The Normalized Graph Cut and Cheeger Constant: from Discrete to Continuous

Ery Arias-Castro, Bruno Pelletier, Pierre Pudlo

The Normalized Graph Cut and Cheeger Constant: from Discrete to Continuous

Ery Arias-Castro[*], Bruno Pelletier[†] and Pierre Pudlo[‡]

June 9, 2011

Abstract. Let M be a bounded domain of $\mathbb{R}^d$ with smooth boundary. We relate the Cheeger constant of M and the conductance of a neighborhood graph defined on a random sample from M. By restricting the minimization defining the latter over a particular class of subsets, we obtain consistency (after normalization) as the sample size increases, and show that any minimizing sequence of subsets has a subsequence converging to a Cheeger set of M.

Index Terms: Cheeger isoperimetric constant of a manifold, conductance of a graph, neighborhood graph, spectral clustering, U-processes, empirical processes.

AMS 2000 Classification: 62G05, 62G20.

1 Introduction and main results

The Cheeger isoperimetric constant may be defined for a Euclidean domain as well as for a graph. In either case it quantifies how well the set can be bisected or 'cut' into two pieces that are as little connected as possible. Motivated by recent developments in spectral clustering and computational geometry, we relate the Cheeger constant of a neighborhood graph defined on a sample from a domain and the Cheeger constant of the domain itself.

Given a graph G with weights $\{\delta_{ij}\}$, the *normalized cut of* a subset $S \subset G$ is defined as

$$h(S; G) = \frac{\sigma(S)}{\min\{\delta(S), \delta(S^c)\}}, \tag{1.1}$$

where S^c denotes the complement of S in G, and

$$\delta(S) = \sum_{i \in S} \sum_{j \neq i} \delta_{ij}, \quad \sigma(S) = \sum_{i \in S} \sum_{j \in S^c} \delta_{ij}, \tag{1.2}$$

[*]Department of Mathematics, University of California, San Diego, USA
[†]Département de Mathématiques, IRMAR – UMR CNRS 6625, Université Rennes II, France
[‡]Département de Mathématiques, I3M – UMR CNRS 5149, Université Montpellier II, France

are the discrete volume and perimeter of S. The *Cheeger constant* or *conductance* of the graph G is defined as the value of the optimal normalized cut over all non-empty subsets of G, i.e.

$$H(G) = \min\{h(S; G) : S \subset G, S \neq \emptyset\}. \tag{1.3}$$

A corresponding quantity can be defined for a domain of a Euclidean space. Let M be a bounded domain (i.e. open, connected subset) of $\mathbb{R}^d$ with smooth boundary ∂M. For an integer $1 \leq k \leq d$, let Vol_k denote the k-dimensional volume (Hausdorff measure) in $\mathbb{R}^d$. For an open subset $A \subset \mathbb{R}^d$, define its normalized cut with respect to M by

$$h(A; M) = \frac{\mathrm{Vol}_{d-1}(\partial A \cap M)}{\min\{\mathrm{Vol}_d(A \cap M), \mathrm{Vol}_d(A^c \cap M)\}},$$

where A^c denotes the complement of A in $\mathbb{R}^d$ and with the convention that $0/0 = \infty$. The Cheeger (isoperimetric) constant of M is defined as

$$H(M) = \inf\{h(A; M) : A \subset M\}.$$

Equivalently, the infimum may be restricted to all open subsets A of M such that $\partial A \cap M$ is a smooth submanifold of co-dimension 1. This quantity was introduced by Cheeger [15] in order to bound the eigengap of the spectrum of the Laplacian on a manifold. A Cheeger set is a subset $A \subset M$ such that $h(A; M) = H(M)$; there is always a Cheeger set and it is unique under some conditions on the domain M [12]. For $A \subset M$, we call $\partial A \cap M$ its relative boundary.

1.1 Consistency of the normalized cut

Suppose that we observe an i.i.d. random sample $\mathcal{X}_n = (X_1, \ldots, X_n)$ from the uniform distribution μ on M. For $r > 0$, let $G_{n,r}$ be the graph with nodes the sample points and edge weights $\delta_{ij} = \mathbf{1}\{\|X_i - X_j\| \leq r\}$, which is an instance of a random geometric graph [33]. Let ω_d denote the d-volume of the unit d-dimensional ball, and define

$$\gamma_d = \int_{\mathbb{R}^d} \max(\langle u, z\rangle, 0)\, \mathbf{1}\{\|z\| \leq 1\}\, \mathrm{d}z, \tag{1.4}$$

where u is any unit-norm vector of $\mathbb{R}^d$. Actually γ_d is the average volume of a spherical cap when the height is chosen uniformly at random. We establish the pointwise consistency of the normalized cut, which yields an asymptotic upper bound on the Cheeger constant of the neighborhood graph based on the Cheeger constant of the manifold. This is the first result we know of that relates these two quantities.

Theorem 1. *Let A be a fixed subset of M with smooth relative boundary. Fix a sequence $r_n \to 0$ with $nr_n^{d+1}/\log n \to +\infty$, and let $S_n = A \cap G_{n,r_n}$. Then with probability one*

$$\frac{\omega_d}{\gamma_d r_n} h(S_n; G_{n,r_n}) \to h(A; M),$$

and, consequently,

$$\limsup_{n\to\infty} \frac{\omega_d}{\gamma_d r_n} H(G_{n,r_n}) \leq H(M).$$

We do not know whether the Cheeger constant of the neighborhood graph, for an appropriate choice of the connectivity radius and properly normalized, converges to the Cheeger constant of the domain.

1.2 Consistent estimation of the Cheeger constant and Cheeger sets

We obtain a consistent estimator of the Cheeger constant $H(M)$ by restricting the minimization defining the conductance of the neighborhood graph (1.3) to subsets associated with subsets of $\mathbb{R}^d$ with controlled reach. The reach of a subset $S \subset \mathbb{R}^d$ [20], denoted reach(S), is the supremum over $\eta > 0$ such that, for each x within distance η of S, there is a unique point in S that is closest to x. We assume here that $M \subset (0,1)^d$. When this is not known and/or not the case, we may always infer a hypercube that contains M—by taking a hypercube containing all the data points, with some lee-way so that the hypercube contains M with high probability when the sample gets large—and then rescale and translate the points so that M is within the unit hypercube. So this assumption is really without loss of generality.

Theorem 2. *Assume that $M \subset (0,1)^d$ and that $r_n \to 0$ such that $nr_n^{2d+1} \to \infty$. Let $\rho_n \to 0$ slowly so that $r_n = o(\rho_n^\alpha)$ and $nr_n^{2d+1}\rho_n^\alpha \to \infty$ for all $\alpha > 0$. Let $\mathcal{R}_n$ be a class of open subsets $R \subset (0,1)^d$ such that* reach$(\partial R) \geq \rho_n$. *Define the functional $h_n^\ddagger$ over $\mathcal{R}_n$ by*

$$h_n^\ddagger(R) = \frac{\omega_d}{\gamma_d r_n} h\left(R \cap \mathcal{X}_n; G_{n,r_n}\right)$$

if both R and R^c contain a ball of radius ρ_n centered at a sample point, and $h_n^\ddagger(R) = \infty$ otherwise.

(*i*) *With probability one,*

$$\min_{R \in \mathcal{R}_n} h_n^\ddagger(R) \to H(M), \quad n \to \infty.$$

(*ii*) *Let $\{R_n\}$ be a sequence satisfying*

$$R_n \in \mathcal{R}_n, \qquad h_n^\ddagger(R_n) = \min\{h_n^\ddagger(R) : R \in \mathcal{R}_n\}. \tag{1.5}$$

Then with probability one, $\{R_n \cap M\}$ admits a subsequence converging in the L^1-metric. Moreover, any subsequence of $\{R_n \cap M\}$ converging in the L^1-metric converges to a Cheeger set of M.

Note that the infimum defining R_n in (1.5) is attained in $\mathcal{R}_n$ since the function $h_n^\ddagger$ takes only a finite number of values.

Part *(ii)* of Theorem 2 hints at a consistent estimate of a Cheeger set of M, but $R_n \cap M$ depends on M, which is unknown. On the other hand reconstructing an unknown set from a random sample of it is an independent problem for which there exists multiple techniques and an important literature—see e.g., [6] and the references therein. In the following result we construct a random discrete measure which does not require the knowledge of M, and prove that, seen as a sequence of random measures indexed by the sample size n, any accumulation point is the uniform measure on a Cheeger set of M.

Theorem 3. *Let $\{R_n\}$ be a sequence as in Theorem 2-(ii), $\{R_{n_k}\}$ a subsequence of $\{R_n\}$ with $R_{n_k} \cap M \to A_\infty$ in L^1. Define the random discrete measure $Q_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{R_n}(X_i)\delta_{X_i}$ and the measure $Q = \mathbf{1}_{A_\infty}(.)\mu$. Then, that Q_n converges weakly to Q is an event which holds with probability one.*

As an example of an estimate of a Cheeger set of M, one can consider a union of balls of radius κ_n centered at the observations falling in R_n. Under appropriate conditions, it is known that this estimate converges in L^1; see [6].

Let us mention that with our result, only the "regular" part of a Cheeger set can be reconstructed. Indeed, in dimension $d \geq 8$, the boundary of a Cheeger set is not necessarily regular and may contain parts of codimension greater than 1.

1.3 Connections to the literature

Our results relating the respective Cheeger constants of a domain and of a neighborhood graph defined from a sample from the domain are the first of their kind, as far as we know. The connections to the literature stem from the concept of normalized cut taking a central place in graph partitioning and related methods in clustering; from a recent trend in computational geometry (and topology) aiming at estimating geometrical (and topological) attributes of a set based on a sample; and from the fact that we can use the conductance to bound the mixing time of a random walk on the neighborhood graph.

Clustering. In spectral graph partitioning, the goal is to partition a graph G into subgraphs based on the eigenvalues and eigenvectors of the Laplacian [36, 16]. It arises as a convex relaxation of the combinatorial search of finding an optimal bisection in terms of the normalized cut. Given a set of points $X_1, \ldots, X_n$ and a dissimilarity measure (or kernel) ϕ, spectral clustering applies spectral graph partitioning to the graph with nodes the data points and edge weight $\delta_{ij} = \phi(X_i, X_j)$ between X_i and X_j [37]. For instance, if the points are embedded in a Euclidean space, the kernel ϕ is often of the form $\phi(x, y) = \psi(\|x - y\|/\sigma)$, where σ is a tuning parameter, and ψ is, e.g., the Gaussian kernel $\psi(t) = \exp(-t^2)$ or the simple kernel $\psi(t) = \mathbf{1}_{[0,1]}(t)$ [30, 3]. The consistency of spectral methods has been analyzed in this context [38, 32, 4, 21, 35]. In particular, [28] proves a result similar to our Theorem 1 in that context.

About cuts, [27] also proves a result similar to our Theorem 1 when the separating surface ∂A is an affine hyperplane. Closer to our Theorem 2, [29] establishes rates for learning a cut for classification purposes—so the setting there is that of supervised learning, with each sample point X_i associated with a class label Y_i.

Computational geometry (and topology). The Cheeger constant $H(M)$, and Cheeger sets, are *bona fide* geometric characteristics of the domain M that we might want to estimate, following a fast developing line of research around the estimation of some geometric and topological characteristics of sets from a sample, e.g., the number of connected components [5], the intrinsic dimensionality [26] and, more generally, the homology [31, 10, 11, 41, 14, 34, 13]; the Minkowski content [17], as well as the perimeter and area (volume) [8].

Random walks. Random geometric graphs are gaining popularity as models for real-life networks. Some protocols for passing information between nodes amounts to performing a random walk and it is important to bound the time it takes for information to spread to the whole network;

see [2] and references therein. It is well-known that, given a graph G, a lower bound on $H(G)$ may be used to bound the mixing time the random walk on G. This is the path taken in [7, 2] when M is the unit hypercube and the graph is $G_{r_n,n}$. However, in both papers the authors reduce the setting to that of a regular grid without rigorous justification, leaving the problem unresolved (in our opinion) even in this particular case.

1.4 Discussion

As we saw, there are only a handful of other papers relating cuts in neighborhood graphs and cuts in the corresponding domain from which the points making the neighborhood graph where sampled from. Our paper is the first one we know of that establishes a relationship between the Cheeger constant (optimal normalized cut) on the neighborhood graph and the Cheeger constant of the domain, and the first one to propose a method that is consistent for the estimation of the latter based on a restricted normalized cut, and also consistent for the estimation of Cheeger sets. Our results generalize with varying amount of effort to other related settings. However, we leave important questions behind.

Generalizations. With some additional work, our results and methodology extend to settings where the kernel (here the simple kernel) is fast decaying and where the data points are sampled from a probability distribution on M that has a non-vanishing density with respect to the uniform distribution. It would also be interesting to consider the setting where M is a d-dimensional smooth submanifold embedded in some Euclidean ambient space. Our arguments seem to carry through using a set of charts for the manifold M, as is done in [9, Lem. 3.4].

