25 Minimum bandwidth: Approximation via volume respecting embeddings

We continue the study of *Volume respecting embeddings*. In the last lecture, we motivated the use of volume respecting embeddings by an approximation algorithm for the minimum bandwidth problem. This lecture continues the proof of the approximation ratio of the algorithm. We prove parts of the following claims.

1. $\sum_{S:|S|=k} 1/\mathsf{Tvol}(S) = \tilde{O}(n(D\log n)^{k-1})$.
2. When the mapped metric space $\phi(v) \in \mathbb{R}^L$ is projected to a line, the probability that an edge is projected to an interval greater than $O(\sqrt{\log n/L})$ is no more than $\frac{1}{2m}$.
3. There exists an (η, k)-well separated contracting map ϕ.

The terms used in the above list were defined in the previous lecture, and will be redefined as we work our way through the proof. Putting together the three claims yields an approximation ratio for minimizing bandwidth of $O(n^{1/k}k\eta D\log n)$. Setting $k = \log n$ and choosing η appropriately yields the desired poly-logarithmic approximation ratio.

25.1 A Lemma about Distances

Claim 25.1 *For all graphs G, we have*

$$\sum_{S\subseteq V:|S|=k} \frac{1}{\mathsf{Tvol}(S)} \leq n \times O(D\log n)^{k-1}$$

Proof. The proof is continued from the last lecture. Recall the definition of the tree volume $\mathsf{Tvol}(S)$. Given a set S of vertices, let T be a minimum spanning tree of S in the graph obtained by taking the metric completion of the original graph. Then, $\mathsf{Tvol}(S) = \Pi_{e\in T} d(e)$, where $d(e)$ is the length of edge e.

Let $\mathfrak{S}_n$ denote the set of all permutations of $[n]$. Recall from the last lecture that in order to prove Claim 1, it suffices to show the following for all S such that $|S| = k$.

$$\frac{2^{k-1}}{\mathsf{Tvol}(S)} \leq \sum_{\pi\in\mathfrak{S}_n} \frac{1}{d(v_{\pi(1)}, v_{\pi(2)})\dots d(v_{\pi(k-1)}, v_{\pi(k)})} \tag{25.1}$$

We prove (25.1) by induction on k. As the base case, consider $k = 2$. In this case, there are only 2 permutations in $\mathfrak{S}_2$ and the MST is unique, so (25.1) holds with equality.

For the inductive step, consider a set $S = \{v_1, v_2, \dots, v_{k+1}\}$, labeled so that v_{k+1} is a leaf in the MST of S. Define $S' = S \setminus \{v_{k+1}\}$; by the induction hypothesis, (25.1) holds for S'.

Let $i = \text{argmin}_{j=1,2,\dots,k}\, d(v_{k+1}, v_j)$. Hence, in the MST of S, the vertex v_{k+1} is connected by an edge to v_i, and thus $\mathsf{Tvol}(S) = \mathsf{Tvol}(S')d(v_{k+1}, v_i)$. Using the induction hypothesis, we get the following.

$$\frac{2^k}{\mathsf{Tvol}(S)} = \frac{2}{d(v_{k+1}, v_i)} \times \frac{2^{k-1}}{\mathsf{Tvol}(S')} \leq \frac{2}{d(v_{k+1}, v_i)} \Big(\sum_{\pi \in \mathfrak{S}_k} \frac{1}{d(v_{\pi(1)}, v_{\pi(2)}) \dots d(v_{\pi(k-1)}, v_{\pi(k)})} \Big)$$

Therefore, it suffices to prove:

$$\frac{2}{d(v_{k+1}, v_i)} \sum_{\pi \in \mathfrak{S}_k} \frac{1}{d(v_{\pi(1)}, v_{\pi(2)}) \dots d(v_{\pi(k-1)}, v_{\pi(k)})} \leq \sum_{\sigma \in \mathfrak{S}_{k+1}} \frac{1}{d(v_{\sigma(1)}, v_{\sigma(2)}) \dots d(v_{\sigma(k)}, v_{\sigma(k+1)})} \tag{25.2}$$

