Composable Coresets for Constrained Determinant Maximization and Beyond

Sepideh Mahabadi
Microsoft Research Redmond

Thuy Duong Vuong
University of California, San Diego

Abstract

We study algorithms for construction of *composable coresets* for the task of *Determinant Maximization* under *partition constraint*. Given a point set $V \subset \mathbb{R}^d$ that is partitioned into s groups $V_1, \cdots, V_s$, and integers $k_1, ..., k_s$, where $k = \sum_i k_i$, the goal is to pick k_i points from group V_i such that the overall determinant of the picked k points is maximized. Determinant Maximization and its constrained variants have gained a lot of interest for modeling diversity, and have found applications in the context of data summarization. When the cardinality k of the selected set is greater than the dimension d, we show a peeling algorithm that gives us a composable coreset of size kd with a provably optimal approximation factor of $d^{O(d)}$. When $k \leq d$, we show a simple coreset construction with optimal size and approximation factor. As a further application of our technique, we get a composable coreset for determinant maximization under the more general laminar matroid constraints, and a composable coreset for unconstrained determinant maximization in a previously unresolved regime. Our results generalize to all strongly Rayleigh distributions and to several other experimental design problems. As an application, we improve the runtime of the practical local-search based algorithm of [Anari-Vuong–COLT'22] for determinantal maximization under partition constraint from $O(n^{2^s} k^{2^s})$ to $O(nk^{2^s})$, making it only linear on the number of points n.

1 INTRODUCTION

Determinant maximization is a fundamental optimization problem that arises in various domains such as data summarization, experimental design, computational geometry, and machine learning. At its core, the problem involves selecting a subset of items, typically vectors, such that the selected set is as diverse or informative as possible. A common way to quantify this diversity is via the determinant of a submatrix derived from the selected vectors.

Before formally defining the problem, let us start by recalling the notation $\det_k$, a generalization of the standard determinant.

Definition 1 ($\det_k$)**.** Given a $d \times d$ matrix $M \in \mathbb{R}^{d \times d}$ and $k \leq d$, $\det_k(M)$ denotes the sum of determinants of all $k \times k$ principal submatrices of M:

$$\det_k(M) = \sum_{|S|=k} \det(M_{S,S}),$$

where $M_{S,S}$ is the principal sub-matrix indexed by $S \subseteq [d]$.

Given a collection of n vectors $V = \{v_1, \cdots, v_n\}$ in $\mathbb{R}^d$, and a target subset size k (not necessarily smaller than d), the *determinant maximization* problem seeks a subset $S \subseteq [n]$, $|S| = k$, that maximizes the quantity

$$\phi(S) := \det_{\min\{k,d\}}(L(S))$$

where $L(S) := A_S A_S^\intercal = \sum_{i \in S} v_i v_i^\top \in \mathbb{R}^{d \times d}$ and A_S is the $d \times k$ matrix whose columns are the vectors v_i for $i \in S$.

Geometrically, if $k \leq d$, the objective function is equal to the volume squared of the parallelepiped spanned by the selected vectors. In this setting, the best approximation guarantee is e^k Nikolov (2015), which is essentially tight Civril and Magdon-Ismail (2013) unless P = NP.

On the other hand, if $k > d$, the objective can be rewritten as $\det(A_S^\intercal A_S)$, and problem is known as the D-optimal design problem. The best known approximation algorithm in this regime achieves a factor of $\min\{e^k, (\frac{k}{k-d})^d\}$, which is always at most $\leq e^{O(d)}$ and becomes a constant when for example $k \geq d^2$ Madan et al. (2019).

Due to its connection with subset diversity, determinant maximization and its variants have been widely used in modern data analysis, particularly for summarization tasks

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

where one aims to select a compact, informative, and representative subset from large-scale data, and thus studied extensively over the last decade Mirzasoleiman et al. (2017); Gong et al. (2014); Kulesza et al. (2012); Chao et al. (2015); Kulesza and Taskar (2011); Yao et al. (2016); Lee et al. (2016).

Determinant Maximization under partition and matroid constraints. Diversity maximization problems, including determinant maximization, have been studied extensively under partition and more generally under matroid constraints Madan et al. (2020); Nikolov and Singh (2016); Abbassi et al. (2013); Moumoulidou et al. (2020); Addanki et al. (2022); Mahabadi and Trajanovski (2023). These constraints are important in real-world applications where certain fairness or grouping criteria must be respected. In the simpler case of a partition constraint, the data set V is partitioned into s groups $V_1, \cdots, V_s$ and we are provided with s numbers $k_1, \cdots, k_s$, and the goal is to pick k_i points from each group i such that the overall determinant (or more generally diversity) of the chosen $k = \sum_i k_i$ points is maximized. This setting allows for fine-grained control over the contribution of each group in the selected summary: for example, limiting the number of movies from each genre in a recommendation system, and has further applications in the context of fair and balanced data summarization (see e.g. Mahabadi and Trajanovski (2023)). More generally, given a matroid $([n], \mathcal{I})$ of rank k, the problem of finding a basis of the matroid that maximizes the determinant admits a $\min\{k^{O(k)}, d^{O(d)}\}$ approximation, and it improves to $\min\{e^{O(k)}, d^{O(d)}\}$ for the *estimation* problem where the goal is to only estimate the value of the optimal solution Madan et al. (2020); Nikolov and Singh (2016); Brown et al. (2022b,a).

Composable Coresets. As one of the main applications of determinant maximization is in data summarization, the problem has been considered extensively in massive data computation models Mirzasoleiman et al. (2017); Wei et al. (2014); Pan et al. (2014); Mirzasoleiman et al. (2013, 2015); Mirrokni and Zadimoghaddam (2015); Barbosa et al. (2015). In this work, we focus on designing *composable coreset* for determinant maximization. A *Coreset* is a *small subset* of the data that is sufficient for computing an approximate solution to a pre-specified optimization problem on the whole dataset Agarwal et al. (2005). More specifically, we present a summarization algorithm $\mathcal{A}$ that, given a data set V, produces a subset of V. Moreover, we want our coresets to be composable Indyk et al. (2014): that is if we have multiple datasets $V^{(1)}, \cdots, V^{(m)}$ (note that in the context of, e.g., a partition constraint, each data set $V^{(i)}$ is itself partitioned into groups $V_1^{(j)}, \cdots, V_s^{(j)}$), then the union of coresets $\mathcal{A}(V^{(1)}) \cup \cdots \cup \mathcal{A}(V^{(m)})$ should be sufficient for computing an approximate solution for the union of the datasets $V^{(1)} \cup \cdots \cup V^{(m)}$ (see Section 2.3 for a formal definition).

As shown in Indyk et al. (2014), having a composable coreset for an optimization task automatically yields a solution for the same task in several massive data models including distributed/parallel and streaming models. For example, in a distributed setting where the whole data is partitioned over multiple machines, each machine can compute a coreset for its own data, and only send this small summary to a single aggregator. The aggregator then processes the union of the summaries and outputs the solution. Due to their applications, several works have focused on designing composable coresets for determinant maximization, and more broadly, diversity maximization, over the past decade Indyk et al. (2014, 2020); Mahabadi et al. (2019); Mirrokni and Zadimoghaddam (2015); Moumoulidou et al. (2020); Ceccarello et al. (2018, 2020); Zadeh et al. (2017); Mahabadi and Narayanan (2023); Mahabadi and Trajanovski (2023).

Prior work. Composable coresets have been designed for the unconstrained determinant maximization problem. More precisely, for $k \leq d$, one can get a $k^{O(k)}$- approximate coreset of size $O(k)$, which is also known to be tight Indyk et al. (2020); Mahabadi et al. (2019) (see the first row of Table 1). Furthermore, for $k \geq d$, if the solution is allowed to pick vectors from V *with repetition* (which we refer to as the "with-repetition" setting), then Mahabadi et al. (2019) gives a coreset of size $\tilde{O}(d)$ with approximation factor $\tilde{O}(d)^d$ (Second row of Table 1). However, the case where the solution is required to pick *distinct* points from V (also known as the "without-repetition" setting) remains open: when $k \gg d$, Mahabadi et al. (2019)'s size-$\tilde{O}(d)$ coreset clearly does not contain enough points to construct a solution that consists of k distinct points. Determinant maximization in the without-repetition setting is generally harder and has received significantly more interest than the with-repetition setting (see Madan et al., 2019; Lau and Zhou, 2021; Brown et al., 2022b), as its solution often provides a much better summary of the original dataset [1].

1.1 Our Results

In this work, we establish the following contributions.

Algorithms. A summary of our coreset construction algorithms is provided in Table 1.

[1] Consider the following illustrative example. Suppose $d = 2$ and the data set V consists of $v_1 \equiv Me_1, v_2 \equiv Me_2, v_3, \ldots, v_n$ where e_1, e_2 are standard basis vectors and $v_3, \cdots, v_n$ are arbitrary vectors in $\mathbb{R}^2$. For large enough M, the coreset C of V for $k \geq 2$ in the "with repetition" setting will only contain the 2 vectors v_1 and v_2, since the set consisting of $k/2$ copies of v_1 and of v_2 has the biggest possible determinant. But we cannot say that this set is a good summary of the data set since we ignore the information provided by $v_3, \ldots, v_n$. On the other hand, a coreset for $k \geq 2$ in the "without repetition" setting would necessarily consider the remaining vectors $v_3, \ldots, v_n$, and thus is a better representation of the data set.

Table 1: Our upper bound results on composable coresets for determinant maximization. Here s is the number of groups in the partition constraints. Note that the third row follows from our results on partition constraint but we spell it out to compare to the previous result.

	size	approx.	constraint
$k \leq d$	k	$O(k)^{2k}$	cardinality Indyk et al. (2020) Mahabadi et al. (2019)
$k \geq d$	$\tilde{O}(d)$	$\tilde{O}(d)^{2d}$	cardinality (with repetition) Indyk et al. (2020)
$k \geq d$	kd	d^{2d}	cardinality (w.o. repetition) This work
$k \leq d$	sk	k^{2k}	partition This work
$k \geq d$	kd	d^{2d}	partition This work
$k \leq d$	k^{2k}	k^{2k}	laminar This work
$k \geq d$	$(kd)^k$	d^{2d}	laminar This work

- In Section 3, we construct coresets for unconstrained determinant maximization in the without-repetition setting (third row in Table 1), previously left open in prior work.
- In Section 4, we develop efficient composable coresets for determinant maximization under *partition* and *laminar* matroid constraints, as shown in rows 4–7 of Table 1, and verified in Theorem 26. Our results extends to *Strongly Rayleigh* distributions (see Theorem 24), and are obtained via an improved exchange inequality for determinant when $k > d$.
- In Appendix C (Theorem 36),we demonstrate an application of our results to design composable coresets for a broader class of experimental design problems in the without-repetition setting, for all regular objective functions. This complements the result of Indyk et al. (2020) for experimental design in the with-repetition setting.

