Nearly Tight Spectral Sparsification of Directed Hypergraphs

Kazusato Oko
Department of Mathematical Informatics, The University of Tokyo, Japan
Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan

Shinsaku Sakaue
Department of Mathematical Informatics, The University of Tokyo, Japan

Shin-ichi Tanigawa
Department of Mathematical Informatics, The University of Tokyo, Japan

Abstract

Spectral hypergraph sparsification, an attempt to extend well-known spectral graph sparsification to hypergraphs, has been extensively studied over the past few years. For undirected hypergraphs, Kapralov, Krauthgamer, Tardos, and Yoshida (2022) have proved an ε-spectral sparsifier of the optimal $O^*(n)$ size, where n is the number of vertices and O^* suppresses the ε^{-1} and $\log n$ factors. For directed hypergraphs, however, the optimal sparsifier size has not been known. Our main contribution is the first algorithm that constructs an $O^*(n^2)$-size ε-spectral sparsifier for a weighted directed hypergraph. Our result is optimal up to the ε^{-1} and $\log n$ factors since there is a lower bound of $\Omega(n^2)$ even for directed graphs. We also show the first non-trivial lower bound of $\Omega(n^2/\varepsilon)$ for general directed hypergraphs. The basic idea of our algorithm is borrowed from the spanner-based sparsification for ordinary graphs by Koutis and Xu (2016). Their iterative sampling approach is indeed useful for designing sparsification algorithms in various circumstances. To demonstrate this, we also present a similar iterative sampling algorithm for undirected hypergraphs that attains one of the best size bounds, enjoys parallel implementation, and can be transformed to be fault-tolerant.

2012 ACM Subject Classification Theory of computation → Sparsification and spanners

Keywords and phrases Spectral sparsification, (Directed) hypergraphs, Iterative sampling

Digital Object Identifier 10.4230/LIPIcs.ICALP.2023.94

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: https://arxiv.org/abs/2204.02537

Funding This work was supported by JST PRESTO Grant Number JPMJPR2126, JST ERATO Grant Number JPMJER1903, and JSPS KAKENHI Grant Number 20H05961.

1 Introduction

Graph sparsification is a fundamental idea for developing efficient algorithms and data structures. One of the earliest developments in this context is a cut sparsifier due to Benczúr and Karger [4], which approximately keeps the size of cuts (by adjusting edge weights). Spielman and Teng [24] introduced a generalized notion called a spectral sparsifier, which approximately preserves the spectrum of the Laplacian matrix of a given graph. Since this seminal work, spectral sparsification of graphs has been extensively studied and used in many applications. See, e.g., [27, 26, 23] for more details on spectral graph sparsification.

This paper studies spectral sparsification of undirected/directed hypergraphs. A hypergraph is a standard tool for generalizing graph-theoretic arguments in a set-theoretic setting, and extending a theory for graphs to hypergraphs is a common theoretical interest.

Besides, many hypergraph-based methods [12, 28, 25, 29, 31] have recently been attracting much attention as extensions of graph-based methods, which also increases the demand for advancing the theory of spectral hypergraph sparsification.

An *undirected hypergraph* is defined by a tuple $H = (V, F, z)$, where V is a finite vertex set, F is a set of subsets of V, and $z\colon F \to \mathbb{R}_+$. Each element in F is called a hyperedge and $z_f := z(f)$ is called the weight of $f \in F$ in H. The Laplacian $L_H\colon \mathbb{R}^V \to \mathbb{R}^V$ of H is defined as a nonlinear operator such that

$$x^\top L_H(x) = \sum_{f \in F} z_f \max_{u,v \in f} (x_u - x_v)^2 \quad \text{for all } x \in \mathbb{R}^V.$$

If x is restricted to $\{0,1\}^V$, $x^\top L_H(x)$ represents the cut function of H. In this sense, the above definition gives a proper extension of the ordinary graph Laplacian. (Here, $x^\top L_H(x)$ is an abuse of notation since $L_H(x)$ is not defined uniquely; nevertheless, this notation is widely used in analogy to the case of ordinary graphs.)

A *directed hypergraph* $H = (V, F, z)$ consists of a finite set V, a set F of hyperarcs, and $z\colon F \ni f \mapsto z_f \in \mathbb{R}_+$, where each *hyperarc* $f \in F$ is a pair $(t(f), h(f))$ of non-empty subsets of V, called the *tail* and the *head* (which may not be disjoint). The Laplacian $L_H\colon \mathbb{R}^V \to \mathbb{R}^V$ of H is defined as a nonlinear operator such that

$$x^\top L_H(x) = \sum_{f \in F} z_f \max_{u \in t(f), v \in h(f)} (x_u - x_v)_+^2 \quad \text{for all } x \in \mathbb{R}^V,$$

where $(\cdot)_+ = \max\{\cdot, 0\}$ (and $(\cdot)_+^2 = (\max\{\cdot, 0\})^2$). If $x \in \{0,1\}^V$, the definition of directed hypergraph Laplacian L_H also captures the cut function of H. Importantly, cut functions of directed hypergraphs can represent a large class of submodular functions [10].[1] Directed hypergraphs are also useful for modeling higher-order directional relations that appear in, e.g., propositional logic [11] and causal inference [14], which have constituted a motivation for studying spectral properties of directed hypergraphs [7].

Given an undirected/directed hypergraph $H = (V, F, z)$ and $\varepsilon \in (0,1)$, a hypergraph $\tilde{H} = (V, \tilde{F}, \tilde{z})$ is called an *ε-spectral sparsifier* of H if it satisfies $\tilde{F} \subseteq F$ and

$$(1-\varepsilon) x^\top L_H(x) \le x^\top L_{\tilde{H}}(x) \le (1+\varepsilon) x^\top L_H(x) \quad \text{for all } x \in \mathbb{R}^V.$$

One of the big motivations for studying spectral sparsification of directed hypergraphs comes from the connection to the representation of submodular functions. Since such a cut-function representation uses $\Omega(2^{|V|})$ hyperarcs in general, a spectral sparsifier of a directed hypergraph can serve as a compact approximate representation (see the full version [20] for more details).

Soma and Yoshida [22] initiated the study of spectral hypergraph sparsification and gave an algorithm for constructing an ε-spectral sparsifier with $O(n^3 \log n/\varepsilon^2)$ hyperedges, where n is the number of vertices. Unlike ordinary graphs, the hypergraph size can be as large as 2^n (and 4^n if directed). Thus, obtaining a polynomial bound is already nontrivial. For undirected hypergraphs, the result by Soma and Yoshida [22] has been improved to $\tilde{O}(nr^3/\varepsilon^2)$ [3] and to $\tilde{O}(nr/\varepsilon^{O(1)})$ [15],[2] where r denotes the maximum size of a hyperedge in the input hypergraph H and is called the *rank* of H. Kapralov et al. [16] has removed the dependence on r and obtained a nearly linear bound of $\tilde{O}(n/\varepsilon^4)$. Very recently, an improved bound of $\tilde{O}(n/\varepsilon^2)$ has been shown in [13, 18] (concurrently to our work). This upper bound is nearly tight since the $\Omega(n/\varepsilon^2)$ lower bound applies even to ordinary graphs [2, 6].

---

[1] In fact, any set function can be represented as a cut function of some directed hypergraph if negative weights are allowed [10].

[2] We use $\tilde{O}$ to hide poly$(\log(n/\varepsilon))$ factors.

Table 1 Bounds on sparsification of directed hypergraphs. In the time complexity, additive poly$(n, 1/\varepsilon)$ terms are omitted. Note that Kapralov et al. [15] assume the unweighted case.

Method	Cut/Spectral	Bound	Time complexity
Soma and Yoshida [22]	Spectral	$O(n^3 \log n/\varepsilon^2)$	$O(mr^2)$
Kapralov et al. [15]	Spectral	$O(n^2 r^3 \log^2 n/\varepsilon^2)$	$O(mr^2)$
Rafiey and Yoshida [21]	Cut	$O(n^2 r^2/\varepsilon^2)$	$O(m2^r)$
This paper	Spectral	$O(n^2 \log^3(n/\varepsilon)/\varepsilon^2)$	$O(mr^2)$

As for spectral sparsification of directed hypergraphs, Soma and Yoshida [22] showed that their algorithm is also applicable, and hence the $O(n^3 \log n/\varepsilon^2)$ bound also holds for directed hypergraphs. Later, Kapralov et al. [15] gave an $O(n^2 r^3 \log^2 n/\varepsilon^2)$ bound for unweighted directed hypergraphs, where the rank r is defined by $r = \max_{f \in F}\{|h(f)| + |t(f)|\}$ in the directed case. Recently, for the case of cut sparsification, Rafiey and Yoshida [21] obtained sparsifiers with $O(n^2 r^2/\varepsilon^2)$ hyperarcs.[3] See Table 1. On the other hand, a well-known $\Omega(n^2)$ lower bound for directed graphs [9] is valid for directed hypergraphs. Therefore, a central open question in this context is: *can we obtain an upper bound of $\tilde{O}(n^2/\varepsilon^{O(1)})$ that has no dependence on the rank r?*

1.1 Main Results and Idea

Our main contribution is the first algorithm that constructs an ε-spectral sparsifier with $\tilde{O}(n^2/\varepsilon^2)$ hyperarcs for a directed hypergraph, thus settling the aforementioned question.

▶ **Theorem 1.** *Let $H = (V, F, z)$ be a directed hypergraph with n vertices. For any $\varepsilon \in (0, 1)$, our algorithm (shown in Algorithm 3) returns an ε-spectral sparsifier $\tilde{H} = (V, \tilde{F}, \tilde{z})$ of H such that $|\tilde{F}| = O\left(\frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon}\right)$ with probability at least $1 - O\left(\frac{1}{n}\right)$. Its time complexity is $O(mr^2)$ with probability at least $1 - O\left(\frac{1}{n}\right)$, where $m = |F|$ and r is the rank of H.*

This bound improves the previous results and is optimal up to the ε^{-1} and logarithmic factors due to the presence of the $\Omega(n^2)$ lower bound for directed graphs. We prove Theorem 1 in Section 4 by providing a concrete algorithm and its analysis.

A natural next question would be whether the ε^{-1} term can be deleted. Our new lower bound shows that the ε^{-1} term is indeed necessary, and an ε-spectral sparsifier of size $O(n^2)$ may not exist in general, thus complementing our upper bound.

▶ **Theorem 2.** *Let $n \in \mathbb{Z}_{>0}$. For any $\varepsilon \in \left(\frac{1}{4n}, 1\right)$, there is a directed hypergraph $H = (V, F, z)$ with $2n$ vertices, $\Omega\left(\frac{n^2}{\varepsilon}\right)$ hyperarcs, and the rank three that has no sub-hypergraph $\tilde{H} = (V, \tilde{F}, \tilde{z})$ such that $\tilde{F} \subsetneq F$ and $(1-\varepsilon)x^\top L_H(x) \le x^\top L_{\tilde{H}}(x) \le (1+\varepsilon)x^\top L_H(x)$ for all $x \in \{0,1\}^V$.*

This gives a lower bound even for the case of cut sparsification and is the first nontrivial lower bound for sparsification of directed hypergraphs. We give the proof in the full version [20].