For a permutation $\pi \in \mathfrak{S}_k$, define

$$X = \frac{1}{d(v_{\pi(1)}, v_{\pi(2)}) \dots d(v_{\pi(k-1)}, v_{\pi(k)})}.$$

Note that there are $k+1$ possible extensions of π to permutations in $\mathfrak{S}_{k+1}$, obtained by the $(k+1)$ ways to insert the new element (i.e., the $(k+1)^{th}$ element) in π; this defines a bijection between $\mathfrak{S}_k$ and $(\mathfrak{S}_{k+1})^{k+1}$. Hence, to prove (25.2), it suffices to show the following for every permutation $\pi \in \mathfrak{S}_k$:

$$\frac{2X}{d(v_{k+1}, v_i)} \leq X \Big(\frac{1}{d(v_{k+1}, v_{\pi(1)})} + \frac{1}{d(v_{k+1}, v_{\pi(k)})} + \sum_{j=1}^{k-1} \frac{d(v_{\pi(j)}, v_{\pi(j+1)})}{d(v_{\pi(j)}, v_{k+1}) d(v_{\pi(j+1)}, v_{k+1})} \Big) \tag{25.3}$$

The rest of the arguyment is dedicated to proving (25.3). Pick any permutation $\pi \in \mathfrak{S}_k$; without loss of generality (and for ease of exposition), assume that π is the identity permutation with $\pi(i) = i$ for all $1 \leq i \leq k$.[1] By the triangle inequality, we know that

$$d(v_j, v_{j+1}) \geq \begin{cases} -d(v_j, v_{k+1}) + d(v_{k+1}, v_{j+1}) & \text{for } j < i, \text{ and} \\ d(v_j, v_{k+1}) - d(v_{k+1}, v_{j+1}) & \text{for } j \geq i. \end{cases}$$

Using these two equations in the RHS of (25.3) yields:

$$\begin{aligned} & X \Big(\frac{1}{d(v_{k+1}, v_{\pi(1)})} + \frac{1}{d(v_{k+1}, v_{\pi(k)})} + \sum_{j=1}^{k-1} \frac{d(v_{\pi(j)}, v_{\pi(j+1)})}{d(v_{\pi(j)}, v_{k+1}) d(v_{\pi(j+1)}, v_{k+1})} \Big) \\ \geq & X \Big(\frac{1}{d(v_{k+1}, v_1)} + \frac{-d(v_1, v_{k+1}) + d(v_{k+1}, v_2)}{d(v_1, v_{k+1}) d(v_2, v_{k+1})} + \dots + \frac{-d(v_{i-1}, v_{k+1}) + d(v_i, v_{k+1})}{d(v_{i-1}, v_{k+1}) d(v_i, v_{k+1})} \\ & \quad + \frac{d(v_i, v_{k+1}) - d(v_{k+1}, v_{i+1})}{d(v_i, v_{k+1}) d(v_{k+1}, v_{i+1})} + \dots + \frac{1}{d(v_k, v_{k+1})} \Big) \\ \geq & \frac{2X}{d(v_i, v_{k+1})} \end{aligned}$$

The last inequality follows via a telescopic cancellation. The last term above is the LHS of (25.3), and this completes the proof of Claim 25.1. ■

[1] Inspired by Henry Ford: "The Model T will be available in any color, so long as the color is black."

25.2 An Upper Bound on Edge Lengths

In this section, we consider random projections of vectors $v \in \mathbb{R}^L$.

Claim 25.2 *If we pick a vector $\vec{r} = (r_1, \ldots, r_L)$ and each $r_i \sim N(0,1)$, then for a vector v of at most unit length,*

$$\mathbf{Pr}\Big[|\langle \vec{r}, v\rangle| > O(\sqrt{\log n})\Big] \leq \frac{1}{2m}.$$

Proof. By spherical symmetry, it follows that $\langle \vec{r}, v\rangle$ behaves like a $N(0, \|v\|_2^2)$ random variable with $\|v\|_2 \leq 1$. Let X be a $N(0,1)$ random variable; now, the probability

$$\begin{aligned}\mathbf{Pr}\Big[|\langle \vec{r}, v\rangle| > 2\sqrt{\log n}\Big] &= \mathbf{Pr}\Big[\|v\|_2 \times |X| > 2\sqrt{\log n}\Big] \\ &\leq \mathbf{Pr}\Big[|X| > 2\sqrt{\log n}\Big]\end{aligned}$$

since $\|v\|_2 \leq 1$. However, we can now use standard tail bounds on Gaussians: Lemma 2 in Chapter VII.1 of Feller (1968) states that $\mathbf{Pr}[|X| > x] \leq \frac{2}{x} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Hence

$$\leq \frac{2}{2\sqrt{\log n} \times \sqrt{2\pi}} e^{-\frac{1}{2}(2\sqrt{\log n})^2} < \frac{1}{n^2} < \frac{1}{2m}$$

■

25.2.1 An alternate proof

We can give a geometric proof of this fact as well; for this, assume that we pick a vector $\vec{r} \in S^{L-1}$ and project the embedding onto the line defined by $\vec{r}$. Though this is a different procedure from the one we analysed last time, the results are almost the same. Let us now give a geometric proof of the fact that the length of an edge is not too large if $\vec{r}$ is chosen in this new fashion.

