Web Algorithms
Roger Wattenhofer

2 Embeddings

Wouldn't it be nice...

- ...to have points in a plane, each point representing a web site, points being close meaning that these web sites are related? In other words, to have a visual form of Google, where one could zoom in arbitrarily?
- ...to place all Internet hosts and routers onto the surface of a ball such that their latency or bandwidth is represented well by their Euclidean distance?
- ...to assign "virtual coordinates" to sensor or ad hoc network nodes such that two nodes are close with respect to their virtual coordinates if and only if the nodes are within direct transmission range, and then use a simple geo-routing algorithm (see Mobile Computing lecture) on top of that?
- ...to build a peer-to-peer system which finds data not only with few "hops," but also with without hopping from continent to continent.

All of these examples (and many more) are applications of a hot research area, known as geometric embeddings. In contrast to the last section on online algorithms, this area of "Web Algorithms" is not yet well understood. This last part of this course will therefore have more of a "seminar" style than previous sections. Similar to ski rental, we start by explaining the idea of embeddings for a very different example.

2.1 The Diameter Problem

You are given a set P with n points in the d-dimensional space. Think of n being huge, and d being small, in fact so huge/small that $n \gg 2^d$. We want to find the two points that have maximum Manhattan distance, that is

$$p, q \in P \text{ with } \|p-q\|_1 = \max_{p',q' \in P} \|p'-q'\|_1.$$

This problem can be solved in time $O(dn^2)$. However, with n being huge we might need an algorithm which is only linear in n. First, this sounds impossible. However, we can do this by taking a detour through embeddings. We first embed all points in d' dimensions with $d' = 2^d$, and then solve a simpler problem in d' dimensions.

Definition 2.1 *An embedding is a mapping* $f : P_a \rightarrow P_b$*, where* P_a *is a set of points in the source space, and* P_b *is a set of points in the target space.*

Algorithm 2.2 *We define the embedding $f(p)$ by specifying all its d' coordinates. Specifically, for each vector $s \in \{-1,+1\}^d$, we define $f_s(p) = s \cdot p$. Then, $f(p)$ is a vector obtained by concatenating the values of $f_s(p)$ for all $s \in \{-1,+1\}^d$. The maximum diameter in P is given by*

$$\max_{i=1\ldots d'} \left(\max_{p \in P} f(p)_i - \min_{q \in P} f(q)_i \right).$$

Theorem 2.3 *The algorithm is correct. The result can be computed in time $O(2^d n)$.*

Proof. For linear functions without additive terms f we have $f(a) - f(b) = f(a-b)$. Moreover we have

$$\|x\|_1 = sgn(x) \cdot x = \max_{\forall s \in \{-1,+1\}^d} s \cdot x = \|f(x)\|_\infty,$$

where $sgn(x)$ is a vector containing the signs of x. Thus, for two points $p, q \in P$ we have

$$\|f(p) - f(q)\|_\infty = \|f(p-q)\|_\infty = \|p-q\|_1.$$

Therefore

$$\begin{aligned}
\max_{p,q \in P} \|p-q\|_1 &= \max_{p,q \in P} \|f(p) - f(q)\|_\infty \\
&= \max_{p,q \in P} \max_{i=1\ldots d'} |f(p)_i - f(q)_i| \\
&= \max_{i=1\ldots d'} \left(\max_{p \in P} f(p)_i - \min_{q \in P} f(q)_i \right).
\end{aligned}$$

So far for correctness. Efficiency analysis has two parts:

- The first part is to compute the embedding f. A naive algorithm needs time $O(nd'd)$, since for each point (n many) and each value of the vector f (d' many) we must multiply/add d values. A better algorithm will reduce the running time to $O(nd')$; this faster algorithm is beyond the scope of in this lecture, however.

- The second part is to solve $\max_{i=1\ldots d'} \left(\max_{p \in P} f(p)_i - \min_{q \in P} f(q)_i \right)$. For each index ($d'$ many) we must study all points (n many), thus the running time is again $O(nd')$.

$\sqcap$

Remarks:

- Note that already the naive embedding algorithm gives a solution which only needs time $O(d2^d n)$. Therefore even the simple embedding solution is much faster for huge n and small d than the algorithm without the embedding detour.

