

Optimal Sparsifiers for Minkowski Sums and Sums of Seminorms

Arpon Basu¹ Joshua Brakensiek² Yeyuan Chen³ Aaron Putterman⁴
 Victor Reis⁵ Zihan Zhang⁶

August 4, 2026

Abstract

We extend the recent work of Reis and Rothvoss on sparsifying sums of ℓ_1 norms to the more general task of sparsifying (Minkowski) sums of centrally symmetric, convex sets. As our main result, we prove that for any $\varepsilon > 0$ and centrally symmetric, convex sets $C_1, \dots, C_m \subseteq \mathbb{R}^n$ there is a choice of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that at most $O(n/\varepsilon^2)$ of the weights are non-zero, and

$$(1 - \varepsilon) \cdot C \subseteq \sum_{i=1}^m \lambda_i \cdot C_i \subseteq (1 + \varepsilon) \cdot C,$$

where $C := C_1 + \dots + C_m$ refers to the Minkowski sums of the sets $C_1, \dots, C_m$, and $\lambda \cdot C$ refers to the dilation of the set C .

As immediate applications of this result, we obtain sparsifiers of size $O(n/\varepsilon^2)$ for sparsifying sums of seminorms in n -dimensional space, improving on the $O\left(\frac{n \log(n/\varepsilon) \cdot \log^{2.5}(n)}{\varepsilon^2}\right)$ size sparsifiers from the work of Jambulapati, Lee, Liu, and Sidford (FOCS 2023). This further yields optimal size hypergraph cut sparsifiers with $O(n/\varepsilon^2)$ hyperedges, improving on the $O(n \log(n)/\varepsilon^2)$ size sparsifiers from the work of Chen, Khanna, and Nagda (FOCS 2020). More generally, this also gives optimal size sparsifiers for sums of symmetric submodular functions.

¹Princeton University. arpon.basu@princeton.edu

²University of California, Berkeley. Supported in part by a Simons Investigator award of Venkatesan Guruswami, and NSF awards CCF-2211972 and DMS-2503280 josh.brakensiek@berkeley.edu

³Department of EECS, University of Michigan, Ann Arbor. Supported in part by National Science Foundation grant No. CCF-2236931. yeyuanch@umich.edu

⁴School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported in part by Simons Investigator Awards to Madhu Sudan and Salil Vadhan, a Jane Street Graduate Research Fellowship, and AFOSR award FA9550-25-1-0112. aputterman@g.harvard.edu

⁵Microsoft Research, Redmond. victorol@microsoft.com

⁶Department of Computer Science and Engineering, The Ohio State University. zhang.13691@osu.edu

Contents

1	Introduction	1
1.1	Our Results	2
1.2	Technical Overview	5
1.3	Organization	8
2	Convex Geometry Preliminaries	8
2.1	Minkowski Sums and Differences	9
2.2	Volumes and Mixed Volumes	10
2.3	Quermassintegrals and Differentiability	10
2.4	Support Functions	12
3	A One-Sided Volume Lemma	12
3.1	The Framework of [RR26]	12
3.2	The Proof of Lemma 3.1	13
4	Building Support Sparsifiers for Sums of Convex Bodies	15
4.1	[RR26] Preliminaries	15
4.2	The Partial Coloring	16
4.3	Building the Sparsifier	17
5	Deriving Seminorm Sparsifiers	20
6	Applications	21
6.1	Hypergraph Cut Sparsifiers	21
6.2	Symmetric Submodular Sparsification	23
	References	24

1 Introduction

Sparsification refers to the general procedure by which a large, complex object is replaced by an ideally much smaller object which nevertheless preserves certain properties of interest. Such sparsification was first introduced (independently) in the work of Benczúr and Karger [BK96] and Talagrand [Tal90]. [BK96] studied the so-called cut sparsification of *graphs* $G = (V, E, w)$ where the goal is to select a subset of the *edges* of the graph, while still preserving all of the cut-sizes. [Tal90] instead studied a continuous generalization of this notion called ℓ_1 -norm sparsification, where one is provided with a sum of ℓ_1 norms of linear functions, and wishes to choose a subset which preserves the sum on every input.

Since these seminal works, sparsification has evolved into a rich field. The works of Spielman and Teng [ST11], Spielman and Srivastava [SS11], and Batson, Spielman, and Srivastava [BSS14] generalized cut sparsification of graphs to a stronger notion called *spectral sparsification*. These notions of graph sparsification have proved to be useful in a plethora of algorithmic applications, ranging from graphs [BK96, ST04, SS11], to linear system solving [ST04], and numerous problems in the graph streaming literature [AGM12a, AGM12b, McG14, ADK⁺16, KLM⁺17, KMM⁺19].

Other works have sought to understand the theoretical limits of *which* objects can be sparsified. This exploration began with the study of *hypergraph* cut sparsification and CSP sparsification in the work of [KK15], which led to works improving sparsifier sizes [CKN20], a study of spectral hypergraph sparsification [SY19, KKTY21a, KKTY21b, Lee23, JLS23], sparsification of codes [KPS24a, KPS25, BG25, BGP26, KPS26, LMW26], as well as other generalizations like sparsifying sums of norms [JLLS23], and sparsifying sums of PSD matrices [HLM⁺26, BKLM26, BKMT26].

Despite this vast attention devoted to sparsification, there is one common thread which plagues *almost* all of these works: random sampling. Indeed, in all instances of sparsification via random sampling, there is an inherent blow-up (in the size of the sparsifier) which scales as $\log(\text{dim})$. For instance, in the cut-sparsification of graphs of [BK96], this led to sparsifiers preserving $O_\varepsilon(|V| \log(|V|))$ many edges, and in the work of [Tal90], this led to ℓ_1 -norm sparsifiers preserving $O_\varepsilon(n \log n)$ many terms. Removing these extra logarithmic factors remained an open question for the ensuing years. Only in the work of Batson, Spielman, and Srivastava [BSS14] was this first resolved for cut (and in fact, the stronger notion of spectral) sparsification, whereby they produced sparsifiers which retained only $O_\varepsilon(n)$ many edges. For ℓ_1 norms, this remained open much longer, until the very recent work of Reis and Rothvoss [RR26], which produced ℓ_1 -norm sparsifiers with only $O_\varepsilon(n)$ many terms. Along the way, the only other distinct instances of sparsification that avoid this inherent logarithmic factor are due to De Carli Silva, Harvey and Sato [SHS15] on sparsifying sums of PSD matrices (and hypergraph cut sparsification for arity $r = 3$), and the (even more) recent work of Basu, Kothari, Meka and Tudose [BKMT26] which extends [RR26] to sparsifying Abelian Cayley graphs.

In this work, we re-visit the framework of Reis and Rothvoss [RR26] and ask just how far it can be pushed in constructing optimal size sparsifiers.

1.1 Our Results

As our main result, we show that the framework of [RR26] can be extended to imply optimal sparsification for sums of *support functions of centrally symmetric, compact, convex sets* a.k.a., symmetric convex bodies¹. More formally, one can consider the space $E = \mathbb{R}^n$. A set $C_i \subseteq E$ is said to be:

1. *Centrally symmetric* if whenever $x \in C_i$, then $-x \in C_i$.
2. *Compact* if C_i is closed and bounded.
3. *Convex* if for any $x, y \in C_i$ and $\lambda \in [0, 1]$, $\lambda \cdot x + (1 - \lambda) \cdot y \in C_i$.

For such a set C_i , the support function is defined as $h_{C_i}(u) = \max_{x \in C_i} \langle x, u \rangle$ for $u \in E$. As our main theorem, we show the following:

Theorem 1.1 (Sparsifying Sums of Support Functions). *There is a universal constant $C_\star > 0$ with the following property. Let E be a finite-dimensional real vector space, and let $C_1, \dots, C_m \subseteq E$ be compact centrally symmetric convex sets. Let*

$$C = C_1 + \dots + C_m, \quad d = \dim \text{span}(C),$$

be the Minkowski sum of the convex bodies. For every $0 < \varepsilon < 1$, there exist coefficients $\lambda_1, \dots, \lambda_m \geq 0$, supported on at most $C_\star \cdot d/\varepsilon^2$ indices, such that for every $u \in E$,

$$(1 - \varepsilon) \cdot \sum_{i=1}^m h_{C_i}(u) \leq \sum_{i=1}^m \lambda_i \cdot h_{C_i}(u) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m h_{C_i}(u).$$

Equivalently, we have the following containment of convex bodies:

$$(1 - \varepsilon) \cdot C \subseteq \sum_{i=1}^m \lambda_i \cdot C_i \subseteq (1 + \varepsilon) \cdot C.$$

Importantly, the size of the sparsifier scales as $O(d/\varepsilon^2)$ *without* any logarithmic factors depending on d or n . This is the key improvement with respect to prior work. To appreciate the expressivity of Theorem 1.1, we consider a few examples. First, as mentioned above, our work hinges on a generalization of the framework of [RR26]; it is therefore natural to ensure that Theorem 4.4 generalizes their main theorem. In their work they consider functions of the form $f_i : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ where $f_i(u) = |\langle a_i, u \rangle|$ for some fixed $a_i \in \mathbb{R}^n$. Given such a collection of functions $f_1, \dots, f_m$ and a parameter $\varepsilon > 0$, the goal is then to find weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that for every $u \in \mathbb{R}^n$,

$$(1 - \varepsilon) \cdot \sum_{i=1}^m f_i(u) \leq \sum_{i=1}^m \lambda_i \cdot f_i(u) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m f_i(u).$$

Reis and Rothvoss [RR26] show that such sparsifiers can be built while having only $O(n/\varepsilon^2)$ many non-zero weights, which improved upon the previous best of $O(n \log(n)/\varepsilon^2)$ from the work of

¹ Some definitions of convex body require the convex body to have an interior point. We follow the presentation of [Gru07, Sch14] where convex bodies need to only be non-empty with a *proper* convex body having non-empty interior.

[Tal90]. We can now observe that this fits *exactly* into our framework above: indeed, for each function f_i , one simply considers the convex set which is $C_i = \{\alpha \cdot a_i : \alpha \in [-1, 1]\}$. In this way, the support function becomes $h_{C_i}(u) = \max_{x \in C_i} \langle x, u \rangle = \max(\langle a_i, u \rangle, \langle -a_i, u \rangle) = |\langle a_i, u \rangle|$.

Thus, the work of [RR26] corresponds exactly with the setting in which each C_i above is a *one-dimensional* convex set. The key expressivity of Theorem 1.1 instead comes from the fact that it works for *arbitrary dimensional* convex sets (again, provided they are compact and centrally symmetric).

Sparsifying Sums of Seminorms To demonstrate this expressivity, we consider the sparsification of sums of seminorms. In this setting, first introduced in the work of Jambulapati, Lee, Liu, and Sidford [JLLS23], one is provided with a collection of seminorms $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$. A function f is a seminorm if it satisfies:

(1) **Positive Homogeneity.** For all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, $f(\lambda x) = |\lambda| \cdot f(x)$.

