

THE NORMALIZED GRAPH CUT AND CHEEGER CONSTANT: FROM DISCRETE TO CONTINUOUS

ERY ARIAS-CASTRO,* *University of California*

BRUNO PELLETIER,** *Université Rennes II*

PIERRE PUDLO,*** *Université Montpellier II*

Abstract

Let M be a bounded domain of \mathbb{R}^d with a 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 .

Keywords: Cheeger isoperimetric constant of a manifold; conductance of a graph; neighborhood graph; spectral clustering; U -process; empirical process

2010 Mathematics Subject Classification: Primary 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

$$h(S; G) = \frac{\sigma(S)}{\min\{\delta(S), \delta(S^c)\}}, \quad (1)$$

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

$$\delta(S) = \sum_{i \in S} \sum_{j \neq i} \delta_{ij}, \quad \sigma(S) = \sum_{i \in S} \sum_{j \in S^c} \delta_{ij} \quad (2)$$

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 nonempty subsets of G , i.e.

$$H(G) = \min\{h(S; G) : S \subset G, S \neq \emptyset\}. \quad (3)$$

A corresponding quantity can be defined for a domain of a Euclidean space. Let M be a bounded domain (i.e. an open, connected subset) of \mathbb{R}^d with a smooth boundary ∂M of class at least C^2 .

Received 16 June 2011; revision received 17 April 2012.

* Postal address: Department of Mathematics, University of California, San Diego, USA.

** Postal address: Département de Mathématiques, IRMAR – UMR CNRS 6625, Université Rennes II, France.

Email address: bruno.pelletier@univ-rennes2.fr

*** Postal address: Département de Mathématiques, I3M – UMR CNRS 5149, Université Montpellier II, France.

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 $A \subset \mathbb{R}^d$, we denote by $P_M(A)$ its de Giorgi perimeter (see [25, Definition 2.3.1] or [23, Definition 1.6]), which is equal to $\text{Vol}_{d-1}(\partial A \cap M)$ when A is a (bounded) open subset of M of class C^1 [25, Proposition 2.3.3], and we define its normalized cut with respect to M by

$$h(A; M) = \frac{P_M(A)}{\min\{\text{Vol}_d(A \cap M), \text{Vol}_d(A^c \cap M)\}},$$

where A^c denotes the complement of A in \mathbb{R}^d and with the convention that $0/0 = \infty$. So, when A is a C^1 open subset of M , we also have

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

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 codimension 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 independent and identically distributed (i.i.d.) random sample $\mathcal{X}_n = (X_1, \dots, X_n)$ 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 [36]. Let ω_d denote the d -volume of the unit d -dimensional ball, and define

$$\gamma_d = \int_{\mathbb{R}^d} \max(\langle u, z \rangle, 0) \mathbf{1}\{\|z\| \leq 1\} dz, \tag{4}$$

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. The volume $V(h, d)$ of a spherical cap of the d -dimensional unit ball at height $0 \leq h \leq 1$ is given by

$$V(h, d) = \frac{\pi^{(d-1)/2}}{\Gamma((d+1)/2)} \int_0^{\arccos(1-h)} \sin^d(t) dt,$$

which leads to

$$\gamma_d = \frac{\pi^{(d-1)/2}}{(d+1)\Gamma((d+1)/2)}.$$

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 $r_n \rightarrow 0$ with $nr_n^{d+1} / \log n \rightarrow +\infty$, and let $S_n = A \cap G_{n,r_n}$. Then, with probability 1,*

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

and, consequently,

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

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 (3) to subsets associated with subsets of \mathbb{R}^d with controlled reach. The reach of a subset $S \subset \mathbb{R}^d$ [21], denoted by $\text{reach}(S)$, 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 leeway 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. We also impose a regularity condition on the boundary of M , namely that ∂M is of class at least C^2 . In particular, the differentiability of ∂M implies that the reach of ∂M is strictly positive.

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

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

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 1,

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

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

$$R_n \in \mathcal{R}_n, \quad h_n^\ddagger(R_n) = \min\{h_n^\ddagger(R) : R \in \mathcal{R}_n\}. \tag{5}$$

Then, with probability 1, $\{R_n \cap M\}$ 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 (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 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), and let $\{R_{n_k}\}$ be a subsequence of $\{R_n\}$ with $R_{n_k} \cap M \rightarrow A_\infty$ in L^1 . Define the random discrete measure $Q_n = (1/n) \sum_{i=1}^n \mathbf{1}_{R_n}(X_i) \delta_{X_i}$ and the measure $Q = \mathbf{1}_{A_\infty}(\cdot)\mu$. Then, that Q_n converges weakly to Q is an event which holds with probability 1.*

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; see, e.g. [9].

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 are aware. 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.

1.3.1. Clustering. In spectral graph partitioning, the goal is to partition a graph G into subgraphs based on the eigenvalues and eigenvectors of the Laplacian [16], [39]. 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, \dots, 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 [40]. 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 $\psi(t) = \mathbf{1}_{[0,1]}(t)$ [3], [33]. The consistency of spectral methods has been analyzed in this context [4], [22], [35], [38], [41]. In particular, Narayanan *et al.* [32] proved a result similar to our Theorem 1 in that context.

About cuts, Maier *et al.* [30] also proved a result similar to our Theorem 1 when the separating surface ∂A is an affine hyperplane. Closer to our Theorem 2, Narayanan and Niyogi [31] established 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 .

1.3.2. 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 [29] and, more generally, the homology [10], [11], [13], [14], [34], [37], [44]; the Minkowski content [17], as well as the perimeter and area (volume) [8]. In the

related field of stochastic geometry, Khmaladze and Weil [27] established limit properties of Poisson point processes in the context of the change-set problem, where the Poisson process has two homogeneous components, with different intensities inside and outside of an unknown convex compact subset of \mathbb{R}^d . Their proof techniques involve integration over tubular neighborhoods similar to those that we use in the present paper.

1.3.3. 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 the 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 of the random walk on G . This is the path taken in [2] and [7] when M is the unit hypercube and the graph is $G_{r_n, n}$. However, in both papers the authors reduced 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

From the above discussion, we see that 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 were sampled from. To the best of the authors' knowledge, our paper is the first to establish a relationship between the Cheeger constant (optimal normalized cut) on the neighborhood graph and the Cheeger constant of the domain, and the first 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 amounts of effort to other related settings. However, we leave important questions behind.

1.4.1. 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 nonvanishing 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, Lemma 3.4].

1.4.2. 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 $nr_n^d \rightarrow \infty$ is necessary, and may be sufficient if 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.