- Our embedding had several special properties. First, it was an *isometry* ($\|f(p) - f(q)\|_\infty = \|p-q\|_1$); usually, embeddings are *distorted*. Also, our embedding was *oblivious*, meaning that for any point $p \in P$, the value $f(p)$ did not depend on the other points in P. Sometimes the points are given online (yes, ski rental!) – in such a case oblivious embeddings are important.

Generally speaking, there are several classes of embeddings:

- Some problems can be solved better when the problem is moved from a “difficult” norm (l_2) to an “easier” norm (l_∞).
- Many embeddings try to reduce the dimension. These embeddings allow us to “visualize” a set, or to run a faster (low-dimensional) algorithm.
- Some graph problems (networking!) can be simplified by moving a metric (see definition below) to a normed space; an example is in the next subsection. Similar arguments hold for other special source spaces (e.g. the edit distance, or the Haussdorff metric).

Definition 2.4 *A metric space (or simply metric) is a pair (X, D) where X is a set of points and D is a distance function such that for all $p, q \in X$:*

- $D(p, q) = 0 \Leftrightarrow p = q$
- $D(p, q) = D(q, p)$
- $D(p, r) + D(r, q) \geq D(p, q)$.

2.2 Bourgain’s Embedding

In this section we present the “mother of all embeddings” by Bourgain. It is an embedding from a high dimension to a low dimension; naturally, the dimension reduction makes it necessary to compromise:

Definition 2.5 *An embedding is a mapping $f : X \to Y$, where (X, D) is a metric space, and Y is a set of points in the normed space l_m^d (d-dimensional space with m-norm; $m = 1$: Manhattan, $m = 2$: Euclidean, etc.). If for any points $p, q \in X$ we have*

$$\frac{1}{c} D(p, q) \leq \|f(p) - f(q)\|_m \leq D(p, q),$$

then the embedding has distortion c.

Remark:

- As you might have guessed there are more general definitions of embeddings.

Theorem 2.6 (Bourgain) *Any finite metric (X, D) can be embedded into l_2^d with $d < \infty$ with distortion $O(\log |X|)$.*

We prove a variant of Bourgain’s Lemma which features a less technical proof:

Theorem 2.7 *For any integer $b > 0$ let $c = 2b - 1$. Let M be a finite metric (X, D). Then M can be embedded into l_∞^d with distortion c and $d = O(bn^{1/b} \log n)$.*

Proof. Studying the special case $b = 1$ gives additional insights. So let’s do this first! Consider the mapping $f : X \to l_\infty^d$ defined as

$$f(q) =< D(q, p_1), \ldots, D(q, p_n) > \text{ where } X = \{p_1, \ldots, p_n\}.$$

Then, using the triangle inequality (in the form $D(a,c) \geq |D(a,b) - D(b,c)|$),

$$\|f(p) - f(q)\|_\infty = \max_{p_i \in X} |D(p,p_i) - D(q,p_i)| \leq D(p,q).$$

On the other hand, for $p_i = p$ (or $p_i = q$) we have

$$\|f(p) - f(q)\|_\infty = \max_{p_i \in X} |D(p,p_i) - D(q,p_i)| \geq |D(p,p) - D(q,p)| = D(p,q).$$

Therefore, f is an isometry:

$$\|f(p) - f(q)\|_\infty = D(p,q).$$

The general case is similar. The main difference is that the individual points p_i are replaced by sets of points $A_i \subset X$, and the distance $D(q,p_i)$ is replaced by $D(q,A_i) = \min_{a \in A_i} D(q,a)$.
Formally, the embedding f is defined as

$$f(q) =< D(q,A_1), \ldots, D(q,A_d) >$$

For any choice of sets A_i, the embedding is a contraction. Let a_p and a_q be points in A_i which minimize $D(p,A_i)$ and $D(q,A_i)$, respectively. Then,

$$\begin{aligned} \|f(p) - f(q)\|_\infty &= \max_{\forall A_i} |D(p,A_i) - D(q,A_i)| \\ &= |D(p,a_p) - D(q,a_q)| \\ &\leq |D(p,a_q) - D(q,a_q)| \leq D(p,q). \end{aligned}$$

