



מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree
Master of Science

עבודת גמר (תזה) לתואר
מוסמך למדעים

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Yotam Kenneth

מאת
יותם קנת

דילול חתכים וייצוג תמציתי של היפגרפים סאבמודולריים
Cut Sparsification and Succinct Representation of
Submodular Hypergraphs

Advisor:
Prof. Robert Krauthgamer

מנחה:
פרופ' רוברט קראוטגמר

September 2023

אלול ה'תשפ"ג

Abstract

Modern data analysis requires sifting through ever-increasing amounts of observations, this increases the computational and storage cost of analyzing the data. One solution to this challenge is sparsification, constructing a smaller data structure that preserves some desired property of the original data. The main challenge in sparsification is constructing efficient algorithms that minimize the size of the resulting data structure.

In cut sparsification, all cuts of a hypergraph $H = (V, E, w)$ are approximated within $1 \pm \epsilon$ factor by a small hypergraph H' . This widely applied method was generalized recently to a setting where the cost of cutting each $e \in E$ is provided by a splitting function, $g_e : 2^e \rightarrow \mathbb{R}_+$. This generalization is called a submodular hypergraph when the functions $\{g_e\}_{e \in E}$ are submodular, and it arises in machine learning, combinatorial optimization, and algorithmic game theory. Previous work focused on the setting where H' is a reweighted sub-hypergraph of H , and measured size by the number of hyperedges in H' . We study such sparsification, and also a more general notion of representing H succinctly, where size is measured in bits.

In the sparsification setting, where size is the number of hyperedges, we present three results: (i) all submodular hypergraphs admit sparsifiers of size polynomial in $n = |V|$; (ii) monotone-submodular hypergraphs admit sparsifiers of size $O(\epsilon^{-2}n^3)$; and (iii) we propose a new parameter, called spread, to obtain even smaller sparsifiers in some cases.

In the succinct-representation setting, we show that a natural family of splitting functions admits a succinct representation of much smaller size than via reweighted subgraphs (almost by factor n). This large gap is surprising because for graphs, the most succinct representation is attained by reweighted subgraphs. Along the way, we introduce the notion of deformation, where g_e is decomposed into a sum of functions of small description, and we provide upper and lower bounds for deformation of common splitting functions.

Acknowledgements

First, I would like to thank Prof. Robert Krauthgammer, my advisor, for his support throughout the research process. His insights on algorithms, theory and (just as important) the research process were invaluable and allowed me to create a thesis that I am proud of. Furthermore, his focus on clear and concise writing helped me make my research more accessible.

I was also fortunate to be in the company of my fellow students, especially Shay Sapir. Life allotted me only a few months of full time research at the institute but they made it feel like home.

I would also like to thank my family for supporting me through the highs and lows of the research. In particular, my partner Shay who was always there for me. Finally, a special thank you to my grandfather, Nino Levy, I would not be here if he did not insist on me memorizing that $7 \times 8 = 56$.

1 Introduction

A powerful tool for many graph problems is sparsification, where one constructs from an input graph a much smaller graph that preserves (perhaps approximately) certain properties, for example the input graph’s cuts [BK96] or its spectrum [ST11]. Algorithms for downstream applications can then be executed on the small graph, which improves the overall running time, and the small graph can also be stored (or sent to another site) instead of the input graph, which improves the memory (or communication) requirements. The extensive research on cut sparsification has started with the seminal work of Benczúr and Karger on cuts in graphs [BK96], and was later extended to hypergraphs [KK15, BST19, CKN20] and to directed hypergraphs [SY19, CCPS21, KKTY21, OST23]. We focus on sparsifying hypergraphs, and in fact more general objects, while preserving all their cuts.

In recent years, the notion of cuts in a weighted hypergraph $H = (V, E, w)$ has been generalized even further, to a setting where each hyperedge $e \in E$ has a *splitting function* $g_e : 2^e \rightarrow \mathbb{R}_+$, and the *value* of a cut $S \subseteq V$ is defined as

$$\text{cut}_H(S) := \sum_{e \in E} g_e(S \cap e). \quad (1)$$

Associating every $e \in E$ with the *all-or-nothing* splitting function, given by $g_e^{\text{aon}} : S \mapsto w_e \cdot \mathbb{1}_{\{S \neq \emptyset, e\}}$, clearly models an ordinary hypergraph $H = (V, E, w)$, where the value of a cut is the total weight of hyperedges that intersect both sides; in fact, a simple extension can model a directed hypergraph. Such a generalized hypergraph $H = (V, E, g)$, where $g = \{g_e\}_{e \in E}$, is called a *submodular hypergraph* if all its splitting functions g_e are submodular. Recall that a set function $g : 2^e \rightarrow \mathbb{R}_+$ is *submodular* if

$$\forall S, T \subseteq e, \quad g(S \cup T) + g(S \cap T) \leq g(S) + g(T).$$

Submodular hypergraphs are useful in clustering data with higher-order relations that are not captured by ordinary hyperedges [LM17, LM18, VBK20, LVS+21, VBK21, ZLS22]. For example, the *small-side splitting* function, given by $g_e^{\text{sm}} : S \mapsto \min(|S|, |\bar{S}|)$, is employed when unbalanced cuts are preferable. Cut functions of submodular hypergraphs were studied also under a different name of *decomposable submodular functions*. A submodular function $f : 2^V \rightarrow \mathbb{R}_+$ is called *decomposable* if it can be written as $f = \sum_i f_i$, where each $f_i : 2^V \rightarrow \mathbb{R}_+$ is submodular. This notion is widely applied in data summarization [GK10, LB11, TIWB14], where each f_i is a submodular similarity function, and the task of summarizing the data under a given budget k is modeled by maximizing $f(S)$ over all $S \subset V$ of size $|S| \leq k$. Decomposable submodular functions arise also in welfare maximization, where each agent has a submodular utility function, for instance in approximation algorithms [Fei09, FV06] and in truthful mechanisms [DS06, AS20].

We study how to succinctly represent all the cuts of a submodular hypergraph H up to $1 \pm \epsilon$ factor. We examine two complementary approaches: (1) *sparsification*, which reduces the number of hyperedges, i.e., H is represented using a sparse H' ; and (2) *deformation*, which replaces large hyperedges or complicated splitting functions by new ones of low space complexity, i.e., H is represented using H' whose hyperedges can be stored succinctly. These approaches can yield (separately and/or together) a sparsifier H' that can be encoded using a small number of bits. More generally, we may consider a general encoding that *does not* rely on a sparsifier H' , e.g., an explicit list of all the $2^{|V|}$ cut values.

Let us introduce some basic notation to make the discussion more precise. Throughout, let $n := |V|$; we write $\tilde{O}(t)$ or $\tilde{\Omega}(t)$ to suppress a polylogarithmic factor in t , and $O_\alpha(t)$ or $\Omega_\alpha(t)$ to hide a factor that depends only on α .

Definition 1.1 (Sparsifier). A cut sparsifier of quality $1 + \epsilon$ for $H = (V, E, g)$, or in short a $(1 + \epsilon)$ -sparsifier, is a submodular hypergraph $H' = (V, E', g')$ such that

$$\forall S \subseteq V, \quad \text{cut}_{H'}(S) \in (1 \pm \epsilon) \cdot \text{cut}_H(S). \quad (2)$$

The size of the sparsifier is $|E'|$, the cardinality of its hyperedge set. We call H' a reweighted subgraph of H if $E' \subseteq E$ and each function g'_e for $e \in E'$ is a scaling of g_e (i.e., $g'_e = s_e g_e$ for some $s_e > 0$).

Question 1.2 (Sparsification). Do all submodular hypergraphs admit a reweighted-subgraph sparsifier with few hyperedges, say $\text{poly}(\epsilon^{-1}n)$? And which families of splitting functions admit even smaller sparsifiers, like $\tilde{O}_\epsilon(n^2)$ or even $\tilde{O}_\epsilon(n)$?

The first part (i.e., the general question) is known for several families of splitting functions (see Section 1.1 for a detailed account). But despite this significant progress, the general question was left open in [RY22], where the bound on the sparsifier size depends on g and is exponential in n in the worst case. We answer the first part positively by proving that every submodular hypergraph admits a polynomial-size sparsifier. We address also the second part by showing that several families of splitting functions admit even smaller sparsifiers.

We further ask about a more general notion, of a data structure that approximates the cuts of H , which can potentially be more succinct than a sparsifier.

Question 1.3 (Succinct Representation). What is the smallest data structure (in bits of space) that stores a submodular hypergraph H so as to report $(1 + \epsilon)$ -approximation to every cut value? In particular, what is the smallest number of bits $s = s(\epsilon, n)$ that suffices to store a sparsifier for H ?

For simplicity, we ask above only about the existence of a sparsifier or a data structure, but we are of course interested also in fast algorithms to build them. Fortunately, an algorithmic solution follows from the existential ones because our proofs are constructive. Furthermore, running times are polynomial under the assumption of integral weights and $\min_{S \subseteq e} g_e(S) \leq \text{poly}(n)$ ¹.

1.1 Sparsification: All Submodular Hypergraphs

We start by addressing Question 1.2. Our first result (proved in Section 2) shows that all submodular hypergraphs admit a sparsifier of polynomial size (in $\epsilon^{-1}n$). This is the first polynomial bound for all submodular splitting functions, and the previously known bound, due to [RY22], was $O_\epsilon(n^2 B_H)$, where $B_H := \max_{e \in E} |\mathcal{B}(g_e)|$ and $\mathcal{B}(g_e)$ is the set of extreme points in the polytope of g_e .² In general, B_H can be exponential in n , for example small-side splitting g_e^{sml} has $B_e = 2^{\Theta(|e|)}$.

Theorem 1.4. Every submodular hypergraph admits a $(1 + \epsilon)$ -sparsifier of size $O(\epsilon^{-2}n^5)$.

The existing literature shows that several families of splitting functions admit even smaller sparsifiers than the above bound. For example, sparsifiers of size $\tilde{O}_\epsilon(n)$ are known for all-or-nothing splitting g_e^{aon} [CKN20] (see also [Qua22]) and for product splitting, given by $g_e^{\text{prd}} : S \mapsto |S| \cdot |\bar{S}|$ [dCSHS16]. In contrast, for the splitting that models cuts in a directed hypergraph, the known construction has size $\tilde{O}_\epsilon(n^2)$ [OST23], which is near-tight with an $\Omega(n^2/\epsilon)$ lower bound [OST23];

¹The running times of Theorem 1.4, Theorem 1.5, and Theorem 1.10 are polynomial in general. Theorem 1.7 is polynomial under the stated assumption.

²A recent manuscript [KZ23] claims that the proof in [RY22] has a flaw and holds only for monotone submodular hypergraphs.

this function, called *directed all-or-nothing splitting*, is given by $g_e^{\text{d-aon}} : S \mapsto \mathbb{1}_{\{e_T \cap S \neq \emptyset \wedge e_H \not\subseteq S\}}$, where $e_H, e_T \subseteq e$ are the hyperedge’s head and tail, respectively.

Very recently, it was shown that the entire family of symmetric splitting functions admits sparsifiers of size $\tilde{O}_\epsilon(n)$ [JLLS23]. We show (in Section 2) that another large family, of *monotone* splitting functions (defined as those where $g_e(S) \leq g_e(T)$ for all $S \subseteq T$), admits an improved sparsifier. Monotone functions arise naturally in many applications, however no sparsification bound was previously known for any function in this family.³

Theorem 1.5. *Every hypergraph with monotone splitting functions admits a $(1 + \epsilon)$ -sparsifier of size $O(\epsilon^{-2}n^3)$.*

Related Work. Previous work on sparsification focused mostly on specific splitting functions. The study of this problem began with sparsifiers for undirected graph cut; the current size bound is $O(\epsilon^{-2}n)$ edges [BSS14], which improves over [BK96] and is known to be tight [ACK⁺16, CKST19]. Later work on undirected hypergraphs has matched this size bound up to $O(\log n)$ factor [KK15, BST19, CKN20, Qua22]. In the directed case, graphs have a trivial upper bound of $O(n^2)$ edges which is tight (regardless of ϵ). Recent work on hypergraphs has culminated with $(1 + \epsilon)$ -sparsifiers with $\tilde{O}_\epsilon(n^2)$ hyperedges [SY19, CCPS21, KKTY21, OST23], which is near-tight with an $\Omega(n^2/\epsilon)$ lower bound [OST23]. Finally, several symmetric functions were known to admit polynomial or near-linear sparsifiers even before the recent result for the entire family of symmetric functions [JLLS23]; these include the aforementioned *small-side splitting* [AGK14, ADK⁺16] and product splitting [dCSHS16]. Figure 1 depicts several families of splitting functions and the sparsification bounds known for them, including our results from above and from Section 1.2.

Techniques. Our sparsification method follows the importance-sampling approach, which has been used extensively in the literature, including in [RY22] for submodular hypergraphs, which uses a straightforward calculation of the importance of a hyperedge, based on all its possible cuts. We devise a more sophisticated approximation of the importance of a hyperedge, based on the minimum directed cut between all pairs $u, v \in e$, given by $g_{\min}^{u \rightarrow v} := \min_{S \subseteq V: v \in S, u \notin S} g_e(S)$. The resulting sampling probabilities produce a much smaller sparsifier, because now each importance is bounded using only $O(n^2)$ directed cuts instead of all the extreme points of its polytope, and the number of relevant terms (directed cuts or extreme points) eventually propagates into the sparsifier size. We then prove that our probabilities suffice to approximate all the cuts by showing that the sum of these directed cuts provide an upper bound on the maximum cut of g_e by an analogy to directed flow in graphs. The proof for the family of all monotone functions is similar, but the monotone structure allows us to calculate the importance based only on singletons, i.e., cuts of the form $\text{cut}_H(\{v\})$, which results in an n^2 improvement in size over the general case.

1.2 Sparsification: Parameterized by Spread

We already know that submodular splitting functions can have very different optimal sparsification bounds, see e.g. the bounds $\tilde{\Theta}_\epsilon(n)$ and $\tilde{\Theta}_\epsilon(n^2)$ mentioned above. However, there are too many submodular functions to analyze each one separately, and we thus seek a parameter that can control the sparsifier size. Our approach is inspired by the notion of imbalance in a directed graph $G = (V, E, w)$, defined as the worst ratio between antiparallel edge weights, i.e., $\beta_G := \max\{w(i, j)/w(j, i) : i, j \in V\}$. This parameter can be used to show that every directed

³The running time of [RY22] was improved in [KZ23], where a sparsifier of size $O(Bn^2)$ for monotone functions with low curvature is constructed in polynomial time.

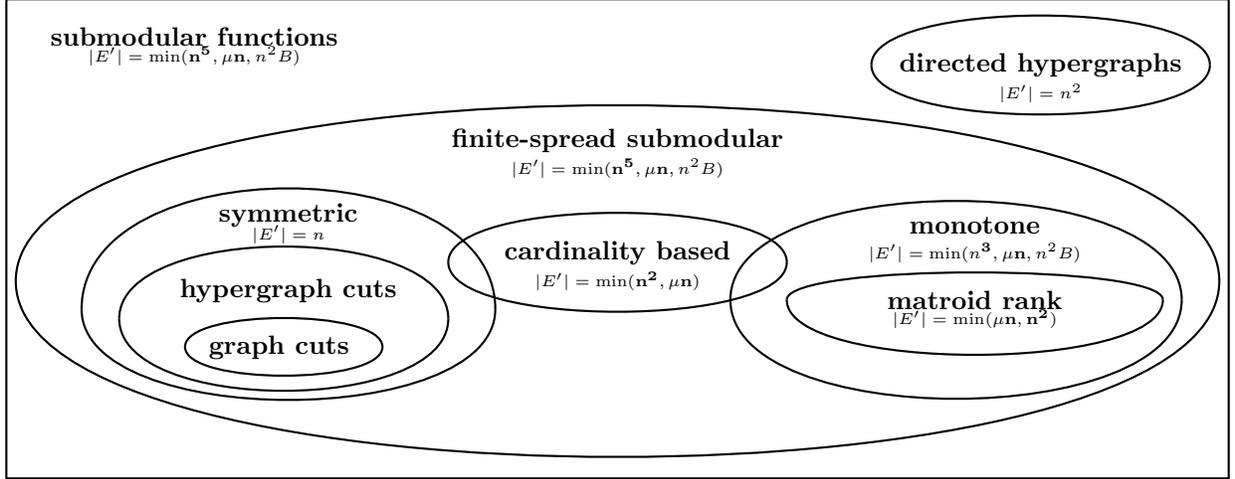


Figure 1: Sparsification bounds for various families of submodular functions, omitting for simplicity $\text{poly}(\epsilon^{-1} \log n)$ factors.

graph admits a sparsifier of size $\tilde{O}_\epsilon(\beta_G n)$.⁴ For submodular hypergraphs, we propose an analogous parameter, which is basically the ratio between the maximum and minimum values of the splitting function, excluding certain trivial cuts.

Definition 1.6 (Spread). *The spread of a hyperedge $e \in E$ with a splitting function g_e is*

$$\mu_e := \frac{\max_{T \subseteq e} g_e(T)}{\min_{S \subseteq e: S \notin W} g_e(S)}, \quad (3)$$

where $W := \{\emptyset\}$, unless $g_e(e) = 0$ in which case $W := \{\emptyset, e\}$.

Our third result (proved in Section 3) constructs a sparsifier whose size depends on the spread of the input H , defined as $\mu_H := \max_{e \in E} \mu_e$. By convention, the spread μ_e is called *finite* if it is well-defined (the denominator in (3) is non-zero), and similarly μ_H is called finite if it is well-defined (all the terms μ_e are finite).

Theorem 1.7 (Sparsifier Parameterized by Spread). *Every submodular hypergraph $H = (V, E, g)$ with finite spread admits a $(1 + \epsilon)$ -sparsifier of size $\tilde{O}(\epsilon^{-2} \mu_H n)$.*

Many natural submodular functions have finite spread, and in many common cases even $\mu_H \leq n$. This can be seen, for example, in an easy application of Theorem 1.7 to approximation of coverage functions, see Appendix B for details. Another example is the sparsification of the capped version of small-side splitting, given by $g_e : S \mapsto \min(|S|, |\bar{S}|, K)$ for $K > 0$, which clearly has spread $\mu_e \leq K$. This function is part of a much larger family, cardinality-based splitting functions, a notion formalized in [VBK22] as follows: A submodular function $g_e : 2^e \rightarrow \mathbb{R}_+$ is called *cardinality-based* if there exists a function $f_e : [|e|] \rightarrow \mathbb{R}_+$ such that $g_e : S \mapsto f_e(|S|)$. Cardinality-based functions, which are commonly used in submodular hypergraph clustering, all have spread $\mu_e \leq n$, which is an easy consequence of the symmetry and subadditivity of g_e . By Theorem 1.7, these splitting function admit a $(1 + \epsilon)$ -sparsifier of size $\tilde{O}(\epsilon^{-2} n^2)$, which is the first bound for this family.

⁴As a side comment, this condition can be relaxed significantly to $\beta_G := \max\{\text{cut}_G(S)/\text{cut}_G(\bar{S}) : S \subset V\}$, and the same sparsification bound still holds [CCPS21].

It is easily verified that for monotone splitting functions, the spread is approximately equal to the imbalance, when we generalize the imbalance from above to hyperedges by $\beta_e := \max\{g_e(S)/g_e(\bar{S}) : S \subset V\}$.⁵ Hence, we immediately obtain the following.

