Spectral Clustering on Large Datasets: When Does it Work? Theory from Continuous Clustering and Density Cheeger-Buser

Timothy Chu	Gary L. Miller	Noel J. Walkington
Google	CMU	CMU
timothyzchu@gmail.com	glmiller@cs.cmu.edu	noelw@andrew.cmu.edu

May 12, 2023

Abstract

Spectral clustering is one of the most popular clustering algorithms that has stood the test of time. It is simple to describe, can be implemented using standard linear algebra, and often finds better clusters than traditional clustering algorithms like k-means and k-centers. The foundational algorithm for two-way spectral clustering, by Shi and Malik, creates a geometric graph from data and finds a spectral cut of the graph.

In modern machine learning, many data sets are modeled as a large number of points drawn from a probability density function. Little is known about when spectral clustering works in this setting – and when it doesn't. Past researchers justified spectral clustering by appealing to the graph Cheeger inequality (which states that the spectral cut of a graph approximates the "Normalized Cut"), but this justification is known to break down on large data sets.

We provide theoretically-informed intuition about spectral clustering on large data sets drawn from probability densities, by proving when a continuous form of spectral clustering considered by past researchers (the *unweighted spectral cut*) finds good clusters of the underlying density itself. Our work suggests that Shi-Malik spectral clustering works well on data drawn from mixtures of Laplace distributions, and works poorly on data drawn from certain other densities, such as a density we call the 'square-root trough'.

Our core theorem proves that *weighted* spectral cuts have low weighted isoperimetry for all probability densities. Our key tool is a new Cheeger-Buser inequality for all probability densities, including discontinuous ones.

Contents

1   Introduction

Spectral clustering [SM97, NJW01] is one of the most popular methods in machine learning for finding clusters in data [VL07, VLBB08, BN04, BN07, BN05, LLKL11, YHJ09, KL19, CCLB21, ZR18, ACSS20, KSAM19, MS22, SSL$^+$18, BGA20], and has inspired modern deep-learning based approaches to clustering and semantic segmentation [LUZ17, SSL$^+$18, MKRLV22]. The foundational paper in spectral clustering, by Shi and Malik [SM97], gave a two-way normalized spectral clustering algorithm based on spectral cuts of a geometric graph from data.

Despite this popularity, little is known theoretically about the quality of clusters given by spectral clustering on large data sets. Researchers previously justified spectral clustering [BN04, MS22] by appealing to the graph Cheeger inequality [AM84], which guarantees that the spectral cut of a geometric graph has a low surface area (sum of edge weights crossing the cut) to volume (sum of edge weights on the smaller side of the cut) ratio. [1] However, graph Cheeger gives very poor guarantees on large or infinite graphs [Tre21, BHL$^+$15, KKRT16]. [2]

One fundamental assumption of machine learning is that data can be modeled as a points in Euclidean space drawn from an underlying probability density. Researchers have analyzed spectral clustering by examining its behavior as the number of samples drawn from a density grows large [VLBB08, TS18, TSVB$^+$16]. Past work showed that normalized Shi-Malik spectral clustering [SM97] on a large sample from probability density ρ is related to its continuous form: the **unweighted spectral cut** of ρ^2 [TS18].
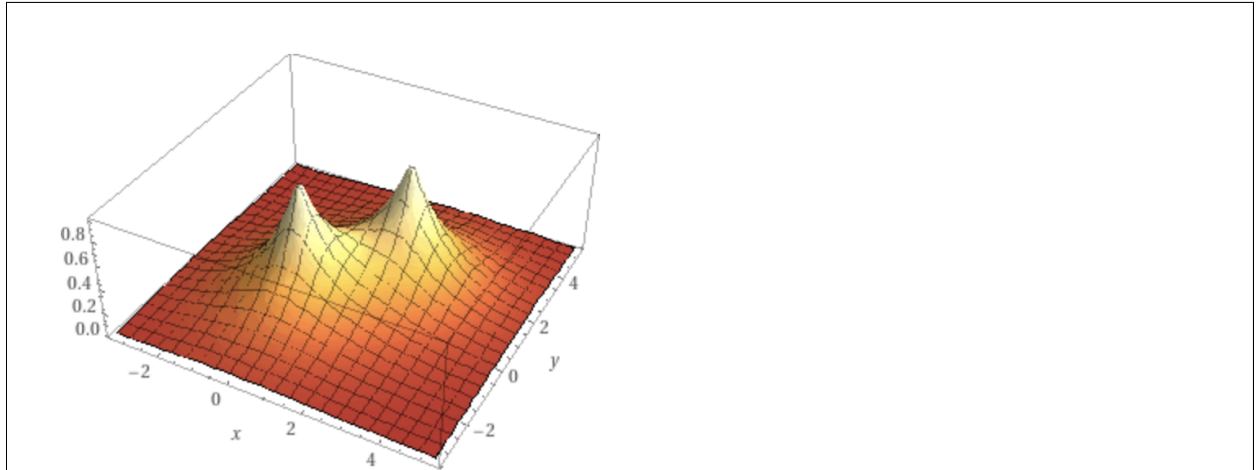
Our Contribution:

Figure 1.1: The mixture of Laplace distributions $e^{-\sqrt{x^2+y^2}} + e^{-\sqrt{(x-2)^2+(y-2)^2}}$ scaled to have integral one. We prove continuous (normalized Shi-Malik) spectral clustering finds a cut with low surface area to volume ratio on mixtures of Laplace distributions. This gives intuition that Shi-Malik spectral clustering may find good clusters on data drawn from these densities.

---

[1]This ratio is known as the *isoperimetry* of the cut, and minimizing this quantity is known as the "Normalized Cut" [SM97, VL07] or "Sparsest Cut" [MS90, LR99, CGR05] problem.

[2]The Cheeger inequality guarantees that spectral cuts on a graph are a $1/\Phi$ approximation to optimal (where Φ is the minimum isoperimetry over all cuts). On geometric graphs like the n point line or the $\sqrt{n} \times \sqrt{n}$ grid, $1/\Phi$ diverges since the surface area scales more slowly with n than the volume. For large or infinite graphs, researchers needed graph *Buser* inequalities for tighter spectral cut guarantees [Tre21, BHL$^+$15, KKRT16].
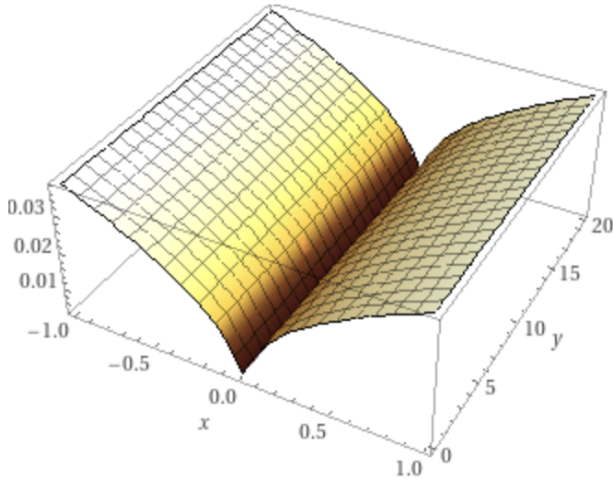
1

Figure 1.2: The 'square-root trough' density $\sqrt{|x| + 1/2^{20}}$ on $[-1, 1] \times [0, 20]$, scaled to have integral one. We show the continuous (normalized Shi-Malik) spectral clustering finds the suboptimal cut $y = 10$, when the cut $x = 0$ is much smaller. This gives intuition that Shi-Malik spectral clustering may find bad clusters on data drawn from these densities.

We prove that continuous (normalized Shi-Malik) spectral clustering works well on mixtures of Laplace distributions. This, combined with [TS18], gives theoretically informed intuition that normalized Shi-Malik spectral clustering works well on large data sets drawn from these densities – but it stops short of formally proving that Shi-Malik spectral clustering on a sample drawn from a density converges to an good cut on these densities as the number of samples grows large. We also show that continuous (normalized Shi-Malik) spectral clustering works poorly on densities similar to the 'square-root trough' density in Figure 1.2.

We prove the unweighted spectral cut on any function f has a good surface area to volume ratio (also known as *isoperimetry*), as long as the probability density does not vary by more than a constant factor on balls of radius 1. Examples of such functions include mixtures of Laplace distributions. Loosely speaking, a cut with good isoperimetry will cut through regions of low probability mass, while dividing the space into two regions of comparatively high probability mass. Some kind of condition on the function is necessary: in Appendix A, we show that unweighted spectral cuts have very bad isoperimetry on the function $f(x, y) = x + \frac{1}{2^K}$ supported on $[-1, 1] \times [0, K]$, where K is large.

Our core theorem proves that weighted spectral cuts have low weighted isoperimetry for all probability densities. This theorem implies the aforementioned results on unweighted spectral cuts. The weights are functions from each point x in the domain to the positive reals, chosen based on how large of a ball can be drawn centered at the point x so that the density at each point in the ball is within a constant factor of the density at any other point in the ball. Our paper gives the first spectral cuts with any isoperimetry guarantee, weighted or otherwise, on all densities: even discontinuous ones.

Our core tool is a new Cheeger [Che70] and Buser [Bus82] inequality on probability densities, which relates weighted isoperimetry of a probability density to its weighted fundamental eigenvalue. Weights are necessary: unweighted Buser inequalities are false for the function $\rho(x) = |x| + \epsilon$ on the

2

interval $[-1, 1]$ (see Section 6 of [CMWW20]). Our paper gives the first Buser-type inequality on all probability densities, including discontinuous ones. Past Buser inequalities on densities relied on log-concavity [KLS95, LV18a], Lipschitzness [CMWW20], or a bound on some analog of Ricci curvature for densities (one analog is based on the Hessian of the log density, see Theorem 1.4 in [Mil09]). Buser inequalities were previously known for manifolds [Bus82, Led04, Mil09] and graphs [KKRT16, BHL+15] with some notion of bounded curvature.

The core difficulty in our work is proving the density Buser inequality: the density Cheeger inequality is straightforward from existing proofs of the Cheeger inequality in manifold [Che70] and graph [AM84] settings. Using weights to ensure a correct Cheeger-Buser inequality is necessary even in the case of graphs [Chu97][3], and should not be considered a major drawback of our approach.

1.1 Road Map

For the remainder of Section 1, we provide definitions and give our main theorem: Theorem 1.6, which states weighted spectral cuts have good weighted isoperimetry. We also give our core tool, Theorem 1.7, a weighted Cheeger-Buser inequality on densities. In Section 1.3.1, we explain how our weighted spectral cut theorem proves our result on unweighted spectral cuts. We survey past work in Section 1.4.

In Section 2, we give a simple proof of the density Buser inequality in Theorem 1.7, in one dimension. In Section 3, we overview proof techniques for the d-dimensional density Buser inequality, which requires considerably heavier computation than our one dimensional proof. We give future directions in Section 4.