The basic idea of our algorithm for Theorem 1 comes from a spanner-based sparsification method for undirected graphs by Koutis and Xu [17], in contrast to the method of [16] for nearly tight sparsification of undirected hypergraphs. The analysis of [16] uses a technique called *weight assignment* [8], which crucially depends on linear algebraic arguments on the

[3] This bound follows from their general result on sparsification of submodular functions.

linear Laplacian of some underlying undirected graph. *Directed* hypergraphs, however, do not have such convenient underlying *undirected* graphs, and hence their idea cannot be utilized. We thus take an alternative route and use the algorithmic framework of Koutis and Xu [17] – iteratively select important edges and sample the remaining edges. Due to its combinatorial nature, we can analyze errors via combinatorial arguments instead of linear algebraic tools. Although our algorithm is as simple as theirs, our analysis for proving Theorem 1 involves novel techniques. Specifically, while building on a recent chaining-based analysis [15, 16], we develop a completely new *discretization scheme* based on a non-trivial combinatorial observation to obtain the optimal upper bound. See Section 3 for an overview of our analysis.

1.2 Additional Results

We also present the following additional results in the full version [20].

Undirected hypergraph sparsification. The iterative sampling approach mentioned above indeed has much potential in hypergraph sparsification. We exhibit its power by presenting a natural extension of the spanner-based algorithm by Koutis and Xu [17] to undirected hypergraphs. The concept of spanners in graphs can be naturally extended to undirected hypergraphs, and accordingly, Koutis and Xu's algorithm can also be extended to undirected hypergraphs. Based on a result by Bansal et al. [3], we show that the resulting algorithm constructs an ε-spectral sparsifier with $O\left(\frac{nr^3}{\varepsilon^2}\log^2 n\right)$ hyperedges, which is nearly optimal if r is constant and matches the bound of [3] (up to a $\log n$ factor). Moreover, our algorithm inherits advantages of the spanner-based approach in that it can be implemented in parallel [17] and can be converted to be fault-tolerant [32], demonstrating that the iterative sampling approach can enjoy various useful extensions.

Application to learning of submodular functions. A notable application of directed hypergraph sparsification due to [22] is agnostic learning of submodular functions. We apply our method to this setting and obtain an $\tilde{O}\left(\frac{n^3}{\varepsilon^4} + \frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right)$ sample complexity bound for agnostic learning of nonnegative hypernetwork-type submodular functions on a ground set of size n, improving the previous $\tilde{O}\left(\frac{n^4}{\varepsilon^4} + \frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right)$ bound in [22]. Note that since the rank r of a hypergraph representing a submodular function can be $O(n)$, eliminating the dependence on r in the sparsifier size (i.e., our improvement from [15]) is crucial in this application. It should be mentioned that this application only requires cut sparsifiers. Nevertheless, since our result gives the first near-optimal bound even on the size of cut sparsifiers of directed hypergraphs, this application serves as a good motivation for our result.

1.3 Related Work

Besides the aforementioned application to agnostic learning of submodular functions, there are many other potential applications that involve the quadratic form $x^\top L_H(x)$ (which is sometimes called the *energy* of hypergraphs), e.g., clustering [25], semi-supervised learning [12, 28, 31, 19], and link prediction [29]. For example, Li et al. [19] use the quadratic form as a smoothness regularizer. Our result on spectral sparsification can be useful when dealing with such regularizers on dense directed hypergraphs.

Cohen et al. [9] studied directed graph sparsification under a different definition of approximation based on *Eulerian scaling*. While their definition is compatible with fast Laplacian solvers, how to extend it to directed hypergraphs seems non-trivial. Our definition is based on a general notion called *submodular transformations* [30] and admits a natural interpretation as a generalization of cut sparsification of directed hypergraphs.

2 Preliminaries

We usually denote a directed hypergraph by $H = (V, F, z)$, the numbers of vertices by n, and the numbers of hyperarcs by m. The Laplacian $L_H : \mathbb{R}^V \to \mathbb{R}^V$ is defined as a nonlinear operator that satisfies $x^\top L_H(x) = \sum_{f \in F} z_f \max_{u \in t(f), v \in h(f)} (x_u - x_v)_+^2$ for all $x \in \mathbb{R}^V$, where $h(f), t(f) \subseteq V$ are the head and the tail of f, respectively. For each $f \in F$, we denote the contribution of f to $x^\top L_H(x)$ by $Q_H^x(f) = z_f \max_{u \in t(f), v \in h(f)} (x_u - x_v)_+^2$, which we call the *energy* of f. Note that $x^\top L_H(x) = \sum_{f \in F} Q_H^x(f)$ holds. For any subset F' of F, we let $Q_H^x(F') = \sum_{f \in F'} Q_H^x(f)$, i.e., the sum of energies over F'. For a hyperarc $f \in F$, we define its *biclique* as an arc set $C(f) = \{ (u, v) \mid u \in t(f), v \in h(f) \}$. For a subset $F' \subseteq F$, we let $C(F') = \bigcup_{f \in F'} C(f)$. Below, we often take $\mathrm{argmax}_{f \in F'} \zeta(f)$ for a function $\zeta : F \to \mathbb{R}$ and a hyperarc subset $F' \subseteq F$. For convenience, we let such argmax (or argmin) operations always return a singleton by using some tie-breaking rule with a pre-defined total order on F. For example, if vertices are labeled by $1, \ldots, n$ and each $f \in F$ is labeled by vertices in f, we may use the lexicographical order on F with respect to the labels. Similarly, we break ties when taking argmax/argmin on any $E' \subseteq V \times V$. We will often use the following Chernoff bound.

▶ **Proposition 3** ([1]). *Let $X_1, X_2, \cdots, X_m$ be independent random variables in the range of $[0, a]$. For any $\delta \in [0, 1]$ and $\mu \geq \mathbb{E}[\sum_{i=1}^m X_i]$, we have*

$$\mathbb{P}\left[\left| \sum_{i=1}^m X_i - \mathbb{E}\left[\sum_{i=1}^m X_i \right] \right| > \delta\mu \right] \leq 2 \exp\left(-\frac{\delta^2 \mu}{3a} \right).$$

3 Technical Overview

Our algorithm is an iterative algorithm whose each step goes as follows: given a hypergraph $H = (V, F, z)$ from a previous iteration, it constructs a set S of heavy hyperarcs, called a *coreset*, which is kept deterministically in this step, and samples the remaining hyperarcs with probability $1/2$, where weights of sampled ones are doubled. This single step yields a hypergraph with fewer hyperarcs, which is taken as input in the next step. We iterate this until a sub-hypergraph of the desired size is obtained. Roughly speaking, the size of the coreset is about $\tilde{O}(n^2/\varepsilon^2)$, and after about $O(\log(m\varepsilon^2/n^2))$ iterations, we obtain a sub-hypergraph of size $\tilde{O}(n^2/\varepsilon^2)$. This algorithmic framework is identical to that of Koutis and Xu [17] for ordinary undirected graph sparsification, which iteratively constructs a bundle of spanners (instead of a coreset) and sample the remaining edges with probability $1/4$.

We then describe how to analyze the sparsification error. Note that if a sub-hypergraph produced in each step is a sparsifier of a hypergraph $H = (V, F, z)$ given from the previous step with a sufficiently large probability, then we can bound the error accumulated over the iterations. Thus, we focus on the analysis of a single step (which is presented in Lemma 6). To bound the sparsification error in $Q_H^x(F) = x^\top L_H(x)$ for all $x \in \mathbb{R}^V$ in each step, we adopted a chaining-type argument [15, 16]; this enables us to derive a desired uniform bound on a continuous domain from a pointwise bound via adaptive scaling of the domain discretization. Here, how to design a discretization scheme crucially affects how sharp the resulting uniform bound is. Therefore, we need to design an appropriate discretization scheme by carefully looking at the structure of directed hypergraphs.

We below sketch our discretization scheme. Inspired by the previous studies [15, 16], we classify hyperarcs $f \in F \setminus S$ based on their energies $Q_H^x(f)$. Here, since the coreset S is always selected, we can exclude it when discussing the following probabilistic arguments. For each $x \in \mathbb{R}^V$, we consider a partition of $F \setminus S$ into F_i^x ($i \in \mathbb{Z}$) such that each F_i^x consists of hyperarcs f with energies $Q_H^x(f) \approx 2^{-i} Q_H^x(F)$. Then, the Chernoff bound offers the following pointwise guarantee for each $x \in \mathbb{R}^V$:

$$\mathbb{P}\left[|Q_{\tilde{H}}^x(\tilde{F}_i^x) - Q_H^x(F_i^x)| \geq \varepsilon Q_H^x(F)\right] \lesssim \exp\left(-\frac{\varepsilon^2 Q_H^x(F)}{2^{-i} Q_H^x(F)}\right) = \exp\left(-\varepsilon^2 2^i\right),$$

where $\tilde{H}$ is a sparsifier obtained from H and $Q_{\tilde{H}}^x(\tilde{F}_i^x)$ denotes the energy of $\tilde{H}$ with hyperarcs restricted to F_i^x. To obtain a desired uniform bound using this inequality, we need to design a discretization scheme that satisfies the following two requirements:

(R1) the discretization error is $O(\varepsilon)$, and

(R2) the number of possible discretized energies is bounded by about $\exp(\varepsilon^2 2^i)$.

Kapralov et al. [15] obtained such a scheme by looking at underlying clique digraphs. By contrast, we obtain a discretization scheme by directly looking at hypergraphs. This strategy enables us to eliminate the extra r^3 factor in their bound, but it also poses a new challenge.

We explain the challenge when designing such a discretization scheme by directly looking at hypergraphs. Once $x \in \mathbb{R}^V$ is fixed, the number of hyperarcs f with $Q_H^x(f) \approx 2^{-i} Q_H^x(F)$ is bounded by about 2^i; on the other hand, we need to prepare at least $\mathrm{poly}(n, 1/\varepsilon)$ possible discretized energies for each f to satisfy requirement (R1). Thus, naive counting implies that the number of total discretized energies for all $f \in F_i^x$ is $(\mathrm{poly}(n, 1/\varepsilon))^{2^i} \approx \exp(\tilde{O}(2^i))$, which is too large to satisfy requirement (R2). To overcome this problem, we need an additional combinatorial idea: we count the number of discretized energies by focusing on the number of possible critical pairs. We say that $(u, v) \in C(F \setminus S)$ is a *critical pair* of f if $(u, v) = \mathrm{argmax}_{u' \in t(f), v' \in h(f)} (x_{u'} - x_{v'})_+^2$ (see also Figure 1b). Suppose that a lot of hyperarcs in F_i^x share a common critical pair for a given $x \in \mathbb{R}^V$, particularly when F_i^x contains as many as 2^i hyperarcs. Then, since the energy of f is determined by the $(x_u - x_v)_+^2$ value of the critical pair (u, v) of f, we may get a sharper bound on the number of discretized energies by defining a discretization scheme based on $(x_u - x_v)_+^2$ values so that hyperarcs with the same critical pairs share the same discretized energies (up to scaling of weights).