Corollary 1.8. *Every finite-spread monotone splitting function admits a $(1 + \epsilon)$ -sparsifier of size $\tilde{O}(\epsilon^{-2}\beta_H n)$.*

Two other examples of commonly used monotone functions with finite spread are set-coverage functions (defined in Appendix B) and the *matroid-rank functions*,⁶ which have $\mu_e = r$ where r is the rank of the matroid.

We remark that spread does not fully characterize the sparsifier size. Indeed, symmetric functions can have a large spread μ_e but still admit $\tilde{O}_\epsilon(n)$ sparsifier due to [JLLS23], consider e.g. product splitting g_e^{prd} which has $\mu_e = O(n)$. Furthermore, directed all-or-nothing splitting $g_e^{\text{d-aon}}$ does not have finite spread, and nevertheless admits a sparsifier of size $\tilde{O}_\epsilon(n^2)$ [OST23]. Figure 1 depicts different families of splitting functions including that of finite spread, and the sparsification bounds known for them.

Techniques. Our technique is based on approximate H as an undirected hypergraph and use the sampling probabilities of [CKN20] but amplified by μ_e for each hyperedge. This is a known technique in generalizing sampling mechanisms. Our main contribution is to identify the spread as a relevant and useful parameter. We remark that the generalization of balance, which is known to control the size of sparsifier in directed graphs, to submodular hypergraphs does not suffice for sparsification. Furthermore, we prove that the spread also characterizes other traits of splitting function, such as the deformation lower bound.

1.3 Succinct Representation

We provide the first example of submodular splitting functions for which sparsifiers that are not subgraphs are provably (much) more succinct than sparsifiers that are reweighted subgraphs.⁷ To be more precise, we exhibit a natural family of splitting functions, where the former $(1 + \epsilon)$ -sparsifiers take only $\tilde{O}_\epsilon(n)$ bits (Corollary 1.11), while the latter $(1 + \epsilon)$ -sparsifiers require $\tilde{\Omega}_\epsilon(n^2)$ bits (Theorem 1.12). It follows that a reweighted subgraph *need not be* the smallest data structure that stores a $(1 + \epsilon)$ -approximation of the cuts values, and by a wide margin!

Our plan for constructing a succinct representation has two stages. The first stage creates a $(1 + \epsilon)$ -sparsifier H' , by deforming each $e \in E$ into multiple small hyperedges. The second stage computes for this H' a $(1 + \epsilon)$ -sparsifier H'' that is a reweighted subgraph. It then follows that H'' is a $(1 + \epsilon)^2$ -sparsifier, and has a few hyperedges that are all small.

Definition 1.9. *A splitting function $g_e : 2^e \rightarrow \mathbb{R}_+$ on hyperedge e is called $(1 + \epsilon)$ -approximable with support size p if there are submodular functions $g_{e_i} : 2^{e_i} \rightarrow \mathbb{R}_+$ for $i = 1, \dots, r$, each on a hyperedge $e_i \subseteq e$ of size $|e_i| \leq p$, such that*

$$\forall S \subseteq e, \quad \sum_{i=1}^r g_{e_i}(S \cap e_i) \in (1 \pm \epsilon)g_e(S).$$

⁵For a monotone g_e , the spread is $\mu_e = g_e(V)/\min_{v \in V} g_e(\{v\})$ and the imbalance is $\beta_e = \max_{v \in V} g_e(V \setminus \{v\})/g_e(\{v\})$, and they differ by at most a constant factor by the subadditivity of g_e .

⁶For a matroid with ground set e and independent sets \mathcal{I} , the rank function is given by $g_e : S \mapsto \max_{T \subseteq S: T \in \mathcal{I}} |T|$. This rank function is submodular and monotone.

⁷Previously, a non-subgraph sparsifier was shown in [ADK⁺16] for small-side splitting, however it optimizes the number of hyperedges and not the encoding size.

(We use the term support size to avoid confusion with sparsifier size.)

Our example is the family of *additive* splitting functions, defined as functions g_e that can be written as either $g_e : S \mapsto \min(|S|, K)$ or $g_e : S \mapsto \min(|S|, |\bar{S}|, K)$ for some $K > 0$. The next theorem (proved in Section 4) achieves the first stage in our plan above; it shows that additive functions can be $(1 + \epsilon)$ -approximated by creating several copies of e and sampling the vertices.

Theorem 1.10 (Deformation of Additive Functions). *Let g_e be an additive splitting function on hyperedge e . Then g_e can be $(1 + \epsilon)$ -approximated with support size $O(\epsilon^{-2}(|e|/K) \log |e|)$.*

Following our plan, suppose that given an input H , we first apply Theorem 1.10 to obtain a sparsifier H' with small support size. The construction of H' also implies that it has small spread, $\mu_{H'} \leq O(\epsilon^{-2} \log n)$. Applying Theorem 1.7 on H' we obtain a succinct representation H'' . A straightforward encoding of H'' then proves the following corollary (see Section 4).

Corollary 1.11 (Additive Functions admit Small Representation). *Let $H = (V, E, \{g_e\})$ be a submodular hypergraph such that every g_e is additive with parameter $K_e > 0$, and let $\hat{K} := \min_{e \in E} K_e / |e|$ be a normalized bound on K_e over all hyperedges. Then H admits a $(1 + \epsilon)$ -sparsifier with encoding size $O(\epsilon^{-6} \hat{K}^{-1} n \log^4 n)$ bits.*

The next theorem (proved in Section 5) shows that reweighted-subgraph sparsifiers of additive functions require $\Omega(n^2)$ bits in the worst-case. Putting this together with our succinct representation from Corollary 1.11, we conclude that relaxing the (natural) restriction to reweighted subgraphs improves the space complexity by a factor of $\tilde{\Omega}_\epsilon(n\hat{K})$, observe that this can be $\tilde{\Omega}_\epsilon(n)$ when $\hat{K} \in \Omega(1)$.

Theorem 1.12 (Reweighted Sparsifiers Require $\Omega(n^2)$ Bits). *There exists a family \mathcal{H} of hypergraphs with additive splitting functions with parameter $1 \leq K \leq n/3$, such that encoding a reweighted-subgraph $(1 + \epsilon)$ -sparsifier for an input $H \in \mathcal{H}$ requires $\Omega(n^2)$ bits.*

This lower bound is surprising because in the case of undirected graphs, the best encoding size is achieved by a reweighted-subgraph sparsifier [BSS14, ACK+16, CKST19]. Our proof is based on a technical lemma that can be applied to many cardinality-based splitting functions. Furthermore, Theorem 1.12 can be extended to the directed all-or-nothing splitting function $g_e^{\text{d-aon}}$, to show a lower bound of $\Omega(n^3/\epsilon)$ bits. For details see Section 5.

Finally, we can also prove a space lower bound for an arbitrary encoding of cuts in a directed hypergraph (arbitrary means that it need not represent a reweighted-subgraph sparsifier, see Section 5 for details). This proof provides an ϵ^{-1} factor improvement over the trivial lower bound of $\Omega(n^2)$ bits. The proof combines the techniques from Theorem 1.12 with a lower bound from [OST23] on the number of edges in a reweighted-subgraph sparsifier.

Theorem 1.13. *There exists a family of directed hypergraphs \mathcal{H} such that encoding a $(1 + \epsilon)$ -approximation of their cuts requires $\Omega(n^2/\epsilon)$ bits.*

Techniques. Our lower bound for the encoding size of reweighted-subgraph sparsifiers (Theorem 1.12) boils down to a counting argument on a large family of hypergraphs \mathcal{H} , that have sufficiently different cut values and thus require distinct encodings. We construct hypergraphs in this family \mathcal{H} by partitioning the vertices into three parts V, U, W , and adding hyperedges that contain vertices from all three parts. We first create hyperedges consisting of a large random subset of vertices from V ; this adds entropy that will differentiate between hypergraphs in \mathcal{H} . We then augment each hyperedge with vertices from U , where each hyperedge is defined by a word in the

Function Family	Example	Support Size	
additive functions	$g_e(S) = \min(S , K)$	$\Omega(n/K)$	Lemma 1.14
polynomial	$g_e(S) = S ^\alpha$ for constant $\alpha \in (0, 1)$	$\Omega(n)$	Corollary 6.3
logarithmic	$g_e(S) = \log(S + 1)$	$\Omega(n)$	Corollary 6.4
cardinality based	$g_e(S) = f(S)$ for concave f	$\Omega(n/\mu_e^{1.5})$	Corollary 6.5
unweighted	$g_e(v) = 1$ for all $v \in V$	$\Omega(n/\mu_e^3)$	Corollary 6.6

Table 1: Our lower bounds on the support size for several families of splitting functions. They are all obtained by applying Lemma 6.1, stated for simplicity for sufficiently small fixed $\epsilon > 0$ and $|e| = n$.

Hadamard code. We use the structure of this code to show that by making cut queries to a hypergraph $H \in \mathcal{H}$, one can recover the random bits encoded in the adjacency matrix of H induced on V . We use W to create an unsparsifiable hypergraph, i.e., one where removing any hyperedge will violate the approximation guarantee. Finally, every hyperedge on $V \cup U$ is combined with a hyperedge on W .

1.4 Deformation Lower Bounds

Our success in finding a small succinct representation for additive functions motivates searching for deformations of other splitting functions.

A similar problem, of approximating a submodular function by functions of small support but over the uniform distribution (i.e., in average-case rather than worst-case), has received significant attention [FKV13, CKKL12, GHRU13, FK14, FV16], and it is known that every submodular function $f : 2^V \rightarrow [0, 1]$ can be approximated within additive error ϵ using support size $O(\epsilon^{-2} \log \epsilon^{-1})$ [FV16]. We show (see Section 6) that a similar result is unfortunately not possible in our setting (multiplicative error for worst-case approximation).

Theorem 1.14 (Approximation Requires Large Support Size). *Let g_e be an additive splitting function on a hyperedge e . Then every 1.1-approximation of g_e must have support size $p \geq \Omega(|e|/K)$.*

Techniques. The proof of Theorem 1.14 is based on Lemma 6.1, a technical result that can be applied to many splitting functions. The main idea is to examine a certain quantity δ_t , which is related the notion of curvature (of a submodular function). The curvature is often used to parameterize approximation guarantees in maximization of submodular optimization [CC84, Von10]. Intuitively, both the curvature and δ_t characterize the locality of the function, i.e., how much error is introduced by decomposing the function into smaller parts and summing them. The main difference between the two quantities is that the curvature looks at the marginal contributions and δ_t characterizes the curvature of the union of two sets of size t . Furthermore, in the approximation setting, a low worst-case curvature is desirable while for our proof it suffices that δ_t is high for many sets of size t . Specifically, we show that if a constant fraction of pairs of subsets of size t have constant positive δ_t , then g_e cannot be approximated with support size smaller than $O(\delta_t^2 n/t)$.

By applying Lemma 6.1, we obtain lower bounds on the support size required to approximate several natural splitting functions, as presented in Table 1.

1.5 Related Work

Submodular functions appear in many applications, and have been studied extensively in the literature. In particular, the problem of finding a simple representation for submodular functions

has been studied in several works. An $O(\sqrt{n} \log n)$ -approximation for monotone submodular functions by functions of the form $f(S) = \sqrt{\sum_{v \in S} c_v}$, where $c_v > 0$ are weights for all $v \in V$, was obtained in [GHIM09]. A later result [DDS⁺13] showed the same approximation using coverage and budget-additive functions. The same paper also provided a lower bound of $\Omega(n^{1/3} \log^{-2} n)$ for approximating monotone submodular functions by coverage and budget additive. Approximating the all-or-nothing splitting function on n vertices using hyperedges with the all-or-nothing function and with support size r must incur approximation factor $\Omega(n/r)$ [Pog17, Section 2.3].

It was previously shown that every symmetric cardinality-based splitting functions can be deformed into a sum of $|e|/2$ hyperedges with capped small-side splitting function, while preserving the value of g_e exactly [VBK22]. Subsequent work by the same authors [VBK21], achieves a similar deformation but with $(1 + \epsilon)$ -approximation and using only $O(\epsilon^{-1} \log |e|)$ hyperedges. Notice the difference from our work, which focuses on an approximation with small support size.

1.6 Concluding Remarks

Our work provides several promising directions for future work. We prove that all submodular hypergraph admit sparsifiers of polynomial size (Theorem 1.4) and provide smaller sparsifiers for several families (monotone and finite-spread), however characterizing the optimal sparsifier size for each family remains open. In the succinct-representation setting, we found a useful deformation only for additive splitting functions (Theorem 1.10), and it would be desirable to find deformations for more families.

Another interesting avenue is to find applications or connections to other problems. For example, we know that Theorem 1.7 can be used to approximate a set-coverage function using a small ground set, see Appendix B for details. Another potential application is constructing succinct representations for terminal-cut functions, see Appendix C.

2 Polynomial Sparsifiers for Submodular Hypergraphs

This section proves Theorems 1.4 and 1.5. Both proofs are based on importance sampling similarly to the method employed in [RY22], where the main idea is to sample hyperedges proportionally to their maximum contribution to any cut. Our proof differs in the way that sampling probabilities are assigned to the hyperedges. In the proof of [RY22], the sampling probabilities are proportional to

$$\rho_e := \max_{S \subseteq e} \frac{g_e(S \cap e)}{\sum_{e \in E} g_e(S \cap e)},$$

and thus the size of the sparsifier is proportional to the total importance $\sum_{e \in E} \rho_e$. The exponential bound results from bounding each ρ_e by a sum over all the extreme points of the polytope of g_e . We show that it is sufficient to consider only a small subset of the cuts, namely the minimum directed cuts between all pairs of vertices in e . Formally, the minimum directed cut between $(u, v) \in e \times e$ is defined as

$$g_e^{v \rightarrow u} := \min_{S \subseteq e: v \in S, u \notin S} g_e(S), \quad (4)$$

where by convention $g_e^{v \rightarrow u} := 0$ whenever the minimization is not defined (e.g., $v \notin e$ or $u = v$). Our sampling probabilities are proportional to

$$\rho_e := \max_{(u,v) \in e \times e} \frac{g_e^{v \rightarrow u}}{\sum_{f \in E} g_f^{v \rightarrow u}},$$

which can be easily bounded by replacing the maximum by a sum over all $O(n^2)$ pairs. The main technical difficulty is in showing that these ρ_e suffice for the sampling to approximate all the cuts of H . We do that by showing that the maximum cut of each hyperedge e is at most n^2 times the largest minimum directed cut in e , which is proved by analogy to directed flows in graphs.

Finally, the directed cuts $g_e^{v \rightarrow u}$ can be computed in polynomial time using standard submodular minimization techniques [McC05]. This improves on the exponential running time required by [RY22] because maximizing the ratio of two submodular functions is NP-hard. In the case of monotone splitting functions, a polynomial running time is achievable by approximating the ratio up to factor \sqrt{n} , which increases the size of the sparsifier by the same factor [RY22]. This was recently improved by [KZ23] to allow polynomial running time without increasing the sparsifier size for the case of monotone functions with low curvature.⁸

Proof of Theorem 1.4. Our construction of a quality $(1 + \epsilon)$ -sparsifier for H uses the importance sampling method, where each hyperedge is sampled independently with some probability p_e , and the splitting functions of the sampled hyperedges are then scaled by factor $1/p_e$. We begin by describing our method for assigning sampling probabilities to the hyperedges of H . Recall that the minimum directed cut on the hyperedge e between $(u, v) \in e \times e$ is given by (4), but it is 0 in exceptional cases. Note that the cost to separate v from u may be strictly positive even if $u \notin e$, for example in monotone functions. Let $P_{v \rightarrow u}^e \subseteq e$ be a set that attains the minimum for $g_e^{v \rightarrow u}$.

We denote $\bar{S} = e \setminus S$ when e is clear from context. Throughout the proof we will use the following claim to bound cuts of H by minimum directed cuts.

Claim 2.1. *For every $e \in E$ and $S \subset e$,*

$$g_e(S \cap e) \leq \sum_{v \in S} \sum_{u \in \bar{S}} g_e^{v \rightarrow u}.$$

The proof of Claim 2.1 appears later. Intuitively, it is similar to bounding the capacity of a cut in a graph by the sum of all pairwise flows between S, \bar{S} . We now proceed assuming this claim.

Denote the set of all pairs whose minimum directed cut on e is strictly positive by $V_e := \{(u, v) \in e \times e : g_e^{v \rightarrow u} > 0\}$. For every hyperedge $e \in E$, set (for $M > 0$ that will be chosen later)

$$p_e := \begin{cases} M \cdot \max \left(\frac{g_e(e)}{\sum_{f \in E} g_f(f)}, \max_{(v, u) \in V_e} \frac{g_e^{v \rightarrow u}}{\sum_{f \in E} g_f^{v \rightarrow u}} \right) & \text{if } g_f(f) > 0 \text{ for some } f \in E; \\ M \cdot \max_{(v, u) \in V_e} \frac{g_e^{v \rightarrow u}}{\sum_{f \in E} g_f^{v \rightarrow u}} & \text{otherwise.} \end{cases}$$

Now sample every hyperedge e independently with probability $\tilde{p}_e = \min(p_e, 1)$ and rescale the splitting functions of every sampled hyperedge by factor $1/\tilde{p}_e$. To prove the theorem, we need to show that the resulting H' accurately approximates all cuts of H , and that it contains at most $O(\epsilon^{-2}n^5)$ hyperedges.

We first prove that the number of hyperedges in the sparsifier H' is at most $O(M \cdot n^2)$. Let I_e be an indicator for the event that the hyperedge e is sampled into H' . The expected number of

⁸The curvature of g_e is defined as $c_{g_e} := 1 - \min_{S \subseteq e, v \in e \setminus S} (g_e(S \cup \{v\}) - g_e(S \cap e)) / (g_e(\{v\}) - g_e(\emptyset))$.

sampled hyperedges is

$$\begin{aligned}
\mathbb{E} \left[\sum_{e \in E} I_e \right] &= \sum_{e \in E} p_e = M \sum_{e \in E} \max \left\{ \frac{g_e(e)}{\sum_{f \in E} g_f(f)}, \max_{(v,u) \in V_e} \frac{g_e^{v \rightarrow u}}{\sum_{f \in E} g_f^{v \rightarrow u}} \right\} \\
&\leq M \frac{\sum_{e \in E} g_e(e)}{\sum_{f \in E} g_f(f)} + M \sum_{e \in E} \sum_{(v,u) \in V_e} \frac{g_e^{v \rightarrow u}}{\sum_{f \in E} g_f^{v \rightarrow u}} \\
&= M + M \sum_{(v,u) \in \bigcup_{e \in E} V_e} \frac{\sum_{e \in E} g_e^{v \rightarrow u}}{\sum_{f \in E} g_f^{v \rightarrow u}} \leq Mn^2,
\end{aligned}$$

where the first inequality is by replacing maximums with sums and the last by $|\bigcup_{e \in E} V_e| \leq |V^2| = n^2$. Therefore, by Markov's inequality, with high constant probability the sparsifier has at most $O(Mn^2)$ hyperedges; this satisfies the size requirement as we later set $M = O(\epsilon^{-2}n^3)$.