Lower bounds. We complement our results with the following lower bounds shown in Appendix B.

- In Lemma 28, we show that for $k \leq d$, any composable coreset for determinant maximization under partition constraint with a finite approximation factor must have size at least $\Omega(sk)$. This shows that our construction (fourth row in Table 1) is essentially tight, since our approximation factor matches that of Indyk et al. (2020) for the unconstrained version of the problem, which is known to be tight.
- In Lemma 29, we show that for $k \geq d$, any composable coreset for the problem under partition constraint with a finite approximation factor must have size at least $k + d(d-1)$. This partially complements our result in the fifth row of Table 1, and shows that for $k = O(d)$, our coreset size cannot be improved. We leave it as future work to find the tight coreset size when $k = \omega(d)$.
- In Theorem 31, we prove that for $d \leq k \leq \text{poly}(d)$ and coreset of polynomial size in k, the approximation factor of $d^{O(d)}$ is essentially the best possible for the unconstrained determinant maximization problem in the without-repetition setting. This matches the approximation factor of our construction in the third row of Table 1.

We leave the question of improving the coreset size of the laminar matroid case (or proving a lower bound) to future work.

Application. Our coreset can be constructed in essentially linear time in n, the number of data points. Hence, by first constructing a coreset then applying any standard determinant maximization algorithm on the coreset, we obtain an algorithm for the determinant maximization problem under partition constraint that runs in time $O(n\text{poly}(k))$ (see Lemma 27). Thus, if we use the multi-step local search algorithm by Anari and Vuong (2022) for determinant maximization under partition constraint, then we obtain a practical local-search-based algorithm whose runtime improves from $O(n^{2^s}k^{2^s})$ in Anari and Vuong (2022) to $O(nk^{2^s})$ where s is the number of parts in the partition.

1.2 Roadmap

We give an overview of our approach in Section 1.3. Section 2 contains the preliminaries. In Section 3, we give our results for the unconstrained version of the problem and present our peeling algorithm. We then use our peeling algorithm in Section 4 to get composable coresets for the problem under the partition and laminar matroid constraints. Deferred proofs are presented in Appendix A. Finally, we present our lower bound results in Appendix B, and our results on other experimental design problems in Appendix C.

1.3 Overview of the Techniques

Let us give a brief overview of our approach. Consider a set of vectors $V = \{v_1, \cdots, v_n\} \subseteq \mathbb{R}^d$. As mentioned earlier, when $k \leq d$, the objective function for any subset $T \in \binom{[n]}{k}$ defined as $\phi(T) = \det_k(\sum_{i \in T} v_i v_i^\intercal)$ corresponds to the square of the volume spanned by the vectors in $V_T := \{v_i \mid i \in T\}$. Mahabadi et al. (2019) showed that in this setting, any local maximum $U \subseteq V$ of size k with respect

to $\det_k(\cdot)$ approximately preserves the *k-directional height* of the set V. More precisely, for any set $V_S \subset \mathbb{R}^d$ of k vectors and any $v \in V_S \cap V$, one can replace v with some u in U so that the distance $d(u, \mathcal{H})$ from u to the $(k-1)$-dimensional subspace $\mathcal{H}$ spanned by $V_S \setminus \{v\}$ is at least $\frac{1}{k} \cdot d(v, \mathcal{H})$. Thus

$$k^2 \det_k(uu^\intercal + \sum_{w \in V_S \setminus \{v\}} ww^\intercal) \geq \det_k(\sum_{w \in V_S} ww^\intercal).$$

Thus, all elements of V_S can be successively replaced by elements of U while only incurring a factor $k^{O(k)}$ increase in the objective function $\det_k(\cdot)$ and thus U is a $k^{O(k)}$ composable coreset w.r.t. $\det(\cdot)$.

Generalization of directional height. In this work, we *extend* the notion of directional height to the regime where $k \geq d$. We show that for any local maximum U of size d with respect to $\det(\cdot)$, the following holds: For any index set $S \subseteq [n]$ of size k, and letting $V_S = \{v_i : i \in S\}$ and $v \in V_S \cap V$, there exists $u \in U$ s.t.

$$\det(d^2 uu^\intercal + \sum_{w \in V_S \setminus \{v\}} ww^\intercal) \geq \det(\sum_{w \in V_S} ww^\intercal)^2.$$

This already implies that U forms a $d^{O(d)}$-composable coreset for determinant maximization in the with-repetition setting, i.e., when the selected subset is allowed to contain duplicate vectors.

The without-repetition setting. The without-repetition case is more delicate, as we must ensure that $(V_S \setminus \{v\}) \cup \{u\}$ is a proper subset, i.e., $u \notin V_S \setminus \{v\}$. To handle this, we apply the idea of peeling coresets, previously used for constructing robust coresets that tolerate outliers Agarwal et al. (2008); Abbar et al. (2013). Our construction repeatedly peels away local optimum solutions from the input set, and takes the union of all the peeled local optimums to be the final coreset. By the pigeon hole principle, for any set V_S not fully contained in the final coreset, there exists at least one peeled-away local optimum that is disjoint from V_S. Consequently, we can replace an element of V_S by an element inside this local optimum without creating a multiset.

Partition and laminar constraints. For determinant maximization under partition constraint, our coreset construction is simple and intuitive when $k \leq d$: we take the union of the coresets for each partition part. When $k \geq d$, we construct a coreset of size kd with approximation factor $d^{O(d)}$ by taking the union of the peeling coresets for each part of the partition. For the laminar matroid constraint case, we apply the peeling coreset idea to ensure that for any subset V_S satisfying the laminar constraint, there exists one peeled-away subset U s.t. replacing an element of V_S by an element of U will not violate the laminar constraint.

[2]This is precisely $\phi(T) = \det_{\min\{k,d\}}(\cdot)$ when $k \geq d$.

Strongly Rayleigh distributions. The fixed-size determinantal point processes (DPP), a distribution over subsets $\binom{[n]}{k}$ defined by $\Pr[S] \propto \det(S)$, belongs to the class of *Strongly Rayleigh* distributions (see Section 2.5 for details). All of our results readily extend to maximum a posteriori problems, i.e., find $\arg\max_S \mu(S)$ for any strongly Rayleigh objective functions $\mu(\cdot)$. This is because our proof relies only on an *exchange inequality* which is satisfied by all strongly Rayleigh distributions. Exchange inequalities offer a unifying framework for analyzing our coreset constructions across all settings. To the best of our knowledge, this is the first work to leverage exchange inequalities in the context of coreset construction.

Experimental design. Our construction also applies to experimental design problems with respect to other, not necessarily strongly Rayleigh, objective functions, such as matrix traces (A-design) or condition number (E-design). By replacing the base-level building blocks in our construction, i.e., the local optimum w.r.t. $\det(\cdot)$, with spectral spanners Indyk et al. (2020), we can guarantee that the union of the coresets contains a feasible fractional solution as a combination of input vectors that achieves a good value. However, the algorithm to round the fractional solution to an integral solution under matroid constraint only exists in limited cases of objective function other than the determinant.

Lower bounds. For lower bound on the size of composable coreset for determinant maximization under partition constraint, when $k \leq d$, we show that any coreset that achieves a finite approximation factor must include at least k vectors from each part of the partition constraints $P_1, \cdots, P_s$, leading to a coreset size of sk. To see why, suppose U is a coreset for a dataset Q, and suppose that $|U \cap P_1| \leq k-1$. Consider a new dataset Q' consisting of vectors of arbitrarily large norm that all lie in $span(U)$, which has dimension $\leq k-1$. Moreover, we assume that Q' consists of k_2 vectors in P_2, k_3 vectors in Q_3,..., and k_s vectors in P_3, so that selecting all vectors in Q' saturates the partition constraints in parts $P_2, \cdots, P_s$. In the combined dataset $Q \cup Q'$, the determinant maximizer will select all vectors in Q' (since their norms dominate those in Q) and at least one vector in Q lying outside $span(U)$. However, in the combined coreset $U \cup Q'$, any maximizer in $U \cup Q'$ must either (1) drop some vectors in Q', which leads to a much smaller determinant or (2) takes all vectors in Q', and the vectors in $U \cap P_1$ (since Q' saturates the constraint in $P_2, \cdots, P_2$), forcing the chosen set to lie entirely in a $(k-1)$-dimensional subspace, and thus has determinant 0.

When $k \geq d$, we show an analogous result: any such coreset must include at least d vectors from at least d d parts of the partition. Finally, to prove that the approximation factor of $d^{O(d)}$ is the best possible when $\text{poly}(d) \geq k \geq d$, we use a similar construction as the one used in Indyk et al. (2020)'s lower bound for the approximation factor when

$k \leq d$.

2 PRELIMINARIES

Let $[n]$ denote the set $\{1, \cdots, n\}$. For a set U, we use $\binom{U}{k}$ to denote the family of all size-k subsets of U. For a set V of vectors $\{v_1, \cdots, v_n\}$ and $S \subseteq [n]$, we use V_S to denote the set $\{v_i \mid i \in S\}$. For sets U, W, we use $U + W$ and $U - W$ to denote $U \cup W$ (union) and $U \setminus W$ (set-exclusion) respectively. For singleton subsets, we abuse notation and write $U - e$ ($U + e$ resp.) for $U - \{e\}$ ($U + \{e\}$ resp.).

For a matrix $M \in \mathbb{R}^{n \times n}$ and $S \subseteq [n]$, we use M_S to denote the principal submatrix of M whose rows and columns are indexed by S. We use $\mathbb{S}_d^+$ to denote the set of all symmetric positive semi-definite matrices in $\mathbb{R}^{d \times d}$.

Definition 2 (Local optima)**.** For a function $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ and $\zeta \geq 1$, we say U is an ζ-approximate local optima of μ iff $\zeta \mu(U) \geq \max_{e \in U, f \in [n] \setminus U} \mu(U - e + f)$.

When $\zeta = 1$, we simply refer to U as a local optima.

2.1 Matroids

We say a family of sets $\mathcal{B} \subseteq \binom{[n]}{k}$ is the family of bases of a matroid if $\mathcal{B}$ satisfies the basis exchange axiom: for any two bases $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \setminus B_2$, there exists $y \in B_2 \setminus B_1$ such that $B_1 - x + y \in \mathcal{B}$. We call k the rank of the matroid, and $[n]$ the ground set of the matroid. We let the family of independent sets of the matroid be $\mathcal{I} = \{I \in 2^{[n]} \mid \exists B \in \mathcal{B} : I \subseteq B\}$.