(2) **Subadditivity.** For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$.

Crucially, *all* seminorms can be written as support functions of centrally symmetric, compact, convex sets. By using this identity in combination with Theorem 1.1, we obtain the following theorem on sparsifying sums of seminorms:

Theorem 1.2. Let $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ be a collection of seminorms. Then, for any $\varepsilon > 0$, there exists a set of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ with only $O(n/\varepsilon^2)$ of the weights non-zero such that for every $x \in \mathbb{R}^n$:

$$(1 - \varepsilon) \cdot \sum_{i=1}^m f_i(x) \leq \sum_{i=1}^m \lambda_i \cdot f_i(x) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m f_i(x).$$

The previous best known sparsifier size for sums of seminorms was due to the aforementioned work of [JLLS23], which retained $O\left(\frac{n \log(n/\varepsilon) \cdot \log^{2.5}(n)}{\varepsilon^2}\right)$ many terms in the sparsifier. Our work shaves off all of these logarithmic factors thereby *settling* the optimal size complexity of seminorm sparsification. Indeed, because graph cut-sparsification can be captured by sparsifying sums of seminorms, the work of Carlson, Kolla, Srivastava, and Trevisan [CKST19] implies that there are sums of seminorms for which $(1 \pm \varepsilon)$ sparsification requires the retention of $\Omega(n/\varepsilon^2)$ many terms.

Sparsifying Hypergraphs Seminorms themselves already generalize many frequently studied sparsification problems. As our first application, we are able to conclude optimal size *hypergraph cut* sparsifiers. In this setting, one is provided with a hypergraph $H = (V, E, w)$, where $w : E \rightarrow \mathbb{R}_{\geq 0}$ is a collection of non-negative hyperedge weights. The goal is to compute a re-weighted sub-hypergraph $H' = (V, E, w')$ with $\text{Supp}(w')$ much smaller than $|E|$ which nevertheless preserves all *cut-sizes* in the hypergraph. In this context, for a set $S \subseteq V$, the cut-size of S is defined as

$$\delta_H(S) := \sum_{e \in E} w(e) \cdot \mathbf{1}[S \cap e \neq \emptyset \text{ and } (V - S) \cap e \neq \emptyset],$$

intuitively capturing the total weight of hyperedges which are crossing from S to $V - S$. A hypergraph H' defined on the same vertex set is then a $(1 \pm \varepsilon)$ cut sparsifier of H if for every $S \subseteq V$, $\delta_{H'}(S) \in (1 \pm \varepsilon) \cdot \delta_H(S)$.

This cut-sparsification was first proposed in the work of Kogan and Krauthgamer [KK15], where they showed there exist $(1 \pm \varepsilon)$ sparsifiers which retain only $O(n \cdot (r + \log(n))/\varepsilon^2)$ many hyperedges, where $n = |V|$ and r is the maximum size of any hyperedge. This was later improved in the work of Chen, Khanna, and Nagda [CKN20] to size $O(n \log(n)/\varepsilon^2)$, which has remained the best ever since, despite extensive follow-up attention [SY19, KKTY21a, KKTY21b, Lee23, JLS23, JLLS23, KPS24a, Qua24].

Nevertheless, because hypergraph cut functions have a natural extension as seminorms, we immediately obtain the following theorem:

Theorem 1.3. *Let $H = (V, E, w)$ be any hypergraph (with non-negative weights). Then, for any $\varepsilon > 0$, there exists a hypergraph $H' = (V, E, w')$ (again, with non-negative weights) such that:*

- (1) H' is a $(1 \pm \varepsilon)$ hypergraph cut-sparsifier of H .
- (2) H' has at most $O(|V|/\varepsilon^2)$ hyperedges with non-zero weight: i.e., $|\{e \in E : w'(e) > 0\}| \leq O(|V|/\varepsilon^2)$.

This removes the final logarithmic factor from the work of [CKN20], and establishes $O(|V|/\varepsilon^2)$ as the *optimal* size for any hypergraph cut sparsifiers (again, with the matching lower bound owed to [CKST19]).

Sparsifying Sums of Submodular Functions Another application of our seminorm sparsification is to sparsifying sums of submodular functions. Recall that given an integer n , a function $f : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ is said to be submodular if for every $S, T \subseteq [n]$

$$f(S) + f(T) \geq f(S \cap T) + f(S \cup T).$$

A well-studied question in sparsification (see, e.g., [RY22, KŽ23, KK23, JLLS23, KPS24b, BCC26]) asks whether, given a collection of submodular functions $f_1, \dots, f_m : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$, one can compute a smaller $(1 \pm \varepsilon)$ sparsifier. In this context, mirroring above, a sparsifier is a collection of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that for every $S \subseteq [n]$ it is the case that

$$(1 - \varepsilon) \cdot \sum_{i=1}^m f_i(S) \leq \sum_{i=1}^m \lambda_i \cdot f_i(S) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m f_i(S).$$

As observed in [JLLS23], whenever the submodular functions f_i are additionally *symmetric* (meaning that $f_i(S) = f_i([n] - S)$) a crucial result of Lovász [Lov83] implies that such submodular functions can be written as continuous seminorms. In combination with Theorem 1.2, we then obtain the following theorem:

Theorem 1.4. *Let n be an integer, and let $f_1, \dots, f_m : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ be symmetric submodular functions. For any parameter $\varepsilon > 0$, there exists a collection of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that:*

- (1) $\lambda_1, \dots, \lambda_m$ constitutes a $(1 \pm \varepsilon)$ sparsifier of $f_1, \dots, f_m$.

(2) At most $O(n/\varepsilon^2)$ of the weights $\lambda_1, \dots, \lambda_m$ are non-zero.

This improves upon the previous best sparsifier size from [JLLS23] of $O\left(\frac{n \log(n/\varepsilon) \cdot \log^{2.5}(n)}{\varepsilon^2}\right)$ many terms. This also immediately yields sparsifiers of the same $O(n/\varepsilon^2)$ many terms for sums of *monotone*, submodular functions (where $f(S \cup \{v\}) \geq f(S)$ for every $S \subseteq [n], v \in [n]$) via a reduction of [KPS24b].

1.2 Technical Overview

The Framework of [RR26] In this section, we now present a brief technical overview for [Theorem 1.1](#). As mentioned above, the proof of [Theorem 1.1](#) follows in the footsteps of [RR26]. The main idea of [RR26] is to apply an iterative “weight perturbation” procedure which slowly reduces the support size of an object, while simultaneously ensuring that the “error accumulation” is not too large.

In more detail, let $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$ denote the collection of functions we wish to sparsify. As mentioned above, in our setting these will be the support functions of compact, centrally symmetric convex sets $C_1, \dots, C_m$. To start, these functions f_i are each given an implicit weight of 1, so we let $w_i^{(1)} = 1$ for $i \in [m]$. The goal of [RR26] is simple: they wish to find a “weight perturbation” vector $z^{(1)} \in [-1, 1]^m$ such that the new weight vector $w^{(2)} = w^{(1)} + z^{(1)}$ satisfies:

(1) $w^{(2)}$ is still a good sparsifier of $w^{(1)}$. I.e., for every $x \in \mathbb{R}^n$,

$$\sum_{i=1}^m w_i^{(2)} \cdot f_i(x) \approx \sum_{i=1}^m w_i^{(1)} \cdot f_i(x).$$

(2) $w^{(2)}$ has a smaller support than $w^{(1)}$. i.e., some entries of $w^{(2)}$ are now equal to 0.

Importantly, this second condition above exactly corresponds with finding a weight perturbation vector z which has many coordinates *equal* to -1 .

The key contribution from the work of [RR26] is a general framework for establishing sufficient conditions for when such vectors z exist. To do this, they define an appropriate “one-sided sparsification polytope”:

$$K_0 = \left\{ z \in [-1, 1]^m : \sum_{i=1}^m z_i \cdot f_i(x) \leq 0 \text{ for every } x \in \mathbb{R}^n \right\}.$$

This is simply the set of all weight perturbation vectors which constitute good *one-sided* sparsifiers, i.e., which yield $w^{(2)} = z + w^{(1)}$ such that for all $x \in \mathbb{R}^n$,

$$\sum_{i=1}^m w_i^{(2)} f_i(x) \leq \sum_{i=1}^m w_i^{(1)} f_i(x).$$

The key fact from [RR26] is that if K_0 has large volume, then *there must* exist weight perturbation vectors which satisfy both of the desired conditions above.

Claim 1.5 (Informal). *Let K_0 be defined as above. If*

$$\frac{\text{Vol}_m(K_0)}{2^m} \geq 2^{-O(d)},$$

then for any $\varepsilon > 0$, there exists $(1 \pm \varepsilon)$ sparsifiers of $f_1, \dots, f_m$ with total support $O(d/\varepsilon^2)$.

This clarifies the target statement that we now need in our setting: we define $f_1, \dots, f_m$ to be the support functions of centrally symmetric, compact, convex sets, and define K_0 to be one-sided sparsification polytope. Can we then lower bound the volume of K_0 by $2^m \cdot 2^{-O(n)}$?

Lower Bounding the Volume Fortunately, [RR26] also presents tools for lower bounding the volume. Our starting point is their Lemma 18:

Fact 1.6 (Lemma 18 of [RR26]). *Let $F : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ be separately convex and not identically zero. Suppose that, for some integer $d \geq 0$,*

$$F(\rho x) = \rho^d F(x) \quad \text{for every } x \in \{-1, 1\}^m \text{ and } \rho \geq 0.$$

Then

$$\Pr_{\sigma \in \{-1, 1\}^m} [F(\sigma) > 0] \geq e^{-2d}.$$

Of key importance is the notion of *separate convexity* of the function F : for a convex set $\Omega \subseteq \mathbb{R}^n$, this essentially requires that for any $i \in [m]$ and any $\omega \in \Omega$, the restricted function $t \mapsto F(\omega + te_i) \in \mathbb{R}$ is convex, where $e_i \in \mathbb{R}^m$ is the i^{th} standard basis vector.

Nevertheless, this still raises new challenges; we must find a function F which is separately convex (and satisfies the desired scaling property with ρ) such that being able to conclude $\Pr_{\sigma \in \{-1, 1\}^m} [F(\sigma) > 0] \geq e^{-2n}$ implies our desired volume lower bound on K_0 . This is where we make our first key contribution; given our collection of compact, convex, centrally symmetric sets $C_1, \dots, C_m$, we assume that the total dimension spanned by the sum of $C_1, \dots, C_m$ is n and define the function $F(t_1, \dots, t_m) = \text{Vol}_n(\sum_{i=1}^m t_i C_i)$. In this notation, $t_i \cdot C_i$ refers to scaling the set C_i by t_i , and summing over bodies refers to the typical Minkowski sum of sets.

It is not hard to see that

$$F(\rho x) = \rho^n F(x) \quad \text{for every } x \in \{-1, 1\}^m \text{ and } \rho \geq 0;$$

this follows from the fact that if one scales all coordinates by a factor of ρ in a space of dimension n , then the resulting volume of the body shrinks by exactly a factor of ρ^n .