Let us prove that the sparsifier H' indeed approximates the cuts of H . Fix some $S \subseteq V$ and note that

$$\mathbb{E} [\text{cut}_{H'}(S)] = \mathbb{E} \left[\sum_{e \in E} I_e \cdot \frac{1}{p_e} g_e(S \cap e) \right] = \sum_{e \in E} \mathbb{E} \left[I_e \cdot \frac{1}{p_e} g_e(S \cap e) \right] = \sum_{e \in E} g_e(S \cap e) = \text{cut}_H(S).$$

Hence, the cut is preserved in expectation. We shall now prove that the value of the cut is concentrated around its expectation. Let $Q_S = \{e \in E : p_e \in (0, 1) \wedge g_e(S \cap e) > 0\}$ be the set of all relevant hyperedges except those sampled with probability 1. Furthermore, denote the maximum contribution of any hyperedge to the cut by $b := \max_{e \in Q_S} \frac{1}{p_e} g_e(S \cap e)$. By the Chernoff bound for bounded variables (Lemma A.2),

$$\Pr [\text{cut}_{H'}(S) \notin (1 \pm \epsilon) \cdot \text{cut}_H(S)] \leq 2 \cdot \exp \left(-\frac{\epsilon^2 \cdot \text{cut}_H(S)}{b} \right). \quad (5)$$

We first analyze the special case when $S = V$. Observe that if $\text{cut}_H(V) = 0$ then the cut is preserved trivially. Otherwise, note that $p_e \geq \frac{Mg_e(e)}{\sum_{f \in E} g_f(f)}$ and hence

$$b = \max_{e \in Q_V} \frac{g_e(e)}{p_e} \leq \max_{e \in Q_V} g_e(e) \frac{\sum_{f \in E} g_f(f)}{Mg_e(e)} = \frac{\text{cut}_H(V)}{M}.$$

Plugging this back into Equation (5), we find $\Pr [\text{cut}_{H'}(V) \notin (1 \pm \epsilon) \cdot \text{cut}_H(V)] \leq 2 \cdot \exp(-\epsilon^2 M)$. Now turning to the general case $S \subset V$, we bound b by

$$\begin{aligned}
b &= \max_{e \in Q_S} \left\{ \frac{g_e(S \cap e)}{p_e} \right\} = \max_{e \in Q_S} \left\{ \frac{1}{M} \cdot \min_{(v,u) \in V_e} \frac{g_e(S \cap e) \cdot \sum_{f \in E} g_f^{v \rightarrow u}}{g_e^{v \rightarrow u}} \cdot \frac{\text{cut}_H(S)}{\text{cut}_H(S)} \right\} \\
&\leq \max_{e \in Q_S} \left\{ \frac{\text{cut}_H(S)}{M} \cdot \min_{(v,u) \in V_e: v \in S, u \notin S} \frac{g_e(S \cap e)}{g_e^{v \rightarrow u}} \frac{\sum_{f \in E} g_f^{v \rightarrow u}}{\sum_{f \in E} g_f(S \cap f)} \right\}, \quad (6)
\end{aligned}$$

where in the first equality we multiply and divide by $\text{cut}_H(S)$; the second inequality is from restricting the set over which the minimum is evaluated and substituting $\text{cut}_H(S) = \sum_{f \in E} g_f(S \cap f)$. Note that we do not lose any hyperedges by the restriction since by Claim 2.1, if $g_e(S \cap e) > 0$ then there exists some $v \in S, u \in e \setminus S$ that satisfy $g_e^{v \rightarrow u} > 0$.

We now show that b is bounded for every $e \in Q_S$. Fix some e , and observe that is sufficient to bound b using a specific pair $v^* \in S, u^* \in \bar{S}$. Let $v^* \in S, u^* \in \bar{S}$ be a pair of vertices that maximize

$g_e^{v^* \rightarrow u^*}$ (the minimum directed cut over e). Since $g_f^{v^* \rightarrow u^*}$ minimizes over subsets of f that contain v^* and not u^* , we have $g_f(S \cap f) \geq g_f^{v^* \rightarrow u^*}$ for all $f \in E$. Therefore, $\sum_{f \in E} g_f(S \cap f) \geq \sum_{f \in E} g_f^{v^* \rightarrow u^*}$, which we use to bound one of the quotients in (6); and it remains to bound $g_e(S \cap e)/g_e^{v^* \rightarrow u^*}$. By Claim 2.1, we have $g_e(S \cap e) \leq \sum_{v \in S} \sum_{u \in \bar{S}} g_e^{v \rightarrow u} \leq |S| |\bar{S}| g_e^{v^* \rightarrow u^*}$, where the last inequality is since $g_e^{v^* \rightarrow u^*}$ is maximal. Therefore, $g_e(S \cap e)/g_e^{v^* \rightarrow u^*} < n^2$. Plugging these two bounds into (6) we find that $b < n^2 \text{cut}_H(S)/M$, and then back into the concentration bound we get

$$\Pr [\text{cut}_{H'}(S) \notin (1 \pm \epsilon) \cdot \text{cut}_H(S)] \leq 2 \cdot \exp\left(-\frac{\epsilon^2 M}{n^2}\right)$$

Note that this is much smaller than the probability for $S = V$, and therefore we assume that both cases have the same probability. Choosing $M := c \cdot \epsilon^{-2} n^3$ for large enough but fixed $c > 0$, we get that H' approximates the cut S up to $1 + \epsilon$ factor with probability at least $1 - 2 \exp(-cn)$. Using a union bound over all $S \subseteq V$ we get that the sparsifier approximates all cuts simultaneously with probability at least $1 - 2 \exp(-cn) \cdot 2^n \geq 1 - 2 \exp(-n)$. This completes the construction of a sparsifier with $O(\epsilon^{-2} n^5)$ hyperedges that approximates H with quality $1 + \epsilon$.

We now turn back to prove Claim 2.1.

Proof of Claim 2.1. Fix some $S \subset V$ and $e \in E$. We show that

$$\sum_{v \in S} \sum_{u \in \bar{S}} g_e^{v \rightarrow u} = \sum_{v \in S} \sum_{u \in \bar{S}} g_e(P_{v \rightarrow u}^e) \geq g_e(S \cap e).$$

Let $v \in S$, $u_1, u_2 \in \bar{S}$, and observe that since g_e is submodular

$$g_e(P_{v \rightarrow u_1}^e) + g_e(P_{v \rightarrow u_2}^e) \geq g_e(P_{v \rightarrow u_1}^e \cap P_{v \rightarrow u_2}^e) + g_e(P_{v \rightarrow u_1}^e \cup P_{v \rightarrow u_2}^e) \geq g_e(P_{v \rightarrow u_1}^e \cap P_{v \rightarrow u_2}^e),$$

where the last inequality is from $g_e(P_{v \rightarrow u_1}^e \cup P_{v \rightarrow u_2}^e) \geq 0$. Summing over all $v \in S$, $u \in \bar{S}$, we get

$$\sum_{v \in S} \sum_{u \in \bar{S}} g_e(P_{v \rightarrow u}^e) \geq g_e\left(\bigcup_{v \in S} \bigcap_{u \in \bar{S}} P_{v \rightarrow u}^e\right).$$

Notice that for every $v \in S$, we must have $v \in \bigcap_{u \in \bar{S}} P_{v \rightarrow u}^e$, hence $S \subseteq \bigcup_{v \in S} \bigcap_{u \in \bar{S}} P_{v \rightarrow u}^e$. In addition, every $\bigcap_{u \in \bar{S}} P_{v \rightarrow u}^e \subseteq S$, therefore $S = \bigcup_{v \in S} \bigcap_{u \in \bar{S}} P_{v \rightarrow u}^e$. We conclude that

$$\sum_{v \in S} \sum_{u \in \bar{S}} g_e(P_{v \rightarrow u}^e) \geq g_e(S \cap e). \quad \square$$

This completes the proof of Theorem 1.4. □

2.1 Sparsification of Monotone Hypergraphs (Theorem 1.5)

The proof for the monotone case is similar to the general case. However, since monotone splitting functions are more structured it suffices to examine the importance of all the singleton cuts for each hyperedge. This results in smaller sampling probabilities and a better bound on the number of hyperedges in the sparsifier. The proof utilizes the following well known property of submodular functions.

Claim 2.2. *Let $g_e : 2^e \rightarrow \mathbb{R}^+$ be a splitting function. Fix $S \subseteq e$ and let $v^* \in S$ be a vertex that maximizes $g_e(\{v^*\})$. Then*

$$g_e(S \cap e) \leq |S| \cdot g_e(\{v^*\}).$$

Proof. Denote $S = \{v_1, \dots, v_{|S|}\}$. By the choice of v^* , we have $\sum_{i=1}^{|S|} g_e(\{v_i\}) \leq \sum_{i=1}^{|S|} g_e(\{v^*\}) = |S| \cdot g_e(\{v^*\})$. To conclude, observe that by the subadditivity of g_e , $g_e(S \cap e) \leq \sum_{i=1}^{|S|} g_e(\{v_i\})$. Combining the two we find $|S| \cdot g_e(\{v^*\}) \geq g_e(S \cap e)$. \square

Proof of Theorem 1.5. Again we start by describing the sampling probabilities for every hyperedge. For every hyperedge, set (for some $M > 0$ that will be chosen later)

$$p_e := M \cdot \max_{v \in V} \frac{g_e(\{v\})}{\sum_{f \in E} g_f(\{v\} \cap f)}.$$

To construct the sparsifier H' , sample each hyperedge with probability $\tilde{p}_e = \min(p_e, 1)$, and reweigh every sampled vertex by factor $\frac{1}{\tilde{p}_e}$. We prove that this sparsifier accurately approximates cuts of H and that it has at most $O(\frac{n^3}{\epsilon^2})$ hyperedges.

We first prove that the number of hyperedges in the sparsifier is at most $O(M \cdot n)$. Let I_e be an indicator for the event that the hyperedge e is picked. The expected number of sampled hyperedges is

$$\begin{aligned} \mathbb{E} \left[\sum_{e \in E} I_e \right] &= \sum_{e \in E} p_e = \sum_{e \in E} M \cdot \max_{v \in e} \frac{g_e(\{v\})}{\sum_{f \in E} g_f(\{v\} \cap f)} \\ &\leq M \cdot \sum_{e \in E} \sum_{v \in e} \frac{g_e(\{v\})}{\sum_{f \in E} g_f(\{v\} \cap f)} = M \cdot \sum_{v \in V} \frac{\sum_{e \in E} g_e(\{v\} \cap e)}{\sum_{f \in E} g_f(\{v\} \cap f)} = M \cdot \sum_{v \in V} 1 \leq Mn, \end{aligned}$$

where the inequality is by replacing a maximum with a sum. Therefore, by Markov's inequality, with high constant probability the sparsifier has at most $O(Mn)$ hyperedges. We later show that $M = O(\frac{n^2}{\epsilon^2})$ and hence the sparsifier has $O(\epsilon^{-2}n^3)$ hyperedges with high constant probability.

We now prove that the sparsifier indeed approximates the terminal cuts of G . Fix some $S \subseteq V$ and note that

$$\mathbb{E} [\text{cut}_{H'}(S)] = \mathbb{E} \left[\sum_{e \in E} I_e \cdot \frac{1}{p_e} g_e(S \cap e) \right] = \sum_{e \in E} \mathbb{E} \left[I_e \cdot \frac{1}{p_e} g_e(S \cap e) \right] = \sum_{e \in E} g_e(S \cap e) = \text{cut}_H(S).$$

Hence, the cut is preserved in expectation. We now prove that the value of the cut is concentrated around its expectation. Let $Q_S = \{e \in E : p_e < 1, g_e(S) > 0\}$ and $b = \max_{e \in Q_S} \frac{1}{p_e} g_e(S \cap e)$. Then by Chernoff's bound we find

$$\Pr [\text{cut}_{H'}(S) \notin (1 \pm \epsilon) \cdot \text{cut}_H(S)] \leq 2 \cdot \exp \left(-\frac{\epsilon^2 \cdot \text{cut}_H(S)}{b} \right)$$

Now note that

$$\begin{aligned} b &= \max_{e \in Q_S} \left\{ \frac{1}{p_e} g_e(S \cap e) \right\} = \max_{e \in Q_S} \left\{ \frac{1}{M} \min_{v \in e} \frac{\sum_{f \in E} g_f(\{v\} \cap f)}{g_e(\{v\})} g_e(S \cap e) \right\} \\ &= \max_{e \in Q_S} \left\{ \frac{1}{M} \cdot \min_{v \in e} \frac{\sum_{f \in E} g_f(\{v\} \cap f)}{g_e(\{v\})} \frac{g_e(S \cap e)}{\sum_{f \in E} g_f(S \cap f)} \sum_{f \in E} g_f(S \cap f) \right\}. \end{aligned} \quad (7)$$

We now bound the maximum value of b . Let $v^* \in S \subseteq V$ be a vertex that maximizes $g_e(\{v^*\})$ over S . Since g is monotone for all $S \subseteq V$ and $f \in E$, we find $\sum_{f \in E} g_f(S \cap f) \geq \sum_{f \in E} g_f(\{v^*\} \cap f)$.

Furthermore, by Claim 2.2 we have $g_e(S \cap e) \leq |S|g_e(\{v^*\}) \leq ng_e(\{v^*\})$. Combining the two we find

$$\frac{\sum_{f \in E} g_f(\{v^*\} \cap f)}{g_e(\{v^*\})} \cdot \frac{g_e(S \cap e)}{\sum_{f \in E} g_f(S \cap f)} \leq n$$

and therefore $b \leq n/M \cdot \sum_{f \in E} g_f(S \cap f) = n \cdot \text{cut}_H(S)/M$. Plugging this back into the concentration bound we find

$$\Pr[\text{cut}_{H'}(S) \notin (1 \pm \epsilon) \cdot \text{cut}_H(S)] \leq 2 \cdot \exp\left(-\frac{\epsilon^2 M}{n}\right).$$

Choosing $M = c \cdot \frac{n^2}{\epsilon^2}$, we get that the event is satisfied with probability at least $1 - 2 \exp(-cn)$. Using union bound over all $S \subseteq V$ and choosing some large enough but constant $c > 0$ we get that the sparsifier approximates all cuts with probability at least $1 - 2 \exp(-cn) \cdot 2^n \geq 1 - 2 \exp(-n)$. This completes the construction of a sparsifier for monotone-submodular hypergraphs with at most $O(\epsilon^{-2}n^3)$ hyperedges. \square

3 Sparsifiers for Finite-Spread Splitting Functions

This section provides a construction of sparsifiers for finite-spread splitting functions (Theorem 1.7). Our construction is based on the method presented in [CKN20] for constructing cut sparsifiers for the all-or-nothing splitting function. The main argument is that by approximating every hyperedge up to a factor of μ_H as the all-or-nothing splitting function, we can follow the algorithm and proof of [CKN20]. This approximation only holds when the functions g_e have finite spread. The main difference is that we need to account for hyperedges contributing different amounts to different cuts. We show that by oversampling hyperedges at a rate higher by a μ_H factor, we can adjust the Chernoff bounds, and then the rest of the proof follows using the all-or-nothing approximation. Throughout the proof we assume that $\min_{S \subseteq e: g_e(S) \neq 0} g_e(S) = 1$ for all $e \in E$. We can make this assumption by recalling that we limited our discussion to splitting functions with integral values, and observing that a splitting function g_e with a higher minimal value can be divided into multiple functions with minimal value 1 without affecting the cuts of H .

We begin by presenting the relevant definitions and results from the existing literature.

Definition 3.1. *Let $G = (V, F, w)$ be a weighted graph. A k -strong component in G is a maximal vertex induced subgraph such that the minimum cut in the component is k .*

Lemma 3.2 ([BK96]). *Given a weighted graph $G = (V, F, w)$ and some $k > 0$, the k -strong components of G partition V . For every $k' > k$ the k' -strong components are a refinement of the k -strong components.*

Definition 3.3. *In a weighted graph $G = (V, F, w)$, the strength of an edge $f \in F$, denoted by k_f , is the maximal $k > 0$ such that f is contained in a k -strong component.*

Claim 3.4 (Corollary 4.9 in [BK15]). *In every weighted graph G on n vertices, there are at most $n - 1$ distinct values of edge strengths.*

Following the proof in [CKN20], the sampling probabilities of the hyperedges of $H = (V, E, g)$ are determined by an auxiliary weighted graph G , where every hyperedge $e \in E$ is represented by a weighted clique F_e in G (with perhaps some weights being zero). Observe that G may have many parallel edges between the same pair of vertices, each induced by a different hyperedge. Denote the set of edges with positive weight in F_e by F_e^+ . The precise construction of the auxiliary graph G is described in [CKN20]. Define the *hyperedge strengths* as $\kappa_e := \min_{f \in F_e} k_f$ and $\kappa_e^{\max} := \max_{f \in F_e^+} k_f$.

Theorem 3.5 (Theorem 3 in [CKN20]). *Let $H = (V, E, g)$ be a hypergraph with a finite-spread splitting function $g : 2^V \rightarrow \mathbb{R}_+$. For every integer $\gamma \geq 2$ there exists an assignment of weights to the edges of F_e such that in the resulting G*

- $\sum_{f \in F_e} w(f) = 1;$
- $\frac{\kappa_e^{\max}}{\kappa_e} \leq \gamma.$

We now turn to proving Theorem 1.7. Throughout the proof we will partition the hyperedges into two sets, those with $g_e(e) = 0$ and those with $g_e(e) > 0$. We sparsify each subgraph independently and then the union of the two sparsifiers to obtain the sparsifier for H . Note that if each sparsifier approximates its hyperedges with quality $1 + \epsilon$, then their union $(1 + \epsilon)$ -approximates all the cuts of H by the additivity of the cuts; this increases the size of the sparsifier by at most a factor of 2. The sampling process for both sets is identical and hence from now on we assume that either all the hyperedges of H have $g_e(e) = 0$ or all have $g_e(e) > 0$.

Start by applying Theorem 3.5 with $\gamma = 2$ to obtain the auxiliary graph G and define the strengths of the hyperedges of H accordingly. Let H' be a sparsifier constructed by sampling every hyperedge $e \in E$ with probability $p_e = \min\{1, \rho/\kappa_e\}$, for $\rho = \epsilon^{-2} t \mu_H \gamma^2 \log n$ with constant $t > 0$ to be determined later. For each sampled edge we assign a rescaled splitting function $g'_e = p_e^{-1} g_e$.

By Claim 6 in [CKN20] the size of the sparsifier resulting from this sampling method is at most $O(\rho \gamma n) = O(\epsilon^{-2} n t \mu_H \gamma^3 \log n)$.

Hence, it remains to show that H' approximates the cuts of H . The proof is based on partitioning the hyperedges of G based on their strength, κ_e , and showing that the additive error for each set of hyperedges is small. Define the following sets. Let $E_{\geq i} = \{e \in E : \kappa_e \geq \rho 2^i\}$ be the set of all hyperedges in H with strength at least $\rho 2^i$, and let the set of all hyperedges with strength in $[\rho 2^i, \rho 2^{i+1})$ be $E_i = E_{\geq i} \setminus E_{\geq i+1}$. Similarly, let $F_{\geq i} = \{f \in F^+ : k_f \geq \rho 2^i\}$ be the set of all edges in G of strength at least $\rho 2^i$. Finally, let $E_{\geq i}^{\max} = \{e \in E : \kappa_e^{\max} \geq \rho 2^i\}$ be the set of all hyperedges with maximum strength at least $\rho 2^i$.

Let $Y \subseteq E$ be a subset of the hyperedges of H and let $\beta \in \mathbb{R}_+^{|Y|}$. Denote by $H[\beta Y]$ the sub-hypergraph with the hyperedges in Y , where for each $e \in Y$ the splitting function g_e is scaled by β_e . Similarly, for $Y \subseteq F$, and $\beta \in \mathbb{R}_+^{|Y|}$, let $G[\beta Y]$ be the subgraph of G with the edges in Y and the weight of each $f \in Y$ scaled by β_f .

