15-859(J): Metric Embeddings and Methods **Lecturer:** R. Ravi, A. Gupta
Topic: Finding closest tree metrics; Dimensionality Reduction **Date:** October 16, 2003
Scribe: Virginia Vassilevska

14.1 Introduction

Recall the following from last class:

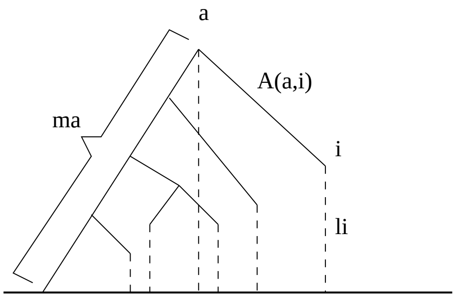
1. Let D be a set of distances between points in X, D not necessarily a metric. Then there exists a unique subdominant ultrametric S_D with generating tree MST_D which gives us

$$\underset{U\ ultrametric}{\arg\min}\ ||U, D||_\infty$$

2. "Interchange Lemma": Additive metric A + a-fixed centroid metric C^a = ultrametric.

3. "Centroid Addition Corollary": Additive metric A + centroid quasimetric C is additive as long as $A + C$ obeys the triangle inequality and is nonnegative. (one can see this since both A and C obey the 4 point condition).

Also recall how we get C^a in 2: Let $m_a = \max_i A(a,i)$, then $C(i,j) = l_i + l_j$, where $l_i = m_a - A(a,i)$. (See Fig. 14.1.)



Definition 14.1.1 *A metric M^a is a-fixed with respect to (X, D) for $a \in X$ if $M^a(a,i) = D(a,i)$.*

Definition 14.1.2 *Let us define the following "optimal" metrics:*

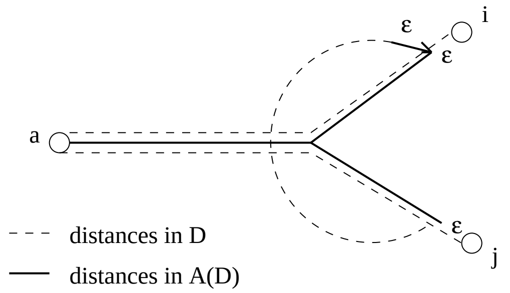
$$\mathcal{A}(D) = \underset{A\ additive}{\arg\min}\ ||D, A||_\infty$$
$$U(D) = \underset{U\ ultrametric}{\arg\min}\ ||D, U||_\infty$$
$$\mathcal{A}^a(D) = \underset{A\ a-fixed\ additive}{\arg\min}\ ||D, A||_\infty$$

14.2 3-approximation for $L_\infty(D, \mathcal{A})$

The following lemma shows that, to find an approximation to the closest additive metric, it is enough to focus our attention on a-fixed additive metrics. The lemma was proven in an earlier lecture on finding the closest path metric to D under the ℓ_∞ norm. We'll do it again for completeness.

Lemma 14.2.1 $||\mathcal{A}^a(D), D||_\infty \leq 3||\mathcal{A}(D), D||_\infty$.

Proof: The idea behind the proof is to note that $\mathcal{A}(D)$ stretched or contracted D-distances between pairs of points by at most some ϵ. To get $\mathcal{A}^a(D)$ for each point we bring the distance back to what it was in D. Then for each distance between pairs of points we pay at most ϵ, and another 2ϵ for the two point distances from a. So overall when we bring the distances from a back, we pay at most 3ϵ (see Fig. 14.2). This is a bit vague, so let us make it more precise:



Derive some tree T^a from some optimal $\mathcal{A}(D)$ by moving points in X independently to match appropriately distances in D from a. Note, that points may cross over and change places, but this turns out not to matter. We have:

$$|D(i,j) - T^a(i,j)| \leq |\mathcal{A}(D)(i,j) - T^a(i,j)| + |D(i,j) - \mathcal{A}(D)(i,j)| \leq \epsilon(a,i) + \epsilon(a,j) + \epsilon(i,j) \leq 3\epsilon.$$

Above, $\epsilon = \max_{i,j} \epsilon(i,j)$, where $\epsilon(i,j)$ is how much we need to move the distance from i to j when we are matching the distances to a, and is clearly no more than $||A(D), D||_\infty$. The inequality above is obtained by the triangle

inequality and proves Lemma 14.2.1. ∎

14.2.1 Polynomial Time Algorithms

It now remains to show that we can, in fact, find $\mathcal{A}^a(D)$.