The more challenging task is to show that this function F is indeed separately convex. This relies crucially on a result in convex geometry related to Minkowski subtraction, which we first formally define. Given convex sets $A, B \subseteq \mathbb{R}^n$, we let their Minkowski difference be defined as

$$A \ominus B := \bigcap_{y \in B} (A - y) = \{y \in \mathbb{R}^n : B + y \subseteq A\}.$$

Note that $A \ominus B$ may potentially be empty, for instance in the case when $A \subsetneq B$. Nevertheless, for a proper convex body A (i.e., A has nonzero volume) and any (bounded) convex set B , there will always *exist* some constant $r(A, B) > 0$ such that $A \ominus r(A, B) \cdot B$ is non-empty.

More importantly, this now lets us build a parameterized function $g_{A,B}(r) = \text{Vol}(A \ominus r \cdot B)$, for which we show the following fundamental property.

Claim 1.7. *For any convex bodies A and B , the function $g_{A,B}$ is convex.*

The proof of [Claim 1.7](#) follows from a formula for the derivative of $g_{A,B}(r)$ in terms of *quermassintegrals* due to Bol [[Bol43](#)] and Matheron [[Mat78](#)]. See [Section 2](#) for details. By carefully re-writing the function F from above, we can directly show that the influence of each coordinate can be written in the form $g_{A,B}(r)$ (for carefully defined A, B), whereby we are able to conclude convexity in that coordinate, and thus all together, we obtain separate convexity. A formal proof that F is separately convex is given in [Section 3](#).

Finally, it remains to show why a bound of the form $\Pr_{\sigma \in \{-1,1\}^m} [F(\sigma) > 0] \geq e^{-2n}$ suffices for our target one-sided volume bound of [Claim 1.5](#). Indeed, let us fix a vector $\sigma \in \{-1,1\}^m$ such that $F(\sigma) > 0$. Letting $C_P = \sum_{i \in [m]: \sigma_i = 1} C_i$ and $C_N = \sum_{i \in [m]: \sigma_i = -1} C_i$, this implies that

$$F(\sigma) = \text{Vol}(C_P \ominus C_N) > 0,$$

or equivalently, that $C_N \subseteq C_P$. By using known characterizations of convex bodies in terms of support functions, this then implies that for every $u \in \mathbb{R}^n$,

$$h_{C_N}(u) \leq h_{C_P}(u),$$

or equivalently, that $h_{C_P}(u) - h_{C_N}(u) \geq 0$. By again using properties of support functions, this is in turn equivalent to the fact that

$$\sum_{i=1}^m \sigma_i \cdot h_{C_i}(u) \geq 0 \text{ for all } u \in \mathbb{R}^n.$$

Thus, we know that

$$\Pr_{\delta \sim \{\pm 1\}^m} \left[\sum_{i=1}^m \sigma_i \cdot h_{C_i}(u) \geq 0 \text{ for all } u \in \mathbb{R}^n \right] \geq e^{-2n},$$

and by simply replacing δ with $-\delta$, we then have that

$$\Pr_{\delta \sim \{\pm 1\}^m} \left[\sum_{i=1}^m \sigma_i \cdot h_{C_i}(u) \leq 0 \text{ for all } u \in \mathbb{R}^n \right] \geq e^{-2n}.$$

A simple transformation of the convex bodies can then be used to show that the same statement holds when sampling from $[-1, 1]^m$ rather than just $\{\pm 1\}^m$, thereby yielding our desired volume bound (and sparsifiers).

We remark that the recent work of [[BKMT26](#)] also proves a bound similar to [[RR26](#), Lemma 18] for the sparsification polytope arising in the context of sparsifying abelian Cayley graphs. Their paper

uses very different techniques from the ones outlined here though, relying on group symmetries to prove such a bound. However, once the volume lower bound is established, the remaining machinery of [RR26] can be invoked as it is to obtain a sparsifier, as is also done in this paper.

It would be interesting to find further such instances in sparsification where optimal sparsifiability follows volume lower bounds analogous to [RR26, Lemma 18].

1.3 Organization

In Section 2 we present most of the necessary convex geometric preliminaries. In Section 3, we present a formal proof of the necessary one-sided volume bound, which we then use in Section 4 to build sparsifiers for sums of support functions of convex bodies. In Section 5, we show this implies optimal size sparsification of sums of seminorms, and in Section 6, we further use these seminorm sparsifiers to build optimal sparsifiers of hypergraph cut functions and sums of symmetric submodular functions.

Concurrent Work An independent and concurrent work posted at <https://yangpliu.github.io/repo/sparsifying-sums-seminorms/paper.pdf> was achieved via ChatGPT 5.6-Sol prompted by Yang P. Liu and provides the slightly weaker bound for seminorm sparsifiers of $O(n\varepsilon^{-2} \log(1/\varepsilon))$.

Acknowledgments The main result of this paper was also discovered through a sequence of discussions with ChatGPT 5.5-Pro and 5.6-Sol. This initially began with an exploration of hypergraph cut sparsification, which the authors later realized could be extended to sparsification of seminorms, and ultimately its final form of “Minkowski sum sparsification” in conversation with ChatGPT. These initial results had size complexity $O\left(\frac{n \log(1/\varepsilon)}{\varepsilon^2}\right)$, which the authors then simplified and improved over the subsequent weeks via careful use of the framework of [RR26], again in conversation with ChatGPT. All of the text in this paper is written by the authors.

A. B. thanks Pravesh Kothari for insightful conversations related to sparsification. J. B. thanks Venkatesan Guruswami for valuable discussions and encouragement. A. P. thanks Sanjeev Khanna and Madhu Sudan for useful conversations related to sparsification.

2 Convex Geometry Preliminaries

We now formally define a number of concepts in convex geometry which we need for our study. We closely follow the presentations of Gruber [Gru07] and Schneider [Sch14]. Given a positive integer $n \in \mathbb{N}$, we say that a set $A \subseteq \mathbb{R}^n$ is *convex* if for all $x, y \in A$ and all scalars $\lambda \in [0, 1]$, we have that $\lambda x + (1 - \lambda)y \in A$. If in addition A is also non-empty and compact (i.e., closed with finite volume), then we say that A is a *convex body*. We let $\mathcal{K}^n$ denote the set of all such convex bodies A in $\mathbb{R}^n$.

We say that a convex body A is *proper* if A has non-zero interior. That is, there is a point $x \in A$ and a radius $r > 0$ such that the Euclidean ball $B_n(x, r) \subseteq A$. We let $\mathcal{K}_n^n \subsetneq \mathcal{K}^n$ denote the set of proper convex bodies in $\mathbb{R}^n$.

We say that $A \in \mathcal{K}^n$ is *symmetric* (a.k.a. *centrally symmetric*) if for all $x \in \mathbb{R}^n$, we have that $x \in A$ if and only if $-x \in A$. We let $\mathcal{K}_{\text{sym}}^n$ denote the set of symmetric convex bodies, and $\mathcal{K}_{n,\text{sym}}^n$ denote the set of proper, symmetric convex bodies.

2.1 Minkowski Sums and Differences

Convex bodies support a number of algebraic operations. Given a scalar $\lambda \in \mathbb{R}$ and a convex body $A \in \mathcal{K}^n$, we let $\lambda A := \{\lambda x \mid x \in A\}$. Furthermore, given a vector $y \in \mathbb{R}^n$, we denote the *translate* $A + y := \{x + y \mid x \in A\}$, with $A - y := A + (-y)$.

For any $A, B \in \mathcal{K}^n$ we can define their *Minkowski sum* to be

$$A + B := \bigcup_{y \in B} (A + y) = \{x + y \mid x \in A, y \in B\} \in \mathcal{K}^n.$$

A more subtle notion one can also define is the *Minkowski difference*, which is defined as (see e.g., [Sch14, p.146]²)

$$A \ominus B := \bigcap_{y \in B} (A - y) = \{y \in \mathbb{R}^n \mid B + y \subseteq A\}. \quad (1)$$

It is easy to see that $A \ominus B$ is always convex, as it is defined to be an intersection of convex bodies. However, $A \ominus B$ may be empty and thus disqualify it from being a convex body. It is easy to observe that $A \ominus B$ is nonempty if and only if there exists a translate of B contained in A . In the special case in which A and B are both centrally symmetric, we can see that $A \ominus B$ (if nonempty) is centrally symmetric as well, as $B + y \subseteq A$ if and only if $-(B + y) = B - y \subseteq -A = A$. In particular, this means that $0 \in A \ominus B$, so $B \subseteq A$ by Eq. (1). As such, we have the following fact (observed in, e.g., [RR26]).

Fact 2.1. *Let $A, B \in \mathcal{K}_{\text{sym}}^n$ have the property that $A \ominus B \neq \emptyset$, then $B \subseteq A$.*

We now state some standard facts about Minkowski addition and subtraction (e.g., Chapter 3 of [Sch14] or Proposition 11 in [RR26]).

Fact 2.2. *For any $A, B, C \in \mathcal{K}^n$ and scalars $\alpha, \beta \geq 0$, we have that.*

$$(1) \quad (A + B) + C = A + (B + C).$$

$$(2) \quad (A \ominus B) \ominus C = A \ominus (B + C).$$

$$(3) \quad (A + B) \ominus B = A.$$

$$(4) \quad \alpha A + \beta A = (\alpha + \beta)A$$

$$(5) \quad \alpha A + \alpha B = \alpha(A + B).$$

$$(6) \quad \alpha A \ominus \beta A = (\alpha - \beta)A \text{ if } \alpha \geq \beta.$$

² Schneider [Sch14] uses the notation $A \div B$ for Minkowski difference.

As previously discussed, for arbitrary $A, B \in \mathcal{K}^n$, $A \ominus B$ may be empty. However, if we scale B by some factor λB , then the *parallel body* $A \ominus \lambda B$ is non-empty for $\lambda \geq 0$ sufficiently small. As such, we can define the *inradius* of A relative to B as follows [Sch14, p.148].

$$r(A; B) := \max\{\lambda \geq 0 \mid \exists x \in \mathbb{R}^n, \lambda B + x \subseteq A\}. \quad (2)$$

Note that the maximum here is well-defined as A and B are compact.