Refinements. Though we focused on sufficient conditions for r_n to enable a consistent estimation of the Cheeger constant of the domain, it may also be of interest to find necessary conditions. Partial work suggests that $nr_n^d \to \infty$ is necessary, and may be sufficient the divergence to infinity is faster than a sufficiently large power of $\log n$. The arguments in support of this, however, are substantially different than those we use in the paper, which hinge on Hoeffding's inequality for U-statistics.

An open problem. Whether the normalized Cheeger constants of some sequence of neighborhood graphs converges to the Cheeger constant of the domain is an intriguing question. To paraphrase the question we leave open, is there a sequence $\{r_n\}$ such that, with probability one,

$$\lim_{n\to\infty} \frac{\omega_d}{\gamma_d r_n} H(G_{n,r_n}) = H(M)?$$

A positive answer would establish the consistency of the normalized cut criterion for graph partitioning. Also, a lower bound on $H(G_{n,r_n})$ would provide a lower bound on the eigengap between the first and second eigenvalue of the Laplacian, which in turn may be used to bound the mixing time of the random walk on G_{n,r_n}, as done in [7, 2] when M is the unit hypercube.

Consistent estimation in polynomial time. Our estimation procedures, though theoretically valid and consistent, are not practical. It would be interesting to know whether there is a consistent estimator for the Cheeger constant that can be implemented in polynomial-time. Note that computing the Cheeger constant of a graph is NP-hard (which motivates the use of spectral methods),

and even the best polynomial-time approximations we are aware of are not precise enough to allow for consistency [1].

1.5 Content

The rest of the paper is devoted to the proofs of the three theorems. In Section 2, we establish the convergence of the discrete volume and perimeter to their continuous counterparts of a fixed subset of M with smooth relative boundary, using Hoeffding's inequality for U-statistics [24]. Then, by the lower semi-continuity of the map $A \mapsto h(A; M)$, we deduce the supremum-limit bound of Theorem 1. In Section 3, we prove Theorems 2 and 3 by utilizing results on empirical U-processes [18] on the one hand, and compactness properties of the L^1-metric [23] on the other hand.

1.6 Notation and background

The uniform measure on M is denoted μ, so that $\mu(A) = \mathrm{Vol}_d(A \cap M)/\mathrm{Vol}_d(M)$; and the normalized perimeter is denoted $\nu(A) = \mathrm{Vol}_{d-1}(\partial A \cap M)/\mathrm{Vol}_d(M)$. Let $\tau_M = \mathrm{Vol}_d(M)$, and define the discrete volume and perimeters as

$$\mu_n(A) = \frac{\tau_M}{\omega_d n(n-1) r_n^d} \delta(A \cap \mathcal{X}_n; G_{n,r_n}), \quad \nu_n(A) = \frac{\tau_M}{\gamma_d n(n-1) r_n^{d+1}} \sigma(A \cap \mathcal{X}_n; G_{n,r_n}), \tag{1.6}$$

where δ, σ are given in (1.2), $\mathcal{X}_n$ is the sample, and G_{n,r_n} the neighborhood graph. Also, define the discrete ratio

$$h_n(A) = \frac{\nu_n(A)}{\min(\mu_n(A), \mu_n(A^c))},$$

and note that

$$h_n(A) = \frac{\omega_d}{\gamma_d r_n} h(A \cap \mathcal{X}_n; G_{n,r_n}),$$

where h is given in (1.1). For further reference, we define the volume $\pi_d(\eta)$ of a spherical cap at height η by

$$\pi_d(\eta) = \mathrm{Vol}_d \{x : \|x\| \leq 1 \text{ and } \langle u, x \rangle \geq \eta\},$$

where u is any unit-norm vector of $\mathbb{R}^d$. Note that the constant γ_d defined in (1.4) may be expressed as

$$\gamma_d = \int_0^1 \pi_d(\eta) \mathrm{d}\eta.$$

The reach coincides with the condition number introduced in [31] for submanifolds without boundary, and the property $\mathrm{reach}(\partial A) \geq r$ is equivalent to A and A^c being both r-convex [39], in the sense that a ball of radius r rolls freely inside A and A^c. (We say that a ball of radius r rolls freely in A if, for all $p \in \partial A$, there is $x \in A$ such that $p \in \partial B(x, r)$ and $B(x, r) \subset A$.) It is well-known that the reach bounds the radius of curvature from below [20, Thm. 4.18]. In particular, if $\mathrm{reach}(\partial A) > 0$, then ∂A is a smooth submanifold (possibly with boundary).

In the rest of the paper, the generic constant C may vary from line to line, except when stated explicitly otherwise.

2 Proof of Theorem 1: Consistency of the normalized cut

For a subset A of M and a real number $r > 0$, define the symmetric kernel

$$\phi_{A,r}(x,y) = \frac{1}{2}\Big\{\mathbf{1}_A(x) + \mathbf{1}_A(y)\Big\}\mathbf{1}\{\|x-y\| \le r\}, \tag{2.1}$$

so that $\mu_n(A)$ may be expressed as the following U-statistic:

$$\mu_n(A) = \frac{\tau_M}{\omega_d n(n-1)r_n^d}\sum_{i\neq j}\phi_{A,r_n}(X_i,X_j).$$

Similarly, $\nu_n(A)$ may be written as

$$\nu_n(A) = \frac{\tau_M}{\gamma_d n(n-1)r_n^{d+1}}\sum_{i\neq j}\bar{\phi}_{A,r_n}(X_i,X_j).$$

with the symmetric kernel

$$\bar{\phi}_{A,r}(x,y) = \frac{1}{2}\Big\{\mathbf{1}_A(x)\mathbf{1}_{A^c}(y) + \mathbf{1}_A(y)\mathbf{1}_{A^c}(x)\Big\}\mathbf{1}\{\|x-y\| \le r\}. \tag{2.2}$$

We shall need the following Hoeffding's Inequality for U-statistics [24], which is a special case of [18, Thm. 4.1.8].

Theorem 4. *Let ϕ be a measurable, bounded kernel on $\mathbb{R}^d\times\mathbb{R}^d$ and let $\{X_k : k\in\mathbb{N}\}$ be i.i.d. random vectors in $\mathbb{R}^d$. Assume that $\mathbb{E}[\phi(X_1,X_2)] = 0$ and that $b := \|\phi\|_\infty < \infty$, and let $\sigma^2 = \mathrm{Var}(\phi(X_1,X_2))$. Then, for all $t > 0$,*

$$\mathbb{P}\left[\frac{1}{n(n-1)}\sum_{i\neq j}\phi(X_i,X_j) \ge t\right] \le \exp\left(-\frac{nt^2}{5\sigma^2+3bt}\right).$$

To prove Theorem 1, we establish the almost-sure convergence of $\mu_n(A)$ to $\mu(A)$ and of $\nu_n(A)$ to $\nu(A)$ for a subset $A\subset M$ with smooth relative boundary. To this aim, we combine upper bounds on bias terms together with exponential inequalities for U-statistics. The bias terms involve volume bounds which we present next, and integrations over some neighborhoods of the boundary of a regular set, namely tubular neighborhoods or simply tubes, which comes after that.

2.1 Volume bounds

For any $r > 0$, define

$$M_r = \{x\in M : \mathrm{dist}(x,\partial M) \ge r\}. \tag{2.3}$$

The following two lemmas provide bounds on the volume of the intersection of balls with some subsets of M.

Lemma 5. *Let R be a bounded open subset of $\mathbb{R}^d$ with* $\mathrm{reach}(\partial R) = \rho > 0$. *Set $A = R \cap M$. For any $r < \min\{\mathrm{reach}(\partial M); \rho\}$, any $0 \leq \eta \leq 1$, and all p in $\partial A \cap M_r$, we have*

$$\left|\mathrm{Vol}_d\left(B(p + \eta r e_p, r) \cap A^c\right) - \pi_d(\eta) r^d\right| \leq 2\omega_{d-1} r^{d+1}/\rho,$$

where e_p denotes the unit normal vector at p pointing inward A.

Proof. For ease of notation, set $B = B(p + \eta r e_p, r)$. Let $(\tilde{e}_1, \ldots, \tilde{e}_d)$ be an orthonormal frame at p, with $\tilde{e}_d = e_p$. Denote by $\tilde{x}_1, \ldots, \tilde{x}_d$ the local coordinates in this frame, such that p has coordinates 0. Then $\partial A \cap M$ can be expressed locally as the set of points $\tilde{x}$ such that $\tilde{x}^d = F(\tilde{x}^1, \ldots, \tilde{x}^{d-1})$ for some function F, and, if we set $\tilde{x}^{(d)} = (\tilde{x}^1, \ldots, \tilde{x}^{d-1})$, then

$$\begin{aligned}
\mathrm{Vol}_d(B \cap A^c) &= \int_B \mathbf{1}\{\tilde{x}^d < F(\tilde{x}^{(d)})\} \mathrm{d}\tilde{x} \\
&= \int_B \left[\mathbf{1}\{\tilde{x}^d < F(\tilde{x}^{(d)})\}\mathbf{1}\{\tilde{x}^d < 0\} + \mathbf{1}\{\tilde{x}^d < F(\tilde{x}^{(d)})\}\mathbf{1}\{\tilde{x}^d > 0\}\right] \mathrm{d}\tilde{x}
\end{aligned}$$

Since

$$\pi_d(\eta) r^d = \int_B \mathbf{1}\{\tilde{x}^d < 0\} \mathrm{d}\tilde{x}$$

it follows that

$$\begin{aligned}
\left|\mathrm{vol}_d\left(B_n \cap A^c\right) - \pi_d(\eta) r_n^d\right| &\leq \int_B \left[\mathbf{1}\{\tilde{x}^d > F(\tilde{x}^{(d)})\}\mathbf{1}\{\tilde{x}^d < 0\} + \mathbf{1}\{\tilde{x}^d < F(\tilde{x}^{(d)})\}\mathbf{1}\{\tilde{x}^d > 0\}\right] \mathrm{d}\tilde{x} \\
&\leq \int_{B_n} \mathbf{1}\left\{|\tilde{x}^d| \leq |F(\tilde{x}^{(d)})|\right\} \mathrm{d}\tilde{x} \leq 2 \int_{\{\|\tilde{x}^{(d)}\| \leq r\}} |F(\tilde{x}^{(d)})| \mathrm{d}\tilde{x}^{(d)}.
\end{aligned}$$

Expanding F at 0, we have, for all $\tilde{x}$ with $\|\tilde{x}\| \leq r$,

$$F(\tilde{x}^{(d)}) = \sum_{i,j=1}^{d-1} G_{ij}(\xi) \tilde{x}^i \tilde{x}^j,$$

for some $\xi := \xi(\tilde{x}^{(d)})$. Since the reach bounds the principal curvatures by $1/\rho$ [20], we have $\sup_{p \in \partial A \cap M_r} \|G(p)\| \leq 1/\rho$. Then, using the change of variable $u = r\tilde{x}$, we deduce that

$$\begin{aligned}
\left|\mathrm{vol}_d\left(B(p + \eta r e_p, r) \cap A^c\right) - \pi_d(\eta) r_n^d\right| &\leq 2\omega_{d-1} \sup_{p \in \partial A \cap M} \|G(p)\| r^{d+1} \\
&\leq 2\omega_{d-1} r^{d+1}/\rho. \qquad \square
\end{aligned}$$

Lemma 6. *There exists some constant $C > 0$ such that, for all r, α satisfying $0 < 2r \leq \alpha \leq \mathrm{reach}(\partial M)$, and all x in M,*

$$\mathrm{Vol}_d(B(x, \alpha) \cap M_r) \geq C\alpha^d.$$

Proof. The main argument is to include a ball of radius $\alpha/4$ into $B(x, \alpha) \cap M_r$. We can proceed the following way. First, because $\rho := \mathrm{reach}(\partial M) > 0$, for any $x \in M$ there is $y \in M$ such that $x \in B(y, \rho) \subset M$. Second, since $\mathrm{dist}(y, \partial M) \geq \rho$ and $\rho \geq 2r$, we have $y \in M_r$ and $B(y, \rho - r) \subset M_r$. Hence

$$B(x, \alpha) \cap B(y, \rho - r) \subset B(x, \alpha) \cap M_r.$$

If $y = x$, the result is trival. Otherwise, let $z := x + (r + \alpha/4)(y - x)/\|y - x\|$ and note that $B(z, \alpha/4)$ is a ball of radius $\alpha/4$ included in $B(x, \alpha) \cap B(y, \rho - r)$. $\square$

2.2 Integration over tubes

We introduce the notion of tubes and some of their properties; see [22] for an extensive treatment. Let S be a submanifold of $\mathbb{R}^d$. The *tubular neighborhood* of radius $r > 0$ about S, denoted $\mathcal{V}(S, r)$, is the set of points x in $\mathbb{R}^d$ for which there exists $s \in S$ with $\|x - s\| < r$ and such that the line joining x and s is orthogonal to S at s. When S is without boundary, $\mathcal{V}(S, r)$ coincides with the set of points x in $\mathbb{R}^d$ at a distance no more than r from S. If S has boundary, then the tube coincides with the set of points at distance no more than r, with the ends removed, corresponding to the points projecting onto ∂S. Assume S is of codimension 1, and oriented, and define e_p as the (unit) normal vector of S at $p \in S$. When $r < \mathrm{reach}(S)$, $\mathcal{V}(S, r)$ admits the following parameterization

$$\mathcal{V}(S, r) = \{x = p + te_p : p \in S, -r \le t \le r\}.$$

Denote by $\mathbb{I}_p$ the second fundamental form of S at $p \in S$. The infinitesimal change of volume function is defined on $S \times (-r; r)$ by $\vartheta(p, t) = \det(I - t\mathbb{I}_p)$; the dependence of ϑ on S is omitted. Given an integrable function g on $\mathcal{V}(S, r)$, we have:

$$\int_{\mathcal{V}(S,r)} g(x)\mathrm{d}x = \int_S \int_{-r}^{r} g(p, t)\vartheta(p, t)\mathrm{d}t\, v_\sigma(\mathrm{d}p),$$

where v_σ is the Riemannian volume measure on S.