Claim 25.3 *When the metric space is projected to a line, the probability that an edge is projected to an interval greater than $O(\sqrt{\frac{\log n}{L}})$ is no more than $\frac{1}{2m}$.*

Proof. Recall that our projection ϕ is constructed by choosing a line randomly in $\mathbb{R}^L$ and projecting V down to it. For an edge e, we may visualize it as follows: one of the end-points of e is the origin, and a line is chosen passing through the origin. This line is chosen randomly, and the following process is equivalent for choosing the line: pick a point v_0 uniformly at random on the surface of the unit sphere in $\mathbb{R}^L$, and let the line be the unique line l_0 passing through this chosen point and the origin. Refer to Figure 25.1.

Since the map is a contraction, we may assume that the length of e is no more than 1. The length of e in the projection is equal to the inner product between e and the line. Given x, we now bound from above the probability that this inner product is more than $\sqrt{x/L}$.

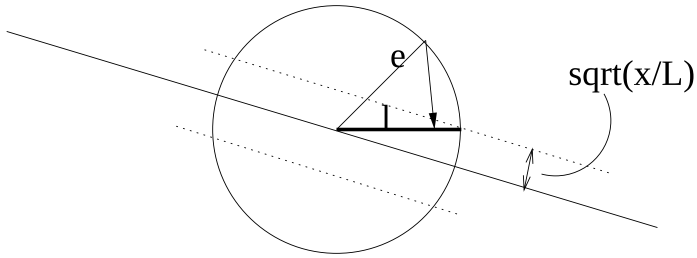

Figure 25.1: Projection to a randomly chosen line

For the inner product to be less than $\sqrt{x/L}$, the point v_0 must lie on the sphere in a narrow band close to the diameter defined by the plane orthogonal to e. In Figure 25.1, this band is the area between the dotted lines. The width of this band is precisely $\sqrt{x/L}$. If we remove this band and move the rest of the sphere together, then the entire "shell" can be enclosed in a sphere of radius $\sqrt{1-x/L}$.

The surface area of this shell is therefore no more than $c(\sqrt{1-x/L})^{L-1}$, for some constant c. Therefore, the measure of this surface area is no more than $(\sqrt{1-x/L})^{L-1} \leq e^{-x/2}$. Hence, the probability that the edge e is projected to an interval of length greater than $\sqrt{x/L}$ is no more than $e^{-x/2}$. Setting $x = 2\log(2m)$ and $L = 2\log(2m)/l^2$ yields Claim 25.3. ∎

26 Existence of Well-Separated Maps

All that remains is to show that there exists an (η, k)-well separated contracting map ϕ. Recall that the following definition:

Definition 26.4 *A map ϕ is (η, k)-well separated if for all S such that $|S| = k$, there exists a permutation $s_0, s_1, \ldots, s_{k-1}$ of the elements of S such that the following holds for all i: if $\mathcal{A}_i(\phi(S))$ is the affine span of the first i points $\{\phi(s_0), \phi(s_1), \ldots, \phi(s_{i-1})\}$, then*

$$||\phi(s_i), \mathcal{A}_i(\phi(S))||_2 \geq \frac{1}{\eta} dist_G(s_i, \{s_0, s_1, \ldots, s_{i-1}\}).$$

Let us prove the following useful sufficient condition for a map to be well-separated.

Claim 26.5 *Let $q_i = dist_G(s_i, \{s_0, s_1, \ldots, s_{i-1}\})$. To show the existence of an (η, k)-well separated map ϕ, it suffices to give a randomized construction such that for* an arbitrary (fixed) point $a \in \mathcal{A}_i(\phi(S))$,

$$\mathbf{Pr}\left[\ ||\phi(s_i) - a||_2 > 3\,\frac{q_i}{\eta}\right] \geq 1 - n^{-3k}. \tag{26.4}$$

Proof. To show well-separatedness, we need to show (for all S and i) that the *closest* point to $\phi(s_i)$ in $\mathcal{A}_i(\phi(S))$ is at least q_i/η from it. Fix a set S and an index i: since ϕ is a contraction, all the points $\phi(s_j)$ for $j < i$ lie in $\mathcal{A}_i(\phi(S))$ at a distance at most $\mathsf{diam}(G) \leq n$ from $\phi(s_i)$. Hence, the closest point to $\phi(s_i)$ in $\mathcal{A}_i(\phi(S))$ also lies at distance at most n from it.