The family $\binom{[n]}{k}$ of all size-k subsets of $[n]$ forms the set of bases of the *uniform matroid* of rank k over $[n]$. We define two simple classes of matroids that are widely used in applications.

Definition 3 (Partition matroid)**.** Given a partition of $[n]$ into $P_1, \cdots, P_s$ and integers $k_1, \cdots, k_s$, the associated partition matroid is defined as follows. We say a set $I \subseteq [n]$ is independent iff $|I \cap P_i| \leq k_i$, for $\forall i \in [s]$. The rank of the matroid is $k := \sum_{i=1}^{s} k_i$.

Definition 4 (Laminar matroid)**.** A family $\mathcal{F}$ of subsets of $[n]$ is laminar iff for any $F_1, F_2 \in \mathcal{F}$ either F_1 and F_2 are disjoint or F_1 contains F_2 or F_2 contains F_1. Given a laminar family $\mathcal{F}$ and integers k_F for each $F \in \mathcal{F}$, the associated laminar matroid is defined by: a set $I \subseteq [n]$ is independent iff $|I \cap F| \leq k_F$ for $\forall F \in \mathcal{F}$. The maximal independent sets have the same cardinality k, and they form the bases of the laminar matroid.

We assume that $k_F > 0$, for $\forall F \in \mathcal{F}$, otherwise we can remove the set F from $\mathcal{F}$ and all elements in F from the ground set. For two sets F_1, F_2 in $\mathcal{F}$ s.t. $F_1 \subseteq F_2$, we can assume $k_{F_2} > k_{F_1}$, otherwise the constraint on $I \cap F_1$ is redundant and F_1 can be removed from the laminar family. We call such a laminar family *non-redundant*.

2.2 Determinant Maximization and Experimental Design Problems

Given vectors $v_1, \cdots, v_n \in \mathbb{R}^d$ and a matroid $\mathcal{M} = ([n], I)$, determinantal point processes (DPP) under matroid constraint samples a basis $S \subseteq [n]$ of $\mathcal{M}$ such that

$$\Pr[S] \sim \det_{\min\{k,d\}}(\sum_{i \in S} v_i v_i^\intercal).$$

This distribution favors diversity, since sets of vectors that are more linearly independent (i.e., different from each other) are assigned higher probabilities. The fundamental optimization problem associated with DPPs, and probabilistic model in general, is to find a "most diverse" subset by computing $\arg\max_{S \text{ is a basis of } \mathcal{M}} \Pr[S]$ i.e. solving the maximum a posteriori (MAP) inference problem.

When the matroid is the uniform matroid, we simply refer to the problem as the *determinant maximization* problem.

When $k \leq d$, $\Pr[S]$ is proportional to the squared volume of the parallelepiped spanned by the elements of S. Thus MAP-inference for DPPs is also known as the volume maximization problem.

Determinant maximization is also known as the D-design problem, since the objective function is the (D)eterminant. Other objective functions have also been studied, for example, matrix traces (A-desing) or condition number (E-design). We discuss these different objective functions in more details in Appendix C.

The setting where S is allowed to be a multiset has also been studied. This is known as the experimental design problem *with-repetition* Allen-Zhu et al. (2017); Madan et al. (2019), as opposed to the *without-repetition* setting where S needs to be a proper subset. The with-repetition setting is generally easier: it can be reduced to the without-repetition setting by duplicating each vector k times.

2.3 Composable Coresets

In the context of the optimization problem on $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, a function c that maps any set $V \subseteq [n]$ to one of its subsets is called an α-composable coreset (Indyk et al. (2014)) if it satisfies the following condition: given any collection of m sets $V^{(1)}, \cdots, V^{(m)} \subseteq [n]$

$$\alpha \cdot \max\{\mu(S) \mid S \subseteq \bigcup_{i=1}^{m} c(V^{(i)})\}$$
$$\geq \max\{\mu(S) \mid S \subseteq \bigcup_{i=1}^{m} V^{(i)}\}$$

If we are further given a matroid $\mathcal{M} = ([n], I)$ to satisfy, we additionally require S to be a basis of $\mathcal{M}$ in both sides of the above inequality. Finally, we say that c is a coreset of *size* t if $|c(V)| \leq t$ for all sets V. Composable coresets

are very versatile; once a composable coreset is designed for a task, it automatically imples efficient streaming and distributed algorithms for the same task.

2.4 Directional Height

For this subsection let $k \leq d$.

Definition 5 (Directional height and k-directional height Mahabadi et al. (2019))**.** For a set $V \subseteq \mathbb{R}^d$ of vectors and a unit vector x, the directional height of V w.r.t. x is $h(V, x) = \max_{v \in V} |\langle v, x \rangle|$.

The k-directional height of V w.r.t. a $(k-1)$-dimensional subspace H is $h(V, H) = \max_{v \in V, x \in H^\top} |\langle v, x \rangle|$ where $H^\top$ is the $(d-k+1)$-dimensional subspace perpendicular to H.

Theorem 6 (Coreset for k-directional height Mahabadi et al. (2019))**.** *Let $k \leq d$ and $V \subseteq \mathbb{R}^d$. Then any size k local optimum U w.r.t. $\det(\cdot)$ inside V approximately preserves the k-directional height. That is, for any $(k-1)$-dimensional subspace H*

$$h(U, H) \geq \frac{1}{k} h(V, H).$$

2.5 Strongly Rayleigh Distribution and Exchange Inequalities

Let $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$, be a distribution over size-ℓ subsets of $[n]$. Its generating polynomial is defined as

$$g_\mu(z_1, \cdots, z_n) = \sum_{S \in \binom{[n]}{\ell}} \nu(S) \prod_{i \in S} z_i.$$

Definition 7 (Strongly Rayleigh)**.** A distribution $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ is *strongly Rayleigh* (or *real-stable*) if its generating polynomial g_ν has no roots in the upper-half of the complex plane. That is, $g_\nu(z_1, \cdots, z_n) \neq 0$, if for $\forall i : \mathrm{Im}(z_i) > 0$.

Strongly Rayleigh distributions satisfy the following *exchange inequality*, which implies that for any local optimum subset U (w.r.t. a strongly Rayleigh distribution ν), and any set $W \in \mathrm{supp}(\nu)$, we can replace an element of W for an element of U while approximately preserving the value of $\nu(\cdot)$. This property will later be used to show that U can serve as a coreset by successively replacing all elements of W with elements of U without significantly losing the objective value.

Lemma 8. *Exchange inequality (Anari et al., 2020, Lemma 26) Let $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ be a strongly Rayleigh distribution. Let $V \subseteq [n]$ be an arbitrary subset. For $\zeta \geq 1$, let $U \subseteq V$ be a ζ-local optimum w.r.t. ν, and assume $\nu(U) \neq 0$. Then for any $W \in \binom{[n]}{\ell}$, and $e \in W \setminus U$*

$$\nu(W)\nu(U) \leq \ell \cdot \sum_{j \in U \setminus W} \nu(W - e + j)\nu(U + e - j)$$

In particular, if $e \in V$, then by approximate local optimality of U within V, we have

$$\nu(W) \leq \zeta\ell \cdot \sum_{j \in U} \nu(W - e + j) \leq (\zeta\ell)^2 \cdot \max_{j \in U} \nu(W - e + j)$$

where we implicitly understand that if $j \in W - e$ then $W - e + j$ is not a proper set, and $\nu(W - e + j) = 0$.

In the context of determinant maximization with given input vectors $v_1, \cdots, v_n \in \mathbb{R}^d$, for any $k \leq d$, $\nu(S) = \det_k(\sum_{i \in S} v_i v_i^\intercal)$ defines a strongly Rayleigh distribution over subsets $S \in \binom{[n]}{k}$ Borcea et al. (2009); Anari et al. (2016).

3 UNCONSTRAINED CASE: PEELING CORESET

As mentioned in the overview of the techniques, if U is a composable coreset for V with respect to the function $\mu(\cdot)$, then for any set S, we can replace any element in $S \cap V$ with an element of U while not significantly reducing $\mu(\cdot)$. We formalize this intuition with the following definition.

Definition 9 (Value-preserving set)**.** Given $\tilde{\mu} : [n]^k \to \mathbb{R}_{\geq 0}$ and $V \subseteq [n]$, we say $U \subseteq V$ is *value-preserving* with respect to $\tilde{\mu}$ if for any $S \in \binom{[n]}{k}$ and $e \in S \cap V$, there exists $f \in U \setminus (S - e)$ s.t. $\tilde{\mu}(S) \leq \tilde{\mu}(S - e + f)$.

The next lemma shows the relationship between value-preserving sets and composable coresets.

Lemma 10. *Suppose functions $\mu, \tilde{\mu} : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ satisfy that $\mu(S) \leq \tilde{\mu}(S) \leq \alpha\mu(S)$ for all S, and let c be a coreset map c such that $U := c(V) \subseteq V$ is* value-preserving *with respect to $\tilde{\mu}$. Then c gives an α-composable coreset with respect to μ.*

Proof. Consider a collection of datasets $V^{(1)}, \cdots, V^{(m)}$, let $U_i := c(V^{(i)})$ for $\forall i \in [m]$, and let S be an arbitrary size-k subset of $\bigcup_{i=1}^m V^{(i)}$. Since each U_i is value-preserving w.r.t. $\tilde{\mu}$, for any $e \in S \cap V^{(i)}$, we can replace e with $f \in U_i \setminus (S - e)$ while keeping $\tilde{\mu}(\cdot)$ non-decreasing. Thus, we successively replace $E := S \setminus \bigcup_{i=1}^m U_i$ with $L \subseteq \bigcup_{i=1}^m U_i$ while ensuring that

$$\tilde{\mu}(S) \leq \tilde{\mu}(S - E + L)$$

Moreover, $S - E + L \in \bigcup_{i=1}^m U_i$ and

$$\mu(S) \leq \tilde{\mu}(S) \leq \tilde{\mu}(S - E + L) \leq \alpha\mu(S - E + L).$$

By choosing S such that $\mu(S) = \max\{\mu(S') \mid S' \subseteq \bigcup_{i=1}^m V^{(i)}, |S'| = k\}$ we get the desired conclusion. □

We now show that an (approximate) local optima with respect to $\det(\cdot)$ is value preserving for suitably chosen functions. When $k \leq d$, Mahabadi et al. (2019) shows that

any size-k local optimum U with respect to $\det(\cdot)$ approximately preserves k-directional height (see Theorem 6), and hence is a value-preserving set with respect to $\tilde{\mu}$, where $\tilde{\mu}$ is defined by

$$\tilde{\mu}(S) = \det_k(\sum_{i\in S} k^{\mathbf{1}[i\in U]} v_i v_i^\intercal)$$

This is easy to see since for $|S| = k \leq d$, as $\mu(S) = \det_k(\sum_{v\in S} vv^\intercal)$ is precisely the square of the volume spanned by vectors in S, i.e., $\mu(S) = \text{Vol}^2(\{v_i | i \in S\})$ and $\tilde{\mu}(S) = k^{2|U\cap S|}\text{Vol}^2(\{v_i | i \in S\})$.