Lemma 7. *Assume S is a submanifold of $\mathbb{R}^d$ of codimension 1, with $\rho := \mathrm{reach}(S) > 0$. Then, for all $r < \rho$,*

$$\sup_{p\in S} \sup_{-r\le t\le r} \vartheta(p, t) \le (1 + r/\rho)^{d-1},$$

and

$$\sup_{p\in S} \sup_{-r\le t\le r} |\vartheta'(p, t)| \le \frac{(d - 1)(1 + r/\rho)^{d-1}}{\rho - r}$$

where ϑ' is the derivative of ϑ with respect to t.

Proof. By [20, Thm. 4.18], the reach bounds the radius of curvature from below so that the principal curvatures $\kappa^{(1)}, \ldots, \kappa^{(d-1)}$ (the eigenvalues of the second fundamental form) are everywhere bounded (in absolute value) from above by $1/\rho$. Therefore, for $r < \rho$ and $-r \le t \le r$,

$$0 \le \vartheta(p, t) = \det(I - t\mathbb{I}_p) = \prod_{i=1}^{d-1} \left(1 - \kappa_p^{(i)} t\right) \le (1 + r/\rho)^{d-1}.$$

For the derivative of ϑ, we have

$$\frac{\vartheta'(p, t)}{\vartheta(p, t)} = -\sum_{i=1}^{d-1} \frac{\kappa_p^{(i)}}{1 - \kappa_p^{(i)} t}.$$

Hence

$$|\vartheta'(p, t)| \le \vartheta(p, t)(d - 1)\frac{1/\rho}{1 - r/\rho} \le \frac{(d - 1)(1 + r/\rho)^{d-1}}{\rho - r}.$$ □

The celebrated Weyl's tube formula [40] provides fine estimates for the volume of a tubular region around a smooth submanifold of $\mathbb{R}^d$. We only require a rough upper bound of the right order of magnitude, which we state and prove here.

Lemma 8. *For any bounded open subset $R \subset \mathbb{R}^d$ with* $\mathrm{reach}(\partial R) = \rho > 0$ *and any* $0 < r < \rho$,

$$\mathrm{Vol}_d(\mathcal{V}(\partial R, r)) \le 2^d \,\mathrm{Vol}_{d-1}(\partial R)\, r.$$

In particular, Lemma 8 implies

$$\mu\left[\mathcal{V}(\partial M, r)\right] \le Cr, \quad \forall r < \mathrm{reach}(\partial M), \tag{2.4}$$

where C is a constant depending only on M.

Proof. Using the uniform bound of the infinitesimal change of volume given in Lemma 7, we have

$$\begin{aligned}\mathrm{Vol}_d\left[\mathcal{V}(\partial R, r)\right] &= \int_{\partial R}\int_{-r}^{r} \vartheta(p,u)\mathrm{d}u\, v_\sigma(\mathrm{d}p) \\ &\le \mathrm{Vol}_{d-1}(\partial R)\, 2r(1+r/\rho)^{d-1} \le 2^d \,\mathrm{Vol}_{d-1}(\partial R)\, r. \end{aligned}$$

□

2.3 Bounds on bias terms

Recall the definition of M_r in (2.3).

Lemma 9. *Let $\phi_{A,r}$ be defined as in* (2.1). *There exists a constant C, depending only on M, such that, for any $A \subset M$ and $r <$* $\mathrm{reach}(\partial M)$,

$$\left|\frac{\tau_M}{\omega_d r^d}\mathbb{E}\left[\phi_{A,r}(X_1,X_2)\right] - \mu(A)\right| \le \mu(A \cap M_r^c).$$

Proof. Assume without loss of generality that $\tau_M = 1$. We first note that

$$\mathbb{E}\left[\phi_{A,r}(X_1,X_2)\right] = \mathbb{E}\left[\mathbf{1}_A(X_1)\mathbf{1}\{\|X_1 - X_2\| \le r\}\right].$$

We partition A into $A \cap M_r$ and $A \cap M_r^c$. By conditioning on X_1, we have

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{A\cap M_r}(X_1)\mathbf{1}\{\|X_1 - X_2\| \le r\}\right] &= \omega_d r^d \mu(A \cap M_r) = \omega_d r^d \mu(A) - \omega_d r^d \mu(A \cap M_r^c); \\ \mathbb{E}\left[\mathbf{1}_{A\cap M_r^c}(X_1)\mathbf{1}\{\|X_1 - X_2\| \le r\}\right] &\le \omega_d r^d \mu(A \cap M_r^c).\end{aligned}$$

Hence the result. □

Lemma 10. *Let $A = R \cap M$, where R is a bounded domain with smooth boundary and* $\mathrm{reach}(\partial R) = \rho > 0$. *Let $\bar{\phi}_{A,r}$ be defined as in* (2.2).

(*i*) *There exists a constant C, depending only on M, such that, for any $A \subset M$ and $r <$* $\min\{\rho/2, \mathrm{reach}(\partial M)\}$,

$$\left|\frac{\tau_M}{\gamma_d r^{d+1}}\mathbb{E}\left[\bar{\phi}_{A,r}(X_1,X_2)\right] - \nu(A)\right| \le C\left(\mathrm{Vol}_{d-1}(\partial R \cap \mathcal{V}(\partial M, r)) + \mathrm{Vol}_{d-1}(\partial R \cap M)\frac{r}{\rho}\right).$$

(*ii*) *There exists a constant C, depending only on M, such that, for any $A \subset M$ and $r < \min\{\rho/2, \mathrm{reach}(\partial M)\}$,*

$$\frac{\tau_M}{\gamma_d r^{d+1}} \mathbb{E}\left[\bar{\phi}_{A,r}(X_1, X_2)\right] - \frac{\mathrm{Vol}_{d-1}(\partial A \cap M_r)}{\mathrm{Vol}_d(M)} \geq -C\nu(A)\frac{r}{\rho}. \tag{2.5}$$

Proof. Assume without loss of generality that $\tau_M = 1$. Let S denote $\partial R \cap M$. Then

$$\mathbb{E}\left[\bar{\phi}_{A,r}(X_1, X_2)\right] = \mathbb{E}\left[\mathbf{1}_A(X_1)\mathbf{1}_{A^c}(X_2)\mathbf{1}\left\{\|X_1 - X_2\| \leq r\right\}\right] = \int_D \mathrm{Vol}_d\left[B(x,r) \cap A^c\right]\mu(\mathrm{d}x),$$

where

$$D = \left\{x \in A \,:\, \mathrm{dist}(x, \partial R) \leq r\right\}.$$

Since $r < \rho$, the projection on ∂R is well-defined on D, and any x in D can be written as $x = p + te_p$, for $p \in \partial R$, and with e_p the unit normal vector of ∂R at p pointing inwards.

We partition D into $D \cap M_r$ and $D \cap M_r^c$. Denote by S_r the projection of $D \cap M_r$ on S. We have

$$\begin{aligned}\int_{D\cap M_r} \mathrm{Vol}_d\left[B(x,r) \cap A^c\right]\,\mathrm{d}x &= \int_{S_r}\int_{-r}^{0} \mathrm{Vol}_d\left[B(p + te_p, r) \cap A^c\right]\vartheta(p,t)\mathrm{d}t\,\nu_\sigma(\mathrm{d}p) \\ &= r\int_{S_r}\int_{0}^{1} \mathrm{Vol}_d\left[B(p - \eta r e_p, r) \cap A^c\right]\vartheta(p, r\eta)\mathrm{d}\eta\,\nu_\sigma(\mathrm{d}p).\end{aligned}$$

Therefore

$$\begin{aligned}&\left|\frac{1}{r^{d+1}}\int_{D\cap M_r} \mathrm{Vol}_d\left[B(x,r) \cap A^c\right]\mathrm{d}x - \gamma_d\nu(A)\right| \\ &\quad\leq \frac{1}{r^d}\int_{S_r}\int_0^1 \left|\mathrm{Vol}_d\left[B(p - \eta r e_p, r) \cap A^c\right] - \pi_d(\eta)r^d\right|\vartheta(p, r\eta)\mathrm{d}\eta\,\nu_\sigma(\mathrm{d}p) \\ &\qquad+ \left|\int_{S_r}\int_0^1 \pi_d(\eta)\vartheta(p, r\eta)\mathrm{d}\eta\,\nu_\sigma(\mathrm{d}p) - \gamma_d\nu(A)\right|.\end{aligned} \tag{2.6}$$

Lemma 5 provides the inequality $\left|\mathrm{Vol}_d\left[B(p - \eta r e_p, r) \cap A^c\right] - \pi_d(\eta)r^d\right| \leq 2\omega_{d-1}r^{d+1}/\rho$, and the first inequality of Lemma 7 states that $\sup_{p\in S}\sup_{-r\leq t\leq t}\vartheta(p,t) \leq (1 + r/\rho)^{d-1}$. Since $r < \rho$, $\sup_{p\in S}\sup_{0\leq\eta\leq 1}\vartheta(p, \eta r) \leq 2^{d-1}$. Hence, the first term on the right-hand side is bounded by

$$2\omega_{d-1}(r/\rho)\int_{S_r}\int_0^1 \vartheta(p, r\eta)\mathrm{d}\eta\,\nu_\sigma(\mathrm{d}p) \leq 2^d\omega_{d-1}(r/\rho)\,\mathrm{Vol}_{d-1}(S_r).$$

To bound the second term, a Taylor expansion leads to the relation $\vartheta(p, r\eta) = 1 + \vartheta'(p, r\xi_\eta)r\eta$ for some $0 < \xi_\eta < 1$. The second inequality of Lemma 7 states that $\sup_{p\in S}\sup_{-r\leq t\leq r}|\vartheta'(p,t)| \leq (d-1)(1 + r/\rho)^{d-1}/(\rho - r)$ so that $\sup_{p\in S}\sup_{0\leq\eta\leq 1}|\vartheta'(p, r\xi_\eta)|$ is bounded by $(d-1)2^d/\rho$ since $r < \rho$.

Recall that the constant γ_d is expressed as $\gamma_d = \int_0^1 \pi_d(\eta)\mathrm{d}\eta$. Then the second term in the right-hand side of (2.6) is bounded by

$$\begin{aligned}\left|\int_{S_r}\int_0^1 \pi_d(\eta)\mathrm{d}\eta\, \nu_\sigma(\mathrm{d}p) - \gamma_d \nu(A)\right| + r\int_{S_r}\int_0^1 \eta\pi_d(\eta)|\vartheta'(p, r\xi_\eta)|\mathrm{d}\eta\, \nu_\sigma(\mathrm{d}p)\\ \leq \gamma_d\,|\mathrm{Vol}_{d-1}(S_r) - \mathrm{Vol}_{d-1}(S)| + (d-1)2^d\gamma_d(r/\rho)\,\mathrm{Vol}_{d-1}(S_r)\\ \leq \gamma_d\,\mathrm{Vol}_{d-1}(S\cap M_r^c) + (d-1)2^d\gamma_d(r/\rho)\,\mathrm{Vol}_{d-1}(S_r),\end{aligned}$$

where we have used the fact that $S\setminus S_r \subset M_r^c$ since $S\cap M_r \subset S_r$. Collecting terms, the term in (2.6) is bounded by

$$\gamma_d\,\mathrm{Vol}_{d-1}(S\cap M_r^c) + C\frac{r}{\rho}\,\mathrm{Vol}_{d-1}(S_r),$$

for some constant C independent of M.