1.4.3. An open problem. An intriguing question is whether along some sequence of neighborhood graphs, the normalized Cheeger constant converges to the Cheeger constant of the domain. To paraphrase the question we leave open: is there a sequence $\{r_n\}$ such that, with probability 1,

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

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 eigenvalues 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 [2] and [7] when M is the unit hypercube.

1.4.4. *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 [26]. Then, by the lower semicontinuity 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 [25] on the other hand.

1.6. Notation and background

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

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

where δ and σ are given in (2), \mathcal{X}_n is the sample, and G_{n,r_n} is 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). For further reference, we define the volume $\pi_d(\eta)$ of a spherical cap at height η by

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

where u is any unit-norm vector of \mathbb{R}^d . Note that the constant γ_d defined in (4) may be expressed as

$$\gamma_d = \int_0^1 \pi_d(\eta) \, d\eta.$$

The reach coincides with the condition number introduced in [34] for submanifolds without boundary, and the property $\text{reach}(\partial A) \geq r$ is equivalent to A and A^c being both r -convex [42], 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 [21, Theorem 4.18]. In particular, if $\text{reach}(\partial A) > 0$ 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

$$\phi_{A,r}(x, y) = \frac{1}{2} \{ \mathbf{1}_A(x) + \mathbf{1}_A(y) \} \mathbf{1} \{ \|x - y\| \leq r \}, \tag{6}$$

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

$$\bar{\phi}_{A,r}(x, y) = \frac{1}{2} \{ \mathbf{1}_A(x) \mathbf{1}_{A^c}(y) + \mathbf{1}_A(y) \mathbf{1}_{A^c}(x) \} \mathbf{1} \{ \|x - y\| \leq r \}. \tag{7}$$

We will need the following Hoeffding inequality for U -statistics [26], which is a special case of [18, Theorem 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 $E[\phi(X_1, X_2)] = 0$ and that $b := \|\phi\|_\infty < \infty$, and let $\sigma^2 = \text{var}(\phi(X_1, X_2))$. Then, for all $t > 0$,*

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

To prove Theorem 1, we establish the almost-sure convergences 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 end, we combine upper bounds on bias terms 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

$$M_r = \{x \in M : \text{dist}(x, \partial M) > r\}. \tag{8}$$

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

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

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

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

Proof. For ease of notation, set $B = B(p + \eta r e_p, r)$. Let $(\tilde{e}_1, \dots, \tilde{e}_d)$ be an orthonormal frame at p , with $\tilde{e}_d = e_p$. Denote by $\tilde{x}_1, \dots, \tilde{x}_d$ 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 $\tilde{x}^d = F(\tilde{x}^1, \dots, \tilde{x}^{d-1})$ for some function F , and, if we set $\tilde{x}^{(d)} = (\tilde{x}^1, \dots, \tilde{x}^{d-1})$ then

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

Since

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

it follows that

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

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 $1/\rho$ [21], 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

$$\begin{aligned} |\text{Vol}_d(B(p + \eta r e_p, r) \cap A^c) - \pi_d(\eta)r_n^d| &\leq 2\omega_{d-1} \sup_{p \in \partial A \cap M} \|G(p)\| r^{d+1} \\ &\leq \frac{2\omega_{d-1}r^{d+1}}{\rho}. \end{aligned}$$

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

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

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

2.2. Integration over tubes

We introduce the notion of tubes and some of their properties; see [24] for an extensive treatment. Let S be a submanifold of \mathbb{R}^d . The *tubular neighborhood* of radius $r > 0$ about S , denoted by $\mathcal{V}(S, r)$, 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 a 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 a boundary then the tube coincides with the set of points at a distance no more than r , with the ends removed, corresponding to the points projecting onto ∂S . Assume that S is of codimension 1, and oriented, and define e_p as the (unit) normal vector of S at $p \in S$. When $r < \text{reach}(S)$, $\mathcal{V}(S, r)$ admits the following parameterization:

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

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

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

where v_σ is the Riemannian volume measure on S .

Lemma 3. *Assume that S is a submanifold of \mathbb{R}^d of codimension 1, with $\rho := \text{reach}(S) > 0$. Then, for all $r < \rho$,*

$$\sup_{p \in S} \sup_{-r \leq t \leq r} \vartheta(p, t) \leq \left(1 + \frac{r}{\rho}\right)^{d-1}$$

and

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

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

Proof. By [21, Theorem 4.18], the reach bounds the radius of curvature from below so that the principal curvatures $\kappa^{(1)}, \dots, \kappa^{(d-1)}$ (the eigenvalues of the second fundamental form) are everywhere bounded (in absolute value) from above by $1/\rho$. Therefore, for $r < \rho$ and $-r \leq t \leq r$,

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

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,

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

The celebrated Weyl’s tube formula [43] 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 4. *For any bounded open subset $R \subset \mathbb{R}^d$ with $\text{reach}(\partial R) = \rho > 0$ and any $0 < r < \rho$,*

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

In particular, Lemma 4 implies that

$$\mu[\mathcal{V}(\partial M, r)] \leq Cr \quad \text{for all } r < \text{reach}(\partial M), \tag{9}$$

where C is a constant depending only on M .

Proof of Lemma 4. Using the uniform bound of the infinitesimal change of volume given in Lemma 3, we have

$$\begin{aligned} \text{Vol}_d(\mathcal{V}(\partial R, r)) &= \int_{\partial R} \int_{-r}^r \vartheta(p, u) \, du v_\sigma(dp) \\ &\leq \text{Vol}_{d-1}(\partial R)2r \left(1 + \frac{r}{\rho}\right)^{d-1} \\ &\leq 2^d \text{Vol}_{d-1}(\partial R)r. \end{aligned}$$

2.3. Bounds on bias terms

Recall the definition of M_r in (8).

Lemma 5. *Let $\phi_{A,r}$ be defined as in (6). There exists a constant C depending only on M such that, for any $A \subset M$ and $r < \text{reach}(\partial M)$,*

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

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

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

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

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

This completes the proof.

