Domain Sparsification of Discrete Distributions using Entropic Independence

Nima Anari[1], Michał Dereziński[2], Thuy-Duong Vuong[1], and Elizabeth Yang[3]

[1]Stanford University, {anari,tdvuong}@stanford.edu
[2]University of Michigan, mderezin@umich.edu
[3]UC Berkeley, elizabeth_yang@berkeley.edu

Abstract

We present a framework for speeding up the time it takes to sample from discrete distributions μ defined over subsets of size k of a ground set of n elements, in the regime where k is much smaller than n. We show that if one has access to estimates of marginals $\mathbb{P}_{S\sim\mu}[i \in S]$, then the task of sampling from μ can be reduced to sampling from related distributions ν supported on size k subsets of a ground set of only $n^{1-\alpha} \cdot \text{poly}(k)$ elements. Here, $1/\alpha \in [1, k]$ is the parameter of entropic independence for μ. Further, our algorithm only requires sparsified distributions ν that are obtained by applying a sparse (mostly 0) external field to μ, an operation that for many distributions μ of interest, retains algorithmic tractability of sampling from ν. This phenomenon, which we dub domain sparsification, allows us to pay a one-time cost of estimating the marginals of μ, and in return reduce the amortized cost needed to produce many samples from the distribution μ, as is often needed in upstream tasks such as counting and inference.

For a wide range of distributions where $\alpha = \Omega(1)$, our result reduces the domain size, and as a corollary, the cost-per-sample, by a $\text{poly}(n)$ factor. Examples include monomers in a monomer-dimer system, non-symmetric determinantal point processes, and partition-constrained Strongly Rayleigh measures. Our work significantly extends the reach of prior work of Anari and Dereziński who obtained domain sparsification for distributions with a log-concave generating polynomial (corresponding to $\alpha = 1$). As a corollary of our new analysis techniques, we also obtain a less stringent requirement on the accuracy of marginal estimates even for the case of log-concave polynomials; roughly speaking, we show that constant-factor approximation is enough for domain sparsification, improving over $O(1/k)$ relative error established in prior work.

1 Introduction

Sparsification has been a crucial idea in designing many fast algorithms; famous examples include cut or spectral graph sparsifiers [ST11] and dimension reduction using sparsified Johnson-Lindenstrauss transforms [DKS10]. In this work, we address the question of sparsifying discrete distributions, with the goal of speeding up the fundamental task associated with distributions: sampling from them.

As an illustrative example building towards our notion of sparsification for distributions, consider the task of sampling a (uniformly) random edge in a graph. Suppose that we have a graph on n non-isolated vertices, and we are allowed to make adjacency queries. How many of the n vertices do we have to "probe" before we observe both endpoints of some edge? The answer to this question depends on the structure of the graph; for a star graph, where a single vertex is connected to the other $n-1$, we would have to find this central vertex to have any chance of observing an edge; so no amount of smart guessing can result in looking at $\ll n$ vertices. However, for regular graphs, where every vertex has the same degree, because of Birthday Paradox phenomena, it is enough to look at a sample of $O(\sqrt{n})$ uniformly random vertices to observe an edge between two of them with overwhelming probability. The bound of $O(\sqrt{n})$ is indeed the best possible, since in a random perfect matching (degree 1 regular graph), the best and only sensible strategy is to pick vertices at random.

This phenomenon generalizes to hypergraphs as well. On a k-uniform hypergraph, with hyperedges representing sets of k vertices, to observe a hyperedge one has to generally look at $\simeq n$ vertices in the worst case. But on regular hypergraphs, a substantially smaller sample, namely $\simeq n^{1-1/k}$ many vertices, will contain a hyperedge with high probability [see, e.g., Suz+06, for forms of Birthday Paradox related to k-sets and k-collisions]. Notice that this improvement quickly deteriorates as k gets large, and becomes meaningless as soon as $k \simeq \log n$.

When can the bound of $n^{1-1/k}$ for regular hypergraphs be improved, ideally to a polynomially small fraction of the n vertices, even for $k \gg 1$? Moreover, suppose that instead of desiring just one of the hyperedges, we want to extract an (approximately) *uniformly random hyperedge*. Can we still produce a hyperedge following this distribution, by only looking at a small subset of vertices? In a nutshell, how small of a (random) vertex set can we look at in order to have a distributionally representative "sparsification" of the entire hypergraph?

In this work, we tie the answer to these questions to notions of high-dimensional expansion, specifically the notion of *entropic independence* introduced by Anari, Jain, Koehler, Pham, and Vuong [Ana+21]. To every measure on size k subsets of $\{1, \ldots, n\}$, a.k.a. weighted hypergraph, denoted by $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, one can associate a parameter of entropic independence $1/\alpha \in [1, k]$, defined formally in Section 2. A larger α corresponds to better high-dimensional expansion. We show that the hypergraph defined by μ, while being k-uniform, behaves almost as if it was $(1/\alpha)$-uniform: informally, we can "sparsify" this hypergraph by looking at only $n^{1-\alpha} \cdot \text{poly}(k)$ vertices, under "regularity assumptions." To avoid confusion with classical concepts of graph and hypergraph sparsification, which primarily keep the vertex set while deleting a subset of the edges, we call this type of sparsification *domain sparsification*.

A long line of recent works have obtained breakthroughs in sampling and counting by viewing combinatorial distributions as (weighted) hypergraphs and studying notions of high-dimensional expansion for them [Ana+19; CGM19; AL20; ALO20; CLV20; GM20; Ana+20; Che+21b; Fen+21; Ali+21; Liu21; Bla+21; JPV21; Ana+21; ALO21; Che+21a]. A central theme in all of the aforemen-

tioned works is the establishment of some form of high-dimensional expansion for a hypergraph encoding the probability distribution of interest. At a high-level, these notions can be viewed as measures of proximity to independent/product distributions. Sampling from distributions extremely close to product distributions is roughly as easy as sampling i.i.d. from the marginal distribution over single elements; it is no surprise then, that for distributions with limited correlations, knowledge of marginals can boost sampling time. This is what we formally establish in this work.

Our main result applies to distributions that have *entropic independence* [Ana+21], a notion stronger than *spectral independence* [ALO20], but weaker than *fractional log-concavity* and *sector-stability* [Ali+21]. Roughly speaking, a background measures μ over k-sized sets is entropically independent, if for any (randomly chosen) set S, the relative entropy of a uniformly random element of S is at most $1/\alpha k$ fraction of the relative entropy of S, where we usually take $\alpha = \Omega(1)$. The main intuition leading to our results is that high correlations in such distributions must be limited to small groups of elements. By sampling enough many elements from the domain, we cover these correlated groups.

Similar to graph sparsification, in domain sparsification we need to reweigh the sparsified object. This is achieved by the standard operation of applying an external field. For a weight vector $\lambda \in \mathbb{R}_{\geq 0}^n$, the λ-external field applied to μ is the distribution $\lambda \star \mu$ defined by

$$\lambda \star \mu(S) \propto \mu(S) \prod_{i \in S} \lambda_i.$$

Theorem 1 (Informal). *Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be $(1/\alpha)$-entropically independent. Suppose that we have access to estimates $p_1, \ldots, p_n$ of the marginals and an oracle that can produce i.i.d. samples $i \in [n]$ with $\mathbb{P}[i] \propto p_i$; suppose that our estimates satisfy $p_1 + \cdots + p_n = k$ and*

$$p_i \geq \Omega(\mathbb{P}[i \in S])$$

for all i. Then we can produce a random sparse external field $\lambda \in \mathbb{R}_{\geq 0}^n$ with at most $n^{1-\alpha} \cdot \mathrm{poly}(k)$ nonzero entries, in time $n^{1-\alpha} \cdot \mathrm{poly}(k)$, such that a random sample S of $\lambda \star \mu$ approximately follows the distribution defined by μ.

Theorem 1 follows directly from Propositions 24 and 25 and Lemma 26. We outline our techniques in Section 1.3.

Algorithm. Our sampling algorithm, or rather a simplification of it, is easiest to describe for the case of uniform $p_i = k/n$ and when the desired total variation distance is a constant: we sample a uniformly random subset T of size roughly $\simeq n^{1-\alpha} \cdot \mathrm{poly}(k)$ elements from the ground set and then sample S from μ conditioned on $S \subseteq T$, i.e., from $\mathbb{1}_T \star \mu$. In order to handle non-uniform p_i, we perform a transformation to μ, which algorithmically translates to selecting elements of T with biases dictated by p_i. See Section 3.1 for details. Finally, in order to achieve logarithmic dependence on the total variation distance, we perform a Markov-chain-based variant of the mentioned process; in each step of this Markov chain, we simply *add* elements to the current set S in order to get T, and then subsample a new set of size k from T. For details see Section 3.2.

Notice that we do not need degree-regularity of μ, which would be equivalent to $\mathbb{P}_{S \sim \mu}[i \in S]$ being *exactly* the same for all i. Instead, it is enough to just have an *estimate* of these marginals. This is because instead of sampling a subset of vertices uniformly at random, we can sample a biased subset of vertices, with probability biases defined by the marginals, and domain sparsification will still work.

3

Prior to our work, domain sparsification was known for distributions with log-concave generating polynomials (the case of $\alpha = 1$) [AD20], based on techniques inspired by earlier algorithms for sampling from determinantal point processes (an even narrower class) [Der19; DCV19]. All of these distributions satisfy forms of negative dependence [BBL09; AD20] that were crucial in obtaining domain sparsification for them. Our work significantly extends the reach of domain sparsification beyond these classes; in particular, for *any distribution μ*, we have $\alpha \geq 1/k$, and as a simple corollary we get nontrivial domain sparsification for *any distribution μ* as long as $k = O(1)$, a result which appears to be nontrivial on its own.