We now define certain subsets that will be used to bound the error. For every $i \in \mathbb{N}$, define a relative weighting function $\beta^i \in \mathbb{R}_+^E$ where $\beta_e^i = 2^{i-j}$ for all hyperedges $e \in E$ (or similarly for edges in G) with strength in $[\rho 2^j, \rho 2^{j+1})$. Observe that this function modifies the weight of each hyperedge (or edge) according to its strength. Finally, let the random vector $\hat{\beta}^i \in \mathbb{R}_+^E$ be the scaling obtained from the sampling process limited to E_i ; i.e. $\hat{\beta}_e^i = p_e^{-1}$ if e is sampled into H' and $\hat{\beta}_e^i = 0$ otherwise. Observe that $H' = H \left[\bigcup_i \hat{\beta}^i E_i \right]$.

Our main technical lemma bounds the additive error between the value of every cut S on $\hat{\beta}^i E_i$ and its expectation. Note that the lemma does not immediately imply that H' is a quality $q = 1 + \epsilon$ sparsifier for H as the error bound might be larger than $\epsilon \cdot \text{cut}_{H[E_i]}(S)$. However, we will show that the sum of all these error bounds is small compared to the value of the cut.

Lemma 3.6. *Fix an integer $i > 0$. Then with probability at least $1 - 8n^{-2}$,*

$$\forall S \subseteq V, \quad \left| \text{cut}_{H[\hat{\beta}^i E_i]}(S) - \text{cut}_{H[E_i]}(S) \right| \leq \frac{\epsilon}{\gamma} \text{cut}_{H[\beta^i E_{\geq i}^{\max}]}(S). \quad (8)$$

We first show that Lemma 3.6 implies Theorem 1.7.

Proof of Theorem 1.7. Sample the hyperedges of H into H' using the sampling method described above. Using the union bound, we bound the probability that the event in Equation (8) holds for all values of i whose corresponding hyperedge set E_i is non-empty. By Claim 3.4 there are at most $n - 1$ such values. Hence, with probability at least $1 - 8n^{-1}$ Equation (8) holds for all such i .

We now show that if Equation (8) holds for all such i , then the sparsifier has quality $q = 1 + \epsilon$. Observe that for every $S \subseteq V$,

$$\begin{aligned} \left| \text{cut}_{H[\hat{\beta}^i E_i]}(S) - \text{cut}_{H[E_i]}(S) \right| &\leq \frac{\epsilon}{\gamma} \text{cut}_{H[\beta^i E_{\geq i}^{\max}]}(S) \\ &\leq \frac{\epsilon}{\gamma} \sum_{j \geq i - \log \gamma} \text{cut}_{H[\beta^i E_j]}(S) = \frac{\epsilon}{\gamma} \sum_{j \geq i - \log \gamma} 2^{i-j} \cdot \text{cut}_{H[E_j]}(S), \end{aligned}$$

where the second inequality is since $E_{\geq i}^{\max} \subseteq E_{\geq i - \log \gamma}$ (this holds since $\frac{\kappa_e^{\max}}{\kappa_e} \leq \gamma$). Summing over all i ,

$$\begin{aligned} |\text{cut}_{H'}(S) - \text{cut}_H(S)| &= \sum_{i \geq 0} \left| \text{cut}_{H[\hat{\beta}^i E_i]}(S) - \text{cut}_{H[E_i]}(S) \right| \\ &\leq \frac{\epsilon}{\gamma} \sum_{i \geq 0} \sum_{j \geq i - \log \gamma} 2^{i-j} \cdot \text{cut}_{H[E_j]}(S) \\ &= \frac{\epsilon}{\gamma} \sum_{j \geq 0} \text{cut}_{H[E_j]}(S) \sum_{i \leq j + \log \gamma} 2^{i-j} \leq 2\epsilon \sum_{j \geq 0} \text{cut}_{H[E_j]}(S) = 2\epsilon \text{cut}_H(S), \end{aligned}$$

where the second inequality is by $\sum_{i \leq j + \log \gamma} 2^{i-j} \leq 2\gamma$. Hence, with probability $1 - 8n^{-1}$ we have that H' is a cut sparsifier of quality $q = 1 + \epsilon$ for H . \square

To conclude the proof, we still need to show Lemma 3.6.

3.1 Proof of Lemma 3.6

To prove the lemma, we first provide several definitions and claims. Define the *sample set of level i* to be $\hat{E}_i = \hat{\beta}^i E_i + \beta^i E_{\geq i+1}$. Since the contribution of each hyperedge to a cut is additive,

$$\text{cut}_{H[\hat{E}_i]}(S) - \text{cut}_{H[\beta^i E_{\geq i}]}(S) = \text{cut}_{H[\hat{\beta}^i E_i]}(S) - \text{cut}_{H[E_i]}(S),$$

notice that in the last term we omitted β^i as for every $e \in E_i$ we have $\beta_e^i = 2^0$. Hence, showing

$$\left| \text{cut}_{H[\hat{E}_i]}(S) - \text{cut}_{H[\beta^i E_{\geq i}]}(S) \right| \leq \frac{\epsilon}{\gamma} \text{cut}_{H[\beta^i E_{\geq i}^{\max}]}(S) \quad (9)$$

would suffice to prove the lemma. The following claims outline useful properties of $G[F_{\geq i}]$.

Claim 3.7 (Claim 8 from [CKN20]). *For any $e \in E_{\geq i}$, the entire vertex set F_e belongs to the same connected component in $G[F_{\geq i}]$.*

Claim 3.8 (Claim 9 from [CKN20]). *Let A_G be a connected component in $G[F_{\geq i}]$. Then the minimum cut size of A_G is at least $\rho 2^i$.*

Our proof will also use the following cut-counting lemma.

Lemma 3.9 (Corollary 8.2 in [Kar93]). *Given a weighted graph $G = (V, F, w)$ with minimum cut size $c > 0$, for all integers $\alpha \geq 1$, the number of cuts of the graph of weight at most αc is at most $|V|^{2\alpha}$.*

To use the cut counting lemma, we show that the cuts of $H [\beta^i E_{\geq i}^{\max}]$ are bounded from below by the cuts of $G [\beta^i F_{\geq i}]$. To do that, we use the following from [CKN20].

Claim 3.10 (Claim 7 from [CKN20]). *Let $\bar{H} = (V, E, \bar{g})$ be a hypergraph with the same vertices and hyperedges as H , but with $\bar{g}_e : 2^e \rightarrow \mathbb{R}_+$ being the all-or-nothing splitting function, i.e.,*

$$\bar{g}_e(S) = \begin{cases} 0, & \text{if } |S \cap e| = 0 \text{ or } |S \cap e| = |e|; \\ 1, & \text{otherwise.} \end{cases}$$

Then for all $i \geq 0$,

$$\forall S \subseteq V, \quad \text{cut}_{\bar{H} [\beta^i E_{\geq i}^{\max}]}(S) \geq \text{cut}_{G [\beta^i F_{\geq i}]}(S).$$

We then obtain the desired lower bound as an immediate corollary. This corollary is where we use our assumption that g_e has finite spread, otherwise the cuts of H are not bounded by the cuts of G , and we cannot use the cut counting lemma (Lemma 3.9).

Corollary 3.11. *For all $i \geq 0$,*

$$\forall S \subseteq V, \quad \text{cut}_{H [\beta^i E_{\geq i}^{\max}]}(S) \geq \text{cut}_{G [\beta^i F_{\geq i}]}(S).$$

Furthermore, if $g_e(e) > 0$ for all $e \in E$ then for every connected component $A_G \subseteq V$ of $G [\beta^i F_{\geq i}]$, the cut A_G on $H [\beta^i E_{\geq i}^{\max}]$ is at least the minimum cut of $G [\beta^i F_{\geq i}]$ in the component A_G . Formally,

$$\text{cut}_{H [\beta^i E_{\geq i}^{\max}]}(A_G) \geq \min_{T \subset A_G, T \neq \emptyset} \text{cut}_{G [\beta^i F_{\geq i}]}(T).$$

Proof. We begin by showing the first part of the corollary. Observe that since g_e has finite spread, whenever $|e| > |S \cap e| \geq 1$ we have $g_e(S \cap e) \geq 1 = \bar{g}_e(S \cap e)$. In addition, $g_e(e) \geq 0 = \bar{g}_e(e)$. Hence, for all $e \in E$ and $S \subseteq e$, $g_e(S \cap e) \geq \bar{g}_e(S \cap e)$. Therefore, for every cut $S \subseteq V$,

$$\text{cut}_{H [\beta^i E_{\geq i}^{\max}]}(S) \geq \text{cut}_{\bar{H} [\beta^i E_{\geq i}^{\max}]}(S) \geq \text{cut}_{G [\beta^i F_{\geq i}]}(S).$$

We now turn to the second part of the corollary. Fix some connected component A_G of $G [\beta^i F_{\geq i}]$ and denote some set achieving the minimum cut over A_G by $T^* \subset A_G, T^* \neq \emptyset$. Note that the value of the cut T^* is strictly positive since A_G is a connected component. Notice that for all S such that $S \cap e \neq \emptyset$ we have $g_e(S \cap e) \geq 1 \geq \bar{g}_e(S \cap e)$ by our assumption. Hence,

$$\begin{aligned} \text{cut}_{H [\beta^i E_{\geq i}^{\max}]}(A_G) &= \sum_{e \in E_{\geq i}^{\max}} \beta_e^i g_e(A_G \cap e) \geq \sum_{e \in E_{\geq i}^{\max}} \beta_e^i \mathbf{1}_{\{A_G \cap e \neq \emptyset\}} \geq \sum_{e \in E_{\geq i}^{\max}} \beta_e^i \mathbf{1}_{\{T^* \cap e \neq \emptyset\}} \\ &\geq \sum_{e \in E_{\geq i}^{\max}} \beta_e^i \bar{g}_e(T^*) = \text{cut}_{\bar{H} [\beta^i E_{\geq i}^{\max}]}(T^*) \geq \text{cut}_{G [\beta^i F_{\geq i}]}(T^*), \end{aligned}$$

where the second inequality is since $T^* \subseteq A_G$, the third is by the definition of the all-or-nothing splitting function, and the last one is by Claim 3.10. \square

To prove Lemma 3.6, we bound the error of each connected component of $H[E_{\geq i}]$ separately. For each connected component, we bound the error by a term proportional to $\text{cut}_{H[\beta^i E_{\geq i}^{\max}]}(S)$, which by Corollary 3.11 bounds the cut in $G[\beta^i F_{\geq i}]$ from above. This allows us to utilize the cut-counting-lemma (Lemma 3.9) to bound the number of small cuts.

Proof of Lemma 3.6. Fix some $i \geq 0$ and let $A_G = (V_{A_G}, F_{A_G})$ be some connected component in the subgraph $G[F_{\geq i}]$. Note that if for all $e \in E$ we have $g_e(e) = 0$ then the error for the cut $S = V_{A_G}$ is 0. By Claim 3.7 the hypergraph induced by the vertices V_{A_G} is a proper sub-hypergraph of $H[E_{\geq i}]$, as every hyperedge in $E_{\geq i}$ is incident only to the vertices in V_{A_G} . Denote this connected component by H_{A_G} . Note, that every cut with value 0 is preserved trivially by the sparsifier.

Fix some cut $S \subseteq V_{A_G}$ such that $\text{cut}_{H_{A_G}}(S) > 0$. We will start by bounding the probability of $H_{A_G}[\hat{E}_i]$ creating a large error,

$$\left| \text{cut}_{H_{A_G}[\hat{E}_i]}(S) - \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S) \right| \geq \frac{\epsilon}{\gamma} \text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S). \quad (10)$$

Notice that,

$$\mathbb{E} \left[\text{cut}_{H_{A_G}[\hat{E}_i]}(S) \right] = \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S).$$

Denote

$$\theta := \frac{\text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S)}{\text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S)} \geq 1.$$

Observe that by the Chernoff bound when $\theta\epsilon/\gamma \geq 1$ the probability of large deviation is at the most the probability of large deviation in the case $\theta\epsilon/\gamma \leq 1$. Hence, we assume henceforth that $\theta\epsilon/\gamma \leq 1$. By the Chernoff bound (Lemma A.2),

$$\begin{aligned} & \Pr \left[\left| \text{cut}_{H_{A_G}[\hat{E}_i]}(S) - \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S) \right| \geq \frac{\epsilon}{\gamma} \text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S) \right] \\ &= \Pr \left[\left| \text{cut}_{H_{A_G}[\hat{E}_i]}(S) - \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S) \right| \geq \theta \frac{\epsilon}{\gamma} \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S) \right] \\ &\leq 2 \exp \left(- \frac{\epsilon^2 \theta^2 \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S)}{3\gamma^2 r} \right) \leq 2 \exp \left(- \frac{\epsilon^2 \text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S)}{3\gamma^2 r} \right), \end{aligned}$$

where r is the maximum possible contribution of any sampled hyperedge to the cut. The second inequality is from $\theta \geq 1$. Observe that r is bounded by

$$r = \max_{e \in E, S \subseteq V} \frac{1}{p_e} g_e(S) \leq \frac{\kappa_e}{\rho} \mu_H.$$

Since all the sampled edges e are in E_i , all of them have strength in $\kappa_e \in [\rho 2^i, \rho 2^{i+1}]$. This implies that $r \leq 2^{i+1} \mu_H$. Let $x = \min_{T \subseteq V_{A_G}} \text{cut}_{A_G}(T)$ be the value of a minimum cut of A_G . We begin by lower bounding the value of the cut S in $H_{A_G}[\beta^i E_{\geq i}^{\max}]$. By Corollary 3.11, if $S \neq V_{A_G}$ then $\text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S) \geq \text{cut}_{A_G}(S)$. Otherwise, by the same corollary we have that $\text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S) \geq x$. Combining these and recalling $\text{cut}_{A_G}(S) = 0$ if $S = V_{A_G}$ we find that $\text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S) \geq \max(\text{cut}_{A_G}(S), x)$.

Let $\alpha = \max(\text{cut}_{A_G}(S), x)/x$ be the ratio between the lower bound on $\text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S)$ and x . By Claim 3.8, we have that $x \geq \rho 2^i$. Therefore,

$$\text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S) \geq \alpha x \geq \alpha \rho 2^i.$$

Plugging in these two bounds

$$\begin{aligned} & \Pr \left[\left| \text{cut}_{H_{A_G}[\hat{E}_i]}(S) - \text{cut}_{H_{A_G}[\beta^i E_{\geq i}]}(S) \right| \geq \frac{\epsilon}{\gamma} \text{cut}_{H_{A_G}[\beta^i E_{\geq i}^{\max}]}(S) \right] \\ & \leq 2 \exp \left(-\frac{\epsilon^2 \alpha \rho 2^i}{3\gamma^2 \mu_H 2^{i+1}} \right) = 2 \exp \left(-\frac{\epsilon^2 \alpha}{6\gamma^2 \mu_H} \frac{t \mu_H \gamma^2 \ln n}{\epsilon^2} \right) = 2 \exp \left(-\frac{\alpha t \ln n}{6} \right). \end{aligned}$$

Therefore, the event occurs with probability at most $2n^{-\frac{\alpha t}{6}}$.

By the cut-counting lemma (Lemma 3.9), there are at most $|V_{A_G}|^{2\alpha}$ cuts of value at most αx in A_G which are different from \emptyset, V_{A_G} . Counting also the cut $S = V_{A_G}$, there are at most $|V_{A_G}|^{2\alpha} + 1 \leq 2|V_{A_G}|^{2\alpha}$ such cuts in total. Hence, the probability that Equation (10) occurs for any cut is at most,

$$\sum_{\alpha=1}^{\infty} 4|V_{A_G}|^{2\alpha(\frac{-t}{6}+2)} \leq 8|V_{A_G}|^{\frac{-t}{6}+1}.$$

Using a union bound over all connected components we find that the probability that Equation (10) occurs for any cut in any connected component A_{G_i} is at most

$$\sum_{\alpha=1}^{\infty} \sum_j 4|V_{A_{G_j}}|^{2\alpha(\frac{-t}{6}+2)} \leq 8n^{\frac{-t}{6}+2},$$

where inequality is since $\sum_j |V_{A_{G_j}}| \leq n$. Choosing $t > 24$ we find that the error term satisfies the requisite bound for all cuts simultaneously with probability at least $1 - 8n^{-2}$. Therefore, Equation (9) holds with high probability. This completes the proof of Lemma 3.6. \square

4 Deformation of Additive Splitting Functions

This section proves Theorem 1.10 by showing that every hyperedge with an additive splitting function $g_e : 2^e \rightarrow \mathbb{R}$ can be decomposed into multiple hyperedges, each with support size at most $O(\epsilon^{-2}K^{-1}|e| \log |e|)$. Our proof is based on sampling, namely, we approximate g_e by a sum of hyperedges $\{e_i\}_i$ with additive splitting functions $\{g_{e_i}\}_i$, such that each $e_i \subseteq e$ is constructed by a uniform independent sample of the vertices of e . The main challenge in the proof is showing that for every $S \subseteq e$, we have $\mathbb{E}[g_{e_i}(S \cap e_i)] \in (1 \pm \epsilon)g_e(S)$. We also prove that Theorem 1.10 implies a succinct representation of hypergraphs with additive splitting functions (Corollary 1.11). We begin by presenting a more detailed version of Theorem 1.10.

Lemma 4.1 (Detailed Statement of Theorem 1.10). *Let e be a hyperedge with an additive splitting function g_e with parameter $K \leq |e|$. Then, for every $\epsilon \in (0, 1)$ there exists a deformation of e into $N = O(\epsilon^{-2}|e|^2)$ hyperedges $\{e_i\}_i$ with additive splitting functions $\{g_{e_i}\}_i$ with parameter K' , such that*

$$\forall S \subseteq V, \quad \sum_{i=1}^N g_{e_i}(S \cap e_i) \in (1 \pm \epsilon)g_e(S).$$

In addition, the cardinality of each hyperedge satisfies $|e_i| \leq O(\epsilon^{-2}(|e|/K) \log |e|)$ and its spread is $\mu_{e_i} \leq O(\epsilon^{-2} \log |e|)$. Moreover, this decomposition can be found by a randomized sampling process with success probability at least $1 - 2|e|^{-4}$.

Using the above decomposition we obtain a succinct representation of hypergraphs with additive splitting functions.

Proof of Corollary 1.11. Let $H = (V, E, g)$ be a hypergraph such that all $g_e \in g$ are additive splitting functions, each with parameter K_e . Denote $\hat{K} = \min_{e \in E} K_e / |e|$. Applying Theorem 1.10 on all the hyperedges, we obtain a new hypergraph H' with additive splitting functions that $(1 + \epsilon)$ -approximates all the cuts of H . Furthermore, $\mu_{H'} \leq O(\epsilon^{-2} \log n)$. Hence, applying Theorem 1.7 on H' , we obtain a sparsifier $H'' = (V, E'')$ with at most $O(\epsilon^{-2} n \mu_{H'} \log n) = O(\epsilon^{-4} n \log^2 n)$ hyperedges. Note that H'' is a $(1 + \epsilon)^2$ -sparsifier of H and that it has additive splitting functions.

Finally, to bound the encoding size of H'' note that $K_{e''} \leq n$ for all $e'' \in E''$ since otherwise it does not affect the splitting function. Hence, we can store $K_{e''}$ using $O(\log n)$ bits. Additionally, the cardinality of each hyperedge e'' in H'' is at most $O(\epsilon^{-2} \hat{K}^{-1} \log n)$ by Lemma 4.1. Hence, we can store it using $O(\epsilon^{-2} \hat{K}^{-1} \log^2 n)$ bits (using $\log n$ bits to store every vertex in e''). Therefore, we find that the total encoding size of H'' is $O(\epsilon^{-6} \hat{K}^{-1} n \log^4 n)$ bits. \square

It is straightforward to adapt the proof of Corollary 1.11 to be algorithmic. Observe that we can assume that $|E| \leq \tilde{O}(\epsilon^{-2} n^2)$, since by the structure of additive splitting functions we have $\mu_e \leq n$, and thus by applying Theorem 1.7 we can first find a $(1 + \epsilon)$ -sparsifier of H with $\tilde{O}(\epsilon^{-2} n^2)$ hyperedges. We can then improve the success probability of Lemma 4.1 to $1 - 2n^{-4}$ by changing the sampling probabilities to use $\log n$ instead of $\log |e|$. Finally, using the union bound on all $\tilde{O}(\epsilon^{-2} n^2)$ hyperedges of H , we obtain that the succinct representation is successfully constructed with very high probability.