Appendix A shows that unweighted spectral cuts can have poor unweighted isoperimetry. Appendix B shows how to prove Theorem 1.6 from Theorem 1.7. Appendix C proves the Cheeger inequality in Theorem 1.7. Appendix D proves the Buser inequality in Theorem 1.7. In Appendix E, we present special cases of Theorem 1.7.

1.2 Definitions

Definition 1.1. *Let $m, c : \mathbb{R}^d \to \mathbb{R}$. The (m, c)-**isoperimetry** of a set $A \subset \mathbb{R}^d$ is defined as*

$$\Phi_{m,c}(A) := \frac{\int_{B(A)} c(x)dx}{\min\left(\int_A m(x)dx, \int_{\mathbb{R}^d - A} m(x)dx\right)}$$

where $B(A)$ is the boundary of A.

The **unweighted isoperimetry** of set A with respect to ρ is defined as $\Phi_{\rho,\rho}(A)$.[4]

Definition 1.2. *Let $m : \mathbb{R}^d \to \mathbb{R}$ and $c : \mathbb{R}^d \to \mathbb{R}$ be integrable functions. Define the (m, c) **isoperimetric constant** $\Phi_{m,c}$ as:*

$$\Phi_{m,c} := \inf_{A \subset \mathbb{R}^d} \Phi_{m,c}(A)$$

The **unweighted isoperimetric constant** on probability density function ρ is defined as $\Phi_{\rho,\rho}$

---

[3]Cheeger inequalities using normalized Laplacians set the weights of each vertex to be its weighted degree.

[4]The unweighted isoperimetry of A corresponds to the ρ-weighted surface area to volume ratio for the set A or $\mathbb{R}^d \backslash A$, depending which has smaller ρ-weighted mass.

Definition 1.3. *Let $m : \mathbb{R}^d \to \mathbb{R}$ and $s : \mathbb{R}^d \to \mathbb{R}$.*

*Define the (m, s) **Rayleigh quotient of** $u : \mathbb{R} \to \mathbb{R}$ to be:*

$$Q_{u,m,s} = \frac{\int_{\mathbb{R}^d} s(x)|\nabla u(x)|^2 dx}{\int_{\mathbb{R}^d} m(x)|u(x)|^2 dx}$$

*Define the (m, s) **fundamental eigenvalue** $\lambda_2^{m,s}$ as:*

$$\lambda_2^{m,s} := \inf_{\int m(x)u(x)dx=0} Q_{u,m,s}$$

*Define the (m, s) **fundamental eigenfunction** $u_2^{m,s}$ be:*

$$u_2^{m,s} := \arg\min_{\int m(x)u(x)dx=0} Q_{u,m,s}$$

This definition is a modeled off the definition of a fundamental eigenvalue of a graph Laplacian with edge and mass weights. The **unweighted fundamental eigenvalue** of probability density function ρ, is defined as $\lambda_2^{\rho,\rho}$, and the **unweighted fundamental eigenfunction** is $u_2^{\rho,\rho}$

Definition 1.4. *Let $m, s, c : \mathbb{R}^d \to \mathbb{R}$. The (m, s, c) **weighted spectral cut** is the set $S_{t'}$ defined as follows: first, find the (m, s) fundamental eigenfunction $u_2^{m,s}$. Next, define $S_t := \{x \in \mathbb{R}^d | u_2^{m,s}(x) \geq t\}$. Finally, set*

$$t' := \arg\min_t \Phi_{m,c}(S_t)$$

Note that the weighted spectral cut is defined as a subset of $\mathbb{R}^d$, similar to how a cut on a graph can be defined as a subset of the graph's vertices.

Definition 1.5. *The **unweighted spectral cut** of a probability density ρ is defined as the (ρ, ρ, ρ) weighted spectral cut.*

1.3 Main Theorems: Weighted Spectral Cuts and Density Cheeger-Buser

The core theorem of our paper is Theorem 1.6. Our core tool is Theorem 1.7. Proving Theorem 1.6 using our core tool is straightforward (Appendix B). The core technical contribution is the Buser inequality in Theorem 1.7.

Theorem 1.6. *(Weighted Spectral Cut) Let $\rho : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$, let C be a constant, and let $R : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ be an L-Lipschitz function such that $\frac{\rho(y)}{\rho(z)} \leq C$ for all y, z in a ball of radius $R(x)$ centered at x. Let A be the $(\rho, R^2\rho, R\rho)$ weighted spectral cut of ρ. Then:*

$$\Phi_{\rho,R\rho} \leq \Phi_{\rho,R\rho}(A) \leq O_C\left(\sqrt{d(L+1)\Phi_{\rho,R\rho}} + \sqrt{d} \cdot \Phi_{\rho,R\rho}\right)$$

This theorem shows that weighted spectral cuts have low weighted isoperimetry. This is the core theorem in our paper. We prove Theorem 1.6 using a new, weighted Cheeger/ Buser inequality:

Theorem 1.7. *(Weighted Cheeger/Buser Inequality for Probability Densities)*
Let $\rho : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ with $\int_{x \in \mathbb{R}^d} \rho(x) < \infty$. For any $R : \mathbb{R}^d \to \mathbb{R}$,

$$\frac{\Phi_{\rho,R\rho}^2}{4} \leq \lambda_2^{\rho,R^2\rho}$$

Suppose furthermore that R has the property that $\frac{1}{2}\rho(x) \le \rho(y) \le 2\rho(x)$ for any y in a ball of radius $R(x)$ around x, for some L-Lipschitz function R. Then:

$$\lambda_2^{\rho,R^2\rho} \le d \cdot O\left((L+1)\Phi_{\rho,R\rho} + \Phi_{\rho,R\rho}^2\right)$$

We call the first inequality the **weighted density Cheeger inequality**, and the second inequality the **weighted density Buser inequality**. Proving Theorem 1.6 from Theorem 1.7 is standard (see Appendix B). The proof for weighted density Cheeger is straightforward (see Appendix C). The core difficulty is proving the Buser inequality of Theorem 1.7, which is in Appendix D. A simplified proof of our Buser inequality in one dimension is given in Section 2.

For Theorems 1.6 and 1.7, the restriction that R is L-Lipschitz is not as restrictive as it seems: for any density ρ, setting $R(x)$ to be the *maximum* possible value so that ρ doesn't vary by more than a constant factor on balls of radius $R(x)$ centered at x, results in R being 1-Lipschitz. Thus, we can state corollaries that do not depend on L:

Corollary 1.8. *(Weighted Spectral Cut, no Lipschitz Condition) Let $\rho : \mathbb{R}^d \to \mathbb{R}^{\ge 0}$, and let $R : \mathbb{R}^d \to \mathbb{R}$ be defined such that for all x, $R(x)$ is the maximum value such that $\frac{\rho(y)}{\rho(z)} \le 2$ for all y, z in a ball of radius $R(x)$ centered at x. Let A be the $(\rho, R^2\rho, R\rho)$ weighted spectral cut of ρ. Then:*

$$\Phi_{\rho,R\rho} \le \Phi_{\rho,R\rho}(A) \le O\left(\sqrt{d\Phi_{\rho,R\rho}} + \sqrt{d} \cdot \Phi_{\rho,R\rho}\right)$$

Corollary 1.9. *(Weighted Buser, no Lipschitz Condition) Let $\rho : \mathbb{R}^d \to \mathbb{R}^{\ge 0}$ with $\int_{x \in \mathbb{R}^d} \rho(x) < \infty$. Let $R : \mathbb{R}^d \to \mathbb{R}$ be defined such that for all x, $R(x)$ is the maximum value such that $\frac{\rho(y)}{\rho(z)} \le 2$ for all y, z in a ball of radius $R(x)$ centered at x. Then*

$$\lambda_2^{\rho,R^2\rho} \le d \cdot O(\Phi_{\rho,R\rho} + \Phi_{\rho,R\rho}^2)$$

Proof. (of Corollaries 1.8 and 1.9) Observe that any ball of radius $R(x)$ centered at x cannot be contained strictly in the interior of a ball of radius $R(y)$ centered at y, by definition of R. Therefore, R is 1-Lipschitz, and our corollaries follow from Theorems 1.6 and 1.7. □

1.3.1 Theorems for Continuous Spectral Clustering

We state and prove our main result on unweighted spectral cuts of ρ (equivalently, continuous Shi-Malik spectral clustering on $\sqrt{\rho}$).

Corollary 1.10. *(Continuous Spectral Clustering / Unweighted Spectral Cut) Let $\rho : \mathbb{R}^d \to \mathbb{R}^{\ge 0}$, which does not vary by more than a constant multiple C on all balls of radius 1. Let A be the unweighted spectral cut of ρ. Then:*

$$\Phi_{\rho,\rho} \le \Phi_{\rho,\rho}(A) \le O_{C,d}(\sqrt{\Phi_{\rho,\rho}} + \Phi_{\rho,\rho})$$

Proof. This follows directly from Theorem 1.6 when R is constant. □

If ρ has the property described in the corollary, so does ρ^2. Thus, the (ρ^2, ρ^2, ρ^2) spectral cut has good (ρ^2, ρ^2) isoperimetry when ρ doesn't vary by more than a constant factor on balls of constant radius. Since (ρ^2, ρ^2, ρ^2) spectral cuts are a continuous form of spectral clustering on ρ [TS18] (see past work in Section 1.4.1), this corollary informs us about when continuous spectral clustering on ρ has a good (ρ^2, ρ^2) isoperimetry. We now state our unweighted Cheeger-Buser inequality on probability densities.

Corollary 1.11. *(Unweighted Cheeger-Buser) If $\rho : \mathbb{R}^d \to \mathbb{R}$ does not vary by more than a constant multiple C on balls of constant radius K, then*

$$\Omega(\Phi_{\rho,\rho}^2) \le \lambda_2^{\rho,\rho} \le O_{C,K}(d\Phi_{\rho,\rho} + \Phi_{\rho,\rho}^2)$$

Proof. This follows directly from Theorem 1.7 when R is constant. □

1.4 Past Work and Contributions

1.4.1 Spectral Clustering on Large Datasets and Continuous Clustering

Shi-Malik normalized spectral clustering [SM97] builds a geometric graph with data as vertices (using either the Gaussian kernel as edge weights, or by connecting the graph locally), and then performing a spectral sweep cut method based on Cheeger cut methods in spectral graph theory [SM97, NJW01, VL07, AM84, Chu97]. Researchers have studied the behavior of spectral clustering on large data sets drawn i.i.d from a probability density [VLBB08, TS18, TSVB+16], and have shown convergence properties of spectral clustering as the data set grows large.

