 2012; Han et al., 2017; Chen et al., 2018; Chao et al., 2015). Since the logarithm of the determinant as a set function $f(S) \triangleq \log\det(\mathbf{A}_S)$ for a positive semi-definite matrix $\mathbf{A}$ is submodular,[3] the Greedy algorithm for monotone submodular maximization (Nemhauser et al., 1978) is widely used,[4] which works pretty well in practice (Yao et al., 2016; Zhang and Ou, 2016) and guarantees an $(\frac{\mathrm{e}}{\mathrm{e}-1})$-factor approximation (with respect to f) under a size constraint if every eigenvalue of $\mathbf{A}$ is greater than 1 (Han and Gillenwater, 2020). Several attempts have been made to scale up Greedy (Han et al., 2017; Chen et al., 2018; Han and Gillenwater, 2020; Gartrell et al., 2020), whose naive implementation requires quartic time in n. Other than Greedy, Gillenwater, Kulesza, and Taskar (2012) propose a gradient-based efficient algorithm that has an approximation factor of 4 (with respect to f). We emphasize that some studies focus on maximizing the *logarithm* of the determinant, e.g., (Gillenwater et al., 2012; Han and Gillenwater, 2020). Such results for a different objective f are "incomparable" to our result in the sense that a multiplicative approximation to $\max_{S \subseteq [n]} f(S)$ does *not* imply a multiplicative approximation to DetMax.

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 and 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 and 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 Mod$_3$P-hardness for every positive even integer $p = 2, 4, 6, \ldots$. 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 $p \geq 10^{10^{13}}$.

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

[4]$f(S) = \log\det(\mathbf{A}_S)$ is not necessarily monotone, for which Greedy has, in fact, a poor approximation guarantee. The tight approximation factor for unconstrained nonmonotone submodular maximization (including $\max_{S \subseteq [n]} \log\det(\mathbf{A}_S)$ as a special case) is 2 (Buchbinder et al., 2015; Buchbinder and Feldman, 2018).

1.3 Notations

For a positive integer n, let $[n] \triangleq \{1, 2, \ldots, n\}$. 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., $\langle \mathbf{v}, \mathbf{w} \rangle \triangleq \sum_{i \in [d]} v(i) w(i)$ for two vectors $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^d$. For an $n \times n$ matrix $\mathbf{A}$ and subset $S \subseteq [n]$ of indices, 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 $\det(\mathbf{A}) \triangleq \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i \in [n]} A_{i,\sigma(i)}$, where $\mathfrak{S}_n$ is the symmetric group on $[n]$. For a set $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of n vectors in $\mathbb{R}^d$, the *volume* of the parallelepiped spanned by $\mathbf{V}$ is defined as $\mathrm{vol}(\mathbf{V}) \triangleq \|\mathbf{v}_1\| \cdot \prod_{2 \leq i \leq n} \left\| \mathrm{proj}_{\{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}\}}(\mathbf{v}_i) \right\|$, where $\mathrm{proj}_{\mathbf{P}}(\cdot)$ is an operator of orthogonal projection onto the subspace spanned by vectors in $\mathbf{P}$.

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 $S \subseteq [n]$ such that the determinant $\det(\mathbf{A}_S)$ of a principal submatrix is maximized. The optimal value of DetMax is denoted $\mathrm{maxdet}(\mathbf{A}) \triangleq \max_{S \subseteq [n]} \det(\mathbf{A}_S)$.

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

$$\det(\text{ALG}(\mathbf{A})) \geq (1/\rho) \cdot \mathrm{maxdet}(\mathbf{A}),$$

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 $\mathrm{maxdet}(\mathbf{A}) \geq c(n)$ or $\mathrm{maxdet}(\mathbf{A}) < s(n)$.[5] If $[s(n), c(n)]$-Gap-DetMax is NP-hard, then so is approximating DetMax within a $\frac{c(n)}{s(n)}$-factor.

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 P = 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 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{Q}^d$. In such a case, we have a simple relation that $\det(\mathbf{A}_S) = \mathrm{vol}(\{\mathbf{v}_i\}_{i \in S})^2$ for every subset $S \subseteq [n]$. DetMax is thus essentially equivalent to the following optimization problem:

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

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 §2. §2.1 introduces projection games to be reduced to VolMax and Raz's parallel repetition theorem. §2.2 reviews the indistinguishability of projection games. §2.3 describes Main Lemma, which is crucial in proving Theorem 2.2, and §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 tuple $\mathfrak{G} = (X, Y, E, \Sigma, \Pi)$ such that

- (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 $\Sigma \to \Sigma$.