Lemma 6. *Let $A = R \cap M$, where R is a bounded domain with smooth boundary and $\text{reach}(\partial R) = \rho > 0$. Let $\bar{\phi}_{A,r}$ be defined as in (7).*

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

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

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

$$\frac{\tau_M}{\gamma_d r^{d+1}} E[\bar{\phi}_{A,r}(X_1, X_2)] - \frac{\text{Vol}_{d-1}(\partial A \cap M_r)}{\text{Vol}_d(M)} \geq -C v(A) \frac{r}{\rho}.$$

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

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

where

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

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 $p \in \partial R$, 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

$$\begin{aligned} \int_{D \cap M_r} \text{Vol}_d(B(x, r) \cap A^c) dx &= \int_{S_r} \int_{-r}^0 \text{Vol}_d(B(p + te_p, r) \cap A^c) \vartheta(p, t) dt v_\sigma(dp) \\ &= r \int_{S_r} \int_0^1 \text{Vol}_d(B(p - \eta r e_p, r) \cap A^c) \vartheta(p, r\eta) d\eta v_\sigma(dp). \end{aligned}$$

Therefore,

$$\begin{aligned} &\left| \frac{1}{r^{d+1}} \int_{D \cap M_r} \text{Vol}_d(B(x, r) \cap A^c) dx - \gamma_d v(A) \right| \\ &\leq \frac{1}{r^d} \int_{S_r} \int_0^1 |\text{Vol}_d(B(p - \eta r e_p, r) \cap A^c) - \pi_d(\eta)r^d| \vartheta(p, r\eta) d\eta v_\sigma(dp) \\ &\quad + \left| \int_{S_r} \int_0^1 \pi_d(\eta) \vartheta(p, r\eta) d\eta v_\sigma(dp) - \gamma_d v(A) \right|. \end{aligned} \tag{10}$$

Lemma 1 provides the inequality $|\text{Vol}_d(B(p - \eta r e_p, r) \cap A^c) - \pi_d(\eta)r^d| \leq 2\omega_{d-1}r^{d+1}/\rho$, and the first inequality of Lemma 3 states that $\sup_{p \in S} \sup_{-r \leq t \leq r} \vartheta(p, t) \leq (1 + r/\rho)^{d-1}$. 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

$$2\omega_{d-1} \left(\frac{r}{\rho}\right) \int_{S_r} \int_0^1 \vartheta(p, r\eta) d\eta v_\sigma(dp) \leq 2^d \omega_{d-1} \left(\frac{r}{\rho}\right) \text{Vol}_{d-1}(S_r).$$

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 3 states that

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

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 $\gamma_d = \int_0^1 \pi_d(\eta) d\eta$. Then the second term on the right-hand side

of (10) is bounded by

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

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 on the left-hand side of (10) is bounded by

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

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

$$\begin{aligned} & \int_{D \cap M_r^c} \text{Vol}_d(B(x, r) \cap A^c) \, dx \\ & \leq \int_{\partial R \cap \mathcal{V}(\partial M, r)} \int_{-r}^0 \text{Vol}_d(B(p + te_p, r) \cap A^c) \vartheta(p, t) \, dt v_\sigma(\mathbf{d}p) \\ & = r \int_{\partial R \cap \mathcal{V}(\partial M, r)} \int_0^1 \text{Vol}_d(B(p - \eta r e_p, r) \cap A^c) \vartheta(p, r\eta) \, d\eta v_\sigma(\mathbf{d}p) \\ & \leq 2^{d-1} \omega_d r^{d+1} \text{Vol}_{d-1}(\partial R \cap \mathcal{V}(\partial M, r)), \end{aligned}$$

where we have used Lemma 3 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

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

which proves the first bound stated in Lemma 6.

To prove (ii), using the bound on (10), we deduce that

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

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

2.4. Exponential inequalities

Proposition 1. Fix a sequence $r_n \rightarrow 0$. 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 large enough n , we have

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

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

Proof. By the triangle inequality we have

$$|\mu_n(A) - \mu(A)| \leq |\mu_n(A) - E[\mu_n(A)]| + |E[\mu_n(A)] - \mu(A)|.$$

For all n large enough such that $r_n \leq \text{reach}(\partial M)$, the second term on the right-hand side (the bias term) is bounded by Cr_n with C depending only on M . Indeed, Lemma 5 states that the bias is lower than $\mu(A \cap M_{r_n}^c)$. 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 (9).

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

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

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

$$\text{var}(\phi(X_1, X_2)) \leq E[\phi_{A,r_n}(X_1, X_2)^2] \leq \frac{\mu(A)\omega_d r_n^d}{\tau_M} \leq \frac{\omega_d r_n^d}{\tau_M}.$$

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

Proposition 2. Fix a sequence $r_n \rightarrow 0$. 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 all large enough n , we have

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

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

$$v_n(A) \rightarrow v(A) \quad \text{as } n \rightarrow \infty \text{ almost surely.}$$

Proof. By the triangle inequality we have

$$|v_n(A) - v(A)| \leq |v_n(A) - E[v_n(A)]| + |E[v_n(A)] - v(A)|.$$

Using the control on the bias in Lemma 6(i), the second term on the right-hand side goes to 0 as $n \rightarrow \infty$. Then, for large enough n , we apply Hoeffding’s inequality, Theorem 4, to the first term on the right-hand side with the kernel

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

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

$$\text{var}(\phi(X_1, X_2)) \leq E[\bar{\phi}_{A,r_n}(X_1, X_2)^2] = E[\bar{\phi}_{A,r_n}(X_1, X_2)] \leq \frac{2\gamma_d v(A) r_n^{d+1}}{\tau_M},$$

where the last inequality follows from the upper bound on the bias of Lemma 6(i) for large enough n . 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 1 and 2.

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

3. Proofs 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 the discrete volume and discrete perimeter. We exploit this control on the complexity of \mathcal{R}_n in building a covering for \mathcal{R}_n , which is done in Section 3.1, later used to obtain uniform versions of Propositions 1 and 2. Then Theorem 2(i), which states the convergence of a penalized graph Cheeger constant towards the Cheeger constant of M , is proved in Section 3.7. Finally, Theorem 2(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}_ρ be the class of open subsets $R \subset (0, 1)^d$ with $\text{reach}(\partial R) \geq \rho$. Let $d_H(R, R')$ be the Hausdorff distance between two sets R and R' , i.e.