For the other side, $\|f(p) - f(q)\|_\infty \geq D(p,q)/c$, we need to choose the sets A_i carefully. We need an additional Lemma with Definition:

Definition 2.8 *A ball $B(p,r)$ of radius $r \geq 0$ around point p is a set of points q for which $D(q,p) \leq r$.*

Lemma 2.9 *For any pair of points $p,q \in X$ there is an $r > 0$ such that*

$$\frac{|B(q, r + D/c)|}{|B(p,r)|} \leq n^{1/b}, \text{ where } D = D(p,q).$$

We call the two balls B_q and B_p respectively. We have $B_q \cap B_p = \emptyset$.

Proof. We start out with $r = 0$ so that $B(p,0) = \{p\}$. If $\frac{|B(q,D/c)|}{|B(p,0)|} \leq n^{1/b}$ we are done. If not, we know that $|B(q,D/c)| > n^{1/b}$. In this case we switch p and q and consider $r = D/c$, then $r = 2D/c$, then $r = 3D/c$, etc. If any of the ratios encountered during this process is small, we are done. However, if not, after b steps we have $|B(q,bD/c)| > n^{1/b^b} = n$ (or a symmetric statement for p), which is a contradiction. Therefore the first part of the Lemma holds. Since $c = 2b - 1$, the sum of the two Ball radii is less than D, therefore $B_q \cap B_p = \emptyset$. ⊓

Back to the proof of Theorem 2.7. Now we are in the position to construct the sets A_i. For each $s = 0, 1, \ldots, b$ we randomly construct $O(n^{1/b} \log n)$ sets A_i. Specifially, in each set A_i, each point in X is included with probability $1/n^{s/b}$. In order to prove Theorem 2.7 we need another helper Lemma:

Lemma 2.10 *Let's fix a specific pair* p, q *with* $B_q > B_p$ *in Lemma 2.9. Let* s *be such that* $n^{s/b} \approx B_q$*. Then, with high probability, for one of the sets* A_i *we have* $A_i \cap B_q = \emptyset$ *and* $A_i \cap B_p > \emptyset$*.*

Proof. To simplify the proof, we assume $n^{s/b} = B_q$. Then, with constant probability $A_i \cap B_q = \emptyset$. Similarly, using $\frac{|B_q|}{|B_p|} \leq n^{1/b}$ from Lemma 2.9, with at least probability $1/n^{1/b}$ we have $A_i \cap B_p > \emptyset$. Since we randomly constructed $O(n^{1/b} \log n)$ sets A_i of type s, the usual formulas ($(1 + t/n)^n \approx e^t$) will reveal that with high probability one of the sets fulfills the properties of the Lemma. $\sqcap$

Back to the proof of Theorem 2.7. Now, for each pair p, q, with high probability we have a set A_i which fulfills Lemma 2.10. Therefore,

$$\|f(p) - f(q)\|_\infty = \max_{\forall A_i} |D(p, A_i) - D(q, A_i)| \geq |D(q, A_i) - D(p, A_i)| \geq r + D/c - r = D/c.$$

This proves the remaining inequality of Definition 2.5. This concludes the proof of Theorem 2.7 (with high probability). $\sqcap$

Remarks:

- Of particular interest is $b = \log n$, which gives a logarithmic distortion in the log-squared dimension. Using $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{d}\|x\|_\infty$, we also get a weaker but simpler proof of Bourgain's Lemma with loq-squared distortion.
- In fact, there is a lower bound argument by Linial, London, and Rabinovich which proves that the logarithmic distortion of Bourgain is tight.
- As of now, many other embeddings have been discovered. Many researchers have studied restricted families of metric spaces which feature a better distortion. For example, it is conjectured that planar graphs using the shortest-path metric can be embedded with only constant distortion.
- Recently, Feige presented a new generalization of embeddings, where not only distances have low distortion, but also volumes.
- In practice, many graphs are embedded using so-called *spring embedding* algorithms. So far, neither running time nor quality (distortion) of these algorithms have been proved formally. Still, for many applications the results look very promising.