A *labeling* σ is defined as a label assignment of each vertex in the bipartite graph, i.e., $\sigma : (X \uplus Y) \to \Sigma$. 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 $\mathrm{val}(\mathfrak{G})$, is defined as the maximum fraction of edges satisfied over all possible labelings σ, i.e.,

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

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

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 $e = ((x_1, x_2), (y_1, y_2)) \in E$, $\pi_e : \Sigma_1 \times \Sigma_2 \to \Sigma_1 \times \Sigma_2$ 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 $\mathfrak{G}^{\otimes \ell} \triangleq \underbrace{\mathfrak{G} \otimes \cdots \otimes \mathfrak{G}}_{k \text{ times}}$. Raz (1998) proved the *parallel repetition theorem*, which states that for every (not necessarily projection) game $\mathfrak{G}$ with $\mathrm{val}(\mathfrak{G}) = 1 - \epsilon$, it holds that $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) \leq (1 - \overline{\epsilon})^{\frac{\ell}{\log |\Sigma|}}$, where $\overline{\epsilon}$ is a constant depending only on ϵ. 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 a tighter, explicit bound derived by Dinur and Steurer (2014).

Theorem 2.7 (Corollary 1 in Dinur and Steurer (2014))**.** *For any projection game $\mathfrak{G}$ with $\mathrm{val}(\mathfrak{G}) \leq 1 - \epsilon$ for some $\epsilon > 0$, $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) \leq \left(1 - \frac{\epsilon^2}{16}\right)^{\ell}$.*

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 whether a projection game has value 1 or has value less than $1 - \epsilon$ for some $\epsilon > 0$. Though its proof is widely known, we include it in Appendix A to describe the value of such ϵ explicitly.

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 $\mathrm{val}(\mathfrak{G}) = 1$ and $\mathrm{val}(\mathfrak{G}) < 1 - \frac{1}{206{,}401}$.*

A projection game satisfying the conditions in Theorem 2.8 is called *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 $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) = 1$ or

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

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 $\ell \triangleq \left\lceil \frac{3}{\alpha} \right\rceil = 1.5 \cdot 10^{12}$. Our main lemma in the following can be thought of as an extension of Çivril and Magdon-Ismail (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 $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ of VolMax such that $N = 2 \cdot (35n)^{\ell}$ for some integer n, each vector of $\mathbf{V}$ is normalized, and the following is satisfied:*

- *(Completeness) If $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there exists a set $\mathbf{S}$ of K vectors from $\mathbf{V}$ with volume $\mathrm{vol}(\mathbf{S}) = 1$, where $K = N/7^{\ell}$.*
- *(Soundness) If $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$, then any set $\mathbf{S}$ of k vectors from $\mathbf{V}$ satisfies the following properties:*
  1. *$0 \leq k < \frac{7}{8}K$: $\mathrm{vol}(\mathbf{S}) \leq 1$.*
  2. *$\frac{7}{8}K \leq k \leq N$: $\mathrm{vol}(\mathbf{S}) < 2^{-\beta^\circ k}$ for $\beta^\circ = 10^{-10^{12.3}}$.*

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

Proof of Theorem 2.2. Let $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ 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 $\mathbf{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_N\}$, 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 $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) = 1$, then there is a set $\mathbf{S}$ of $K = N/7^{\ell}$ vectors from $\mathbf{W}$ such that $\mathrm{vol}(\mathbf{S}) = 2^{\beta^\circ K}$. On the other hand, if $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$, $\mathrm{maxvol}(\mathbf{W})$ is (strictly) bounded from above by $2^{\frac{7}{8}\beta^\circ K}$ through the following case analysis on $\mathbf{S} \subseteq \mathbf{W}$:

- **1.** $0 \leq |\mathbf{S}| < \frac{7}{8}K$: $\mathrm{vol}(\mathbf{S}) < 2^{\frac{7}{8}\beta^\circ K}$.
- **2.** $\frac{7}{8}K \leq |\mathbf{S}| \leq N$: $\mathrm{vol}(\mathbf{S}) < 2^{(\beta^\circ - \beta^\circ)|\mathbf{S}|} \leq 1$.