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

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

Lemma 7. (i) *There exists a constant C depending only on d such that, for any $\epsilon > 0$ and any $\rho > 0$,*

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

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

Proof. Let x_1, \dots, x_n be a ϵ -covering of $(0, 1)^d$, so $\bigcup_{i=1}^n B(x_i, \epsilon)$ covers $(0, 1)^d$ and $n \leq C \epsilon^{-d}$ for some constant C depending only on d . For any set R in \mathcal{R}_ρ , define

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

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

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

Therefore,

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

Since, when R ranges in \mathcal{R}_ρ , the cardinality of sets of the form $\bigcup_{i \in I_\varepsilon(R)} B(x_i, \varepsilon)$ is bounded by 2^n , the collection of Hausdorff balls of radius 2ε and centered at sets of the form $\bigcup_{i \in I} B(x_i, \varepsilon)$, where I is any subset of $\{1, \dots, n\}$, covers \mathcal{R}_ρ . By doubling the radius of the balls, we can take centers in \mathcal{R}_ρ , 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}_ρ is rather weak. Standard results by Kolmogorov and Tikhomirov [28] 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. [20]):

$$d\omega_d^{1/d} \text{Vol}_d(R)^{1-1/d} \leq \text{Vol}_{d-1}(\partial R). \tag{11}$$

But, in the case where ∂R has positive reach, the perimeter may in turn be bounded by the volume, as stated in Lemma 8 below. The proof uses the following inequality: for all Borel sets R and S ,

$$\text{Vol}_{d-1}(\partial(R \cup S)) + \text{Vol}_{d-1}(\partial(R \cap S)) \leq \text{Vol}_{d-1}(\partial R) + \text{Vol}_{d-1}(\partial S). \tag{12}$$

Lemma 8. *Let R be a bounded open subset of \mathbb{R}^d with $\text{reach}(\partial R) = \rho > 0$. Then,*

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

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 $R_n = \bigcup_{i=1}^n B_i$, 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

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

and, on the other hand, using inequality (12), we have

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

Consequently,

$$\text{Vol}_{d-1}(\partial R_{n+1}) - \frac{d}{\rho} \text{Vol}_d(R_{n+1}) \leq \text{Vol}_{d-1}(\partial R_n) - \frac{d}{\rho} \text{Vol}_d(R_n) + \left[\frac{d}{\rho} \text{Vol}_d(R_n \cap B_{n+1}) - \text{Vol}_{d-1}(\partial(R_n \cap B_{n+1})) \right].$$

But, using the isoperimetric inequality (11), we may write

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

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

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

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

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

Since R_n converges to R in L^1 , it follows from the lower semicontinuity of the perimeter, see, e.g. [25, Proposition 2.3.6], that $\liminf_n \text{Vol}_{d-1}(\partial R_n) \geq \text{Vol}_{d-1}(\partial R)$. This concludes the proof.

3.3. Exponential inequalities

We prove the uniform versions of Propositions 1 and 2 for the class \mathcal{R}_ρ .

Proposition 3. *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*

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

Proof. The bias term is dealt with exactly as in Proposition 1, yielding

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

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

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

Define the kernel class

$$\mathcal{F} = \{\phi_{R,r} : R \in \mathcal{R}_\rho\}, \tag{13}$$

where $\phi_{R,r}$ is defined in (6). 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

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

Consider a minimal covering of \mathcal{R}_ρ of cardinal K by balls centered at elements R_1, \dots, R_K of \mathcal{R}_ρ , and of radius $\eta < \rho$ for the Hausdorff distance. By Lemma 7,

$$\log(K) \leq C_1 \left(\frac{1}{\eta}\right)^d.$$

For any R in \mathcal{R}_ρ , 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 4, there exists a constant C_2 depending only on the dimension d such that $\text{Vol}_d(\mathcal{V}(\partial R_k, \eta)) \leq C_2 \eta / \rho$ for all $1 \leq k \leq K$, which implies that

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

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

We have

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

Next, consider the inequality

$$\begin{aligned} |U_n(\phi_{R,r}) - \mu^{\otimes 2}(\phi_{R,r})| &\leq |U_n(\phi_{R,r}) - U_n(\phi_{R_k,r})| + |U_n(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R_k,r})| \\ &\quad + |\mu^{\otimes 2}(\phi_{R_k,r}) - \mu^{\otimes 2}(\phi_{R,r})|. \end{aligned}$$

For the double expectations, we have,

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

with C_4 still depending only on M . The last inequality is a consequence of Lemmas 4 and 8, and the fact that $\text{Vol}_d(R_k) \leq 1$ since $R_k \subset (0, 1)^d$.

For the empirical averages, we have

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

Therefore,

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

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

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

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

$$\text{var}(\phi_{\mathcal{V}(\partial R_k, \eta)}(X_1, X_2)) \leq \mathbb{E}[\phi_{\mathcal{V}(\partial R_k, \eta)}(X_1, X_2)^2] \leq \mathbb{E}[\phi_{\mathcal{V}(\partial R_k, \eta)}(X_1, X_2)],$$

with

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

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 Hoeffding’s inequality for U -statistics (Theorem 4), we have

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

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

$$\text{var}(\phi_{R_k,r}(X_1, X_2)) \leq \mathbb{E}[\phi_{R_k,r}(X_1, X_2)] \leq \frac{\omega_d r^d}{\tau_M},$$

we may apply Lemma 4 again to obtain the bound

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

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

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

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}_ρ . 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, \dots, 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 4. *There exists a constant C depending only on M such that, for any $\varepsilon > 0$, $\rho < 1$, $r < \min(\text{reach}(M), \rho/2)$, and all n satisfying $nr^{2d+1} \rho^{d+1} \varepsilon^{d+2} > C$, we have*

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

Proof. The proof follows that of Proposition 3, with the symmetric kernel $\bar{\phi}_{R,r}$ defined in (7) and the class $\bar{\mathcal{F}}$ defined in (13) with $\phi_{R,r}$ replaced by $\bar{\phi}_{R,r}$. Observe that

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

As in the proof of Proposition 3, we start with a minimal covering of \mathcal{R}_ρ of cardinal K by balls of radius η for the Hausdorff distance. For any R in \mathcal{R}_ρ at a Hausdorff distance no more than η of an element R_k of the covering, we have

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

Hence,

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

and, therefore, following the same arguments,

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

for a constant C_1 depending only on M ; also,

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

Hence,

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

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

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

for some constant $C_2 > 0$ depending only on M . For the second term, since, by Lemma 6, when $r \leq \rho/2$,

$$\text{var}(\bar{\phi}_{R_k, r}) \leq \frac{C_3 r^{d+1}}{\rho}$$

for a constant C_3 depending only on M , we have

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

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

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

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

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

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 completes the proof.

