PyTDA

Kirk P. Gardner[*] Donald R. Sheehy[†]

1 Sensor Networks and Simplicial Complexes

We consider the problem of determining coverage in a coordinate free sensor network. That is, we would like to determine if an unknown domain is covered by a collection of sensors without their precise coordinates. Let $\mathcal{D}$ denote our unknown domain and $P \subset \mathcal{D}$ be a collection of points, each representing a sensor in our network. At the very least each sensor is capable of detecting nodes which are sufficiently "close" in the sense that, if we endow our domain with a metric $\mathrm{d} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ there is some radius of communication $\alpha > 0$ such that two nodes $p, q \in P$ such that $\mathrm{d}(p, q) \leq \alpha$ are capable of communication. Note that, although sensors can communicate within this distance they are not able to measure the distance itself.

We say that $D \subseteq \mathcal{D}$ is covered by P at scale $\varepsilon > 0$ if every point $x \in D$ is within distance α at least one point in P. Let $\mathrm{ball}_\varepsilon(p) = \{x \in \mathcal{D} \mid \mathrm{d}(x, p)\}$ denote the **coverage region** of $p \in P$ at scale ε. We will use the notation P^ε to denote set of points in $\mathcal{D}$ within distance ε of at least one point in P:

$$P^\varepsilon = \bigcup_{p \in P} \mathrm{ball}_\varepsilon(p).$$

That is, $D \subseteq P^\varepsilon$ then the subset D is **covered** by P at scale ε.

With this limited capability we can construct an undirected graph $G = (V, E)$ with vertices $V = P$ and edges $E = \{\{p, q\} \subset P \mid \mathrm{d}(p, q) \leq \alpha\}$. In order to determine coverage we must at least assert that the coverage domain spanned by the points in P does not contain any holes. Assuming the coverage radius of our sensors is equal to their communication radius *alpha* we may define a hole in coverage as a cycle that cannot be "filled" with triangles. This leads us to a more natural construction known as a simplicial complex.

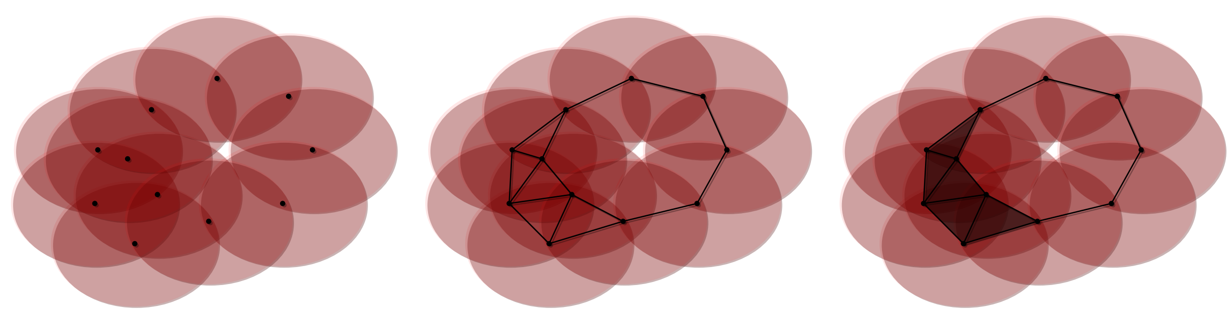

Figure 1:

Definition 1. *A **simplicial complex** K is a collection of subsets, called **simplices**, of a vertex set V such that for all $\sigma \in K$ and $\tau \subset \sigma$ it must follow that $\tau \in K$.*

Simplices may be thought of as a generalization of edges in a graph to higher dimensions. The **dimension** of a simplex $\sigma \in K$ is defined as $\dim(\sigma) := |\sigma| - 1$ where $|\cdot|$ denotes set cardinality. The dimension of a

[*]University of Connecticut.
[†]University of Connecticut.

simplicial complex K is the maximum dimension of any simplex in K. That is, a graph is a 1-dimensional simplicial complex in which vertices and edges are 0 and 1-dimensional simplices, respectively.

Recall, we have defined a hole in our graph as a cycle that cannot be filled with triangles. If we instead construct a 2-dimensional simplicial complex in which a triangle, or 2-simplex, whenever 3 points are all within pairwise distance α a hole is simply a 1-cycle in the simplicial complex. This particular simplicial complex is known as the Vietoris-Rips complex.