It is thus NP-hard to decide whether $\mathrm{maxvol}(\mathbf{W}) \geq 2^{\frac{\beta^\circ}{7^\ell} N}$ or $\mathrm{maxvol}(\mathbf{W}) < 2^{\frac{7\beta^\circ}{8 \cdot 7^\ell} N}$ by Theorems 2.7 and 2.8. Owing to the relation between VolMax and DetMax, $[2^{\frac{7\beta^\circ}{4 \cdot 7^\ell} n}, 2^{\frac{2\beta^\circ}{7^\ell} n}]$-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 $\lambda_s = \frac{7\beta^\circ}{4 \cdot 7^\ell}$. Observing that $\beta = 10^{-10^{13}} < 10^{-10^{12.6}} < \lambda_c - \lambda_s$ suffices to complete the proof. $\square$

Remark 2.10. *The above proof implies that it is* NP*-hard to decide whether* $\log_2 \operatorname{maxdet}(\mathbf{A}) \geq \lambda_c n$ *or* $\log_2 \operatorname{maxdet}(\mathbf{A}) < \lambda_s n$*, i.e., to approximately maximize the "logarithm" of determinant within* $\frac{\lambda_c}{\lambda_s} = \frac{8}{7}$.

2.4 Proof of Main Lemma

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

Lemma 2.11 (Union Lemma (Çivril and Magdon-Ismail, 2013, Lemma 6))**.** *Let* $\mathbf{P}$ *and* $\mathbf{Q}$ *be two (finite) sets of vectors in* $\mathbb{R}^d$*. Then, we have the following:*

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

where $\operatorname{d}(\mathbf{v}, \mathbf{Q})$ *denotes the distance of* $\mathbf{v}$ *to the subspace spanned by* $\mathbf{Q}$*; i.e.,* $\operatorname{d}(\mathbf{v}, \mathbf{Q}) \triangleq \|\mathbf{v} - \operatorname{proj}_{\mathbf{Q}}(\mathbf{v})\|$.

Lemma 2.12 ((Çivril and Magdon-Ismail, 2013, Lemma 13))**.** *For any positive integer* ℓ*, there exists a set of* 2^ℓ *vectors* $\mathbf{B}^{(\ell)} = \{\mathbf{b}_1, \ldots, \mathbf{b}_{2^\ell}\}$ *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* $\overline{\mathbf{b}_j} = 2^{-\frac{\ell}{2}} \cdot \mathbf{1} - \mathbf{b}_j$*. 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 $|X| = |Y| = (5n)^\ell$ and $|E| = (75n)^\ell$, 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 $\mathbf{B}^{(3\ell)} = \{\mathbf{b}_1, \ldots, \mathbf{b}_{2^{3\ell}}\}$ 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:

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

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

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. We first prove the completeness.

Lemma 2.13. *If* $\operatorname{val}(\mathfrak{G}^{\otimes \ell}) = 1$*, then there exists a set* $\mathbf{S}$ *of* K *vectors from* $\mathbf{V}$ *such that* $\operatorname{vol}(\mathbf{S}) \geq 1$.

Proof. Let $\sigma : (X \uplus Y) \to \Sigma$ 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 $\pi_e(\sigma(x)) = \sigma(y)$. Furthermore, $\langle \mathbf{v}_{x_1,\sigma(x_1)}, \mathbf{v}_{x_2,\sigma(x_2)} \rangle = 0$ 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 $\operatorname{vol}(\mathbf{S}) = 1$. $\square$

Soundness. We then prove the 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.** $0 \leq |\mathbf{S}| < \frac{7}{8}K$ and **2.** $\frac{7}{8}K \leq |\mathbf{S}| \leq N$.

Soundness 1. $0 \leq |\mathbf{S}| < \frac{7}{8}K$**.**

Lemma 2.14. *Suppose* $\operatorname{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha \ell}$*. For any set* $\mathbf{S}$ *of less than* $\frac{7}{8}K$ *vectors from* $\mathbf{V}$*,* $\operatorname{vol}(\mathbf{S}) \leq 1$.