2.2 Volumes and Mixed Volumes

Given a convex body $A \in \mathcal{K}^n$, we let $\text{Vol}_n(A)$ denote the standard Euclidean volume of A . Note that $\text{Vol}_n(A) = 0$ if and only if $A \in \mathcal{K}^n \setminus \mathcal{K}_n^n$. By convention, we also let $\text{Vol}_n(\emptyset) = 0$. For some of our computations, we also need a notion of *mixed volume*. In particular, for any $A_1, \dots, A_n \in \mathcal{K}^n$, we define their mixed volume to be (see [Gru07, Proposition 6.7])

$$\text{Vol}_n(A_1, \dots, A_n) := \frac{1}{d!} \left[\sum_{S \subseteq [n]} (-1)^{n-|S|} \text{Vol}_n \left(\sum_{i \in S} A_i \right) \right].$$

The key defining property of mixed volumes is that for any $m \in \mathbb{N}$ and $A_1, \dots, A_m \in \mathcal{K}_n^n$ and scalars $\lambda_1, \dots, \lambda_m \geq 0$, we have that ([Gru07, Theorem 6.5])

$$\text{Vol}_n(\lambda_1 A_1 + \dots + \lambda_m A_m) := \sum_{(i_1, \dots, i_n) \in [m]^n} \lambda_{i_1} \dots \lambda_{i_n} \text{Vol}_n(A_{i_1}, \dots, A_{i_n}).$$

As such, we have that $\text{Vol}_n(A, \dots, A) = \text{Vol}_n(A)$ for any $A \in \mathcal{K}^n$. Further observe that both volume and mixed volume are translation invariant [Gru07, Proposition 6.5]. That is, $\text{Vol}_n(A_1, \dots, A_n) = \text{Vol}_n(A_1 + x_1, \dots, A_n + x_n)$ for all $x_1, \dots, x_n \in \mathbb{R}^n$.

Fact 2.3 (Theorem 6.9 and Corollary 6.1 [Gru07]). *Mixed volume is always nonnegative and is monotone in all of its arguments. That is, for any $A_1, \dots, A_n, B_1, \dots, B_n \in \mathcal{K}^n$ with $A_i \subseteq B_i$ for all $i \in [n]$, we have that*

$$0 \leq V(A_1, \dots, A_n) \leq V(B_1, \dots, B_n).$$

2.3 Quermassintegrals and Differentiability

In this work, we are only concerned with mixed volumes $\text{Vol}_n(A_1, \dots, A_n)$ where the set $\{A_1, \dots, A_n\}$ has at most two elements. This special case has the name *quermassintegrals* and has the following special notation. For convex bodies $A, B \in \mathcal{K}^n$ and integer $i \in \{0, 1, \dots, n\}$, we denote the i th quermassintegral as

$$W_i(A; B) := \text{Vol}_n(\underbrace{A, \dots, A}_{n-i}, \underbrace{B, \dots, B}_i).$$

In particular, $W_0(A; B) = \text{Vol}_n(A)$ and $W_n(A; B) = \text{Vol}_n(B)$.

A key application of these quermassintegrals is differentiating volumes of Minkowski differences. More precisely, given convex bodies with $A, B \in \mathcal{K}^n$, we define for all $\lambda \geq 0$ the quantity

$$f_{A,B}(\lambda) := \text{Vol}_n(A \ominus \lambda B).$$

We observe that this function is continuous at $\lambda = r(A; B)$.

Proposition 2.4. $\text{Vol}_n(A \ominus r(A; B)B) = 0$.

Proof. Assume for sake of contradiction that $\text{Vol}_n(A \ominus r(A; B)B) > 0$. Thus, there exists a Euclidean ball $B_n(x, \varepsilon) \subseteq A \ominus r(A; B)B$. Since B is compact, there exists some $\lambda > 0$ such that $\lambda B + x \subseteq B_n(x, \varepsilon)$. That is, $\lambda B + x \subseteq A \ominus r(A; B)B$. By [Fact 2.2](#), this is equivalent to $x \in A \ominus (r(A; B) + \lambda)B$. By [Eq. \(2\)](#), this implies that $r(A; B) \geq r(A; B) + \lambda$, a contradiction. $\square$

Rather importantly, this function is (usually) differentiable.

Fact 2.5 ([\[Bol43, Mat78\]](#), adapted from Proposition 2.6 of [\[RSG22\]](#)). Assume $A \in \mathcal{K}^n$ and $B \in \mathcal{K}_n^n$, then $f_{A,B}$ is differentiable in the range $(0, r(A; B))$ with

$$f'_{A,B}(\lambda) := -nW_1(A \ominus \lambda B; B). \quad (3)$$

Furthermore, (3) gives the right-derivative of $f_{A,B}(\lambda)$ at $\lambda = 0$ and the left-derivative at $\lambda = r(A; B)$.

For further context and generalizations of [Fact 2.5](#), see some papers by Hernández Cifre and Saorín [\[HCS10b, HCS10a\]](#).

Extending [Fact 2.5](#), we show that $f_{A,B}$ is convex on the entire interval $[0, \infty)$, even when B is improper.

Lemma 2.6. For any $A, B \in \mathcal{K}^n$, we have that $f_{A,B}$ is convex on $[0, \infty)$.

Proof. First assume B is proper (i.e., $B \in \mathcal{K}_n^n$), then by [Fact 2.5](#), we know that the derivative $f'_{A,B}(\lambda) = -nW_1(A \ominus \lambda B; B)$ for all $\lambda \in [0, r(A; B)]$. Note that this derivative is negative and decreasing by [Fact 2.3](#). Thus, $f_{A,B}$ is convex, non-negative, and decreasing on $[0, r(A; B)]$. Since $f_{A,B}(\lambda) = 0$ for all $\lambda \geq r(A; B)$, we can immediately deduce that $f_{A,B}$ is convex on all of $[0, \infty)$. Alternatively, notice that the left derivative of $f_{A,B}$ at $r(A; B)$ is $-nW_1(A \ominus r(A; B)B; B) \leq 0$, while the right derivative is 0, so convexity is preserved along the entirety of $f_{A,B}$.

If $B \in \mathcal{K}^n \setminus \mathcal{K}_n^n$, for any $\varepsilon > 0$, consider $B_\varepsilon := B + B_n(0, \varepsilon)$. Then, f_{A,B_ε} is convex on $[0, \infty)$. Thus, it suffices then to show for all $\lambda \in [0, \infty)$, we have pointwise convergence to $f_{A,B}$ as $\varepsilon \rightarrow 0$. More formally, we claim that

$$\lim_{\varepsilon \rightarrow 0^+} f_{A,B_\varepsilon}(\lambda) = f_{A,B}(\lambda). \quad (4)$$

This fact is trivial for $\lambda = 0$, as $f_{A,B_\varepsilon}(0) = \text{Vol}_n(A)$, so assume $\lambda > 0$. Observe that we always have that since $B \subseteq B_\varepsilon$ for any $\varepsilon > 0$, we have that $A \ominus \lambda B_\varepsilon \subseteq A \ominus \lambda B$. If $\text{Vol}_n(A \ominus \lambda B) = 0$ or $\lambda > r(A; B)$, then convergence is guaranteed by monotonicity and non-negativity. Thus, assume that $\text{Vol}_n(A \ominus \lambda B) > 0$. Using a second invocation of [Fact 2.5](#), we have that $\text{Vol}_n((A \ominus \lambda B) \ominus B_n(0, \delta))$ is a continuous function of δ . Thus, by [Fact 2.2](#), we have that

$$\lim_{\delta \rightarrow 0^+} \text{Vol}_n(A \ominus \lambda B_{\delta/\lambda}) = \lim_{\delta \rightarrow 0^+} \text{Vol}_n((A \ominus \lambda B) \ominus B_n(0, \delta)) = \text{Vol}_n(A \ominus \lambda B).$$

Thus, setting $\varepsilon = \delta/\lambda$, we get [Eq. \(4\)](#). Thus, $f_{A,B}$ is convex. $\square$

2.4 Support Functions

Our sparsification theorem will crucially rely on *support functions* of convex bodies. For a nonempty compact convex set $K \subseteq \mathbb{R}^n$, its support function is

$$h_K(u) = \max_{x \in K} \langle x, u \rangle, \quad u \in \mathbb{R}^n.$$

We have the following fact:

Fact 2.7 (e.g., Chapter 13 of [Roc70] or [DK00]). *Support functions satisfy*

$$h_{K+L} = h_K + h_L, \quad h_{tK} = th_K \quad (t \geq 0),$$

and

$$K \subseteq L \iff h_K(u) \leq h_L(u) \text{ for every } u \in \mathbb{R}^n.$$

Likewise, we also have the following relationship between support functions and Minkowski subtraction:

Fact 2.8 (e.g., Proposition 2 of [DK00]). *For compact convex sets A, B ,*

$$z \in A \ominus B \iff \langle z, u \rangle \leq h_A(u) - h_B(u) \text{ for every } u \in \mathbb{R}^n.$$

3 A One-Sided Volume Lemma

With the above preliminaries established, we now proceed to the key technical lemma which underlies our sparsification theorem. Just as in [RR26] and [BKMT26], this relies on a “one-sided volume lemma”. In our case this one-sided volume lemma will correspond with appropriate one-sided sparsifiers of *sums of support functions of convex bodies*:

Lemma 3.1. *Let $C_1, \dots, C_m$ be compact centrally symmetric convex sets whose Minkowski sum has dimension d in $\mathbb{R}^n$. Define*

$$K_0 := \left\{ z \in [-1, 1]^m : \sum_{i=1}^m z_i h_{C_i}(u) \leq 0 \text{ for every } u \right\}.$$

Then

$$\frac{\text{vol}_m(K_0)}{2^m} \geq e^{-2d}.$$

3.1 The Framework of [RR26]

Let $\Omega \subseteq \mathbb{R}^m$ be a convex set. A function $F : \Omega \rightarrow \mathbb{R}$ is said to be *separately convex* if for any $i \in [m]$ and any $\omega \in \Omega$, the restricted function $t \mapsto F(\omega + te_i) \in \mathbb{R}$ is convex, where $e_i \in \mathbb{R}^m$ is the i^{th} standard basis vector. Separately convex functions satisfy the following convenient multi-variate generalization of Jensen’s theorem [RR26, Lemma 12]: Let $X_1, \dots, X_m$ be independent $\mathbb{R}$ -valued

random variables such that $(X_1, \dots, X_m)$ is supported on Ω . Then for any separately convex function $F : \mathbb{R}^m \rightarrow \mathbb{R}$ we have $\mathbb{E}F(X_1, \dots, X_m) \geq F(\mathbb{E}X_1, \mathbb{E}X_2, \dots, \mathbb{E}X_m)$.

Importantly, the work of [RR26] provides a powerful framework for bounding one-sided volume by using the notion of separate convexity:

Fact 3.2 (Lemma 18 of [RR26]). *Let $F : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ be separately convex and not identically zero. Suppose that, for some integer $d \geq 0$,*

$$F(\rho x) = \rho^d F(x) \quad \text{for every } x \in \{-1, 1\}^m \text{ and } \rho \geq 0.$$

Then

$$\Pr_{\sigma \in \{-1, 1\}^m} [F(\sigma) > 0] \geq e^{-2d}.$$

Note that it is not a priori obvious exactly what the function F should be. In what follows, we show that for an appropriate choice of this function, we can simultaneously (a) satisfy the conditions of the preceding fact, and (b) take advantage of the conclusion of the above fact to conclude our desired volume lemma.