Definition 2. *The **(Vietoris-)Rips complex** is defined for a set P at scale $\varepsilon > 0$ as*

$$\mathcal{R}^{\varepsilon}(P) = \{\sigma \subseteq P \mid \forall p, q \in \sigma,\ \mathrm{d}(p,q) \leq \varepsilon\}.$$

This construction generalizes to higher dimensions, allowing us to identify not only holes in planar graphs, "holes" in any dimension k as $(k-1)$-cycles in high dimensional simplicial complexes.

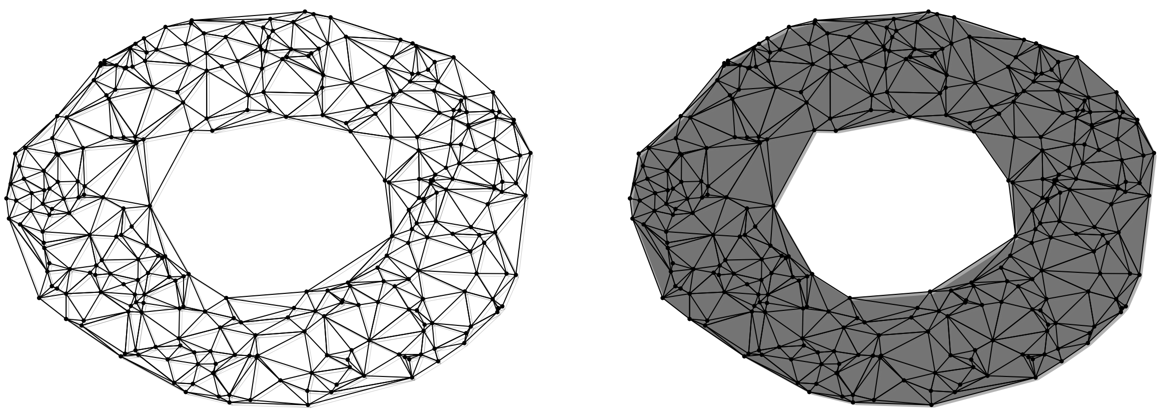

Figure 2:

Once we have constructed a Rips complex $K = \mathcal{R}^{\alpha}(P)$ we may identify gaps in coverage as "unfilled" cycles in the dimension of our domain. However, even if we know there are no gaps in coverage how can we assert that our network sufficiently samples the domain? Moreover, if there are gaps in coverage, are they due to insufficient sampling or a gap in the domain itself? Both of these issues may be reconciled by allowing our sensors to detect the **boundary** $\mathcal{B} \subset \mathcal{D}$ of the domain. Let $Q = \{p \in P \mid \min_{x \in \mathcal{B}} \mathrm{d}(x,p)\}$ be the set of boundary nodes in P and let $K\mid_Q = \mathcal{R}^{\alpha}(Q)$ be the subcomplex of K restricted to nodes in Q. Now, way may say that a subset $D \subset \mathcal{D}$ is covered by our network at scale α if there are no holes in our network and no path from a simplex in K to a point in $\mathcal{D} \setminus D$ without crossing a simplex in $K\mid_Q$.

Note that the communication radius is far from a tight bound on the coverage radius. In section 5 we will detail precisely how to verify these conditions with a suitable coverage radius.