4.1 Proof of Lemma 4.1

We will use the following claim.

Claim 4.2. *Let $S \subseteq e$ be a subset of the vertices of the hypergraph, and let \hat{S} be a random subset of S where each $v \in e$ is sampled independently with probability $p = \min(\frac{c \log |e|}{K \epsilon^2}, 1)$ for some $K \leq |e|$ and $c > 0$. Then,*

$$\forall \delta \geq 0, \quad \Pr \left[\left| |\hat{S}| - p|S| \right| \geq \delta p|S| \right] \leq 2|e|^{-\frac{c\delta^2|S|}{(2+\delta)K\epsilon^2}}.$$

The claim follows immediately from the Chernoff bound (Lemma A.1).

Proof of Lemma 4.1. Denote $\epsilon' = \epsilon/4$ and let $N = q\epsilon'^{-2}|e|^2$ for some $q \geq 1$. Let e be some hyperedge with splitting function $g_e(S) = \min(|S|, K)$ or $g_e(S) = \min(|S|, |\bar{S}|, K)$. If either $K \leq 100\epsilon^{-2} \log |e|$ or $\epsilon^{-2} > |e|$, then $|e| \leq O(\epsilon^{-2}(|e|/K) \log |e|)$ and we are done.

Generate the hyperedges $\{e_i\}_{i=1}^N$ by sampling the vertices of e independently with probability $p = c\epsilon'^{-2}K^{-1} \log |e|$ for some constant $100 > c > 0$ to be determined later. Note that $p < 1$ since $K > 100\epsilon^{-2} \log |e|$. The splitting functions of the new hyperedges are then given by $g_{e_i}(S) = \frac{1}{N} \min(|S \cap e_i|/p, K)$ or $g_{e_i}(S) = \frac{1}{N} \min(|S \cap e_i|/p, |\bar{S} \cap e_i|/p, K)$ in accordance with g_e . Observe that these functions are additive by factoring out $1/p$ from all terms in the minimum.

We begin by showing that e_i has small spread and small support. Observe that

$$\mu_{e_i} = \frac{\max_{S \subseteq e} g_i(S)}{\min_{T \subseteq e} g_i(T)} \leq Kp \leq c \cdot \frac{\log |e|}{\epsilon'^2}.$$

By Claim 4.2 the cardinality of each hyperedge is no more than $2p|e| = 2c\epsilon'^{-2}(|e|/K) \log |e|$ with probability $1 - 2|e|^{-c\delta\epsilon'^{-2}|e|/(3K)} \geq 1 - 2|e|^{-c\epsilon'^{-2}/3}$, where the inequality is by $K \leq |e|$. Choosing $c \geq 20$ and applying the union bound over all N hyperedges we obtain that the cardinality of

all hyperedges is at most $2c\epsilon'^{-2}(|e|/K) \log |e|$ with probability at least $1 - 2|e|^{-c\epsilon'^{-2}/3} q\epsilon'^{-2}|e|^2 \geq 1 - 2|e|^{-5}$; the inequality holds as $\epsilon'^{-2} < |e|$, $q \geq 1$ and $\epsilon < 1$ which implies $\epsilon' \leq 1/4$.

We now show that the sum $\sum_i g_{e_i}$ $(1 + \epsilon)$ -approximates g_e . Denote $h(S) = \sum_{i=1}^N g_{e_i}(S)$. We will show that for every $S \subseteq e$, with high probability,

$$h(S) \in (1 \pm \epsilon) \cdot g_e(S).$$

Fix some $S \subseteq e$ such that $g_e(S) > 0$ and assume without loss of generality $|S| \leq |\bar{S}|$. The analysis of the expectation of g_{e_i} is split into two claims for the monotone and symmetric cases, that we will prove shortly.

Claim 4.3. *Let $g_e(S) = \min(|S|, K)$ and suppose $c > 0$ is a sufficiently large constant. Then*

$$\mathbb{E}[N \cdot g_{e_i}(S)] \in (1 \pm 2\epsilon')g_e(S).$$

Claim 4.4. *Let $g_e(S) = \min(|S|, |\bar{S}|, K)$ and suppose $c > 0$ is a sufficiently large constant. Then*

$$\mathbb{E}[N \cdot g_{e_i}(S)] \in (1 \pm 2\epsilon')g_e(S).$$

By Claims 4.3 and 4.4, $\mathbb{E}[h(S)] \in (1 \pm 2\epsilon')g_e(S)$. Now using the Chernoff bound for bounded random variables (Lemma A.2),

$$\Pr[h(S) \notin (1 \pm \epsilon')\mathbb{E}[h(S)]] \leq 2 \exp\left(-\frac{\epsilon'^2 \mathbb{E}[h(S)]}{3r}\right),$$

where r is the maximum possible contribution of a single $g_{e_i}(S)$ to the sum. By the definition of g_{e_i} we have $r = K/N \leq |e|/N = q^{-1}\epsilon'^2/|e|$. Note also that $g_e(S) \geq 1$ and hence $\mathbb{E}[h(S)] \geq 1 - 2\epsilon'$. Plugging these back in we find that

$$\Pr[h(S) \notin (1 \pm \epsilon')\mathbb{E}[h(S)]] \leq 2 \exp(-q(1 - 2\epsilon')|e|/3).$$

Choosing $q > 12$, and noting $1 - 2\epsilon' > 1/2$, gives us that with probability at least $1 - 2^{-2|e|}$,

$$h(S) \in (1 \pm \epsilon') \cdot ((1 \pm 2\epsilon')g_e(S)) \subseteq (1 \pm 4\epsilon')g_e(S).$$

Using a union bound over all $2^{|e|}$ possible cuts of e we find that $h(S)$ is a quality $1 + \epsilon$ approximation for $g_e(S)$ with probability at least $1 - 2^{-|e|}$. Finally, observe that by the union bound we have that both the cardinality of all hyperedges is bounded and the quality of the approximation is $1 + \epsilon$ simultaneously with probability at least $1 - 2|e|^{-4}$.

We now return to prove the claim about the expectation of $h(S)$ when g_e is monotone. The proof for the symmetric case (Claim 4.4) is similar to Claim 4.3 and appears in Appendix D.

Proof of Claim 4.3. Let $S_i = S \cap e_i$ be the intersection between S and the sampled hyperedge. Observe that the function $\min(x, K)$ is concave, and hence by Jensen's inequality

$$\mathbb{E}[\min(|S_i|/p, K)] \leq \min(\mathbb{E}[|S_i|/p], K) = \min(|S|, K) = g_e(S).$$

It remains to prove that $\mathbb{E}[Ng_{e_i}(S)] \geq (1 - 2\epsilon')g_e(S)$. We split the analysis into two cases, when $|S| < K/2$ and its complement. Starting with the case when $|S|$ is small, observe that setting $\delta = K/|S| > 2$ we have by Claim 4.2 that

$$\Pr[|S_i| \geq pK] \leq 2|e|^{-\frac{c\delta^2|S|}{(2+\delta)K\epsilon'^2}} \leq 2|e|^{-\frac{c\delta|S|}{2K}} = 2|e|^{-\frac{c}{2}},$$

where the second inequality is by $\delta/(2 + \delta) > 1/2$ when $\delta > 2$. Choosing $c > 5$ we find that the probability is at most $2|e|^{-5/2}$. Now rewriting the expectation of $f(S)$ we have

$$\begin{aligned} \mathbb{E}[Ng_{e_i}(S)] &= \sum_{j=0}^{|S|} \min\left(\frac{j}{p}, K\right) \Pr[|S_i| = j] \geq \sum_{j=0}^{\lfloor pK \rfloor} \frac{j}{p} \Pr[|S_i| = j] \\ &\geq \sum_{j=0}^{|S|} \frac{j}{p} \Pr[|S_i| = j] - \sum_{j=\lfloor pK \rfloor + 1}^{|S|} \frac{j}{p} \Pr[|S_i| = j], \end{aligned}$$

where the first inequality is by dropping all elements with $j \geq \lfloor pK \rfloor + 1$ and the second is by adding and subtracting the same terms for $j \geq \lfloor pK \rfloor + 1$. Since $j \leq |S|$ and $\sum_{j=\lfloor pK \rfloor + 1}^{|S|} \Pr[|S_i| = j] = \Pr[|S_i| \geq \lfloor pK \rfloor + 1]$ we find

$$\mathbb{E}[f(S)] \geq \frac{\mathbb{E}[|S_i|]}{p} - \frac{|S|}{p} \Pr[|S_i| \geq pK] \geq |S| - \frac{|S|}{p|e|^{5/2}}.$$

Observe that $|S|/p = |S|K\epsilon^2/(c \log |e|) < |e||S|$ by $K \leq |e|$ and $c > 1$. Therefore, we find

$$\mathbb{E}[Ng_{e_i}(S)] \geq |S| - \frac{|e|^2}{10|e|^{5/2}} \geq |S| - \frac{1}{10|e|^{1/2}} \geq (1 - \epsilon')g_e(S),$$

where the last inequality is by $|S| \geq 1$ for all nontrivial cuts and $\epsilon^{-2} < |e|$.

Now we turn to the case $|S| > K/2$. Observe that for $Ng_{e_i}(S) = \min(|S_i|/p, K) < (1 - \epsilon')g_e(S)$ we must have $|S_i| < (1 - \epsilon')p|S|$. By Claim 4.2 this event happens with probability at most $2|e|^{-\frac{c\epsilon'^2|S|}{3K\epsilon'^2}} \leq 2|e|^{-\frac{c}{6}}$. Therefore,

$$\mathbb{E}[Ng_{e_i}(S)] \geq \left(1 - 2|e|^{-\frac{c}{6}}\right) (1 - \epsilon')p|S|/p \geq (1 - 2\epsilon)|S| = (1 - 2\epsilon)g_e(S),$$

where the last inequality is by $c > 9$ and $\epsilon' > 1/|e|$. □

This concludes the proof of Lemma 4.1. □

5 Encoding-Size Lower Bounds

This section shows that for additive splitting functions, $\Omega(n^2)$ bits are needed to represent sparsifiers that are reweighted-subgraphs, proving Theorem 1.12. We also prove two results on the encoding size of directed hypergraphs (see Section 5.2): (1) reweighted-subgraph sparsifiers for directed hypergraphs require encoding size of $\Omega(n^3/\epsilon)$ bits, and (2) any data structure that stores the cuts of a directed hypergraph requires $\Omega(n^2/\epsilon)$ bits.

To show Theorem 1.12, we actually prove a stronger version of the theorem for all cardinality-based splitting functions with $\mu_e < n/3$.

Lemma 5.1 (Stronger Version of Theorem 1.12). *Let $\hat{g} : [n] \rightarrow \mathbb{R}_+$ define a cardinality-based splitting function g_e with spread $\mu_e < n/3$. There exists a family of hypergraphs \mathcal{H} with splitting function $g_e(S) = \hat{g}(|S|)$ for all $e \in E$, such that encoding a reweighted subgraph $(1 + \epsilon)$ -sparsifier for an input $H \in \mathcal{H}$ requires $\Omega(n^2)$ bits.*

5.1 Proof of Lemma 5.1

Denote the *gradient* of a function $f : [n] \rightarrow \mathbb{R}_+$ be defined as $\Delta_i(f) := f(i+1) - f(i)$, we will omit the function f when it is clear from the context. The proof follows immediately from the proceeding technical claim.

Claim 5.2. *Let $\hat{g} : [n] \rightarrow \mathbb{R}_+$ and let t be the smallest integer such that $\Delta_t < \Delta_0$. If $t < n/3$, then there exists a family of hypergraphs \mathcal{H} , with splitting function $g_e(S) = \hat{g}(|S \cap e|)$ for all $e \in E$, such that encoding a reweighted-subgraph $(1 + \epsilon)$ -sparsifier for an input $H \in \mathcal{H}$ requires $\Omega(n^2)$ bits.*

Note that Δ_i satisfies the following easy property.

Observation 5.3. *Let $g_e : 2^e \rightarrow \mathbb{R}_+$ be a cardinality based splitting function which is defined by the function $\hat{g} : [e] \rightarrow \mathbb{R}_+$. Then the series defined by $\Delta_0, \Delta_1, \dots, \Delta_{|e|-1}$ is non-increasing.*

Proof. Let $S \subseteq e$ be a set of size i and $T \subset S$ be a set of size $j < i$ and $x \in e \setminus \{S\}$. Then by the submodularity of g_e we have

$$\Delta_i = g_e(S \cup \{x\}) - g_e(S) \leq g_e(T \cup \{x\}) - g_e(T) = \Delta_j.$$

□

Lemma 5.1 follows immediately using Observation 5.3 and Claim 5.2.

Proof. Assume without loss of generality that $\hat{g}(1) = 1$, then $\Delta_0 = 1$. If there exists some $k < n/3$ such that $\Delta_k \leq 0$ we can apply Claim 5.2 for the first k with $\Delta_k < \Delta_0$. Since $\Delta_0 = 1$ and Δ_i are non-increasing we have that $\sum_{i=0}^{\mu_e} \Delta_i \leq \mu_e$ and hence the average Δ_i is at most $\mu_e / (\mu_e + 1) < 1$. Therefore, there must exist at least one $\Delta_i < 1$ for some $i \in [\mu_e + 1] \subseteq [n/3]$. Apply Claim 5.2 to the first such i . □

The proof of Claim 5.2 boils down to a counting argument, we create a family of hypergraphs with $g_e(S) = \hat{g}(|S|)$ as their splitting function. The vertices of each hypergraph are partitioned into three sets V, U, W . Each hyperedge is defined by a union of three parts: (1) a random subset of the vertices of V , (2) a subset of U that is defined by the Hadamard code, and (3) an unsparifiable part over the vertices W .

Using the unsparifiable part we show that any reweighted-subgraph sparsifier must contain all the hyperedges. We then show that it is possible to exactly recover the adjacency matrix over the vertices of V from any hypergraph containing the same hyperedges as H (up to reweighing) using only cut queries, hence $\Omega(n^2)$ bits are required to represent it.

The recovery process is based on observing that the difference between any two cuts $S, S \cup \{v\}$ for $v \notin S$ is given by

$$\text{cut}_H(S \cup \{v\}) - \text{cut}_H(S) = \sum_{e \in E} g_e(S \cup \{v\}) - g_e(S) = \sum_{e \in E} \mathbf{1}_{\{v \in e\}} \Delta_{|S \cap e|} w_e,$$

where w_e is the weight of e . Notice that if we find some d such that $\Delta_d < \Delta_{d/2}$, and a cut where exactly one hyperedge e^* has $|S \cap e^*| = d$ while all the hyperedges have $|S \cap e| \leq d/2$ then we can recover whether $v \in e^*$. We create a hypergraph where such cuts exist for every hyperedge $e \in E$ and vertex $v \in V$ using the Hadamard code.

Proof of Claim 5.2. Let $d = 2^k$ be the smallest power of two such that $d \geq t$. Assume for simplicity that $d \leq n/3$. We start by defining the family \mathcal{H} of hypergraphs. Let $H \in \mathcal{H}$ be a hypergraph over n vertices. Split the vertices into three sets, V, U, W with $|V| = |W| = n/6$ and $|U| = 2n/3$. Denote the vertices in each set by $V = \{v_i\}_{i=1}^{n/6}$, $W = \{w_i\}_{i=1}^{n/6}$ and $U = \{u_i\}_{i=1}^{2n/3}$. Notice that since the splitting functions g_e are all cardinality based, they are defined by the hyperedges. H includes exactly $n/6$ hyperedges as described below.

We start by describing the Hadamard code words which we will use in the proof. If $d = 1$ then for all $i \in [n/6]$ set p_i to be strings of length $2n/3$ with 1 in the i -th position and 0 elsewhere. Otherwise, denote the words of the Hadamard code (without the all zeros and all ones words) of length $2d$ by h_1, \dots, h_{2d-2} . Since $2d$ could be much smaller than $2n/3$ we pad h_i with zeros to get words of length $2n/3$. Furthermore, we wish to create a hypergraph with $n/6$ hyperedges and hence if $2d < n/6$ we create $\lceil n/(12d) \rceil$ copies of each word and denote them by $p_{i,j}$, where $p_{i,j}$ is the padded version of h_i shifted by $2d \cdot j$ bits to the right. Observe that by the properties of the Hadamard code $p_{i,j}^2 = d$ and that $p_{i,j} \cdot p_{i',j'} \in \{d/2, 0\}$. For simplicity, we rename $p_{i,j}$ to $p_1, \dots, p_{n/6}$ where $p_k = p_{\lfloor k/(2d) \rfloor, k \bmod 2d}$ dropping extra words if needed.

For each $i \in [n/6]$ let e_i be a union of three sets: $P_i = \{u_j \in U : p_i(j) = 1\}$, a random subset of $n/12$ vertices of V and the singleton $\{w_i\}$.

Every reweighted-subgraph sparsifier H' for H must contain every hyperedge $e \in E$ with weight in $[1 - \epsilon, 1 + \epsilon]$, otherwise the cuts of the singletons, $S = \{w_i\}$ for $i \in [n/6]$, would not be preserved. Fix some reweighted-subgraph sparsifier H' of H , and denote the weights of the hyperedges in H' by w'_e .

Denote the incidence matrix of H corresponding to the vertices in V by B . We show that it is possible to recover B from the cuts of H' this means that every $H \in \mathcal{H}$ requires a unique sparsifier. In addition, there are $\binom{n/6}{n/12}^{n/6}$ possible choices for B and hence the encoding size is $\Omega(n^2)$ bits.

Denote the set of hyperedges containing v_i by $E_i = \{e \in E : v_i \in e\}$. We recover each element B_{ij} by examining the difference

$$\beta_{ij} = \text{cut}_{H'}(P_j \cup \{v_i\}) - \text{cut}_{H'}(P_j) = \sum_{e \in E_i} g_e(P_j) - g_e(P_j \cup \{v_i\}).$$

Recall that by the construction of the code part of the incidence matrix, for every $k \neq j$, $|P_j \cap P_k| \in \{0, d/2\}$. Furthermore, by the definition of d we have $\Delta_{d/2} = \Delta_0 > \Delta_d$. Hence,

$$\beta_{ij} = \begin{cases} \beta_{ij}^1 := \sum_{e \in E_i \setminus \{e_j\}} \Delta_0 w'_e + \Delta_d w'_{e_j}, & \text{if } B_{ij} = 1 \\ \beta_{ij}^0 := \sum_{e \in E_i} \Delta_0 w'_e, & \text{if } B_{ij} = 0. \end{cases}$$

Hence, $\beta_{ij}^1 < \beta_{ij}^0$. In addition, observe that

$$\text{cut}_{H'}(\{v_i\}) = \sum_{e \in E_i} w'_e \hat{g}(1) = \Delta_0 \sum_{e \in E_i} w'_e.$$

Therefore, if $\text{cut}_{H'}(\{v_i\}) = \beta_{ij}$ then $B_{ij} = 0$ and otherwise $B_{ij} = 1$. □

5.2 Directed Hypergraph Encoding Size

In [OST23], the authors provide an $\Omega(n^2/\epsilon)$ lower bound for the number of hyperedges in a reweighted-subgraph sparsifier of directed hypergraphs. We improve on this result in two different ways: (1) Lemma 5.4 shows that in the reweighted-subgraph sparsifier setting, encoding directed

hypergraph cuts requires $\Omega(n^3/\epsilon)$ bits. (2) Theorem 1.13 proves that any encoding of directed hypergraph cuts requires $\Omega(n^2/\epsilon)$ bits (rather than hyperedges).