To accomplish the counting based on this idea, we use the existence of a coreset kept in each iteration. As we will see shortly from the definition, a λ-coreset $S \subseteq F$ contains λ heaviest hyperarcs for each $(u, v) \in C(F)$ (see also Figure 1a). Roughly speaking, important properties of λ-coresets are as follows:

(P1) $|S| \leq \lambda n^2$,

(P2) for any fixed $x \in \mathbb{R}^V$, many hyperarcs with large energies are included in S, and

(P3) for any fixed $x \in \mathbb{R}^V$, the number of critical pairs of hyperarcs in F_i^x is at most $2^i/\lambda$.[4]

If we set $\lambda = \tilde{O}(\varepsilon^{-2})$, the size of λ-coreset S is $\tilde{O}(n^2/\varepsilon^2)$ by property (P1), which is small enough that the output size decreases geometrically in each iteration until we obtain an $\tilde{O}(n^2/\varepsilon^2)$ size sparsifier. Property (P2) bounds the range of i such that F_i^x is non-empty. Most importantly, property (P3) implies that if we count possible discretized energies over F_i^x, the total number is at most $(\mathrm{poly}(n, 1/\varepsilon))^{2^i/\lambda} \approx \exp\left(\tilde{O}(\varepsilon^2 2^i)\right)$, satisfying requirement (R2).

In summary, once the coreset is selected, we can categorize the remaining hyperarcs in each F_i^x based on a moderate number of critical pairs, which yields a sharp bound on the number of possible discretized energies of the remaining hyperarcs. This is the key idea of our discretization scheme, which, together with the chaining-type argument, provides the desired uniform bound on the sparsification error.

---

[4] For ease of exposition, λ is used differently from Section 4. In Section 4.2, we will instead define F_i^x based on $2^{-i} Q_H^x(F)/\lambda$ values and, accordingly, bound the number of critical pairs by 2^i (Lemma 11).

4 Spectral Sparsification of Directed Hypergraphs

We prove Theorem 1 by presenting a concrete algorithm. Section 4.1 presents our algorithm and key lemmas. Section 4.2 focuses on the analysis of a single iteration, and Section 4.3 bounds the overall sparsification error and the resulting sparsifier size, thus proving Theorem 1. Section 4.4 shows the $O(mr^2)$ time complexity bound of our algorithm.

4.1 Algorithm Description

Our algorithm consists of CoresetFinder (Algorithm 1), DH-OneStep (Algorithm 2), and DH-Sparsify (Algorithm 3). DH-Sparsify iteratively calls DH-OneStep, which uses CoresetFinder as a subroutine. We below explain them one by one.

Algorithm 1 CoresetFinder(H, λ): greedy algorithm for coreset construction.

Input: $H = (V, F, z)$ and $\lambda > 0$
Output: $S \subseteq F$

$S \leftarrow \emptyset$ and $S^{uv} \leftarrow \emptyset$ for each $(u, v) \in C(F)$
$A^{uv} \leftarrow \{ f \in F \mid (u, v) \in C(f) \}$ for each $(u, v) \in C(F)$
for each $(u, v) \in C(F)$ **:**
  if $|A^{uv} \setminus S| \geq \lambda$ **:**
    Find the first λ heaviest hyperarcs $f_1^{uv}, f_2^{uv}, \cdots, f_\lambda^{uv} \in A^{uv} \setminus S$
    Add $f_1^{uv}, f_2^{uv}, \cdots, f_\lambda^{uv}$ to S^{uv}
  else
    $S^{uv} \leftarrow A^{uv} \setminus S$
  $S \leftarrow S \cup S^{uv}$
return S

Algorithm 2 DH-OneStep(H, λ): sampling algorithm called in each iteration in Algorithm 3.

Input: $H = (V, F, z)$ and $\lambda > 0$
Output: $\tilde{H} = (V, \tilde{F}, \tilde{z})$

$S \leftarrow$ CoresetFinder(H, λ)
$\tilde{F} \leftarrow S$ and $\tilde{z}_f \leftarrow z_f$ for $f \in S$
for each $f \in F \setminus S$ **:**
  With probability $\frac{1}{2}$, add f to $\tilde{F}$ and set $\tilde{z}_f \leftarrow 2z_f$
return $\tilde{H} = (V, \tilde{F}, \tilde{z})$

The first building block of our algorithm is CoresetFinder(H, λ) given in Algorithm 1. It takes a hypergraph H and a parameter λ as input, constructs a set, S^{uv}, of up to λ hyperarcs for each $(u, v) \in C(F)$, and outputs $S = \bigcup_{(u,v)\in C(F)} S^{uv}$. For each pair (u, v) (in arbitrary order), S^{uv} is obtained by selecting up to the λ heaviest hyperarcs f with $(u, v) \in C(f)$ among those not selected yet. The parameter λ controls the size of output S.

▶ **Lemma 4.** *Let H be a directed hypergraph and λ be a positive integer.* CoresetFinder(H, λ) *returns a set S of at most λn^2 hyperarcs that can be partitioned into disjoints subsets $\{ S^{uv} \mid (u, v) \in C(F) \}$ satisfying the following conditions:*
1. *for any $(u, v) \in C(F)$, every $f \in S^{uv}$ satisfies $(u, v) \in C(f)$,*
2. *if $(u, v) \in C(F \setminus S)$, $|S^{uv}| = \lambda$ holds, and*
3. *for any $(u, v) \in C(F)$, $f \in S^{uv}$, and $f' \in F \setminus S$ such that $(u, v) \in C(f')$, $z_f \geq z_{f'}$ holds.*

Algorithm 3 DH-SPARSIFY(H,ε): iterative algorithm that computes an ε-spectral sparsifier.

Input: $H=(V,F,z)$ with $|V|=n$ and $|F|=m$, and $\varepsilon>0$
Output: $\tilde{H}=(V,\tilde{F},\tilde{z})$

```
1: m* ← (n²/ε²) log³(n/ε)        ▷ This is the (asymptotic) target size of the resulting sparsifier.
2: T ← ⌈log_{4/3}(m/m*)⌉
3: i ← 0, H̃_0 = (V, F̃_0, z̃_0) ← H, and m_0 ← |F̃_0|
4: while i < T and m_i ≥ C_2 m* :        ▷ C_2 is a constant that is explained in Section 4.3.
5:     ε_i ← ε / (4 log²_{4/3}(m_i/m*)) and λ_i ← ⌈C_1 log³ m_i / ε_i²⌉        ▷ ε_i is used in the analysis.
6:     H̃_{i+1} = (V, F̃_{i+1}, z̃_{i+1}) ← DH-ONESTEP(H̃_i, λ_i)
7:     m_{i+1} ← |F̃_{i+1}|
8:     i ← i + 1
9: i_end ← i and H̃ ← H̃_{i_end}
10: return H̃ = (V, F̃, z̃)
```

Proof. Since CORESETFINDER(H,λ) constructs S^{uv} for each $(u,v)\in C(F)$ by selecting up to the λ heaviest hyperarcs f with $(u,v)\in C(f)$ among those that have not been selected yet, S^{uv} for $(u,v)\in C(F)$ are mutually disjoint. This also implies $|S|=\sum_{(u,v)\in C(F)}|S^{uv}|\le\lambda n^2$ and the first and third conditions. After S is constructed, if there is a hyperarc $f'\in F\setminus S$ such that $(u,v)\in C(f')$, then λ hyperarcs must have been added to S^{uv}. Hence $|S^{uv}|=\lambda$ if $(u,v)\in C(F\setminus S)$, implying the second condition. ◀

We call the set S shown in Lemma 4 a *coreset*, which plays a key role in the analysis.

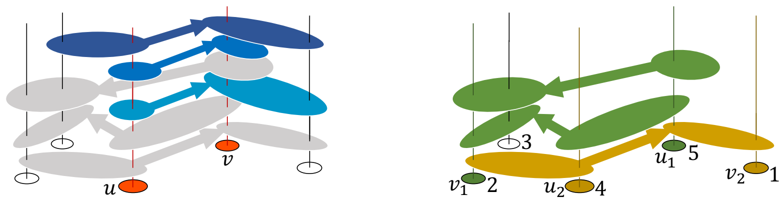
▶ **Definition 5.** *Given a directed hypergraph $H=(V,F,z)$, a subset $S\subseteq F$, and a positive integer λ, we say S is a* λ-coreset *of H if S can be partitioned into disjoints subsets $\{S^{uv}\mid(u,v)\in C(F)\}$ satisfying the three conditions in the statement of Lemma 4.*

In short, if there is a hyperarc $f'\notin S$ with $(u,v)\in C(f')$, S^{uv} contains (at least) λ hyperarcs that are at least as heavy as $z_{f'}$. Figure 1a illustrates an example of a coreset. We use this coreset as a counterpart of a bundle of spanners in the spanner-based sparsification.

Next, we explain DH-ONESTEP(H,λ) given in Algorithm 2, which is the main subroutine in our algorithm. The algorithm first computes a λ-coreset S by calling CORESETFINDER(H,λ). The hyperarcs in the coreset S are deterministically added to the output. Then, it randomly chooses the remaining hyperarcs with probability 1/2 and doubles the weights if sampled, thus preserving the expected total weight. The main technical observation is that, under an appropriate choice of λ, the output of DH-ONESTEP(H,λ) is an ε-spectral sparsifier of H. Formally, we can show the following lemma, which is the main technical contribution and will be proved in Section 4.2.

▶ **Lemma 6.** *Let $H=(V,F,z)$ be a directed hypergraph with $|V|=n$ and $|F|=m$. For any $\varepsilon\in(0,1)$ and $\lambda\ge\frac{C_1\log^3 m}{\varepsilon^2}$, where C_1 is a sufficiently large constant,* DH-ONESTEP(H,λ) *returns an ε-spectral sparsifier $\tilde{H}=(V,\tilde{F},\tilde{z})$ of H satisfying $|\tilde{F}|\le\frac{m}{2}+(3m\log n)^{\frac{1}{2}}+\lambda n^2$ with probability at least $1-O\left(\frac{1}{n^2}\right)$.*

Finally, we present our sparsification algorithm DH-SPARSIFY(H,ε) in Algorithm 3. In the algorithm description, C_1 denotes the constant given in the statement of Lemma 6, and C_2 is a sufficiently large constant (which we can compute explicitly by carefully expanding the analysis in Section 4.3). The algorithm iteratively calls DH-ONESTEP$(\tilde{H}_i,\lambda_i)$, where $\tilde{H}_i$



(a) Coreset. **(b)** Critical pairs.