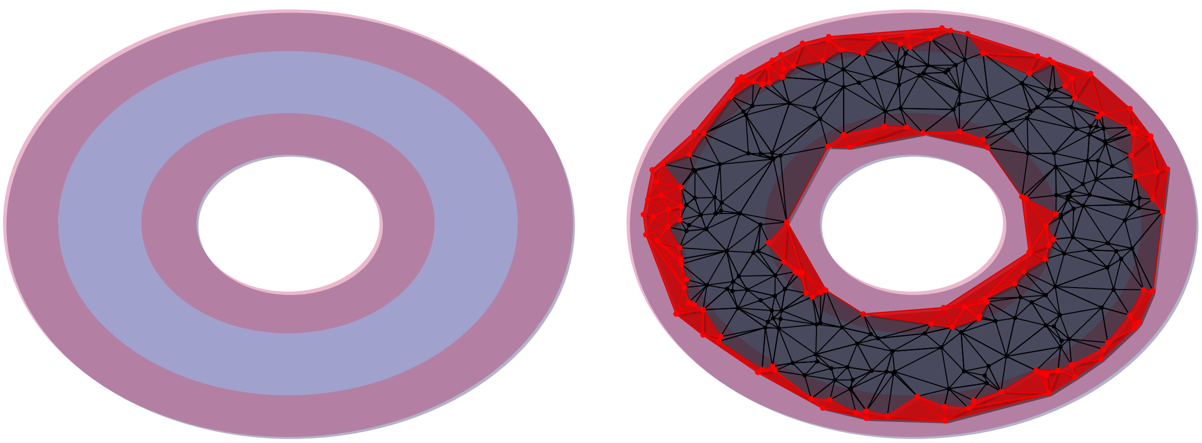

Figure 3:

2 Functions on Complexes

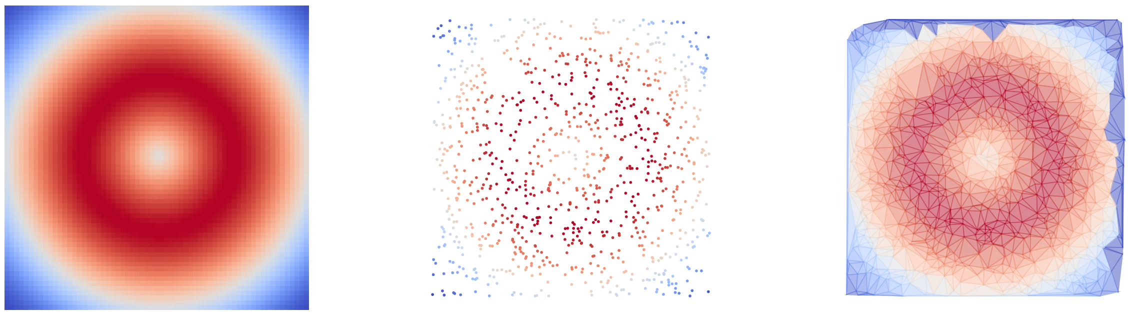

Figure 4: (Left) Function on the plane. (Middle) Function values on random sample. (Right) Function values on the 2-simplices of a simplicial complex on the random sample.

3 Homology and Cohomology

Simplicial complexes not only provide a comprehensive discretization of continuous domains, but are the primary tool for concrete calculations in algebraic topology. In particular, the study of simplicial homology groups and its dual cohomology groups rely on simplicial complexes in order to study important topological invariants of a discretized space.

3.1 Homology

The following vector spaces may be defined over any field $\mathbb{F}$, however we will assume the field $\mathbb{F}_2$ in order to avoid orienting the simplices in K. Let $C_k(K)$ denote the vector space over some field $\mathbb{F}$ consisting of linear combinations of k-simplices in K, which form a basis for $C_k(K)$, known as k**-chains**. These vector spaces are connected by **boundary maps** $\partial_k : C_k(K) \to C_{k+1}(K)$ which are linear transformations taking basis elements of $C_k(K)$ to the abstract sum of basis $(k+1)$-simplex faces. The collection of chains and boundary maps forms a sequence of vector spaces known as the **chain complex** $\mathcal{C} = (C_*, \partial_*)$

$$\ldots \xrightarrow{\partial_{k+1}} C_k(K) \xrightarrow{\partial_k} C_{k-1}(K) \xrightarrow{\partial_{k-1}} \ldots \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0.$$

An important property of the boundary maps ∂_k is that the composition of subsequent boundary maps is null. That is, for all k

$$\partial_k \circ \partial_{k-1} = 0.$$

As a result the image of ∂_{k+1}, denoted im $\partial_{k+1} = \{\partial_{k+1} c \mid c \in C_{k+1}(K)$ is a subspace of the kernel, $\ker \partial_k = \{c \in C_k(K) \mid \partial_k c = 0\}$, of ∂_k. A k**-cycle** of $\mathcal{C}$ is a k-chain with empty boundary - an element of $\ker \partial_k$. Two cycles in $\ker \partial_k$ are said to be **homologous** if they differ by an element of im ∂_{k+1}. This leads us to the definition of the **homology groups** of K as quotient vector spaces $H_k(K)$ over $\mathbb{F}$, defined for $k \in \mathbb{N}$ as