Lemma 14.2.2 *$\mathcal{A}^a(D)$ is computable in polynomial time.*

Proof: The following definition will be useful for the proof:

Definition 14.2.3 *U is an a-fixed ultrametric (w.r.t. D) if*

1. $U(i,j) \geq 2\max\{l_i, l_j\}$
2. $U(a,i) = 2m_a$, $\forall i$

where $m_a = \max_j D(j,a)$*, and* $l_i = m_a - D(i,a)$*.*

To get some intuition for this definition, consider the case when D is some additive metric A. From the interchange lemma, we know $U = A + C^a$ is an ultrametric; it is easy to verify that U is also an a-fixed ultrametric w.r.t. A.

The proof proceeds by proving the following claims:

Claim 14.2.4 $T \doteq U(D + C^a) - C^a$ *is an* a*-fixed additive metric.*

Claim 14.2.5 $A^a(D) + C^a$ *is an ultrametric.*

Before we prove the claims, let us complete the proof of Lemma 14.2.2:

$$\begin{aligned} ||T, D||_\infty &\geq ||\mathcal{A}^a(D), D||_\infty = ||\mathcal{A}^a(D) + C^a, D + C^a||_\infty \\ &\geq ||U(D + C^a), D + C^a||_\infty = ||U(D + C^a) - C^a, D + C^a - C^a||_\infty \\ &= ||T, D||_\infty. \end{aligned}$$

The first inequality is because by Claim 14.2.4 T is in the class of a-fixed additive metrics, and $A^a(D)$ is the closest one to D by definition. The second and the final conclusions follow since adding the same amounts to each metric does not change the $||\cdot||_\infty$. The second inequality comes from Claim 14.2.5 and the fact that $U(D + C^a)$ is the closest ultrametric to $D + C^a$. The above inequality shows that actually

$$||T, D||_\infty = ||\mathcal{A}^a(D), D||_\infty.$$

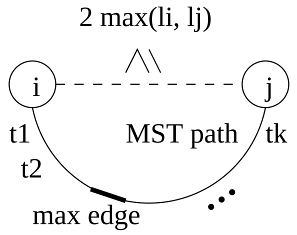
Therefore, to compute $\mathcal{A}^a(D)$ we only need to find T. ■

Now to prove the claims: note that Claim 14.2.5 follows directly from the interchange lemma. Therefore, we will concentrate on Claim 14.2.4:

Lemma 14.2.6 $U(D + C^a)$ *is* a*-fixed.*

Proof: Say, without loss of generality, $l_i = \max\{l_i, l_j\}$. Then,

$$\begin{aligned} (D + C^a)(i,j) &= D(i,j) + l_i + l_j \\ &= D(i,j) + m_a - D(i,a) + m_a - D(j,a) \\ &= m_a - D(i,a) + m_a - (D(j,a) - D(i,j)) \\ &\geq m_a - D(i,a) + m_a - D(i,a) \\ &= 2l_i = 2\max\{l_i, l_j\}. \end{aligned}$$



Now let $\{i = t_0, t_1, \dots, t_k, t_{k+1} = j\}$ be the path in the MST from i to j. From the above we know that for each p, $(D + C^a)(t_p, t_{p+1}) \geq 2\max\{l_{t_p}, l_{t_{p+1}}\}$. But then

$$(D + C^a)(t_p, t_{p+1}) \geq 2(m_a - \min\{D(t_p, a), D(t_{p+1}, a)\}).$$

(See Fig. 14.2.1). Since this is true for all p, and as $U(D + C^a)(i, j) = \max_p (D + C^a)(t_p, t_{p+1})$, we have that

$$U(D + C^a)(i, j) \geq 2(m_a - \min\{D(t_0, a), D(t_{k+1}, a)\}) = 2\max\{l_i, l_j\}$$