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

Soundness 2. $\frac{7}{8}K \leq |\mathbf{S}| \leq N$**.** For $\mathbf{S} \subseteq \mathbf{V}$, define

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

Analogous notations are used for $\mathbf{S}_Y$, $Y(\mathbf{S})$, and $\operatorname{rep}(\mathbf{S}_Y)$. Here, $\operatorname{rep}(\mathbf{S}_X)$ and $\operatorname{rep}(\mathbf{S}_Y)$ 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 $\mathrm{rep}(\mathbf{S}_X)$ and $\mathrm{rep}(\mathbf{S}_Y)$, respectively, which is proved in Appendix B for the sake of completeness.

Lemma 2.15 (Lemma 16 of Çivril and Magdon-Ismail (2013)). *For any set* $\mathbf{S}$ *of vectors from* $\mathbf{V}$,

$$\mathrm{vol}(\mathbf{S}_X) \leq (\sqrt{3}/2)^{\mathrm{rep}(\mathbf{S}_X)} \text{ and } \mathrm{vol}(\mathbf{S}_Y) \leq (\sqrt{3}/2)^{\mathrm{rep}(\mathbf{S}_Y)}.$$

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

Claim 2.16. *For any set* $\mathbf{S}$ *of* $k \geq \frac{7}{8}K$ *vectors from* $\mathbf{V}$, *if* $\mathrm{vol}(\mathbf{S}) \geq 2^{-ck}$ *for some number* $c > 0$, *then* $|X(\mathbf{S})| > \left(\frac{3}{7} - 10c\right)k$ *and* $|Y(\mathbf{S})| > \left(\frac{3}{7} - 10c\right)k$.

Proof. Observe first that $\mathrm{vol}(\mathbf{S}_X) \geq \mathrm{vol}(\mathbf{S}) \geq 2^{-ck}$. By Lemma 2.15, we have $(\sqrt{3}/2)^{\mathrm{rep}(\mathbf{S}_X)} \geq \mathrm{vol}(\mathbf{S}_X) \geq 2^{-ck}$, implying that $\mathrm{rep}(\mathbf{S}_X) \leq ck/\log_2(2/\sqrt{3}) < 5ck$. Similarly, $\mathrm{rep}(\mathbf{S}_Y) < 5ck$. Using the facts that $|\mathbf{S}_X| = |X(\mathbf{S})| + \mathrm{rep}(\mathbf{S}_X)$, $|\mathbf{S}_Y| = |Y(\mathbf{S})| + \mathrm{rep}(\mathbf{S}_Y)$, and $k = |\mathbf{S}_X| + |\mathbf{S}_Y|$, we bound $|X(\mathbf{S})|$ from below as follows:

$$\begin{aligned}|X(\mathbf{S})| &= k - |\mathbf{S}_Y| - \mathrm{rep}(\mathbf{S}_X) > k - |Y| - 10ck \\ &= \left(1 - \frac{(5n)^\ell}{k} - 10c\right)k \geq \left(\frac{3}{7} - 10c\right)k.\end{aligned}$$

Similarly, $|Y(\mathbf{S})| > (\frac{3}{7} - 10c)k$. □

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

Lemma 2.17. *Suppose* $\mathrm{val}(\mathfrak{G}^{\otimes \ell}) < 2^{-\alpha\ell}$. *For any set* $\mathbf{S}$ *of* k *vectors from* $\mathbf{V}$ *with* $k \geq \frac{7}{8}K$, *it holds that* $\mathrm{vol}(\mathbf{S}) < 2^{-\beta^\circ k}$, *where* $\beta^\circ = 10^{-10^{12.3}}$.

Proof. The proof is by contradiction. Suppose there exists a set $\mathbf{S}$ of $k \geq \frac{7}{8}K$ vectors from $\mathbf{V}$ such that $\mathrm{vol}(\mathbf{S}) \geq 2^{-\beta^\circ k}$.