For the integral over $D\cap M_r^c$, since D is included in the intersection of tubes of radius r about ∂R and ∂M, i.e., $D\subset \mathcal{V}(\partial R, r)\cap\mathcal{V}(\partial M, r)$, we have

$$\begin{aligned}\int_{D\cap M_r^c}\mathrm{Vol}_d\left[B(x,r)\cap A^c\right]\mathrm{d}x &\leq \int_{\partial R\cap\mathcal{V}(\partial M,r)}\int_{-r}^0 \mathrm{Vol}_d\left[B(p+te_p,r)\cap A^c\right]\vartheta(p,t)\mathrm{d}t\,\nu_\sigma(\mathrm{d}p)\\ &= r\int_{\partial R\cap\mathcal{V}(\partial M,r)}\int_0^1 \mathrm{Vol}_d\left[B(p-\eta re_p,r)\cap A^c\right]\vartheta(p,r\eta)\mathrm{d}\eta\,\nu_\sigma(\mathrm{d}p)\\ &\leq 2^{d-1}\omega_d r^{d+1}\,\mathrm{Vol}_{d-1}(\partial R\cap\mathcal{V}(\partial M,r)),\end{aligned}$$

where we have used Lemma 7 again to bound $|\vartheta(p,r\eta)|$ by $(1+r/\rho)^{d-1}\leq 2^{d-1}$ in the last inequality.

Combining the two inequalities on the integrals over $D\cap M_r$ and $D\cap M_r^c$, we obtain that

$$\begin{aligned}&\left|\frac{1}{\gamma_d r^{d+1}}\mathbb{E}\left[\bar{\phi}_{A,r}(X_1,X_2)\right] - \nu(A)\right|\\ &\quad\leq \mathrm{Vol}_{d-1}(S\cap M_r^c) + C\frac{r}{\rho}\,\mathrm{Vol}_{d-1}(S_r) + 2^{d-1}\omega_d\,\mathrm{Vol}_{d-1}(\partial R\cap\mathcal{V}(\partial M,r))\\ &\quad\leq C\left(\mathrm{Vol}_{d-1}(\partial R\cap\mathcal{V}(\partial M,r)) + \mathrm{Vol}_{d-1}(S)\frac{r}{\rho}\right),\end{aligned}$$

which proves the first bound stated in Lemma 10.

To prove (*ii*), using the bound on (2.6), we deduce that

$$\begin{aligned}\frac{1}{\gamma_d r^{d+1}}\mathbb{E}\left[\bar{\phi}_{A,r}(X_1,X_2)\right] &\geq \frac{1}{\gamma_d r^{d+1}}\int_{D\cap M_r}\mathrm{Vol}_d\left[B(x,r)\cap A^c\right]\mathrm{d}x\\ &\geq \mathrm{Vol}_{d-1}(S) - \left[\mathrm{Vol}_{d-1}(S\cap M_r^c) + \frac{C}{\gamma_d}\frac{r}{\rho}\,\mathrm{Vol}_{d-1}(S_r)\right]\\ &\geq \mathrm{Vol}_{d-1}(S\cap M_r) - C\frac{r}{\rho}\,\mathrm{Vol}_{d-1}(S_r),\end{aligned}$$

and since $S_r\subset S$, the result follows. □

2.4 Exponential inequalities

Proposition 11. *Fix a sequence $r_n \to 0$. Let $A \subset M$ be an arbitrary open subset of M. There exists a constant C depending only on M such that, for any $\varepsilon > 0$, and all n large enough, we have*

$$\mathbb{P}\left[|\mu_n(A) - \mu(A)| \geq \varepsilon\right] \leq 2\exp\left(-\frac{nr_n^d\varepsilon^2}{C(1+\varepsilon)}\right).$$

In particular, if $nr_n^d/\log n \to \infty$, then $\mu_n(A)$ converges almost surely to $\mu(A)$ when $n \to \infty$.

Proof. By the triangle inequality, we have

$$|\mu_n(A) - \mu(A)| \leq \left|\mu_n(A) - \mathbb{E}\left[\mu_n(A)\right]\right| + \left|\mathbb{E}\left[\mu_n(A)\right] - \mu(A)\right|.$$

For all n large enough such that $r_n \leq \mathrm{reach}(\partial M)$, the second term on the right-hand side (the bias term) is bounded by Cr_n with C depending only on M. Indeed, Lemma 9 states that the bias is lower than $\mu(A \cap M_{r_n}^c)$. And the tubular neighborhood of ∂M of radius r_n, which contains $A \cap M_{r_n}^c$, has a volume bounded by Cr_n by (2.4).

Assume that n is large enough such that $2Cr_n \leq \varepsilon$. We then apply Theorem 4, which is Hoeffding's Inequality for U-statistics, to the first term (the deviation term) on the right-hand side with the kernel

$$\phi := \phi_{A,r_n} - \mathbb{E}\left[\phi_{A,r_n}(X_1, X_2)\right]$$

and $t = \omega_d r^d \varepsilon/2$. The kernel satisfies $\|\phi\|_\infty \leq 1$, and simple calculations yields

$$\mathrm{Var}(\phi(X_1, X_2)) \leq \mathbb{E}\left[\phi_{A,r_n}(X_1, X_2)^2\right] \leq \mu(A)\omega_d r_n^d/\tau_M \leq \omega_d r_n^d/\tau_M.$$

From this we obtain the large deviation bound. The almost sure convergence is then a simple consequence of the Borel-Cantelli Lemma. □

Proposition 12. *Fix a sequence $r_n \to 0$. Let A be an open subset of M with smooth relative boundary and positive reach. There exists a constant C depending only on M such that, for any $\varepsilon > 0$, and for all n large enough, we have*

$$\mathbb{P}\left[|\nu_n(A) - \nu(A)| \geq \epsilon\right] \leq 2\exp\left(-\frac{nr_n^{d+1}\epsilon^2}{C(\nu(A)+\epsilon)}\right).$$

In particular, if $nr_n^{d+1}/\log n \to \infty$, then

$$\nu_n(A) \to \nu(A), \quad n \to \infty, \quad \textit{almost surely}.$$

Proof. By the triangle inequality, we have

$$|\nu_n(A) - \nu(A)| \leq |\nu_n(A) - \mathbb{E}\left[\nu_n(A)\right]| + |\mathbb{E}\left[\nu_n(A)\right] - \nu(A)|.$$

Using the control on the bias in Lemma 10-(i), the second term on the right-hand side goes to 0 as $n \to \infty$. Then for n large enough, we apply Hoeffding's inequality of Theorem 4 to the first term on the right-hand side with the kernel

$$\phi := \bar{\phi}_{A,r_n} - \mathbb{E}\left[\bar{\phi}_{A,r_n}(X_1, X_2)\right]$$

and $t := \gamma_d r^{d+1} \nu(A)\epsilon/2$. The kernel satisfies $\|\phi\|_\infty \leq 1$, hence

$$\mathrm{Var}(\phi(X_1, X_2)) \leq \mathbb{E}\left[\bar{\phi}_{A,r_n}(X_1, X_2)^2\right] = \mathbb{E}\left[\bar{\phi}_{A,r_n}(X_1, X_2)\right] \leq 2\gamma_d \nu(A) r_n^{d+1}/\tau_M,$$

where the last inequality follows from upper bound on the bias of Lemma 10-(*i*) for n large enough. From this we obtain the large deviation bound, and the almost sure convergence is a consequence of the Borel-Cantelli Lemma. □

2.5 Proof of Theorem 1

The first statement of Theorem 1 is an immediate consequence of the exponential inequalities of Propositions 11 and 12.

To prove the second statement, under the conditions of Theorem 1, for any subset A with smooth relative boundary, with probability one $\lim_n h_n(A) = h(A; M)$ while $h_n(A) \geq \frac{\omega_d}{\gamma_d r_n} H(G_{n,r_n})$, so that $\limsup_n \frac{\omega_d}{\gamma_d r_n} H(G_{n,r_n}) \leq h(A; M)$. Then we obtain the upper bound of Theorem 1 by taking the infimum over all such subsets A.

3 Proof of Theorems 2 and 3: consistent estimation

Consistent estimation in the context of Theorem 2 is possible because the class $\mathcal{R}_n$ is sufficiently rich as to include sets that approach Cheeger sets of M and its complexity is controlled, so as to allow for a uniform convergence both in terms of discrete volume and discrete perimeter. This control on the complexity of $\mathcal{R}_n$ we exploit in building a covering for $\mathcal{R}_n$, which is done in Section 3.1, later used to obtain uniform versions of Propositions 11 and 12. Then Part *(i)* of Theorem 2, which states the convergence of a penalized graph Cheeger constant towards the Cheeger constant of M, is proved in Section 3.7. Finally, Part *(ii)*, which characterizes the accumulation points of a sequence of minimizing sets, is proved in Section 3.8. The convergence of the discrete measures associated with a sequence of minimizing sets (Theorem 3) is proved in Section 3.9.

3.1 Covering numbers

For $\rho > 0$, let $\mathcal{R}_\rho$ be the class of open subsets $R \subset (0,1)^d$ with $\mathrm{reach}(\partial R) \geq \rho$. Let $d_H(R, R')$ be the Hausdorff distance between two sets R and R', i.e.,

$$d_H(R, R') = \inf\{r > 0 \,:\, R \subset R' \oplus B(r) \quad \text{and} \quad R' \subset R \oplus B(r)\}.$$

Denote by $\mathcal{N}\left(\varepsilon, \mathcal{R}_\rho, d_H\right)$ be the covering number of $\mathcal{R}_\rho$ for the Hausdorff distance, i.e., the minimal number of balls of radius ε for the Hausdorff distance, centered at elements in $\mathcal{R}_\rho$ that are needed to cover $\mathcal{R}_\rho$.

Lemma 13. *(i) There exists a constant C depending only on d such that, for any $\varepsilon > 0$ and any $\rho > 0$:*

$$\log \mathcal{N}(\varepsilon, \mathcal{R}_\rho, d_H) \leq C\left(\frac{1}{\varepsilon}\right)^d.$$

(ii) If $0 < \varepsilon < \rho$*, then for any* R *and* R' *in* $\mathcal{R}_\rho$*, if* $d_H(R, R') \leq \varepsilon$*, then* $R \Delta R' \subset \mathcal{V}(\partial R, \varepsilon) \cap \mathcal{V}(\partial R', \varepsilon)$*.*

Proof. Let $x_1, \ldots, x_n$ be an ε-packing of $(0,1)^d$, so $\cup_{i=1}^n B(x_i, \varepsilon)$ covers $(0,1)^d$ and $n \leq C\varepsilon^{-d}$ for some constant C depending only on d. For any set R in $\mathcal{R}_\rho$, define

$$I_\varepsilon(R) = \{i = 1, \ldots, n \,:\, B(x_i, \varepsilon) \cap R \neq \emptyset\}.$$

Then clearly, by definition of the covering, $R \subset \cup_{i \in I_\varepsilon(R)} B(x_i, \varepsilon)$, and

$$\cup_{i \in I_\varepsilon(R)} B(x_i, \varepsilon) \subset R \oplus B(2\varepsilon).$$

Therefore

$$d_H\left(\cup_{i \in I_\varepsilon(R)} B(x_i, \varepsilon), R\right) \leq 2\varepsilon.$$

Since when R ranges in $\mathcal{R}_\rho$, the cardinality of sets of the form $\cup_{i \in I_\varepsilon(R)} B(x_i, \varepsilon)$ is bounded by 2^n, then the collection of Hausdorff balls of radius 2ε and centered set of the form $\cup_{i \in I} B(x_i, \varepsilon)$, where I is any subset of $\{1, \ldots, n\}$, covers $\mathcal{R}_\rho$. By doubling the radius of the balls, we can take centers in $\mathcal{R}_\rho$, which proves the first part of the lemma.

The second part follows from the fact that if $\text{reach}(\partial R) > \rho$, then $\partial R \oplus B(\rho) = \mathcal{V}(\partial R, \rho)$, assuming, without loss of generality, that ∂R has no boundary. □

We mention that the bound on the ε-entropy of $\mathcal{R}_\rho$ is rather weak. Standard results by Kolmogorov and Tikhomirov [25] suggest a bound of the form $C(\rho\varepsilon)^{-(d-1)/2}$. Such a result would change the exponent for r_n in Theorem 2 to $(3d+1)/2$.

3.2 Perimeter bounds of a regular set

The classical isoperimetric inequality provides a bound of the volume of a Borel set R in terms of its perimeter (see e.g., Evans and Gariepy, 1992):

$$d\omega_d^{1/d} \,\text{Vol}_d(R)^{1-1/d} \leq \text{Vol}_{d-1}(\partial R). \tag{3.1}$$

But, in the case where ∂R has positive reach, the perimeter may in turn be bounded by the volume, as stated in Lemma 14 below. The proof uses the following inequality: for every Borel sets R, S

$$\text{Vol}_{d-1}\left(\partial(R \cup S)\right) + \text{Vol}_{d-1}\left(\partial(R \cap S)\right) \leq \text{Vol}_{d-1}(\partial R) + \text{Vol}_{d-1}(\partial S). \tag{3.2}$$

Lemma 14. *Let* R *be a bounded open subset of* $\mathbb{R}^d$ *with* $\text{reach}(\partial R) = \rho > 0$*. Then,*

$$\text{Vol}_{d-1}(\partial R) \leq d \,\text{Vol}_d(R)/\rho.$$

Proof. Since $\text{reach}(\partial R) = \rho > 0$, a ball of radius ρ rolls freely in R. Consequently R can be written as a countable union of balls of radius ρ, i.e.,

$$R = \bigcup_{i=1}^{\infty} B(x_i, \rho).$$

Set $R_n = \cup_{i=1}^n B_i$ where $B_i = B(x_i, \rho)$.