3.2 The Proof of Lemma 3.1

To begin, we will define a function which is separately convex and not identically 0, thereby satisfying the conditions of Fact 3.2.

To start, as in the statement of Lemma 3.1, we let $C_1, \dots, C_m$ be centrally symmetric convex bodies which span a d dimensional subspace of $\mathbb{R}^n$. For $t = (t_1, \dots, t_m) \in \mathbb{R}^m$, we let

$$\mathcal{Z}(t) = \left\{ z \in \mathbb{R}^n : \langle z, u \rangle \leq \sum_{i=1}^m t_i h_{C_i}(u) \text{ for every } u \in \mathbb{R}^n \right\}$$

Importantly, we have the following facts about $\mathcal{Z}(t)$:

Claim 3.3. *$\mathcal{Z}(t)$ is closed and convex. It is also bounded whenever it is nonempty. Moreover, for $a \in \mathbb{R}_{\geq 0}^m$,*

$$\mathcal{Z}(a) = \sum_{i=1}^m a_i C_i.$$

Proof. To see the final item, observe that for any $a \in \mathbb{R}_{\geq 0}^m$, the right-hand side of the definition of $\mathcal{Z}(a)$ is the support function of $\sum_i a_i C_i$. It then immediately follows that $\mathcal{Z}(a) = \sum_{i=1}^m a_i C_i$. $\square$

Now, we define the function

$$\Phi(t) = \text{Vol}_d(\mathcal{Z}(t)),$$

with the understanding that $\Phi(t) = 0$ if $\mathcal{Z}(t) = \emptyset$. Crucially, Φ is separately convex:

Claim 3.4. *For $C_1, \dots, C_m$ and $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ as defined above, Φ is separately convex.*

Proof. To prove this, we rely crucially on [Lemma 2.6](#). We fix any $R > 0$, we set $c = R \cdot \mathbf{1} \in \mathbb{R}^m$, and take $t \in [-R, R]^m$. By [Fact 2.8](#), we then see that

$$\mathcal{Z}(t) = \mathcal{Z}(c) \ominus \mathcal{Z}(c - t).$$

Now, we fix all coordinates except t_j , for any choice of $j \in [m]$. We write

$$D = \sum_{i \neq j} (R - t_i) C_i, \quad A = \mathcal{Z}(c) \ominus D.$$

By [Fact 2.2](#),

$$\mathcal{Z}(t) = A \ominus (R - t_j) C_j.$$

If $A = \emptyset$, then this section of Φ is identically zero. Otherwise A is a nonempty compact convex set and the map

$$t_j \mapsto \Phi(t) = \text{Vol}_d(A \ominus (R - t_j) C_j)$$

is convex by [Lemma 2.6](#). Thus Φ is separately convex on $[-R, R]^m$. Since R is arbitrary, it is separately convex on all of $\mathbb{R}^m$. $\square$

We can then proceed to a proof of [Lemma 3.1](#):

Proof of Lemma 3.1. We let E denote the span of the bodies $C_1, \dots, C_m$, and let d denote the dimension of E . If $d = 0$, the claim is immediate. Otherwise, we define Φ as above. Note that for any $x \in \{\pm 1\}^m$ and any $\rho \geq 0$, we have that

$$\begin{aligned} \Phi(\rho x) &= \text{Vol}_d(\mathcal{Z}(\rho x)) = \text{Vol}_d\left(\sum_{i=1}^m \rho \cdot x_i \cdot C_i\right) = \text{Vol}_d\left(\rho \cdot \sum_{i=1}^m x_i \cdot C_i\right) \\ &= \rho^d \cdot \text{Vol}_d\left(\sum_{i=1}^m x_i \cdot C_i\right) = \rho^d \cdot \Phi(x). \end{aligned}$$

Now, in conjunction with [Claim 3.4](#), this then implies that

$$\Pr_{\delta \sim \{\pm 1\}^m} [\Phi(\delta) > 0] \geq e^{-2d}.$$

Now, consider any vector $\delta \in \{\pm 1\}^m$ such that $\Phi(\delta) > 0$. Let $P = \{i \in [m] : \delta_i = 1\}$, and let $N = [m] - P$. We let C_P denote the sum of the convex bodies corresponding to P ; i.e., $C_P = \sum_{i \in P} C_i$, and we let $C_N = \sum_{i \in N} C_i$. By [Claim 3.3](#), we then see that $\mathcal{Z}(\delta) = C_P \ominus C_N$. Because $\Phi(\delta) > 0$, this then also means that $\mathcal{Z}(\delta)$ is non-empty. Because C_P and C_N are centrally symmetric, it is then the case (by [Fact 2.1](#)) that $C_N \subseteq C_P$. In the view of support functions, this then means that for every $u \in E$,

$$\sum_{i \in N} h_{C_i}(u) \leq \sum_{i \in P} h_{C_i}(u),$$

or equivalently, that for every $u \in E$,

$$\sum_{i=1}^m \delta_i h_{C_i}(u) \geq 0$$

This implies that

$$\Pr_{\delta \sim \{\pm 1\}^m} [\delta_i h_{C_i}(u) \geq 0 \text{ for all } u \in E] \geq e^{-2d};$$

by considering $-\delta$ instead of δ , this then also implies that

$$\Pr_{\delta \sim \{\pm 1\}^m} [\delta_i h_{C_i}(u) \leq 0 \text{ for all } u \in E] \geq e^{-2d}.$$

Finally, it remains to translate this result from the setting of $\{\pm 1\}^m$ vectors to $[-1, 1]^m$ vectors. For this, we sample a random vector $Z \in [-1, 1]^m$. We write the vector Z as $Z_i = R_i \cdot \varepsilon_i$, where $R_i = |Z_i|$ and $\varepsilon_i = \text{sgn}(Z_i)$. The variables R_i are uniform on $[0, 1]$ and ε_i is uniform from $\{\pm 1\}$. Now, we can simply invoke the same reasoning from above on the collection of convex bodies $R_1 \cdot C_1, \dots, R_m \cdot C_m$. On these bodies, we then see that for every choice of R_i ,

$$\Pr_{\varepsilon \sim \{\pm 1\}^m} [\varepsilon_i \cdot R_i h_{C_i}(u) \leq 0 \text{ for all } u \in E] \geq e^{-2d}.$$

Thus,

$$\begin{aligned} & \Pr_{Z \sim [-1, 1]^m} [Z_i \cdot h_{C_i}(u) \leq 0 \text{ for all } u \in E] \\ &= \Pr_{R \sim [0, 1]^m} \Pr_{\varepsilon \sim \{\pm 1\}^m} [\varepsilon_i \cdot R_i h_{C_i}(u) \leq 0 \text{ for all } u \in E] \geq e^{-2d}. \end{aligned}$$

This concludes the claim. $\square$

4 Building Support Sparsifiers for Sums of Convex Bodies

With [Lemma 3.1](#) established, we now invoke the framework of [\[RR26\]](#) for building linear-size sparsifiers. Note that the material of this section is almost entirely from the work of [\[RR26\]](#).

4.1 [\[RR26\]](#) Preliminaries

We start by recording a few facts from [\[RR26\]](#). To do this, we first require the notion of the restriction of a convex-set: for $K \subseteq [-1, 1]^m$, we let $K_S = \{x' \in \mathbb{R}^S : (x', \mathbf{0}^{[m]-S}) \in K\}$.

Fact 4.1 (Reis–Rothvoss symmetrization, Theorem 16 from [\[RR26\]](#)). *Let $p, \eta \in (0, 1/2]$, and let $K \subseteq [-1, 1]^m$ be convex. Suppose that $[-\eta, \eta]^m \subseteq K$ and*

$$\frac{\text{Vol}_S(K_S)}{2^{|S|}} \geq p \quad \text{for every } S \subseteq [m].$$

If

$$m \geq \frac{\log_2(1/p)}{\eta^2},$$

then

$$\text{Vol}_m(K \cap (-K)) \geq 2^{-5m}.$$

We also use the following fact which shows that (appropriate scalings of) large enough symmetric convex bodies contain vectors with many integral coordinates:

Fact 4.2 (Large-body partial coloring, Theorem 7 from [RR26]). *For every constant $c > 0$, there is a constant $s = s(c) > 0$ such that the following holds. If $Q \subseteq [-1, 1]^m$ is a symmetric convex body with*

$$\text{Vol}_m(Q) \geq c^m,$$

then there is a vector

$$z \in sQ \cap [-1, 1]^m$$

for which at least $m/2$ coordinates satisfy $|z_i| = 1$.

4.2 The Partial Coloring

In this section, we now prove the following lemma:

Lemma 4.3. *There are universal constants $A, \kappa > 0$ with the following property. Let $C_1, \dots, C_m$ be compact centrally symmetric convex sets contained in a d -dimensional space, and put $C = \sum_i C_i$. If $m \geq \kappa d$, then there is $z \in [-1, 1]^m$ such that*

(1) *at least $m/4$ coordinates of z equal -1 ; and*

(2) *for every u ,*

$$\left| \sum_{i=1}^m z_i h_{C_i}(u) \right| \leq A \sqrt{\frac{d}{m}} h_C(u).$$

Proof. The proof relies on both Fact 4.1 and Fact 4.2 from above. To start, we can observe that when $d = 0$, the claim is trivial. Thus, we focus only on the case when $d \geq 1$. We set

$$p_0 = e^{-2d}, \quad \beta = 2 \log_2 e, \quad \eta = \sqrt{\frac{\beta d}{m}}.$$

Next, choose $\kappa \geq 4\beta$; in this way if $m \geq \kappa \cdot d$, it must be that $\eta \leq 1/2$. Now, we define a new convex set K_η :

$$K_\eta = \left\{ x \in [-1, 1]^m : \sum_{i=1}^m x_i h_{C_i}(u) \leq \eta h_C(u) \text{ for every } u \right\}.$$

This set is a relaxation of the body considered in Lemma 3.1 as we now allow for some slack on the right-hand side. Importantly, this set now ensures that $[-\eta, \eta]^m \subseteq K_\eta$. Indeed, for any $x \in [-\eta, \eta]^m$,

$$\sum_{i=1}^m x_i \cdot h_{C_i}(u) \leq \sum_{i=1}^m |x_i| \cdot h_{C_i}(u) \leq \eta \cdot h_C(u).$$

Now, fix $S \subseteq [m]$. Because $(K_0)|_S \subseteq (K_\eta)|_S$, we can invoke Lemma 3.1 to see that

$$\frac{\text{Vol}_{|S|}((K_\eta)|_S)}{2^{|S|}} \geq e^{-2d}.$$

Furthermore,

$$\log_2(1/p_0) = 2d \log_2 e = \beta d \quad \text{and} \quad m = \frac{\log_2(1/p_0)}{\eta^2}.$$

Thus [Fact 4.1](#) applies and gives

$$\text{Vol}_m(Q_\eta) \geq 2^{-5m},$$

where

$$Q_\eta = K_\eta \cap (-K_\eta) = \left\{ x \in [-1, 1]^m : \left| \sum_{i=1}^m x_i h_{C_i}(u) \right| \leq \eta h_C(u) \text{ for every } u \right\}.$$