The main application of domain sparsification is in accelerating the time it takes to produce multiple samples from a distribution μ. Suppose that an algorithm $\mathcal{A}$ can produce (approximate) samples from a distribution μ and any distribution obtained from it by an external field, in time $T(n, k)$, which usually depends polynomially on n.[1] Then after a preprocessing step, where we use $\mathcal{A}$ to estimate the marginals of μ, we can produce new samples in time $T(n^{1-\alpha} \cdot \text{poly}(k), k)$ per sample, which is polynomially smaller than $T(n, k)$, as long as k is smaller than some $\text{poly}(n)$ threshold. Notice that the preprocessing step has to be done only once, and its cost gets amortized when we are interested in obtaining multiple samples from μ. A careful implementation, directly adapted from what was done for log-concave polynomials by Anari and Dereziński [AD20], can bootstrap domain sparsification with estimation of marginals to complete the preprocessing step in roughly $\simeq T(n, k) + n \cdot \text{poly}(k, \log n) \cdot T(n^{1-\alpha} \cdot \text{poly}(k), k)$ time.

Corollary 2 (Informal, adapted from [AD20]). *Suppose that we have an algorithm $\mathcal{A}$ that can produce approximate samples from any external field λ applied to μ in time $T(m, k)$, where m is the sparsity of λ. Then we can produce the marginal estimates p_i and the i.i.d. sampling oracle required in* Theorem 1 *in time*

$$O\big(T(n, k) + n \cdot \text{poly}(k, \log n) \cdot T(n^{1-\alpha} \cdot \text{poly}(k), k)\big).$$

Further, for any desired t, we can produce t i.i.d. approximate samples from μ in time

$$O\big(T(n, k) + \max\{t, n \cdot \text{poly}(k, \log n)\} \cdot T(n^{1-\alpha} \cdot \text{poly}(k), k)\big).$$

Sampling is often used to solve the problem of approximate counting, that is computing the partition function

$$\sum_S \mu(S).$$

To obtain an ϵ-relative error approximation, known reductions between counting and sampling [JVV86] introduce at least a multiplicative factor of $1/\epsilon^2$ to the sampling time. Directly adapting the same technique for log-concave polynomials [AD20] and combining with our new domain sparsification result, we obtain an ϵ-relative error of the counts in time $\simeq T(n, k) + \max\{n, 1/\epsilon^2\} \cdot \text{poly}(k, \log n) \cdot T(n^{1-\alpha} \cdot \text{poly}(k), k)$. Notice that here $1/\epsilon^2$ is multiplied by the term $T(n^{1-\alpha} \cdot \text{poly}(k), k)$ that can be substantially smaller than $T(n, k)$; as a result, we can get a substantially improved running time for the high-precision regime where ϵ is inverse-polynomially small.

Corollary 3 (Informal, adapted from [AD20]). *Suppose that we have an algorithm $\mathcal{A}$ that can produce approximate samples from any external field λ applied to μ in time $T(m, k)$, where m is the sparsity of λ. Then we can compute an ϵ relative error approximation of $\sum_S \mu(S)$ in time*

$$O\big(T(n, k) + \max\{n, 1/\epsilon^2\} \cdot \text{poly}(k, \log n) T(n^{1-\alpha} \cdot \text{poly}(k), k)\big).$$

---

[1]Typically the running time has logarithmic dependencies on the approximation error and potentially magnitude of external fields, but for simplicity of exposition we hide them here.

Remark 4. For many applications, we can derive entropic independence of μ from a stronger property called fractional log-concavity [Ali+21; Ana+21]. For an α-fractionally log-concave distribution, recent work of Anari, Jain, Koehler, Pham, and Vuong [Ana+21] established Modified Log-Sobolev Inequalities for natural (multi-step) down-up random walks. For simplicity of exposition, assume that $1/\alpha \in \mathbb{Z}$ and $\alpha = \Omega(1)$. Then, these random walks produce approximate samples from μ in the following number of steps:

$$O\left(k^{1/\alpha} \cdot \log\log \frac{1}{\mathbb{P}_\mu[S_0]}\right),$$

where S_0 is the starting point of the random walk. Further, each step of the random walk requires querying μ at $n^{1/\alpha}$ points, leading to a total runtime of

$$O\left((kn)^{1/\alpha} \cdot \log\log \frac{1}{\mathbb{P}_\mu[S_0]}\right).$$

In most settings, such as when the bit-complexity of μ is bounded by $\mathrm{poly}(n)$, the extra $\log\log(1/\mathbb{P}_\mu[S_0])$ can be safely ignored, as long as we make sure S_0 is in the support. So one can think of this Markov chain as an algorithm $\mathcal{A}$ that, up to this initial step of finding a suitable starting point, satisfies

$$T(n,k) = \tilde{O}\left((nk)^{1/\alpha}\right).$$

Note that for this choice of the algorithm $\mathcal{A}$ and running time T, the bounds in Corollaries 2 and 3 simplify as

$$n \cdot \mathrm{poly}(k, \log n) \cdot T(n^{1-\alpha} \cdot \mathrm{poly}(k), k) \simeq \mathrm{poly}(k, \log n) \cdot T(n,k).$$

However, our results apply to any choice of a base sampling algorithm $\mathcal{A}$.

A challenging part of obtaining our results is the lack of negative dependence inequalities, which were used by the prior work of Anari and Dereziński [AD20]. These negative dependence inequalities result in domain sparsification with sparsified domain size solely depending on k, with no dependence on n. We show in Section 4 that our analysis of our domain sparsification scheme is tight. An intriguing question is if we can find other domain sparsification schemes, perhaps using higher-order marginals, that sparsify domains to size $\mathrm{poly}(k, \log n)$? We make the following conjecture.

Conjecture 5 (Informal). *Let μ be an α-fractionally-log-concave distribution for some $\alpha = \Omega(1)$. Given access to estimates for high-order marginals of the form $\mathbb{P}_{S\sim\mu}[T \subseteq S]$ for all T of size $\ell \simeq 1/\alpha$, and an oracle that produces i.i.d. samples from these marginals, there is a domain sparsification scheme for μ which reduces the domain size to only $\mathrm{poly}(k)$.*

Despite the attractiveness of a bound independent of n, we give evidence that obtaining these domain sparsification schemes requires entirely new ideas; we show in Section 4 that if we replace fractional log-concavity by entropic independence (which is sufficient for our main result, Theorem 1) in the above conjecture, the conjecture becomes false.

1.1 Applications

Here we mention examples of distributions to which our results can be applied beyond those covered by prior work of Anari and Dereziński [AD20]. Our examples satisfy fractional log-concavity [Ali+21] which entails both entropic independence [Ana+21], and the existence of the base sampling algorithm $\mathcal{A}$.

For a distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ we define the generating polynomial g_μ to be

$$g_\mu(z_1, \ldots, z_n) := \sum_S \mu(S) \prod_{i \in S} z_i.$$

We say that the distribution μ or the polynomial g_μ is α-fractionally-log-concave for some parameter $\alpha \leq 1$ if $\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is concave as a function over the positive orthant $(z_1, \ldots, z_n) \in \mathbb{R}_{\geq 0}^n$.

Notice that any multi-affine homogeneous polynomial with nonnegative coefficients is the generating polynomial of a distribution μ. Throughout the paper, we often equate these polynomials with the distributions they represent, and freely talk about fractional log-concavity of either the generating polynomial or the distribution. For more details, see [Ali+21].

Example 6. If g is a degree-k homogeneous multi-affine polynomial, then it is $\frac{1}{k}$-log-concave. Every monomial $\prod_{i \in S} z_i^{1/k}$ is concave, since by Hölder's inequality

$$\prod_{i \in S}(\lambda z_i + (1 - \lambda)y_i) \geq \left(\lambda \prod_{i \in S} z_i^{1/k} + (1 - \lambda) \prod_{i \in S} y_i^{1/k} \right)^k.$$

Now, $g(z_1^{1/k}, \ldots, z_n^{1/k})$ is concave as (weighted)-sum of concave functions $\prod_{i \in S} z_i^{1/k}$. Thus is also log-concave, as $\log$ is a monotone and concave function.

Example 7. We present another toy class of α-fractionally-log-concave polynomials that provides some intuition despite not having many applications. Let μ be an α-fractionally-log-concave polynomial over the variables $z_1, \ldots, z_n$. If we replace each z_i with the monomial $\prod_{j=1}^m z_i^{(j)}$, we obtain a degree mk, $\frac{\alpha}{m}$-fractionally-log-concave polynomial over the variables $\{z_i^{(j)} \mid i \in [n], j \in [m]\}$ [Ali+21]. For example, if the starting distribution μ is the uniform distribution over bases of a matroid, then $\alpha = 1$, and the resulting distribution will be $1/m$-fractionally log-concave.

Notice that if we normalize this blown-up polynomial to convert it into to a distribution, for any $i \in [n]$, the elements $i^{(1)}, \ldots, i^{(m)}$ are all perfectly correlated. On the other hand, if $i \neq j$, any two elements $i^{(m_i)}, j^{(m_j)}$ inherit the correlations from the log-concave distribution.

Example 8. Let G be a graph and $k \in \mathbb{N}$. For each set $S \subseteq \binom{V}{2k}$, set $\mu(S)$ to be proportional to the number of perfect matchings on S. Sampling from μ allow us to approximately count the number of k-matching, i.e., matchings using k edges. [Ali+21] proved that for any value of k, this distribution is fractionally log-concave with $\alpha \geq 1/4$.

Not all choices of G result in efficient sampling algorithms. The implementation of each iteration of the Markov chain involves counting perfect matchings over $S \subseteq V$, and we do not have a $\mathrm{poly}(k)$ time algorithm for counting matchings in general graphs. We thus only consider downward closed graph families with an FPRAS for counting perfect matchings, e.g., bipartite graphs [JSV04], planar graphs [Kas67], certain minor-free graphs [EV19], and small genus graphs [GL99]. Our main results imply that as long as we estimate the probability of every vertex being part of a random k-matching, we can reduce the task of sampling k-matchings on an n vertex graph to graphs with only $n^{3/4} \cdot \mathrm{poly}(k)$ many vertices.

Example 9. Let L be a nonsymmetric positive semidefinite matrix, i.e., an $n \times n$ matrix L that satisfies $L + L^\intercal \succeq 0$. Then, the nonsymmetric k-determinantal point process [see, e.g., Gar+19; Gar+20; AV21] with kernel L, defined by

$$\mu(S) = \det(L_{S,S})$$

for all $S \in \binom{[n]}{k}$ is fractionally log-concave with $\alpha \geq 1/4$ [Ali+21].