Using the decomposition $R_{n+1} = R_n \cup B_{n+1}$, on the one hand we have

$$\mathrm{Vol}_d(R_{n+1}) = \mathrm{Vol}_d(R_n \cup B_{n+1}) = \mathrm{Vol}_d(R_n) + \omega_d \rho^d - \mathrm{Vol}_d(R_n \cap B_{n+1}),$$

and on the other hand, using inequality (3.2), we have

$$\mathrm{Vol}_{d-1}(\partial R_{n+1}) = \mathrm{Vol}_{d-1}(\partial(R_n \cup B_{n+1})) \leq \mathrm{Vol}_{d-1}(\partial R_n) + d\omega_d \rho^{d-1} - \mathrm{Vol}_{d-1}(\partial(R_n \cap B_{n+1})).$$

Consequently

$$\begin{aligned}\mathrm{Vol}_{d-1}(\partial R_{n+1}) - \frac{d}{\rho}\mathrm{Vol}_d(R_{n+1}) \quad &\leq \quad \mathrm{Vol}_{d-1}(\partial R_n) - \frac{d}{\rho}\mathrm{Vol}_d(R_n) \\ &\quad + \left[\frac{d}{\rho}\mathrm{Vol}_d(R_n \cap B_{n+1}) - \mathrm{Vol}_{d-1}(\partial(R_n \cap B_{n+1}))\right].\end{aligned}$$

But, using the isoperimetric inequality (3.1), we may write

$$\begin{aligned}&\frac{d}{\rho}\mathrm{Vol}_d(R_n \cap B_{n+1}) - \mathrm{Vol}_{d-1}\left(\partial(R_n \cap B_{n+1})\right) \\ &\quad \leq \frac{d}{\rho}\mathrm{Vol}_d(R_n \cap B_{n+1}) - d\omega_d^{1/d}\Big(\mathrm{Vol}_d(R_n \cap B_{n+1})\Big)^{1-1/d} \\ &\quad \leq \Big(\mathrm{Vol}_d(R_n \cap B_{n+1})\Big)^{1-1/d}\left[\frac{d}{\rho}\mathrm{Vol}_d(R_n \cap B_{n+1})^{1/d} - d\omega_d^{1/d}\right] \leq 0\end{aligned}$$

since, in the last bracket, $\mathrm{Vol}_d(R_n \cap B_{n+1}) \leq \mathrm{Vol}_d(B_{n+1}) = \omega_d \rho^d$. Therefore, for all $n \geq 1$, we have

$$\mathrm{Vol}_{d-1}(\partial R_{n+1}) - \frac{d}{\rho}\mathrm{Vol}_d(R_{n+1}) \leq \mathrm{Vol}_{d-1}(\partial R_n) - \frac{d}{\rho}\mathrm{Vol}_d(R_n).$$

But since R_1 is a ball of radius ρ, we have $\mathrm{Vol}_{d-1}(\partial R_1) - d\,\mathrm{Vol}_d(R_1)/\rho = 0$ and so

$$\mathrm{Vol}_{d-1}(\partial R_n) - \frac{d}{\rho}\mathrm{Vol}_d(R_n) \leq 0 \quad \text{for all } n \geq 1.$$

Since R_n converges to R in L^1, it follows from the lower semi-continuity of the perimeter, see e.g. [23, Prop. 2.3.6], that $\liminf_n \mathrm{Vol}_{d-1}(\partial R_n) \geq \mathrm{Vol}_{d-1}(\partial R)$. This concludes the proof. □

3.3 Exponential inequalities

We prove the uniform versions of Propositions 11 and 12 for the class $\mathcal{R}_\rho$.

Proposition 15. *There exists a constant C depending only on M such that, for any $\varepsilon, r > 0$ and all n satisfying $nr^d\rho^d\varepsilon^{d+2} > C$ and $\varepsilon > Cr$, we have*

$$\mathbb{P}\left[\sup_{R \in \mathcal{R}_\rho} |\mu_n(R) - \mu(R)| \geq \varepsilon\right] \leq 2\exp\left(-\frac{nr^d\varepsilon^2}{C(1+\varepsilon)}\right). \tag{3.3}$$

Proof. The bias term is dealt exactly as in Proposition 11, obtaining

$$\left|\mathbb{E}\left[\mu_n(R)\right] - \mu(R)\right| \leq C_0 r,$$

valid for all $R \in \mathcal{R}_\rho$, so assuming $\varepsilon > 2C_0 r$, we may focus on bounding the variance term

$$\mu_n(R) - \mathbb{E}\left[\mu_n(R)\right].$$

Define the kernel class

$$\mathcal{F} = \{\phi_{R,r} \,:\, R \in \mathcal{R}_\rho\}, \tag{3.4}$$

where $\phi_{R,r}$ is defined in (2.1). Let $U_n(\phi)$ be the U-process over $\mathcal{F}$ defined by

$$U_n(\phi) = \frac{1}{n(n-1)} \sum_{i \neq j} \phi(X_i, X_j).$$

Observe that

$$\sup_{R \in \mathcal{R}_\rho} \left|\mu_n(R) - \mathbb{E}\left[\mu_n(R)\right]\right| = \frac{\tau_M}{\omega_d r^d} \sup_{\phi \in \mathcal{F}} \left|U_n(\phi) - \mu^{\otimes 2}(\phi)\right|.$$

Consider a minimal covering of $\mathcal{R}_\rho$ of cardinal K by balls centered at elements $R_1, \ldots, R_K$ of $\mathcal{R}_\rho$, and of radius $\eta < \rho$ for the Hausdorff distance. By Lemma 13,

$$\log(K) \leq C_1 (1/\eta)^d.$$

For any R in $\mathcal{R}_\rho$, there exists $1 \leq k \leq K$ such that $d_H(R, R_k) \leq \eta$, which implies that $R \Delta R_k \subset \mathcal{V}(\partial R_k, \eta)$. Also, by Lemma 8, there exists a constant C_2 depending only on the dimension d such that $\mathrm{Vol}_d(\mathcal{V}(\partial R_k, \eta)) \leq C_2 \eta/\rho$, for all $1 \leq k \leq K$, which implies that

$$\mu\left(\mathcal{V}(\partial R_k, \eta)\right) \leq C_3 \eta/\rho, \quad \text{for all } 1 \leq k \leq K,$$

since $\eta < \rho$, and where C_3 now depends on M.

We have

$$\begin{aligned}
\left|\phi_{R,r}(x,y) - \phi_{R_k,r}(x,y)\right| &= \frac{1}{2}\left|\mathbf{1}_R(x) + \mathbf{1}_R(y) - \mathbf{1}_{R_k}(x) - \mathbf{1}_{R_k}(y)\right| \mathbf{1}\left\{\|x-y\| \leq r\right\} \\
&\leq \frac{1}{2}\left(\mathbf{1}_{R \Delta R_k}(x) + \mathbf{1}_{R \Delta R_k}(y)\right) \mathbf{1}\left\{\|x-y\| \leq r\right\}.
\end{aligned}$$

Next, consider the inequality

$$\left|U_n(\phi_{R,r}) - \mu^{\otimes 2}(\phi_{R,r})\right| \leq \left|U_n(\phi_{R,r}) - U_n(\phi_{R_k,r})\right| + \left|U_n(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R_k,r})\right| + \left|\mu^{\otimes 2}(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R,r})\right|.$$

For the double expectations, we have,

$$\begin{aligned}
\left|\mu^{\otimes 2}(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R,r})\right| &\leq \mu^{\otimes 2}\left|\phi_{R_k,r} - \phi_{R,r}\right| \\
&= \mathbb{E}\left[1_{R \Delta R_k}(X_1) \mathbf{1}\left\{\|X_1 - X_2\| \leq r\right\}\right] \\
&= \int_{R \Delta R_k} \mu\left(B(x,r)\right) \mu(\mathrm{d}x) \\
&\leq \int_{\mathcal{V}(\partial R_k, \eta)} \mu\left(B(x,r)\right) \mu(\mathrm{d}x) \\
&\leq \frac{\omega_d r^d}{\tau_M} \mu\left(\mathcal{V}(\partial R_k, \eta)\right) \\
&\leq C_4 r^d \eta/\rho,
\end{aligned}$$

with C_4 still depending only on M. The last inequality is a consequence of Lemmas 8 and 14, and the fact that $\mathrm{Vol}_d(R_k) \leq 1$ since $R_k \subset (0,1)^d$.

For the empirical averages, we have

$$\begin{aligned}
\left|U_n(\phi_{R,r}) - U_n(\phi_{R_k,r})\right| &\leq \frac{1}{2}\frac{1}{n(n-1)}\sum_{i\neq j}\left(\mathbf{1}_{R\Delta R_k}(X_i) + \mathbf{1}_{R\Delta R_k}(X_j)\right)\mathbf{1}\left\{\|X_i - X_j\| \leq r\right\} \\
&\leq \frac{1}{2}\frac{1}{n(n-1)}\sum_{i\neq j}\left(\mathbf{1}_{\mathcal{V}(\partial R_k,\eta)}(X_i) + \mathbf{1}_{\mathcal{V}(\partial R_k,\eta)}(X_j)\right)\mathbf{1}\left\{\|X_i - X_j\| \leq r\right\} \\
&= U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right).
\end{aligned}$$

Therefore,

$$\sup_{R\in\mathcal{R}_\rho}\left|U_n(\phi_{R,r}) - \mu^{\otimes 2}(\phi_{R,r})\right| \leq \max_{1\leq k\leq K} U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) + C_4\frac{r^d\eta}{\rho} + \max_{1\leq k\leq K}\left|U_n(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R_k,r})\right|.$$

Consequently, for any $\varepsilon > 0$, we may write

$$\begin{aligned}
&\mathbb{P}\left(\sup_{R\in\mathcal{R}_\rho}\left|\mu_n(R) - \mathbb{E}\left[\mu_n(R)\right]\right| \geq \varepsilon\right) \\
&= \mathbb{P}\left(\sup_{\phi\in\mathcal{F}}\left|U_n(\phi) - \mu^{\otimes 2}(\phi)\right| \geq \frac{\omega_d r^d\varepsilon}{\tau_M}\right) \\
&\leq \mathbb{P}\left(\max_{1\leq k\leq K} U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) \geq \frac{\omega_d r^d\varepsilon}{2\tau_M} - C_4\frac{r^d\eta}{\rho}\right) + \mathbb{P}\left(\max_{1\leq k\leq K}\left|U_n(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R_k,r})\right| \geq \frac{\omega_d r^d\varepsilon}{2\tau_M}\right) \\
&\leq K\max_{1\leq k\leq K}\mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) \geq \frac{\omega_d r^d\varepsilon}{2\tau_M} - C_4\frac{r^d\eta}{\rho}\right) + K\max_{1\leq k\leq K}\mathbb{P}\left(\left|U_n(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R_k,r})\right| \geq \frac{\omega_d r^d\varepsilon}{2\tau_M}\right),
\end{aligned}$$

by the union bound. To bound the first term, note first that

$$\mathrm{Var}\left(\phi_{\mathcal{V}(\partial R_k,\eta)}(X_1,X_2)\right) \leq \mathbb{E}\left[\phi_{\mathcal{V}(\partial R_k,\eta)}(X_1,X_2)^2\right] \leq \mathbb{E}\left[\phi_{\mathcal{V}(\partial R_k,\eta)}(X_1,X_2)\right],$$

with

$$\mathbb{E}\left[\phi_{\mathcal{V}(\partial R_k,\eta)}(X_1,X_2)\right] \leq \frac{\omega_d r^d}{\tau_M}\mu\left(\mathcal{V}(\partial R_k,\eta)\right) \leq C_4\frac{r^d\eta}{\rho},$$

for the same reasons as above. Now take $\eta = \rho\min(\omega_d\varepsilon/(8C_4\tau_M), 1)$. Then, for any $1 \leq k \leq K$, by Hoedffding's inequality for U-statistics (Theorem 4), we have,

$$\begin{aligned}
\mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) \geq \frac{\omega_d r^d\varepsilon}{2\tau_M} - C_4\frac{r^d\eta}{\rho}\right) &\leq \mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) - \mathbb{E}\left[U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right)\right] \geq \frac{\omega_d r^d\varepsilon}{4\tau_M}\right) \\
&\leq \exp\left(-\frac{n(\omega_d r^d\varepsilon/4\tau_M)^2}{5(C_4 r^d\eta/\rho) + 3(\omega_d r^d\varepsilon/4\tau_M)}\right) \\
&\leq \exp\left(-\frac{nr^d\varepsilon}{C_5}\right),
\end{aligned}$$

for a constant $C_5 > 0$ depending only on M. To bound the second term, since

$$\mathrm{Var}\left(\phi_{R_k,r}(X_1, X_2)\right) \leq \mathbb{E}\left[\phi_{R_k,r}(X_1, X_2)\right] \leq \omega_d r^d / \tau_M,$$

we may apply Lemma 4 again to obtain the bound

$$\begin{aligned}\mathbb{P}\left(\left|U_n(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R_k,r})\right| \geq \frac{\omega_d r^d \varepsilon}{2\tau_M}\right) &\leq \exp\left(-\frac{n(\omega_d r^d \varepsilon/2\tau_M)^2}{5\omega_d r^d + 3(\omega_d r^d \varepsilon/2\tau_M)}\right) \\ &\leq \exp\left(-\frac{nr^d\varepsilon^2}{C_6(1+\varepsilon)}\right),\end{aligned}$$

for a constant $C_6 > 0$ depending only on M.