Previous work established robust connections between widely used two-way spectral clustering algorithms (like the one in [SM97]), and spectral cuts applied to probability density [VLBB08, TS18, TSVB+16]. Loosely speaking, [TS18] shows that normalized Shi-Malik spectral clustering[5] on datasets drawn i.i.d from a probability density, behaves similarly to (ρ^2, ρ^2, ρ^2) weighted spectral cuts on the density when the number of samples grows large. They also show that unnormalized spectral clustering in the limit behaves similarly to (ρ, ρ^2, ρ^2) spectral cuts [6]. Corollary 1.11 show notions of isoperimetry guaranteed by normalized spectral cut methods. Our work provides theoretical justifications for normalized Shi-Malik spectral clustering on large datasets, for certain classes of densities.

Researchers have previously suggested that Shi-Malik clustering works due to the graph Cheeger inequality [BN04, MS22]. As shown in the introduction, this argument breaks down for large datasets: the quality of approximation given by the graph Cheeger and Buser inequalities gets arbitrarily bad as the number of points gets large [Tre21], which is undesirable in the large data setting. On some densities, spectral cuts can give cuts with undesirable isoperimetry that defy machine learning intuition. See Appendix A for more details. Other work on spectral clustering has shown its limitations on certain data sets: however, this past work focuses more on the limits of the "normalized cut" approach that spectral clustering is based on [NG06], whereas our work accepts that the normalized cut is a reasonable approach.

1.4.2 Cheeger and Buser Inequalities

Cheeger and Buser inequalities have been foundational to spectral graph theory [Chu97, AM84], manifold theory [Bus82, Led04, Mil09], and probability densities [LV18a, KLS95]. Our work significantly expands the scope of Buser inequalities for densities.

For Probability Densities: Past researchers have considered Buser-like inequalities on restricted probability densities, using unweighted isoperimetry and unweighted fundamental eigenvalues. Log concave densities allow for a tighter bound of $\Phi^2 = \Theta_d(\lambda_2)$ [KLS95, LV18a], where d is

---

[5]Normalized spectral clustering uses a normalized Laplacian of an intermediate geometric graph, while unnormalized spectral clustering uses an unnormalized Laplacian. See [VL07] for more information.

[6]These proofs currently rely on some initial assumptions about ρ - for more details, refer to [TS18]

the dimension of the density's domain. Log concave densities are relevant to volume estimation and sampling algorithms for convex bodies [DF89, LV18a], since the indicator function for any convex body is log-concave [LV18a]. This line of work is related to the KLS conjecture – a fundamental conjecture in convex geometry [KLS95, LV18b, Che20, JLV22]. While the bounds for log concave densities are tighter than the bounds we obtain, these densities are very restricted. For example, log concave densities cannot have two local maxima. While this is not an issue for convex body sampling, past results on this line are less applicable to spectral clustering.

The work of [CMWW20] shows a weighted Buser inequality relating (ρ, ρ^2) isoperimetry with the (ρ, ρ^3) fundamental eigenvalue for Lipschitz densities. This work suffers from a core drawback that both quantities are 0 for densities like the Gaussian or Laplace distribution. The spectral cut algorithm from their work, when applied to *any* mixture of two Gaussians in one dimension, results in cuts arbitrarily far out on the tail. This is because the (ρ, ρ^2) isoperimetry of cuts far out on the tail goes to 0. Our theorems overcome this barrier: it can be shown that the $(\rho, R\rho)$ isoperimetry of a Gaussian mixture is non-zero, when $R : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ is chosen so that ρ does not vary by more than a constant multiple on balls of radius $R(x)$ centered at x, and $R(x)$ is as large as possible.

The manifold Buser inequality (Theorem 1.4 in [Mil09], first shown in [Led04]) addresses the case of probability densities on manifolds, but is limited to density functions whose Hessian of the log density is bounded. In contrast, our work applies to all densities, including discontinuous ones.

For Graphs: The Cheeger inequality for graphs [AM84] states that up to constant factors, the fundamental eigenvalue of the normalized Laplacian is bounded below by the square of the graph isoperimetry[7], and bounded above by the graph isoperimetry. This inequality is the foundation of spectral graph theory [AM84, Chu97]. Insights from these inequalities have been applied to algorithmic tasks including multi-way graph clustering [LRTV12, LGT14], Laplacian system solving [ST04, KMP10, CKM+14], balanced cuts [OSV12, CGL+20], expander decomposition [ST04, SW19], maximum flow [CKM+11], sparsest cut [KLL+13, AM84, ST04], and more [GM95, OSVV08, OV11, KW16, KLL16, Sch18, KLT22]. For a deeper exposition on Cheeger's inequality and its proof, see [AM84, Chu97, ST04].

It is known that the upper bound of the Cheeger inequality graphs loses meaning as the number of vertices grows large [Tre21]. Researchers have established tighter upper bounds on the fundamental eigenvalue via *Buser* inequalities for large or infinite graphs, which require a bound on some discrete analog of Ricci curvature [Tre21, BHL+15, KKRT16].

Cheeger and Buser Inequality for Manifolds: The Cheeger [Che70] and Buser [Bus82] inequalities on manifolds are foundational. They were first discovered in manifolds, and only later were they applied to graph theory [AM84] and other settings. The Cheeger inequality applies to all compact manifolds, while the Buser inequality only applies to compact manifolds with Ricci curvature bounded from below.

Manifold Cheeger inequalities and Buser inequalities are useful to understand heat diffusion and mixing rates of random walks on manifolds [Bus82, Led04]. For more on Cheeger and Buser inequalities on manifolds, see [Led04].

---

[7]Isoperimetry in graphs is also known as sparsity or the normalized edge expansion

2 A simple proof of density Buser in one dimension

We give a simple proof of Theorem 1.7 in one dimension using the Hardy-Muckenhoupt inequality [Muc72, Sch18, MWW18]. The Hardy-Muckenhoupt Inequality can be viewed as a bound on the fundamental eigenvalue of a mass-weighted, edge-weighted graph Laplacian, from a three-way split of the graph. See [MWW18, Sch18] for details.
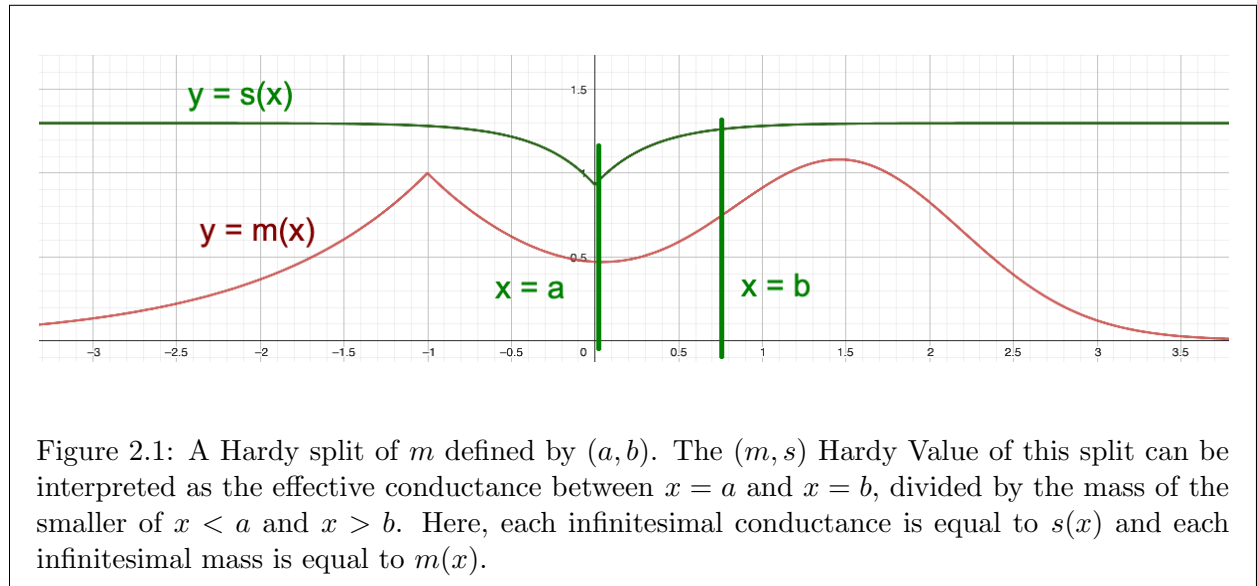
This section illustrates *why* we choose our weights in Theorem 1.7 the way that we do. Our choice of weights leads to a simple proof for Theorem 1.7 in one dimension. We conjecture (but do not prove) that the localization lemma [KLS95, LV17, JLV22] can generalize this proof to all higher dimensions. This paper's full proof of Theorem 1.7 in higher dimension uses a different set of tools, and requires heavier computation compared to the simple proof in this section.

2.1 Preliminaries: Hardy Splits and Inequality

Consider a one dimensional mass function $m : \mathbb{R} \to \mathbb{R}^{\geq}$, cut function $c : \mathbb{R} \to \mathbb{R}^{\geq 0}$, and spring function $s : \mathbb{R} \to R^{\geq 0}$, where $\int_{-\infty}^{\infty} m(x)dx < \infty$. Define a **Hardy Split** to be two reals (a, b) with $a \leq b$. We define the (m, s) **Hardy Value** of the Hardy Split (a, b) to be:

$$\frac{1/\int_a^b (dx/s(x))}{\min\left(\int_{-\infty}^a m(x)dx, \int_b^{\infty} m(x)dx\right)}$$

The work of [Sch18, MWW18] showed that this Hardy Value can be viewed in terms of effective and mass weights on graphs. Loosely speaking, the numerator of the Hardy Value can be seen as the effective conductance between the sets $\{x \leq a\}$ and $\{x \geq b\}$ where s is the infinitesimal conductance, and the denominator can be seen as the smaller of the masses of these two sets where m is the infinitesimal mass.