3.4. A uniform control on $h_n(A)$

As 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. $\text{Vol}_{d-1}(\partial R \cap M_{r_n})$, 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 9. *Under the conditions of Theorem 2, with probability 1, we have*

$$\liminf_{n \rightarrow \infty} \inf_{R \in \mathcal{R}_n} (h_n(R) - h(R; M_{r_n})) \geq 0. \tag{14}$$

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

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

as well as

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

Then

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

Define the event

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

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

Bounding $\zeta_n(R)$. By the definition of \mathcal{R}_n , the sets R and R^c each contain a ball of radius ρ_n , and, by Lemma 2, 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,

$$\lambda_n^*(R) \geq C_1 \rho_n^d. \tag{15}$$

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

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

for a constant $C_2 = C_1/2 > 0$. Using the bias bounds of Lemma 6 together with the perimeter bound in Lemma 8(ii), we have

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

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

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

where the second inequality comes from the fact that $\mathcal{R}_n \subset \mathcal{R}_{\rho_n}$. By the fact that $nr_n^{2d+1} \rho_n^\alpha \rightarrow \infty$ for any $\alpha > 0$, the conditions of Proposition 4 are satisfied, so

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

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

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

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

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

Bounding $\xi_n(R)$. (We reset the constants, except for C_1 .) By the perimeter bound of Lemma 8, we have

$$v_n^*(R) \leq \frac{\text{Vol}_{d-1}(\partial R)}{\tau_M} \leq d \frac{\text{Vol}_d(R)}{\tau_M \rho_n} = \frac{C_2}{\rho_n}$$

for a constant $C_2 > 0$ depending only on M . So, together with (15) and the fact that, on Ω_n , $\lambda_n(R) \geq \lambda_n^*(R)/2$,

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

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

$$I_2 := \mathbb{P}\left[\left[\inf_{R \in \mathcal{R}_n} \xi_n(R) < -\varepsilon\right] \cap \Omega_n\right] \leq \mathbb{P}\left[\sup_{R \in \mathcal{R}_n} |\lambda_n(R) - \lambda_n^*(R)| > \frac{\rho_n^{2d+1} \varepsilon}{C_3}\right]. \tag{16}$$

Define

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

Then

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

with $\mu(M_{r_n}^c) \leq C_4 r_n$ by (9). For fixed $\varepsilon > 0$ and large enough n , $2C_4 r_n \leq \rho_n^{2d+1} \varepsilon / C_3$, again by the fact that $\rho_n \rightarrow 0$ subpolynomially in r_n . We therefore obtain

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

where we have 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 3, whose conditions are satisfied for fixed $\varepsilon > 0$ and large enough n , again because $\rho_n \rightarrow 0$ 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 large enough n . As before, when ε is fixed, the exponent is a positive power of n , so

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

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 (15)), we have

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

We then proceed as in bounding (16), obtaining

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

Conclusion. We have

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

so that the left-hand side is summable. Therefore, we complete the proof by applying the Borel–Cantelli lemma.

3.5. Some continuity of the Cheeger constant

Our proof of Theorem 2 relies on the continuity properties of the normalized cut and of the Cheeger constant. Lemma 10 below compares the conductance function on M and on a bi-Lipschitz deformation of M . For a Lipschitz map f , let $\|f\|_{\text{Lip}}$ denote its Lipschitz constant. If f is bi-Lipschitz, we define its condition number by $\text{cond}(f) := \|f\|_{\text{Lip}} \|f^{-1}\|_{\text{Lip}}$. Lemma 11 below states that M_r is a bi-Lipschitz deformation of M ; hence, Lemma 10 yields the continuity property of Proposition 5.

Lemma 10. *Let f be a bi-Lipschitz and C^1 map on M . Then, for any C^1 open subset A of M ,*

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

Proof. For any C^1 open subset A of M , $f(A)$ is also C^1 since f is bi-Lipschitz and C^1 . Moreover, $\partial f(A) = f(\partial A)$ and $f(A)^c \cap f(M) = f(A^c \cap M)$, and so, for $k = 1, \dots, d$,

$$\|f^{-1}\|_{\text{Lip}}^{-k} \text{Vol}_k(A) \leq \text{Vol}_k(f(A)) \leq \|f\|_{\text{Lip}}^k \text{Vol}_k(A).$$

Therefore,

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

and vice versa.

Lemma 11. Fix $r < s < \rho_M$, where $\rho_M = \text{reach}(\partial M)$. Then there is a bi-Lipschitz C^1 map between M_r and M that leaves M_s unchanged, and with condition number at most

$$\left(1 + \frac{2r}{s-r} + \frac{s}{\rho_M - s}\right) \left(1 + \frac{2r}{s+r} + \frac{s}{\rho_M - s}\right).$$

Proof. Let δ denote the distance function to ∂M on M . For x in M such that $\delta(x) < s$, let $\xi(x) \in M$ be its metric projection onto ∂M and let u_x be the unit normal vector of M at $\xi(x)$ pointing outwards. Let $g: [r, s] \rightarrow [0, s]$ be a nonnegative, increasing, real-valued function of class C^1 such that

$$g(r) = 0, \quad g(s) = s, \quad g'(s) = 1, \quad \|g'\|_\infty > 1, \quad g'(u) \geq 1 \quad \text{for all } u \in [r, s].$$

Let $f: M_r \rightarrow M$ be the map defined for all x in M_r by

$$f(x) = \begin{cases} \xi(x) - g(\delta(x))u_x & \text{if } \delta(x) < s, \\ x & \text{if } x \in \bar{M}_s. \end{cases}$$

By construction, f is of class C^1 on M_r , bijective, and with inverse $f^{-1}: M \rightarrow M_r$ of class C^1 defined for all y in M by

$$f^{-1}(y) = \begin{cases} \xi(y) - g^{-1}(\delta(y))u_y & \text{if } \delta(x) < s, \\ y & \text{if } y \in \bar{M}_s. \end{cases}$$

By [21, Theorem 4.8(1)], $x \mapsto \delta(x)$ is Lipschitz with constant at most 1 on M and is C^1 on $M \setminus \bar{M}_{\rho_M}$ by [21, Theorem 4.8(5)]. By [21, Theorem 4.8(8)], $x \mapsto \xi(x)$ is Lipschitz with constant at most $\rho_M/\rho_M - s$ on $M \setminus M_s$, on which it is also C^1 since ∂M is C^2 by assumption. Since the reach bounds the radius of curvature from below [21, Theorem 4.18], for any x and y in $M \setminus M_s$, we have

$$\|u_x - u_y\| \leq \frac{1}{\rho_M} \|\xi(x) - \xi(y)\|$$

(see also [42, Theorem 1]), and the map $x \mapsto u_x$ is C^1 on $M \setminus M_s$ for ∂M is C^2 . From these facts, it follows that f is of class C^1 on M_r . We now bound the Lipschitz constants of f and f^{-1} .