With the choice of η as above, the cardinal K of the covering is such that $\log(K) \leq C_7(\varepsilon\rho)^{-d}$, for some constant C_7 depending only on M, and we obtain the bound

$$\begin{aligned}&\mathbb{P}\left(\sup_{R\in\mathcal{R}_\rho}\left|\mu_n(R) - \mathbb{E}\left[\mu_n(R)\right]\right| \geq \varepsilon\right) \\ &\quad\leq K\exp\left(-\frac{nr^d\varepsilon}{C_5}\right) + K\exp\left(-\frac{nr^d\varepsilon^2}{C_6(1+\varepsilon)}\right) \\ &\quad\leq 2\exp\left(C_7(\varepsilon\rho)^{-d} - \frac{nr^d\varepsilon^2}{C_8(1+\varepsilon)}\right) \\ &\quad\leq 2\exp\left(-\frac{nr^d\varepsilon^2}{C_9(1+\varepsilon)}\right),\end{aligned}$$

if $nr^d\varepsilon^{d+2}\rho^d > C_9$, for a constant C_9 depending only on M. □

For the perimeter, we only control the variance, as the bias may not be controlled uniformly over $\mathcal{R}_\rho$. Indeed, consider the case where M is a hypercube with rounded corners so as to satisfy the condition on its reach, and let R be another hypercube with rounded corners included in M sharing one of its faces with M. Then given a sample $X_1, \ldots, X_n$, it is possible to translate R inside M just enough that the translate does not share a boundary with M, while its discrete volume and perimeter are left equal to those of R.

Proposition 16. *There exists a constant C depending only on M such that, for any $\varepsilon > 0$, $\rho < 1$, $r < \min(\mathrm{reach}(M), \rho/2)$ and all n satisfying $nr^{2d+1}\rho^{d+1}\varepsilon^{d+2} > C$, we have*

$$\mathbb{P}\left[\sup_{R\in\mathcal{R}_\rho}|\nu_n(R) - \mathbb{E}[\nu_n(R)]| \geq \varepsilon\right] \leq 2\exp\left(-\frac{nr^{d+1}\rho\varepsilon^2}{C(1+\rho\varepsilon)}\right).$$

Proof. The proof follows that of Proposition 15, with the symmetric kernel $\bar{\phi}_{R,r}$ defined in (2.2) and the class $\bar{\mathcal{F}}$ defined in (3.4) with $\phi_{R,r}$ replaced by $\bar{\phi}_{R,r}$. Observe that

$$|\nu_n(R) - \mathbb{E}\left[\nu_n(R)\right]| = \frac{\tau_M}{\gamma_d r^{d+1}}\sup_{\phi\in\bar{\mathcal{F}}}\left|U_n(\phi) - \mu^{\otimes 2}(\phi)\right|.$$

As in the proof of Proposition 15, we start with a minimal covering of $\mathcal{R}_\rho$ of cardinal K by balls of radius η for the Hausdorff distance. For any R in $\mathcal{R}_\rho$ at a Hausdorff distance no more than η of an element R_k of the covering, we have

$$\begin{aligned}\left|\mathbf{1}_R(x)\mathbf{1}_{R^c}(y) - \mathbf{1}_{R_k}(x)\mathbf{1}_{R_k^c}(y)\right| &\leq \left|\mathbf{1}_R(x) - \mathbf{1}_{R_k}(x)\right|\mathbf{1}_{R^c}(y) + \mathbf{1}_{R_k}(x)\left|\mathbf{1}_{R^c}(y) - \mathbf{1}_{R_k^c}(y)\right| \\ &= \mathbf{1}_{R\Delta R_k}(x)\mathbf{1}_{R^c}(y) + \mathbf{1}_{R\Delta R_k}(y)\mathbf{1}_{R_k}(x) \\ &\leq \mathbf{1}_{R\Delta R_k}(x) + \mathbf{1}_{R\Delta R_k}(y).\end{aligned}$$

Hence,

$$\left|\bar{\phi}_{R,r}(x,y) - \bar{\phi}_{R_k,r}(x,y)\right| \leq 2\phi_{R\Delta R_k,r}(x,y) \leq 2\phi_{\mathcal{V}(\partial R_k,\eta)}(x,y),$$

and therefore, following the same arguments,

$$\left|\mu^{\otimes 2}(\bar{\phi}_{R,r}) - \mu^{\otimes 2}(\bar{\phi}_{R_k,r})\right| \leq 2\mu^{\otimes 2}(\phi_{\mathcal{V}(\partial R_k,\eta)}) \leq C_1 r^d \eta/\rho,$$

for a constant C_1 depending only on M; and also,

$$\left|U_n(\bar{\phi}_{R,r}) - U_n(\bar{\phi}_{R_k,r})\right| \leq 2U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right).$$

Hence

$$\begin{aligned}&\mathbb{P}\left(\sup_{R\in\mathcal{R}_\rho}|\nu_n(R) - \mathbb{E}[\nu_n(R)]| \geq \varepsilon\right) \\ &\quad\leq K\max_{1\geq k\geq K}\mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) \geq \frac{\gamma_d r^{d+1}\varepsilon}{4\tau_M} - C_1\frac{r^d\eta}{2\rho}\right) \\ &\qquad + K\max_{1\geq k\geq K}\mathbb{P}\left(\left|U_n(\bar{\phi}_{R_k,r}) - \mu^{\otimes 2}(\bar{\phi}_{R_k,r})\right| \geq \frac{\gamma_d r^{d+1}\varepsilon}{2\tau_M}\right).\end{aligned}$$

Take $\eta = \rho\min(\gamma_d r\varepsilon/(4C_1\tau_M), 1)$. For the first term, for any $1 \leq k \leq K$, we have,

$$\begin{aligned}\mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) \geq \frac{\gamma_d r^{d+1}\varepsilon}{4\tau_M} - C_1\frac{r^d\eta}{2\rho}\right) &\leq \mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) - \mathbb{E}\left[U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right)\right] \geq \frac{\gamma_d r^{d+1}\varepsilon}{8\tau_M}\right) \\ &\leq \exp\left(-\frac{n(\gamma_d r^{d+1}\varepsilon/8\tau_M)^2}{5(C_1 r^d\eta/\rho) + 3(\gamma_d r^{d+1}\varepsilon/8\tau_M)}\right) \\ &= \exp\left(-\frac{nr^{d+1}\varepsilon^2}{C_2(1+\varepsilon)}\right),\end{aligned}$$

for some constant $C_2 > 0$ depending only on M. For the second term, since by Lemma 10, when $r \leq \rho/2$,

$$\mathrm{Var}(\bar{\phi}_{R_k,r}) \leq C_3 r^{d+1}/\rho,$$

for a constant C_3 depending only on M, we have

$$\begin{aligned}\mathbb{P}\left(\left|U_n(\bar{\phi}_{R_k,r}) - \mu^{\otimes 2}(\bar{\phi}_{R_k,r})\right| \geq \frac{\gamma_d r^{d+1}\varepsilon}{2\tau_M}\right) &\leq \exp\left(-n\frac{(\gamma_d r^{d+1}\varepsilon/2\tau_M)^2}{5(C_3 r^{d+1}/\rho) + 3(\gamma_d r^{d+1}\varepsilon/2\tau_M)}\right) \\ &\leq \exp\left(-\frac{nr^{d+1}\rho\varepsilon^2}{C_4(1+\rho\varepsilon)}\right),\end{aligned}$$

for a constant $C_4 > 0$ depending only on M. Finally, with the choice of η as above, the cardinal K of the covering is such that $\log(K) \leq C_5(r\rho\varepsilon)^{-d}$, for C_5 depending only on M.

Then

$$K \max_{1\geq k\geq K} \mathbb{P}\left(U_n\left(\phi_{\mathcal{V}(\partial R_k,\eta)}\right) \geq \frac{\gamma_d r^{d+1}\varepsilon}{4\tau_M} - C_1\frac{r^d\eta}{\rho})\right) \leq \exp\left(-\frac{nr^{d+1}\varepsilon^2}{C_6(1+\varepsilon)}\right),$$

if $nr^{2d+1}\rho^d\varepsilon^{d+2} > C_6$, and

$$K \max_{1\geq k\geq K} \mathbb{P}\left(\left|U_n(\bar{\phi}_{R_k,r}) - \mu^{\otimes 2}(\bar{\phi}_{R_k,r})\right| \geq \frac{\gamma_d r^{d+1}\varepsilon}{2\tau_M}\right) \leq \exp\left(-\frac{nr^{d+1}\rho\varepsilon^2}{C_7(1+\rho\varepsilon)}\right),$$

if $nr^{2d+1}\rho^{d+1}\varepsilon^{d+2} > C_7$, where C_6 and C_7 depend on M only. Combining these inequalities, we conclude. □

3.4 A uniform control on $h_n(A)$

As we argued earlier, the boundary of M makes a uniform convergence of the perimeters of sets in $\mathcal{R}_n$ impossible. Our way around that is to compare the discrete perimeter of a set R with its perimeter inside M_{r_n}, thus avoiding the boundary of M, i.e., $\mathrm{Vol}_{d-1}(\partial R \cap M_{r_n})$, leading to a comparison between $h_n(R)$ and $h(R; M_{r_n})$. We relate the latter to $h(R; M)$ in Section 3.5.

Lemma 17. *Under the conditions of Theorem 2, with probability one, we have:*

$$\liminf_{n\to\infty} \inf_{R\in\mathcal{R}_n} (h_n(R) - h(R; M_{r_n})) \geq 0. \tag{3.5}$$

Proof. Take $R \in \mathcal{R}_n$ and define

$$\lambda_n(R) = \min(\mu_n(R), \mu_n(R^c)), \quad \lambda_n^*(R) = \frac{1}{\tau_M}\min(\mathrm{Vol}_d(R \cap M_{r_n}), \mathrm{Vol}_d(R^c \cap M_{r_n})),$$

as well as

$$\nu_n^*(R) = \frac{1}{\tau_M}\mathrm{Vol}_{d-1}(\partial R \cap M_{r_n}).$$

Then

$$\begin{aligned} h_n(R) - h(R; M_{r_n}) &= \frac{1}{\lambda_n(R)}(\nu_n(R) - \nu_n^*(R)) + \frac{\nu_n^*(R)}{\lambda_n(R)\lambda_n^*(R)}(\lambda_n^*(R) - \lambda_n(R)) \\ &=: \zeta_n(R) + \xi_n(R). \end{aligned}$$

Define the event

$$\Omega_n = \left\{\frac{1}{2} \leq \frac{\lambda_n(R)}{\lambda_n^*(R)} \leq \frac{3}{2}, \forall R \in \mathcal{R}_n\right\}.$$

We will see that $\mathbb{P}[\Omega_n] \to 1$.