So consider the set $X = \mathcal{A}_i(\phi(S)) \cap B(\phi(s_i), n)$, and pick an $(2/\eta)$-net Y for X—this is just a maximal set of points in X that are at least $(2/\eta)$ from each other. Note that Y can have at most $(n/\eta)^{i-1} = O(n^{2k})$ points; indeed, note that the $1/\eta$ balls around the points in Y are pairwise disjoint, and they can all be packed inside a set X of radius at most n. We now apply (26.4) to each of the points in Y, and applying the union bound, get that $\phi(s_i)$ is $3\,q_i/\eta$ far from the set Y with probability $1 - n^{-k}$. But now the triangle inequality implies that, given *any* point $b \in X$ (that is a candidate closest point), there is a point $a \in Y$ with $\|a - b\|_2 < (2/\eta)$, and hence

$$\|\phi(s_i) - b\|_2^2 \geq \|\phi(s_i) - a\|_2^2 - \|a - b\|_2^2 \geq (\frac{3q_i}{\eta})^2 - (\frac{2}{\eta})^2 \geq (\frac{q_i}{\eta})^2$$

with the same probability $1 - n^{-k}$. Finally, applying a union bound over all sets S of size k and all indices i, we get that ϕ is (η, k)-well separated with constant probability. ■

26.1 The Embedding

Instead of the value of $\eta = O(\sqrt{\log n(\log n + k \log k)})$ given by Feige (1998), we will prove a slightly simpler $O(\log^{3/2} n)$ bound on η. (Note that this bound is independent of k, and hence better for large values of k; however, for our application where we set $k = O(\log n)$, this bound will be worse by a $\sqrt{\log n / \log \log n}$ factor.)

To recap, we will now show a randomized map which, for an arbitrary point $a \in A_i(\phi(S))$, ensures that

$$\|\phi(s_i) - a\|_2 \geq \frac{3\,q_i}{\eta}$$

with probability greater than $1 - n^{-3k}$.

This will be a random subset embedding with $O(k \log n \log D)$ coordinates, where D is the diameter of the graph. Recall that we are working with a graph with unit-length edges. For a constant c to be specified later, the embedding ϕ is given by the procedure in Figure 26.2.

For some value of t and j, it remains to describe the procedure "Generate-Coordinate" that maps all the vertices of G into the real line, which we call f_{tj}. We will create a random subset S_{tj} of the vertices, by the following randomized procedure. (See Figure 26.3 for an illustration.)

We start with S_{tj} being empty and $G' = G$. We pick an arbitrary node $v \in G'$, and look at a BFS rooted at v. Let $l(u)$ be the distance of a node u from the root v; all vertices u that have the same value of $l(u)$ form a *level set*. We define a "swaths" to be a collection of $r = \Delta/(4 \log n)$ consecutive level sets, with the k-th swath consisting of the vertices $\{u \mid l(u) \in [(k-1)r, kr)\}$. We pick one of these swaths randomly, with the k^{th} swath chosen with probability $p_k = 2^{-k}$, and then pick a level l uniformly at random *within* this swath, adding the nodes at this distance l to the set S_{tj} and deleting them from G'. This disconnects a connected component C containing the root v from G'; we remove this component C from the current graph G' and repeat until $G' = \emptyset$.

let $L = c \log n \log D$.
for $t = 1$ **to** $\log D$
  let $\Delta = 2^t$
  for $j = 1$ to $ck \log n$
    f_{tj} = Generate-Coordinate(G, Δ)
  end for
end for
let $f = \oplus_{tj} f_{tj}$.
let $\phi = f/\sqrt{4L}$ for all t, j.

Figure 26.2: The embedding algorithm

Finally, for each component C created during the course of the above procedure, we independently choose a value $\gamma_{tj}(C)$ uniformly at random from the interval $[1, 2]$, and set the lengths of all edges within C to $\gamma_{tj}(C)$. The coordinate value for a vertex $x \in V(G)$ is now its distance from the set S_{tj} with respect to these lengths; i.e., if $x \in C$, $f_{tj}(x) = \gamma_{tj}(C)\, d_G(S_{tj}, v)$.