Example 10. Suppose that we start with a measure μ_0 on $\binom{[n]}{k}$ that is Strongly Rayleigh [see BBL09, for definition], such as a (symmetric) determinant point process, or the uniform distribution over spanning trees of a graph. Suppose that we partition the ground set into a constant number $c = O(1)$ of parts: $[n] = A_1 \cup A_2 \cup \cdots \cup A_c$, and fix cardinalities $k_1, \ldots, k_c \in \mathbb{Z}_{\geq 0}$, with $k_1 + \cdots + k_c = k$. Then the partition-constrained version of μ_0 can be defined as

$$\mu(S) \propto \mu_0(S) \cdot \mathbb{1}[|S \cap A_i| = k_i \text{ for } i = 1, \ldots, c].$$

As long as $c = O(1)$, this distribution μ will be $\Omega(1)$-fractionally-log-concave [Ali+21]. For some discussion of partition-constrained Strongly Rayleigh measures, see [Cel+16].

1.2 Related Work

Log-Concavity

Log-concavity has been a well-studied concept in continuous sampling since it captures many common distributions like uniform distributions over convex bodies, and Gaussian distributions. Discrete notions of log-concavity we work with in this paper have been introduced by [Gur09; AOV18; BH19; Ana+19]. The formulation of [Ana+19] is what we refer to in this paper as "log-concave," and is motivated by examples such as the uniform distribution over bases of a matroid and the special subcase of the uniform distribution over spanning trees.

We have a nearly complete picture for MCMC-based sampling algorithms for homogeneous log-concave distributions. For degree-k distributions, Anari, Liu, Oveis Gharan, and Vinzant [Ana+19] analyzed the down-up walk, which occurs between sets of size k and sets of size $(k-1)$; in the case of matroid bases, this walk is also known as a form of the "basis exchange walk." Furthermore, Cryan, Guo, and Mousa [CGM19] proved a Modified Log-Sobolev Inequality (MLSI) for this walk, and Anari, Liu, Oveis Gharan, Vinzant, and Vuong [Ana+20] further reduced the runtime of sampling by analyzing a warm start to the down-up walk algorithm. Most recently, Anari and Dereziński [AD20] devised an algorithm for sampling a log-concave distribution when we are given the single-element marginals; we will elaborate upon their contributions more in Section 3.3, where we compare their algorithm to ours.

Intermediate Sampling and Determinantal Point Processes

A class of domain sparsification algorithms, related to the algorithms we used here, called intermediate sampling was first proposed by [DWH18; Der19] in the context of sampling from Determinantal Point Processes (DPPs, [DM21]), also known as Volume Sampling [Des+06; DR10; GS12]. DPPs are a family of distributions (a small, but important, subset of distributions with log-concave generating polynomials) which arise for instance when sampling random spanning trees [Gué83], as well as in randomized linear algebra [DW17; Der+19], machine learning [KT11; KT12; DKM20], optimization [NST19; Der+20; MDK20], and other areas [Mac75; Hou+06; Bar+17].

The complexity of intermediate sampling for DPPs was further improved by [DCV19; CDV20], and the approach was extended to DPPs over continuous domains by [DWH19]. Crucially, these algorithms take advantage of the additional structure in DPPs, to enable *distortion-free* intermediate sampling: instead of using a Markov chain, this uses rejection sampling to draw *exactly* from the target distribution. This approach is not possible more generally, since μ typically does not have

a tractable partition function. However, [AD20] showed that the original analysis of distortion-free intermediate sampling can largely be retained for distributions with log-concave generating polynomials, as long as we switch to a Markov chain implementation.

On the other hand, in this work, we largely abandon the original analysis in favor of a new one which is specific to the Markov chain and requires less precision in marginal estimates. As a result, we show that the preprocessing cost for Markov chain intermediate sampling is substantially smaller than for distortion-free intermediate sampling. This leads to significant improvements in time complexity even for DPPs, e.g., by reducing the preprocessing cost in [DCV19] from $\tilde{O}(nk^6 + k^9)$ to $\tilde{O}(nk^2 + k^3)$, where $\tilde{O}$ hides polylogarithmic terms.

1.3 Overview of Techniques

Given an entropically independent distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, we first preprocess it using *isotropic transformation*, which is further detailed in Section 3.1. This converts our distribution μ into a related distribution μ' whose single element marginals $\mathbb{P}_{S' \sim \mu'}[i \in S']$ are approximately uniform. We prove various properties of μ' in Proposition 24, including the fact that μ' has ground set size linear in n. This preprocessing step may be of independent interest for other discrete sampling problems outside of entropically independent and fractionally log-concave distributions.

After this step, we may assume our distribution μ has already undergone isotropic transformation. Next, we design a Markov chain M_μ^t that has μ as its stationary distribution, where taking a step requires sampling from a sparsified distribution ν. We refer to this algorithm as *Markov Chain Intermediate Sampling*. The benefit of this Markov chain over other natural Markov chains (e.g., down-up random walks [see, e.g., Ali+21]) is that each step requires paying attention only to a subset of elements as opposed to all. We show that $n^{1-\alpha} \cdot \text{poly}(k)$ size is sufficient to ensure that M_μ^t mixes rapidly. Specifically, in Lemma 26, we prove that a single step of M_μ^t from any $S \in \text{supp}(\mu)$ satisfies $\|P(S, \cdot) - \mu\| \leq \frac{1}{4}$.

The mixing time analysis of M_μ^t is novel and improves upon the methods used in [AD20]. The improvement is discussed and concretely illustrated with an example distribution in Section 3.3. The proof of Lemma 26 relies heavily on new negative dependence inequalities that are "average-case" rather than "worst-case", since the worst-case inequalities simply do not hold for fractionally-log-concave distributions. We show that these "average-case" inequalities suffice for fast mixing of M_μ^t, thus opening the intermediate sampling framework to wider families of distributions.

1.4 Acknowledgements

2 Preliminaries

We use $[n]$ to denote the set $\{1, \ldots, n\}$. For a set S, we use $\binom{S}{k}$ to denote the family of subsets of S of size k. For a distribution μ, we use $X \sim \mu$ to denote that X is a random variable distributed according to μ. For a set U, we abuse notation and let $X \sim U$ denote X following the *uniform* distribution over U.

For a distribution μ over size k sets and a set T of size potentially larger than k, we abuse notation and use $\mu(T)$ to denote:

$$\mu(T) := \sum_{S \subseteq T} \mu(S).$$

2.1 Markov Chains and Mixing Time

Definition 11. Let μ, ν be two discrete probability distributions over the same event space Ω. The *total variation distance*, or TV-*distance*, between μ and ν is given by

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|$$

Definition 12. Let P be an ergodic Markov chain on a finite state space Ω and let μ denote its (unique) stationary distribution. For any probability distribution ν on Ω and $\epsilon \in (0, 1)$, we define

$$t_{\mathrm{mix}}(P, \nu, \epsilon) = \min\{t \geq 0 \mid \|\nu P^t - \mu\|_{\mathrm{TV}} \leq \epsilon\},$$

and

$$t_{\mathrm{mix}}(P, \epsilon) = \max\{t_{\mathrm{mix}}(P, \mathbb{1}_x, \epsilon) \mid x \in \Omega\},$$

where $\mathbb{1}_x$ is the point mass distribution supported on x.

We will drop P and ν if they are clear from context. Moreover, if we do not specify ϵ, then it is set to $1/4$. This is because the growth of $t_{\mathrm{mix}}(P, \epsilon)$ is at most logarithmic in $1/\epsilon$ (cf. [LP17]).

The modified log-Sobolev constant of a Markov chain, defined next, provides control on its mixing time. For a detailed coverage see [LP17].

Definition 13. Let P denote the transition matrix of an ergodic, reversible Markov chain on Ω with stationary distribution μ.

- The *Dirichlet form* of P is defined for $f, g \in \Omega \to \mathbb{R}$ by

$$\mathcal{E}_P(f, g) = \langle f, (I - P)g \rangle_\mu = \langle (I - P)f, g \rangle_\mu.$$

- The *modified log-Sobolev constant* of P is defined to be

$$\rho_0(P) = \inf\left\{\frac{\mathcal{E}_P(f, \log f)}{2 \cdot \mathrm{Ent}_\mu[f]} \,\middle|\, f : \Omega \to \mathbb{R}_{\geq 0}, \mathrm{Ent}_\mu[f] \neq 0\right\},$$

  where

$$\mathrm{Ent}_\mu[f] = \mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log \mathbb{E}_\mu[f].$$

  Note that, by rescaling, the infimum may restrict our attention to functions $f : \Omega \to \mathbb{R}_{\geq 0}$ satisfying $\mathrm{Ent}_\mu[f] \neq 0$ and $\mathbb{E}_\mu[f] = 1$.

The relationship between the modified log-Sobolev constant and mixing times is captured by the following well-known lemma.

Lemma 14 (cf. [BT06]). *Let P denote the transition matrix of an ergodic, reversible Markov chain on Ω with stationary distribution μ and let $\rho_0(P)$ denote its modified log-Sobolev constant. Then, for any probability distribution ν on Ω and for any $\epsilon \in (0, 1)$*

$$t_{\mathrm{mix}}(P, \nu, \epsilon) \leq \left\lceil \rho_0(P)^{-1} \cdot \left(\max_{x \in \Omega} \log \log \left(\frac{\nu(x)}{\mu(x)} \right) + \log \left(\frac{1}{2\epsilon^2} \right) \right) \right\rceil.$$

In particular,

$$t_{\mathrm{mix}}(P, \epsilon) \leq \left\lceil \rho_0(P)^{-1} \cdot \left(\log \log \left(\frac{1}{\min_{x \in \Omega} \mu(x)} \right) + \log \left(\frac{1}{2\epsilon^2} \right) \right) \right\rceil.$$

Theorem 15 (cf. [LP17]). *If an irreducible aperiodic Markov chain with stationary distribution μ and transition matrix P satisfies $\|P^t(S, \cdot) - \mu\|_{TV} \leq 1/4$ for all $S \in \mathrm{supp}(\mu)$ and some $t \geq 1$, then for any $\epsilon \in (0, 1/4]$,*

$$t_{mix}(P, \epsilon) \leq t \log(1/\epsilon).$$

2.2 Fractional Log-Concavity and Entropic Independence

We recall the notion of fractional log-concavity [Ali+21] and entropic independence [Ana+21].