Bounding $\zeta_n(R)$. By definition of $\mathcal{R}_n$, the sets R and R^c contain each a ball of radius ρ_n, and by Lemma 6, the volume of the intersection of this ball with M_{r_n} is bounded from below by $C_1\rho_n^d$, for a constant C_1 depending only on M. Hence,

$$\lambda_n^*(R) \geq C_1\rho_n^d. \tag{3.6}$$

Also, on Ω_n, $\lambda_n(R) \geq \lambda_n^*(R)/2$. These last two inequalities being valid for all $R \in \mathcal{R}_n$, for $\varepsilon > 0$ we have

$$\begin{aligned} I_1 := \mathbb{P}\left[\left[\inf_{R\in\mathcal{R}_n} \zeta_n(R) < -\varepsilon\right] \cap \Omega_n\right] &\leq \mathbb{P}\left[\inf_{R\in\mathcal{R}_n} (\nu_n(R) - \nu_n^*(R)) < -C_2\varepsilon\rho_n^d\right] \\ &\leq \mathbb{P}\left[\inf_{R\in\mathcal{R}_n} (\nu_n(R) - \mathbb{E}[\nu_n(R)]) + \inf_{R\in\mathcal{R}_n} (\mathbb{E}[\nu_n(R)] - \nu_n^*(R)) < -C_2\varepsilon\rho_n^d\right], \end{aligned}$$

for a constant $C_2 = C_1/2 > 0$. Using the bias bounds of Lemma 10 together with the perimeter bound in Lemma 14(ii), we have

$$\inf_{R\in\mathcal{R}_n} (\mathbb{E}[\nu_n(R)] - \nu_n^*(R)) \geq -C_3\frac{r_n}{\rho_n^2}.$$

Hence, since $r_n = o(\rho_n^\alpha)$ for any $\alpha > 0$, for ε fixed and n large enough, we have by assumption, for all n large enough,

$$I_1 \leq \mathbb{P}\left[\inf_{R\in\mathcal{R}_n} (\nu_n(R) - \mathbb{E}[\nu_n(R)]) < -C_2\varepsilon\rho_n^d/2\right] \leq \mathbb{P}\left[\sup_{R\in\mathcal{R}_{\rho_n}} |\nu_n(R) - \mathbb{E}[\nu_n(R)]| > C_2\varepsilon\rho_n^d/2\right],$$

where the second inequality comes from the fact that $\mathcal{R}_n \subset \mathcal{R}_{\rho_n}$. By the fact that $nr_n^{2d+1}\rho_n^\alpha \to \infty$ for any $\alpha > 0$, the conditions of Proposition 16 are satisfied, so that

$$I_1 \leq C_4 \exp\left(-\frac{nr_n^{d+1}\rho_n^{2d+1}\varepsilon^2}{C_4(1+\varepsilon)}\right),$$

for some constant $C_4 > 0$ and all n large enough. At last, we have

$$\frac{nr_n^{d+1}\rho_n^{2d+1}}{\log(n)} = nr_n^{2d+1}\frac{\rho_n^{2d+1}r_n^{-d}}{\log(n)} \to +\infty,$$

since $r_n = o(\rho_n^\alpha)$ for any $\alpha > 0$ and $r_n \to 0$ polynomially in n, we deduce that, for all $\varepsilon > 0$,

$$\sum_n \mathbb{P}\left[\left[\inf_{R\in\mathcal{R}_n} \zeta_n(R) < -\varepsilon\right] \cap \Omega_n\right] < \infty. \tag{3.7}$$

Bounding $\xi_n(R)$. (We reset the constants, except for C_1.) By the perimeter bound of Lemma 14, we have

$$\nu_n^*(R) \leq \frac{\mathrm{Vol}_{d-1}(\partial R)}{\tau_M} \leq d\frac{\mathrm{Vol}_d(R)}{\tau_M\rho_n} = C_2/\rho_n,$$

for a constant $C_2 > 0$ depending only on M. So, together with (3.6) and the fact that, on Ω_n, $\lambda_n(R) \geq \lambda_n^*(R)/2$,

$$\frac{\nu_n^*(R)}{\lambda_n(R)\lambda_n^*(R)} \leq C_3 \rho_n^{-2d-1},$$

for all R in $\mathcal{R}_n$. It follows that

$$I_2 := \mathbb{P}\left[\left[\inf_{R\in\mathcal{R}_n} \xi_n(R) < -\varepsilon\right] \cap \Omega_n\right] \leq \mathbb{P}\left(\sup_{R\in\mathcal{R}_n} \left|\lambda_n(R) - \lambda_n^*(R)\right| > \frac{\rho_n^{2d+1}\varepsilon}{C_3}\right). \tag{3.8}$$

Define

$$\mu_n^*(R) = \frac{\mathrm{Vol}_d(R \cap M_{r_n})}{\tau_M}.$$

Then

$$\begin{aligned} \left|\lambda_n(R) - \lambda_n^*(R)\right| &\leq \left|\mu_n(R) - \mu_n^*(R)\right| + \left|\mu_n(R^c) - \mu_n^*(R^c)\right| \\ &\leq |\mu_n(R) - \mu(R)| + |\mu_n(R^c) - \mu(R^c)| + 2\mu(M_{r_n}^c), \end{aligned}$$

with $\mu(M_{r_n}^c) \leq C_4 r_n$ by (2.4). For $\varepsilon > 0$ fixed and n large enough, $2C_4 r_n \leq \rho_n^{2d+1}\varepsilon/C_3$, again by the fact that $\rho_n \to 0$ sub-polynomially in r_n. We therefore obtain that

$$I_2 \leq 2\mathbb{P}\left(\sup_{R\in\mathcal{R}_{\rho n}} |\mu_n(R) - \mu(R)| > \frac{\rho_n^{2d+1}\varepsilon}{4C_3}\right),$$

where we used the fact that $R^c \in \mathcal{R}_n$ when $R \in \mathcal{R}_n$, together with $\mathcal{R}_n \subset \mathcal{R}_{\rho_n}$. We then apply Proposition 15, whose conditions are satisfied for $\varepsilon > 0$ fixed and n large enough, again because $\rho_n \to 0$ very slowly, arriving at

$$I_2 \leq C_4 \exp\left(-\frac{n r_n^d \rho_n^{4d+2}\varepsilon^2}{C_4(1+\varepsilon)}\right),$$

for some constant $C_4 > 0$ and all n large enough. As before, when ε is fixed, the exponent is a positive power of n, so that

$$\sum_n \mathbb{P}\left[\left[\inf_{R\in\mathcal{R}_n} \xi_n(R) < -\varepsilon\right] \cap \Omega_n\right] < \infty. \tag{3.9}$$

Bounding $\mathbb{P}[\Omega_n^c]$. Since $\lambda_n^*(R) > C\rho_n^d$ for some C uniformly over $R \in \mathcal{R}_n$ (see (3.6) above), we have

$$\begin{aligned} \mathbb{P}(\Omega_n^c) &= \mathbb{P}\left(\sup_{R\in\mathcal{R}_n} \frac{|\lambda_n(R) - \lambda_n^*(R)|}{\lambda_n^*(R)} > \frac{1}{2}\right) \\ &\leq \mathbb{P}\left(\sup_{R\in\mathcal{R}_n} \left|\lambda_n(R) - \lambda_n^*(R)\right| > C\rho_n^d\right). \end{aligned}$$

We then proceed as in bounding (3.8), obtaining

$$\sum_n \mathbb{P}(\Omega_n^c) < \infty. \tag{3.10}$$

Conclusion. We have

$$\mathbb{P}\left[\inf_{R\in\mathcal{R}_n}(h_n(R)-h(R;M_{r_n}))<-2\varepsilon\right]\le\mathbb{P}\left[\left[\inf_{R\in\mathcal{R}_n}\zeta_n(R)<-\varepsilon\right]\cap\Omega_n\right]+\mathbb{P}\left[\left[\inf_{R\in\mathcal{R}_n}\xi_n(R)<-\varepsilon\right]\cap\Omega_n\right]+\mathbb{P}\left[\Omega_n^c\right],$$

so that the left-hand side is summable. Therefore, we conclude by applying the Borel-Cantelli lemma. □

3.5 Some continuity of the Cheeger constant

Our proof of Theorem 2 relies on continuity properties of the normalized cut and of the Cheeger constant. Lemma 18 below compares the conductance function on M and on a bi-Lipschitz deformation of M. For a Lipschitz map f, let $\|f\|_{\mathrm{Lip}}$ denote its Lipschitz constant. If f is bi-Lipschitz, we define its condition number by $\mathrm{cond}(f) := \|f\|_{\mathrm{Lip}}\,\|f^{-1}\|_{\mathrm{Lip}}$. Lemma 19 below states that M_r is a bi-Lipschitz deformation of M, hence Lemma 18 yields the continuity property of Proposition 20.

Lemma 18. *Let f be a bi-Lipschitz on M. Then for any $A\subset M$ measurable,*

$$\max\left\{\frac{h(f(A);f(M))}{h(A;M)},\frac{h(A;M)}{h(f(A);f(M))}\right\}\le\mathrm{cond}(f)^d.$$

Proof. For any $A\subset M$, $\partial f(A)=f(\partial A)$ and $f(A)^c\cap f(M)=f(A^c\cap M)$, and if A is measurable, for $k=1,\dots,d$,

$$\|f^{-1}\|_{\mathrm{Lip}}^{-k}\,\mathrm{Vol}_k(A)\le\mathrm{Vol}_k(f(A))\le\|f\|_{\mathrm{Lip}}^{k}\,\mathrm{Vol}_k(A).$$

Therefore,

$$\begin{aligned}h(f(A);f(M)) &= \frac{\mathrm{Vol}_{d-1}(f(\partial A\cap M))}{\min\{\mathrm{Vol}_d(f(A)),\mathrm{Vol}_d(f(A^c\cap M))\}}\\ &\le \frac{\|f\|_{\mathrm{Lip}}^{d-1}\,\mathrm{Vol}_{d-1}(\partial A\cap M)}{\|f^{-1}\|_{\mathrm{Lip}}^{-d}\min\{\mathrm{Vol}_d(A),\mathrm{Vol}_d(A^c\cap M)\}}\\ &\le \mathrm{cond}(f)^d\,h(A;M).\end{aligned}$$

And vice-versa. □

Lemma 19. *Fix $r<s\le\mathrm{reach}(\partial M)$. Then there is a bi-Lipschitz map between M_r and M that leaves M_s unchanged, and with condition number at most $(1+2r/(s-r))^2$.*

Proof. For x in M such that $\delta(x):=\mathrm{dist}(x,\partial M)<s$, let $\xi(x)\in M$ be its metric projection onto ∂M and u_x be the unit normal vector of M at $\xi(x)$ pointing outwards. We define the map

$$f_r: M_r\mapsto M,\qquad f_r(x)=x+\frac{r(s-\delta(x))_+}{s-r}\,u_x,$$

where a_+ denotes the positive part of $a\in\mathbb{R}$. By construction, f is one-to-one, with inverse

$$f_r^{-1}: M\mapsto M_r,\qquad f_r^{-1}(x)=x-\frac{r(s-\delta(x))_+}{s}\,u_x.$$

By [20, Thm. 4.8(1)], δ is Lipschitz with constant at most 1, therefore so is $x \mapsto (s - \delta(x))_+$; and since the reach bounds the radius of curvature from below [20, Thm. 4.18], $x \mapsto u_x$ is Lipschitz with constant at most $1/\operatorname{reach}(\partial M)$. Therefore, using the fact that $(s - \delta(x))_+ \leq s$ and $\|u_x\| = 1$, f_r and f_r^{-1} are Lipschitz with constants at most $1 + 2r/(s - r)$ and $1 + 2r/s$ respectively. □

Proposition 20. *We have*

$$H(M_r) = (1 + O(r))\, H(M), \quad r \to 0.$$

Proof. From Lemmas 18 and 19, we deduce that

$$\max\left\{\frac{H(M_r)}{H(M)}, \frac{H(M)}{H(M_r)}\right\} \leq (1 + 2r/(\rho_M - r))^{2d},$$

for any $r < \rho_M := \operatorname{reach}(\partial M)$, which immediately yields the desired result. □

3.6 L^1-metric on Borel sets

We will use the L^1-metric on Borel subsets of $\mathbb{R}^d$, defined by $\mathrm{Vol}_d(A \Delta B) = \int |\mathbf{1}_A(x) - \mathbf{1}_B(x)|\, \mathrm{d}x$. This metric comes from the bijection between Borel sets A and their indicator functions $\mathbf{1}_A$, endowed with the L^1-topology. Strictly speaking, this is a semi-metric on Borel subsets of $\mathbb{R}^d$ since $\mathrm{Vol}_d(A \Delta B) = 0$ if and only if $A \Delta B$ is a null set.

The following propositions are adapted from [23, Thm. 2.3.10] and [23, Prop. 2.3.6] respectively. Proposition 21 is a compactness criterion, and Proposition 22 results from lower semi-continuity of the perimeter measure with respect to L^1-metric.

Proposition 21. *Let (E_n) be a sequence of measurable subsets of M. Suppose that*

$$\limsup_{n \to \infty} \mathrm{Vol}_{d-1}(\partial E_n \cap M) < \infty.$$

Then (E_n) admits a subsequence converging for the L^1-metric.