Below, we show that for $k \geq d$, a size-d local optimum U is value-preserving with respect to $\tilde{\mu}$ defined by

$$\begin{aligned}\tilde{\mu}(S) &= \det(\sum_{i\in S} d^{2\times \mathbf{1}[i\in U]} v_i v_i^\intercal) \\ &= \sum_{W\in\binom{S}{d}} d^{2|W\cap U|} \det(\sum_{i\in W} v_i v_i^\intercal).\end{aligned} \tag{1}$$

where the second equality is due to Cauchy-Binet's formula, i.e., $k \geq d$ and $|S| = k$.

$$\mu(S) = \det(\sum_{i\in S} v_i v_i^\intercal) = \sum_{W\in\binom{S}{d}} \det(\sum_{i\in W} v_i v_i^\intercal)$$

We can generalize this setting by considering $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$[3] and $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ by

$$\mu(S) = \sum_{W\in\binom{S}{\ell}} \nu(W). \tag{2}$$

We will assume that ν is strongly Rayleigh, and consequently satisfies the exchange inequality in Lemma 8. Because Lemma 8 only applies to local optima with $\nu(S) > 0$, it is more convenient to work with a full-support distribution, i.e., $\nu(S) > 0$ for all $S \in \binom{[n]}{\ell}$. Fortunately, we can approximate any strongly Rayleigh distribution ν with a full-support strongly Rayleigh distribution a.k.a. strictly real-stable distribution. In other words, for any $\epsilon > 0$, there exists a strongly Rayleigh distribution $\tilde{\nu} : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ such that for all S, $\tilde{\nu}(S) > 0$ and $|\tilde{\nu}(S) - \nu(S)| \leq \epsilon$. Moreover, $\tilde{\nu}$ can be efficiently computed given ν (see the main theorem of Nuij (1968), (Brändén and Huh, 2019, Proof of Proposition 2.2) and (Brändén, 2020, page 7)). We include a formal statement, and leave the proof to Appendix A.

Proposition 11. *Let $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ be strongly Rayleigh. For any $\epsilon > 0$, there exists strongly Rayleigh $\tilde{\nu} : \binom{[n]}{\ell} \to \mathbb{R}_{>0}$ such that $|\nu(S) - \tilde{\nu}(S)| \leq \epsilon$. for all S.*

Remark 12. We remark that a $O(1)$-approximate local optima of size ℓ w.r.t. $\det(\cdot)$ can be found in time $O(n \cdot \text{poly}(\ell))$ using a combination of simple heuristics such as greedy and local search (also known as Fedorov exchange algorithm). The same algorithmic result holds more generally for all strongly Rayleigh distributions $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ (see Anari and Vuong, 2022; Anari et al., 2020, for details).

In the remainder of the paper, we will consider the problem of maximizing the function $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ under a matroid constraint where $\mu(S) = \sum_{W\in\binom{S}{\ell}} \nu(W)$ and $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{>0}$. We assume ν is strongly Rayleigh, which implies that μ is also strongly Rayleigh. (see Appendix A for the proof).

Proposition 13. *For $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ being strongly Rayleigh and $k \geq \ell$, define $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ by*

$$\mu(S) = \sum_{W\in\binom{S}{\ell}} \nu(W)$$

then μ is also strongly Rayleigh.

Remark 14. This set-up of μ encompasses determinant maximization for both cases of $k \leq d$ and $k \geq d$. More concretely, for the former case, we set $\ell = k$, and for the latter case, we set $\ell = d$. We will explain this in more detail in Theorem 26.

For some constant $\zeta \geq 1$ to be specified later, let $\phi(W) = (\zeta\ell)^{2|W\cap U|}$. We define $\tilde{\mu} : [n]^k \to \mathbb{R}_{\geq 0}$ by:

$$\tilde{\mu}(S) = \sum_{W\in\binom{S}{\ell}} \phi(W)\nu(W)$$

This is the proper generalization of Eq. (1). We observe the following simple fact.

Fact 15.

$$\mu(S) \leq \tilde{\mu}(S) \leq (\zeta\ell)^{2\ell}\mu(S).$$

Lemma 16. *Let $V \subseteq [n]$ and let U be a ζ-approximate local optimum inside V with respect to ν, for $\zeta = O(1)$. Then for any $e \in (V \cap S) \setminus U$, there exists $f \in U$ s.t.*

$$\tilde{\mu}(S) \leq \tilde{\mu}(S - e + f).$$

See Appendix A for the proof.

3.1 The Peeling Algorithm

We have just shown how to exchange $e \in (S \cap V) \setminus U$ for $f \in U$ while keeping $\tilde{\mu}$ non-decreasing. However, we still need to ensure that $S - e + f$ is a proper set, i.e., ensure that $f \notin S - e$. To achieve this, we need a slightly more elaborate coreset construction.

Definition 17 (Peeling coreset)**.** Given $V \subseteq [n]$, and a number $k_V \geq 1$, define the (V, k_V, ζ)-peeling coreset U as follows:

[3] For now think of ℓ as being equal to d, but more generally we are introducing parameter ℓ to unify the two cases of $k \leq d$ and $d \leq k$.

- Let $U_0 = \emptyset$. For $i = 1, \cdots, k_V$, let $V_i := V \setminus \bigcup_{j=0}^{i-1} U_j$, and let $U_i \subseteq V_i$ be a ζ-approximate local optimum w.r.t. ν inside V_i.
- Let $U = \bigcup_{i=1}^{k_V} U_i$.

Note that the U_i's are disjoint and $|U| \leq k_V \ell$. We further remark that the peeling algorithm is a generic algorithm and its runtime is simply k_V times the runtime of finding an approximately local optimum w.r.t. ν.

Lemma 18. *The (V, k_V)-peeling coreset U constructed in Definition 17 is a value-preserving subset of V with respect to $\hat{\mu} : [n]^k \to \mathbb{R}_{\geq 0}$ defined as*

$$\hat{\mu}(S) = \mathbf{1}[S \in \binom{[n]}{k} \wedge |S \cap V| \leq k_V]\tilde{\mu}(S)$$

Proof. Fix $S \in \binom{[n]}{k}$ such that $|S \cap V| \leq k_V$ and $e \in (S \cap V) \setminus U$. Since S has at most $k_V - 1$ elements inside[4] $U = \bigcup_{j=1}^{k_V} U_j$, there exists some $j \in [k_V]$ such that $S \cap U_j = \emptyset$. Note that $e \in (S \cap V) \setminus U \subseteq (S \cap V_j) \setminus U_j$,. Thus, there exists $f \in U_j$ such that $\tilde{\mu}(S) \leq \tilde{\mu}(S - e + f)$. Since $S \cap U_j = \emptyset$, we are guaranteed that f is not in $S - e$. □

Lemma 19. *For $V \subseteq [n]$ and $\zeta \geq 1$, the ζ-approximate local optimum U w.r.t. μ is a value-preserving subset of V w.r.t. $\hat{\mu}$ defined by $\hat{\mu}(S) = (\zeta k)^{2|U \cap S|}\mu(S)$.*

Proof. We use the fact that μ is strongly Rayleigh, and Lemma 8. For any $S \in \binom{[n]}{k}$ and $e \in (S \cap V) \setminus U$, there exists $j \in U \setminus S$ s.t.

$$\mu(S) \leq (\zeta k)^2 \mu(S - e + j)$$

Multiplying both sides by $(\zeta k)^{2|U \cap S|}$ and using the fact that $|(S - e + j) \cap U| = |S \cap U| + 1$ we have $\hat{\mu}(S) \leq \hat{\mu}(S - e + j)$ □

Lemma 20 (Composability of value-preserving subsets)**.** *Consider datasets $V^{(1)}, \cdots, V^{(m)}$ with U_i being a value preserving subset of $V^{(i)}$ w.r.t. $\tilde{\mu}$. Then $U := \bigcup_{i=1}^m U_i$ is a value-preserving subset of $V := \bigcup_{i=1}^m V^{(i)}$ w.r.t. $\tilde{\mu}$. See Appendix A for the proof.*

4 COMPOSABLE CORESETS FOR PARTITION AND LAMINAR MATROIDS

We construct composable coresets for determinant maximization under laminar matroid constraint. To build intuition, we first describe composable coresets for the simpler case of partition matroid. The idea is to build a peeling coreset of suitable size for each part of the partition which define the parition matroid.

[4] $|S \cap U| \leq |(S \cap V) \setminus \{e\}| \leq k_V - 1$.

As in section Section 3, given a matroid $\mathcal{M}$ with the set of bases $\mathcal{B}$, we consider the problem of maximizing $\mu(S)$ (under matroid constraint) where $\mu(S) = \sum_{W \in \binom{S}{\ell}} \nu(W)$ and $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ is strongly Rayleigh. Let $\mu_{\mathcal{M}}$ be the restriction of μ to the set of bases of $\mathcal{M}$ i.e. $\mu_{\mathcal{M}}(S) = \mathbf{1}[S \in \mathcal{B}(\mathcal{M})]\mu(S)$.

Definition 21. Consider a partition matroid $\mathcal{M} = ([n], \mathcal{I})$ defined by the partition $P_1, \cdots, P_s$ of $[n]$ and $k_1, \cdots, k_s \in \mathbb{N}$. Fix constant $\zeta \geq 1$. For $V \subseteq [n]$, the composable coreset U for V w.r.t. $\mu_{\mathcal{M}}$ is constructed as follows:

- When $k > \ell$: U is the union of $(V \cap P_i, k_i, \zeta)$-peeling coresets for each $i \in [s]$, thus $|U| = k\ell$.
- When $k = \ell$: U is the union over $i \in [s]$ of the ζ-approximate local optimum w.r.t. ν in $V \cap P_i$, thus $|U| = s\ell = sk$.

Lemma 22. *The coreset constructed in Definition 21 has an approximation factor of $(\zeta\ell)^{2\ell}$. See Appendix A for the proof.*

We generalize the above to all laminar matroids.