Figure 2.1: A Hardy split of m defined by (a, b). The (m, s) Hardy Value of this split can be interpreted as the effective conductance between $x = a$ and $x = b$, divided by the mass of the smaller of $x < a$ and $x > b$. Here, each infinitesimal conductance is equal to $s(x)$ and each infinitesimal mass is equal to $m(x)$.
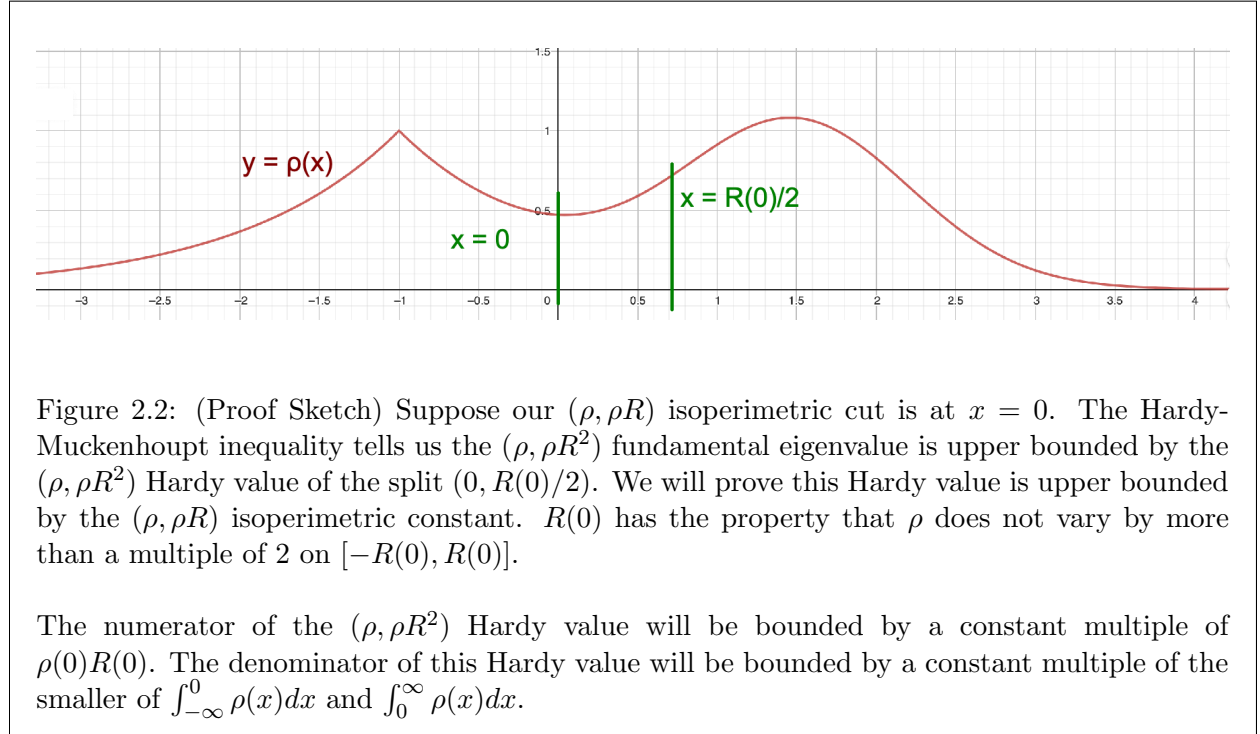
The **Hardy-Muckenhoupt Inequality** [Muc72, MWW18] states that, up to constant factors, the (m, s) fundamental eigenvalue is bounded above by the (m, s) Hardy value of any Hardy split,

and bounded below by the infimum (m, s) Hardy value over all Hardy splits. In our proof, we will only use the upper bound on the fundamental eigenvalue.[8]

2.2 Density Cheeger-Buser in one dimension

Proof. (of Theorem 1.7 in one dimension)

Let ρ be any one-dimensional L^1 function $\mathbb{R} \to \mathbb{R}$. Let $R : \mathbb{R} \to \mathbb{R}^{\geq 0}$ be a 1-Lipschitz function, where for all z, $\rho(x)/\rho(y)$ is bounded above by 2 for all x, y in balls of radius $R(z)$ centered at z.

Figure 2.2: (Proof Sketch) Suppose our $(\rho, \rho R)$ isoperimetric cut is at $x = 0$. The Hardy-Muckenhoupt inequality tells us the $(\rho, \rho R^2)$ fundamental eigenvalue is upper bounded by the $(\rho, \rho R^2)$ Hardy value of the split $(0, R(0)/2)$. We will prove this Hardy value is upper bounded by the $(\rho, \rho R)$ isoperimetric constant. $R(0)$ has the property that ρ does not vary by more than a multiple of 2 on $[-R(0), R(0)]$.

The numerator of the $(\rho, \rho R^2)$ Hardy value will be bounded by a constant multiple of $\rho(0)R(0)$. The denominator of this Hardy value will be bounded by a constant multiple of the smaller of $\int_{-\infty}^{0} \rho(x)dx$ and $\int_{0}^{\infty} \rho(x)dx$.

First, we prove the Buser inequality, by upper bounding the $(\rho, \rho R^2)$ fundamental eigenvalue by a $(\rho, \rho R^2)$ Hardy value of some Hardy split, and then upper bounding that Hardy value by the $(\rho, \rho R)$ isoperimetric constant. In one dimension, any isoperimetric cut can be represented by a single point x, where the two sides of the cut are $\{y : y \leq x\}$ and $\{y : y > x\}$. Without loss of generality, suppose the $(\rho, \rho R)$ isoperimetric cut is at $x = 0$.

The $(\rho, \rho R^2)$ fundamental eigenvalue is upper bounded by the $(\rho, \rho R^2)$ Hardy value for any Hardy split, by the Hardy-Muckenhoupt inequality. Without loss of generality, suppose that $\int_{-\infty}^{0} \rho(x) \leq \int_{0}^{\infty} \rho(x)$. Now consider the Hardy split defined by $(0, R(0)/2)$. The Hardy Value of the split equals:

$$\frac{1/\left(\int_{0}^{R(0)/2} \frac{dx}{\rho(x)R(x)^2}\right)}{\min\left(\int_{\infty}^{0} \rho(x)dx, \int_{R(0)/2}^{\infty} \rho(x)dx\right)} \tag{2.1}$$

---

[8]The upper bound is straightforward from simple bounds on the Rayleigh quotient formulation of the fundamental eigenvalue [Muc72]. The difficulty in proving the Hardy-Muckenhoupt inequality is in the lower bound, which we do not use in this proof.

The numerator of the Hardy value of this split is equal to:

$$\frac{1}{\int_0^{R(0)/2} \frac{dx}{\rho(x)R(x)^2}} \tag{2.2}$$

By the definition of R, we know that $\rho(x)$ doesn't change by more than a constant factor in the interval $[0, R(0)/2]$. Since R is 1-Lipschitz, $R(x)$ also doesn't change by more than a constant factor on this interval. Therefore, the value in Equation 2.2 is a constant multiple of

$$\frac{1}{\int_0^{R(0)/2} \frac{dx}{\rho(0)R(0)^2}}$$

which equals $\rho(0)R(0)/2$. Thus, the numerator of the Hardy Value is $\Theta(\rho(0)R(0))$.

Now we bound the denominator of the Hardy Value. We claim that $\int_{R(0)/2}^\infty \rho(x)dx = \Theta\left(\int_0^\infty \rho(x)dx\right)$. This is because $\int_{R(0)/2}^{R(0)} \rho(x)dx = \Theta\left(\int_0^{R(0)} \rho(x)dx\right)$, by the definition of $R(0)$. Putting together the numerator and denominator bound, the $(\rho, \rho R^2)$ Hardy value of the Hardy Split $(0, R(0)/2)$ is

$$\Theta\left(\frac{\rho(0)R(0)}{\min\left(\int_{-\infty}^0 \rho(x)dx, \int_0^\infty \rho(x)dx\right)}\right)$$

which is a constant multiple of the isoperimetry of the $(\rho, \rho R)$ cut at 0, which we assumed without loss of generality was optimal. Therefore, the $(\rho, \rho R^2)$ fundamental eigenvalue of any density is bounded above by a constant multiple of the $(\rho, \rho R)$ cut in one dimension, which proves the Buser inequality in Theorem 1.7 for one dimension. The Cheeger inequality in this case is a straightforward derivation from the original proof [Che70], and has been deferred to Appendix C.

□

3   Technique for Density Cheeger-Buser in d dimensions

We prove the Buser inequality in Theorem 1.7 by using mollification over balls of radius R. Consider an indicator function of the $(\rho, \rho R)$ isoperimetric cut: the indicator function is 1 on one side of the cut and 0 on the other. For each point x, we create a new function that is the weighted average of our indicator function in a ball of radius $R(x)$ around x (this is also known in the literature as 'mollification' on a ball of radius $R(x)$ around x [Fri44]). We will show the $(\rho, \rho R^2)$ Rayleigh quotient of this new function is bounded above by the weighted isoperimetry of the cut.

Although the idea is simple, the details for proving that this idea works are technical, and have been deferred to Appendix (the entire appendix is a proof of Theorem 1.7). These computational details are *not* standard: past proof techniques for Buser type inequalities generally involve mollification of balls of constant radius (or relied on heat kernel arguments with a constant specific heat capacity [Led04]), rather than with balls of varying radii. Mollifying with balls of varying radii introduces computational issues from extra terms not present when the radii are all the same [9], and much of our proof is devoted to bounding these extra terms. Despite the technical complications,

---

[9]For example, the integral of the function after mollification is *not* equal to the integral of the function before mollification when the molliifcation balls have varying radii, in contrast to the case when the radius does not vary.

we have organized our proof in the appendix in a clear and organized way, and we believe it is accessible to the interested reader with limited background on mollification.

This technical argument is similar to that presented in [CMWW20], but differs in a key way: their work only applied only to the case when ρ was L-Lipschitz, and their work can be interpreted as a strict corollary of ours when R is forced to be ρ/L (see Corollary E.1). Forcing R in this fashion leads to spectral cuts for *any* mixture of two one-dimensional Gaussians being arbitrarily far out on the tail, which is undesirable. Our work avoids such issue. Compared to [CMWW20], our proof is cleaner and disentangles ρ and R in a careful way.

We are allowed to have $\rho(x)$ equal to 0 for some x, which differs from nearly every other proof about spectral cuts [VLBB08, TSVB+16, GTS16, GTS15, TS18]. No non-empty ball of radius $R(x)$ around point x contains y where $\rho(y) = 0$. This is key: if ρ has a cut with 0 isoperimetry, it is necessary for our proof method that any ball of mollification does not 'cross' this 0 cut.

As mentioned earlier, our technique for the Cheeger component of Theorem 1.7 is straightforward from existing work, and can be found in Appendix C.

4   Conclusion and Open Questions

We give the first result on isoperimetry of spectral cuts, weighted or otherwise, on general probability densities. This informs us when continuous spectral clustering finds clusters with good isoperimetry, and when it doesn't. We list some future directions of work:

One Cheeger-Buser to Rule Them All.   Is there be a single, unified inequality for probability densities supported on manifolds, that covers all of: the Cheeger-Buser inequality in Theorem 1.7, the tighter Cheeger inequality for log concave densities (see Section 1.4), and the manifold Buser inequality [Bus82]? This setting is relevant to manifold learning [BN04, BN07], which assumes data can be modeled as samples from a density supported on a manifold.

A Simpler Proof of Density Buser using Hardy Inequalities.   In Section 2, we use the Hardy-Muckenhoupt inequality to show a Buser inequality (Theorem 1.7) on probability densities in one dimension. Can we use the Localization Lemma [KLS95] to generalize this proof to arbitrary dimension? This would simplify our full proof in Appendix D through D.6. Can the localization lemma improve dimension dependence?