Figure 1 Illustration of a coreset and critical pairs on (a part of) a given hypergraph. A circle is a vertex, and a hyperarc is indicated by an arrow and two ellipses representing a head and a tail. A hyperarc contains a vertex if the line originating from the vertex pierces its head or tail. Figure 1a presents an image of a coreset, focusing on a vertex pair (u, v). Suppose that the hyperarcs are aligned in decreasing order of their weights from top to bottom. The blue hyperarcs are the three heaviest ones having u and v as elements of their tails and heads, respectively, and they are included in a subset S^{uv} of a coreset S. We suppose that gray hyperarcs are not in S. While the bottommost gray hyperarc f also satisfies $u \in t(f)$ and $v \in h(f)$, the three blue hyperarcs are heavier than it. Thus, the conditions of the λ-coreset with $\lambda = 3$ are satisfied for (u, v). Figure 1b presents an image of critical pairs of three hyperarcs, which are missed by the coreset S in Figure 1a. Suppose that vertices v have x_v values of 2, 3, 4, 5, and 1 from left to right, respectively, as shown nearby the vertices. Then, the green and yellow hyperarcs have (u_1, v_1) and (u_2, v_2), respectively, as x-ctirical pairs. If the three hyperarcs constitute $F_i^x \subseteq F \setminus S$, we have $E_i^x = \{(u_1, v_1), (u_2, v_2)\}$, and F_i^x is partitioned into F_i^{x,u_1v_1} and F_i^{x,u_2v_2}, shown in green and yellow, respectively.

is the sub-hypergraph obtained in the previous step. Here, the parameter λ_i is defined as in Line 3, which makes $\tilde{H}_{i+1}$ an ε_i-spectral sparsifier of $\tilde{H}_i$ by the condition in Lemma 6.[5] The algorithm repeatedly calls DH-OneStep$(\tilde{H}_i, \lambda_i)$ until the size of $\tilde{H}_i$ becomes $\tilde{O}(n^2/\varepsilon^2)$ or the maximum number of iterations, T, is reached. With this choice of ε_i, we will show that the size of $\tilde{H}_i$ decreases geometrically and that the accumulated sparsification error is bounded by ε. Consequently, the final output is an ε-spectral sparsifier of the desired size, which completes the proof of Theorem 1. We present the analysis in Section 4.3.

4.2 Proof of Lemma 6

We prove Lemma 6, which ensures the correctness of DH-OneStep. In this section, we let $H = (V, F, z)$, $\lambda \geq \frac{C_1 \log^3 m}{\varepsilon^2}$, and $\varepsilon \in (0, 1)$ be as given in the statement of Lemma 6, and let $\tilde{H} = (V, \tilde{F}, \tilde{z})$ be the output of DH-OneStep(H, λ).

To prove Lemma 6, we bound the size and sparsification error of $\tilde{H}$ from above. The former is an easy consequence of the Chernoff bound. We below prove it assuming $m > 12 \log n$; otherwise, an input hypergraph is already sparsified and we do not run DH-OneStep.

▶ **Lemma 7.** *Let $H = (V, F, z)$ be a directed hypergraph with $|V| = n$ and $|F| = m$, and let λ be a positive integer. If $m > 12 \log n$,* DH-OneStep(H, λ) *outputs a sub-hypergraph $\tilde{H} = (V, \tilde{F}, \tilde{z})$ of H satisfying $|\tilde{F}| \leq \frac{m}{2} + (3m \log n)^{\frac{1}{2}} + \lambda n^2$ with probability at least $1 - \frac{2}{n^2}$.*

Proof. Let S be a λ-coreset constructed in Line 2 in DH-OneStep(H, λ). By Lemma 4, S has at most λn^2 hyperarcs. To bound $|\tilde{F} \setminus S|$, for each $f \in F \setminus S$, we let X_f be a random variable that takes 1 if f is sampled and 0 otherwise. Note that $|\tilde{F} \setminus S| = \sum_{f \in F \setminus S} X_f$ holds.

[5] Unlike the existing spanner-based algorithm [17], we need to change ε_i adaptively since fixing $\varepsilon_i = \frac{\varepsilon}{T}$ does not yield a sparsifier of the desired size when the input hypergraph is exponentially large in n.

Since we have $\mathbb{E}\left[\sum_{f \in F \setminus S} X_f\right] = (m - |S|)/2 \le m/2$, for any $t \in (0,1)$, the Chernoff bound (Proposition 3) implies

$$\mathbb{P}\left[\sum_{f \in F \setminus S} X_f - \mathbb{E}\left[\sum_{f \in F \setminus S} X_f\right] > \frac{m}{2} t\right] \le 2\exp\left(-\frac{mt^2}{6}\right).$$

By setting $t = \left(\frac{12\log n}{m}\right)^{\frac{1}{2}}$, which is smaller than 1 due to the lemma assumption, we obtain

$$\mathbb{P}\left[\sum_{f \in F \setminus S} X_f \le \frac{m}{2} + (3m\log n)^{\frac{1}{2}}\right] \ge 1 - \frac{2}{n^2}.$$

Thus, we have $|\tilde{F}| = |S| + \sum_{f \in F \setminus S} X_f \le \frac{m}{2} + (3m\log n)^{\frac{1}{2}} + \lambda n^2$ with probability at least $1 - \frac{2}{n^2}$. ◀

The rest of this section focuses on showing that $\tilde{H}$ is an ε-spectral sparsifier of H, i.e., $(1-\varepsilon)x^\top L_H(x) \le x^\top L_{\tilde{H}}(x) \le (1+\varepsilon)x^\top L_H(x)$ for any $x \in \mathbb{R}^V$. Since this relation is invariant under scaling of x, it suffices to prove the relation for any x satisfying $x^\top L_H(x) = 1$. Let $\mathbb{S}_H = \left\{ x \in \mathbb{R}^V \;\middle|\; x^\top L_H(x) = 1 \right\}$. A similar normalization is used in [15] with respect to the total energy of the corresponding underlying clique digraphs. By contrast, we directly normalize the total energy of a hypergraph, $x^\top L_H(x)$. This difference is a key to eliminating the extra r^3 factor, while it requires a new discretization scheme, as described later.

Since we analyze the contribution of each hyperarc to the energy of H, it is convenient to use the notation of $Q_H^x(f)$ and $Q_H^x(F')$ for $f \in F$ and $F' \subseteq F$, respectively, defined in Section 2. Our goal is to prove $(1-\varepsilon)Q_H^x(F) \le Q_{\tilde{H}}^x(\tilde{F}) \le (1+\varepsilon)Q_H^x(F)$ for all $x \in \mathbb{S}_H$.

Given $x \in \mathbb{S}_H$ and a λ-coreset $S \subseteq F$, our strategy is to partition $F \setminus S$ into subsets based on the energies and evaluate the error caused by sparsification for each subset. Specifically, we classify hyperarcs $f \in F \setminus S$ into subsets F_i^x defined for each $i \in \mathbb{Z}$ as follows:

$$F_i^x := \left\{ f \in F \setminus S \;\middle|\; Q_H^x(f) \in \left[\frac{1}{2^i\lambda}, \frac{1}{2^{i-1}\lambda}\right) \right\}.$$

We also define $\tilde{F}_i^x := F_i^x \cap \tilde{F}$ for each $i \in \mathbb{Z}$.

Since $Q_H^x(F \setminus S) = \sum_{i \in \mathbb{Z}} Q_H^x(F_i^x)$ and $Q_{\tilde{H}}^x(\tilde{F} \setminus S) = \sum_{i \in \mathbb{Z}} Q_{\tilde{H}}^x(\tilde{F}_i^x)$, our goal is to prove that $|Q_{\tilde{H}}^x(\tilde{F}_i^x) - Q_H^x(F_i^x)|$ is sufficiently small for all $i \in \mathbb{Z}$ and $x \in \mathbb{S}_H$. This is not difficult if i is sufficiently large, as in the following lemma.

▶ **Lemma 8.** *Let $I = \lceil \log_2(9m) \rceil$. For any $x \in \mathbb{S}_H$, $\left| Q_H^x(\cup_{i \ge I+1} F_i^x) - Q_{\tilde{H}}^x\left(\cup_{i \ge I+1} \tilde{F}_i^x\right) \right| \le \frac{\varepsilon}{3}$.*

Proof. Due to the assumption in Lemma 6, $\lambda\varepsilon \ge \frac{C_1 \log^3 m}{\varepsilon} \ge 1$ holds for sufficiently large C_1. By the definition of F_i^x, the energy of each hyperarc in $\cup_{i \ge I+1} F_i^x$ is less than $\frac{1}{2^I\lambda}$, which is at most $\frac{\varepsilon}{9m}$ by $I = \lceil \log_2(9m) \rceil$ and $\lambda\varepsilon \ge 1$. Thus, it holds that

$$Q_H^x(\cup_{i \ge I+1} F_i^x) = \sum_{f \in \cup_{i \ge I+1} F_i^x} Q_H^x(f) \le m \cdot \frac{\varepsilon}{9m} \le \frac{\varepsilon}{9}. \tag{1}$$

As for $\tilde{F} \subseteq F$, since the weight of each hyperarc in $\tilde{F}$ is doubled in DH-Onestep, we have

$$Q_{\tilde{H}}^x\left(\cup_{i \ge I+1} \tilde{F}_i^x\right) \le 2 \cdot Q_H^x(\cup_{i \ge I+1} F_i^x) \le \frac{2\varepsilon}{9}. \tag{2}$$

Combining eqs. (1) and (2), we obtain the claim. ◀

We then introduce additional definitions for the convenience of describing our discretization scheme and analyzing the sparsification error.

▶ **Definition 9.** *For $x \in \mathbb{S}_H$, we say $(u,v) \in V \times V$ is an x-critical pair of $f \in F$ if we have $(u,v) = \operatorname{argmax}_{(u,v) \in C(f)} (x_u - x_v)_+^2$, breaking ties as in Section 2. For $i \in \mathbb{Z}$ and $x \in \mathbb{S}_H$, let*

$$E_i^x = \{ (u,v) \in C(F) \mid (u,v) \text{ is an } x\text{-critical pair of some } f \in F_i^x \}$$

and, for each $(u,v) \in E_i^x$, let

$$F_i^{x,uv} = \{ f \in F_i^x \mid (u,v) \text{ is an } x\text{-critical pair of } f \}.$$

Note that the collection of $F_i^{x,uv}$ for $(u,v) \in E_i^x$ forms a partition of F_i^x. Figure 1b presents an example of x-critical pairs.

We now discuss how to bound $|Q_{\tilde{H}}^x(\tilde{F}_i^x) - Q_H^x(F_i^x)|$ for i that is not covered in Lemma 8. By using the Chernoff bound, it is easy to evaluate the probability that $|Q_{\tilde{H}}^x(\tilde{F}_i^x) - Q_H^x(F_i^x)|$ is small for each $x \in \mathbb{S}_H$. To convert it to a uniform bound over all $x \in \mathbb{S}_H$, we construct an appropriate discretization scheme, as follows.