$$H_k(K) := \ker \partial_k / \text{im } \partial_{k+1}.$$

3.2 Cohomology

A **cochain complex** is a sequence $\mathcal{C} = (C^*, \delta_*)$ of $\mathbb{R}$-modules C^k consisting of **cochains** and module homomorphisms known as **coboundary maps** $\delta_k : C^k \to C^{k+1}$. As in homology we have the property that $\delta_{k+1} \circ \delta_k = 0$ for all k, leading to a familiar definition of the **cohomology** of $\mathcal{C}$

$$H^k(\mathcal{C}) = \ker \delta_k / \text{im } \delta_{k-1}.$$

The equivalence classes of $H^k(\mathcal{C})$ consist of k**-cocycles**: elements of $\ker \delta_k$ that differ by a k**-coboundary** in im ∂_{k-1}. Such cocycles are said to be **cohomologous** if they belong to the same equivalence class in $H^k(\mathcal{C})$.

The simplest construction of a cochain complex is to dualize a chain complex. For a simplicial complex K with chain complex (C_*, ∂_*) define $C^k(K)$ to be the module of homomorphisms $\psi : C_k \to \mathbb{R}$. The coboundary maps δ_k are defined for cochains $\psi : C_k \to \mathbb{R}$ and k-simplices $\sigma \in K$ as

$$\delta_k \psi(\sigma) = \psi(\partial_k \sigma).$$

3.3 Relative Homology and Cohomology

3.4 Representative Cycles and Cocycles

We can find a basis for each homology group $H_p(K)$ and cohomology group $H^p(K)$ consisting of p-cycles and p-cocycles, respectively. In general, p-cycles represent p-dimensional holes in the simplicial complex K, where p-cocycles can be understood as “blocking chains.”

3.5 Persistent Homology and Cohomology

4 The Topological Coverage Criterion

There is a specific simplicial complex defined for a collection of sets that precisely captures the coverage information we require. The **Čech complex** of a finite collection of points P at scale $\varepsilon > 0$ is defined

$$\check{\mathcal{C}}^{\varepsilon}(P) := \left\{ \sigma \subseteq P \mid \bigcap_{p \in \sigma} \mathrm{ball}_{\varepsilon}(p) \neq \emptyset \right\}.$$

The Čech complex is a special case of a more general construction known as a the **nerve** $\mathbb{N}(\mathcal{U})$ of a collection of sets $\mathcal{U} = \{U_i\}_{i \in I}$, where I is any indexing set. The nerve of $\mathcal{U}$ is defined as the simplicial complex with vertex set I such that $\sigma \subseteq I$ is a simplex if and only if

$$\bigcap_{i \in \sigma} U_i \neq \emptyset.$$

The collection $\mathcal{U}$ is a **good cover** if for each $\sigma \subset I$ the set $\bigcap_{i \in \sigma} U_i$ is contractible if it is nonempty. The **nerve lemma** states that if $\mathcal{U}$ is a good cover then its nerve $\mathbb{N}(\mathcal{U})$ is homotopy equivalent to $\bigcup_{i \in I} U_i$. That is, for a set of nodes $P \subset \mathcal{D}$ such that $\mathcal{U} = \{\mathrm{ball}_{\varepsilon}(p) \mid p \in P\}$ is a good cover the nerve $\mathbb{N}(\mathcal{U})$ is homotopy equivalent to $P^{\varepsilon} = \bigcup_{p \in P} \mathrm{ball}_{\varepsilon}(p)$. It follows that the Čech complex $\check{\mathcal{C}}_{\varepsilon}(P)$ of P at scale ε is a suitable representation of the coverage region P^{ε}.

While the Čech complex captures the topology of the coverage region in question exactly it can only be computed when the precise pairwise distances between nodes is known. If instead we are only provided pairwise *proximity* information indicating when nodes are within some fixed distance we may use the **(Vietoris-)Rips complex**, defined for a set P at scale $\varepsilon > 0$ as

$$\mathcal{R}^{\varepsilon}(P) := \{\sigma \subseteq P \mid \forall p, q \in \sigma,\ \mathrm{d}(p, q) \leq \varepsilon\}.$$

An important result about the relationship of Čech and Rips complexes follows from Jung's Theorem [2] relating the diameter of a point set P and the radius of the minimum enclosing ball:

$$\check{\mathcal{C}}^{\varepsilon}(P) \subseteq \mathcal{R}^{\varepsilon}(P) \subseteq \check{\mathcal{C}}^{\vartheta_d \varepsilon}(P), \tag{1}$$

where the constant $\vartheta_d = \sqrt{\frac{2d}{d+1}}$ (see [1]).