Convergence of Spectral Clustering to Spectral Cuts.   Our paper combined with [TS18], gives intuition that spectral clustering on a large number of samples drawn from a mixture of Laplace distributions, converges to a cut of the mixture with low isoperimetry. However, this intuition is not yet a formal proof. Currently, the convergence proofs of [TS18] (linking spectral clustering on large datasets to unweighted spectral cuts) require that the density be of bounded support: can we do away with this assumption?

New Spectral Clustering Algorithms.   Our work suggests there exist weighted variants of spectral clustering algorithms on samples from a density, which approximate our weighted spectral cut when the number of samples grows large. Do these algorithms work well in practice? What are these algorithms, and can we prove they converge to weighted spectral cuts on the underlying density? What if we know something about the structure of the underlying density, such as knowing it is the mixture of Gaussians?

Manifold Buser *without* lower bounds on Ricci Curvature.   Buser's inequality on manifolds [Bus82] requires a lower bound on the Ricci curvature. This is necessary as long as the isoperimetry and eigenvalue of the manifold are unweighted. If one allows for weighted isoperimetry

and weighted eigenvalues, can we find a weighted Buser inequality that applies to all manifolds (possibly by including point-wise Ricci curvature into the weights), for which the manifold Buser inequality is a corollary?

4.1 Acknowledgements

References

[ACSS20]   Josh Alman, Timothy Chu, Aaron Schild, and Zhao Song. Algorithms and hardness for linear algebra on geometric graphs. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 541–552. IEEE, 2020.

[AM84]   N. Alon and V. D. Milman. Eigenvalues, expanders and superconcentrators. In *Proceedings of the 25th Annual Symposium on Foundations of Computer Science, 1984*, FOCS '84, pages 320–322, Washington, DC, USA, 1984. IEEE Computer Society.

[BGA20]   Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. Spectral clustering with graph neural networks for graph pooling. In *International Conference on Machine Learning (ICML)*, pages 874–883. PMLR, 2020.

[BHL$^+$15]   Frank Bauer, Paul Horn, Yong Lin, Gabor Lippner, Dan Mangoubi, and Shing-Tung Yau. Li-yau inequality on graphs. *Journal of Differential Geometry*, 99(3):359–405, 2015.

[BN04]   Mikhail Belkin and Partha Niyogi. Semi-supervised learning on riemannian manifolds. *Mach. Learn.*, 56(1-3):209–239, June 2004.

[BN05]   Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. In *International Conference on Computational Learning Theory*, pages 486–500. Springer, 2005.

[BN07]   Mikhail Belkin and Partha Niyogi. Convergence of laplacian eigenmaps. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 129–136, 2007.

[Bus82]   Peter Buser. A note on the isoperimetric constant. *Annales scientifiques de l'École Normale Supérieure*, Ser. 4, 15(2):213–230, 1982.

[CCLB21]   Romain Couillet, Florent Chatelain, and Nicolas Le Bihan. Two-way kernel matrix puncturing: towards resource-efficient pca and spectral clustering. In *International Conference on Machine Learning (ICML)*, pages 2156–2165. PMLR, 2021.

[CGL$^+$20]   Julia Chuzhoy, Yu Gao, Jason Li, Danupon Nanongkai, Richard Peng, and Thatchaphol Saranurak. A deterministic algorithm for balanced cut with applications to dynamic connectivity, flows, and beyond. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1158–1167. IEEE, 2020.

[CGR05]   Shuchi Chawla, Anupam Gupta, and Harald Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 102–111. SIAM, 2005.

[Che70]   Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. In *Problems in Analysis: A Symposium in Honor of Salomon Bochner*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.

[Che20]   Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the kls conjecture, 2020.

[Chu97]   F.R.K. Chung. *Spectral Graph Theory*, volume 92 of *Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[CKM$^+$11]   Paul Christiano, Jonathan A. Kelner, Aleksander Madry, Daniel A. Spielman, and Shang-Hua Teng. Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 273–282. ACM, 2011.

[CKM$^+$14]   Michael B. Cohen, Rasmus Kyng, Gary L. Miller, Jakub W. Pachocki, Richard Peng, Anup B. Rao, and Shen Chen Xu. Solving SDD linear systems in nearly $m\log^{1/2}n$ time. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 343–352. ACM, 2014.

[CMWW20]   Timothy Chu, Gary L. Miller, Noel J. Walkington, and Alex L. Wang. Weighted cheeger and buser inequalities, with applications to clustering and cutting probability densities. `https://arxiv.org/abs/2004.09589`. *CoRR*, abs/2004.09589, 2020.

[DF89]   M. Dyer and A. Frieze. A random polynomial time algorithm for approximating the volume of convex bodies. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 375–381, New York, NY, USA, 1989. ACM.

[Fri44]   Kurt Otto Friedrichs. The identity of weak and strong extensions of differential operators. *Transactions of the American Mathematical Society*, 55:132–151, 1944.

[GM95]   Stephen Guattery and Gary L. Miller. On the performance of the spectral graph partitioning methods. In *SODA'95*, pages 233–242. ACM-SIAM, 1995.

[GTS15]   Nicolás García Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

[GTS16]   Nicolás García Trillos and Dejan Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, 04 2016.

[JLV22]   Arun Jambulapati, Yin Tat Lee, and Santosh S. Vempala. A slightly improved bound for the KLS constant. *CoRR*, abs/2208.11644, 2022.

[KKRT16] Bo'az Klartag, Gady Kozma, Peter Ralli, and Prasad Tetali. Discrete curvature and abelian groups. *Canadian Journal of Mathematics*, 68(3):655–674, 2016.

[KL19] Ioannis Koutis and Huong Le. Spectral modification of graphs for improved spectral clustering. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

[KLL+13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 11–20. ACM, 2013.

[KLL16] Tsz Chiu Kwok, Lap Chi Lau, and Yin Tat Lee. Improved cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 1848–1861, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics.

[KLS95] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemama. *Discrete & Computational Geometry*, 13:541–559, 1995.

[KLT22] Tsz Chiu Kwok, Lap Chi Lau, and Kam Chuen Tung. Cheeger inequalities for vertex expansion and reweighted eigenvalues. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 366–377. IEEE, 2022.

[KMP10] Ioannis Koutis, Gary L. Miller, and Richard Peng. Approaching optimality for solving SDD linear systems. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 235–244. IEEE Computer Society, 2010.

[KSAM19] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In *International Conference on Machine Learning (ICML)*, pages 3458–3467. PMLR, 2019.

[KW16] Robert Krauthgamer and Tal Wagner. Cheeger-type approximation for sparsest *st*-cut. *ACM Trans. Algorithms*, 13(1):14:1–14:21, 2016.

[Led04] Michel Ledoux. Spectral gap, logarithmic Sobolev constant, and geometric bounds. *Surveys in differential geometry*, 9(1):219–240, 2004.

[LGT14] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM*, 61(6)(37), 2014.

[LLKL11] Mu Li, Xiao-Chen Lian, James T. Kwok, and Bao-Liang Lu. Time and space efficient spectral clustering via column sampling. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 2297–2304. IEEE Computer Society, 2011.

[LR99]   Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM). Prelim version in ACM STOC 1988*, 46(6):787–832, 1999.

[LRTV12]   Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, STOC 2012, pages 1131–1140. ACM, 2012.

[LUZ17]   Marc T Law, Raquel Urtasun, and Richard S Zemel. Deep spectral clustering learning. In *International conference on machine learning (ICML)*, pages 1985–1994. PMLR, 2017.

[LV17]   Y. T. Lee and S. S. Vempala. Eldan's stochastic localization and the kls hyperplane conjecture: An improved lower bound for expansion. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science*, FOCS 2017, pages 998–1007, Oct 2017.

[LV18a]   Yin Tat Lee and Santosh S. Vempala. The Kannan-Lovász-Simonovits conjecture, 2018.

[LV18b]   Yin Tat Lee and Santosh S. Vempala. Stochastic localization + stieltjes barrier = tight bound for log-Sobolev. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 1122–1129, New York, NY, USA, 2018. ACM.

[Mil09]   Emanuel Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Inventiones mathematicae*, 177(1):1–43, 2009.

[MKRLV22] Luke Melas-Kyriazi, Christian Rupprecht, Iro Laina, and Andrea Vedaldi. Deep spectral methods: A surprisingly strong baseline for unsupervised semantic segmentation and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8364–8375, 2022.

[MS90]   David W Matula and Farhad Shahrokhi. Sparsest cuts and bottlenecks in graphs. *Discrete Applied Mathematics*, 27(1-2):113–123, 1990.

[MS22]   Peter Macgregor and He Sun. A tighter analysis of spectral clustering, and beyond. In *International Conference on Machine Learning (ICML)*, pages 14717–14742. PMLR, 2022.

[Muc72]   Benjamin Muckenhoupt. Hardy's inequality with weights. *Studia Mathematica*, 44(1):31–38, 1972.

[MWW18]   Gary L. Miller, Noel J. Walkington, and Alex L. Wang. Hardy-muckenhoupt bounds for laplacian eigenvalues. *RANDOM/APPROX. http://arxiv.org/abs/1812.02841*, 2018.

[NG06]   Boaz Nadler and Meirav Galun. Fundamental limitations of spectral clustering. *Advances in neural information processing systems*, 19, 2006.

[NJW01]   Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, pages 849–856, Cambridge, MA, USA, 2001. MIT Press.

[OSV12]   Lorenzo Orecchia, Sushant Sachdeva, and Nisheeth K Vishnoi. Approximating the exponential, the lanczos method and an o (m)-time spectral algorithm for balanced separator. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing (STOC)*, pages 1141–1160, 2012.

[OSVV08]   Lorenzo Orecchia, Leonard J Schulman, Umesh V Vazirani, and Nisheeth K Vishnoi. On partitioning graphs via single commodity flows. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, STOC '08, pages 461–470. ACM, 2008.

[OV11]   Lorenzo Orecchia and Nisheeth K. Vishnoi. Towards an sdp-based approach to spectral methods: A nearly-linear-time algorithm for graph partitioning and decomposition. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 532–545, Philadelphia, PA, USA, 2011. Society for Industrial and Applied Mathematics.

[Sch18]   Aaron Schild. A schur complement cheeger inequality. In *ITCS '19*, 2018.

[SM97]   J.B. Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, 22:888–905, 01 1997.

[SSL+18]   Uri Shaham, Kelly Stanton, Henry Li, Boaz Nadler, Ronen Basri, and Yuval Kluger. Spectralnet: Spectral clustering using deep neural networks. *International Conference on Learning Representations (ICLR)*, 2018.

[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*, STOC '04, pages 81–90, New York, NY, USA, 2004. ACM.

[SW19]   Thatchaphol Saranurak and Di Wang. Expander decomposition and pruning: Faster, stronger, and simpler. 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 2616–2635. SIAM, 2019.