We begin with Lemma 5.4. This result is based a similar construction to Claim 5.2 without the Hadamard code part. We define a random family \mathcal{H} whose vertices are partitioned into three parts V, U, W . Each hyperedge is defined by a union of two parts: (1) a random subset of the vertices of V , and (2) an unsparifiable part on the vertices of U, W that is based on the construction of [OST23].

To recover the random part of the hypergraph we again turn to comparing different cuts of H' . However, since the hypergraph is much denser we need to isolate the contribution of each hyperedge using the intersection of several cuts.

Lemma 5.4. *There exists a family of hypergraphs \mathcal{H} with the directed all-or-nothing splitting function, such that for every $1/(4\epsilon) < n/3$ encoding a reweighted-subgraph $(1 + \epsilon)$ -sparsifier for an input $H \in \mathcal{H}$ requires $\Omega(n^3)$ bits.*

Proof. Define a hypergraph H over n vertices as follows. Partition the vertices of H into three sets of equal cardinality V, U, W . Denote $V = \{v_i\}_{i=1}^{n/3}$ and similarly for U, W . For every $i, j \in [n/3]$ and $r \in [1, 2, \dots, \frac{1}{8\epsilon}]$ add a hyperedge $e_{i,i+r,j}$ with tail $(e_{i,i+r,j}^T) = \{u_i, u_{i+r \bmod n/3}\}$ and head $(e_{i,i+r,j}^H) = \{w_j\}$. Note that this is the same construction as in [OST23]. Then augment the tail of every hyperedge with a random subset of V , where every vertex $v \in V$ is sampled independently with probability $1/2$.

Observe that for any cut $S \subseteq U \cup W$ the value of the cut is independent of the random bits in the head of the hyperedges. Hence, following the argument in [OST23] any reweighted subgraph sparsifier of H must include all its hyperedges. Let H' be some reweighted subgraph sparsifier for H , we will show that we can recover the random part of the tail of every hyperedge from the cuts of H' . Since H has $\Theta(n^2/\epsilon)$ hyperedges, and each one encodes $n/3$ random bits, storing the any reweighted subgraph sparsifier requires $\Omega(n^3/\epsilon)$ bits.

Denote the modified weights of the hyperedges in H' by $w'(e) : E \rightarrow \mathbb{R}_+$, we will also write $w'(F) = \sum_{e \in F} w'(e)$ for $F \subseteq E$. For every cut set S , denote the set of hyperedges $e \in E$ with $g_e(S) > 0$ by $E(S)$. Examine the cut set $S_{i,j} := \{u_i\} \cup (W \setminus \{w_j\})$ and notice that $E(S_{i,j}) = \{e_{i,i+r,j} : r \in [1/(8\epsilon)]\} \cup \{e_{i-r,i,j} : r \in [1/(8\epsilon)]\}$.

We now describe the process for determining whether v_k is in $e_{i,i+x,j}^T$ for some $x \in [1, 2, \dots, \frac{1}{8\epsilon}]$. Observe that $E(S_{i,j}) \cap E(S_{i+x,j}) = \{e_{i,i+x,j}\}$, and since $e_{i,i+x,j} \in H'$, we find that $v_k \in e_{i,i+x,j}^T$ if $w'(E(S_{i,j}) \cap E(S_{i+x,j}) \cap E(\{v_k\})) > 0$. To find the value of $w'(E(S_{i,j}) \cap E(S_{i+x,j}) \cap E(\{v_k\}))$, observe that for every $S \subseteq U \cup W$ we can find $w'(E(S) \cap E(\{v_k\}))$ by the following method.

Begin by noting that,

$$\begin{aligned} \text{cut}_{H'}(S \cup \{v_k\}) - \text{cut}_{H'}(S) &= w'(E(S \cup \{v_k\}) \setminus E(S)) = w'(E(\{v_k\}) \setminus E(S)) \\ &= w'(E(\{v_k\})) - w'(E(S) \cap E(\{v_k\})), \end{aligned}$$

where the last equality is by the directed all-or-nothing splitting function and that V is disjoint from the heads of the hyperedges. Furthermore, observe that $\text{cut}_{H'}(\{v_k\}) = w'(E(\{v_k\}))$ and hence,

$$w'(E(S) \cap E(\{v_k\})) = \text{cut}_{H'}(\{v_k\}) - (\text{cut}_{H'}(S \cup \{v_k\}) - \text{cut}_{H'}(S)). \quad (11)$$

To conclude the proof observe that

$$\begin{aligned} w'(E(S_{i,j}) \cap E(S_{i+x,j}) \cap E(\{v_k\})) &= w'(E(S_{i,j}) \cap E(\{v_k\})) + w'(E(S_{i+x,j}) \cap E(\{v_k\})) \\ &\quad - w'(E(S_{i,j} \cup S_{i+x,j}) \cap E(\{v_k\})). \end{aligned}$$

We can find all the terms on the right-hand side using Equation (11), and hence we can determine for every k if v_k is in $e_{i,i+x,j}^T$. Therefore, representing any reweighted subgraph sparsifier for H requires $\Omega(n^3/\epsilon)$ bits. \square

We now turn to proving that representing directed hypergraph cuts requires $\Omega(n^2/\epsilon)$ bits in any data structure (Theorem 1.13). The proof constructs of hypergraphs \mathcal{H} based on the construction in [OST23], with an added sampling step. We then show that every $(1 + \epsilon)$ -sparsifier of some hypergraph $H \in \mathcal{H}$ does not $(1 + \epsilon)$ -approximate the cuts of any other member of the family. Therefore, every hypergraph in the family requires a unique sparsifier. The lower bound follows by showing there are $2^{\Omega(n^2/\epsilon)}$ hypergraphs in the family.

Proof of Theorem 1.13. Define a family of hypergraphs \mathcal{H} as follows. Let $H \in \mathcal{H}$ be a hypergraph with $2n$ vertices, and partition its vertex set into two disjoint sets of equal cardinality U, W . Throughout the proof we assume that $1/(16\epsilon)$ is an integer for simplicity. For every $i, j \in [n]$ sample a uniform subset of size $\frac{1}{16\epsilon}$ from $[\frac{1}{8\epsilon}]$ and denote it by $V_{i,j}$. For every $x \in V_{i,j}$ add the hyperedges $e_{i,i+x,j}$ with tail $e_{i,i+x,j}^T = \{u_i, u_{i+x \bmod n}\}$ and head $e_{i,i+x,j}^H = \{w_j\}$ to H .

Fix some $H \in \mathcal{H}$. We will show that every quality $(1 + \epsilon)$ -sparsifier for H does not $(1 + \epsilon)$ -approximate any other hypergraph $\hat{H} \in \mathcal{H}$. Therefore, this family implies the existence of at least $\Omega(2^{n^2/\epsilon})$ distinct sparsifiers. Hence, representing a $(1 + \epsilon)$ -approximation of the cuts of any $H \in \mathcal{H}$ requires $\Omega(n^2/\epsilon)$ bits.

Assume without loss of generality that the hyperedge $e_{1,2,1}$ is in H but not in \hat{H} . Let $S_i = \{u_i\} \cup \{W \setminus \{w_1\}\}$. Observe that by the symmetry of the construction,

$$\text{cut}_H(S_i) \in \left\{ \frac{1}{16\epsilon}, \frac{1}{16\epsilon} + 1, \dots, \frac{1}{8\epsilon} \right\}.$$

If there exists some i such that $\text{cut}_H(S_i) \neq \text{cut}_{\hat{H}}(S_i)$ then any quality $1 + \epsilon$ sparsifier for \hat{H} does not approximate H , since

$$\left| \frac{\text{cut}_{\hat{H}}(S_i)}{\text{cut}_H(S_i)} - 1 \right| \geq \left| \frac{\frac{1}{8\epsilon} - 1}{\frac{1}{8\epsilon}} - 1 \right| = 8\epsilon.$$

Hence, we proceed with the case where H, \hat{H} have the same value for every cut S_i . Assume there exists some quality $(1 + \epsilon)$ -sparsifier H' for both H, \hat{H} . Now observe that for H

$$\text{cut}_H(S_1 \cup S_2) = \text{cut}_H(S_1) + \text{cut}_H(S_2) - 1,$$

and since H' is a quality $(1 + \epsilon)$ -sparsifier for H , then

$$\text{cut}_{\hat{H}}(S_1 \cup S_2) \geq (1 - \epsilon)(\text{cut}_H(S_1) + \text{cut}_H(S_2) - 1).$$

We can also get an upper bound on the cut $S_1 \cup S_2$ in \hat{H} by observing that since $e_{1,2,1} \notin \hat{H}$ then

$$\text{cut}_{\hat{H}}(S_1 \cup S_2) = \text{cut}_{\hat{H}}(S_1) + \text{cut}_{\hat{H}}(S_2) = \text{cut}_H(S_1) + \text{cut}_H(S_2),$$

where the second equality is by our assumption that H, \hat{H} have the same value for every cut S_i . Since H' also $(1 + \epsilon)$ -approximates the cuts of \hat{H} we have

$$\text{cut}_{H'}(S_1 \cup S_2) \leq (1 + \epsilon)(\text{cut}_H(S_1) + \text{cut}_H(S_2)).$$

However,

$$(1 - \epsilon)(\text{cut}_H(S_1) + \text{cut}_H(S_2) - 1) \leq (1 + \epsilon)(\text{cut}_H(S_1) + \text{cut}_H(S_2)),$$

whenever $\text{cut}_H(S_1) + \text{cut}_H(S_2) \leq \frac{1}{4\epsilon}$. Therefore H' does not $(1 + \epsilon)$ approximate both H, \hat{H} and every hypergraph in \mathcal{H} requires a unique sparsifier. \square

6 Deformation Lower Bounds

In this section we prove lower bound on the support size for approximating several families of splitting functions. In particular, we show a lower bound for additive splitting function (Theorem 1.14). The results are all based on the following technical lemma, which we prove at the end of the section.

Lemma 6.1. *Let e be a hyperedge with a splitting function $g_e : 2^e \rightarrow \mathbb{R}_+$. For every $S, T \subseteq e$ such that $|S| = |T| = t$, denote*

$$\delta_t(S, T) := 1 - \frac{g_e(S \cup T)}{g_e(S) + g_e(T)}.$$

If for some $t < |e|/2$ at least a ρ -fraction of the pairs $(S, T) \in \binom{e}{t}$ satisfy $\delta_t(S, T) > \hat{\delta}$ for some $\hat{\delta}$ such that $\rho \hat{\delta}^2 \geq \Omega(|e|^{-1/2})$, then every $(1 + \hat{\delta}/2)$ -approximation of e must have support size at least $\Omega(\rho \hat{\delta}^2 |e|/t)$.

Informally, the lemma states that if a splitting function is far from linear on a large enough fraction of pairs of sets of size t , then it cannot be closely approximated by a sum of hyperedges with small support. The lemma is based on identifying

$$\delta_t(S, T) := 1 - \frac{g_e(S \cup T)}{g_e(S) + g_e(T)}$$

as a quantity that describes how close to linear is the function g_e for subsets of size t , $S, T \subseteq e$. We then show that if the function is far from linear on a large enough fraction, of pairs of sets of size t , then it cannot be closely approximated by a sum of hyperedges with small support.

Note that δ_t is related to the notion of curvature of submodular functions, the *total curvature* of a submodular function g_e is given by

$$c_{g_e} := 1 - \min_{S \subseteq e, v \in e \setminus S} \frac{g_e(S \cup \{v\}) - g_e(S)}{g_e(\{v\})}.$$

Intuitively, the curvature describes how far from linear the function g_e is in the worst case. The curvature is used to parametrize the quality of approximation in maximization of submodular functions; where if a function has low curvature, hence it is close to linear, then it is possible to achieve a better approximation [CC84, Von10].

The quantity δ_t differs from the curvature in two regards. First, δ_t describes a relation two sets of size t and not the marginal contribution of adding a single element. Second, in the optimization setting the guarantees depend on the worst case curvature, while our lemma requires δ_t to be large only on a constant fraction of subsets of size t .

6.1 Support Size Lower Bounds for Approximating Splitting Functions

This section proves support size lower bounds for approximating several families of common splitting functions. A summary of the results is provided in Table 1.

We begin by presenting several results for different families of cardinality based splitting functions. For cardinality based splitting functions the value $\delta_t(S, T)$ only depends on $|S \cup T|$, therefore it is possible to find a uniform bound on $\delta_t(S, T)$ for all sets of size t . This idea is formalized in the following corollary of Lemma 6.1.

Corollary 6.2 (Lemma 6.1 for Cardinality Based Splitting Functions). *Let e be a hyperedge with a cardinality based splitting function $g_e : 2^e \rightarrow \mathbb{R}_+$. For every $t \leq |e|$ denote*

$$\bar{\delta}_t := 1 - \max_{S_1, S_2 \subseteq V: |S_1|=|S_2|=t} \frac{g_e(S_1 \cup S_2)}{g_e(S_1) + g_e(S_2)}.$$

Suppose $\bar{\delta}_t > 0$ for some $t \leq |e|/2$, then every $(1 + \bar{\delta}_t/2)$ -approximation of e must have support size at least $\Omega(\bar{\delta}_t^2 |e|/t)$.

Proof. Let $\rho = 1$, and observe that $\delta_t(S, T) \geq \bar{\delta}_t$ for every $S, T \subseteq e$ of size t . Applying Lemma 6.1 concludes the proof. \square

We begin with the lower bound for additive splitting functions (Theorem 1.14).

Proof. Let e be a hyperedge with an additive splitting function g_e , with parameter K . We will show that every 1.1-approximation of g_e requires support size $\Omega(|e|/K)$. Note that if $K > |e|/2$ then $\Omega(|e|/K) = \Omega(1)$ and the lower bound is trivial. Otherwise, note that for $t = K$ we have $\bar{\delta}_t \geq 1/2$. Applying Corollary 6.2 we find that every 1.1-approximation of e requires support size at least $\Omega(|e|/K)$. \square

We also provide results for polynomial and logarithmic cardinality based splitting functions. Both results are based on identifying a constant t such that $\bar{\delta}_t$ is strictly positive constant.

Corollary 6.3 (Lower Bound for Polynomial Cardinality Based Splitting Functions). *Let $\hat{g}(S) = |S|^\beta$ or $\hat{g}(S) = \min(|S|^\beta, |\bar{S}|^\beta)$ with $\beta \in (0, 0.999)$ and let e be a hyperedge with $g_e(S) = \hat{g}(|S|)$ as its splitting function. Every $(1 + (2^{-1} - 2^{\beta-2}))$ -approximation of g_e must have support size at least $\Omega((2^{-1} - 2^{\beta-2})^2 |e|)$.*

Proof. Notice that for any $t \geq 1$, we have

$$\bar{\delta}_t = 1 - \frac{(2t)^\beta}{2t^\beta} = 1 - 2^{\beta-1}.$$

Applying Corollary 6.2 with $t = 1$ concludes the proof. \square

Corollary 6.4. *Let $\hat{g}(S) = \log(|S|)$ or $\hat{g}(S) = \min(\log(|S| + 1), \log(|\bar{S}| + 1))$ and let e be a hyperedge with $g_e(S) = \hat{g}(|S|)$ as its splitting function. Then every $1 + 1/5$ -approximation of g_e must have support size at least $\Omega(|e|)$.*

Proof. Observe that for $t = e^5$, we have

$$\bar{\delta}_t = 1 - \frac{\log(2e^5)}{2\log(e^5)} > 2/5.$$

Applying Corollary 6.2 with $t = e^5$ concludes the proof. \square

We also present a general lower bound for all cardinality based splitting functions that is characterized by the spread. The proof is based on showing that if δ_t is small for all $t < r$, for $r \in \mathbb{N}$, then $g_e(2^r) \geq c^r$ with $c > 1$. Hence, there exists some $t < \log(\mu_e)$ with a large δ_t . We can then apply Corollary 6.2 for this t .

Corollary 6.5. *Let e be a hyperedge with cardinality based splitting function $g_e(S) = \hat{g}(|S|)$, for some $\hat{g} : [|e|] \rightarrow \mathbb{R}_+$. For every $\epsilon < 1/4$, every $(1 + \epsilon)$ -approximation of e requires support size at least $\Omega\left(\epsilon^2|e|/\mu_e^{\log_2^{-1}(2-4\epsilon)}\right)$.*

Proof. Assume without loss of generality that $\hat{g}(1) = 1$. Note that for every t , if $\bar{\delta}_t \leq 2\epsilon$ then

$$2\epsilon \geq 1 - \frac{\hat{g}(2t)}{2\hat{g}(t)},$$

and hence $\hat{g}(2t) \geq (2 - 4\epsilon)\hat{g}(t)$. Therefore, if $\bar{\delta}_t \leq 2\epsilon$ for all $t \leq r$ with $r \in \mathbb{N}$, then $\hat{g}(2^r) \geq (2 - 4\epsilon)^r$. However, since the maximum of the splitting function $\max_{i \in [|e|]} \hat{g}(i) = \mu_e$ there exists some $t < 2^{\log_2^{-1}(2-4\epsilon)} + 1 \leq \mu_e^{\log_2^{-1}(2-4\epsilon)} + 1$ such that $\bar{\delta}_t > 2\epsilon$. The lower bound follows from applying Corollary 6.2 for this t . \square

Finally, we present a generalization of Corollary 6.5 to all *unweighted splitting functions*. A splitting function is called unweighted if all its singleton cuts are equal to 1, i.e. $g_e(\{v\}) = 1$ for all $v \in e$. One natural example of a family of unweighted splitting functions are matroid rank functions.

The proof is similar to the cardinality based case, but in the unweighted case we have an additional complication as not all cuts of size t have the same value. Therefore, we lower bound both the value of the splitting function for sets of size t (as in the cardinality based case) and the fraction of sets S of size t for which $g_e(S)$ is at least this value. To simplify the proof we focus on pairs of sets of size t that are disjoint. Specifically, we show that for some $t \leq \mu_e^{\Omega(1)}$, at least a $\mu_e^{-\Omega(1)}$ fraction of disjoint pairs of sets have $\delta_t(S, T) > 2\epsilon$. This technique introduces an additional $\mu_e^{\Omega(1)}$ factor in the lower bound in comparison to Corollary 6.5 as we apply Lemma 6.1 for only $\rho = \mu_e^{-\Omega(1)}$ fraction of pairs.

Corollary 6.6. *Let $\epsilon < 1/4$ and denote $\gamma = \log_2^{-1}(2 - 4\epsilon)$. In addition, let e be a hyperedge with an unweighted splitting function $g_e(S)$ such that $\mu_e < |e|^{1/(2\gamma)}$. Then, every $(1 + \epsilon)$ -approximation of e must have support size at least $\Omega\left(\epsilon^2|e|/\mu_e^{2\gamma}\right)$.*

Proof. Let $D_i = \{(S, T) \in \binom{e}{i} \times \binom{e}{i} : S \cap T = \emptyset\}$ be the set of disjoint pairs of subsets of size i of e . Also, let $P_i = \{(S, T) \in D_i : \delta_i(S, T) < 2\epsilon\}$ be subset of D_i composed of all pairs with $\delta_i(S, T) < 2\epsilon$. Denote $\alpha = 2 - 4\epsilon$ and let $R_i = \{(S, T) \in P_i : f(T) \geq \alpha^i, f(S) \geq \alpha^i\}$ be the subset of P_i such that $f(S), f(T) \geq \alpha^i$. Observe that by the definition of μ_e , $R_{\log_\alpha(2\mu_e)}$ must be empty.