Let $\Delta = \frac{\varepsilon}{9m}$. For $(u,v) \in E_i^x$, we define the *discretization width* as $\Delta_i^{uv} := \frac{\Delta}{\max_{f \in F_i^{x,uv}} z_f}$. Note that $F_i^{x,uv} \neq \emptyset$ holds for $(u,v) \in E_i$ by the definitions of E_i and $F_i^{x,uv}$, and hence Δ_i^{uv} is well-defined. The denominator plays the role of scaling the width. Given any $x \in \mathbb{S}_H$, we consider discretizing $(x_u - x_v)_+^2$ for each $(u,v) \in E_i^x$, not the energy itself. Specifically, for each $i \in \mathbb{Z}$ and $(u,v) \in E_i^x$, we use $\left\lfloor \frac{(x_u - x_v)_+^2}{\Delta_i^{uv}} \right\rfloor \Delta_i^{uv}$ as a discretized value of $(x_u - x_v)_+^2$. Then, for each $f \in F_i^{x,uv}$ such that $(u,v) \in E_i^x$, we define the *discretized energy* $D_H^x(f)$ by

$$D_H^x(f) := z_f \left\lfloor \frac{(x_u - x_v)_+^2}{\Delta_i^{uv}} \right\rfloor \Delta_i^{uv}.$$

It should be noted that the discretized energy of $f \in F_i^{x,uv}$ is defined by first discretizing $(x_u - x_v)_+^2$ and then scaling it by z_f. This somewhat indirect discretization scheme will turn out important when bounding the number of possible discretized energies.

For each sampled hyperarc $f \in F_i^{x,uv} \cap \tilde{F}$ with $(u,v) \in E_i^x$, we define the *discretized energy after sampling* by $D_{\tilde{H}}^x(f) := 2D_H^x(f)$. We also let $D_H^x(F_i^x) = \sum_{f \in F_i^x} D_H^x(f)$ and $D_{\tilde{H}}^x(\tilde{F}_i^x) = \sum_{f \in \tilde{F}_i^x} D_{\tilde{H}}^x(f)$. We can ensures that discretization errors are small as follows.

▶ **Lemma 10.** *For any $x \in \mathbb{S}_H$, we have*

$$\sum_{i \in \mathbb{Z}} |D_H^x(F_i^x) - Q_H^x(F_i^x)| \leq \frac{\varepsilon}{9} \quad \text{and} \quad \sum_{i \in \mathbb{Z}} \left| D_{\tilde{H}}^x(\tilde{F}_i^x) - Q_{\tilde{H}}^x(\tilde{F}_i^x) \right| \leq \frac{2\varepsilon}{9}.$$

Proof. Recall that the discretized energy $D_H^x(f)$ of each $f \in F_i^{x,uv}$ is obtained by discretizing $(x_u - x_v)_+^2$ with the width Δ_i^{uv} and scaling it by z_f. Therefore, the discretization error for each f is bounded by $z_f \Delta_i^{uv}$. From the definition of Δ_i^{uv}, we have $z_f \Delta_i^{uv} = z_f \frac{\Delta}{\max_{f \in F_i^{x,uv}} z_f} \leq \Delta$. Hence, the total discretization error over all $f \in F \setminus S$ is bounded by $m\Delta$, which is at most $\frac{\varepsilon}{9}$ since $\Delta = \frac{\varepsilon}{9m}$. Thus, we obtain the first inequality. The second inequality follows from the fact that the weights of sampled hyperarcs are doubled. ◀

From Lemma 10, we can bound the sparsification error $|Q_{\tilde{H}}^x(\tilde{F}_i^x) - Q_H^x(F_i^x)|$ for all $x \in \mathbb{S}_H$ by bounding $|D_{\tilde{H}}^x(\tilde{F}_i^x) - D_H^x(F_i^x)|$ for all $x \in \mathbb{S}_H$. Since the number of possible discretized energies is finite, we can use the standard Chernoff bound and union bound to evaluate the sparsification error. Thus, what remains is to prove that the number of discretized energies is small enough so that we can obtain the desired uniform bound. To this end, we first bound the size of E_i^x and then bound the number of possible discretized values. The following lemma bounds the size of E_i^x, in which the existence of a λ-coreset plays an important role.

▶ **Lemma 11.** *For $i \in \mathbb{Z}$, we have $|E_i^x| < 2^i$.*

Proof. By the definition of E_i^x, for each $(u,v) \in E_i^x$, there is a hyperarc $f^{uv} \in F_i^x \subseteq F \setminus S$ such that (u,v) is an x-critical pair of f^{uv}. Since S is a λ-coreset, S admits a partition $\{ S^{uv} \mid (u,v) \in C(F) \}$ satisfying the three conditions in Lemma 4. Since $f^{uv} \notin S$, the third condition in Lemma 4 implies $z_f \geq z_{f^{uv}}$ for any $f \in S^{uv}$. Hence, for any $f \in S^{uv}$, we have

$$Q_H^x(f^{uv}) = z_{f^{uv}}(x_u - x_v)_+^2 \leq z_f(x_u - x_v)_+^2 \leq \max_{(u',v') \in C(f)} z_f(x_{u'} - x_{v'})_+^2 = Q_H^x(f). \quad (3)$$

Since the second condition in Lemma 4 implies $|S^{uv}| = \lambda$ for $(u,v) \in E_i^x \subseteq C(F \setminus S)$,

$$\begin{aligned} Q_H^x(F) &\geq \sum_{(u,v) \in E_i^x} (Q_H^x(S^{uv}) + Q_H^x(f^{uv})) \quad \text{(since all } S^{uv} \text{ and } f^{uv} \notin S \text{ are disjoint)} \\ &\geq \sum_{(u,v) \in E_i^x} (\lambda + 1) \cdot Q_H^x(f^{uv}) \qquad \text{(by eq. (3) and } |S^{uv}| = \lambda) \\ &\geq \sum_{(u,v) \in E_i^x} (\lambda + 1) \cdot (2^i \lambda)^{-1} \qquad \text{(by } f^{uv} \in F_i^x). \end{aligned}$$

holds, hence $Q_H^x(F) > 2^{-i}|E_i^x|$. Since $Q_H^x(F) = 1$ by $x \in \mathbb{S}_H$, we obtain $|E_i^x| < 2^i$. ◀

From Lemma 11, if $i \leq 0$, we have $|E_i^x| < 2^i \leq 1$, which implies $E_i^x = \emptyset$ and $F_i^x = \emptyset$. Thus, the following corollary holds.

▶ **Corollary 12.** *If $i \leq 0$, we have $F_i^x = \emptyset$.*

Due to Corollary 12 and Lemma 8, we can focus on $i \in \mathbb{Z}$ with $1 \leq i \leq I = \lceil \log_2 9m \rceil$. In this range, we have the following bound on the number of possible discretized values.

▶ **Lemma 13.** *For each positive integer i, let $L_i = \left\{ \left(F_i^x, \{D_H^x(f)\}_{f \in F_i^x}\right) \mid x \in \mathbb{S}_H \right\}$, where $\{D_H^x(f)\}_{f \in F_i^x}$ is the list of the discretized energies over all hyperarcs in F_i^x. If $1 \leq i \leq I = \lceil \log_2 9m \rceil$, we have $|L_i| \leq \left(\frac{648 n^4 m^4}{\lambda \varepsilon}\right)^{2^i}$.*

Since the proof of Lemma 13 is not short, we first complete the proof of Lemma 6 assuming that Lemma 13 is true; then, we prove Lemma 13 in Section 4.2.1.

Proof of Lemma 6. Let $I = \lceil \log_2(9m) \rceil$ as in Lemma 8 and define L_i as in Lemma 13. Fix $i \in \{1, 2, \ldots, I\}$ and consider any element of L_i, which we denote by $(F_i^y, \{D_H^y(f)\}_{f \in F_i^y})$ for some $y \in \mathbb{S}_H$. Since the discretized energy of each hyperarc is obtained by rounding down, we have $D_H^y(f) \leq Q_H^y(f)$. Thus, for every $f \in F_i^y$, it holds that

$$D_H^y(f) \leq Q_H^y(f) < \frac{1}{2^{i-1}\lambda}. \quad (4)$$

For each $f \in F \setminus S$, let X_f be a random variable that takes 1 with probability 1/2 and 0 otherwise, which represents the randomness of sampling and hence $D_{\tilde{H}}^y(\tilde{F}_i^y) = \sum_{f \in F_i^y} 2X_f D_H^y(f)$. By $D_H^y(f) \leq Q_H^y(f)$ again, we have

$$\mathbb{E}\left[\sum_{f \in F_i^y} 2X_f D_H^y(f)\right] = \sum_{f \in F_i^y} D_H^y(f) = D_H^y(F_i^y) \leq Q_H^y(F_i^y) \leq Q_H^y(F) = 1. \quad (5)$$

Due to eqs. (4) and (5), the Chernoff bound (Proposition 3) with $\mu = 1$, $a = \frac{1}{2^{i-2}\lambda}$, and $\delta = \frac{\varepsilon}{3I}$ implies

$$\mathbb{P}\Big[\Big|D^y_{\tilde{H}}(\tilde{F}^y_i) - D^y_H(F^y_i)\Big| > \frac{\varepsilon}{3I}\Big] = \mathbb{P}\left[\left|\sum_{f \in F^y_i} 2X_f D^y_H(f) - \mathbb{E}\left[\sum_{f \in F^y_i} 2X_f D^y_H(f)\right]\right| > \frac{\varepsilon}{3I}\right]$$
$$\le 2\exp\left(-\frac{2^i \cdot \varepsilon^2 \lambda}{108 I^2}\right).$$

This bound is true for each $(F^y_i, \{D^y_H(f)\}_{f \in F^y_i}) \in L_i$, and we can convert it to a uniform bound over all $(F^y_i, \{D^y_H(f)\}_{f \in F^y_i}) \in L_i$ by using Lemma 13 and the union bound as follows:

$$\mathbb{P}\Big[\exists (F^y_i, \{D^y_H(f)\}_{f \in F^y_i}) \in L_i, \big|D^y_{\tilde{H}}(\tilde{F}^y_i) - D^y_H(F^y_i)\big| > \frac{\varepsilon}{3I}\Big] \le 2\exp\left(-\frac{2^i \cdot \varepsilon^2 \lambda}{108 I^2}\right) \cdot \left(\frac{648 n^4 m^4}{\lambda \varepsilon}\right)^{2^i}.$$

We may assume $nm \ge 648$ (otherwise Lemma 6 is trivial for a sufficiently large C_1). Letting C_1 be sufficiently large, we have $\lambda \ge \frac{C_1 \log^3 m}{\varepsilon^2} \ge \frac{108 I^2}{\varepsilon^2}(6\log n + 5\log m)$ and $\lambda\varepsilon \ge 1$. Thus, we can further bound the right-hand side from above by

$$2\exp\left(-\frac{2^i \cdot \varepsilon^2 \lambda}{108 I^2}\right) \cdot \big(n^5 m^5\big)^{2^i} \le 2\exp\big(-2^i \cdot (6\log n + 5\log m)\big) \cdot \big(n^5 m^5\big)^{2^i} \le \frac{2}{n^{2^i}}.$$