Definition 23. Consider a laminar matroid over the ground set $[n]$ defined by a laminar family $\mathcal{F}$ and the associated integers $(k_S)_{S \in \mathcal{F}}$. Fix constant $\zeta \geq 1$. For $V \subseteq [n]$, the coreset for V is constructed as follows:

1. For each maximal set $F \in \mathcal{F}$, construct a coreset $U_F \subseteq V \cap F$ by:
   - Let $D_0 = \emptyset, V_0 = V \cap F$. For $i = 1, \cdots, k_S$, let U_i be the ζ-approximate local optimal w.r.t. ν in $V_i = V_{i-1} \setminus D_{i-1}$. For $e \in U_i$, let $F^e \in \mathcal{F}$ be the maximal *proper* subset of F containing e or $\{e\}$ if no such F^e exists. Let $D_i := \bigcup_{e \in U_i} F^e$. For each $e \in U_i$ with $F^e \neq \{e\}$, recursively construct a coreset $U_{F^e} \subseteq V \cap F^e$. Observe that if no proper subset of F is inside $\mathcal{F}$, then the coreset U_F is precisely the $(V \cap F, k_F)$ peeling-coreset.
   - The coreset for F is the union of all U_i and U_{F^e} for $e \in U_i$.
2. The coreset U of V is the union of all coresets U_F for maximal sets $F \in \mathcal{F}$ and all elements $e \in V$ that do not belong to any set $F \in \mathcal{F}$.

Theorem 24. *Consider $\nu : \binom{[n]}{\ell} \to \mathbb{R}_{\geq 0}$ that is strongly Rayleigh and $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ s.t.*

$$\mu(S) = \sum_{W \in \binom{S}{\ell}} \nu(W).$$

Consider laminar matroid constraint $\mathcal{M}$ of rank k defined by non-redundant family $\mathcal{F}$ with cover number r i.e.

$$r := \max_{e \in [n]} |\{F \in \mathcal{F} : e \in F\}|.$$

Definition 23 gives a $(\zeta\ell)^{2\ell}$-composable coreset w.r.t μ under matroid constraint $\mathcal{M}$ of size at most $(\zeta k\ell)^r \leq (\zeta k\ell)^k$.

Proof of the approximation factor. Let

$$\tilde{\mu}(S) = \sum_{W \in \binom{S}{\ell}} (\zeta\ell)^{2|W\cap U|}\nu(W).$$

U is a value-preserving subset of V w.r.t the restriction $\tilde{\mu}_{\mathcal{M}}$ of $\tilde{\mu}$ to the set of bases of the laminar matroid i.e. $\tilde{\mu}_{\mathcal{M}}(S) = \mathbf{1}[S \in \mathcal{B}(\mathcal{M})]\tilde{\mu}(S)$. This combined with Lemma 10 immediately imply that U is a $\ell^{2\ell}$-composable coreset w.r.t $\mu_{\mathcal{M}}$.

We only need to show that U_F is value preserving for each $F \in \mathcal{F}$. Fix $S \in \mathcal{B}$ and $h \in (S \cap V \cap F) \setminus U_F$. We claim that there exists $f \in U_F$ s.t. $S - h + f \in B$ and $\hat{\mu}(S) \leq \hat{\mu}(S - h + f)$.

We prove this by induction on F. For the base case when F has no proper subset inside $\mathcal{F}$, then U_F is the $(V \cap F, k_F)$ peeling-coreset, and the claim follows from Lemma 18. If $h \in D_i$ for some i, then h must be contained in a proper subset $F^e \in \mathcal{F}$ of F where $e \in U_i$ and $F^e \neq \{e\}$[5] and we can use the induction hypothesis. Now, assume $h \notin D_i$ for $\forall i \in [k_F]$. In particular, this means $h \in V_{k_F} \subseteq \cdots \subseteq V_1 = V \cap F$ and D_i and U_i are non-empty for all $i \in [k_F]$. Note that since D_i's are disjoint, and S contains at most $k_F - 1$ elements inside F, $S \cap D_i = \emptyset$ for some i. In particular, $S \cap U_i = \emptyset$ and $h \in (V_i \cap S) \setminus U_i$, so Lemma 16 implies that there exists $f \in U_i$ s.t. $\hat{\mu}(S) \leq \hat{\mu}(S - h + f)$. Replacing h with f only affects the constraints for sets $F' \in \mathcal{F}$ containing f. Consider such a set F'. F' must be contained inside D_i by the definition of D_i, thus $S \cap F' = \emptyset$, and $|(S - h + f) \cap F'| \leq 1 \leq k_{F'}$. We just verify that $S - h + f$ is also a base of the laminar matroid, thus

$$\hat{\mu}(S - h + f) = \tilde{\mu}(S - h + f) \geq \tilde{\mu}(S) = \hat{\mu}(S).$$

□

Proof of upper bound on the size of the coreset. For a set H let $r_H := \max_{e\in H}|\{F \in \mathcal{F} : F \subseteq H \wedge e \in F\}|$. We show that $|U_F| \leq (k_F\ell)^{r_F}$ for each $F \in \mathcal{F}$ by induction on r_F. For the base case $r_F = 1$, we have $|U_F| = k_F\ell$, by Definition 17. Fix $F \in \mathcal{F}$ with $r_F \geq 2$ and suppose the induction hypothesis holds for $r < r_F$. Using the definition of U_F, we can bound

$$\begin{aligned}|U_F| &\leq \sum_{i=1}^{k_F}|U_i| + \sum_{e\in\bigcup_{i=1}^{k_F} U_i} |U_{F^e}| \\ &\leq_{(1)} k_F d + (k_F\ell)(d \max_{F'\subseteq F: F'\in\mathcal{F}} k_{F'})^{r_F-1} \\ &\leq_{(2)} (k_F\ell)^{r_F}\end{aligned}$$

where in (1) we use the fact that $r_{F^e} < r_F$ since F^e is a proper subset of F, and in (2) we use

$$(\max_{F'\subseteq F: F'\in\mathcal{F}} k_{F'})^{r_F-1} + 1 \leq (k_F - 1)^{r_F-1} + 1 \leq k_F^{r_F-1}.$$

Thus the induction hypothesis holds for all r.

Suppose the maximal set(s) in $\mathcal{F}$ are $F_1, \cdots, F_t$, and let $R := [n] \setminus \bigcup_{i=1}^t F_t$. Then the rank of the laminar matroid is $k = |R| + \sum_{i=1}^t k_{F_t}$, and

$$|U| = |R \cup \sum_{i=1}^t U_{F_t}| \leq |R| + \sum_{i=1}^t (k_{F_t}\ell)^r \leq (k\ell)^r.$$

□

Remark 25. For any laminar family $\mathcal{F}$ of rank k, we can construct a $d^{O(d)}$-coreset of size $|\mathcal{F}|dk$ by taking the union of all value-preserving subsets of $V \cap (F \setminus \bigcup_{F'\subseteq F, F'\in\mathcal{F}} F')$. However, the size of the coreset might be as bad as linear in n. Indeed, consider the laminar family defined by: $F_i = \{2i+1, 2i+2\}, k_{F_i} = 1$ for $\forall i \in [n/2]$ and $F_0 = [n]$, $k_{F_0} = k$ then Definition 23 gives a coreset of size[6] $\leq k^2d^2$ whereas the naive construction gives a coreset of size $\geq (n/2)d$.

We immediately obtain the following corollary about determinant maximization under matroid constraints.

Theorem 26. *For the determinant maximization matroid constraints with input vectors $v_1, \cdots, v_n \in \mathbb{R}^d$, we obtain the following results:*

1. *Partition matroid defined by partition $P_1, \cdots, P_s$ of $[n]$, Definition 21 gives:*
   - *For $k \leq d$: k^{2k}-composable coreset of size $O(sk)$.*
   - *For $k \geq d$: d^{2d}-composable coreset of size $O(kd)$.*
2. *Laminar matroid:*
   - *For $k \leq d$: k^{2k}-composable coreset of size $O(k^{2k})$.*
   - *For $k \geq d$: d^{2d}-composable coreset of size $O((kd)^k)$.*

See Appendix A for the proof.

Recall that $O(1)$-approximate local optima can be found in time $O(n\text{poly}(k))$ (see Remark 12). Thus, our coreset construction is highly efficient: it takes time $O(n\text{poly}(k))$ for the case of partition matroid constraint. As a corollary, we obtain a quasilinear algorithm for MAP-inference for DPP under partition matroid constraint.

Lemma 27. *Consider a partition matroid $\mathcal{M} = ([n], \mathcal{I})$ of rank k defined by the partition $P_1, \cdots, P_s$ of $[n]$ and $k_1, \cdots, k_s \in \mathbb{N}$. Given input vectors $v_1, \cdots, v_n \in \mathbb{R}^d$, there exists a $O(n\text{poly}(k))$ algorithm that outputs a $\min(k^{O(k)}, d^{O(d)})$-approximation for the determinant maximization under partition matroid constraint $\mathcal{M}$. See Appendix A for the proof.*

[5] if $F^e = \{e\}$ then $h = e \in U_i \subseteq U_F$, a contradiction

[6] We can improve the bound to kd^2 by a more careful analysis.

References

Abbar, S., Amer-Yahia, S., Indyk, P., Mahabadi, S., and Varadarajan, K. R. (2013). Diverse near neighbor problem. In *Proceedings of the twenty-ninth annual symposium on Computational geometry*, pages 207–214.

Abbassi, Z., Mirrokni, V. S., and Thakur, M. (2013). Diversity maximization under matroid constraints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 32–40.

Addanki, R., McGregor, A., Meliou, A., and Moumoulidou, Z. (2022). Improved approximation and scalability for fair max-min diversification. *arXiv preprint arXiv:2201.06678*.

Agarwal, P. K., Har-Peled, S., Varadarajan, K. R., et al. (2005). Geometric approximation via coresets. *Combinatorial and computational geometry*, 52(1).

Agarwal, P. K., Har-Peled, S., and Yu, H. (2008). Robust shape fitting via peeling and grating coresets. *Discrete & Computational Geometry*, 39(1):38–58.

Allen-Zhu, Z., Li, Y., and Wang, Y. (2017). Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, 186.

Anari, N., Gharan, S. O., and Rezaei, A. (2016). Monte carlo markov chain algorithms for sampling strongly rayleigh distributions and determinantal point processes. In *Conference on Learning Theory*, pages 103–115. PMLR.

Anari, N., Liu, K., Gharan, S. O., and Vinzant, C. (2020). Log-concave polynomials IV: exchange properties, tight mixing times, and faster sampling of spanning trees. *CoRR*, abs/2004.07220.

Anari, N. and Vuong, T.-D. (2022). From sampling to optimization on discrete domains with applications to determinant maximization. In *Conference on Learning Theory*, pages 5596–5618. PMLR.

Barbosa, R., Ene, A., Nguyen, H., and Ward, J. (2015). The power of randomization: Distributed submodular maximization on massive datasets. In *International Conference on Machine Learning*, pages 1236–1244.

Borcea, J., Brändén, P., and Liggett, T. (2009). Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567.

Borcea, J. and Brändé n, P. (2009). The lee-yang and pólya-schur programs. II. theory of stable polynomials and applications. *Communications on Pure and Applied Mathematics*, 62(12):1595–1631.