Next, we invoke [Fact 4.2](#) with $c = 2^{-5}$. This implies that there is a universal constant $s > 0$ and a vector

$$z \in s \cdot Q_\eta \cap [-1, 1]^m$$

with at least $m/2$ saturated coordinates (i.e., such that $z_i \in \{\pm 1\}$). Moreover, by taking $-z$ if necessary, we additionally are guaranteed that at least $m/4$ of these coordinates equal -1 . Because $\frac{z}{s} \in Q_\eta$, this implies that for every u ,

$$\left| \sum_{i=1}^m z_i h_{C_i}(u) \right| \leq s \cdot \eta \cdot h_C(u) = s \cdot \sqrt{2 \log_2 e} \cdot \sqrt{\frac{d}{m}} \cdot h_C(u).$$

This implies the claimed result with $A = s \sqrt{2 \log_2 e}$. □

4.3 Building the Sparsifier

We now prove our desired sparsification statement:

Theorem 4.4 (Minkowski-sum sparsification). *There is a universal constant $C_\star > 0$ with the following property. Let E be a finite-dimensional real vector space, and let $C_1, \dots, C_m \subseteq E$ be compact centrally symmetric convex sets. Put*

$$C = C_1 + \dots + C_m, \quad d = \dim \text{span}(C).$$

For every $0 < \varepsilon < 1$, there exist coefficients $\lambda_1, \dots, \lambda_m \geq 0$, supported on at most $C_\star \cdot d/\varepsilon^2$ indices, such that for every $u \in E$,

$$(1 - \varepsilon) \cdot \sum_{i=1}^m h_{C_i}(u) \leq \sum_{i=1}^m \lambda_i \cdot h_{C_i}(u) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m h_{C_i}(u).$$

Proof. Our proof closely follows the strategy used to prove [\[RR26, Theorem 26\]](#). To start, we discard all indices for which $h_{C_i}(u)$ is identically zero. Note that if $d = 0$, then $\sum_{i=1}^m h_{C_i}(u) \equiv 0$, and choosing every $\lambda_i = 0$ proves the theorem. Hence assume $d \geq 1$. Note also that it is enough to consider $0 < \varepsilon \leq 1/2$. Indeed, for $1/2 < \varepsilon < 1$, applying the result with accuracy $1/2$ gives a stronger approximation and uses $O(d) = O(d/\varepsilon^2)$ summands.

Let A and κ be the constants from [Lemma 4.3](#), and set

$$q = \sqrt{\frac{3}{4}}, \quad B = \frac{A}{1-q}.$$

Next, we choose a universal constant $C_\star$ so large that

$$\tau_\star \geq 4\kappa, \quad \frac{B}{\sqrt{C_\star}} \leq \frac{1}{2}. \quad (5)$$

Define the target support size

$$M = \left\lceil C_\star \cdot \frac{d}{\varepsilon^2} \right\rceil. \quad (6)$$

The guiding intuition for building the sparsifier is that [Lemma 4.3](#) provides a “weight perturbation” to be applied to the sparsifier (whose weights start off as being all 1’s). In this way, when a support functions receive a weight perturbation of -1 , this zeroes out the contribution from that support function. We then iteratively apply this procedure until we reach our desired sparsity.

We maintain nonnegative weights $w_i^{(t)}$ and the current Minkowski sum

$$C^{(t)} = \sum_{i=1}^m w_i^{(t)} C_i.$$

Initially $w_i^{(0)} = 1$, so $C^{(0)} = C$. Let

$$S_t = \{i : w_i^{(t)} > 0\}, \quad m_t = |S_t|.$$

If $m_t \leq M$, we terminate the algorithm, as this implies we have reached our desired sparsity. Otherwise $m_t > M \geq \kappa d$, and we apply [Lemma 4.3](#) to the active summands

$$D_i^{(t)} = w_i^{(t)} C_i, \quad i \in S_t.$$

By doing this, we obtain $z^{(t)} \in [-1, 1]^{S_t}$ with at least $m_t/2$ saturated coordinates such that for every u :

$$\left| \sum_{i \in S_t} z_i^{(t)} h_{D_i^{(t)}}(u) \right| \leq \delta_t h_{C^{(t)}}(u), \quad \delta_t = A \sqrt{\frac{d}{m_t}}. \quad (7)$$

By $m_t > M$ and (5),

$$\delta_t < \frac{A}{\sqrt{C_\star}} \varepsilon \leq \frac{B}{\sqrt{C_\star}} \varepsilon \leq \frac{\varepsilon}{2} < 1.$$

As proved in [Lemma 4.3](#), at least $m_t/4$ of the saturated coordinates are equal to -1 , i.e.,

$$\left| \left\{ i \in S_t : z_i^{(t)} = -1 \right\} \right| \geq \frac{m_t}{4}.$$

We now apply this z as our weight perturbation: we update

$$w_i^{(t+1)} = (1 + z_i^{(t)}) w_i^{(t)} \quad (i \in S_t), \quad (8)$$

and retain weight zero outside S_t . The weights remain nonnegative, and every coordinate with $z_i^{(t)} = -1$ disappears. Consequently,

$$m_{t+1} \leq \frac{3}{4}m_t. \quad (9)$$

The second detail is that we must manage the deterioration of the sparsifier accuracy. For this, we can observe that

$$\begin{aligned} h_{C^{(t+1)}}(u) &= \sum_{i \in S_t} (1 + z_i^{(t)}) h_{D_i^{(t)}}(u) \\ &= h_{C^{(t)}}(u) + \sum_{i \in S_t} z_i^{(t)} h_{D_i^{(t)}}(u). \end{aligned}$$

Thus (7) and the support-function characterization of inclusion imply that for every u ,

$$(1 - \delta_t) \cdot h_{C^{(t+1)}}(u) \leq h_{C^{(t)}}(u) \leq (1 + \delta_t) \cdot h_{C^{(t+1)}}(u). \quad (10)$$

The preceding equation bounds the accuracy deterioration in a single round of weight updates. We now bound the total result of these updates across all iterations. We let T be the first index for which $m_T \leq M$. If $T = 0$, there is nothing to prove. Otherwise $m_{T-1} > M$. From (9), for $0 \leq t \leq T-1$,

$$m_t \geq \left(\frac{4}{3}\right)^{T-1-t} m_{T-1}.$$

Therefore

$$\begin{aligned} \sum_{t=0}^{T-1} \delta_t &\leq A \sqrt{\frac{d}{m_{T-1}}} \sum_{j=0}^{\infty} \left(\sqrt{\frac{3}{4}}\right)^j \\ &< \frac{A}{1 - \sqrt{3/4}} \sqrt{\frac{d}{M}} = B \sqrt{\frac{d}{M}} \leq \frac{B}{\sqrt{C_\star}} \varepsilon \leq \frac{\varepsilon}{2}. \end{aligned} \quad (11)$$

In particular, every $\delta_t < 1$. Iterating (10) gives

$$\left(\prod_{t=0}^{T-1} (1 - \delta_t)\right) \cdot h_{C^{(t)}}(u) \leq h_C(u) \leq \left(\prod_{t=0}^{T-1} (1 + \delta_t)\right) \cdot h_{C^{(t)}}(u).$$

Since $\prod_{t=0}^{T-1} (1 - \delta_t) \geq 1 - \sum_{t=0}^{T-1} \delta_t$,

$$\prod_{t=0}^{T-1} (1 - \delta_t) \geq 1 - \varepsilon.$$

Likewise, by (11),

$$\prod_{t=0}^{T-1} (1 + \delta_t) \leq \exp\left(\sum_{t=0}^{T-1} \delta_t\right) \leq e^{\varepsilon/2} \leq 1 + \varepsilon,$$

where the last inequality holds for $0 \leq \varepsilon \leq 1$.

All together, this then implies that for every u ,

$$(1 - \varepsilon) \cdot h_C(u) \subseteq h_{C^{(T)}}(u) \subseteq (1 + \varepsilon) h_C(u).$$

Finally,

$$C^{(T)} = \sum_{i=1}^m w_i^{(T)} C_i$$

is supported on $m_T \leq M = O(d/\varepsilon^2)$ indices. Taking $\lambda_i = w_i^{(T)}$ completes the proof. $\square$

5 Deriving Seminorm Sparsifiers

In this section, we now adapt the results from the previous section to building *seminorm* sparsifiers. We first recall the definition of a seminorm:

Definition 5.1. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ is said to be a seminorm if it satisfies the following properties:

- (1) **Positive Homogeneity.** For all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, $f(\lambda x) = |\lambda| \cdot f(x)$.
- (2) **Subadditivity.** For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$.

Importantly, we have the following claim which relates seminorms to support functions of centrally symmetric convex bodies:

Fact 5.2 (Theorem 1.7.1 of [Sch14]). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ be a seminorm. Then, there exists a compact, convex, centrally symmetric convex set $C \subseteq \mathbb{R}^n$ such that*

$$f(x) = h_C(x).$$

Remark. Note that [Sch14] Theorem 1.7.1 uses the notion of “sublinear functions” which refer to functions that satisfy both positive homogeneity and subadditivity, thereby including seminorms (see Definition 5.1). Theorem 1.7.1 does not explicitly say that $C \subseteq \mathbb{R}^n$ is centrally symmetric in the case of seminorms, but this follows from the proof [Sch14, p.46] which shows that

$$C := \{x \in \mathbb{R}^n \mid \langle x, u \rangle \leq f(u) \text{ for every } u \in \mathbb{R}^n\}$$

which is centrally symmetric. $\square$

With this, we can now conclude with the statement of sparsifying sums of seminorms:

Theorem 5.3. *Let $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ be a collection of seminorms. Then, for any $\varepsilon > 0$, there exists a set of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ with only $O(n/\varepsilon^2)$ of the weights non-zero such that for every $x \in \mathbb{R}^n$:*

$$(1 - \varepsilon) \cdot \sum_{i=1}^m f_i(x) \leq \sum_{i=1}^m \lambda_i \cdot f_i(x) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m f_i(x).$$

Proof. Indeed, for each seminorm f_i , let C_i denote the corresponding centrally symmetric, convex set as guaranteed by [Fact 5.2](#). It is then the case that for every $x \in \mathbb{R}^n$ and every $i \in [m]$ that

$$h_{C_i}(x) = f_i(x).$$

We can then invoke [Theorem 4.4](#) on the collection $C_1, \dots, C_m$. Note that the dimension of the span of these bodies is bounded by n , as they are contained in $\mathbb{R}^n$. This guarantees $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that at most $O(n/\varepsilon^2)$ of the weights are non-zero, and yet still, for every $u \in \mathbb{R}^n$

$$(1 - \varepsilon) \cdot \sum_{i=1}^m h_{C_i}(u) \leq \sum_{i=1}^m \lambda_i \cdot h_{C_i}(u) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m h_{C_i}(u).$$

This then immediately implies that the same weights $\lambda_1, \dots, \lambda_m$ also satisfy

$$(1 - \varepsilon) \cdot \sum_{i=1}^m f_i(x) \leq \sum_{i=1}^m \lambda_i \cdot f_i(x) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m f_i(x)$$

for every $x \in \mathbb{R}^n$. This concludes the proof. $\square$

6 Applications

With [Theorem 5.3](#) in hand, we are now able to derive the sparsifiability of many concrete objects. In this section, we discuss two such applications, both improving over the state of the art.