Fix $\beta = \mu_e^{-\gamma}/16$. We will show that if $|P_{2^j}| \geq (1 - \beta)|D_{2^j}|$ for all $j \leq \log_\alpha(2\mu_e)$ then $R_{\log_\alpha(2\mu_e)}$ is nonempty and hence this leads to contradiction. Therefore, there exists some $q < 2^{\log_\alpha(2\mu_e)} \leq (2\mu_e)^\gamma$ such that $|P_q| \leq (1 - \beta)|D_q|$. Hence, at least $\beta|D_q|$ disjoint pairs of size q have $\delta_t(S, T) > 2\epsilon$. Furthermore, for all $q < \sqrt{|e|}$, we have that $|D_q|$ is at least $1/10$ fraction of all pairs of sets of size t by the following claim, which we prove later.

Claim 6.7. *For all $q < \sqrt{|e|}$, $|D_q| > \binom{|e|}{q}/10$.*

Note that we can apply the claim as $q \leq \mu_e^\gamma < \sqrt{|e|}$ by the theorem statement. Hence, at least $\beta/10$ fraction of pairs (S, T) of size q have $\delta_q(S, T) > 2\epsilon$. Finally, using Lemma 6.1 with $t = q \leq (2\mu_e)^\gamma$, $\hat{\delta} = 2\epsilon$ and $\rho = \beta/10 = \mu_e^{-\gamma}/160$ we find that every $(1 + \epsilon)$ -approximation of e requires support size $\Omega\left(\epsilon^2|e|/\mu_e^{2\gamma}\right)$. Note that we can apply the theorem as $\rho\hat{\delta}^2 \geq \Omega(|e|^{-1/2})$.

To finish the proof we now show that $|P_{2j}| \geq (1 - \beta)|D_{2j}|$ for all $j \leq \log_\alpha(2\mu_e)$ implies that $R_{\log_\alpha(2\mu_e)}$ is nonempty. Denote the size of $|R_{2j}| = (1 - f(j))|D_{2j}|$, we will define $f(j)$ recursively, noting that $f(0) = 0$ since g_e is unweighted. Let $S \in \binom{e}{2j}$, and observe that S can be partitioned into two disjoint subsets S_1, S_2 of size 2^{j-1} . If both these subsets are in the intersection of $R_{2^{j-1}}$ and $P_{2^{j-1}}$ then

$$f(S) = f(S_1 \cup S_2) \geq \alpha(g_e(S_1) + g_e(S_2)) \geq \alpha^{r+1},$$

where the first inequality is by $\delta_{2^{j-1}}(S_1, S_2) < 2\epsilon$ as $(S_1, S_2) \in P_{2^{j-1}}$ and the second inequality is since for every $(S_1, S_2) \in R_{2^{j-1}}$ we have $g_e(S_1), g_e(S_2) \geq \alpha^{2^{j-1}}$. Hence, S can be a member in pairs of R_{2j} . By our assumption $|P_{2^{j-1}}| > (1 - \beta)|D_{2^{j-1}}|$, and hence $R_{2^{j-1}} \cap P_{2^{j-1}} \geq (1 - f(j-1) - \beta)|D_{2^{j-1}}|$. Therefore, the fraction of disjoint pairs $(S, T) \in D_{2j}$ where both S, T can be partitioned into disjoint subsets S_1, S_2 and T_1, T_2 that are in $R_{2^{j-1}} \cap P_{2^{j-1}}$ is at least $(1 - f(j-1) - \beta)^2$. We can now bound $f(j)$ by,

$$f(j) = 2f(j-1) + 2\beta - (f(j-1) + \beta)^2 \leq 2(f(j-1) + \beta).$$

Solving this recursive formula we find $f(j) \leq 2^{j+1}\beta$. Recalling $\beta = \mu_e^{-\gamma}/16$ we get $f(\log_\alpha(2\mu_e)) > 1/2$ and hence $R_{\log_\alpha(2\mu_e)}$ is nonempty in contradiction. Therefore, there exists some $q < \mu_e^\gamma$ such that $|P_q| \leq (1 - \beta)|D_q|$.

To finish the proof we turn back to proving Claim 6.7.

Proof. Let S, T be two random subsets of size j of e . Denote the event that S, T are disjoint by $D_{S,T}$. Observe that

$$\begin{aligned} \Pr[D_{S,T}] &= \binom{|e|-j}{|e|} \binom{|e|}{j}^{-1} = \frac{(|e|-j)! (|e|-j)!}{|e|! (|e|-2j)!} = \prod_{i \in \{|e|, |e|-1, \dots, |e|-j+1\}} \frac{i-j}{i} \\ &= \prod_{i \in \{|e|, |e|-1, \dots, |e|-j+1\}} 1 - \frac{j}{i} \geq \left(1 - \frac{2j}{|e|}\right)^j \geq \left(1 - \frac{2}{\sqrt{|e|}}\right)^{\sqrt{|e|}} \geq e^{-2} \left(1 - \frac{4}{\sqrt{|e|}}\right), \end{aligned}$$

where the first inequality is by $|e| - j > n/2$ for all $j < \sqrt{|e|}$, the second by $j < \sqrt{|e|}$ and the third by $(1 - x/k)^k \geq e^{-x}(1 - x^2/|e|)$. This expression is larger than $1/10$ for all $|e| > 250$. Therefore, the fraction of disjoint pairs out of all pairs of subsets of size j is at least $1/10$. \square

This concludes the proof of Corollary 6.6. \square

6.2 Proof of Lemma 6.1

We now return to proving Lemma 6.1.

Proof of Lemma 6.1. Throughout the proof we denote $|e| = n$. Let $p = \alpha n/t$ for $\alpha > 0$ to be determined later. Assume on the contrary that there exists a set of hyperedges of cardinality at most p such that the sum of their splitting functions approximates g_e with quality $q = 1 + \hat{\delta}/2$. Let $\{e_i\}_{i=1}^k$ be all the possible subsets of e of size p and denote $k = \binom{n}{p}$. Note that we can assume that $p \geq 2$ as otherwise the lower bound is trivial. Since a sum of submodular functions is submodular, this is the most general case for decomposing e into hyperedges with maximal support p as any two hyperedges e_1, e_2 with $e_1 \subseteq e_2$ can be combined into a single hyperedge with $g_{e_1} + g_{e_2}$ as its splitting function.

Starting with some notation let $I_T := \{i \in [k] : e_i \cap T \neq \emptyset\}$ and $h(S, I) := \sum_{i \in I} g_{e_i}(S \cap e_i)$. Choose some subsets of size t of e $S_1, S_2 \subseteq e$ with $\delta_t(S_1, S_2) \geq \hat{\delta}$. Notice that $h(S_1 \cup S_2, I_{S_1} \setminus I_{S_2}) = h(S_1, I_{S_1} \setminus I_{S_2})$, and hence we can write

$$\begin{aligned} h(S_1 \cup S_2, [k]) &= h(S_1, I_{S_1} \setminus I_{S_2}) + h(S_2, I_{S_2} \setminus I_{S_1}) + h(S_1 \cup S_2, I_{S_2} \cup I_{S_1}) \\ &\leq (1 + \hat{\delta}/2)g_e(S_1 \cup S_2) \leq (1 + \hat{\delta}/2)(1 - \hat{\delta})(g_e(S_1) + g_e(S_2)), \end{aligned} \quad (12)$$

where the first inequality is by our assumption that $h(S, [k])$ $(1 + \hat{\delta}/2)$ -approximates $g_e(S)$ and the second is since $\delta_t(S_1, S_2) \geq \hat{\delta}$. Observe that also

$$h(S_1, [k]) = h(S_1, I_{S_1} \setminus I_{S_2}) + h(S_1, I_{S_1} \cap I_{S_2}) \geq (1 - \hat{\delta}/2)g_e(S_1),$$

where the inequality is again by our assumption that $h(S, [k])$ $(1 + \hat{\delta}/2)$ -approximates $g_e(S)$. Let $\beta_1 := h(S_1, I_{S_1} \cap I_{S_2})/g_e(S_1)$, we can then write $h(S_1, I_{S_1} \setminus I_{S_2}) \geq (1 - \hat{\delta}/2 - \beta_1)g_e(S_1)$. Similarly observe that $h(S_2, I_{S_2} \setminus I_{S_1}) \geq (1 - \hat{\delta}/2 - \beta_2)g_e(S_2)$ and denote $\beta := \max\{\beta_1, \beta_2\}$. Substituting back into Equation (12) we find

$$(1 - \hat{\delta}/2 - \beta)g_e(S_2) + (1 - \hat{\delta}/2 - \beta)g_e(S_1) \leq (1 + \hat{\delta}/2)(1 - \hat{\delta})(g_e(S_1) + g_e(S_2)).$$

Therefore, we find $\beta = \max\{\beta_1, \beta_2\} \geq \hat{\delta}^2/2$. Recalling the definition of β this implies that either $h(S_1, I_{S_1} \cap I_{S_2}) > \hat{\delta}^2 g_e(S_1)/2$ or $h(S_2, I_{S_1} \cap I_{S_2}) > \hat{\delta}^2 g_e(S_2)/2$. By the theorem assumption there are $\rho \binom{n}{2}$ pairs of sets $A, B \subseteq e$ with $\delta_t(A, B) \geq \hat{\delta}$, and following the same argument in each pair at least one of A, B satisfies this lower bound. Hence, there must be at least one set $T \subseteq e$ of size t for which the lower bound is satisfied in at least $\rho \binom{n}{t}^{-1} \binom{n}{2} \geq \rho \binom{n}{t}/4$ pairs.

For every $i \in [k]$ denote the set $Q_i = \{P \subseteq e : |P| = t, P \cap e_i \neq \emptyset, P \neq T\}$. Note that by symmetry $|Q_i| = |Q|$ for all $i \in I_T$. Furthermore, we can use the following lemma to bound the size of $|Q_i|$, the proof of the lemma is provided later.

Claim 6.8. *If $\alpha \leq \rho \hat{\delta}^2/16$, then $|Q_i| < \frac{\rho \hat{\delta}^2}{8(1 + \hat{\delta}/2)} \binom{n}{t}$ for every $i \in I_T$.*

Using this notation we can rewrite $h(T, [k])$ as

$$\begin{aligned} h(T, [k]) &= \sum_{i \in I_T} g_{e_i}(T) = \sum_{i \in I_T} \frac{1}{|Q_i|} \sum_{P \in Q_i} g_{e_i}(T) \\ &= \sum_{P \in \binom{e}{t} \setminus \{T\}} \sum_{i \in I_T \cap I_P} \frac{1}{|Q_i|} g_{e_i}(T) = \frac{1}{|Q|} \sum_{P \in \binom{e}{t} \setminus \{T\}} h(T, I_T \cap I_P), \end{aligned}$$

where the second equality is since for every g_{e_i} in the inner sum, $i \in I_T, I_P$. By the definition of T we know that for at least $\rho \binom{n}{t}/4$ of the sets $P \in \binom{e}{t}$ satisfy $h(T, I_T \cap I_P) > \hat{\delta}^2/2 \cdot g_e(T)$. Hence,

$$h(T, [k]) \geq \frac{\rho \hat{\delta}^2}{8|Q|} \binom{n}{t} g_e(T) > (1 + \hat{\delta}/2)g_e(T).$$

Where the last inequality is by Claim 6.8. Therefore, $h(T, [k])$ does not $(1 + \hat{\delta}/2)$ -approximates $g_e(T)$ in contradiction to our assumption. It remains to prove Claim 6.8.

Proof of Claim 6.8. Recall $Q_i = \{P \subseteq V : |P| = t, P \cap e_i \neq \emptyset, P \neq T\}$, hence the number of elements in Q_i is the total number of sets of size t minus the number of sets that don't intersect e_i minus 1. Formally, this is equal to $|Q_i| = \binom{n}{t} - \binom{n-p}{t} - 1$. Examine,

$$\begin{aligned} \binom{n}{t}^{-1} \cdot \binom{n-p}{t} &= \frac{t!(n-t)!}{n!} \cdot \frac{(n-p)!}{t!(n-p-t)!} = \frac{(n-p)!}{n!} \frac{(n-t)!}{(n-p-t)!} = \prod_{i=n}^{n-p+1} \frac{i-t}{i} \\ &= \prod_{i=n}^{n-p+1} 1 - \frac{t}{i} \geq \prod_{i=n}^{n-p+1} 1 - \frac{t}{n-p} = \left(1 - \frac{t}{n-p}\right)^{p-1}. \end{aligned}$$

Substituting $p = \alpha n/t$ and observing $\alpha = \rho \hat{\delta}^2/16 < 1/2$, we find

$$\begin{aligned} \binom{n}{t}^{-1} \cdot \binom{n-p}{t} &\geq \left(1 - \frac{t}{n - \alpha n/t}\right)^{\alpha n/(2t)} = \left(1 - \frac{t}{n} \frac{1}{1 - \alpha/t}\right)^{\alpha n/(2t)} \\ &\geq \exp(-2\alpha) \left(1 - \left(\frac{1}{1 - \alpha/t}\right)^2 \frac{t}{n}\right)^\alpha, \end{aligned}$$

where the first inequality is from $p \geq 2$. The second inequality stems from $(1 - x/k)^k \geq e^x(1 - x^2/k)$ and $(1 - \alpha/t)^{-1} < 2$ since $\alpha < 1/2, t \geq 1$. We now split the analysis into the case where $t < n^{1/3}$ and its complement. When $t < n^{1/3}$, we have

$$\binom{n}{t}^{-1} \cdot \binom{n-p}{t} \geq \exp(-2\alpha) \left(1 - 4 \frac{1}{n^{2/3}}\right) \geq \exp(-3\alpha),$$

where the first inequality is from $t < n^{1/3}$, and the second is from $(1 - 4/n^{2/3}) \geq e^{-\alpha}$ whenever $\alpha \geq \Omega(1/\sqrt{n})$ and n is large enough. For the case when $t \geq n^{1/3}$ observe that

$$1 - \left(\frac{1}{1 - \alpha/t}\right)^2 \frac{t}{n} > \frac{1}{3},$$

for all $n > N$ for some $N > 0$. Hence,

$$\binom{n}{t}^{-1} \cdot \binom{n-p}{t} \geq \exp(-(2 + \log(3))\alpha) \geq \exp(-4\alpha),$$

Overall, we find that in both cases $\binom{n}{t}^{-1} \cdot \binom{n-p}{t} \geq e^{-4\alpha}$. Plugging this back to bound the size of $|Q_i|$ we find

$$|Q_i| = \binom{n}{t} - \binom{n-p}{t} - 1 < (1 - e^{-4\alpha}) \binom{n}{t} - 1.$$

To finish proving the claim we choose α such that

$$\frac{\rho \hat{\delta}^2}{8(1 + \hat{\delta}/2)} \geq 1 - e^{-4\alpha} > 1 - e^{-4\alpha} - \binom{n}{t}^{-1}.$$

Moving sides,

$$e^{-4\alpha} \geq 1 - \frac{\hat{\delta}^2}{8(1 + \hat{\delta}/2)},$$

and using $\log(1 - x) < -x$ for $x \in (0, 1)$, we find that

$$\alpha \leq \frac{\rho \hat{\delta}^2}{16}$$

satisfies the requirement. □

This concludes the proof of Lemma 6.1. □

References

- [ACK⁺16] Alexandr Andoni, Jiecao Chen, Robert Krauthgamer, Bo Qin, David P. Woodruff, and Qin Zhang. On sketching quadratic forms. In *Innovations in Theoretical Computer Science*, ITCS'16, pages 311–319. ACM, 2016. doi:10.1145/2840728.2840753.
- [ADK⁺16] Ittai Abraham, David Durfee, Ioannis Koutis, Sebastian Krinninger, and Richard Peng. On fully dynamic graph sparsifiers. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS*, pages 335–344. IEEE Computer Society, 2016.
- [AGK14] Alexandr Andoni, Anupam Gupta, and Robert Krauthgamer. Towards $(1 + \epsilon)$ -approximate flow sparsifiers. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 279–293. SIAM, 2014.
- [AS20] Sepehr Assadi and Sahil Singla. Improved truthful mechanisms for combinatorial auctions with submodular bidders. *SIGecom Exch.*, 18(1):19–27, 2020. doi:10.1145/3440959.3440964.
- [BK96] András A. Benczúr and David R. Karger. Approximating s - t minimum cuts in $\tilde{O}(n^2)$ time. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 47–55. ACM, 1996. doi:10.1145/237814.237827.
- [BK15] András A. Benczúr and David R. Karger. Randomized approximation schemes for cuts and flows in capacitated graphs. *SIAM J. Comput.*, 44(2):290–319, 2015. doi:10.1137/070705970.
- [BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Rev.*, 56(2):315–334, 2014. doi:10.1137/130949117.
- [BST19] Nikhil Bansal, Ola Svensson, and Luca Trevisan. New notions and constructions of sparsification for graphs and hypergraphs. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019*, pages 910–928. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00059.
- [CC84] Michele Conforti and Gérard Cornuéjols. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discret. Appl. Math.*, 7(3):251–274, 1984. doi:10.1016/0166-218X(84)90003-9.
- [CCPS21] Ruoxu Cen, Yu Cheng, Debmalya Panigrahi, and Kevin Sun. Sparsification of directed graphs via cut balance. In *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021*, volume 198 of *LIPICs*, pages 45:1–45:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPICs.ICALP.2021.45.
- [CKKL12] Mahdi Cheraghchi, Adam R. Klivans, Pravesh Kothari, and Homin K. Lee. Submodular functions are noise stable. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012*, pages 1586–1592. SIAM, 2012. doi:10.1137/1.9781611973099.126.
- [CKN20] Yu Chen, Sanjeev Khanna, and Ansh Nagda. Near-linear size hypergraph cut sparsifiers. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 61–72. IEEE, 2020.
- [CKST19] Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. Optimal lower bounds for sketching graph cuts. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2565–2569, 2019.
- [dCSHS16] Marcel Kenji de Carli Silva, Nicholas J. A. Harvey, and Cristiane M. Sato. Sparse sums of positive semidefinite matrices. *ACM Trans. Algorithms*, 12(1):9:1–9:17, 2016. doi:10.1145/2746241.
- [DDS⁺13] Nikhil R. Devanur, Shaddin Dughmi, Roy Schwartz, Ankit Sharma, and Mohit Singh. On the approximation of submodular functions. *CoRR*, abs/1304.4948, 2013. arXiv:1304.4948.
- [DS06] Shahar Dobzinski and Michael Schapira. An improved approximation algorithm for combinatorial auctions with submodular bidders. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006*, pages 1064–1073. ACM Press, 2006.