Brown, A., Laddha, A., Pittu, M., and Singh, M. (2022a). Efficient determinant maximization for all matroids.

Brown, A., Laddha, A., Pittu, M., Singh, M., and Tetali, P. (2022b). Determinant maximization via matroid intersection algorithms. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 255–266. IEEE.

Brändén, P. (2020). Spaces of lorentzian and real stable polynomials are euclidean balls.

Brändén, P. and Huh, J. (2019). Lorentzian polynomials.

Ceccarello, M., Pietracaprina, A., and Pucci, G. (2018). Fast Coreset-based Diversity Maximization under Matroid Constraints. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 81–89.

Ceccarello, M., Pietracaprina, A., and Pucci, G. (2020). A General Coreset-Based Approach to Diversity Maximization under Matroid Constraints. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(5):1–27.

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

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

Gong, B., Chao, W.-L., Grauman, K., and Sha, F. (2014). Diverse sequential subset selection for supervised video summarization. In *Advances in Neural Information Processing Systems*, pages 2069–2077.

Indyk, P., Mahabadi, S., Gharan, S. O., and Rezaei, A. (2020). Composable core-sets for determinant maximization problems via spectral spanners. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1675–1694. SIAM.

Indyk, P., Mahabadi, S., Mahdian, M., and Mirrokni, V. S. (2014). Composable core-sets for diversity and coverage maximization. In *Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 100–108. ACM.

Kulesza, A. and Taskar, B. (2011). Learning determinantal point processes.

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

Lau, L. C. and Zhou, H. (2021). A local search framework for experimental design. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1039–1058. SIAM.

Lee, D., Cha, G., Yang, M.-H., and Oh, S. (2016). Individualness and determinantal point processes for pedestrian detection. In *European Conference on Computer Vision*, pages 330–346. Springer.

Madan, V., Nikolov, A., Singh, M., and Tantipongpipat, U. T. (2020). Maximizing determinants under matroid constraints. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 565–576.

Madan, V., Singh, M., Tantipongpipat, U., and Xie, W. (2019). Combinatorial algorithms for optimal design. In *Conference on Learning Theory*, pages 2210–2258. PMLR.

Mahabadi, S., Indyk, P., Gharan, S. O., and Rezaei, A. (2019). Composable core-sets for determinant maximization: A simple near-optimal algorithm. In *International Conference on Machine Learning*, pages 4254–4263. PMLR.

Mahabadi, S. and Narayanan, S. (2023). Improved diversity maximization algorithms for matching and pseudoforest. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*.

Mahabadi, S. and Trajanovski, S. (2023). Core-sets for fair and diverse data summarization. *Advances in Neural Information Processing Systems*, 36:78987–79011.

Mirrokni, V. and Zadimoghaddam, M. (2015). Randomized composable core-sets for distributed submodular maximization. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 153–162. ACM.

Mirzasoleiman, B., Jegelka, S., and Krause, A. (2017). Streaming non-monotone submodular maximization: Personalized video summarization on the fly. *arXiv preprint arXiv:1706.03583*.

Mirzasoleiman, B., Karbasi, A., Badanidiyuru, A., and Krause, A. (2015). Distributed submodular cover: Succinctly summarizing massive data. In *Advances in Neural Information Processing Systems*, pages 2881–2889.

Mirzasoleiman, B., Karbasi, A., Sarkar, R., and Krause, A. (2013). Distributed submodular maximization: Identifying representative elements in massive data. In *Advances in Neural Information Processing Systems*, pages 2049–2057.

Moumoulidou, Z., Mc Gregor, A., and Meliou, A. (2020). Diverse data selection under fairness constraints. *arXiv preprint arXiv:2010.09141*.

Nikolov, A. (2015). Randomized rounding for the largest simplex problem. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 861–870.

Nikolov, A. and Singh, M. (2016). Maximizing determinants under partition constraints. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 192–201.

Nuij, W. (1968). A note on hyperbolic polynomials. *Mathematica Scandinavica*, 23:69–72.

Pan, X., Jegelka, S., Gonzalez, J. E., Bradley, J. K., and Jordan, M. I. (2014). Parallel double greedy submodular maximization. In *Advances in Neural Information Processing Systems*, pages 118–126.

Wei, K., Iyer, R., and Bilmes, J. (2014). Fast multi-stage submodular maximization. In *International conference on machine learning*, pages 1494–1502.

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

Zadeh, S., Ghadiri, M., Mirrokni, V., and Zadimoghaddam, M. (2017). Scalable feature selection via distributed diversity maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes, as mentioned in the paper our coreset construction algorithms are generic algorithms whose runtime depends on the runtime of the algorithm one uses for finding an approximately local optimum solution.]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable, this is a theoretical paper.]
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes, if not in the main body, they are included in the appendix.]
   (c) Clear explanations of any assumptions. [Yes]
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable, this is a theoretical paper.]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]

(c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Not Applicable]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Composable Coresets for Constrained Determinant Maximization and Beyond: Supplementary Materials

A OMITTED PROOFS

Proof of Proposition 11. By Nuij (1968), for $i, j \in [n]$ and $s \in \mathbb{R}_{\geq 0}$, the following operator preserves strongly Rayleigh/real-stability of polynomials

$$T_{i,j,s} g = g + s z_j \frac{\partial g}{\partial i}$$

for $g \in \mathbb{R}[z_1, \ldots, z_n]$.

If $\nu(S) = 0$, for $\forall S$, then we can let $\tilde{\nu}(S) = \epsilon$ for $\forall S$. W.l.o.g. assume $\nu(S) \neq 0$ for $S = \{1, \ldots, \ell\}$. Let g be the generating polynomial for ν and let

$$f = \prod_{i \in [n], j \in [n]} T_{i,j,s} g = T_{1,1,s} \circ \cdots \circ T_{1,n}$$
$$\circ\, T_{2,1,s} \circ \ldots T_{2,n,s} \circ \cdots \circ T_{n,n,s} g$$

then f is strongly Rayleigh. It is easy to see that for small enough s, f approximates g. One practical choice for s is $s = \epsilon(\sum_S \nu(S))^{-c}$ for some $c > 1$; computing the partition function $\sum_S \nu(S)$ can be done efficiently and even in $\tilde{O}(1)$-parallel time for distributions of interest e.g. DPP. The map from f to the multi-affine part f^{MAP} of f preserves real stability Borcea et al. (2009) (recall that for $f(z_1, \ldots, z_n) = \sum_{(\alpha_i)_{i=1}^n \in \mathbb{N}^n} c(\vec{\alpha}) \prod_{i=1}^n z_i^{\alpha_i}$, the multiaffine part of f is

$$f^{MAP}(z_1, \ldots, z_n) = \sum_{(\alpha_i)_{i=1}^n \in 0,1^n} c(\vec{\alpha}) \prod_{i=1}^n z_i^{\alpha_i}.$$

Now we only need to check that f^{MAP} has positive coefficients. Indeed, for $S = \{i_1, \ldots, i_\ell\}$ consider the coefficient of the monomial $z^S = \prod_{i \in S} z_i$ in f and f^{MAP} : it is a sum which includes the term

$$\prod_{j=1}^{\ell} (s z_{i_j} \frac{\partial}{\partial z_j}) \nu([\ell]) z^{[\ell]} = z^S s^k \nu([\ell]) > 0$$

thus the coefficient of z^S in f^{MAP} is positive. □

Proof of Proposition 13. Consider the elementary symmetric polynomial

$$e_{k-\ell}(z_1, \cdots, z_n) = \sum_{L \in \binom{[n]}{k-\ell}} \prod_{i \in L} z_i$$

then $e_{k-\ell}$ is real stable i.e. has no roots in the upper half plane. Since the same is true for g_ν, the product $g_\nu e_{k-\ell}$ also has no root on the upper half plane (see e.g. Borcea et al., 2009, Proposition 3.1 for a proof). Consider the linear map $\phi : \mathbb{R}[z_1, \cdots, z_n] \to \mathbb{R}[z_1, \cdots, z_n]$ that maps monomial $z_1^{\alpha_1} \cdots z_n^{\alpha}$ to itself if $\alpha_i \leq 1$ for $\forall i$, and to 0 otherwise. This map preserves real-stability of polynomial Borcea and Brändé n (2009), and $\phi(g_\nu e_{k-\ell}) = g_\mu$, thus g_μ is real stable, and μ is strongly Rayleigh. □

Proof of Lemma 16. Consider $W \in \binom{S}{\ell}$ with $e \in W$. Using Lemma 8 with $\ell' = \zeta \ell$, we have

$$\nu(W) \leq \ell' \sum_{f \in U} \nu(W - e + f)$$

Summing over all such W, we get

$$\begin{aligned}
&\sum_{W\in\binom{S}{\ell}:e\in W}\phi(W)\nu(W)\\
\le&\ell'\sum_{f\in U}\sum_{W:e\in W}\phi(W)\nu(W-e+f)\\
\le&(\zeta\ell)^2\max_{f\in U}\sum_{W\in\binom{S}{\ell}:e\in W}\phi(W)\nu(W-e+f)\\
=&\sum_{W\in\binom{S}{\ell}:e\in W}\phi(W-e+f^*)\nu(W-e+f^*)
\end{aligned}$$

with f^* being the maximizer of the second line. Finally, adding $\sum_{W\in\binom{S-e}{\ell}}\phi(W)\nu(W)$ to both sides gives the desired inequality. □

Proof of Lemma 20. Consider $S\subseteq[n]$ and $e\in(S\cap V)\setminus U$. Clearly, $e\in(S\cap V_i)\setminus U_i$ for some $i\in[m]$. Since U_i is value-preserving w.r.t. $\tilde{\mu}$, there exists $j\in U_i\subseteq U$ s.t. $\tilde{\mu}(S)\le\tilde{\mu}(S-e+f)$. □

Proof of Lemma 22. Note that in both cases, by Lemmas 18 and 19, U is the union of value-preserving subsets U_i of $V\cap P_i$ w.r.t $\tilde{\mu}_{\mathcal{M}}(S)=\mathbf{1}[S\in\mathcal{B}]\sum_{W\in\binom{S}{\ell}}(\zeta\ell)^{2|W\cap U|}\nu(W)$. Thus, by Lemma 20, U is a value-preserving subset of V w.r.t $\tilde{\mu}_{\mathcal{M}}$. Fact 15 and Lemma 10 together imply that U is $\ell^{2\ell}$-composable coreset w.r.t. $\mu_{\mathcal{M}}$. □

Proof of Theorem 26. We show how to adapt the setting of $\nu:\binom{[n]}{\ell}\to\mathbb{R}_{\ge0}$ and $\mu:\binom{[n]}{k}\to\mathbb{R}_{\ge0}$ where $\mu(S)=\sum_{W\in\binom{S}{\ell}}\nu(W)$ with ν being strongly Rayleigh to the determinant maximization setting.