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

$$\sigma(z) = \begin{cases} \text{any } i \text{ s.t. } \mathbf{v}_{z,i} \in \mathbf{S}_X & \text{if } z \in X(\mathbf{S}), \\ \text{any } i \text{ s.t. } \mathbf{v}_{z,i} \in \mathbf{S}_Y & \text{if } z \in Y(\mathbf{S}), \\ \text{any element of } \Sigma & \text{otherwise.} \end{cases}$$

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 2.11, we bound the distance of the vectors of $\mathbf{P}$ to $\mathbf{Q}$. Since $\mathbf{Q}$ forms an orthonormal basis by construction and for each $x \in X(\mathbf{S})$, $\|\mathbf{v}_{x,\sigma(x)}\|^2 = \|\mathrm{proj}_{\mathbf{Q}}(\mathbf{v}_{x,\sigma(x)})\|^2 + \mathrm{d}(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q})^2$, we have the following:

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

If an edge $(x,y) \in E$ between $X(\mathbf{S})$ and $Y(\mathbf{S})$ is not satisfied by σ, 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, $\mathrm{d}(\mathbf{v}_{x,\sigma(x)}, \mathbf{Q}) = (1 - \frac{U(x)}{4 \cdot 15^{2\ell}})^{\frac{1}{2}}$, where $U(x)$ is the number of unsatisfied edges between x and $Y(\mathbf{S})$. Using Lemma 2.11 and the fact that $\mathrm{vol}(\mathbf{Q}) \leq 1$, we have

$$\begin{aligned}\mathrm{vol}(\mathbf{P} \cup \mathbf{Q}) &\leq \left(\prod_{x \in X(\mathbf{S})} \left(1 - \frac{U(x)}{4 \cdot 15^{2\ell}}\right)\right)^{\frac{1}{2}} \\ &\leq \left(\frac{1}{|X(\mathbf{S})|} \sum_{x \in X(\mathbf{S})} \left(1 - \frac{U(x)}{4 \cdot 15^{2\ell}}\right)\right)^{\frac{|X(\mathbf{S})|}{2}}, \end{aligned} \tag{1}$$

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 2.16 derives $|X(\mathbf{S})| > (\frac{3}{7} - 10\beta^\circ)k$ and $|Y(\mathbf{S})| > (\frac{3}{7} - 10\beta^\circ)k$. Because less than $2^{-\alpha\ell}$-fraction of edges in E can be satisfied by any labeling (including σ) by assumption, and more than $(\frac{3}{7} - 10\beta^\circ)k \cdot 15^\ell$ 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

$$\left[\left(\frac{3}{7} - 10\beta^\circ\right)\frac{k}{(5n)^\ell} - 2^{-\alpha\ell}\right](75n)^\ell. \tag{2}$$

Consequently, the number of unsatisfied edges *between* $X(\mathbf{S})$ and $Y(\mathbf{S})$ is at least twice Eq. (2) minus "the number of unsatisfied edges incident to $X(\mathbf{S})$ *or* $Y(\mathbf{S})$" (which is at most $(75n)^\ell$); namely,

$$\sum_{x \in X(\mathbf{S})} U(x) \geq \left[\left(\frac{1}{7} - 10\beta^\circ\right)\frac{2k}{(5n)^\ell} - 2^{-\alpha\ell+1}\right](75n)^\ell,$$

where we have used the fact that $k \geq \frac{7}{8}K$. With this inequality, we further expand Eq. (1) as follows:

$$\begin{aligned}\mathrm{vol}(\mathbf{P} \cup \mathbf{Q}) &\leq \left(1 - \frac{\sum_{x \in X(\mathbf{S})} U(x)}{|X(\mathbf{S})|} \frac{1}{4 \cdot 15^{2\ell}}\right)^{\frac{|X(\mathbf{S})|}{2}} \\ &\leq \exp\left(-\frac{\left[(\frac{1}{7} - 10\beta^\circ)\frac{2k}{(5n)^\ell} - 2^{-\alpha\ell+1}\right](75n)^\ell}{|X(\mathbf{S})|} \frac{|X(\mathbf{S})|}{8 \cdot 15^{2\ell}}\right) \\ &\leq \exp\left(-\left[\left(\frac{1}{7} - 10\beta^\circ\right)\frac{1}{4 \cdot 15^\ell} - 2^{-\alpha\ell+1}\frac{1}{14 \cdot 15^\ell}\right]k\right) \\ &= \exp\left(-\frac{1 - 70\beta^\circ - 2^{-\alpha\ell+2}}{28 \cdot 15^\ell}k\right).\end{aligned}$$