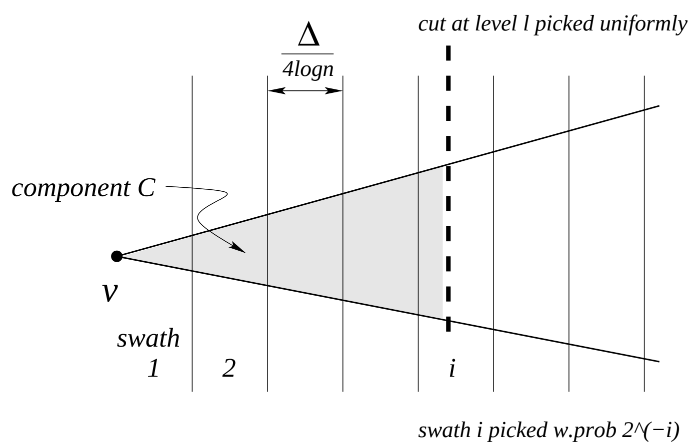

Figure 26.3: Generate-Coordinate: Creating a coordinate

We will now show that the algorithm given above satisfies (26.4). Define a particular coordinate f_{tj} created by the algorithm to be *eligible* if all the components created during its creation have a diameter of at most Δ.

Lemma 26.6 *A coordinate is eligible with probability* $1 - \frac{1}{n}$.

Proof. If a coordinate is not eligible, then we must have chosen a level greater than $\Delta/2$, and hence a swath greather than $(2 \log n)$. Since the probabilities p_k decrease geometrically, the probability of this happening is at most $2^{-2\log n} = 1/n^2$. At each step we remove at least one node from the graph, so at most n cuts are made for the definition of the coordinate. Now the

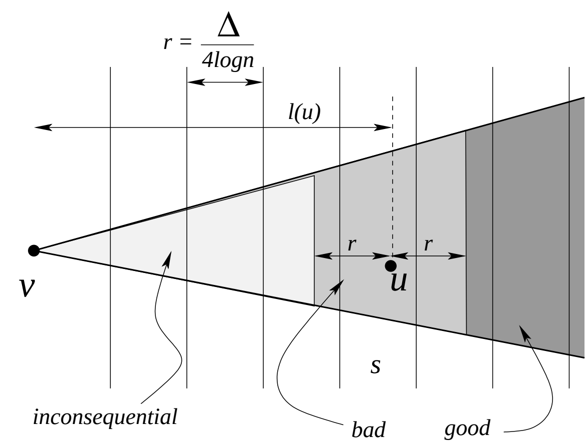

Figure 26.4: Cuts are $\Delta/(4\log n)$-good with constant probability

union bound implies that the probability that some component has diameter greater than Δ is at most $1/n$. ∎

For a given set S_{tj} defining a coordinate f_{tj}, let a node u be δ-good if $d(S_{tj}, u) \geq \delta$.

Lemma 26.7 *Each node $u \in G$ is $r = \Delta/(4\log n)$-good with constant probability.*

Proof. Consider some BFS tree rooted at v created in the above procedure, and let the cut be chosen at level l. Recall that $l(u) = d_{G'}(v, u)$. As pictured in Figure 26.4, the cut at level l is classified in one of three ways: it is *bad* if it lies between levels $l(u) - r$ and $l(u) + r$ and u is not r-bad already (since this cut causes u to become r-bad); the cut is *good* if $l > l(u) + r$ and u is not already r-bad, since u will be removed in this iteration and its distance from S cannot decrease with future cuts; otherwise, the cut does not matter. Note that there will eventually be a good or bad cut; thus the chance that u is r-bad is bounded by the chance that a bad cut happens before a good cut.

However, the probability p_{bad} that a cut is bad is bounded above by the probability that either u's swath in this BFS (say swath s) or one of the adjoining swaths ($s \pm 1$) is chosen, and hence $p_{bad} \leq 3.5\, p_s$. The chance that a cut is good is $p_{good} \geq \sum_{i \geq 2} p_{s+i} = p_s/2 \geq p_{bad}/7$. Hence the probability that we see a bad cut before a good one is at most $p_{bad}/(p_{bad} + p_{good} \leq 7/8$. Hence, with probability at least $1/8$, any node u will be $r = \Delta/(4\log n)$-good. ∎

Now we have the main ingredients of the proof in place.