First, for any x and y in M_s , $\|f(x) - f(y)\| = \|x - y\|$. Second, for any $x \in M_s$ and $y \in M_r \setminus M_s$, we have

$$\|f(y) - f(x)\| = \|f(y) - x\| \leq \|x - y\| + \|y - f(y)\| \leq \|x - y\| + \delta(y) - g(\delta(y)),$$

where we have used the fact that, under the assumptions on g , $g(u) \leq u$ for any u in $[r, s]$. Expanding g at s , we have $g(\delta(y)) = s + g'(\tilde{u})[\delta(y) - s]$ for some $\tilde{u} \in [r, s]$. Thus,

$$\begin{aligned} \|f(y) - f(x)\| &\leq \|x - y\| + (\|g'\|_\infty - 1)(s - \delta(y)) \\ &\leq \|x - y\| + (\|g'\|_\infty - 1)(\delta(x) - \delta(y)) \\ &\leq \|x - y\| + (\|g'\|_\infty - 1)\|x - y\| \\ &= \|g'\|_\infty \|x - y\|. \end{aligned}$$

Third, for any x and y in $M_r \setminus M_s$, we have

$$\begin{aligned} \|f(x) - f(y)\| &= \|\xi(x) - g(\delta(x))u_x - \xi(y) + g(\delta(y))u_y\| \\ &= \|x + [\delta(x) - g(\delta(x))]u_x - y - [\delta(y) - g(\delta(y))]u_y\| \\ &\leq \|x - y\| + |\delta(x) - g(\delta(x)) - \delta(y) + g(\delta(y))| + s\|u_x - u_y\|. \end{aligned}$$

Expanding g at $\delta(x)$, we may write $g(\delta(y)) = g(\delta(x)) + g'(\tilde{u})[\delta(y) - \delta(x)]$ for some \tilde{u} in the segment connecting $\delta(x)$ to $\delta(y)$, which is included in the segment $[r, s]$. Thus,

$$\begin{aligned} \|f(x) - f(y)\| &\leq \|x - y\| + (\|g'\|_\infty - 1)|\delta(y) - \delta(x)| + \frac{s}{\rho_M - s}\|x - y\| \\ &\leq \left[1 + (\|g'\|_\infty - 1) + \frac{s}{\rho_M - s} \right] \|x - y\| \\ &= \left[\|g'\|_\infty + \frac{s}{\rho_M - s} \right] \|x - y\|. \end{aligned}$$

Consequently, we have shown that f is Lipschitz with Lipschitz constant bounded as

$$\|f\|_{\text{Lip}} \leq \|g'\|_\infty + \frac{s}{\rho_M - s}.$$

Similarly, for any x and y in M_s , $\|f^{-1}(x) - f^{-1}(y)\| = \|x - y\|$. For any $x \in M_s$ and $y \in M \setminus M_s$,

$$\|f^{-1}(y) - f^{-1}(x)\| = \|f^{-1}(y) - x\| \leq \|x - y\| + g^{-1}(\delta(y)) - \delta(y),$$

where we have used the fact that, under the assumptions on g , $g^{-1}(u) \geq u$ for any $u \in [0, s]$. Expanding g^{-1} at s , we have

$$g^{-1}(\delta(y)) = s + \frac{1}{g'(g^{-1}(\tilde{u}))}[\delta(y) - s]$$

for some \tilde{u} in $[r, s]$. Thus,

$$\begin{aligned} \|f^{-1}(y) - f^{-1}(x)\| &\leq \|x - y\| + \left[1 - \frac{1}{\|g'\|_\infty} \right] (s - \delta(y)) \\ &\leq \|x - y\| + \left[1 - \frac{1}{\|g'\|_\infty} \right] (\delta(x) - \delta(y)) \\ &\leq \left[2 - \frac{1}{\|g'\|_\infty} \right] \|x - y\|. \end{aligned}$$

Next, for any x and y in $M \setminus M_s$,

$$\begin{aligned} \|f^{-1}(x) - f^{-1}(y)\| &= \|\xi(x) - g^{-1}(\delta(x))u_x - \xi(y) + g^{-1}(\delta(y))u_y\| \\ &= \|x + [\delta(x) - g^{-1}(\delta(x))]u_x - y - [\delta(y) - g^{-1}(\delta(y))]u_y\| \\ &\leq \|x - y\| + |\delta(x) - g^{-1}(\delta(x)) - \delta(y) + g^{-1}(\delta(y))| + s\|u_x - u_y\|. \end{aligned}$$

Expanding g^{-1} at $\delta(x)$, we have

$$g^{-1}(\delta(y)) = g^{-1}(\delta(x)) + \frac{1}{g'(g^{-1}(\tilde{u}))}[\delta(y) - \delta(x)]$$

for some \tilde{u} in the segment connecting $g^{-1}(\delta(x))$ to $g^{-1}(\delta(y))$, which is included in $[0, s]$. Thus,

$$\begin{aligned} \|f^{-1}(x) - f^{-1}(y)\| &\leq \|x - y\| + \left(1 - \frac{1}{\|g'\|_\infty}\right)|\delta(y) - \delta(x)| + \frac{s}{\rho_M - s}\|x - y\| \\ &\leq \left[2 - \frac{1}{\|g'\|_\infty} + \frac{s}{\rho_M - s}\right]\|x - y\|. \end{aligned}$$

Consequently, we have shown that f^{-1} is Lipschitz with Lipschitz constant bounded as

$$\|f^{-1}\|_{\text{Lip}} \leq 2 - \frac{1}{\|g'\|_\infty} + \frac{s}{\rho_M - s}.$$

Now, consider the function $g: [r, s] \rightarrow [0, s]$ defined by

$$g(u) = -\frac{r}{(s-r)^2}(u-r)^2 + \frac{s+r}{s-r}(u-r).$$

Then

$$g(r) = 0, \quad g(s) = s, \quad g'(s) = 1, \quad g'' < 0, \quad 1 \leq g'(u) \leq g'(r) = \frac{s+r}{s-r},$$

so that we may apply the above bounds to deduce that

$$\|f\|_{\text{Lip}} \leq 1 + \frac{2r}{s-r} + \frac{s}{\rho_M - s} \quad \text{and} \quad \|f^{-1}\|_{\text{Lip}} \leq 1 + \frac{2r}{s+r} + \frac{s}{\rho_M - s}.$$

Proposition 5. *We have*

$$H(M_r) = (1 + o(1))H(M) \quad \text{as } r \rightarrow 0.$$

Proof. Consider the bi-Lipschitz C^1 transformation f defined in Lemma 11, with s as a function $s(r)$ such that $s(r) \rightarrow 0$ and $s(r)/r \rightarrow \infty$ as $r \rightarrow 0$. Then, under these conditions, the bound of Lemma 11 implies that $\text{cond}(f) = 1 + o(1)$. Then, using Lemma 10, we deduce that

$$\max\left\{\frac{H(M_r)}{H(M)}, \frac{H(M)}{H(M_r)}\right\} = 1 + o(1) \quad \text{as } r \rightarrow 0,$$

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 $\text{Vol}_d(A \Delta B) = \int |\mathbf{1}_A(x) - \mathbf{1}_B(x)| dx$. 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 semimetric on Borel subsets of \mathbb{R}^d since $\text{Vol}_d(A \Delta B) = 0$ if and only if $A \Delta B$ is a null set.