- For $k\le d$: we let $\ell=k$ and $\mu(S)=\nu(S)=\det(\sum_{i\in S}v_iv_i^\intercal)$ for $|S|=k$. By replacing ℓ with k we get the stated result.
- For $k\ge d$: we let $\ell=d$, $\nu(W)=\det(\sum_{i\in W}v_iv_i^\intercal)$ for $|W|=\ell$ and $\mu(S)=\nu(S)=\det(\sum_{i\in S}v_iv_i^\intercal)$ for $|S|=k$.

□

Proof of Lemma 27. W.l.o.g. we can assume $k_i\ge1$ for all i. We construct coreset U as in Theorem 26. Note that since $k_1+\cdots+k_s=k$ and $k_i\ge1$, we have that $s\le k$ thus the size of U is $O(k^2)$ for both cases $k\le d$ and $k\ge d$. We can restrict the ground set to U and use the existing efficient algorithms Brown et al. (2022b) to get a $\min\{k^{O(k)},d^{O(d)}\}$-approximation for constrained determinant maximization with input vectors from U, which is also a $\min\{k^{O(k)},d^{O(d)}\}$-approximation for the original constrained determinant maximization problem. □

B LOWER BOUNDS

In this section, we show that the coreset we constructed essentially attains the best possible size and approximation factor. We first show that for determinant maximization in $\mathbb{R}^d$ when $k\le d$ under partition matroid constraint, our coreset size is optimal.

Lemma 28. *Suppose $k\le d$. Consider a partition matroid $\mathcal{M}=([n],\mathcal{I})$ defined by a partition $P_1\cup\cdots\cup P_s=[n]$ and constraint $k_1,\cdots,k_s$. Let $k:=\mathrm{rank}(\mathcal{M})=\sum_{i=1}^s k_i$. Any α-composable coreset for the determinant maximization problem under partition matroid constraint $\mathcal{M}$ with size $t<sk$ must incur an arbitrarily large approximation factor.*

Proof. Consider n vectors $v_1,\cdots,v_n\in\mathbb{R}^d$ to be chosen later. For set $Q\subseteq[n]$, let

$$\mathrm{OPT}(Q):=\max_{S\subseteq Q,|S|=k}\det(\sum_{i\in S}v_iv_i^\intercal).$$

Consider a partitioning of $[n]$ into two sets Q,Q' such that for each i, $\{v_j:j\in Q\cap P_i\}=\{e_1,\cdots,e_d\}$ where $e_1,\cdots,e_d$ is the standard basis for $\mathbb{R}^d$. We need to show that for any subset $U\subseteq S$ of size $t<sk$, we can choose the vectors in Q' s.t.

$\text{OPT}(Q \cup Q') \gg \text{OPT}(U \cup Q')$. Indeed, fix one such subset U where $|U| \leq sk-1$. Since $\sum_{i\in[s]} |U \cap P_i| \leq |U| \leq sk-1$, there must exist $i \in [s]$ s.t. $|U \cap P_i| \leq k-1$. W.l.o.g., we can assume that $\{v_j : j \in U \cap P_1\} \subseteq \{e_1, \cdots, e_{k-1}\}$. Choose Q' s.t. $Q' \cap P_1 = \emptyset$, and

$$\{v_j : j \in Q' \cap P_i\} = \{Me_{\sum_{j=1}^{i-1} k_j}, \cdots, Me_{\sum_{j=1}^{i} k_j - 1}\}$$

for some arbitrarily large $M > 0$. Consider $S \subseteq Q \cup Q'$ s.t. $\{v_j : j \in S \cap P_1\} = \{e_1, \cdots, e_{k_1-1}, e_k\}$ and $\{v_j : j \in S \cap P_i\} = Q' \cap P_i$ for $i = 2, \cdots, s$, then $S \in \mathcal{M}$ and $\mu(S) := \det(\sum_{i\in S} v_i v_i^\intercal) = M^{2\sum_{j=2}^{s} k_j}$, thus $\text{OPT}(Q \cup Q') \geq M^{2\sum_{j=2}^{s} k_j}$. On the other hand, for any $S' \in \binom{U\cup Q'}{k}$, either:

- $S' \cap P_i \subseteq Q' \forall i \geq 2$: in this case $\mu(S') = 0$[7] because all the vectors in S' are contained in the $(k-1)$-dimensional subspace spanned by $e_1, \cdots, e_{k-1}$.
- $S' \cap P_i \not\subseteq Q'$ for some $i \geq 2$: in this case $\mu(S') \leq M^{2(\sum_{j=2}^{s} k_j - 1)}$ since there are at most $\sum_{j=2}^{s} k_j - 1$ vectors in S' that are from V' and thus have norm M, while the remaining vectors are from V and have norm 1.

In either case, we have $\text{OPT}(U \cup Q') \leq M^{2(\sum_{j=2}^{s} k_j - 1)} \leq \text{OPT}(Q \cup Q')/M^2$, thus $\text{OPT}(Q \cup Q')$ can be arbitrarily large compared to $\text{OPT}(U \cup Q')$. □

For $k \geq d$, using similar arguments, we can show that any α-coreset for determinant maximization under partition constraint with finite approximation factor α must have size $t \geq k + d(d-1)$.

Lemma 29. *Suppose $k \geq d$. Consider the partition matroid $\mathcal{M} = ([n], \mathcal{I})$ of rank k defined by a partition $P_1, \cdots, P_k$ and constraint $k_1 = \cdots = k_k = 1$. Any composable coreset for the determinant maximization problem under partition matroid constraint $\mathcal{M}$ with size $t < k + d(d-1)$ must incur an arbitrarily large approximation factor.*

Proof. The construction is similar to the proof of Lemma 28. Consider n vectors $v_1, \cdots, v_n \in \mathbb{R}^d$ to be chosen later. For set $Q \subseteq [n]$, let $\text{OPT}(Q) := \max_{S\subseteq Q, |S|=k} \det(\sum_{i\in S} v_i v_i^\intercal)$.

Choose set $Q \subseteq [n]$ and $\{v_j : j \in Q\}$ s.t.

$$\{v_j : j \in Q \cap P_i\} = \{M_i e_1, \cdots, M_i e_d\}$$

with $M_1 \geq M_2 \geq M_d \gg M_{d+1} \cdots \geq M_k$ to be chosen later. Let U be a coreset for Q with finite approximation factor. Clearly, $|Q \cap P_i| \geq 1$. We will show that $|U \cap P_i| = d$ for $i = 1, \cdots, d$, and thus conclude that $|U| \geq (k-d) + d^2 = k + d(d-1)$.

For the base case of $i = 0$, the claim holds trivially. Suppose that the claim holds for $i-1$ with $i \geq 1$. Then we show that it holds for i. We assume for contradiction that $|U \cap P_i| \leq d-1$. W.l.o.g., assume $\{v_j : j \in U \cap P_i\} \subseteq M_i e_1, \cdots, M_i e_{i-1}, M_i e_{i+1}, \cdots, M_i e_d$. Indeed, define Q' such that $\{v_j : j \in Q' \cap P_t\} = \{Me_t\}$ for $t \in \{1, \cdots, i-1, i+1, \cdots, d\}$ and $Q' \cap P_t = \emptyset$ otherwise. By choosing $M \gg M_1$ and $M_i \gg M_d \gg M_{d+1}$, we can ensure that the optimal instance in $Q \cup Q'$ must contain $d-1$ vectors in Q' and $M_i e_i \in Q \cap P_i$, thus

$$\text{OPT}(Q \cup Q') \geq (M^{d-1} M_i)^2.$$

On the other hand, for any $S \subseteq U \cup Q'$, either

- $|S \cap Q'| \leq d-1$: in this case
$$\mu(S) \leq \binom{k}{d} (M^{d-2} M_1^2)^2$$
since any $W \in \binom{S}{d}$ must contain at most $d-2$ vectors of norm M from V', and the remaining vectors have norm at most M_1.
- $|S \cap Q'| = d-1$: since $\{v_j : U \cap P_i\}$ is in the span of $\{v_j : S \cap Q'\}$, any $W \in \binom{S}{d}$ with $\det(\sum_{i\in W} v_i v_i^\intercal) \neq 0$ must satisfy that $V_W := \{v_j : j \in W\}$ consist of at most $d-1$ vectors of norm M from Q', and the remaining vectors must have norm at most M_{d+1}, thus
$$\mu(S) \leq \binom{k}{d} (M^{d-1} M_{d+1})^2.$$

[7] Even when we replace μ by a full-support $\tilde{\mu}$ that approximates μ within distance ϵ, we will have $\tilde{\mu}(S) < \epsilon < \text{OPT}(V \cup V')/M^2$ if we choose ϵ small enough

In either case, $\text{OPT}(U \cup Q')$ can be arbitrarily smaller than $\text{OPT}(Q \cup Q')$. □

Finally, we show that for $k \geq d$, the approximation factor of $d^{O(d)}$ is the best possible even under no constraints. For $k \leq d$, Indyk et al. (2020) shows that approximation factor of $k^{O(k)}$ is optimal.

The following construction is from (Indyk et al., 2020, section 7.1). We include this for completeness.

Definition 30 (Hard input for composable coreset)**.** Let $\beta = o(d/\log^2 d)$, $m = d/\log d$ so that $d^{d/m} = O(1)$. Consider $G \subseteq \mathbb{R}^{m+1}$ of $d^{\beta+2}$ vectors s.t. for every two vectors $p, q \in G$, we have $\langle p, q\rangle \leq O(\frac{\sqrt{\beta}\log d}{\sqrt{d}})$.

For $i = 1, \cdots, d-m$, construct X_i as follows: pick a random index $\pi_i \in [n]$. Embed G into the subspace spanned by $\{e_1, \cdots, e_m, e_{m+i}\}$ s.t. the $\pi(i)^{th}$ vectors in G is mapped into e_{m+i}.

Choose a random rotation matrix Q, and return $QX_1, \cdots, QX_{d-m}$ and $QY_1, \cdots, QY_m$ with $Y_i = \{Me_i\}$ for a large enough M.

Theorem 31. *For $d \leq k \leq d^\gamma$ and γ' s.t. $\gamma\gamma' = o(d/\log^2 d)$, any composable coreset of size $k^{\gamma'}$ must incur an approximation of $(\frac{d}{\gamma\gamma'})^{d(1-o(1))}$. For example, the theorem applies when γ, γ' are constant, i.e. $d \leq k \leq$ poly(d) and the coreset has size poly(k).*

Proof. For a set V of vectors, let $V^{\times t}$ be the set where each vector in V is duplicated t times. Let $\beta = \gamma\gamma'$. We use the construction in Definition 30 where every vector is duplicated $t = k/d$ times. Let $QX_1^{\times t}, \cdots, QX_{d-m}^{\times t}, QY_1^{\times t}, \cdots, QY_m^{\times t}$ be the input sets. Let S be s.t.

$$V_S := \{v_j : j \in S\} = \{Me_1, \cdots, Me_m, e_{m+1}, \cdots, e_d\}$$

then $V_S^{\times t}$ has value $\mu(S) \geq (k/d)^d (M^m)^2$.