Definition 16 ([Ali+21]). A probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is *α-fractionally-log-concave* if $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is log-concave for $z_1, \ldots, z_n \in \mathbb{R}_{\geq 0}^n$. If $\alpha = 1$, we say μ is log-concave.

To define entropic independence we need the definition of the "down" operator.

Definition 17 (Down Operator). For $\ell \leq k$ define the row-stochastic matrix $D_{k \to \ell} \in \mathbb{R}_{\geq 0}^{\binom{[n]}{k} \times \binom{[n]}{\ell}}$ by

$$D_{k \to \ell}(S, T) = \begin{cases} 0 & \text{if } T \not\subseteq S, \\ \frac{1}{\binom{k}{\ell}} & \text{otherwise.} \end{cases}$$

Note that for a distribution μ on size k sets, $\mu D_{k \to \ell}$ will be a distribution on size ℓ sets. In particular, $\mu D_{k \to 1}$ will be the vector of normalized marginals of μ: $(\mathbb{P}[i \in S]/k)_{i \in [n]}$.

Definition 18 ([Ana+21, Definition 2, Theorem 3]). A probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is *$(1/\alpha)$-entropically-independent* for $\alpha \in (0, 1]$, if for all probability distributions ν on $\binom{[n]}{k}$,

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k \to 1} \| \mu D_{k \to 1}) \leq \frac{1}{\alpha k} \mathcal{D}_{\mathrm{KL}}(\nu \| \mu).$$

Or equivalently,

$$\forall (z_1, \ldots, z_n) \in \mathbb{R}_{\geq 0}^n : g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{1/k\alpha} \leq \sum_{i=1}^n p_i z_i, \tag{1}$$

where $p = (p_1, \ldots, p_n) := \mu D_{k \to 1}$.

We note that α-fractionally log concavity implies $(1/\alpha)$-entropic independence [Ana+21, Theorem 3].

An important part of our sparsification scheme is a process to transform distributions into a "near-isotropic" position (defined as having roughly equal marginals) by subdividing the elements of the

ground set. More precisely, let μ be a distribution generated by $g_\mu(z_1, \ldots, z_n)$, then the distribution μ' obtained by subdividing z_i into t_i copies has generating polynomial

$$g_{\mu'}(z_1^{(1)}, \ldots, z_n^{(t_n)}) = g_\mu\left(\frac{z_1^{(1)} + \ldots + z_1^{(t_1)}}{t_1}, \ldots, \frac{z_n^{(1)} + \ldots + z_n^{(t_n)}}{t_n}\right).$$

Subdivision preserves both entropic independence and fractional log-concavity.

Proposition 19. *If μ is $(1/\alpha)$-entropically-independent distribution then μ' is also $(1/\alpha)$-entropic independence.*

Proposition 20. *If μ is α-fractionally-log-concave distribution then μ' is also α-fractionally-log-concave.*

We leave the proofs to [Section 5](#).

3 Intermediate Sampling Algorithm

3.1 Isotropic Transformation

We define, similar to [AD20], a distribution μ to be isotropic if for all $i \in [n]$, the marginal probability $\mathbb{P}_{S \sim \mu}[i \in S]$ is $\frac{k}{n}$. We remark that this is only similar in name and spirit, but different in nature, to the analogous notion of isotropy for continuous distributions; the latter is defined based on the covariance matrix of the distribution, while the former is defined based on marginals. In this paper, isotropy captures "uniformity" over the elements of $[n]$ in their marginal probabilities. Below we discuss a subdivision process [AD20] that transforms an arbitrary distribution μ over $\binom{[n]}{k}$ into a distribution μ' that is nearly-isotropic.

Definition 21. Let $\mu : \binom{n}{k} \to \mathbb{R}_{\geq 0}$ be an arbitrary probability distribution, and assume that we have estimates $p_1, \ldots, p_n$ of the marginals with $p_1 + \cdots + p_n = k$ and $p_i \geq \Omega(\mathbb{P}_{S \sim \mu}[i \in S])$ for all i. Let $t_i := \lceil \frac{n}{k} p_i \rceil$. We will create a new distribution out of μ: For each $i \in [n]$, create t_i copies of the element i and let the collection of all these copies be the new ground set: $U = \bigcup_{i=1}^n \{i^{(1)}, \ldots, i^{(t_i)}\}$. Define the following distribution $\mu' : \binom{U}{k} \to \mathbb{R}_{\geq 0}$ from μ:

$$\mu'\left(\left\{i_1^{(j_1)}, \ldots, i_k^{(j_k)}\right\}\right) := \frac{\mu(\{i_1, \ldots, i_k\})}{t_1 \cdots t_k}.$$

We call μ' the *isotropic transformation* of μ. Another way we can think of μ' is that to produce a sample from it, we can first generate a sample $\{i_1, \ldots, i_k\}$ from μ, and then choose a copy $i_m^{(j_m)}$ for each element i_m uniformly at random.

Remark 22. We note that subdivision or isotropic transformation and external fields behave well together. In particular, a sample from an external field λ applied to μ' can be obtained by first applying an appropriate external field (summing the field values over duplicate elements) to μ and then replacing each element with a copy of it with probability proportional to λ. In fact, subdivision is mostly a tool for analysis. In our algorithms, we never have to formally perform subdivision, and we can just sample from distributions defined as $\lambda \star \mu$ for appropriate external fields λ.

Remark 23. To obtain the estimates $\{p_i\}$ for all i, we can apply the proof of [AD20, Lem. 23], with ϵ constant, rather than $\epsilon = O(\frac{1}{k})$. This provides a running time reduction for our preprocessing step even in the case of log-concave polynomials.

There are three desirable properties of μ' we need to establish for subdivision to be an effective preprocessing step. The first is that subdivision preserves $(1/\alpha)$-entropic independence, which is shown in Proposition 19. The next is for the marginals $\mathbb{P}_{S \sim \mu'}[i^{(j)} \in S]$ to all be close to $\frac{k}{|U|}$ for all $i^{(j)} \in U$; in other words, μ' is actually close to being isotropic. The last is for $|U| \leq O(n)$, so if we ran a sampling algorithm on μ', the increased size of our ground set does not accidentally inflate our desired asymptotic running times. We remark however, that this last concern can be avoided by simply not running the sampling algorithm on the subdivided distribution, but rather on $\lambda \star \mu$ for an appropriate external field λ.

Proposition 24. *Let* $\mu : \binom{n}{k} \to \mathbb{R}_{\geq 0}$, *and let* $\mu' : \binom{U}{k} \to \mathbb{R}_{\geq 0}$ *be the subdivided distribution from Definition 21. The following hold for* μ':

1. *Near-isotropy: For all* $i^{(j)} \in U$, *the marginal* $\mathbb{P}_{S \sim \mu'}[i^{(j)} \in S] \leq O(k/|U|)$.

2. *Linear ground set size: The number of elements* $|U| \leq O(n)$.

Proof. First, we verify that $|U|$ is at most $O(n)$:

$$|U| = \sum_{i=1}^{n} t_i \leq \sum_{i=1}^{n} \left(1 + \frac{n}{k} p_i\right) = n + \frac{n}{k} \sum_{i=1}^{n} p_i = 2n.$$

Next, we check that for any $i^{(j)}$, the marginal probabilities $\mathbb{P}_{S \sim \mu'}[i^{(j)} \in S]$ are at most $O(k/|U|)$. Here, we interpret the sampling from μ' as first sampling from μ, and then choosing a copy for each element.

$$\mathbb{P}_{S \sim \mu'}[i^{(j)} \in S] = \sum_{S \ni i} \mathbb{P}[\text{we chose copy } j \mid \text{we sampled } S \text{ from } \mu] \cdot \mathbb{P}[\text{we sampled } S \text{ from } \mu]$$

$$= \sum_{S \ni i} \frac{1}{t_i} \cdot \mu(S) = \frac{1}{t_i} \sum_{S \ni i} \mu(S) = \frac{1}{t_i} \cdot \mathbb{P}_{S \sim \mu}[i \in S].$$

Since $t_i \geq \frac{n}{k} p_i \geq \frac{n}{k} \cdot \Omega(\mathbb{P}_{S \sim \mu}[i \in S])$, we get that

$$\mathbb{P}_{S \sim \mu'}[i^{(j)} \in S] \leq O\left(\frac{\mathbb{P}_{S \sim \mu}[i \in S]}{\frac{n}{k} \cdot \mathbb{P}_{S \sim \mu}[i \in S]}\right) = O(k/n) \leq O(k/|U|). \qquad \square$$

3.2 Domain Sparsification via Markov Chain Intermediate Sampling

Here, we first describe, for any general distributions μ, a Markov chain based on generating intermediate samples $T \subseteq [n]$, that mixes to μ. Then, in Lemma 26 and Proposition 27, we state our main result that for distributions μ which are $(1/\alpha)$-entropically independent and nearly-isotropic, the size of T only needs to be $n^{1-\alpha} \cdot \text{poly}(k)$ for the mixing to occur in one step.

Take distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, and consider the following Markov chain M_μ^t defined for any positive integer t, with the state space $\text{supp}(\mu)$. Starting from $S_0 \in \text{supp}(\mu)$, one step of the chain is given by:

1. Sample $T \sim \binom{[n] \setminus S_0}{t - k}$.

2. Downsample $S_1 \sim \mu_{S_0 \cup T}$, where $\mu_{S_0 \cup T}$ is μ restricted to $S_0 \cup T$, a.k.a. $\mathbb{1}_{S_0 \cup T} \star \mu$, and update S_0 to be S_1.

We note that the requirement $S_0 \in \mathrm{supp}(\mu)$ is not strictly necessary for this step to be defined.

Proposition 25. *For any distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, the chain M_μ^t for $t \geq 2k$ is irreducible, aperiodic and has stationary distribution μ.*

Proof. Let P denote the transition probability matrix of M_μ^t. Since $t \geq 2k$, for any $S, S' \in \mathrm{supp}(\mu)$, there is a positive probability that we sample $T \supseteq S \cup S'$. Thus, we have $P(S, S') > 0$, and P is both irreducible and aperiodic.