Therefore, $\mathbb{P}[\forall (F^y_i, \{D^y_H(f)\}_{f \in F^y_i}) \in L_i, |D^y_{\tilde{H}}(\tilde{F}^y_i) - D^y_H(F^y_i)| \le \frac{\varepsilon}{3I}] \ge 1 - \frac{2}{n^{2^i}}$ holds. Since $(F^x_i, \{D^x(f)\}_{f \in F^x_i}) \in L_i$ holds for all $x \in \mathbb{S}_H$, we can equivalently rewrite the bound as

$$\mathbb{P}\Big[\forall x \in \mathbb{S}_H, \big|D^x_{\tilde{H}}(\tilde{F}^x_i) - D^x_H(F^x_i)\big| \le \frac{\varepsilon}{3I}\Big] \ge 1 - \frac{2}{n^{2^i}}.$$

By the union bound over $1 \le i \le I = \lceil \log_2(9m) \rceil$ and $\sum_{i=1}^{I} \frac{2}{n^{2^i}} \le \sum_{i=1}^{\infty} \frac{2}{n^{2i}} \le \frac{2}{n^2 - 1} \le \frac{3}{n^2}$ (for $n \ge 2$), we obtain

$$\mathbb{P}\left[\forall x \in \mathbb{S}_H, \sum_{i=1}^{I} \big|D^x_{\tilde{H}}(\tilde{F}^x_i) - D^x_H(F^x_i)\big| \le \frac{\varepsilon}{3}\right] \ge 1 - \frac{3}{n^2}. \tag{6}$$

Thus, for all $x \in \mathbb{S}_H$, we can bound $|x^\top L_{\tilde{H}}(x) - x^\top L_H(x)| = |Q^x_{\tilde{H}}(\tilde{F}) - Q^x_H(F)|$ as follows:

$$\begin{aligned}
&\big|Q^x_{\tilde{H}}(\tilde{F}) - Q^x_H(F)\big| \\
&= \frac{\varepsilon}{3} + \sum_{i=1}^{I} |Q^x_{\tilde{H}}(\tilde{F}^x_i) - Q^x_H(F^x_i)| \qquad \text{(by Lemma 8 and Corollary 12)} \\
&\le \frac{\varepsilon}{3} + \sum_{i=1}^{I} \big[|Q^x_{\tilde{H}}(\tilde{F}^x_i) - D^x_{\tilde{H}}(\tilde{F}^x_i)| + |D^x_{\tilde{H}}(\tilde{F}^x_i) - D^x_H(F^x_i)| + |D^x_H(F^x_i) - Q^x_H(F^x_i)|\big] \\
&\le \frac{\varepsilon}{3} + \frac{\varepsilon}{9} + \frac{\varepsilon}{3} + \frac{2\varepsilon}{9} \qquad \text{(by Lemma 10 and eq. (6))} \\
&= \varepsilon,
\end{aligned}$$

which holds with probability at least $1 - \frac{3}{n^2}$. Hence, $\tilde{H}$ is an ε-spectral sparsifier of H. Combining this with the size bound in Lemma 7, we obtain Lemma 6. ◀

4.2.1 Proof of Lemma 13

We present the proof of Lemma 13. Our goal is to bound the size of L_i defined in Lemma 13 for $i \in \mathbb{Z}$ with $1 \le i \le I = \lceil \log_2 9m \rceil$. To this end, we proceed in two steps: we first bound the number of possible combinations of $(F^x_i, E^x_i, \{F^{x,uv}_i\}_{f \in E^x_i})$ over all $x \in \mathbb{S}_H$, and then bound the number of possible lists $\{D^x_H(f)\}_{f \in F^x_i}$ of discretized energies. For convenience, we define the following notion.

▶ **Definition 14.** *Let $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ be a tuple such that $E \subseteq V \times V$, $\{f_{uv}\}_{(u,v)\in E}$ is a list of hyperarcs indexed by $(u, v) \in E$, and π_E is a total ordering on E. For $i \in \{1, 2, \dots, I\}$, we say $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ is* i-realized *by $x \in \mathbb{S}_H$ if the following conditions hold:*

1. *$E = E_i^x$,*
2. *$f_{uv} = \operatorname{argmin}_{f\in F_i^{x,uv}} z_f$ for each $(u, v) \in E_i^x$, and*
3. *π_E is the increasing order of the values of $(x_u - x_v)_+^2$, i.e., (u, v) is smaller than (u', v') in π_E if and only if $(x_u - x_v)_+^2 \le (x_{u'} - x_{v'})_+^2$ (where the tie-breaking rule explained in Section 2 is used when the equality holds).*

The following lemma says that the i-realizability determines E_i^x, F_i^x, and $F_i^{x,uv}$, implying that we can reduce the problem of counting the number of possible $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{f\in E_i^x})$ to that of counting the number of possible tuples $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$.

▶ **Lemma 15.** *Let $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ be a tuple as defined in Definition 14 and $x, y \in \mathbb{S}_H$. If both x and y i-realize $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ and $\bigcup_{j=1}^{i-1} F_j^x = \bigcup_{j=1}^{i-1} F_j^y$ holds, then, for every $(u, v) \in E$, we have $E_i^x = E_i^y$, $F_i^x = F_i^y$, and $F_i^{x,uv} = F_i^{y,uv}$.*

Proof. By the definition of the i-realizability, we have $E_i^x = E = E_i^y$. If we can assume $F_i^{x,uv} = F_i^{y,uv}$ for every $(u, v) \in E$, we have $F_i^x = \bigcup_{(u,v)\in E} F_i^{x,uv} = \bigcup_{(u,v)\in E} F_i^{y,uv} = F_i^y$ since $\{F_i^{x,uv} \mid (u, v) \in C(F)\}$ and $\{F_i^{y,uv} \mid (u, v) \in C(F)\}$ are partitions of F_i^x and F_i^y, respectively. Therefore, we below focus on proving $F_i^{x,uv} = F_i^{y,uv}$ for every $(u, v) \in E$.

For a contradiction, suppose $F_i^{x,u_1v_1} \ne F_i^{y,u_1v_1}$ for some $(u_1, v_1) \in E$. Without loss of generality, we assume there is a hyperarc $f^* \in F_i^{x,u_1v_1} \setminus F_i^{y,u_1v_1}$. Since both x and y i-realize $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ and $(u_1, v_1) \in E$, the second condition of the i-realizability implies

$$\operatorname*{argmin}_{f\in F_i^{x,u_1v_1}} z_f = f_{u_1v_1} = \operatorname*{argmin}_{f\in F_i^{y,u_1v_1}} z_f. \tag{7}$$

In particular, we have $z_{f_{u_1v_1}} \le z_{f^*}$ for $f^* \in F_i^{x,u_1v_1}$. Hence

$$\begin{aligned} Q_H^y(f^*) &= z_{f^*} \max_{(u,v)\in C(f^*)} (y_u - y_v)_+^2 \\ &\ge z_{f_{u_1v_1}} (y_{u_1} - y_{v_1})_+^2 \qquad \big(\text{by } (u_1, v_1) \in C(f^*) \text{ and } z_{f^*} \ge z_{f_{u_1v_1}}\big) \\ &= z_{f_{u_1v_1}} \max_{(u,v)\in C(f_{u_1v_1})} (y_u - y_v)_+^2 \qquad (\text{by } f_{u_1v_1} \in F_i^{y,u_1v_1} \text{ as in eq. (7)}) \\ &\ge \frac{2^{-i}}{\lambda} \qquad (\text{by } f_{u_1v_1} \in F_i^y). \end{aligned}$$

From $Q_H^y(f^*) \ge \frac{2^{-i}}{\lambda}$ and $f^* \in F_i^{x,u_1v_1} \subseteq F \setminus S$, it must hold that $f^* \in \bigcup_{j=1}^i F_j^y$. Moreover, since $\bigcup_{j=1}^{i-1} F_j^x = \bigcup_{j=1}^{i-1} F_j^y$ by the lemma assumption and $f^* \notin \bigcup_{j=1}^{i-1} F_j^x$ by $f^* \in F_i^{x,u_1v_1}$, we have $f^* \notin \bigcup_{j=1}^{i-1} F_j^y$, hence $f^* \in F_i^y$. Since the orderings of E with respect to $(x_u - x_v)_+^2$ and $(y_u - y_v)_+^2$ are both equal to π_E and (u_1, v_1) is an x-critical pair of f^*, we have

$$(u_1, v_1) = \operatorname*{argmax}_{(u,v)\in C(f^*)\cap E} (y_u - y_v)_+^2. \tag{8}$$

Since $f^* \in F_i^y$, eq. (8) implies $f^* \in F_i^{y,u_1v_1}$, contradicting the assumption of $f^* \notin F_i^{y,u_1v_1}$. Therefore, $F_i^{x,uv} = F_i^{y,uv}$ holds for every $(u, v) \in E$. ◀

Lemma 15 enables us to bound the number of possible $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{f\in E_i^x})$ for $x \in \mathbb{S}_H$.

▶ **Lemma 16.** *For each $i \ge 1$, $\left|\left\{ (F_i^x, E_i^x, \{F_i^{x,uv}\}_{f\in E_i^x}) \;\middle|\; x \in \mathbb{S}_H \right\}\right| \le \left(2^i n^2 m\right)^{2^{i+1}}$ holds.*

Proof. First, we suppose that F_j^x for $j = 1, \dots, i-1$ are fixed. Then, due to Lemma 15, we can bound the number of possible combinations of $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{f \in E_i^x})$ for all $x \in \mathbb{S}_H$ by counting the number of possible tuples $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ that can be i-realized by some $x \in \mathbb{S}_H$. Since $|E| < 2^i$ by Lemma 11, the number of possible E is $\sum_{k=1}^{|E|} \binom{n^2}{k} \le \sum_{k=1}^{2^i-1} \binom{n^2}{k}$. Once E is specified, there are up to m possible choices of f_{uv} for each $(u,v) \in E$. Furthermore, the number of possible total orderings π_E of E is at most $(|E|)! \le (2^i)!$. Thus, the number of possible tuples $(E, \{f_{uv}\}_{(u,v)\in E}, \pi_E)$ that can be i-realized by some $x \in \mathbb{S}_H$ is at most $\left(\sum_{k=1}^{2^i-1} \binom{n^2}{k}\right) \cdot m^{2^i} \cdot (2^i)!$. This is further upper bounded by $\left(2^i n^2 m\right)^{2^i}$ by a simple calculation.

We now remove the assumption that F_j^x for $j = 1, \dots, i-1$ are fixed. By inductively using the above bound in increasing order of j, the number of possible combinations of $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{f \in E_i^x})$ over all $x \in \mathbb{S}_H$ is at most $\prod_{j=1}^{i} \left(2^j n^2 m\right)^{2^j} \le \left(2^i n^2 m\right)^{\sum_{j=1}^{i} 2^j} \le \left(2^i n^2 m\right)^{2^{i+1}}$, thus completing the proof. ◀

We then fix any tuple $(F_i^y, E_i^y, \{F_i^{y,uv}\}_{f \in E_i^y})$ for some representative $y \in \mathbb{S}_H$ and upper bound the number of possible lists of discretized energies, $\{D_H^x(f)\}_{f \in F_i^x}$, over a subspace of $\mathbb{S}_H$ that consists of x with $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{f \in E_i^x}) = (F_i^y, E_i^y, \{F_i^{y,uv}\}_{f \in E_i^y})$.