6.1 Hypergraph Cut Sparsifiers

To start, we consider the setting of *hypergraph cut sparsifiers*. In this setting, one is given a hypergraph H and a hypergraph cut-sparsifier is a re-weighted subset of hyperedges which preserves all cut-sizes:

Definition 6.1. Let $H = (V, E, w)$ be a hypergraph, where $w : E \rightarrow \mathbb{R}_{\geq 0}$ is a set of non-negative edge-weights. For a set $S \subseteq V$, the cut-size of S is defined as

$$\delta_H(S) := \sum_{e \in E} w(e) \cdot \mathbf{1}[S \cap e \neq \emptyset \text{ and } (V - S) \cap e \neq \emptyset].$$

For a parameter $\varepsilon > 0$, hypergraph H' is said to be a $(1 \pm \varepsilon)$ cut-sparsifier of H if for every $S \subseteq V$ it is the case that

$$(1 - \varepsilon)\delta_H(S) \leq \delta_{H'}(S) \leq (1 + \varepsilon)\delta_H(S).$$

We now define an auxiliary relaxation of the hypergraph cut function:

Definition 6.2. Let n be an integer, and let $e \subseteq [n]$. For $x \in \mathbb{R}^n$, define $f_e(x) = \max_{u,v \in e} |x_u - x_v|$.

Importantly, we have the following immediate consequences:

Claim 6.3. Let n be an integer, and let $e \subseteq [n]$. Then, for f_e as defined in [Definition 6.2](#), we have that:

(1) f_e is a seminorm.

(2) For any $S \subseteq [n]$, $f_e(\mathbf{1}_S) = \mathbf{1}[S \cap e \neq \emptyset \text{ and } ([n] - S) \cap e \neq \emptyset]$.

Proof. Indeed, we can see that $f_e(x) \geq 0$. Likewise, for any $\lambda \in \mathbb{R}$, $f_e(\lambda \cdot x) = \max_{u,v \in e} |\lambda \cdot x_u - \lambda \cdot x_v| = |\lambda| \cdot \max_{u,v \in e} |x_u - x_v| = |\lambda| \cdot f_e(x)$. For subadditivity, we can observe that

$$f_e(x + y) = \max_{u,v \in e} |(x + y)_u - (x + y)_v| \leq \max_{u,v \in e} (|x_u - x_v| + |y_u - y_v|) \leq f_e(x) + f_e(y).$$

Now, for the second condition, we can observe that

$$f_e(\mathbf{1}_S) = \max_{u,v \in e} |(\mathbf{1}_S)_u - (\mathbf{1}_S)_v|.$$

This expression evaluates to 1 if and only if there is some choice of u, v such that $u \in e \cap S$, $v \in e \cap ([n] - S)$, and otherwise the expression equals 0. This exactly matches $\mathbf{1}[S \cap e \neq \emptyset \text{ and } ([n] - S) \cap e \neq \emptyset]$. $\square$

With this established, we now have the following theorem:

Theorem 6.4. *Let $H = (V, E, w)$ be any hypergraph (with non-negative weights). Then, for any $\varepsilon > 0$, there exists a hypergraph $H' = (V, E, w')$ (again, with non-negative weights) such that:*

(1) H' is a $(1 \pm \varepsilon)$ hypergraph cut-sparsifier of H .

(2) H' has at most $O(|V|/\varepsilon^2)$ hyperedges with non-zero weight: i.e., $|\{e \in E : w'(e) > 0\}| \leq O(|V|/\varepsilon^2)$.

Proof. We consider the collection of seminorms $w(e) \cdot f_e : \mathbb{R}^V \rightarrow \mathbb{R}_{\geq 0}$ for f_e as defined in [Definition 6.2](#). Letting m denote the number of hyperedges, by invoking [Theorem 5.3](#), we immediately obtain that there exists a set of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ with only $O(|V|/\varepsilon^2)$ of the weights non-zero such that for every $x \in \mathbb{R}^V$:

$$(1 - \varepsilon) \cdot \sum_{e \in E} w(e) f_e(x) \leq \sum_{e \in E} \lambda_e \cdot w(e) \cdot f_e(x) \leq (1 + \varepsilon) \cdot \sum_{e \in E} w(e) f_e(x).$$

Importantly, this implies that for every $S \subseteq V$ that

$$(1 - \varepsilon) \cdot \sum_{e \in E} w(e) f_e(\mathbf{1}_S) \leq \sum_{e \in E} \lambda_e \cdot w(e) \cdot f_e(\mathbf{1}_S) \leq (1 + \varepsilon) \cdot \sum_{e \in E} w(e) f_e(\mathbf{1}_S).$$

By the equivalence of [Claim 6.3](#), we then see that the preceding equation is equivalent to guaranteeing that for every $S \subseteq V$,

$$(1 - \varepsilon) \delta_H(S) \leq \delta_{H'}(S) \leq (1 + \varepsilon) \delta_H(S),$$

where H' is the hypergraph with edge set E and weights $w'(e) = \lambda(e) \cdot w(e)$. This then concludes the claim, as at most $O(|V|/\varepsilon^2)$ edges have non-zero weight under w' . $\square$

6.2 Symmetric Submodular Sparsification

Following the work of [JLLS23], in this section, we show how [Theorem 5.3](#) leads to better sparsifiers for sums of symmetric submodular functions.

Recall that for a set $[n]$, the function $f : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ is said to be submodular if for all $S, T \subseteq [n]$,

$$f(S) + f(T) \geq f(S \cap T) + f(S \cup T).$$

Such a submodular function is additionally said to be *symmetric* if $f(S) = f([n] - S)$ for every $S \subseteq V$.

We now define sparsification for sums of symmetric submodular functions:

Definition 6.5. Let n be an integer, and let $f_1, \dots, f_m : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ be symmetric submodular functions. For a parameter $\varepsilon > 0$, we say that a collection of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ is a $(1 \pm \varepsilon)$ sparsifier of $f_1, \dots, f_m$ if for every $S \subseteq [n]$ it is the case that

$$(1 - \varepsilon) \cdot \sum_{i=1}^m f_i(S) \leq \sum_{i=1}^m \lambda_i \cdot f_i(S) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m f_i(S).$$

As a direct consequence of [Theorem 5.3](#), we obtain the following:

Theorem 6.6. Let n be an integer, and let $f_1, \dots, f_m : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ be symmetric submodular functions. For any parameter $\varepsilon > 0$, there exists a collection of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that:

- (1) $\lambda_1, \dots, \lambda_m$ constitutes a $(1 \pm \varepsilon)$ sparsifier of $f_1, \dots, f_m$.
- (2) At most $O(n/\varepsilon^2)$ of the weights $\lambda_1, \dots, \lambda_m$ are non-zero.

The key ingredient in proving [Theorem 6.6](#) (paralleling the key ingredient of the analogous statement in [JLLS23]) is the following:

Fact 6.7 (See, for instance, [Lov83, JLLS23]). Let n be an integer and let $f : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ be a symmetric submodular function. Then, the Lovász extension of f , denoted $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ is a continuous seminorm such that for any $S \subseteq [n]$, $f(S) = \hat{f}(\mathbf{1}_S)$.

With this fact, we can now prove the stated theorem.

Proof of Theorem 6.6. For each symmetric submodular function f_i , we replace it with its corresponding Lovász extension $\hat{f}_i$. We then invoke [Theorem 5.3](#) on the collection $\hat{f}_i : i \in [m]$. This returns $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ with only $O(n/\varepsilon^2)$ of the weights non-zero such that for every $x \in \mathbb{R}^n$:

$$(1 - \varepsilon) \cdot \sum_{i=1}^m \hat{f}_i(x) \leq \sum_{i=1}^m \lambda_i \cdot \hat{f}_i(x) \leq (1 + \varepsilon) \cdot \sum_{i=1}^m \hat{f}_i(x).$$

In particular, by taking $x = \mathbf{1}_S$ for any $S \subseteq [n]$, this implies that the same set of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ is also a $(1 \pm \varepsilon)$ sparsifier of the collection $f_1, \dots, f_m$. This concludes the theorem. $\square$

As an immediate corollary, this also implies $O(n/\varepsilon^2)$ -size sparsifiers for sums of *monotone* submodular functions:

Definition 6.8. A submodular function $f : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ is said to be monotone if for any $S \subseteq [n]$ and $v \in [n] - S$,

$$f(S \cup \{v\}) \geq f(S).$$

We then have the following corollary:

Corollary 6.9. Let n be an integer, and let $f_1, \dots, f_m : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$ be monotone submodular functions. For any parameter $\varepsilon > 0$, there exists a collection of weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that:

- (1) $\lambda_1, \dots, \lambda_m$ constitutes a $(1 \pm \varepsilon)$ sparsifier of $f_1, \dots, f_m$.
- (2) At most $O(n/\varepsilon^2)$ of the weights $\lambda_1, \dots, \lambda_m$ are non-zero.