On the other hand, let $c(QX_i^{\times t})$ be an arbitrary coreset of size $k^{\gamma'} \leq d^\beta$ for $QX_i^{\times t}$.

As observed in (Indyk et al., 2020, Lemma 7.2), the probability that $C_i := c(QX_i^{\times t})$ contains Qe_{m+i} is bounded by

$$|c(QX_i^{\times t})|/|QX_i^{\times t})| \leq 1/d^2.$$

Thus, with probability $\geq 1 - 1/d$, we have $Qe_{m+i} \notin c(QX_i^{\times t})$ for all $i \in [d-m]$. Assume that this happens. Then for any $u \in C_i$

$$\langle \sum_{i=m+1}^{d} e_i e_i^\intercal, uu^\intercal \rangle \leq O(\frac{\beta \log^2 d}{d})$$

thus for any $u_1, \cdots, u_m$ in

$$\mathcal{C} := \bigcup_{i=1}^{m} c(QY_i^{\times t}) \cup \bigcup_{i=1}^{d-m} c(QX_i^{\times t}),$$

$$\begin{aligned}
&\det(\sum_{i=m+1}^{d} (Me_i)(Me_i)^\intercal + \sum_{i=1}^{m} u_i u_i^\intercal) \\
\leq & M^{2m} (\max \langle \sum_{i=m+1}^{d} e_i e_i^\intercal, uu^\intercal \rangle)^{d-m} \\
\leq & M^{2m} \left(\frac{O(\sqrt{\beta})\log d}{d}\right)^{2(d-m)}.
\end{aligned}$$

Hence, with probability at least $1 - 1/d$, any size-d subset W in $\mathcal{C}$ has

$$\det(\sum_{v \in W} vv^\intercal) \leq M^{2m} \left(\frac{O(\sqrt{\beta})\log d}{d}\right)^{2(d-m)},$$

thus by Cauchy Binet, any size-k subset S in $\mathcal{C}$ has

$$\mu(S) \leq \binom{k}{d} M^{2m} \left(\frac{O(\sqrt{\beta}) \log d}{d} \right)^{2(d-m)}.$$

Thus the approximation factor is at least

$$\frac{1}{e^d} (d/(O(\sqrt{\beta}) \log d)^2)^{d-m}$$

with $m = o(d)$.

□

C OTHER EXPERIMENTAL DESIGN PROBLEMS

In this section, we generalize our composable coreset construction to other experimental design problems such as A-design and E-design.

The main idea is to replace the local optimum in the coreset construction with an α-spectral spanner (see Definition 32). By replacing the local optimum with a spectral spanner, we can ensure that the coreset contains a high-valued feasible fractional x in the convex hull $P(\mathcal{M}) \subseteq [0,1]^n$ of the matroid polytope of $\mathcal{M}$, which can be rounded to an integral solution for uniform matroid constraint and certain class of laminar matroid constraint.

Definition 32 (Indyk et al. (2020))**.** For a set of vectors $V \subseteq \mathbb{R}^d$, a subset $U \subseteq V$ is a α-spectral spanner of V iff for any $v \in V$, there exists a distribution μ_v of vectors in U s.t.

$$vv^\intercal \preceq \alpha \mathbb{E}_{u \sim \mu_v} uu^\intercal$$

Theorem 33 ((Indyk et al., 2020, Proposition 4.2,Lemma 4.6))**.** *Given $V \subseteq \mathbb{R}^d$, there exists an efficient algorithm that constructs $\tilde{O}(d)$-spectral spanner of size $\tilde{O}(d)$.*

Recall that the goal of experimental design problem is to select a set S[8] that maximizes $f(\sum_{xxi \in S} v_i v_i^\intercal)$ for some objective function f. The most popular and well-studied objective functions include:

- D(eterminant)-design: $f(A) = \det(A)^{1/d}$.
- A(verage)-design: $f(A) = -\text{Tr}(A^{-1})/d$
- E(igen)-design: $f(A) = -\|A^{-1}\|_2$
- T(race)-design: $f(A) = d/\text{Tr}(A)$

Each of the above objective functions satisfies the properties of a *regular* function (see Definition 34). Allen-Zhu et al. (2017) shows that under uniform matroid i.e. cardinality constraint, any fractional feasible solution of a regular function can be rounded into an integral solution while incurring only $O(1)$ loss in the objective function. For laminar matroids, Lau and Zhou (2021) shows the same results for D-design and A-design when $k_F \geq Cd$ for $\forall F \in \mathcal{F}$ and for some absolute constant C. For general matroid and $f(\cdot) = \det(\cdot)$, Madan et al. (2020) shows that a fractional feasible solution can be rounded into an integral solution while suffering a $d^{O(d)}$ loss *in expectation*.

Definition 34. A function $f : \mathbb{S}_d^+ \to \mathbb{R}$ is regular if it satisfies the following properties

- Monotonicity: for any $A, B \in \mathbb{S}_d^+$, if $f(A) \leq f(B)$ for $A \preceq B$.
- Concavity: for $A, B \in \mathbb{S}_d^+$ and $t \in [0,1]$, we have $f(tA + (1-t)B) \geq tf(A) + (1-t)f(B)$.

  In particular, this implies the existence of an efficient algorithm that solves the continuous relaxation

$$\max_{s_1, \cdots, s_n} f(\sum_{i \in 1}^{n} s_i v_i v_i^\intercal) \text{ s.t. } s_i \in [0,1] \text{ and } \sum_{i=1}^{n} s_i \leq k.$$

[8] S might need to satisfy additional constraints such as S is a basis of a given matroid

- Reciprocal linearity: for any $A \in \mathbb{S}_d^+$ and $t \in (0,1)$, $f(tA) = t^{-1}f(A)$.

Theorem 35 (Rounding for experimental design, Madan et al. (2020))**.** *Consider the experimental design problem with objective function $f(\cdot)$ and input vectors $v_1, \cdots, v_n \in \mathbb{R}^d$ under matroid constraint $\mathcal{M}$ of rank k. For any fractional $x \in P(\mathcal{M}) \subseteq [0,1]^n$, there exists $z \in \mathcal{B}(\mathcal{M}) \subseteq \{0,1\}^n$ s.t.*

- *When $f(A) = \det(A)$:*

$$\min\{d^{O(d)}, 2^{O(k)}\} f(\sum_{i=1}^n z_i v_i v_i^\intercal) \geq f(\sum_{i=1}^n x_i v_i v_i^\intercal)$$

  The factor $d^{O(d)}$ can be improved to $2^{O(d)}$ when $\mathcal{M}$ is a partition matroid.

- *When $k \geq d$, $\mathcal{M}$ is the uniform matroid and f is regular:*

$$O(1) f(\sum_{i=1}^n z_i v_i v_i^\intercal) \geq f(\sum_{i=1}^n x_i v_i v_i^\intercal)$$

- *When $k \geq d$, $\mathcal{M}$ is a laminar matroid defined by the laminar family $\mathcal{F}$ and $(k_F)_{F \in \mathcal{F}}$ with $k_F \geq Cd$ for $\forall F \in \mathcal{F}$ for some large absolute constant C, and $f(A) = -Tr(A^{-1})/d$:*

$$O(1) f(\sum_{i=1}^n z_i v_i v_i^\intercal) \geq f(\sum_{i=1}^n x_i v_i v_i^\intercal)$$

We show $\tilde{O}(d)$-composable coreset of size $\tilde{O}(dk)$ for experimental design problems in the without repetition setting.

Theorem 36. *Given input vectors $v_1, \cdots, v_n \in \mathbb{R}^d$, $V \subseteq \{v_1, \cdots, v_n\}$ and a number $k_V \geq 1$, the (V, k_V)-spectral peeling coreset U is defined by the same procedure as in Definition 17, but replacing the local optimal U_i by a $O(d)$-spectral spanner U_i of V_i (see Definition 32 and Theorem 33). Then $|U| \leq \tilde{O}(k_V d)$. For any S with $|S \cap V| \leq k_V$, there exists a distribution μ_v for $v \in S \cap V$ with disjoint supports s.t. supp$(\mu_v) \subseteq U$ and for any regular objective function f :*

$$f(\sum_{v \in S} vv^\intercal) \leq f(\sum_{v \in S} \mathbb{E}_{u \sim \mu_v} d u u^\intercal)$$

Consequently, for $k_V = k$, the set U serves as a $\tilde{O}(d)$-composable coreset for the experimental design problem under cardinality constraint k w.r.t. f.

Proof. Consider the set $(S \cap V) \setminus U$. Since $|S \cap V| \leq k_V$, there is an injective map $\pi : (S \cap V) \setminus U \to [k_V]$ s.t. $S \cap U_{\pi(v)} = \emptyset$ for each $v \in (S \cap V) \setminus U$. For each $v \in (S \cap V) \setminus U$, since $v \in V_{\pi(v)}$, we can use the fact that $U_{\pi(v)}$ is a spectral spanner of V_i to deduce that there exists μ_v supported on $U_{\pi(v)}$ where

$$vv^\intercal \preceq d \mathbb{E}_{u \sim \mu_v} u u^\intercal.$$

Note that μ_v are disjoint by injectivity of π. The claim then follows from the monotonicity of f.

For sets $V_1, \cdots, V_m$ let U_i' be the (V_i, k)-peeling coreset for V_i. Let $V' := \bigcup_{i=1}^m V_i$ and $U' := \bigcup U_i'$.

Let $S \in \binom{V'}{k}$ be a subset that maximizes $f(\sum_{v \in S} vv^\intercal)$. Using the above argument, we obtain that U' contains a fractional solution $s \in [0,1]^{U'}$ s.t. $\sum s_i = k$ and

$$f(\sum_{v \in S} vv^\intercal) \leq d f(\sum_{i \in U'} s_i v_i v_i^\intercal) \leq O(d) f(\sum_{u \in \tilde{S}} uu^\intercal)$$

for some $\tilde{S} \in \binom{U'}{k}$, where the second inequality follows from Theorem 35. □

Using similar construction and proof technique, we obtain $\tilde{O}(d)$-composable coreset of size $\tilde{O}(dk)$ and $\tilde{O}((dk)^k)$ respectively for A-design under certain laminar and partition matroid constraint $\mathcal{M}$ where $k_F \geq Cd$ for $\forall F \in \mathcal{F}$.