[Tre21]   Luca Trevisan. Buser inequalities in graphs. `https://lucatrevisan.wordpress.com/2021/10/07/buser-inequalities-in-graphs/`, 2021. Accessed: 2023-01-04.

[TS18]   Nicolas Garcia Trillos and Dejan Slepčev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239–281, 2018.

[TSVB+16]   Nicolás García Trillos, Dejan Slepčev, James Von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of cheeger and ratio graph cuts. *J. Mach. Learn. Res.*, 17(1):6268–6313, Jan 2016.

[VL07]   Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[VLBB08] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

[YHJ09]   Donghui Yan, Ling Huang, and Michael I Jordan. Fast approximate spectral clustering. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 907–916, 2009.

[ZR18]   Yilin Zhang and Karl Rohe. Understanding regularized spectral clustering via graph conductance. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

Appendix

A Density whose Unweighted Spectral Cut has poor Isoperimetry

In this section, we show that for any K, there exists a density function whose unweighted isoperimetric constant is $O(1/2^K)$, and whose unweighted spectral cut has unweighted isoperimetry $O(1/K)$.

Consider function ρ supported on $[-1, 1] \times [0, K]$ for constant K, and let:

$$\rho(x, y) = |x| + \frac{1}{2^K}$$

ρ is constructed to be the Cartesian product of $|x| + \frac{1}{2^K}$ from $[-1, 1]$, and function that is 1 on $[0, K]$. The unweighted fundamental eigenvalue of $|x| + \frac{1}{2^K}$ on $[-1, -1]$ is $O(1/K)$, by the Hardy-Muckenhoupt inequality (see Section 2). The unweighted fundamental eigenvalue of the function that is 1 on $[0, K]$ is $O(1/K^2)$.

By symmetry arguments, it can be shown that the (ρ, ρ) fundamental eigenfunction is either constant on the lines $x = C$ for all constants C, or else constant on the lines $y = C$ for all constants C. From this, we can deduce that for large K, the unweighted spectral cut is defined by $y \leq K/2$, which has unweighted isoperimetry $O(1/K)$. The isoperimetric constant is $O(1/2^K)$, which occurs for the cut defined by $x \leq 0$. Thus, the unweighted spectral cut can have unweighted isoperimetry that is exponentially worse than optimal. The computation is exactly the same when we scale $\sqrt{\rho}$ to have integral 1, so that $\sqrt{\rho}$ a true probability density function (continuous spectral clustering on probability density $\sqrt{\rho}$ is the same as unweighted spectral cut on ρ, see Section 1.4.1 and [TS18]).

This work combined with with [TS18] (see Section 1.4.1), gives evidence that two-way normalized spectral clustering may give undesirable clusters when applied to $\sqrt{\rho}$.

B Spectral Cut Theorems follow from Density Cheeger-Buse

We prove that Theorem 1.7 implies Theorem 1.6. This derivation is straightforward from past work, but we include it here for completeness.

Proof. (Of Theorem 1.6)

Let A be the $(\rho, \rho R^2, \rho R)$ weighted spectral cut, where ρ, R, L, C, d are defined as in Theorem 1.6. Then Lemma C.2 and the Buser inequality in Theorem 1.7 imply:

$$\Phi_{m,c}^2 \leq \Phi_{m,c}^2(A) \leq O(\lambda_2^{m,s}) \leq O(d(L+1)(\Phi_{\rho,\rho R}) + d\Phi_{\rho,\rho R}^2)$$

and Theorem 1.6 follows. □

C Proof of Cheeger Inequality on Densities

In this section, we prove the Cheeger inequality portion of Theorem 1.7. This inequality is straightforward from existing work like [Che70, AM84], but we include a derivation here for completeness.

Theorem C.1. *Let $m, c, s : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$. Then:*

$$\Phi m, c^2 \leq 4\|c/\sqrt{ms}\|_\infty^2 \lambda_2^{m,s}$$

18

Proof. Let $g : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a differentiable function, and let $g = u^2$. By Cauchy Schwarz:

$$\int_{\mathbb{R}^d} c(x)|\nabla g(x)|dx$$

$$= 2 \int_{\mathbb{R}^d} c(x)|u(x)||\nabla u(x)|dx$$

$$\leq 2\sqrt{\int_{\mathbb{R}^d} c(x)^2/m(x)|\nabla u(x)|^2 dx}\sqrt{\int_{\mathbb{R}^d} m(x)u(x)^2 dx}$$

$$\leq 2\|c/\sqrt{ms}\|_\infty\sqrt{\int_{\mathbb{R}^d} s(x)|\nabla u(x)|^2 dx}\sqrt{\int_{\mathbb{R}^d} m(x)u(x)^2 dx}$$

Dividing through by $\int_{\mathbb{R}^d} m(x)g(x)dx$, setting g as members of the sequence of differentiable functions that converge to the indicator function given by the isoperimetric cut, and squaring both sides gives us our desired result.

$\square$

When we set $m = \rho, c = \rho R, s = \rho R^2$ for any $\rho, R : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, the Cheeger inequality of Theorem 1.7 follows. We can prove a slightly stronger theorem with the same proof, which will be useful in proving Theorem 1.6 on weighted spectral cut isoperimetry.

Lemma C.2. *Let A be the (m, s, c) spectral sweep cut. Then:*

$$\Phi_{m,c}(A)^2 \leq 4\lambda_2^{m,s}$$

D   Proof of Buser Inequality on Densities

The remainder of the appendix is devoted to the technical proof of the Buser inequality in Theorem 1.7. The Cheeger portion of the inequality is proven in Appendix C.

- Section D.1 contains notation and conventions that we use throughout the Appendix.

- Section D.2 introduces two Lemmas to bound the numerator and denominator of a key Rayleigh quotient. We will show how to prove Buser's inequality from these two lemmas. The numerator lemma will be proven in Section D.4, and the denominator lemma will be proven in Section D.5.

- Section D.3 gives two tools we use to prove the numerator and denominator bounds.

- Section D.4 proves the numerator bound Lemma from Section D.2.

- Section D.5 proves the denominator bound Lemma from Section D.2.

D.1   Notation

In this section, we list notation to simplify our calculations.

- $u_\theta(x) := \int_{y \in \mathbb{R}^d} u(x - \theta R(x)y)\phi(y)dy$ for some function $R : \mathbb{R}^d \to R$ and some mollifier $\phi : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ supported on the unit ball centered at 0, which we notate as $B(0, 1)$. Here, we force $\int \phi(y) = 1$. In our write-up, θ is always a constant.

  - This is equivalent to taking the ball of radius $\theta R(x)$ around x, and finding the weighted average of u in that ball.
  - u_θ depends on both R and ϕ.

- $\overline{u}$ denotes the constant $\int_{\mathbb{R}^d} \rho(y)u(y)dy / \int_{\mathbb{R}^d} \rho(y)dy$, which is equal to the ρ-weighted average of u.

- $a \lesssim b$ will mean $a <= Cb$ for some constant C.

- $a \approx b$ will mean $cb \leq a \leq Cb$ for some positive constants c and C.

- $\nabla_x(f(x, y))$ means that $f(x, y)$ will be treated as a function in x with constant y, and the gradient will be taken accordingly.

- $\nabla u(f(x))$ means that the gradient of u will be applied to $f(x)$. This is in contrast with $\nabla(u(f(x)))$, where the gradient is applied to the function $u \circ f$

- The notation $\rho R(f(x))$ is equivalent to $\rho(f(x))R(f(x))$. This shorthand will be convenient in simplifying certain equations.

- $\|u\|_q = \left(\int \|u(y)\|^q dy \right)^{1/q}$

- $R(x)$ in this paper will generally have the property that ρ doesn't vary by more than a factor of 2 on the ball of radius $R(x)$ centered at x.

- The notation $(\rho R|\nabla u|_2)_\theta$ is equivalent to defining $f(x) := \rho(x)R(x)|\nabla u(x)|_2$ and computing $f_\theta(x)$.

- $J_x(f)(x')$ refers to the Jacobian of function $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, as a function of variable $x \in \mathbb{R}^d$, applied at point $x' \in \mathbb{R}^d$.

No serious attempts were made to optimize the non-dimensionally sensitive constants.

D.2   Rayleigh Quotient Bound

Lemma D.1. *(Buser Inequality)*

For all indicator functions u that is 1 on $A \subset \mathbb{R}^d$ and 0 elsewhere, $\theta < \min(1, \frac{1}{2L})$, ρ, and L-Lipschitz R where $\rho(y)R(y)$ varies by at most a factor of 4 for y in ball of radius $R(x)$ around x:

$$\lambda_2 \leq \frac{\|\rho R^2 \|\nabla u_\theta\|_2^2\|_1}{\|\rho(u_\theta - \overline{u_\theta})^2\|_1} \lesssim \frac{(dL + d/\theta)\Phi_u}{\frac{1}{4} - \theta\Phi_u} \tag{D.1}$$

where

$$\Phi_u := \frac{\|\rho|\nabla u|_1\|_1}{\|\rho(u - \overline{u})\|_1} \approx \frac{\partial_{\rho R}(A)}{\min(Vol_\rho(A), Vol_\rho(A^c))}$$

Setting $\theta = \min(1, 1/(2L), 1/(8\Phi_u))$ and set u to be the indicator function of A corresponding to the optimal $(\rho, \rho R)$ cut:

$$\lambda_2 \leq O(\max(\Phi(3dL + d), \Phi^2))$$

To prove this, we bound the numerator and denominator of the Rayleigh quotient in Equation D.1.

Lemma D.2. *(Numerator bound)*
For ρ and R where R is L-Lipschitz, ρR doesn't change by more than a factor of 4 on a ball of radius R, and any θ where $\theta L \leq 1/2$.

$$\|\rho R^2 |\nabla u_\theta|_2^2\|_1 \leq \frac{6d(2 + 3L)}{\theta} \|u\|_\infty \|\rho R |\nabla u|_2\|_1$$

Lemma D.3. *(Denominator Bound)*
For ρ and R where R is L-Lipschitz, ρR doesn't change by more than a factor of 4 on a ball of radius R, and any θ where $\theta L \leq 1/2$.

$$\|\rho(u_\theta - \overline{u_\theta})^2\|_1 \geq \frac{1}{4}\|\rho(u - \overline{u})\|_1 - 2\theta\|\rho R \nabla u\|_1$$

Proof. (of Lemma D.1) Lemma D.2 and D.3 tell us that, for indicator functions u (guaranteeing $\|u\|_\infty \leq 1$):

$$\frac{\|\rho R^2 \|\nabla u_\theta\|_2^2\|_1}{\|\rho(u_\theta - \overline{u_\theta})^2\|_1} \leq \frac{6d(2 + 3L)\|\rho R|\nabla u|_2\|_1}{\theta\left(\frac{1}{4}\|\rho(u - \overline{u})\|_1 - \theta\|\rho R|\nabla u|_2\|_1\right)}$$

Divide the numerator and denominator by $\|\rho(u - \overline{u})\|_1$, and Lemma D.1 follows. □

Now we prove Lemma D.2 and D.3.