As we will see the Rips complex may be used to verify coverage of a domain satisfying some minimal assumptions when we allow the sensors in our network to communicate at two radii α and β. In short, if the the inclusion of Rips complexes at scales $\alpha < \beta$ resembles the structure of a subset of the domain then the Čech complex at scale α, and therefore P^{α}, does as well.

4.1 Implementation

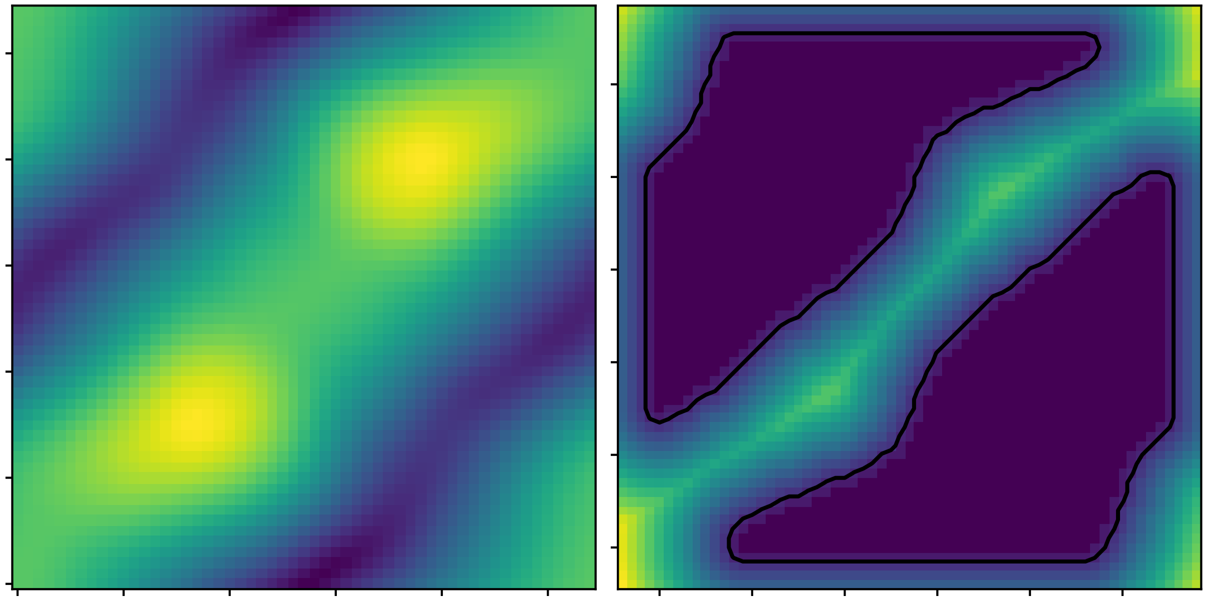

Figure 5:

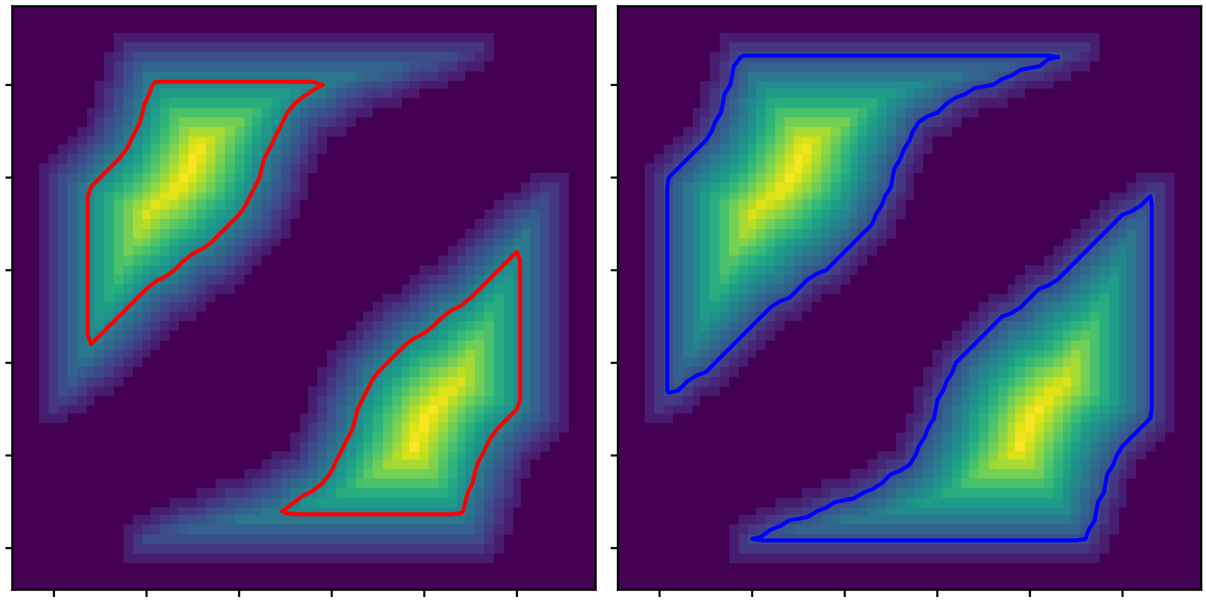

Figure 6:

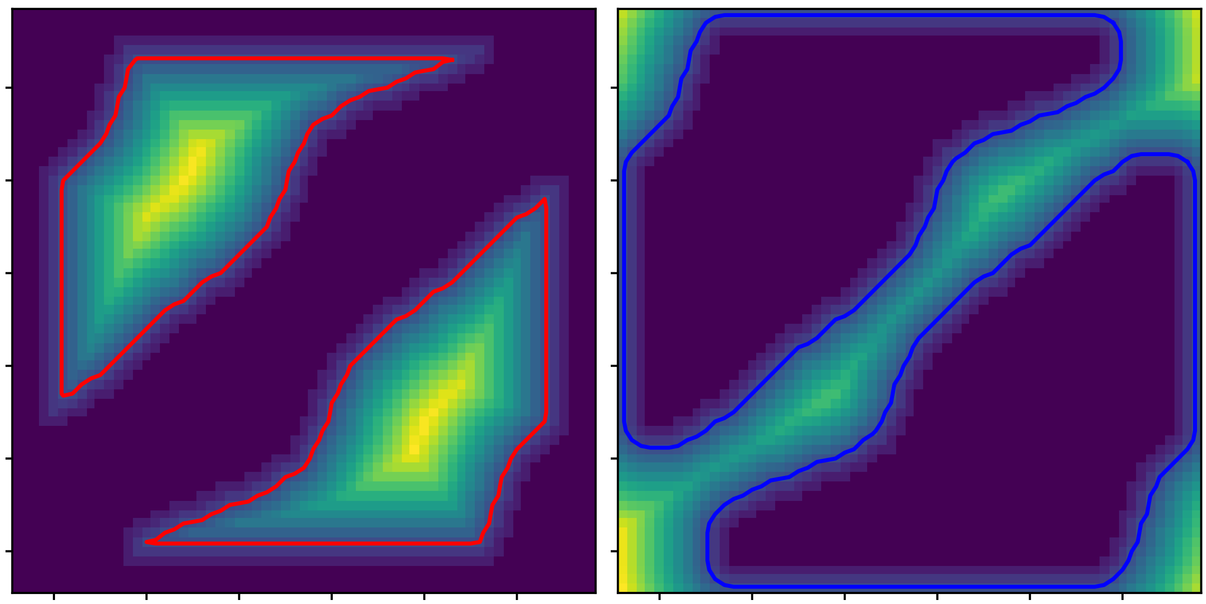

Figure 7:

References

[1] Mickaël Buchet, Frédéric Chazal, Steve Y. Oudot, and Donald R. Sheehy. Efficient and robust persistent homology for measures. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 168–180, 2015.

[2] Heinrich Jung. Über die kleinste kugel, die eine räumliche figur einschließt. *J. Reine Angew. Math*, 123:214–257, 1901.