The following propositions are adapted from Theorem 2.3.10 and Proposition 2.3.6 of [25], respectively, and are also proved in Theorems 1.19 and 1.9 of [23]. Proposition 6 is a compactness criterion, and Proposition 7 results from the lower semicontinuity of the perimeter measure with respect to the L^1 -metric. We recall that, for an open subset A of M , $P_M(A)$ denotes its de Giorgi perimeter.

Proposition 6. *Let (E_n) be a sequence of measurable subsets of M . Suppose that*

$$\limsup_{n \rightarrow \infty} P_M(E_n) < \infty.$$

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

Proposition 7. *Let E_n and E be bounded measurable subsets of M such that $E_n \rightarrow E$ in L^1 . Then*

$$\lim_{n \rightarrow \infty} \text{Vol}_d(E_n) = \text{Vol}_d(E) \quad \text{and} \quad \liminf_{n \rightarrow \infty} P_M(E_n) \geq P_D(E).$$

Consequently,

$$\liminf_{n \rightarrow \infty} h(E_n; M) \geq h(E; M).$$

3.7. Proof of Theorem 2(i)

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

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

since $[h(R_n; M_{r_n}) - H(M_{r_n})] \geq 0$ by the definition of $H(M_{r_n})$. On the last line, by Lemma 9, the first term has a nonnegative inferior limit, and, by Proposition 5, the second term tends to 0. Hence,

$$\liminf_{n \rightarrow \infty} \min_{R \in \mathcal{R}_n} h_n^\ddagger(R) \geq H(M) \quad \text{almost surely.} \tag{17}$$

Upper bound. To obtain the matching upper bound, fix a subset $A \subset M$ with smooth relative boundary and such that $0 < \text{Vol}_d(A) \leq \text{Vol}_d(M \setminus A) < \text{Vol}_d(M)$. Then, for large enough n , 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, $h_n(A) \rightarrow h(A; M)$ almost surely, so

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

By minimizing over A , we obtain

$$\limsup_{n \rightarrow \infty} \min_{R \in \mathcal{R}_n} h_n^\ddagger(R) \leq H(M) \quad \text{almost surely.} \tag{18}$$

Combining the lower and upper bounds, (17) and (18), we conclude that

$$\lim_{n \rightarrow \infty} \min_{R \in \mathcal{R}_n} h_n^\ddagger(R) = H(M) \quad \text{almost surely.} \tag{19}$$

3.8. Proof of Theorem 2(ii)

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 large enough n , 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 $h_n(A^0) \rightarrow h(A^0)$ by Theorem 1, we have

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

Therefore, by compactness of the class of sets with bounded perimeters (Proposition 6), with probability 1, $\{A_n\}$ 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 nonnegative by definition, while the second difference term tends to 0 by Proposition 5. So, with probability 1,

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

On the other hand,

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

so

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

which goes to 0 as $n \rightarrow \infty$ from (14) and (19). Hence,

$$\lim_{n \rightarrow \infty} h(A_n; M_{r_n}) \rightarrow H(M) \quad \text{almost surely.}$$

Now let f_n denote the bi-Lipschitz function mapping M_{r_n} to M defined in Lemma 11 with r and s replaced by r_n and s_n , and where the sequences (r_n) and (s_n) are such that $r_n \rightarrow 0$, $s_n \rightarrow 0$, and $s_n/r_n \rightarrow \infty$ as $n \rightarrow \infty$. Define $B_n = f_n(A_n \cap M_{r_n})$. By Lemmas 10 and 11, we have

$$h(B_n; M) \leq \left(1 + \frac{2r}{s-r} + \frac{s}{\rho_M - s}\right)^d \left(1 + \frac{2r}{s+r} + \frac{s}{\rho_M - s}\right)^d,$$

so $h(B_n; M) \rightarrow H(M)$ almost surely as $n \rightarrow \infty$. Moreover, by Proposition 6, with probability 1, 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 semicontinuous by Proposition 7, with probability 1,

$$\liminf_{n \rightarrow \infty} h(B_n; M) \geq h(B_\infty; M).$$

Since we also have $\liminf_{n \rightarrow \infty} h(B_n; M) = H(M)$ almost surely, it follows that $h(B_\infty; M) = H(M)$ almost surely and so B_∞ is a Cheeger set of M .

Moreover, since f_n leaves M_{s_n} unchanged,

$$\text{Vol}_d(A_n \Delta B_n) \leq \text{Vol}_d(M \setminus M_{s_n}) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence, with probability 1, $\mathbf{1}_{A_n} - \mathbf{1}_{B_n} \rightarrow 0$ 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 $A_n \rightarrow A_\infty$ in L^1 . For all $n \geq 1$, and all f in the class of bounded and continuous functions on M , say $\mathcal{C}_b(M)$, we have

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

where P_n is the empirical measure of the sample X_1, \dots, X_n . Using the bound on the covering numbers in Lemma 7, 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

$$\left| Q_n f - \int_M f(x) \mathbf{1}_{R_n}(x) \mu(dx) \right| \rightarrow 0 \text{ almost surely as } n \rightarrow \infty.$$

Next,

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

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

$$P[\text{for all } f \in \mathcal{C}_b(M), Q_n f \rightarrow Qf] = 1,$$

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

Acknowledgements

Ery Arias-Castro was partially supported by grants from the US National Science Foundation (DMS-06-03890, DMS-09-15160) and wishes 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. We thank an anonymous referee and the Associate Editor for valuable comments and insightful suggestions.