Proposition 22. *Let E_n and E be bounded measurable subsets of M such that $E_n \to E$ in L^1, and $h(E; M) < \infty$. Then*

$$\liminf_n h(E_n; M) \geq h(E; M).$$

3.7 Proof of (i) in Theorem 2

Lower bound. For each n, let $R_n \in \mathcal{R}_n$ be such that

$$h_n^{\ddagger}(R_n) = \min_{R \in \mathcal{R}_n} h_n^{\ddagger}(R).$$

Then

$$\begin{aligned} h_n^{\ddagger}(R_n) - H(M) &= \left[h_n^{\ddagger}(R_n) - h(R_n; M_{r_n})\right] + \left[h(R_n; M_{r_n}) - H(M_{r_n})\right] + \left[H(M_{r_n}) - H(M)\right] \\ &\geq \inf_{R \in \mathcal{R}_n} \left(h_n(R) - h(R; M_{r_n})\right) + \left[H(M_{r_n}) - H(M)\right], \end{aligned}$$

since $[h(R_n; M_{r_n}) - H(M_{r_n})] \geq 0$ by definition of $H(M_{r_n})$. On the last line, by Lemma 17, the first term has a non-negative inferior limit, and by Proposition 20, the second term tends to zero. Hence,

$$\liminf_{n\to\infty} \min_{R\in\mathcal{R}_n} h_n^{\ddagger}(R) \geq H(M) \quad \text{a.s.} \tag{3.11}$$

Upper bound. To obtain the matching upper bound, fix a subset $A \subset M$ with smooth relative boundary and such that $0 < \mathrm{Vol}_d(A) \leq \mathrm{Vol}_d(M\backslash A) < \mathrm{Vol}_d(M)$. Then, for n large enough, there exists R_n in $\mathcal{R}_n$ such that $R_n \cap M = A$, implying that

$$\min_{R\in\mathcal{R}_n} h_n^{\ddagger}(R) \leq h_n(A).$$

By Theorem 1, $h_n(A) \to h(A; M)$ almost surely, so that

$$\limsup_{n\to\infty} \min_{R\in\mathcal{R}_n} h_n^{\ddagger}(R) \leq h(A; M) \quad \text{a.s.}$$

By minimizing over A, we obtain

$$\limsup_{n\to\infty} \min_{R\in\mathcal{R}_n} h_n^{\ddagger}(R) \leq H(M) \quad \text{a.s.} \tag{3.12}$$

Combining the lower and upper bounds, (3.11) and (3.12), we conclude that

$$\lim_{n\to\infty} \min_{R\in\mathcal{R}_n} h_n^{\ddagger}(R) = H(M) \quad \text{a.s.} \tag{3.13}$$

3.8 Proof of (*ii*) in Theorem 2

Let R_n be a sequence in $\mathcal{R}_n$ satisfying

$$h_n^{\ddagger}(R_n) = \min_{R\in\mathcal{R}_n} h_n^{\ddagger}(R),$$

and set $A_n = R_n \cap M$. Fix a subset $A^0 \subset M$ with smooth relative boundary and such that $h(A^0) < \infty$. Then for n large enough, there exists R in $\mathcal{R}_n$ such that $A^0 = R \cap M$. Hence $h_n(A_n) \leq h_n(A^0)$ and since $h_n(A^0) \to h(A^0)$ by Theorem 1, we have

$$\limsup_{n\to\infty} \mathrm{Vol}_{d-1}(A_n) \leq \limsup_{n\to\infty} h(A_n) \min\{\mathrm{Vol}_d(A_n), \mathrm{Vol}_d(A_n^c \cap M)\} \leq h(A^0)\, \mathrm{Vol}_d(M)/2.$$

Therefore by compactness of the class of sets with bounded perimeters (Proposition 21), with probability one, $\{A_n\}$ admits a subsequence converging in the L^1-metric.

On the one hand,

$$h(A_n; M_{r_n}) - H(M) = [h(A_n; M_{r_n}) - H(M_{r_n})] + [H(M_{r_n}) - H(M)],$$

where the first difference term on the right-hand side is non-negative by definition, while the second difference term tends to zero by Proposition 20. So that with probability one:

$$\liminf_{n\to\infty} h(A_n; M_{r_n}) \geq H(M).$$

On the other hand,

$$\begin{aligned} h(A_n; M_{r_n}) - H(M) &= \left[h(A_n; M_{r_n}) - h_n^{\ddagger}(A_n)\right] + \left[h_n^{\ddagger}(A_n) - H(M)\right] \\ &\leq -\inf_{R\in\mathcal{R}_n}\left(h_n^{\ddagger}(R) - h(R; M_{r_n})\right) + \left[h_n^{\ddagger}(A_n) - H(M)\right] \end{aligned}$$

so that

$$\limsup_{n\to\infty} h(A_n; M_{r_n}) - H(M) \leq -\liminf_{n\to\infty}\inf_{R\in\mathcal{R}_n}\left(h_n^{\ddagger}(R) - h(R; M_{r_n})\right) + \left[h_n^{\ddagger}(A_n) - H(M)\right]$$

which goes to 0 as $n \to \infty$ from (3.5) and (3.13). Hence

$$\lim_{n\to\infty} h(A_n; M_{r_n}) \to H(M) \quad \text{a.s.}$$

Now let f_n denote the bi-Lipschitz function mapping M_{r_n} to M defined in Lemma 19 with r and s replaced by r_n and s_n, where $s_n/r_n \to \infty$. Define $B_n = f_n(A_n \cap M_{r_n})$. By Lemmas 18 and 19, we have

$$h(B_n; M) \leq \left(1 + \frac{2r_n}{s_n - r_n}\right)^{2d} h(A_n; M_{r_n}),$$

so that $h(B_n; M) \to H(M)$ almost surely as $n \to \infty$. Moreover, by Proposition 21, with probability one, there exists a subset B_∞ of M and a subsequence $\{B_{n_k}\}$ such that B_{n_k} converges to B_∞ in the L^1-metric. Since $h(\cdot; M)$ is lower-semi-continuous by Proposition 22, with probability one, $\liminf_{n\to\infty} h(B_n; M) \geq h(B_\infty; M)$. Since we also have $\liminf_{n\to\infty} h(B_n; M) = H(M)$ a.s., it follows that $h(B_\infty; M) = H(M)$ a.s. and so B_∞ is a Cheeger set of M.

Moreover, since f_n leaves M_{s_n} unchanged,

$$\mathrm{Vol}_d(A_n \Delta B_n) \leq \mathrm{Vol}_d(M \backslash M_{s_n}) \to 0 \quad \text{as } n \to \infty.$$

Hence with probability one, $\mathbf{1}_{A_n} - \mathbf{1}_{B_n} \to 0$ in L^1. Consequently, the sequences $\{A_n\}$ and $\{B_n\}$ have the same accumulation points, and so any convergent subsequence of $\{A_n\}$ converges to a Cheeger set of M.

3.9 Proof of Theorem 3

Let $A_n = R_n \cap M$ and assume, without loss of generality, that $A_n \to A_\infty$ in L^1. For all $n \geq 1$, and all f in the class of bounded and continuous functions on M, say $C_b(M)$, we have

$$\left|Q_n f - \int_M f(x)\mathbf{1}_{R_n}(x)\mu(\mathrm{d}x)\right| \leq \sup_{R\in\mathcal{R}_n} |P_n(f\mathbf{1}_R) - \mu(f\mathbf{1}_R)|,$$

where P_n is the empirical measure of the sample $X_1, \ldots, X_n$. Using the bound on the covering numbers in Lemma 13, it is a classical exercise to prove that the collection of functions $x \mapsto f(x)\mathbf{1}_R(x)$ where R ranges over $\mathcal{R}_n$ is a Glivenko-Cantelli class, whence

$$\left| Q_n f - \int_M f(x)\mathbf{1}_{R_n}(x)\mu(\mathrm{d}x) \right| \to 0 \quad \text{a.s. as } n \to \infty.$$

Next,

$$\left| \int_M f(x)\mathbf{1}_{R_n}(x)\mu(\mathrm{d}x) - Qf \right| = \left| \int_M f(x)\mathbf{1}_{A_n}(x)\mu(\mathrm{d}x) - Qf \right| \leq \|f\|_\infty \mu\left(A_n \Delta A_\infty\right),$$

which tends to 0 by definition of A_∞. Thus, we have shown that, for all f in $C_b(M)$, $\mathbb{P}\left(Q_n f \to Qf\right) = 1$. Using the separability of $C_b(M)$ [19, p. 131], we deduce that

$$\mathbb{P}[\forall f \in C_b(M),\ Q_n f \to Qf] = 1,$$

so that the event "Q_n converge weakly to Q" is of probability 1.

Acknowledgments

Ery Arias-Castro was partially supported by a grant from the US National Science Foundation (DMS-06-03890) and wish to thank Lei Ni for stimulating discussions. Bruno Pelletier and Pierre Pudlo were supported by the French National Research Agency (ANR) under grant ANR-09-BLAN-0051-01. Bruno Pelletier also thanks Michel Pierre for discussions related to Cheeger sets.

References

[1] S. Arora, E. Hazan, and S. Kale. $O(\sqrt{\log n})$-approximation to sparsest cut in $\tilde{O}(n^2)$ time. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 238–247, Washington, DC, USA, 2004. IEEE Computer Society.

[2] C. Avin and G. Ercal. On the cover time and mixing time of random geometric graphs. *Theor. Comput. Sci.*, 380(1-2):2–22, 2007.

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in neural information processing systems*, 1:585–592, 2002.

[4] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. System Sci.*, 74(8):1289–1308, 2008.

[5] G. Biau, B. Cadre, and B. Pelletier. A graph-based estimator of the number of clusters. *ESAIM Probab. Stat.*, 11:272–280, 2007.

[6] G. Biau, B. Cadre, and B. Pelletier. Exact rates in density support estimation. *Journal of Multivariate Analysis*, 99(10):2185–2207, 2008.

[7] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Mixing times for random walks on geometric random graphs. In C. Demetrescu, R. Sedgewick, and R. Tamassia, editors, *SIAM Workshop on Analytic Algorithmics & Combinatorics (ANALCO)*, pages 240–249. SIAM, 2005.

[8] H. Bräker and T. Hsing. On the area and perimeter of a random convex hull in a bounded convex set. *Probab. Theory Related Fields*, 111(4):517–550, 1998.

[9] P. Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)*, 15(2):213–230, 1982.

[10] G. Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[11] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete Comput. Geom.*, 42(1):71–93, 2009.

[12] V. Caselles, A. Chambolle, and M. Novaga. Some remarks on uniqueness and regularity of Cheeger sets. *Rend. Sem. Mat. Univ. Padova*, 2009.

[13] F. Chazal, L. Guibas, S. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1021–1030. Society for Industrial and Applied Mathematics, 2009.

[14] F. Chazal and A. Lieutier. Weak feature size and persistant homology: computing homology of solids in $\mathbb{R}^n$ from noisy data samples. In *Computational geometry (SCG'05)*, pages 255–262. ACM, New York, 2005.

[15] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.

[16] F. R. K. Chung. *Spectral graph theory*. Number 92 in Regional Conference Series in Mathematics. Amer. Math. Soc., Providence, 1997.

[17] A. Cuevas, R. Fraiman, and A. Rodríguez-Casal. A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.*, 35(3):1031–1051, 2007.

[18] V. H. de la Peña and E. Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

[19] J. L. Doob. *Measure theory*, volume 143 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.

[20] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[21] E. Giné and V. Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 238–259. Inst. Math. Statist., Beachwood, OH, 2006.

[22] A. Gray. *Tubes*, volume 221 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, second edition, 2004. With a preface by Vicente Miquel.

[23] A. Henrot and M. Pierre. *Variation et optimisation de formes*, volume 48 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Berlin, 2005. Une analyse géométrique. [A geometric analysis].

[24] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[25] A. N. Kolmogorov and V. M. Tikhomirov. ε-entropy and ε-capacity of sets in functional space. *Amer. Math. Soc. Transl. (2)*, 17:277–364, 1961.

[26] E. Levina and P. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, volume 17, pages 777–784. MIT Press, Cambridge, Massachusetts, 2005.

[27] M. Maier, U. Von Luxburg, and M. Hein. Influence of graph construction on graph-based clustering measures. *Advances in neural information processing systems (NIPS)*, 22:1025–1032, 2009.

[28] H. Narayanan, M. Belkin, and P. Niyogi. On the relation between low density separation, spectral clustering and graph cuts. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 19. MIT Press, Cambridge, MA, 2007.

[29] H. Narayanan and P. Niyogi. On the sample complexity of learning smooth cuts on a manifold. In *22nd Annual Conference on Learning Theory (COLT)*, 2009.

[30] A. Y. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems (NIPS)*, volume 14, pages 849–856, 2001.

[31] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.

[32] B. Pelletier and P. Pudlo. Operator norm convergence of spectral clustering on level sets. *Journal of Machine Learning Research*, 12:385–416, 2011.

[33] M. Penrose. *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[34] V. Robins. Towards computing homology from finite approximations. In *Proceedings of the 14th Summer Conference on General Topology and its Applications (Brookville, NY, 1999)*, volume 24, pages 503–532, 1999.

[35] A. Singer. From graph to manifold Laplacian: the convergence rate. *Appl. Comput. Harmon. Anal.*, 21(1):128–134, 2006.

[36] D. A. Spielman and S.-H. Teng. Spectral partitioning works: planar graphs and finite element meshes. *Linear Algebra Appl.*, 421(2-3):284–305, 2007.

[37] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[38] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

[39] G. Walther. Granulometric smoothing. *Ann. Statist.*, 25(6):2273–2299, 1997.

[40] H. Weyl. On the volume of tubes. *Amer. J. Math.*, 61(2):461–472, 1939.

[41] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.