Since $\beta^\circ = 10^{-10^{12.3}} < \frac{1 - 2^{-\alpha\ell+2}}{28 \cdot 15^\ell \cdot \log_e(2) + 70}$ (recall that $\ell = \lceil \frac{3}{\alpha} \rceil$) and $\beta^\circ > 0$, we finally have $\mathrm{vol}(\mathbf{S}) \leq \mathrm{vol}(\mathbf{P} \cup \mathbf{Q}) < 2^{-\beta^\circ k}$, a contradiction. □

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 by ¶Our Reduction. Then, the completeness follows from Lemma 2.13 and the soundness follows from Lemmas 2.14 and 2.17. $\square$

3 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 $\propto \det(\mathbf{A}_S)^p$. 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 $\mathcal{Z}^p \leq \widehat{\mathcal{Z}}^p \leq \rho \cdot \mathcal{Z}^p$. 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 3.1. *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* RP = 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 3.1, we have the following inapproximability result on E-DPPs, whose proof is deferred to Appendix B.

Corollary 3.2. *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* RP = NP.

[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 RP $\neq$ NP.

Proof of Theorem 3.1. 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 $\operatorname{maxdet}(\mathbf{A}) \geq 2^{\lambda_c n}$. 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 $(1 - \frac{2^{\lambda_s p n + n}}{2^{\lambda_s p n + n} + 2^{\lambda_c p n}}) - \frac{1}{3} \geq \frac{2}{3} - \frac{1}{1 + 2^{qn - n}} \geq \frac{1}{2}$ (as long as $n \geq \frac{3}{q-1}$), implying that RP = NP. $\square$

Remark 3.3. *Theorem 3.1 holds even when λ_c and λ_s are functions in n. If we apply a $(\frac{9}{8} - \epsilon)$-factor inapproximability by 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 3.2. 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), whose proof is deferred to Appendix B. This means that Corollary 3.2 is tight *up to* a constant in the exponent (when $p \geq 10^{10^{13}}$).

Observation 3.4. *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 $(2 \cdot \mathrm{e}^p)^n$.*

4 Open Questions

We have established exponential inapproximability results for unconstrained MAP inference on DPPs and the normalizing constant for E-DPPs. We conclude this paper with two 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.
- **Smallest exponent p for which $\mathcal{Z}^p$ is inapproximable.** Our upper bound $10^{10,000,000,000,000}$ on p in Corollary 3.2 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. The author would like to thank Leonid Gurvits for pointing out to us the reference (Gurvits, 2009).

References

N. Anari and S. O. Gharan. A generalization of permanent inequalities and applications in counting and optimization. In *STOC*, pages 384–396, 2017.

N. Anari, K. Liu, S. O. Gharan, and C. Vinzant. Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC*, pages 1–12, 2019.

E. Bıyık, K. Wang, N. Anari, and D. Sadigh. Batch active learning using determinantal point processes. *CoRR*, abs/1906.07975, 2019.

A. Borodin and E. M. Rains. Eynard-Mehta theorem, Schur process, and their Pfaffian analogs. *J. Stat. Phys.*, 121(3–4):291–317, 2005.

N. Buchbinder and M. Feldman. Deterministic algorithms for submodular maximization problems. *ACM Trans. Algorithms*, 14(3):1–20, 2018.

N. Buchbinder, M. Feldman, J. S. Naor, and R. Schwartz. A tight linear time (1/2)-approximation for unconstrained submodular maximization. *SIAM J. Comput.*, 44(5):1384–1402, 2015.

W.-L. Chao, B. Gong, K. Grauman, and F. Sha. Large-margin determinantal point processes. In *UAI*, pages 191–200, 2015.

L. Chen, G. Zhang, and E. Zhou. Fast greedy MAP inference for determinantal point process to improve recommendation diversity. In *NeurIPS*, pages 5622–5633, 2018.

A. Çivril and M. Magdon-Ismail. On selecting a maximum volume sub-matrix of a matrix and related problems. *Theor. Comput. Sci.*, 410(47–49):4801–4811, 2009.

A. Çivril and M. Magdon-Ismail. Exponential inapproximability of selecting a maximum volume sub-matrix. *Algorithmica*, 65(1):159–176, 2013.

M. Di Summa, F. Eisenbrand, Y. Faenza, and C. Moldenhauer. On largest volume simplices and sub-determinants. In *SODA*, pages 315–323, 2014.