D.2.1 Numerator Bound Proof

We prove the numerator bound in Lemma D.2 by introducing two lemmas, whose proofs we defer to Section D.4:

Lemma D.4. *Given any L-Lipschitz function $R : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ and any function u,*

$$\|R(x)|\nabla u_\theta(x)|_2\|_\infty \leq \left(\frac{O(d)}{\theta} + 2dL\right)\|u\|_\infty$$

Lemma D.5. *For ρ and L-Lipschitz R where ρR doesn't change by more than a factor of 4 on a ball of radius R, and any θ where $\theta L \leq 1/2$:*

$$\|\rho R|\nabla u_\theta|_2\|_1 \leq 6\|\rho R|\nabla u|_2\|_1$$

Proof. (of Lemma D.2, numerator bound) We know for all a, b that $\|ab\|_1 \leq \|a\|_\infty \|b\|_1$. Set $a := R|\nabla u_\theta|_2$ and $b := \rho R|\nabla u_\theta|_2$, and apply Lemmas D.4 and D.5. □

D.2.2 Denominator Bound Proof

Now we prove the denominator bound in Lemma D.3, by introducing two lemmas. The proof of the latter is long, and deferred to Section D.4.

Lemma D.6. *For all ρ and R (no restrictions) and indicator functions u*

$$\|\rho(u_\theta - \overline{u_\theta})^2\|_1 \geq \frac{1}{4}\|\rho(u - \overline{u})\|_1 - \|\rho(u_\theta - u)\|_1$$

Proof. (of Lemma D.6) We know by triangle inequality:

$$\|\sqrt{\rho}(u_\theta - \overline{u_\theta})\|_2 \geq \|\sqrt{\rho}(u - \overline{u})\|_2 - \|\sqrt{\rho}\left((u_\theta - \overline{u_\theta}) - (u - \overline{u})\right)\|_2$$

If $a \geq b - c$, then $a^2 \geq b^2 - 2bc + c^2 \geq b^2/2 - c^2$. Thus:

$$\|\sqrt{\rho}(u_\theta - \overline{u_\theta})\|_2^2 \geq \frac{1}{2}\|\sqrt{\rho}(u - \overline{u})\|_2^2 - \|\sqrt{\rho}\left((u_\theta - \overline{u_\theta}) - (u - \overline{u})\right)\|_2^2 \tag{D.2}$$

First, note that all functions f, $\int \rho(f - \overline{f})^2 \leq \int \rho f^2$. This follows from basic calculus: this can be seen by replacing $\overline{f}$ in the previous computation with a constant c and optimizing for c. In particular, this holds if $f = (u_\theta - u)$ and $\hat{f} = \overline{u_\theta - u}$ Second, note that $\|u_\theta - u\|_\infty \leq 1$ when u is an indicator function. Therefore:

$$\|\sqrt{\rho}\left((u_\theta - \overline{u_\theta}) - (u - \overline{u})\right)\|_2^2 \leq \|\sqrt{\rho}(u_\theta - u)\|_2^2 \leq \|\rho(u_\theta - u)\|_1\|u_\theta - u\|_\infty \leq \|\rho(u_\theta - u)\|_1 \tag{D.3}$$

Now let A be the set that u is an indicator function for. Note that $\int_A \rho(u - \overline{u}) = -\int_{A^C} \rho(u - \overline{u})$, which follows since $\int_A \rho(u - \overline{u}) = 0$ by the definition of $\overline{u}$. We note that on $A, u \geq \overline{u}$ and on $A^C, u \leq \overline{u}$, and thus $\int_A \rho(u - \overline{u}) = -\int_{A^C} \rho(u - \overline{u}) = \frac{1}{2}\int |\rho(u - \overline{u})|$. Now notice that $\int \rho(u - \overline{u})^2 = (1 - \overline{u})\int_A \rho(u - \overline{u}) - \overline{u}\int_{A^C} \rho(u - \overline{u})) = \frac{1}{2}\int |\rho(u - \overline{u})|$. Thus:

$$\frac{1}{2}\|\sqrt{\rho}(u - \overline{u})\|_2^2 = \frac{1}{4}\|\rho(u - \overline{u}\|_1 \tag{D.4}$$

Combining Inequalities D.2, D.3, and D.4 completes the proof of our lemma. □

Lemma D.7. *For ρ and R where R is L-Lipschitz, ρR doesn't change by more than a factor of 4 on a ball of radius R, and any θ where $\theta L \leq 1/2$,*

$$\|\rho(u_\theta - u)\|_1 \leq 2\theta\|\rho R \nabla u\|_1$$

Proof. (of Lemma D.3, denominator bound) This follows immediately from Lemma D.6 and D.7. □

Therefore, we have proven the Buser inequality in Lemma D.1, conditional on proofs for Lemma D.4, D.5, D.6, and D.7. See Section D.5 for these proofs.

D.3 Useful Tools

Lemma D.8. *(**Bounding the L_1 norm of a mollified function**) Let R be any L-Lipschitz function, and let $f : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ be a non-negative function. Then $\frac{1}{1+\theta L}\|f\|_1 \leq \|f_\theta\|_1 \leq \frac{1}{1-\theta L}\|f\|_1$ for any f when $\theta L < 1$.*

We note that in the case of constant R, the inequalities become equality. We defer the proof to Appendix D.6

Lemma D.9. *There exists a smooth mollifier $\phi(y)$ supported on the unit ball where $\int \phi(y) = 1$ and $\int_{y \in \mathbb{R}^d} |\nabla \phi(y)|_2 dy \leq 2d$.*

Proof. This fact about mollifiers is known in the folklore, and one such mollifier can be constructed by letting ϕ be smooth approximations of the indicator function on the unit ball, scaled so that the integral of ϕ is 1. $\square$

Lemma D.10. *For any differentiable radially symmetric $\phi : \mathbb{R}^d \to \mathbb{R}^{\geq 0}$ satisfying $\int \phi(y) = 1$ and ϕ decreasing radially outwards:*

$$\left| \int_{\mathbb{R}^d} f(y) \nabla \phi(y) dy \right| \leq O(d)\|f\|_\infty$$

for any $f : \mathbb{R}^d \to \mathbb{R}$.

Proof. This follows directly from Lemma D.9. $\square$

D.4 Proofs for Numerator Bound Lemmas

Proof. (Of Lemma D.4) Recall the multivariable integration by parts formula:

$$\int_V (\nabla f)g = \int_V (\nabla(fg) - f\nabla(g)) = \oint_S fg\hat{n} - \int_V f\nabla(g)$$

where S is the surface of V and $\hat{n}$ is the outward unit normal of S. The key to our proof is use of this formula. (Proof continued on next page).

$$R(x)|\nabla u_\theta(x)|_2 = R(x)\left|\nabla_x\left(\int_{y\in B(0,1)} u(x-\theta R(x)y)\phi(y)dy\right)\right|_2 \qquad \text{(def. of } u_\theta)$$

$$= R(x)\left|\int_{y\in B(0,1)}\nabla_x(u(x-\theta R(x)y))\phi(y)dy\right|_2 \qquad \text{(linearity of integral)}$$

$$= R(x)\left|\int_y (I-\theta\nabla R(x)\otimes y)\cdot\nabla u(x-\theta R(x)y)\phi(y)\right|_2 dy$$

$$\text{(Chain rule, using } J_x(x-\theta R(x)y) = I-\theta\nabla R(x)\otimes y)$$

$$= R(x)\left|\int_y (I-\theta\nabla R(x)\otimes y)\cdot\frac{\theta R(x)\nabla u(x-\theta R(x)y)}{\theta R(x)}\phi(y)\right|_2 dy$$

$$= R(x)\left|\int_y (I-\theta\nabla R(x)\otimes y)\cdot\frac{-\nabla_y(u(x-\theta R(x)y))}{\theta R(x)}\phi(y)\right|_2 dy$$

$$\text{(from chain rule: } \nabla_y(u(x-\theta R(x)y) = -\theta R(x)\nabla u(x-\theta R(x)y))$$

$$= \left|\int_y (I-\theta\nabla R(x)\otimes y)\cdot\frac{-\nabla_y u(x-\theta R(x)y)}{\theta}\phi(y)\right|_2 dy$$

$$= \left|\int_y (I-\theta\nabla R(x)\otimes y)\cdot\frac{u(x-\theta R(x)y)}{\theta}\nabla\phi(y)\right|_2$$

$$\text{(Integrate by parts, since } \phi(y) \text{ vanishes on the surface of } B(0,1))$$

$$= \left|\frac{\int_y u(x-\theta R(x))\nabla\phi(y)dy}{\theta} - \int_y(\nabla R(x)\otimes y)\cdot u(x-\theta R(x)y)\nabla\phi(y)dy\right|_2$$

$$\leq \frac{O(d)\|u\|_\infty}{\theta} + \left|\int_y(\nabla R(x)\otimes y)\cdot u(x-\theta R(x)y)\nabla\phi(y)dy\right|_2 \qquad \text{(By Lemma D.10)}$$

$$\leq \frac{O(d)\|u\|_\infty}{\theta} + L\int_y|u(x-\theta R(x)y)||\nabla\phi(y)|_2 dy|_2$$

$$\text{(When } \nabla\phi(y)\neq 0, \text{ then } |y|_2 \leq 1 \text{ and operator norm of } \nabla R(x)\otimes y \leq L)$$

$$\leq \left(\frac{O(d)}{\theta} + 2dL\right)\|u\|_\infty \qquad \text{(By Lemma D.9, bounding } \int_y|\nabla\phi(y)|_2 \text{ with } 2d).$$

$$\square$$

This completes our proof of Lemma D.4. Now we prove Lemma D.5 via a series of claims:

Claim D.11.
$$|\nabla u_\theta(x)|_2 \lesssim (|\nabla u|_2)_\theta(x)$$

(for L-Lipschitz function R, for all u and all x).

Claim D.12.
$$(\rho R|\nabla u_\theta|_2)(x) \lesssim (\rho R|\nabla u|_2)_\theta(x)$$

(for all x, where we have restrictions on ρR and θL.)

Claim D.13.
$$\|(\rho R|\nabla u|_2)_\theta\|_1 \lesssim \|\rho R|\nabla u|_2\|_1$$
(for all R with L-Lipschitz R and $\theta L \leq \frac{1}{2}$. No assumptions on ρ).

First, we show how to use these claims to prove our numerator bound. Then we provide full proofs of these claims.

Proof. (of Lemma D.5, numerator bound) Combine Claim D.12 and Claim D.13.[10].