Theorem 26.8 *For any graph G of diameter D, the above described procedure gives (whp) a (k, η)-volume respecting embedding with $\eta = (\log n\sqrt{\log D})$*

Proof. We first show that ϕ is a contraction. Since the random scaling parameters γ are at most 2, $|f_{tj}(u) - f_{tj}(v)| \leq 2\, d_G(u,v)$ in each coordinate. Now adding and scaling by $\sqrt{4L}$, we get $||\phi(u) - \phi(v)||_2 \leq \sqrt{\sum(4\, d(u,v)^2/(4L))} = d(u,v)$.

Recall that we want to show (26.4) for any ordered set $S = \{s_1, \ldots, s_k\}$, for any $1 \leq i \leq k$ and any fixed point a in the affine space $\mathcal{A}_k(\phi(S))$ generated by the first $i-1$ points of the ordering. To prove this, consider the coordinates f_{tj} created when the procedure "Generate-Coordinate" is called with parameter $\Delta \in [q_i/2, q_i]$; recall that $t = \log_2 \Delta$.

Fix such a coordinate f_{tj}; it is eligible if the diameter of each component is at most Δ, which in turn implies that s_i lies in a different component C_i from the points $\{s_1, \ldots, s_{i-1}\}$. Lemma 26.6 and 26.7 now prove that f_{tj} is both eligible and $\Delta/(4\log n)$-good with constant probability.

Now let us condition on all other random choices made in this coordinate except $\gamma_{tj}(C_i)$; this fixes $f_{tj}(s_1), \ldots, f_{tj}(s_{i-1})\}$. But now the fact that $\gamma_{tj}(C) \in_R [1,2]$, implies that the position of $f_{tj}(s_i)$ varies over an interval of length at least $\Delta/(4\log n)$.

Now using a Chernoff-Hoeffding bound, and taking c sufficiently large, we get that for a constant fraction of the $ck\log n$ coordinates f_{tj}, the coordinate $f_{tj}(s_i)$ varies over an interval of length least $\Delta/(4\log n)$ with probability $(1 - n^{-3k})$, independent of the positions of the first i points. Hence, $|f(s_i) - a|$ is at least $\sqrt{(\Delta/(4\log n))^2 \times \Theta(k\log n)} = \Omega(q_i \times \sqrt{k/\log n})$ with probability $(1 - n^{-3k})$. Finally, dividing by $\sqrt{4L} = \sqrt{4ck\log n \log D}$ to get ϕ from f, we get

$$\|\phi(s_i) - a\| \geq q_i/O(\log n\sqrt{\log D})$$

with the same probability, showing (26.4) and hence the theorem. ∎

27 Coda: Relationships to Euclidean Volumes

We finish the lecture by proving a lower bound on the Euclidean volume $\mathsf{Evol}(\phi(S))$ of any k-set of vertices S, under the embedding ϕ.

Corollary 27.9 *The Euclidean volume of the simplex formed by the images of a k-set S under a (k,η)-volume respecting embedding is*

$$\mathsf{Evol}(S) \geq \frac{\mathsf{Tvol}(s)}{(k-1)!\eta^{k-1}}$$

Proof. By assumption, we have $\|\phi(s_i), \mathcal{A}_{i-1}\phi(S)\|_2 \geq \frac{q_i}{\eta}$. Consider the simplex K_i formed by the images of $\{s_1, \ldots, s_i\}$. Since point $\phi(s_i)$ is at least q_i/η away from images of $\{s_1, \ldots, s_{i-1}\}$, it can be shown (by using standard facts about volumes of simplices) that the following inequality is true for K_i

$$\mathsf{Evol}(K_i) \geq \frac{q_i \times \mathsf{Evol}(K_{i-1})}{\eta(i-1)}$$

This inductively implies that

$$\mathsf{Evol}(K_k) \geq \frac{\prod_i q_i}{\eta^{k-1}\,(k-1)!} \geq \frac{\mathsf{Tvol}(S)}{\eta^{k-1}\,(k-1)!}$$

■

It follows that

$$\text{“Volume distortion of } S\text{”} \equiv \left(\frac{\mathsf{Tvol}(S)}{(k-1)!\ \mathsf{Evol}(\phi(s))} \right)^{1/(k-1)} \leq \eta$$

References

[Fei00] Uriel Feige. Approximating the bandwidth via volume respecting embeddings. *Journal of Computer and System Sciences*, 60(3):510–539, 2000. Also in *Proc. 30th STOC*, 1998, pp. 90–99.