- [Fei09] Uriel Feige. On maximizing welfare when utility functions are subadditive. *SIAM J. Comput.*, 39(1):122–142, 2009. doi:10.1137/070680977.
- [FK14] Vitaly Feldman and Pravesh Kothari. Learning coverage functions and private release of marginals. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 679–702. JMLR.org, 2014.
- [FKV13] Vitaly Feldman, Pravesh Kothari, and Jan Vondrák. Representation, approximation and learning of submodular functions using low-rank decision trees. In *COLT 2013 - The 26th Annual Conference on Learning Theory*, volume 30 of *JMLR Workshop and Conference Proceedings*, pages 711–740. JMLR.org, 2013.
- [FV06] Uriel Feige and Jan Vondrák. Approximation algorithms for allocation problems: Improving the factor of $1 - 1/e$. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 667–676. IEEE Computer Society, 2006. doi:10.1109/FOCS.2006.14.
- [FV16] Vitaly Feldman and Jan Vondrák. Optimal bounds on approximation of submodular and XOS functions by juntas. *SIAM J. Comput.*, 45(3):1129–1170, 2016. doi:10.1137/140958207.
- [GHIM09] Michel X. Goemans, Nicholas J. A. Harvey, Satoru Iwata, and Vahab S. Mirrokni. Approximating submodular functions everywhere. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2009*, pages 535–544. SIAM, 2009.
- [GHRU13] Anupam Gupta, Moritz Hardt, Aaron Roth, and Jonathan R. Ullman. Privately releasing conjunctions and the statistical query barrier. *SIAM J. Comput.*, 42(4):1494–1520, 2013. doi:10.1137/110857714.
- [GK10] Ryan Gomes and Andreas Krause. Budgeted nonparametric learning from data streams. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 391–398. Omnipress, 2010.
- [HKNR98] Torben Hagerup, Jyrki Katajainen, Naomi Nishimura, and Prabhakar Ragde. Characterizing multiterminal flow networks and computing flows in networks of small treewidth. *J. Comput. Syst. Sci.*, 57(3):366–375, 1998. doi:10.1006/jcss.1998.1592.
- [JLLS23] Arun Jambulapati, James R. Lee, Yang P. Liu, and Aaron Sidford. Sparsifying sums of norms. *CoRR*, abs/2305.09049, 2023. arXiv:2305.09049.
- [Kar93] David R. Karger. Global min-cuts in rnc, and other ramifications of a simple min-cut algorithm. In *Proceedings of the Fourth Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms*, pages 21–30. ACM/SIAM, 1993.
- [KG11] Andreas Krause and Carlos Guestrin. Submodularity and its applications in optimized information gathering. *ACM Trans. Intell. Syst. Technol.*, 2(4):32:1–32:20, 2011. doi:10.1145/1989734.1989736.
- [KK15] Dmitry Kogan and Robert Krauthgamer. Sketching cuts in graphs and hypergraphs. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015*, pages 367–376. ACM, 2015. doi:10.1145/2688073.2688093.
- [KPTY21] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Towards tight bounds for spectral sparsification of hypergraphs. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 598–611. ACM, 2021. doi:10.1145/3406325.3451061.
- [KPZ19] Nikolai Karpov, Marcin Pilipczuk, and Anna Zych-Pawlewicz. An exponential lower bound for cut sparsifiers in planar graphs. *Algorithmica*, 81(10):4029–4042, 2019. doi:10.1007/s00453-018-0504-8.
- [KR13] Robert Krauthgamer and Inbal Rika. Mimicking networks and succinct representations of terminal cuts. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '13*, page 1789–1799. SIAM, 2013.

- [KZ23] Jannik Kudla and Stanislav Zivný. Sparsification of monotone k -submodular functions of low curvature. *CoRR*, abs/2302.03143, 2023. [arXiv:2302.03143](https://arxiv.org/abs/2302.03143).
- [LB11] Hui Lin and Jeff A. Bilmes. A class of submodular functions for document summarization. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference 2011*, pages 510–520. The Association for Computer Linguistics, 2011.
- [LM17] Pan Li and Olgica Milenkovic. Inhomogeneous hypergraph clustering with applications. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pages 2308–2318, 2017.
- [LM18] Pan Li and Olgica Milenkovic. Submodular hypergraphs: p-laplacians, cheeger inequalities and spectral clustering. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 3020–3029. PMLR, 2018.
- [LVS+21] Meng Liu, Nate Veldt, Haoyu Song, Pan Li, and David F. Gleich. Strongly local hypergraph diffusions for clustering and semi-supervised learning. In *WWW '21: The Web Conference 2021*, pages 2092–2103. ACM / IW3C2, 2021. [doi:10.1145/3442381.3449887](https://doi.org/10.1145/3442381.3449887).
- [McC05] S Thomas McCormick. Submodular function minimization. *Handbooks in operations research and management science*, 12:321–391, 2005.
- [OST23] Kazusato Oko, Shinsaku Sakaue, and Shin-ichi Tanigawa. Nearly tight spectral sparsification of directed hypergraphs. In *50th International Colloquium on Automata, Languages, and Programming, ICALP 2023*, volume 261 of *LIPICs*, pages 94:1–94:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. [doi:10.4230/LIPICs.ICALP.2023.94](https://doi.org/10.4230/LIPICs.ICALP.2023.94).
- [Pog17] Yosef Pogrow. Solving symmetric diagonally dominant linear systems in sublinear time (and some observations on graph sparsification). Master’s thesis, Weizmann Institute of Science, 2017. URL: https://www.wisdom.weizmann.ac.il/~robi/files/YosefPogrow-MScThesis-2017_12.pdf.
- [Qua22] Kent Quanrud. Quotient sparsification for submodular functions. Manuscript, 2022. URL: <https://kentq.s3.amazonaws.com/submodular-sparsification.pdf>.
- [RY22] Akbar Rafiey and Yuichi Yoshida. Sparsification of decomposable submodular functions. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI*, pages 10336–10344. AAAI Press, 2022.
- [ST11] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40(4):981–1025, 2011. [doi:10.1137/08074489X](https://doi.org/10.1137/08074489X).
- [SY19] Tasuku Soma and Yuichi Yoshida. Spectral sparsification of hypergraphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019*, pages 2570–2581. SIAM, 2019. [doi:10.1137/1.9781611975482.159](https://doi.org/10.1137/1.9781611975482.159).
- [TIWB14] Sebastian Tschiatschek, Rishabh K. Iyer, Haochen Wei, and Jeff A. Bilmes. Learning mixtures of submodular functions for image collection summarization. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, pages 1413–1421, 2014.
- [VBK20] Nate Veldt, Austin R. Benson, and Jon M. Kleinberg. Minimizing localized ratio cut objectives in hypergraphs. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1708–1718. ACM, 2020. [doi:10.1145/3394486.3403222](https://doi.org/10.1145/3394486.3403222).
- [VBK21] Nate Veldt, Austin R. Benson, and Jon M. Kleinberg. Approximate decomposable submodular function minimization for cardinality-based components. In *Advances in Neural Information Processing Systems 34, NeurIPS2021*, pages 3744–3756, 2021.

- [VBK22] Nate Veldt, Austin R. Benson, and Jon M. Kleinberg. Hypergraph cuts with general splitting functions. *SIAM Rev.*, 64(3):650–685, 2022. doi:10.1137/20m1321048.
- [Von10] Jan Vondrák. Submodularity and curvature: The optimal algorithm (combinatorial optimization and discrete algorithms). *RIMS Kokyuroku Bessatsu*, 23:253–266, 2010. URL: <http://hdl.handle.net/2433/177046>.
- [Yam16] Yutaro Yamaguchi. Realizing symmetric set functions as hypergraph cut capacity. *Discret. Math.*, 339(8):2007–2017, 2016. doi:10.1016/j.disc.2016.02.010.
- [ZLS22] Yu Zhu, Boning Li, and Santiago Segarra. Hypergraph 1-spectral clustering with general submodular weights. In *56th Asilomar Conference on Signals, Systems, and Computers, ACSSC 2022*, pages 935–939. IEEE, 2022. doi:10.1109/IEEECONF56349.2022.10052065.

A Chernoff Bounds

We use the following versions of the Chernoff bound throughout the paper.

Lemma A.1 (Chernoff bound for bernoulli random variables). *Let X_1, \dots, X_m be independent random variables taking values in $\{0, 1\}$. Let X denote their sum and $\mu = \mathbb{E}[X]$. Then,*

$$\forall \delta \geq 0, \quad \Pr[|X - \mu| \geq \delta\mu] \leq 2 \cdot \exp\left(-\frac{\delta^2\mu}{2 + \delta}\right).$$

Lemma A.2 (Chernoff bound for bounded random variables, Theorem 6.1 in [AGK14]). *Let $X_1, \dots, X_m \geq 0$ be independent random variables such that either X_i is deterministic or $X_i \in [0, b]$. Let X denote their sum and $\mu = \mathbb{E}[X]$, then,*

$$\forall \delta > 0, \quad \Pr[|X - \mu| \geq \delta\mu] \leq 2 \cdot \exp\left(-\frac{\delta^2\mu}{(2 + \delta)b}\right).$$

Additionally,

$$\forall \delta \in [0, 1], \quad \Pr[|X - \mu| \geq \delta\mu] \leq 2 \cdot \exp\left(-\frac{\delta^2\mu}{3b}\right).$$

B Approximating Coverage Functions

Due to the wide prevalence of submodular hypergraph cut functions, our results have several applications. One example is finding a succinct representation for coverage functions.

Definition B.1. *A function $f : 2^V \rightarrow \mathbb{R}_+$ is called a coverage function over ground set W with weight function $\rho : W \rightarrow \mathbb{R}_+$ if there exists a collection $\{A_v\}_{v \in V}$ where each $A_v \subseteq W$, such that*

$$\forall S \subseteq V, \quad f(S) = \sum_{w \in \bigcup_{v \in S} A_v} \rho(w).$$

Coverage functions are commonly used as objective functions, for example in sensor-placement problems [KG11]. The running time of algorithms for these problems may be improved considerably by decreasing the size of the ground set $|W|$. To this end, the next corollary shows that every coverage function can be approximated by a coverage function with ground-set size $\tilde{O}_\epsilon(n)$. In general, the given ground set might have size 2^n , hence the decrease in size may be exponential.

Corollary B.2. *Let e be a hyperedge whose splitting function g_e is a coverage function, and let $K := \max_{w \in W} |\{v \in V : w \in A_v\}|$. The e can be $(1 + \epsilon)$ -approximated by $O(\epsilon^{-2}n \log n)$ hyperedges with support size at most K . Furthermore, the resulting sparsifier is a coverage function on $O(\epsilon^{-2}n \log n)$ elements.*

This result was recently obtained independently in [Qua22], using two different proof methods. One of them is by reduction to (sparsification of) undirected hypergraph cuts. Our proof is simpler, and designs a reduction to (sparsification of) additive splitting, for which we can apply Theorem 1.7.

Proof. For $w \in W$, let $V_w := \{v \in V : w \in A_v\}$. Observe that $f(S)$ can be written as

$$f(S) = \sum_{w \in \bigcup_{v \in S} A_v} \rho(w) = \sum_{w \in W} \rho(w) \cdot \mathbb{1}_{|V_w \cap S| > 0} = \sum_{w \in W} \rho(w) \cdot \min(|V_w \cap S|, 1).$$

Hence, f can be written as a sum of $|W|$ splitting functions of the form $g_w : S \mapsto \rho(w) \cdot \min(|V_w \cap S|, 1)$. Observe that the spread of each g_w is $\mu_{g_w} = 1$, therefore Theorem 1.7 yields the desired result. Finally, note that the resulting sparsifier is a reweighted subgraph, hence it is a coverage function with $O(\epsilon^{-2}n \log n)$ elements. \square

C Application to Terminal Cut Functions

Terminal Cut Functions. Let $G = (V, E, w)$ be some undirected graph and let $\mathcal{T} \subseteq V$ be a special set of vertices called the terminals of G . Denote $|\mathcal{T}| = k$. The terminal cut function of G is defined as

$$\forall S \subseteq \mathcal{T}, \quad \text{cut}_G(S, \bar{S}) = \min_{U \subseteq V : U \cap \mathcal{T} = S} \sum_{e \in \delta(U)} w(e),$$

where $\delta(U)$ is the set of edges with exactly one endpoint in U . Note that the terminal cut function $\text{cut}_G(S) : 2^{\mathcal{T}} \rightarrow \mathbb{R}_+$, of G is submodular. In [HKNR98], the authors show that it is possible to construct a graph with $O(2^{2^k})$ vertices that preserves the terminal cut function of G exactly. On the other hand, [KR13, KPZ19] showed that a minimum of $2^{\Omega(k)}$ vertices is necessary to preserve the terminal cut function of G , even for planar graphs.

In the approximate case, a construction of quality $1 + \epsilon$ with $\tilde{O}(\text{poly}(k, \epsilon^{-1}))$ vertices was shown for bipartite graphs by [AGK14, ADK⁺16]. However, there is no known upper or lower bound for the size of a data structure approximating the cuts of general graphs. One such possible data structure would be to represent the terminal cut function as cuts of a submodular hypergraph with simple splitting functions. Then, using Theorem 1.4 we can achieve a small data structure to represent the graph cuts. However, it turns out that this is not possible using the all-or-nothing splitting function.

Theorem C.1 (Theorem 3.3 in [Yam16]). *If a symmetric submodular function $f : 2^V \rightarrow \mathbb{R}_+$ can be realized as a cut capacity function of an undirected hypergraph with nonnegative capacities, then*

$$\forall i \in [V], \forall S \in \binom{V}{i}, \quad (-1)^i f^{(i)}(S) \leq 0,$$

where

$$f^{(i)}(S) = \sum_{X \subseteq S} (-1)^{|S \setminus X|} f(X).$$

Consider the following counter-example, let $G = (V, E)$ be the star graph with 4 leaves and a central vertex, where the terminals $\mathcal{T} = \{t_1, t_2, t_3, t_4\}$ are the leaves. Denoting its terminal cut function as f_G , observe that $f_G^{(3)}(\{t_1, t_2, t_3\}) = -2 < 0$. Hence, by Theorem C.1, f cannot be realized as a cut capacity function of an undirected hypergraph. However, in this case f_G is exactly the small-side splitting function. Therefore, we pose the following question - is there a class of simple splitting functions that can represent the terminal cut function of any graph?

D Proof of Expectation for Symmetric Additive Decomposition

Proof of Claim 4.4. Let $S_i = S \cap e_i, \bar{S}_i = \bar{S} \cap e_i$ be the intersection between S and the sampled hyperedge. Assume without loss of generality that $|S| < |\bar{S}|$. Observe that the function $\min(x, y, K)$ is concave, and hence by Jensen's inequality

$$\mathbb{E} [\min(|S_i|/p, |\bar{S}_i|/p, K)] \leq \min(\mathbb{E}[|S_i|/p], \mathbb{E}[|\bar{S}_i|/p], K) = \min(|S|, |\bar{S}|K) = g_e(S),$$

where the second inequality is since S_i, \bar{S}_i are independent. Hence, it only remains to prove that $\mathbb{E}[Ng_{e_i}(S)] \geq (1 - 2\epsilon')g_e(S)$. We split the analysis into two cases, when $|S| < K/2$ and its complement. Starting with the case when $|S|$ is small, observe that setting $\delta = K/|S| > 2$ we have by Claim 4.2 that

$$\Pr[|S_i| \geq pK] \leq 2|e|^{-\frac{c\delta^2|S|}{(2+\delta)K\epsilon'^2}} \leq 2|e|^{-\frac{c\delta|S|}{2K}} = 2|e|^{-\frac{c}{2}},$$

where the second inequality is by $\delta/(2+\delta) > 1/2$ when $\delta > 2$. Choosing $c > 10$ we find that the probability is at most $|e|^{-5}$. Note that $K < |e|/2$, otherwise it doesn't affect the splitting function. Hence, $|\bar{S}| \geq 3K/2$ therefore using Claim 4.2 again we have (setting $\delta = 1/2$),

$$\Pr[|\bar{S}_i| < pK] \leq 2|e|^{-\frac{c\delta^2|\bar{S}|}{3K\epsilon'^2}} \leq 2|e|^{-\frac{3\epsilon K/2}{12K\epsilon'^2}} \leq 2|e|^{-4c},$$

where the last inequality is by $\epsilon' = \epsilon/4$ and $\epsilon < 1$. Choosing $c > 2$ we find that the probability that both events not occur is at most $4|e|^{-5}$. Note that the expectation of $Ng_{e_i}(S)$ is given by

$$\mathbb{E}[Ng_{e_i}(S)] = \sum_{j=0}^{\lfloor pK \rfloor} \frac{j}{p} (\Pr[|S_i| = j, |\bar{S}_i| \geq j] + \Pr[|S_i| > j, |\bar{S}_i| = j]) + K \Pr[|S_i| > pK, |\bar{S}| > pK].$$

Hence, we can bound the expectation from below by keeping only the most significant terms

$$\begin{aligned} \mathbb{E}[Ng_{e_i}(S)] &\geq \sum_{j=0}^{\lfloor pK \rfloor} \frac{j}{p} \Pr[|S_i| = j, |\bar{S}_i| \geq j] \geq \sum_{j=0}^{\lfloor pK \rfloor} \frac{j}{p} \Pr[|S_i| = j] \Pr[|\bar{S}_i| \geq j] \\ &\geq \Pr[|\bar{S}_i| \geq pK] \sum_{j=0}^{\lfloor pK \rfloor} \frac{j}{p} \Pr[|S_i| = j] \geq (1 - 4|e|^{-5}) \sum_{j=0}^{\lfloor pK \rfloor} \frac{j}{p} \Pr[|S_i| = j], \end{aligned}$$

where the second inequality is by the independence of S_i, \bar{S}_i and the last by substituting the bound $\Pr[|\bar{S}_i| < pK]$. Now adding and subtracting the rest of the possible values of $|S_i|$ we get

$$\begin{aligned} \mathbb{E}[Ng_{e_i}(S)] &\geq (1 - 4|e|^{-5}) \left(\sum_{j=0}^{|S|} \frac{j}{p} \Pr[|S_i| = j] - \sum_{j=\lfloor pK \rfloor+1}^{|S|} \frac{j}{p} \Pr[|S_i| = j] \right) \\ &\geq (1 - 4|e|^{-5}) \left(\frac{\mathbb{E}[|S_i|]}{p} - \frac{|S|^2}{p} \Pr[|S_i| > pK] \right), \end{aligned}$$

where the last inequality is by $j, |S| - \lfloor pK \rfloor < |S|$. Note that $\mathbb{E}[|S_i|]/p = |S| = g_e(S)$ since $|S| < K < |\bar{S}|$. Recall that $p = ce'^2 K^{-1} \log |e|$ and observe,

$$\frac{|S|^2}{p} \Pr[|S_i| > pK] \leq \frac{|e|^2 K}{c|e|^5 \log |e|} \leq \frac{1}{10|e|^2},$$

where the first inequality is by substituting the bound for $\Pr[|S_i| > pK]$ and the second by $K < |e|$, $\log |e| > 1$ and $c > 10$. Hence, we obtain $\mathbb{E}[Ng_{e_i}(S)] \geq (1 - 4|e|^{-5})(g_e(S) - |e|^{-2}) \geq (1 - 2\epsilon')g_e(S)$ with the last inequality by $\epsilon^{-2} < |e|$.

Now we turn to the case $|S| > K/2$. Observe that for $Ng_{e_i}(S) = \min(|S_i|/p, |\bar{S}_i|/p, K) < (1 - \epsilon')g_e(S)$ we must have $\min(|S_i|, |\bar{S}_i|) < (1 - \epsilon')p|S|$. By Claim 4.2 this event happens with probability at most $4|e|^{-\frac{c\epsilon'^2|S|}{3K\epsilon'^2}} \leq 4|e|^{-\frac{c}{6}}$. Therefore,

$$\mathbb{E}[Ng_{e_i}(S)] \geq \left(1 - 4|e|^{-\frac{c}{6}}\right) (1 - \epsilon')p|S|/p \geq (1 - 2\epsilon')|S| = (1 - 2\epsilon')g_e(S),$$

where the last inequality is by setting $c > 15$ and $\epsilon^2 > 1/|e|$. □