To derive the stationary distribution, suppose that we perform one step of the chain starting from $S_0 \sim \mu$. We first derive the distribution of the intermediate set $R := S_0 \cup T$.

For any $\tilde{R} \in \binom{[n]}{t}$, the probability of sampling $\tilde{R}$ for the intermediate set R is

$$\mathbb{P}[R = \tilde{R}] = \sum_{S_0 \in \binom{\tilde{R}}{k}} \mu(S_0) \cdot \mathbb{P}[T = \tilde{R} \setminus S_0] = \frac{1}{\binom{n-k}{t-k}} \cdot \mu(\tilde{R}).$$

For any $\tilde{S}_1 \in \mathrm{supp}(\mu)$, the probability of sampling $\tilde{S}_1$ is

$$\mathbb{P}[S_1 = \tilde{S}_1] = \sum_{\tilde{R} \in \binom{[n]}{r}} \mathbb{P}[S_1 = \tilde{S}_1 \mid R = \tilde{R}] \mathbb{P}[R = \tilde{R}] = \sum_{\tilde{R} \in \binom{[n]}{t}: \tilde{R} \supseteq \tilde{S}_1} \frac{\mu(\tilde{S}_1)}{\mu(\tilde{R})} \cdot \frac{1}{\binom{n-k}{t-k}} \cdot \mu(\tilde{R})$$

$$= \mu(\tilde{S}_1) \sum_{(\tilde{R} \setminus \tilde{S}_1) \in \binom{[n] \setminus \tilde{S}_1}{t-k}} \frac{1}{\binom{n-k}{t-k}} = \mu(\tilde{S}_1).$$

Above, we summed over all $\tilde{R}$ that contain the target set $\tilde{S}_1$. $\qquad\square$

The following lemma is the key to analyzing the sampling algorithm, since it quantifies the decrease in TV distance after running one step of M_μ^t. It will be proven in Section 3.4.

Lemma 26. *Let $\mu : \binom{[n]}{k}$ be a $(1/\alpha)$-entropically independent distribution. Suppose that for all $i \in [n]$, we have $\mathbb{P}_{S \sim \mu}[i \in S] \leq \frac{Ck}{n}$. Then, for any constant $\epsilon \in (0, \frac{1}{4}]$, and $t = \Omega\big(n^{1-\alpha}(Ck^2 \log \frac{1}{1-\epsilon})^\alpha\big)$, the output of a single step of the Markov chain M_μ^t starting from S_0 satisfies*

$$\forall S \in \binom{[n]}{k} : \mathbb{P}[S_1 = S] \geq \mu(S)(1 - \epsilon).$$

Recall that if we used the marginal estimates required by Theorem 1, then by applying Proposition 24 we get an equivalent distribution μ' on a ground set of size $O(n)$ that satisfies the above assumption of $\mathbb{P}_{S \sim \mu'}[i \in S] \leq Ck/n$ for some $C = O(1)$ (see Lemma 26).

Proposition 27. *Let $\mu : \binom{[n]}{k}$ be a $(1/\alpha)$-entropically independent distribution, and let $\epsilon \in (0, \frac{1}{4}]$ be a constant. Suppose $\mathbb{P}_{S \sim \mu}[i \in S] \leq \frac{Ck}{n}$ for all i. Choose the intermediate sample size t according to Lemma 26. Then*

$$\|P(S_0, \cdot) - \mu\|_{\mathrm{TV}} \leq \epsilon,$$

and the Markov chain M_μ^t mixes to a distribution that has TV distance $\epsilon' < \epsilon$ from μ in $O(\log(1/\epsilon'))$ steps.

Proof. The bound on TV distance follows via

$$\|P(S_0, \cdot) - \mu\|_{\mathrm{TV}} = \sum_{S \in \binom{[n]}{k} : \mathbb{P}[S_1 = S] < \mu(S)} (\mu(S) - \mathbb{P}[S_1 = S]) \le \epsilon \sum_{S \in \binom{[n]}{k} : \mathbb{P}[S_1 = S] < \mu(S)} \mu(S) \le \epsilon.$$

The mixing time bound follows from Theorem 15. $\qquad\square$

We have shown that M_μ^t is fast mixing (in fact, mixing in one step for appropriately large t). Next, we show that for a wide class of distributions, namely, the class of α-fractionally-log-concave distributions with $\alpha = \Omega(1)$ [see Ali+21, for examples], each step of M_μ^t can be implemented in $\mathrm{poly}(n, k)$ time via a local Markov chain, i.e., the (muti-step) down-up random walk [Ali+21, Def. 1].

Lemma 28 (Runtime analysis). *Suppose μ is α-fractionally-log-concave, and we start with $S_0^{(0)}$ such that $\mu(S_0^{(0)}) \ge 2^{-n^c}$ for some constant $c > 1$ and we run the chain for τ steps. The down-sampling takes time*

$$O\left((t - k)^{\lceil 1/\alpha \rceil} k^{1/\alpha} \left(c \log n + \log \tau + \log \log \frac{1}{1 - \epsilon} \right) \right),$$

and the total runtime is

$$O\left(\tau (t - k)^{\lceil 1/\alpha \rceil} k^{1/\alpha} \left(c \log n + \log \tau + \log \log \frac{1}{1 - \epsilon} \right) \right).$$

Proof. We first show that with probability $\ge 1 - \tau 2^{-n}$, for all $0 \le i \le \tau$, the ith-step starting point, denoted by $S_0^{(i)}$, satisfies $\mu(S_0^{(i)}) \ge 2^{-(n^c + 2ni)}$. This can be shown via induction on i. Conditioned on $\mu(S_0^{(i)}) \ge 2^{-n^c - 2ni}$, we have

$$\mathbb{P}[\mu(S_0^{(i+1)}) \le 2^{-(n^c + 2(i+1)n)}] = \mu(S_0^{(i)} \cup T)^{-1} \sum_{S \subseteq (S_0^{(i)} \cup T) : \mu(S) \le 2^{-(n^c + 2(i+1)n)}} \mu(S)$$

$$\le_{(1)} \frac{2^{-(n^c + 2(i+1)n)} \cdot 2^n}{\mu(S_0^{(i)})} \le 2^{-n}.$$

where in (1) we use the crude bound $\left| \left\{ S \subseteq (S_0^{(i)} \cup T) \, \middle| \, \mu(S) \le 2^{-(n^c + 2(i+1)n)} \right\} \right| \le 2^n$.

Suppose that this good event happens, i.e.

$$\forall i \in [0, \tau] : \mu(S_0^{(i)}) \ge 2^{-n^c - 2ni}.$$

We observe that α-fractional-log-concavity is preserved by subdividing Proposition 20 and restricting to a subset of the ground set [Ali+21]. In the down-sampling step, we run the (multi-step) down-up walk starting at $S_0^{(i)}$, and use [Ana+21, Thm. 4] to bound the runtime and obtain the

desired result. To this end, we need to bound

$$\mathbb{E}_{T \sim \binom{[n] \setminus S_0^{(i)}}{t-k}} \left[\log \left(1 + \log \frac{\mu(S_0^{(i)} \cup T)}{\mu(S_0^{(i)})} \right) \right] \leq_{(1)} \log \left(1 + \log \mathbb{E}_{T \sim \binom{[n] \setminus S_0^{(i)}}{t-k}} \left[\frac{\mu(S_0^{(i)} \cup T)}{\mu(S_0^{(i)})} \right] \right)$$

$$= \log \left(1 + \log \frac{1}{\mathbb{P}[S_1 = S_0^{(i)}]} \right)$$

$$\leq_{(2)} \log \left(1 + \log \frac{1}{\mu(S_1 = S_0^{(i)})(1 - \epsilon)} \right)$$

$$\leq_{(3)} c \log n + \log \tau + \log \log \frac{1}{1 - \epsilon}.$$

where (1) follows from Jensen's inequality for the concave function $f(x) = \log(1 + \log(x))$ on $[1, \infty)$, (2) from Lemma 26 and (3) from lower bound on $\mu(S_0^{(i)})$. □

Remark 29. As a slight optimization, we can replace $(t - k)^{\lceil 1/\alpha \rceil} k^{1/\alpha}$ with $k^{\lceil 1/\alpha \rceil} (t - k)^{1/\alpha}$ when both μ and its complement μ^{comp} are α-fractionally-log concave, by down-sampling from $\mu_{S_0 \cup T}^{\text{comp}}$ then output the complement as S_1, where $\mu^{\text{comp}} : \binom{[n]}{n-k} \to \mathbb{R}_{\geq 0}$ is the complement of μ, defined by $\mu^{\text{comp}}([n] \setminus S) = \mu(S) \forall S \in \binom{[n]}{k}$. In all important instances of α-fractionally-log-concavity, $\frac{1}{\alpha} \in \mathbb{N}$ and this optimization is unnecessary. The bound on total runtime can be simplified into $O(n^{1/\alpha - 1} \text{poly}(k, \log \frac{1}{\epsilon}))$.

3.3 Advantage Over Rejection Sampling

While we use a similar intermediate sampling framework as [AD20], our novel analysis of Markov chain intermediate sampling improves the runtime and applies to wider families of distributions. In order to fully understand the advantages realized by our intermediate sampling framework, we first need an overview of a rejection sampling-based implementation of intermediate sampling [Der19], which inspired the analysis of [AD20]. We then provide an example of $\frac{1}{2}$-log-concave distributions where Markov chain intermediate sampling succeeds using a smaller intermediate sample size than what is required for rejection sampling.

Let $S_0 \in \text{supp}(\mu)$. One step of rejection sampling is given by:

1. Sample $T \sim \binom{[n] \setminus S_0}{t-k}$.

2. Accept the set $S_0 \cup T$ with probability

$$\frac{\mu(S_0 \cup T)}{\max_{T' \in \binom{[n] \setminus S_0}{t-k}} \mu(S_0 \cup T')}.$$

3. Downsample $S_1 \sim \mu_{S_0 \cup T}$.

The key difference between rejection sampling and our Markov chain intermediate sampling algorithm is the rejection step, which is necessary if we want our chain to mix to the correct stationary distribution μ. In order to have a sufficiently large acceptance probability, and assuming μ is isotropic, we require that for all T,

$$\mu(T) \leq \left(\frac{t}{n} \right)^k \cdot (1 + \epsilon)^k.$$

15

Here, ϵ is a parameter related to the guarantee on $\|P(S_0, \cdot) - \mu\|_{\mathrm{TV}}$. Using this bound, we can ensure that the expected acceptance probability is $1 - O(\epsilon k)$.