▶ **Lemma 17.** *Let $i \ge 0$ and fix $y \in \mathbb{S}_H$ arbitrarily. The number of possible lists $\{D_H^x(f)\}_{f \in F_i^x}$ for all $x \in \mathbb{S}_H$ with $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{(u,v) \in E_i^x}) = (F_i^y, E_i^y, \{F_i^{y,uv}\}_{(u,v) \in E_i^y})$ is at most $\left(\frac{9m}{2^{i-2}\lambda\varepsilon}\right)^{2^i}$.*

Proof. Let $x \in \mathbb{S}_H$ satisfy the condition in the lemma statement and fix $(u,v) \in E_i^x$. Since every $f \in F_i^{x,uv} \subseteq F_i^x$ satisfies $z_f(x_u - x_v)_+^2 = Q_H^x(f) < \frac{1}{2^{i-1}\lambda}$, the range of $(x_u - x_v)_+^2$ is restricted to $\left[0, \frac{1}{2^{i-1}\lambda \min_{f \in F_i^{x,uv}} z_f}\right)$. Hence, the number of possible discretized $(x_u - x_v)_+^2$ values, $\lfloor (x_u - x_v)_+^2 / \Delta_i^{uv} \rfloor \Delta_i^{uv}$, over all $x \in \mathbb{S}_H$ under the lemma condition is at most

$$\frac{1}{\Delta_i^{uv} 2^{i-1}\lambda \min_{f \in F_i^{x,uv}} z_f} = \frac{1}{\Delta 2^{i-1}\lambda} \cdot \frac{\max_{f \in F_i^{x,uv}} z_f}{\min_{f \in F_i^{x,uv}} z_f} \le \frac{1}{\Delta 2^{i-2}\lambda}, \tag{9}$$

where the equality is due to $\Delta_i^{uv} = \Delta / \max_{f \in F_i^{x,uv}} z_f$ and the inequality comes from $z_f(x_u - x_v)_+^2 = Q_H^x(f) \in \left[\frac{1}{2^i\lambda}, \frac{1}{2^{i-1}\lambda}\right)$ for $f \in F_i^{x,uv} \subseteq F_i^x$, i.e., $\max_{f \in F_i^{x,uv}} z_f \le 2 \min_{f \in F_i^{x,uv}} z_f$.

Since the discretized energy of $f \in F_i^{x,uv}$ is defined by $D_H^x(f) = z_f \lfloor (x_u - x_v)_+^2 / \Delta_i^{uv} \rfloor \Delta_i^{uv}$, fixing the discretization of $(x_u - x_v)_+^2$ determines discretized energies of all $f \in F_i^{x,uv}$. Therefore, the number of possible lists $\{D_H^x(f)\}_{f \in F_i^{x,uv}}$ is also bounded by eq. (9) for each $(u,v) \in E_i^x$. Since $|E_i^x| < 2^i$ by Lemma 11, the number of possible lists $\{D_H^x(f)\}_{f \in F_i^x}$ is at most $\left(\frac{1}{\Delta 2^{i-2}\lambda}\right)^{2^i}$. By substituting $\Delta = \frac{\varepsilon}{9m}$ into it, we obtain the lemma. ◀

We are now ready to prove Lemma 13.

Proof of Lemma 13. We can uniquely specify any element of L_i by first fixing $(F_i^x, E_i^x, \{F_i^{x,uv}\}_{f \in E_i^x})$ and then $\{D_H^x(f)\}_{f \in F_i^x}$. Therefore, we have $|L_i| \le (2^i n^2 m)^{2^{i+1}} \cdot \left(\frac{9m}{2^{i-2}\lambda\varepsilon}\right)^{2^i} = \left(\frac{36 \cdot 2^i n^4 m^3}{\lambda\varepsilon}\right)^{2^i}$ by Lemmas 16 and 17. Combining this with $i \le I = \lceil \log_2 9m \rceil$ completes the proof. ◀

4.3 Proof of Theorem 1

Let $H = (V, F, z)$ be a directed hypergraph with $|V| = n$ and $|F| = m$, $\varepsilon \in (0,1)$, and $\tilde{H} = (V, \tilde{F}, \tilde{z})$ the output of DH-Sparsify(H, ε). Our goal is to prove that $\tilde{H}$ is an ε-spectral sparsifier of H and $|\tilde{F}| = O\left(\frac{n^2}{\varepsilon^2}\log^3 \frac{n}{\varepsilon}\right)$. We here use m^*, T, i_{end}, $(\tilde{H}_i = (V, \tilde{F}_i, \tilde{z}_i), \lambda_i)$, m_i, and ε_i given in the description of DH-SPARSIFY(H, ε) (Algorithm 3), where $m^* = \frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon}$ is the target sparsifier size, $T = \left\lceil \log_{4/3}\left(\frac{m}{m^*}\right) \right\rceil$ is the maximum number of iterations, i_{end} is the number of iterations performed, $(\tilde{H}_i = (V, \tilde{F}_i, \tilde{z}_i), \lambda_i)$ is the input of DH-ONESTEP at the ith iteration, $m_i = |\tilde{F}_i|$, and $\varepsilon_i = \frac{\varepsilon}{4 \log^2_{4/3}\left(\frac{m_i}{m^*}\right)}$, as in Line 3 of Algorithm 3.

We first show that the number of hyperarcs decreases geometrically in each step.

▶ **Lemma 18.** *Let m_i be the number of hyperarcs in $\tilde{H}_i$. Assume $m_i \geq C_2 m^* = C_2 \frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon}$ for a sufficiently large constant C_2. Then, we have $(3 m_i \log n)^{\frac{1}{2}} + \lambda_i n^2 \leq \frac{m_i}{4}$.*

Proof. It is easy to show that $(3 m_i \log n)^{\frac{1}{2}} \leq \frac{m_i}{8}$ holds if $m_i \geq 192 \log n$, which is true if C_2 is sufficiently large. Hence, the desired inequality holds if $\lambda_i n^2 \leq \frac{m_i}{8}$, which we show below.

By Line 3 in Algorithm 3, we have $\varepsilon_i = \frac{\varepsilon}{4 \log^2_{4/3} \frac{m_i}{m^*}}$ and $\lambda_i = \left\lceil \frac{C_1 \log^3 m_i}{\varepsilon_i^2} \right\rceil$. Hence,

$$\frac{m_i}{8} - \lambda_i n^2 \geq \frac{m_i}{8} - \frac{2500 C_1 n^2}{\varepsilon^2} \log^3 m_i \log^4 \frac{m_i}{m^*} \qquad (\text{by } 4^2/\log^4(4/3) < 2500).$$

Let $m_i = \alpha m_*$ and $g(\alpha)$ be the right-hand side of the above inequality, which we regard as a function of α. Since $m^* = (n/\varepsilon)^2 \log^3(n/\varepsilon)$, we have

$$\begin{aligned} g(\alpha) &= m_* \left(\frac{\alpha}{8} - \frac{2500 C_1}{\log^3(n/\varepsilon)} \log^3(\alpha m_*) \log^4 \alpha \right) \\ &\geq m_* \left(\frac{\alpha}{8} - \frac{10000 C_1}{\log^3(n/\varepsilon)} (\log^3 \alpha + \log^3 m_*) \log^4 \alpha \right) \quad (\text{by } (a+b)^3 \leq 4(a^3 + b^3)) \\ &\geq m_* \left(\frac{\alpha}{8} - 10000 C_1 \left(\frac{\log^3 \alpha}{\log^3(n/\varepsilon)} + 125 \right) \log^4 \alpha \right) \quad \left(\text{by } m_* = \frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon} \leq \left(\frac{n}{\varepsilon}\right)^5 \right). \end{aligned}$$

Thus, there exists a sufficiently large constant C_2, which is independent of n and ε, such that $g(\alpha) \geq 0$ holds for all $\alpha \geq C_2$. Using this constant C_2, for all $m_i \geq C_2 m^*$, we have $\lambda_i n^2 \leq \frac{m_i}{8}$ as desired. ◀

Proof of Theorem 1. We say DH-ONESTEP(H_i, λ_i) is *successful* if $\tilde{H}_{i+1}$ is an ε_i-spectral sparsifier of $\tilde{H}_i$ and $m_{i+1} \leq \frac{3}{4} m_i$ holds. DH-SPARSIFY(H, ε) calls DH-ONESTEP(H_i, λ_i) only when $m_i \geq C_2 m^*$ and $i \leq T$. Therefore, by Lemmas 6 and 18, with probability at least $1 - O\left(\frac{T}{n^2}\right) \gtrsim 1 - O\left(\frac{1}{n}\right)$, DH-ONESTEP$(H_i, \lambda_i)$ is successful for all i with $0 \leq i \leq i_{\text{end}}$. Hence, assuming all DH-ONESTEP(H_i, λ_i) to be successful, we below prove that the output hypergraph $\tilde{H}$ has $O\left(\frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon}\right)$ hyperarcs and that $\tilde{H}$ is an ε-spectral sparsifier of H.

We first discuss the size of $\tilde{H}$. If $m_i \leq C_2 m^* = \frac{C_2 n^2 \log^3(n/\varepsilon)}{\varepsilon^2}$ occurs for some $i \leq T - 1$, then m_i gives the size of $\tilde{H}$ by the termination rule of DH-SPARSIFY, which is already small enough. Hence we below assume $m_i \geq C_2 m^*$ for all $i < T$. Since every DH-ONESTEP(H_i, λ_i) is successful, $m_{i+1} \leq \frac{3}{4} m_i$ holds for all $i = 0, 1, \cdots, T-1$. Thus, it holds that

$$m_T \leq m \cdot \left(\frac{3}{4}\right)^T \leq m \cdot \left(\frac{3}{4}\right)^{\log_{4/3} \frac{m \varepsilon^2}{n^2 \log^3(n/\varepsilon)}} = \frac{n^2 \log^3(n/\varepsilon)}{\varepsilon^2}.$$

Therefore, we have $|\tilde{F}| = O\left(\frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon}\right)$.