References

- [1] ARORA, S., HAZAN, E. AND KALE, S. (2004). $O(\sqrt{\log n})$ -approximation to sparsest cut in $\tilde{O}(n^2)$ time. In *Proc. 45th Ann. IEEE Symp. on Foundations of Computer Science*, IEEE Computer Society, Washington, DC, pp. 238–247.
- [2] AVIN, C. AND ERCAL, G. (2007). On the cover time and mixing time of random geometric graphs. *Theoret. Comput. Sci.* **380**, 2–22.
- [3] BELKIN, M. AND NIYOGI, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, Vol. 1, MIT Press, Cambridge, pp. 585–592.
- [4] BELKIN, M. AND NIYOGI, P. (2008). Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. System Sci.* **74**, 1289–1308.
- [5] BIAU, G., CADRE, B. AND PELLETIER, B. (2007). A graph-based estimator of the number of clusters. *ESAIM Prob. Statist.* **11**, 272–280.
- [6] BIAU, G., CADRE, B. AND PELLETIER, B. (2008). Exact rates in density support estimation. *J. Multivariate Anal.* **99**, 2185–2207.
- [7] BOYD, S. P., GHOSH, A., PRABHAKAR, B. AND SHAH, D. (2005). Mixing times for random walks on geometric random graphs. In *SIAM Workshop on Analytic Algorithmics & Combinatorics (ANALCO)*, eds C. Demetrescu, R. Sedgewick, and R. Tamassia, SIAM, pp. 240–249.
- [8] BRÄKER, H. AND HSING, T. (1998). On the area and perimeter of a random convex hull in a bounded convex set. *Prob. Theory Relat. Fields* **111**, 517–550.
- [9] BUSER, P. (1982). A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)* **15**, 213–230.
- [10] CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N.S.)* **46**, 255–308.
- [11] CARLSSON, G. AND ZOMORODIAN, A. (2009). The theory of multidimensional persistence. *Discrete Comput. Geom.* **42**, 71–93.
- [12] CASELLES, V., CHAMBOLE, A. AND NOVAGA, M. (2010). Some remarks on uniqueness and regularity of Cheeger sets. *Rend. Sem. Mat. Univ. Padova* **123**, 191–201.
- [13] CHAZAL, F. AND LIEUTIER, A. (2005). Weak feature size and persistent homology: computing homology of solids in \mathbb{R}^n from noisy data samples. In *Computational Geometry (SCG'05)*, ACM, New York, pp. 255–262.
- [14] CHAZAL, F., GUIBAS, L. J., OUDOT, S. Y. AND SKRABA, P. (2009). Analysis of scalar fields over point cloud data. In *Proc. 20th Ann. ACM-SIAM Symp. on Discrete Algorithms*, SIAM, Philadelphia, PA, pp. 1021–1030.
- [15] CHEEGER, J. (1970). A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in Analysis (Papers dedicated to Salomon Bochner, 1969)*, Princeton University Press, Princeton, NJ, pp. 195–199.
- [16] CHUNG, F. R. K. (1997). *Spectral Graph Theory*. (COMS Regional Conf. Ser. Math. **92**). American Mathematical Society, Providence, RI.
- [17] CUEVAS, A., FRAIMAN, R. AND RODRÍGUEZ-CASAL, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.* **35**, 1031–1051.
- [18] DE LA PEÑA, V. H. AND GINÉ, E. (1999). *Decoupling*, Springer, New York.
- [19] DOOB, J. L. (1994). *Measure Theory* (Graduate Texts Math. **143**). Springer, New York.
- [20] EVANS, L. C. AND GARIEPY, R. F. (1992). *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL.
- [21] FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93**, 418–491.
- [22] GINÉ, E. AND KOLTCHINSKII, V. (2006). Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High Dimensional Probability* (IMS Lecture Notes Monogr. Ser. **51**), Institute of Mathematical Statistics, Beachwood, OH, pp. 238–259.
- [23] GIUSTI, E. (1984). *Minimal Surfaces and Functions of Bounded Variations* (Monogr. Math. **80**). Birkhäuser, Basel.
- [24] GRAY, A. (2004). *Tubes* (Progress Math. **221**), 2nd edn. Birkhäuser, Basel.
- [25] HENROT, A. AND PIERRE, M. (2005). *Variation et Optimisation de Formes* (Math. Appl. (Berlin) **48**), Springer, Berlin.
- [26] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.
- [27] KHMALADZE, E. AND WEIL, W. (2008). Local empirical processes near boundaries of convex bodies. *Ann. Inst. Statist. Math.* **60**, 813–842.
- [28] KOLMOGOROV, A. N. AND TIKHOMIROV, V. M. (1961). ε -entropy and ε -capacity of sets in functional space. *Amer. Math. Soc. Transl. (2)* **17**, 277–364.
- [29] LEVINA, E. AND BICKEL, P. J. (2005). Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, Vol. 17, MIT Press, Cambridge, pp. 777–784.
- [30] MAIER, M., VON LUXBURG, U. AND HEIN, M. (2009). Influence of graph construction on graph-based clustering measures. In *Advances in Neural Information Processing Systems*, Vol. 22, MIT Press, Cambridge, pp. 1025–1032.

- [31] NARAYANAN, H. AND NIYOGI, P. (2009). On the sample complexity of learning smooth cuts on a manifold. In *22nd Ann. Conf. on Learning Theory (COLT)*.
- [32] NARAYANAN, H., BELKIN, M. AND NIYOGI, P. (2007). On the relation between low density separation, spectral clustering and graph cuts. In *Advances in Neural Information Processing Systems*, Vol. 19, MIT Press, Cambridge.
- [33] NG, A. Y., JORDAN, M. I. AND WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, Vol. 14, MIT Press, Cambridge, pp. 849–856.
- [34] NIYOGI, P., SMALE, S. AND WEINBERGER, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39**, 419–441.
- [35] PELLETIER, B. AND PUDLO, P. (2011). Operator norm convergence of spectral clustering on level sets. *J. Mach. Learning Res.* **12**, 385–416.
- [36] PENROSE, M. (2003). *Random Geometric Graphs* (Oxford Stud. Prob. 5). Oxford University Press.
- [37] ROBINS, V. (1999). Towards computing homology from finite approximations. In *Proc. 14th Summer Conf. on General Topology and Its Applications* (Brookville, NY, 1999; Topology Proc. **24**), pp. 503–532.
- [38] SINGER, A. (2006). From graph to manifold Laplacian: the convergence rate. *Appl. Comput. Harmon. Anal.* **21**, 128–134.
- [39] SPIELMAN, D. A. AND TENG, S.-H. (2007). Spectral partitioning works: planar graphs and finite element meshes. *Linear Algebra Appl.* **421**, 284–305.
- [40] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Statist. Comput.* **17**, 395–416.
- [41] VON LUXBURG, U., BELKIN, M. AND BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36**, 555–586.
- [42] WALTHER, G. (1997). Granulometric smoothing. *Ann. Statist.* **25**, 2273–2299.
- [43] WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61**, 461–472.
- [44] ZOMORODIAN, A. AND CARLSSON, G. (2005). Computing persistent homology. *Discrete Comput. Geom.* **33**, 249–274.