This inequality describes a "worst-case" condition on T. This "worst-case" type analysis originated from earlier works that introduced intermediate sampling for Determinantal Point Processes [Der19]. The proof of our worst-case inequality on $\mu(T)$ relies heavily on the fact that the KL divergence between a log-concave distribution μ and an arbitrary distribution ν contracts by a precise amount when applying the down operator $D_{k \to m}$:

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k \to \ell} \parallel \mu D_{k \to \ell}) \le \frac{\ell}{k} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

This contraction is well-known for log-concave distributions [CGM19], but does not hold with the factor ℓ/k for α-fractionally-log-concave distributions. On the other hand, the inequality we need to show (from the proof Lemma 26) is "average-case" in nature, and when μ is isotropic, it takes the form:

$$\mathbb{E}_{T \sim \binom{[n]}{t}}[\mu(T)] \le \left(\frac{t}{n}\right)^k \cdot \frac{1}{1 - \epsilon}.$$

To concretely illustrate the advantage of Markov chain intermediate sampling, let us consider an example where the worst-case inequality fails to hold. Suppose that $k = 2$, n is even, and μ samples a set from $\{1, \frac{n}{2} + 1\}, \{2, \frac{n}{2} + 2\}, \ldots$ uniformly at random, so that $\mu(\{i, \frac{n}{2} + i\}) = \frac{2}{n}$. This distribution is isotropic, $\frac{1}{2}$-sector stable [Ali+21], and $\frac{1}{2}$-fractionally-log-concave, and yet, according to the worst-case analysis, it does not yield enough acceptance probability when $t = o(n)$. For any set T, we have that

$$\mu(T) \le \frac{t}{2} \cdot \frac{2}{n} = \frac{t}{n}.$$

Equality is achieved by selecting a subset T that contains as many pairs of the form $\{i, \frac{n}{2} + i\}$ as possible, i.e., at least $(t - 1)/2$. Thus, the worst-case analysis would suggest that no non-trivial intermediate sampling is possible for the distribution μ; this is because $t/n \gg (t/n)^2$ for small values of t.

However, our relaxed average-case analysis captures the fact that realistically, not every element of T will be paired up. In fact, we expect only a constant number of pairs when $t = O(\sqrt{n})$, so for this example, we have:

$$\mathbb{E}_{T \sim \binom{[n]}{t}}[\mu(T)] \le C \cdot \frac{2}{n} \le O\left(\frac{1}{n}\right) = O\left(\frac{t^2}{n^2}\right).$$

3.4   Proof of Lemma 26

In this section, we will prove Lemma 26.

Lemma 30. *Let U, V be a sets of size $u, v \le k$ respectively with $U \cap V = \emptyset$. We have:*

$$\left(\frac{t - (u + v)}{n - (u + v)}\right)^u \le \mathbb{P}_{T \in \binom{[n] \setminus V}{t - v}}[U \subseteq T] \le \left(\frac{t - v}{n - v}\right)^u.$$

Proof. Since we are sampling the elements of T uniformly at random from $[n]$,

$$\mathbb{P}_{T \in \binom{[n] \setminus V}{t - v}}[U \subseteq T] = \frac{\binom{n - v - u}{t - v - u}}{\binom{n - v}{t - v}} = \frac{(t - v)(t - v - 1) \cdots (t - v - u + 1)}{(n - v)(n - v - 1) \cdots (n - v - u + 1)} \le \left(\frac{t - v}{n - v}\right)^u.$$

Similarly, we also have:

$$\mathbb{P}_{T\in\binom{[n]\backslash V}{t-v}}[U\subseteq T] \geq \left(\frac{t-v-u+1}{n-v-u+1}\right)^u \geq \left(\frac{t-(u+v)}{n-(u+v)}\right)^u.$$

$\square$

Proof of Lemma 26. Let $R = S_0 \cup S$, and let $r = |S_0 \cup S|$. Note that $|S \backslash S_0| = r - k$ and

$$\mathbb{P}[S_1 = S] = \mathbb{E}_{T\sim\binom{[n]\backslash S_0}{t-k}}\left[\frac{\mu(S)}{\sum_{S'\subseteq(S_0\cup T)}\mu(S')}\,\middle|\, S\subseteq T\right]\cdot\mathbb{P}_{T\sim\binom{[n]\backslash S_0}{t-k}}[(S\backslash S_0)\subseteq T]$$

$$= \mathbb{E}_{T'\sim\binom{[n]\backslash R}{t-r}}\left[\frac{\mu(S)}{\sum_{S'\subseteq(R\cup T')}\mu(S')}\right]\cdot\mathbb{P}_{T\sim\binom{[n]\backslash S_0}{t-k}}[(S\backslash S_0)\subseteq T]$$

$$\geq_{(1)} \frac{\mu(S)}{\mathbb{E}_{T'\sim\binom{[n]\backslash R}{t-r}}[\sum_{S'\subseteq(R\cup T')}\mu(S')]}\cdot\mathbb{P}_{T\sim\binom{[n]\backslash S_0}{t-k}}[(S\backslash S_0)\subseteq T]$$

$$\geq_{(2)} \frac{\mu(S)}{\mathbb{E}_{T'\sim\binom{[n]\backslash R}{t-r}}[\sum_{S'\subseteq(R\cup T')}\mu(S')]}\cdot(\frac{t-r}{n-r})^{r-k}.$$

Inequality (1) is an application of Jensen's inequality to the function $f(x) = \frac{c}{x}$, which is convex when $x > 0$. Inequality (2) use Lemma 30 with $U = S \backslash S_0$ and $V = S_0$.

Now if we bound $\mathbb{E}_{T'\sim\binom{[n]\backslash R}{t-r}}[\sum_{S'\subseteq(R\cup T')}\mu(S')]$ by $(\frac{t-r}{n-r})^{r-k}\frac{1}{1-\epsilon}$ then we are done.

$$\mathbb{E}_{T'\sim\binom{[n]\backslash R}{t-r}}[\sum_{S'\subset(T'\cup R)}\mu(S')] = \sum_{S'\in\binom{[n]}{k}}\mu(S')\cdot\mathbb{P}_{T'\sim\binom{[n]\backslash R}{t-r}}[(S'\backslash R)\subseteq T']] \leq \sum_{S'\in\binom{[n]}{k}}\mu(S')\cdot\left(\frac{t-r}{n-r}\right)^{|S'\backslash R|}.$$

In the very last line, we applied Lemma 30 with $U = |S' \backslash R|$ and $V = R$.

If we set

$$z_i = \begin{cases} (\frac{n-r}{t-r})^{1/\alpha} & \text{if } i \in (S_0 \cup S), \\ 1 & \text{else,} \end{cases}$$

then we can rewrite

$$\sum_{S'\in\binom{[n]}{k}}\mu(S')\cdot\left(\frac{t-r}{n-r}\right)^{|S'\backslash(S_0\cup S)|} = \left(\frac{t-r}{n-r}\right)^k\sum_{S'\in\binom{[n]}{k}}\mu(S')\cdot\left(\frac{n-r}{t-r}\right)^{|S'\cap(S_0\cup S)|}$$

$$= \left(\frac{t-r}{n-r}\right)^k\cdot g_\mu\left(z_1^\alpha,\ldots,z_n^\alpha\right).$$

17

Applying Eq. (1) to $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ and noting that $p_i = \frac{\mathbb{P}_{S\sim\mu}[i\in S]}{k}$, we obtain

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \le \left(\sum_{i=1}^n \frac{\mathbb{P}_{S\sim\mu}[i \in S]}{k} \cdot z_i \right)^{k\alpha}$$

$$\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \le k\alpha \log\left(\sum_{i=1}^n \frac{\mathbb{P}_{S\sim\mu}[i \in S]}{k} \cdot z_i \right)$$

$$\le_{(1)} k\alpha\left(-1 + \sum_{i=1}^n \frac{\mathbb{P}_{S\sim\mu}[i \in S]}{k} \cdot z_i \right)$$

$$=_{(2)} k\alpha\left(\sum_{i=1}^n \frac{\mathbb{P}_{S\sim\mu}[i \in S]}{k} \cdot (z_i - 1) \right)$$

$$= \alpha \sum_{i=1}^n \mathbb{P}_{S\sim\mu}[i \in S] \cdot (z_i - 1),$$

where in (1) we use $\log x \le x - 1$ for $x \in (0, \infty)$ and in (2) we use $\sum_{i=1}^n \frac{\mathbb{P}_{S\sim\mu}[i\in S]}{k} = 1$.

Substitute z_i as specified above into the final inequality, we get

$$\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \le \alpha \sum_{i \in (S_0 \cup S)} \frac{Ck}{n}\left(\frac{n-r}{t-r} \right)^{1/\alpha}$$

$$= \frac{C\alpha kr}{n}\left(\frac{n-r}{t-r} \right)^{1/\alpha}$$

$$\le \frac{2Ck^2}{n}\left(\frac{n-r}{t-r} \right)^{1/\alpha}. \qquad \square$$

4   Lower Bound on Intermediate Sampling

We first show that the dependence of our sparsification analysis on n is optimal. Consider the uniform distribution μ_0 over singletons of a ground set of n/k elements. Any distribution on singletons is log-concave as the generating polynomial is linear and thus log-concave. Now apply the construction of Example 7 with $m = k$ to μ_0 in order to obtain a new distribution μ on $\binom{[n]}{k}$. This distribution is uniform over parts of a particular partition of the ground set $[n]$ into n/k sets $S_1, \ldots, S_{n/k}$. As such, the distribution is also isotropic. Note that this distribution is also $1/k$-entropically independent.

If we sample a uniformly random set T of size t, then the chance that S_i is contained in T can be upperbounded by

$$\binom{n-k}{t-k} / \binom{n}{t} \simeq (t/n)^k.$$

Thus the chance that any of the S_i are contained in T can be upperbounded (via a union bound) by roughly

$$n \cdot (t/n)^k.$$

Thus, as long as $t \ll n^{1-1/k}$, the above is negligible. Without having any S_i in the support with high probability, we obviously cannot faithfully produce a sample of μ from subsets of T.