I. Dinur and D. Steurer. Analytical approach to parallel repetition. In *STOC*, pages 624–633, 2014.

U. Feige. A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4):634–652, 1998.

M. Gartrell, I. Han, E. Dohmatob, J. Gillenwater, and V.-E. Brunel. Scalable learning and MAP inference for nonsymmetric determinantal point processes. *CoRR*, abs/2006.09862, 2020.

J. Gillenwater, A. Kulesza, and B. Taskar. Near-optimal MAP inference for determinantal point processes. In *NIPS*, pages 2735–2743, 2012.

J. A. Gillenwater. *Approximate Inference for Determinantal Point Processes*. PhD thesis, University of Pennsylvania, 2014.

B. Gong, W. Chao, K. Grauman, and F. Sha. Diverse sequential subset selection for supervised video summarization. In *NIPS*, pages 2069–2077, 2014.

L. Gurvits. On the complexity of mixed discriminants and related problems. In *MFCS*, pages 447–458, 2005.

L. Gurvits. On complexity of the mixed volume of parallelograms. In *EuroCG*, pages 337–341, 2009.

I. Han and J. Gillenwater. MAP inference for customized determinantal point processes via maximum inner product search. In *AISTATS*, pages 2797–2807, 2020.

I. Han, P. Kambadur, K. Park, and J. Shin. Faster greedy MAP inference for determinantal point processes. In *ICML*, pages 1384–1393, 2017.

J. Håstad. Clique is hard to approximate within $n^{1-\varepsilon}$. *Acta Math.*, 182(1):105–142, 1999.

J. Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

C.-W. Ko, J. Lee, and M. Queyranne. An exact algorithm for maximum entropy sampling. *Oper. Res.*, 43(4):684–691, 1995.

I. Koutis. Parameterized complexity and improved inapproximability for computing the largest j-simplex in a V-polytope. *Inf. Process. Lett.*, 100(1):8–13, 2006.

A. Kulesza and B. Taskar. Learning determinantal point processes. In *UAI*, pages 419–427, 2011.

A. Kulesza and B. Taskar. Determinantal point processes for machine learning. *Found. Trends Mach. Learn.*, 5(2–3):123–286, 2012.

D. Lee, G. Cha, M.-H. Yang, and S. Oh. Individualness and determinantal point processes for pedestrian detection. In *ECCV*, pages 330–346, 2016.

O. Macchi. The coincidence approach to stochastic point processes. *Adv. Appl. Probab.*, 7(1):83–122, 1975.

Z. E. Mariet, S. Sra, and S. Jegelka. Exponentiated strongly Rayleigh distributions. In *NeurIPS*, pages 4464–4474, 2018.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of the approximations for maximizing submodular set functions—I. *Math. Program.*, 14: 265–294, 1978.

A. Nikolov. Randomized rounding for the largest simplex problem. In *STOC*, pages 861–870, 2015.

N. Ohsaka and T. Matsuoka. On the (in)tractability of computing normalizing constants for the product of determinantal point processes. In *ICML*, pages 6600–6609, 2020.

R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

J. Robinson, S. Sra, and S. Jegelka. Flexible modeling of diversity with strongly log-concave distributions. In *NeurIPS*, pages 15199–15209, 2019.

S. Tamaki. Parallel repetition of two-prover one-round games: An exposition. *Interdiscip. Inf. Sci.*, 21(4): 289–306, 2015.

L. Trevisan. Inapproximability of combinatorial optimization problems. *CoRR*, abs/cs/0409043, 2004.

V. V. Vazirani. *Approximation Algorithms*. Springer Science & Business Media, 2013.

M. Wilhelm, A. Ramanathan, A. Bonomo, S. Jain, E. H. Chi, and J. Gillenwater. Practical diversified recommendations on YouTube with determinantal point processes. In *CIKM*, pages 2165–2173, 2018.

J. Yao, F. Fan, W. X. Zhao, X. Wan, E. Y. Chang, and J. Xiao. Tweet timeline generation with determinantal point processes. In *AAAI*, pages 3080–3086, 2016.

M. J. Zhang and Z. Ou. Block-wise MAP inference for determinantal point processes with application to change-point detection. In *SSP*, pages 1–5, 2016.

J. Y. Zou and R. P. Adams. Priors for diversity in generative latent variable models. In *NIPS*, pages 2996–3004, 2012.