<div style="text-align:right">□</div>

Proof. (of Claim D.11, used in proving Claim D.12)

$$|\nabla u_\theta(x)|_2$$
$$= \left|\nabla_x \int_y u(x - \theta R(x)y)\phi(y)dy\right|_2 \qquad \text{(Definition of } u_\theta)$$
$$= \left|\int_y \nabla_x(u(x - \theta R(x)y)\phi(y)dy\right|_2 \qquad \text{(Linearity of gradient)}$$
$$= \left|\int_y (I - \theta\nabla R(x) \otimes y) \cdot \nabla u(x - \theta R(x)y)\phi(y)dy\right|_2 \qquad \text{(Multivariable chain rule)}$$
$$\leq (1 + \theta L)\int_y |\nabla u(x - \theta R(x)y)|_2\,\phi(y)$$
$$\qquad (R \text{ is } L\text{-Lipschitz, and } |y|_2 \leq 1 \text{ when } \phi(y) \neq 0)$$
$$= (1 + \theta L)(|\nabla u|_2)_\theta$$

<div style="text-align:right">□</div>

Proof. (of Claim D.12)

$$(\rho R|\nabla u_\theta|_2)(x)$$
$$\lesssim \rho R(x)(|\nabla u|_2)_\theta(x) \qquad \text{(by Claim D.11)}$$
$$= \int_y \rho R(x)|(\nabla u)(x - \theta R(x)y)|_2\phi(y)dy$$
$$\leq \max_{y \in B(0,1)} \left(\frac{\rho R(x)}{\rho R(x - \theta Ry)}\right) \int_y \rho R(x - \theta Ry)|(\nabla u)(x - \theta R(x)y)|_2\phi(y)dy$$
$$= \max_{y \in B(0,1)} \left(\frac{\rho R(x)}{\rho R(x - \theta Ry)}\right) (\rho R|\nabla u|_2)_\theta(x)$$
$$\leq \frac{2}{1 - \theta L}(\rho R|\nabla u|_2)_\theta(x)$$

<div style="text-align:right">□</div>

Proof. This is a direct consequence of Lemma D.8 applied to the function $\rho R|\nabla u|_2$ □

---

[10] Note that the weight we attached to ∇u_θ was ρR in these claims, but we could have used any monomial in ρ and R (at the cost of some constant factors)

D.5 Proofs for Denominator Bound Lemmas

We prove Lemma D.7, which we've shown suffices to prove Lemma D.3. We do so by introducing an intermediate claim.

Claim D.14. *For L-Lipschitz R and $\rho(x)$ varying by no more than a factor of two in balls of radius $R(x)$ centered at R, given $\theta L \leq \frac{1}{2}$:*

$$\rho(x)(u_\theta(x) - u(x)) \leq 4\theta \int_0^1 (\rho R |\nabla u|_2)_{t\theta}(x) dt$$

Proof.

$$\rho(x)(u_\theta(x) - u(x))$$

$$= \rho \int_y (u(x - \theta R(x)y) - u(x))\phi(y)dy \qquad \text{(definition of } u_\theta)$$

$$= \theta\rho \int_y \int_{t \in [0,1]} -R(x)y \cdot \nabla u(x - t\theta R(x)y)\phi(y)dtdy$$

$$\text{(Chain rule and fundamental theorem of calculus)}$$

$$= \theta\rho \int_{t \in [0,1]} \int_y -R(x)y \cdot \nabla u(x - t\theta Ry)\phi(y)dydt \qquad \text{(Swap order of integration)}$$

$$= \theta\rho \int_{t \in [0,1]} \int_y -\frac{R(x)}{\rho R(x - t\theta Ry)}\rho R(x - t\theta Ry)\left(y \cdot \nabla u(x - t\theta Ry)\right)\phi(y)dydt$$

$$\leq \theta \max_{y \in B(0,1), t \in [0,1]} \frac{\rho R(x)}{\rho R(x - t\theta Ry)} \left| \int_{t \in [0,1]} \int_y \rho R(x - t\theta Ry)(y \cdot \nabla u(x - t\theta Ry))\phi(y)dydt \right|$$

$$\leq \theta \frac{2}{1 - \theta L} \left| \int_{t \in [0,1]} \int_y \rho R(x - t\theta Ry)(y \cdot \nabla u(x - t\theta Ry))\phi(y)dydt \right|$$

$$\text{(variation in } \rho \text{ is bounded in balls of radius } R, \text{ and } R \text{ is } L\text{-Lipschitz)}$$

$$\leq 4\theta \left| \int_{t \in [0,1]} \int_y \rho R(x - t\theta Ry)(y \cdot \nabla u(x - t\theta Ry))\phi(y)dydt \right| \qquad \text{(since } \theta L \leq 1/2)$$

$$\leq 4\theta \int_{t \in [0,1]} \int_y \rho R(x - t\theta Ry)|\nabla u(x - t\theta Ry)|_2 \phi(y)dydt \qquad \text{(since } |y|_2 \leq 1)$$

$$= 4\theta \int_{t \in [0,1]} (\rho R |\nabla u|_2)_{t\theta}(x)dt \qquad \text{(Definition of } f_{t\theta} \text{ when } f(x) := \rho(x)R(x)|\nabla u(x)|_2)$$

$\square$

Proof. (of Lemma D.7, key Lemma for proving the denominator bound)

$$\|\rho(u_\theta - u)\|_1$$

$$\leq \left\| 4\theta \int_0^1 (\rho R |\nabla u|_2)_{t\theta} \right\|_1 dt \qquad \text{(by Claim D.14)}$$

$$= 4\theta \int_0^1 \|(\rho R |\nabla u|_2)_{t\theta}\|_1 \, dt$$

$$\leq 4\theta \left(\int_0^1 \frac{1}{1 - t\theta L} dt \right) \|\rho R |\nabla u|_2\|_1 \qquad \text{(by Lemma D.8)}$$

$$\leq 4\theta \frac{1}{1 - \theta L} \|\rho R |\nabla u|_2\|_1$$

$$\leq 8\theta \|\rho R |\nabla u|_2\|_1 \qquad \text{(since } \theta L \leq 1/2)$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

D.6   Proofs for Useful Tools

Proof. (of Lemma D.8)

The change of variables formula in multivariable calculus, combined with the inverse function theorem, says that for invertible, surjective functions $g : \mathbb{R}^d \to \mathbb{R}^d$:

$$\int_{x \in \mathbb{R}^d} f(g(x)) dx = \int_{z \in \mathbb{R}^d} \frac{f(z)}{\det(J_x(g)(g^{-1}(z)))} dz \tag{D.5}$$

Now let $g(x) := g_y(x) := x - \theta R(x) y$ where y is a constant. Note that the Jacobian $J_x(g)(x) = I - \theta \nabla R(x) \otimes y$, and the determinant (by the Matrix determinant lemma) is $1 - y \cdot \theta \nabla R(x)$. We note that g is invertible due to the Jacobian determinant being non-zero when $\theta L < 1$, and it is not difficult to see g must be surjective.

Set $z = x - \theta R(x) y$.

$$\|f_\theta\|_1 \tag{D.6}$$

$$= \int_{x \in \mathbb{R}^d} \int_{y \in \mathbb{R}^d} f(x - \theta R(x) y) \phi(y) dx dy \tag{D.7}$$

$$= \int_{y \in \mathbb{R}^d} \phi(y) \int_{x \in \mathbb{R}^d} f(x - \theta R(x) y) dx dy \tag{D.8}$$

$$= \int_{y \in \mathbb{R}^d} \phi(y) \int_{z \in \mathbb{R}^d} \frac{f(z)}{1 - y \cdot \theta \nabla R(g_y^{-1}(z))} \tag{D.9}$$

$$\text{(Multivariable Change of variables and inverse function theorem)} \tag{D.10}$$

$$= \int_{z \in \mathbb{R}^d} f(z) \int_{y \in \mathbb{R}^d} \frac{\phi(y)}{1 - y \cdot \theta \nabla R(g_y^{-1}(z))} \tag{D.11}$$

which is bounded above and below by $\frac{1}{1+\theta L} \|f\|_1$ and $\frac{1}{1-\theta L} \|f\|_1$, since:

- $\int_{y \in \mathbb{R}^d} \phi(y) = 1$

- $|y|_2 \leq 1$ when $\phi(y) \neq 0$, and $|\nabla R(x)| \leq L$

- $\theta L < 1$

by our initial assumptions on $\phi, R, L,$ and θ. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

E   Special Cases of Weighted Spectral Cuts

Corollary E.1. *If $\rho : \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz function, then*

$$\Omega(\Phi^2_{\rho,\rho^2}) \leq \lambda_2^{\rho,\rho^3} \leq O(d\Phi_{\rho,\rho^2} + \Phi^2_{\rho,\rho^2})$$

Corollary E.1 is the main result of [CMWW20] (Theorem 1.5 in that paper). But, it is a simple consequence of our Theorem 1.7, implying our new theorem is far more general.

Corollary E.2. *If ρ is the square of a Lipschitz function, then*

$$\Omega(\Phi^2_{\rho,\rho^{3/2}}) \leq \lambda_2^{\rho,\rho^2} \leq O(d\Phi_{\rho,\rho^{3/2}} + \Phi^2_{\rho,\rho^{3/2}})$$

The eigenvalue λ_2^{ρ,ρ^2} (in Corollary E.2) appears when studying unnormalized spectral clustering, when a large number of datapoints are drawn from a density ρ [VLBB08, TS18, GTS15]. This corollary shows that the $(\rho, \rho^2, \rho^{3/2})$ spectral cut of the probability density has good $(\rho, \rho^{3/2})$ isoperimetry. This is almost, but not quite the same, as the spectral cut corresponding to un-normalized spectral clustering (which does a (ρ, ρ^2, ρ^2) cut) [TS18]. This theorem suggests that unnormalized spectral clustering is theoretically un-sound, and a new clustering algorithm can be built which converges to the $(\rho, \rho^2, \rho^{3/2})$ cut of the density.

Corollary E.3. *(Weighted Cheeger-Buser for Mixtures of Gaussians) Let ρ be the mixture of a finite number of Gaussians in d dimensions with variance bounded away from 0.*
    Let $R(x) = \max(1, 1/\sqrt{-\log \rho(x)})$. Then:

$$\Omega(\Phi^2_{\rho,\rho R}) \leq \lambda_2^{\rho,\rho R^2} \leq O(d\Phi_{\rho,\rho R} + \Phi^2_{\rho,\rho R})$$

We note that Corollary E.3 is not true if we instead set $R(x) := C$ for any constant C. A counterexample can be found by taking two unit variance Gaussians in one dimension, and spacing them arbitrarily far apart. This corollary suggests that a variant of spectral clustering may be more effective on mixtures of Gaussians, compared to standard methods in use today.