Next we construct an example showing that even higher-order marginals cannot remove the dependence on n for entropically independent distributions (in sharp contrast with Conjecture 5). Our construction is based on Reed-Solomon codes.

Lemma 31. *Let q be a prime number and $\mathbb{F}_q$ the finite field of size q. Fix k points $\{x_1, \ldots, x_k\} \subseteq \mathbb{F}_q$ where k is a constant and choose a set of k random permutations from $\mathbb{F}_q$ to $\mathbb{F}_q$ and call them $\pi_1, \ldots, \pi_k$. Let $\mu : \binom{\Omega}{k} \to \mathbb{R}_{\geq 0}$ be the uniform distribution over sets $\{(x_i, y_i) \mid i \in [k]\}$ s.t. $p(x_i) = \pi_i(y_i)$ for some polynomial g of $\deg(p) \leq d$. Note that the ground set Ω is $\{x_1, \ldots, x_k\} \times \mathbb{F}_q$. Then*

1. *μ satisfies $(1/\alpha)$-entropic independence with $\alpha = \frac{d+1}{k}$.*

2. *Any domain sparsification scheme to sample from μ requires $t = \tilde{\Omega}(n^{1-\alpha})$, even when we are allowed to sample higher order marginals.*

Proof. The distribution $\mu D_{k \to (d+1)}$ is uniform over $\{(x_j, y_j) : j \in J \subseteq [k], |J| = (d+1)\}$, because for any such set, there exists a unique polynomial p of degree at most d such that $p(x_j) = \pi_j(y_j)$ for all $j \in J$. Thus, high-order marginals, up to order $d + 1$, are independent of the choice of permutations $\pi_1, \ldots, \pi_k$.

The support of $\mu D_{k \to (d+1)}$ forms the basis of a partition matroid: for each $x = x_i$, we have a block consisting of all points $\{(x, y) : y \in \mathbb{F}_q\}$, and for each set in the support of $\mu D_{k \to (d+1)}$, we have at most one element per block. Since we have a uniform distribution over matroid bases, $\mu D_{k \to (d+1)}$ is log-concave, and thus it satisfies 1-entropic independence. We use this to upper bound $\mathcal{D}_{\mathrm{KL}}(\nu D_{k \to 1} \| \mu D_{k \to 1})$, and from here, conclude $\frac{d+1}{k}$-entropic independence of μ:

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k \to 1} \| \mu D_{k \to 1}) = \mathcal{D}_{\mathrm{KL}}((\nu D_{k \to (d+1)}) D_{(d+1) \to 1} \| (\mu D_{k \to (d+1)}) D_{(d+1) \to 1})$$

$$\leq \frac{1}{d+1} \cdot \mathcal{D}_{\mathrm{KL}}(\nu D_{k \to (d+1)} \| \mu D_{k \to (d+1)})$$

$$\leq \frac{1}{d+1} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \| \mu) = \frac{1}{\frac{d+1}{k} \cdot k} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \| \mu).$$

The second line follows from $\mu D_{k \to (d+1)}$ satisfying 1-entropic independence, and the third line comes from the data-processing inequality.

We now prove that for all $t \leq o(n^{1-\alpha})$, no domain sparsification scheme exists, even with access to higher order marginals. For $d' \leq d + 1$, the distribution $\mu D_{k \to d'}$ is uniform over the size d' independent sets of the partition matroid defined above. One consequence of the independence of high-order marginals from the choice of permutations $\pi_1, \ldots, \pi_k$ is that the higher order marginals do not provide any information about the identity of the distribution μ.

Suppose that we choose our sample in domain sparsification from a distribution whose ground set is the sparse subset T. We want an upper bound on the probability (over the choice of permutations) that T contains some $S \in \operatorname{supp}(\mu)$.

In order for T to contain a valid S, there must be some subset in $S \in \binom{T}{k}$ associated to a degree $\leq d$ polynomial p satisfying $p(x_i) = \pi_i(y_i)$. However it is easy to see that for any particular set S the probability (over the choice of permutations) that S is in the support of μ is $\simeq 1/q^{k-d-1}$.

We can upper bound $\mathbb{P}[S \subseteq T \text{ for some } S \in \operatorname{supp}(\mu)]$ by a union bound as follows:

$$\binom{t}{k} \cdot \frac{1}{q^{k-d-1}} \leq \frac{t^k}{q^{k-d-1}}.$$

This implies that for any $t \leq o(q^{(k-d-1)/k}) \leq o(n^{(k-d-1)/k})$, the probability of containing a set in the support is negligible. Note that we have $\alpha = \frac{d+1}{k}$, so $1 - \alpha = \frac{k-d-1}{k}$, which completes the lower bound. $\qquad\square$

5 Missing Proofs

In this section we prove Propositions 19 and 20.

Proof of Proposition 19. We have a characterization of entropic independence [Ana+21] as

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{1/k\alpha} \leq \sum_{i \in S} p_i z_i,$$

where $p_i = \mathbb{P}_{S \sim \mu}[i \in S]/k$ are the normalized marginals of μ.

Subdivision replaces the generating polynomial by a new polynomial, where z_i is replaced by an average of t_i new variables $z_i^{(1)}, \ldots, z_i^{(t_i)}$. It is easy to see that the normalized marginal of each of the new variables is simply p_i/t_i. Thus z_i in the r.h.s. of the above expression gets replaced by the same average of the t_i new variables. Plugging in $(z_i^{(1)} + \cdots + z_i^{(t_i)})/t_i$ for z_i in the above inequality we get

$$g_\mu\left(\left\{\left(\frac{z_i^{(1)} + \cdots + z_i^{(t_i)}}{t_i}\right)^\alpha\right\}_{i=1}^n\right)^{1/k\alpha} \leq \sum_{i=1}^n \sum_{j=1}^{t_i} p_i/t_i z_i^{(j)}.$$

We can apply the inequality

$$\frac{x_1^\alpha + \cdots + x_t^\alpha}{t} \leq \left(\frac{x_1 + \cdots + x_t}{t}\right)^\alpha,$$

to each set of variables $\{x_1, \ldots, x_t\} = \{z_i^{(1)}, \ldots, z_i^{(t_i)}\}$ to get

$$g_{\mu'}\left((z_1^{(1)})^\alpha, \ldots, (z_n^{(t_n)})^\alpha\right)^{1/k\alpha} \leq g_\mu\left(\left\{\left(\frac{z_i^{(1)} + \cdots + z_i^{(t_i)}}{t_i}\right)^\alpha\right\}_{i=1}^n\right)^{1/k\alpha}.$$

Combining the above we get

$$g_{\mu'}\left((z_1^{(1)})^\alpha, \ldots, (z_n^{(t_n)})^\alpha\right)^{1/k\alpha} \leq \sum_{i=1}^n \sum_{j=1}^{t_i} p_i/t_i z_i^{(j)},$$

which is precisely entropic independence for μ'.

We remark that all of this proof also works for non-uniform subdivisions: if instead of replacing z_i by the uniform average of $z_i^{(1)}, \ldots, z_i^{(t_i)}$, we replace it by a weighted average, the resulting polynomial/distribution still remains $1/\alpha$-entropically independent. $\qquad\square$

To prove Proposition 20, we observe that fractional log-concavity is equivalent to entropic independence under arbitrary external fields [Ana+21].

Proof of Proposition 20. Suppose that we apply an external field $\lambda_i^{(j)}$ to the distribution μ'. We claim that this is the result of two operations: an external field being applied to μ (which preserves fractional-log-concavity and hence implies entropic independence), followed by a non-uniform subdivision (which by the proof of Proposition 19 preserves entropic indpendence). The external field applied to μ is given by multiplying each z_i by

$$\lambda_i^{(1)} + \cdots + \lambda_i^{(t_i)}.$$

The non-uniform subdivision is given by replacing each z_i with

$$\frac{\lambda_i^{(1)} z_i^{(1)} + \cdots + \lambda_i^{(t_i)} z_i^{(t_i)}}{\lambda_i^{(1)} + \cdots + \lambda_i^{(t_i)}}. \qquad \square$$

6 Conclusion

A natural direction to extend our research is to better understand the trade-offs between the run-time of approximate sampling algorithms and how much information we are given in the form of marginals for α-fractionally-log-concave distributions. Recall that α-fractional-log-concavity implies $(1/\alpha)$-entropic independence, but still captures many interesting distributions (see Section 1.1), so while we exhibit a lower bound against domain sparsification for $(1/\alpha)$-entropically independent distributions, even with access to higher-order marginals, the stronger assumption of fractional log-concavity may avoid this barrier.

For a k-uniform distribution, the entire distribution itself is indeed captured by the order-k marginals. So if one is to believe that α-fractionally log-concave distributions behave like distributions over $(1/\alpha)$-sized sets, does that mean order-$(1/\alpha)$ marginals are sufficient to get rid of all dependence on n in domain sparsification? For $\alpha = 1$, the answer is affirmative as was shown by Anari and Dereziński [AD20]. A fascinating open question is whether we can extend this to fractionally log-concave polynomials as suggested by Conjecture 5.

References

[AD20]   Nima Anari and Michał Dereziński. "Isotropy and Log-Concave Polynomials: Accelerated Sampling and High-Precision Counting of Matroid Bases". In: *Proceedings of the 61st Annual Symposium on Foundations of Computer Science*. 2020.

[AL20]   Vedat Levi Alev and Lap Chi Lau. "Improved analysis of higher order random walks and applications". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. 2020, pp. 1198–1211.

[Ali+21]   Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. "Fractionally Log-Concave and Sector-Stable Polynomials: Counting Planar Matchings and More". In: *arXiv preprint arXiv:2102.02708* (2021).

[ALO20]   Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model". In: *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, Nov. 2020.

[ALO21]   Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. "A Matrix Trickle-Down Theorem on Simplicial Complexes and Applications to Sampling Colorings". In: *arXiv preprint arXiv:2106.03845* (2021).