Proof. This follows from Theorem 1.5 of [KPS24b], which states that for any collection of monotone submodular functions $f_1, \dots, f_m : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$, one can define a collection of symmetric submodular functions $\tilde{f}_1, \dots, \tilde{f}_m : 2^{[n] \cup \{\star\}}$ where for every $i \in [m]$ and $S \subseteq [n]$:

$$f_i(S) = \tilde{f}_i(S) = \tilde{f}_i([n] \cup \{\star\} - S).$$

Sparsifying this collection of $\tilde{f}_1, \dots, \tilde{f}_m$ then immediately implies sparsification of $f_1, \dots, f_m$. $\square$

References

- [ADK⁺16] Ittai Abraham, David Durfee, Ioannis Koutis, Sebastian Krinninger, and Richard Peng. On fully dynamic graph sparsifiers. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, Hyatt Regency, New Brunswick, New Jersey, USA, October 9-11, 2016*, pages 335–344. IEEE Computer Society, 2016. doi:10.1109/FOCS.2016.44. (pg. 1)
- [AGM12a] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 459–467. SIAM, 2012. doi:10.1137/1.9781611973099.40. (pg. 1)
- [AGM12b] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Graph sketches: sparsification, spanners, and subgraphs. In Michael Benedikt, Markus Krötzsch, and Maurizio Lenzerini, editors, *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2012, Scottsdale, AZ, USA, May 20-24, 2012*, pages 5–14. ACM, 2012. doi:10.1145/2213556.2213560. (pg. 1)
- [BCC26] Zhengting Bao, Dongrun Cai, and Xue Chen. Sparsify submodular functions under cardinality constraints. *arXiv preprint arXiv:2606.20777*, 2026. (pg. 4)

- [BG25] Joshua Brakensiek and Venkatesan Guruswami. Redundancy is all you need. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC '25*, page 1614–1625, New York, NY, USA, 2025. Association for Computing Machinery. doi:[10.1145/3717823.3718212](https://doi.org/10.1145/3717823.3718212). (pg. 1)
- [BGP26] Joshua Brakensiek, Venkatesan Guruswami, and Aaron Putterman. Classification of non-redundancy of boolean predicates of arity 4. In Nicolas Beldiceanu, editor, *32nd International Conference on Principles and Practice of Constraint Programming, CP 2026, Lisbon, Portugal, July 20-23, 2026*, volume 379 of *LIPICs*, pages 8:1–8:24. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2026. URL: <https://doi.org/10.4230/LIPICs.CP.2026.8>, doi:[10.4230/LIPICs.CP.2026.8](https://doi.org/10.4230/LIPICs.CP.2026.8). (pg. 1)
- [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 Theory of Computing, STOC '96*, page 47–55, New York, NY, USA, 1996. Association for Computing Machinery. doi:[10.1145/237814.237827](https://doi.org/10.1145/237814.237827). (pg. 1)
- [BKLM26] Arpon Basu, Pravesh K. Kothari, Yang P. Liu, and Raghu Meka. *Sparsifying Sums of Positive Semidefinite Matrices*, pages 6042–6064. 2026. doi:[10.1137/1.9781611978971.216](https://doi.org/10.1137/1.9781611978971.216). (pg. 1)
- [BKMT26] Arpon Basu, Pravesh K. Kothari, Raghu Meka, and Stefan Tudose. Optimal sparsifiers for abelian cayley graphs, 2026. [arXiv:2607.08261](https://arxiv.org/abs/2607.08261). (pg. 1, 7, 12)
- [Bol43] G. Bol. Beweis einer Vermutung von H. Minkowski. *Abh. Math. Sem. Hansischen Univ.*, 15:37–56, 1943. doi:[10.1007/BF02941073](https://doi.org/10.1007/BF02941073). (pg. 7, 11)
- [BSS14] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Review*, 56(2):315–334, 2014. doi:[10.1137/130949117](https://doi.org/10.1137/130949117). (pg. 1)
- [CKN20] Yu Chen, Sanjeev Khanna, and Ansh Nagda. Near-linear size hypergraph cut sparsifiers. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 61–72. IEEE, 2020. doi:[10.1109/FOCS46700.2020.00015](https://doi.org/10.1109/FOCS46700.2020.00015). (pg. 1, 4)
- [CKST19] Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. Optimal lower bounds for sketching graph cuts. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2565–2569. SIAM, 2019. doi:[10.1137/1.9781611975482.158](https://doi.org/10.1137/1.9781611975482.158). (pg. 3, 4)
- [DK00] Vladimir I. Danilov and Gleb A. Koshevoy. Cores of cooperative games, superdifferentials of functions, and the Minkowski difference of sets. *J. Math. Anal. Appl.*, 247(1):1–14, 2000. doi:[10.1006/jmaa.2000.6756](https://doi.org/10.1006/jmaa.2000.6756). (pg. 12)

- [Gru07] Peter M. Gruber. *Convex and discrete geometry*, volume 336 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Berlin, 2007. (pg. 2, 8, 10)
- [HCS10a] M. A. Hernández Cifre and E. Saorín. On the volume of inner parallel bodies. *Adv. Geom.*, 10(2):275–286, 2010. doi:[10.1515/ADVGEOM.2010.004](https://doi.org/10.1515/ADVGEOM.2010.004). (pg. 11)
- [HCS10b] María A. Hernández Cifre and Eugenia Saorín. On differentiability of quermassintegrals. *Forum Math.*, 22(1):115–126, 2010. doi:[10.1515/FORUM.2010.006](https://doi.org/10.1515/FORUM.2010.006). (pg. 11)
- [HLM⁺26] Jun-Ting Hsieh, Daniel Z. Lee, Sidhanth Mohanty, Aaron Putterman, and Rachel Yun Zhang. Sparsifying cayley graphs on every group. In Kasper Green Larsen and Barna Saha, editors, *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2026, Vancouver, BC, Canada, January 11-14, 2026*, pages 6029–6041. SIAM, 2026. doi:[10.1137/1.9781611978971.215](https://doi.org/10.1137/1.9781611978971.215). (pg. 1)
- [JLLS23] Arun Jambulapati, James R Lee, Yang P Liu, and Aaron Sidford. Sparsifying sums of norms. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1953–1962. IEEE, 2023. (pg. 1, 3, 4, 5, 23)
- [JLS23] Arun Jambulapati, Yang P. Liu, and Aaron Sidford. Chaining, group leverage score overestimates, and fast spectral hypergraph sparsification. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 196–206. ACM, 2023. doi:[10.1145/3564246.3585136](https://doi.org/10.1145/3564246.3585136). (pg. 1, 4)
- [KK15] Dmitry Kogan and Robert Krauthgamer. Sketching cuts in graphs and hypergraphs. In Tim Roughgarden, editor, *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 367–376. ACM, 2015. doi:[10.1145/2688073.2688093](https://doi.org/10.1145/2688073.2688093). (pg. 1, 4)
- [KK23] Yotam Kenneth and Robert Krauthgamer. Cut sparsification and succinct representation of submodular hypergraphs. *arXiv preprint arXiv:2307.09110*, 2023. (pg. 4)
- [KKTY21a] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Spectral hypergraph sparsifiers of nearly linear size. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 1159–1170. IEEE, 2021. doi:[10.1109/FOCS52979.2021.00114](https://doi.org/10.1109/FOCS52979.2021.00114). (pg. 1, 4)
- [KKTY21b] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Towards tight bounds for spectral sparsification of hypergraphs. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 598–611. ACM, 2021. doi:[10.1145/3406325.3451061](https://doi.org/10.1145/3406325.3451061). (pg. 1, 4)

- [KLM⁺17] Michael Kapralov, Yin Tat Lee, CN Musco, Christopher Paul Musco, and Aaron Sidford. Single pass spectral sparsification in dynamic streams. *SIAM Journal on Computing*, 46(1):456–477, 2017. (pg. 1)
- [KMM⁺19] Michael Kapralov, Aida Mousavifar, Cameron Musco, Christopher Musco, and Navid Nouri. Faster spectral sparsification in dynamic streams. *arXiv preprint arXiv:1903.12165*, 2019. (pg. 1)
- [KPS24a] Sanjeev Khanna, Aaron Putterman, and Madhu Sudan. Code sparsification and its applications. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 5145–5168. SIAM, 2024. doi:[10.1137/1.9781611977912.185](https://doi.org/10.1137/1.9781611977912.185). (pg. 1, 4)
- [KPS24b] Sanjeev Khanna, Aaron Louie Putterman, and Madhu Sudan. Almost-tight bounds on preserving cuts in classes of submodular hypergraphs. In *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*, pages 98–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024. (pg. 4, 5, 24)
- [KPS25] Sanjeev Khanna, Aaron Putterman, and Madhu Sudan. Efficient algorithms and new characterizations for csp sparsification. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC '25*, page 407–416, New York, NY, USA, 2025. Association for Computing Machinery. doi:[10.1145/3717823.3718205](https://doi.org/10.1145/3717823.3718205). (pg. 1)
- [KPS26] Sanjeev Khanna, Aaron Putterman, and Madhu Sudan. A unified theory of sparsification. *arXiv preprint arXiv:2607.16126*, 2026. (pg. 1)
- [KŽ23] Jannik Kudla and Stanislav Živný. Sparsification of monotone k -submodular functions of low curvature. *arXiv preprint arXiv:2302.03143*, 2023. (pg. 4)
- [Lee23] James R. Lee. Spectral hypergraph sparsification via chaining. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 207–218. ACM, 2023. doi:[10.1145/3564246.3585165](https://doi.org/10.1145/3564246.3585165). (pg. 1, 4)
- [LMW26] Shachar Lovett, Raghu Meka, and Yimeng Wang. Moonflowers and efficient code sparsification. *arXiv preprint arXiv:2605.08676*, 2026. (pg. 1)
- [Lov83] László Lovász. Submodular functions and convexity. In *Mathematical Programming The State of the Art: Bonn 1982*, pages 235–257. Springer, 1983. (pg. 4, 23)
- [Mat78] G. Matheron. La formule de Steiner pour les érosions. *J. Appl. Probability*, 15(1):126–135, 1978. doi:[10.2307/3213242](https://doi.org/10.2307/3213242). (pg. 7, 11)
- [McG14] Andrew McGregor. Graph stream algorithms: a survey. *ACM SIGMOD Record*, 43(1):9–20, 2014. (pg. 1)

- [Qua24] Kent Quanrud. Quotient sparsification for submodular functions. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5209–5248. SIAM, 2024. (pg. 4)
- [Roc70] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, NJ, 1970. (pg. 12)
- [RR26] Victor Reis and Thomas Rothvoss. Linear-size ℓ_1 sparsifiers, 2026. URL: <https://arxiv.org/abs/2606.28147>, [arXiv:2606.28147](https://arxiv.org/abs/2606.28147). (pg. , 1, 2, 3, 5, 6, 7, 8, 9, 12, 13, 15, 16, 17)
- [RSG22] Christian Richter and Eugenia Saorín Gómez. On the monotonicity of the isoperimetric quotient for parallel bodies. *J. Geom. Anal.*, 32(1):Paper No. 15, 19, 2022. doi:[10.1007/s12220-021-00766-2](https://doi.org/10.1007/s12220-021-00766-2). (pg. 11)
- [RY22] Akbar Rafiey and Yuichi Yoshida. Sparsification of decomposable submodular functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10336–10344, 2022. (pg. 4)
- [Sch14] Rolf Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014. (pg. 2, 8, 9, 10, 20)
- [SHS15] Marcel K De Carli Silva, Nicholas JA Harvey, and Cristiane M Sato. Sparse sums of positive semidefinite matrices. *ACM Transactions on Algorithms (TALG)*, 12(1):1–17, 2015. (pg. 1)
- [SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011. doi:[10.1137/080734029](https://doi.org/10.1137/080734029). (pg. 1)
- [ST04] Daniel A Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90, 2004. (pg. 1)
- [ST11] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011. doi:[10.1137/08074489X](https://doi.org/10.1137/08074489X). (pg. 1)
- [SY19] Tasuku Soma and Yuichi Yoshida. Spectral sparsification of hypergraphs. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2570–2581. SIAM, 2019. doi:[10.1137/1.9781611975482.159](https://doi.org/10.1137/1.9781611975482.159). (pg. 1, 4)
- [Tal90] Michel Talagrand. Embedding subspaces of ℓ_1 into ℓ_n . *Proceedings of the American Mathematical Society*, pages 363–369, 1990. (pg. 1, 3)