We then show that $\tilde{H}$ is an ε-spectral sparsifier of H. Since $\tilde{H}_{i+1}$ is an ε_i-spectral sparsifier of $\tilde{H}_i$ for all $i = 0, 1, \cdots, i_{\text{end}} - 1$, the output hypergraph $\tilde{H} = \tilde{H}_{i_{\text{end}}}$ is an $\tilde{\varepsilon}$-spectral sparsifier of H, where

$$\tilde{\varepsilon} = \max \left\{ \prod_{i=0}^{i_{\text{end}}-1} (1 + \varepsilon_i) - 1, 1 - \prod_{i=0}^{i_{\text{end}}-1} (1 - \varepsilon_i) \right\}.$$

A simple calculation yields the following upper bound on $\tilde{\varepsilon}$:

$$\tilde{\varepsilon} \leq \sum_{j=1}^{i_{\mathrm{end}}} \sum_{0 \leq i_1 < \cdots < i_j \leq i_{\mathrm{end}}-1} \varepsilon_{i_1} \varepsilon_{i_2} \cdots \varepsilon_{i_j} \leq \sum_{j=1}^{i_{\mathrm{end}}} \left(\sum_{i=0}^{i_{\mathrm{end}}-1} \varepsilon_i \right)^j . \tag{10}$$

Since $m_{i+1} \leq \frac{3}{4} m_i$ and $m_{i_{\mathrm{end}}-1} \geq C_2 m^*$, we have $m_{i_{\mathrm{end}}-j} \geq \left(\frac{4}{3}\right)^{j-1} C_2 m^* \geq \left(\frac{4}{3}\right)^j m^*$ for sufficiently large $C_2 \geq \frac{4}{3}$, hence $\log_{4/3}\left(\frac{m_{i_{\mathrm{end}}-j}}{m^*}\right) \geq j$. Using $\sum_{j=1}^{\infty} \frac{1}{j^2} \leq \frac{\pi^2}{6}$, we obtain

$$\sum_{i=0}^{i_{\mathrm{end}}-1} \varepsilon_i = \sum_{i=0}^{i_{\mathrm{end}}-1} \frac{\varepsilon}{4 \log_{4/3}^2\left(\frac{m_i}{m^*}\right)} \leq \sum_{j=1}^{\infty} \frac{\varepsilon}{4j^2} \leq \frac{\varepsilon}{4} \cdot \frac{\pi^2}{6} \leq \frac{\varepsilon}{2}.$$

Putting this into the right-hand side of eq. (10), we have

$$\sum_{j=1}^{i_{\mathrm{end}}} \left(\sum_{i=0}^{i_{\mathrm{end}}-1} \varepsilon_i \right)^j \leq \sum_{j=1}^{i_{\mathrm{end}}} \left(\frac{\varepsilon}{2}\right)^j \leq \frac{\frac{\varepsilon}{2}}{1 - \frac{\varepsilon}{2}} \leq \varepsilon. \tag{11}$$

By eqs. (10) and (11), $\tilde{H} = \tilde{H}_{i_{\mathrm{end}}}$ is an ε-spectral sparsifier of H.

To conclude, with probability at least $1 - O\left(\frac{1}{n}\right)$, DH-Sparsify$(H, \varepsilon)$ outputs an ε-spectral sparsifier of H with $O\left(\frac{n^2}{\varepsilon^2} \log^3 \frac{n}{\varepsilon}\right)$ hyperarcs. ◀

4.4 Total Time Complexity

We show that our algorithm runs in $O(r^2 m)$ time with probability at least $1 - O(1/n)$.

▶ **Theorem 19.** *For any directed hypergraph $H = (V, F, z)$ with the rank r and m hyperarcs and $\varepsilon \in (0,1)$,* DH-SPARSIFY(H, ε) *runs in $O(r^2 m)$ time with probability at least $1 - O(1/n)$.*

Proof. We first discuss the running time of DH-ONESTEP$(\tilde{H}_i, \lambda_i)$, where $\tilde{H}_i = (V, \tilde{F}_i, \tilde{z}_i)$ and $|\tilde{F}_i| = m_i$. It first constructs a λ_i-coreset by calling CORESETFINDER$(\tilde{H}_i, \lambda_i)$. CORESETFINDER first constructs $A^{uv} = \{ f \in F \mid C(f) \ni (u,v) \}$ for $(u,v) \in C(F)$, which is done in $O(r^2 m_i)$ time since we have $|C(f)| = O(r^2)$ for each $f \in \tilde{F}_i$. Then, for each $(u,v) \in C(F)$, it selects the λ_i heaviest hyperarcs from $A^{uv} \setminus S$ in $O(|A^{uv} \setminus S|)$ time by using a selection algorithm [5], thus taking $O\left(\sum_{(u,v) \in C(F)} |A^{uv} \setminus S|\right) = O(r^2 m_i)$ time in total. Therefore, CORESETFINDER$(\tilde{H}_i, \lambda_i)$ takes $O(r^2 m_i)$ time. After that, DH-ONESTEP samples the remaining hyperarcs in $O(m_i)$ time. Thus, DH-ONESTEP$(\tilde{H}_i, \lambda_i)$ takes $O(r^2 m_i)$ time.

We then bound the total time complexity. Since DH-SPARSIFY(H, ε) calls DH-ONESTEP$(\tilde{H}_i, \lambda_i)$ for $i = 0, 1, \ldots, T-1$ (or stops earlier), the total time complexity is at most $O\left(r^2 \sum_{i=0}^{T-1} m_i\right)$. From Lemmas 7 and 18, whenever DH-ONESTEP is called, we have $m_{i+1} \leq \frac{3}{4} m_i$ with probability at least $1 - O(1/n^2)$. This implies that $\sum_{i=0}^{T-1} m_i \leq m \sum_{i=0}^{T-1} \left(\frac{3}{4}\right)^i \leq 4m$ holds with probability at least $1 - O(T/n^2) \gtrsim 1 - O(1/n)$. Therefore, the total time complexity is bounded by $O(r^2 m)$ with probability at least $1 - O(1/n)$. ◀

References

1 Noga Alon and Joel H. Spencer. *The Probabilistic Method*. John Wiley & Sons, 2016.

2 Alexandr Andoni, Jiecao Chen, Robert Krauthgamer, Bo Qin, David P. Woodruff, and Qin Zhang. On sketching quadratic forms. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 311–319, 2016.

3 Nikhil Bansal, Ola Svensson, and Luca Trevisan. New notions and constructions of sparsification for graphs and hypergraphs. In *Proceedings of the 2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, pages 910–928. IEEE, 2019.

4 András A. Benczúr and David R. Karger. Approximating s-t minimum cuts in $\tilde{O}(n^2)$ time. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 47–55. ACM, 1996.

5 Manuel Blum, Robert W. Floyd, Vaughan R. Pratt, Ronald L. Rivest, and Robert Endre Tarjan. Time bounds for selection. *Journal of Computer and System Sciences*, 7(4):448–461, 1973.

6 Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. Optimal lower bounds for sketching graph cuts. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2565–2569. SIAM, 2019.

7 T.-H. Hubert Chan, Zhihao Gavin Tang, Xiaowei Wu, and Chenzi Zhang. Diffusion operator and spectral analysis for directed hypergraph Laplacian. *Theoretical Computer Science*, 784:46–64, 2019.

8 Yu Chen, Sanjeev Khanna, and Ansh Nagda. Near-linear size hypergraph cut sparsifiers. In *Proceedings of the 2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 61–72. IEEE, 2020.

9 Michael B. Cohen, Jonathan Kelner, John Peebles, Richard Peng, Anup B. Rao, Aaron Sidford, and Adrian Vladu. Almost-linear-time algorithms for Markov chains and new spectral primitives for directed graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 410–419. ACM, 2017.

10 Satoru Fujishige and Sachin B. Patkar. Realization of set functions as cut functions of graphs and hypergraphs. *Discrete Mathematics*, 226(1-3):199–210, 2001.

11 Giorgio Gallo, Giustino Longo, Stefano Pallottino, and Sang Nguyen. Directed hypergraphs and applications. *Discrete Applied Mathematics*, 42(2-3):177–201, 1993.

12 Matthias Hein, Simon Setzer, Leonardo Jost, and Syama Sundar Rangapuram. The total variation on hypergraphs - learning on hypergraphs revisited. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

13 Arun Jambulapati, Yang P Liu, and Aaron Sidford. Chaining, group leverage score overestimates, and fast spectral hypergraph sparsification. *arXiv:2209.10539*, 2022.

14 Mohammad Ali Javidian, Zhiyu Wang, Linyuan Lu, and Marco Valtorta. On a hypergraph probabilistic graphical model. *Annals of Mathematics and Artificial Intelligence*, 88(9):1003–1033, 2020.

15 Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Towards tight bounds for spectral sparsification of hypergraphs. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 598–611. ACM, 2021.

16 Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Spectral hypergraph sparsifiers of nearly linear size. In *Proceedings of the 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science*, pages 1159–1170, 2022.

17 Ioannis Koutis and Shen Chen Xu. Simple parallel and distributed algorithms for spectral graph sparsification. *ACM Transactions on Parallel Computing*, 3(2):1–14, 2016.

18 James R. Lee. Spectral hypergraph sparsification via chaining. *arXiv:2209.04539*, 2022.

19 Pan Li, Niao He, and Olgica Milenkovic. Quadratic decomposable submodular function minimization: Theory and practice. *Journal of Machine Learning Research*, 21(106):1–49, 2020.

20 Kazusato Oko, Shinsaku Sakaue, and Shin ichi Tanigawa. Nearly tight spectral sparsification of directed hypergraphs by a simple iterative sampling algorithm. *arXiv:2204.02537*, 2022.

21 Akbar Rafiey and Yuichi Yoshida. Sparsification of decomposable submodular functions. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, volume 36, pages 10336–10344, 2022.

22 Tasuku Soma and Yuichi Yoshida. Spectral sparsification of hypergraphs. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2570–2581. SIAM, 2019.

23 Daniel A. Spielman. Spectral and Algebraic Graph Theory, 2019. (visited on 02/03/2022). URL: `http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf`.

24 Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

25 Yuuki Takai, Atsushi Miyauchi, Masahiro Ikeda, and Yuichi Yoshida. Hypergraph clustering based on pagerank. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1970–1978. ACM, 2020.

26 Shang-Hua Teng. Scalable algorithms for data and network analysis. *Foundations and Trends in Theoretical Computer Science*, 12(1-2):1–274, 2016. `doi:10.1561/0400000051`.

27 Nisheeth K. Vishnoi. *$Lx = b$*. *Foundations and Trends in Theoretical Computer Science*, 8(1-2):1–141, 2013. `doi:10.1561/0400000054`.

28 Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. HyperGCN: A new method of training graph convolutional networks on hypergraphs. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

29 Naganand Yadati, Vikram Nitin, Madhav Nimishakavi, Prateek Yadav, Anand Louis, and Partha Talukdar. NHP: Neural hypergraph link prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1705–1714. ACM, 2020.

30 Yuichi Yoshida. Cheeger inequalities for submodular transformations. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2582–2601. SIAM, 2019.

31 Chenzi Zhang, Shuguang Hu, Zhihao Gavin Tang, and T.-H. Hubert Chan. Re-revisiting learning on hypergraphs: Confidence interval, subgradient method, and extension to multiclass. *IEEE Transactions on Knowledge and Data Engineering*, 32(3):506–518, 2020.

32 Chunjiang Zhu, Sabine Storandt, Kam-Yiu Lam, Song Han, and Jinbo Bi. Improved dynamic graph learning through fault-tolerant sparsification. In *Proceedings of the 36th International Conference on Machine Learning*, pages 7624–7633. PMLR, 2019.