[Ana+19]   Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-Concave Polynomials II: High-Dimensional Walks and an FPRAS for Counting Bases of a Matroid". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. ACM, June 2019.

[Ana+20]   Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant, and Thuy Duong Vuong. "Log-Concave Polynomials IV: Approximate Exchange, Tight Mixing Times, and Near-Optimal Sampling of Forests". In: *arXiv preprint arXiv:2004.07220* (2020).

[Ana+21]   Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic Independence I: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Distributions and High-Temperature Ising Models". In: *arXiv (to appear)* (2021).

[AOV18]   Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids". In: *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2018, pp. 35–46.

[AV21]   Nima Anari and Thuy-Duong Vuong. "Simple and Near-Optimal MAP Inference for Nonsymmetric DPPs". In: *CoRR* abs/2102.05347 (2021). arXiv: 2102.05347.

[Bar+17]   Bardenet, Rémi, Lavancier, Frédéric, Mary, Xavier, and Vasseur, Aurélien. "On a few statistical applications of determinantal point processes". In: *ESAIM: Procs* 60 (2017), pp. 180–202. DOI: 10.1051/proc/201760180.

[BBL09]   Julius Borcea, Petter Brändén, and Thomas Liggett. "Negative dependence and the geometry of polynomials". In: *Journal of the American Mathematical Society* 22.2 (2009), pp. 521–567.

[BH19]   Petter Brändén and June Huh. "Lorentzian polynomials". In: *arXiv preprint arXiv:1902.03719* (2019).

[Bla+21]   Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. "On mixing of Markov chains: Coupling, spectral independence, and entropy factorization". In: *arXiv preprint arXiv:2103.07459* (2021).

[BT06]   Sergey G Bobkov and Prasad Tetali. "Modified logarithmic Sobolev inequalities in discrete settings". In: *Journal of Theoretical Probability* 19.2 (2006), pp. 289–336.

[CDV20]   Daniele Calandriello, Michał Dereziński, and Michal Valko. "Sampling from a k-DPP without looking at all items". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020.

[Cel+16]   L Elisa Celis, Amit Deshpande, Tarun Kathuria, Damian Straszak, and Nisheeth K Vishnoi. "On the complexity of constrained determinantal point processes". In: *arXiv preprint arXiv:1608.00554* (2016).

[CGM19]   Mary Cryan, Heng Guo, and Giorgos Mousa. "Modified log-Sobolev inequalities for strongly log-concave distributions". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2019, pp. 1358–1370.

[Che+21a]   Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. "Rapid mixing of Glauber dynamics via spectral independence for all degrees". In: *arXiv preprint arXiv:2105.15005* (2021).

[Che+21b]   Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Rapid mixing for colorings via spectral independence". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 1548–1557.

[CLV20]   Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Rapid mixing of Glauber dynamics up to uniqueness via contraction". In: *arXiv preprint arXiv:2004.09083* (2020).

[DCV19]   Michał Dereziński, Daniele Calandriello, and Michal Valko. "Exact sampling of deter-
minantal point processes with sublinear time preprocessing". In: *Advances in Neural
Information Processing Systems 32*. 2019, pp. 11542–11554.

[Der+19] Michał Dereziński, Kenneth L. Clarkson, Michael W. Mahoney, and Manfred K. War-
muth. "Minimax experimental design: Bridging the gap between statistical and worst-
case approaches to least squares regression". In: *Proceedings of the Thirty-Second Confer-
ence on Learning Theory*. Ed. by Alina Beygelzimer and Daniel Hsu. Vol. 99. Proceedings
of Machine Learning Research. Phoenix, USA, 2019, pp. 1050–1069.

[Der+20] Michał Dereziński, Burak Bartan, Mert Pilanci, and Michael W Mahoney. "Debias-
ing Distributed Second Order Optimization with Surrogate Sketching and Scaled
Regularization". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020.

[Der19]   Michał Dereziński. "Fast determinantal point processes via distortion-free intermediate
sampling". In: *Proceedings of the 32nd Conference on Learning Theory*. 2019.

[Des+06] Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang. "Matrix
Approximation and Projective Clustering via Volume Sampling". In: *Proceedings of the
Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*. Miami, FL, USA, Jan.
2006, pp. 1117–1126.

[DKM20]   Michał Dereziński, Rajiv Khanna, and Michael W Mahoney. "Improved guarantees
and a multiple-descent curve for Column Subset Selection and the Nyström method".
In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 4953–4964.

[DKS10]   Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. "A sparse johnson: Lindenstrauss
transform". In: *Proceedings of the forty-second ACM symposium on Theory of computing*.
2010, pp. 341–350.

[DM21]   Michał Dereziński and Michael W Mahoney. "Determinantal Point Processes in Ran-
domized Numerical Linear Algebra". In: *Notices of the American Mathematical Society*
68.1 (2021), pp. 34–45.

[DR10]   Amit Deshpande and Luis Rademacher. "Efficient Volume Sampling for Row/Column
Subset Selection". In: *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations
of Computer Science*. Las Vegas, USA, Oct. 2010, pp. 329–338.

[DW17]   Michał Dereziński and Manfred K. Warmuth. "Unbiased estimates for linear regression
via volume sampling". In: *Advances in Neural Information Processing Systems 30*. Long
Beach, CA, USA, 2017, pp. 3087–3096.

[DWH18]   Michał Dereziński, Manfred K. Warmuth, and Daniel Hsu. "Leveraged volume sam-
pling for linear regression". In: *Advances in Neural Information Processing Systems 31*.
2018, pp. 2510–2519.

[DWH19]   Michał Dereziński, Manfred K Warmuth, and Daniel Hsu. "Unbiased estimators for
random design regression". In: *arXiv preprint arXiv:1907.03411* (2019).

[EV19]   David Eppstein and Vijay V Vazirani. "NC Algorithms for Computing a Perfect Match-
ing, the Number of Perfect Matchings, and a Maximum Flow in One-Crossing-Minor-
Free Graphs". In: *The 31st ACM Symposium on Parallelism in Algorithms and Architectures*.
2019, pp. 23–30.

[Fen+21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. "Rapid mixing from spectral
independence beyond the Boolean domain". In: *Proceedings of the 2021 ACM-SIAM
Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 1558–1577.

[Gar+19] Mike Gartrell, Victor-Emmanuel Brunel, Elvis Dohmatob, and Syrine Krichene. "Learn-
ing nonsymmetric determinantal point processes". In: *arXiv preprint arXiv:1905.12962*
(2019).

[Gar+20]   Mike Gartrell, Insu Han, Elvis Dohmatob, Jennifer Gillenwater, and Victor-Emmanuel Brunel. "Scalable Learning and MAP Inference for Nonsymmetric Determinantal Point Processes". In: *CoRR* abs/2006.09862 (2020). arXiv: 2006.09862.

[GL99]   Anna Galluccio and Martin Loebl. "On the theory of Pfaffian orientations. I. Perfect matchings and permanents". In: *the electronic journal of combinatorics* (1999), R6–R6.

[GM20]   Heng Guo and Giorgos Mousa. "Local-to-Global Contraction in Simplicial Complexes". In: *arXiv preprint arXiv:2012.14317* (2020).

[GS12]   Venkatesan Guruswami and Ali K. Sinop. "Optimal Column-based Low-rank Matrix Reconstruction". In: *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*. Kyoto, Japan, Jan. 2012, pp. 1207–1214.

[Gué83]   A Guénoche. "Random spanning tree". In: *Journal of Algorithms* 4.3 (1983), pp. 214–220. ISSN: 0196-6774. DOI: https://doi.org/10.1016/0196-6774(83)90022-6.

[Gur09]   Leonid Gurvits. "On multivariate Newton-like inequalities". In: *Advances in combinatorial mathematics*. Springer, 2009, pp. 61–78.

[Hou+06] J. Ben Hough, Manjunath Krishnapur, Yuval Peres, and Bálint Virág. "Determinantal processes and independence". In: *Probability surveys* 3 (2006), pp. 206–229.

[JPV21]   Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. "Spectral independence, coupling with the stationary distribution, and the spectral gap of the Glauber dynamics". In: *arXiv preprint arXiv:2105.01201* (2021).

[JSV04]   Mark Jerrum, Alistair Sinclair, and Eric Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *Journal of the ACM (JACM)* 51.4 (2004), pp. 671–697.

[JVV86]   Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoretical computer science* 43 (1986), pp. 169–188.

[Kas67]   Pieter Kasteleyn. "Graph theory and crystal physics". In: *Graph theory and theoretical physics* (1967), pp. 43–110.

[KT11]   Alex Kulesza and Ben Taskar. "k-DPPs: Fixed-Size Determinantal Point Processes". In: *Proceedings of the 28th International Conference on Machine Learning*. June 2011, pp. 1193–1200.

[KT12]   Alex Kulesza and Ben Taskar. *Determinantal Point Processes for Machine Learning*. Hanover, MA, USA: Now Publishers Inc., 2012.

[Liu21]   Kuikui Liu. "From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk". In: *arXiv preprint arXiv:2103.11609* (2021).

[LP17]   David A Levin and Yuval Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[Mac75]   Odile Macchi. "The Coincidence Approach to Stochastic Point Processes". In: *Advances in Applied Probability* 7.1 (1975), pp. 83–122. ISSN: 00018678.

[MDK20]   Mojmir Mutny, Michał Dereziński, and Andreas Krause. "Convergence Analysis of Block Coordinate Algorithms with Determinantal Sampling". In: *International Conference on Artificial Intelligence and Statistics*. 2020, pp. 3110–3120.

[NST19]   Aleksandar Nikolov, Mohit Singh, and Uthaipon Tao Tantipongpipat. "Proportional Volume Sampling and Approximation Algorithms for A-optimal Design". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2019, pp. 1369–1386.

[ST11]   Daniel A Spielman and Shang-Hua Teng. "Spectral sparsification of graphs". In: *SIAM Journal on Computing* 40.4 (2011), pp. 981–1025.

[Suz+06]   Kazuhiro Suzuki, Dongvu Tonien, Kaoru Kurosawa, and Koji Toyota. "Birthday Paradox for Multi-collisions". In: *Information Security and Cryptology – ICISC 2006*. Ed. by Min Surp Rhee and Byoungcheon Lee. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 29–40. ISBN: 978-3-540-49114-9.