To show that the second property for a-fixed ultrametrics holds, note that for every $i \in X, i \neq a$ we have $(D + C^a)(i, a) = 2m_a$. Also note that for every i and j we have $(D + C^a)(i, j) \leq 2m_a$. So when we construct the MST for $D + C^a$ the largest edge on the path from a to each i will have weight $2m_a$, and so $U(D + C^a)(i, a) = 2m_a$. Therefore $U(D + C^a)$ is an a-fixed ultametric. ■

Proof of Claim 14.2.4: Since $U(D + C^a)$ is an a-fixed ultametric $U(D + C^a)(a, i) = 2m_a$ for every $i \neq a$. Thus,

$$T(a, i) = 2m_a - C^a(a, i) = 2m_a - (m_a - D(a, i) + m_a) = D(a, i).$$

And T is a-fixed.

To check that T is additive, we only need to check whether the triangle inequality holds for $U(D + C^a) - C^a$: since $U(D + C^a)$ is an ultrametric, we know that in any triangle the maximum weight edge is nonunique. Therefore:

$$U(D + C^a)(i, j) \leq \max\{U(D + C^a)(i, k), U(D + C^a)(k, j)\}. \tag{14.2.1}$$

Furthermore, since $U(D + C^a)$ is a-fixed,

$$\min\{U(D + C^a)(i, k), U(D + C^a)(k, j)\} \geq 2l_k \tag{14.2.2}$$

Thus, overall we get

$$\begin{aligned} U(D + C^a)(i, j) \leq & \max\{U(D + C^a)(i, k), U(D + C^a)(k, j)\} \\ & + \min\{U(D + C^a)(i, k), U(D + C^a)(k, j)\} - 2l_k \end{aligned} \tag{14.2.3}$$

But this is the same as

$$U(D + C^a)(i, j) \leq U(D + C^a)(i, k) + U(D + C^a)(k, j) - 2l_k \tag{14.2.4}$$

Then,

$$\begin{aligned} U(D + C^a)(i, j) &\leq U(D + C^a)(i, k) + U(D + C^a)(k, j) - l_i - l_k - l_k - l_j + l_i + l_j \\ &= U(D^a)(i, k) + U(D^a)(k, j) - C^a(i, k) - C^a(k, j) + C^a(i, j) \end{aligned}$$

The above expression shows that

$$T(i, j) = (U(D + C^a) - C^a)(i, j) \leq (U(D + C^a) - C^a)(i, k) + (U(D + C^a) - C^a)(k, j)$$

and thus $T(i, j)$ obeys the triangle inequality. Note that $T(i, j)$ is clearly nonnegative since $U(D + C^a)$ is an a-fixed ultrametric (and thus $U(D + C^a)(i, j) \geq 2\max\{l_i, l_j\}$). Hence T is an a-fixed additive metric. ■

14.3 Dimension Reduction in l_p spaces

The question that we will address in this discussion is the following: given a set of points in some $\ell_p^{k'}$, can we reduce the number of dimensions required without distorting distances too much? Formally,

Problem 14.3.1 *Given $(X, ||\cdot||_p) \subseteq \mathbf{R}^{k'}$, what is the least $k = k(n, D)$ such that*

$$(X, ||\cdot||_p) \overset{D}{\hookrightarrow} \ell_p^k?$$

It turns out that the answer depends on p. E.g., for ℓ_∞, almost no dimension reduction is possible. For l_1, a lower bound was recently shown by Brinkman and Charikar (2003), and much simpler exposition was given by Lee and Naor (2003):

Theorem 14.3.2 *There exists a set of n points in ℓ_1 which require $k = n^{\Omega(1/D^2)}$.*

However, the situation is very good for Euclidean space ℓ_2, as a seminal paper of Johnson and Lindenstrauss (1984) shows:

Theorem 14.3.3 *For any set of n points and $D = 1 + \epsilon$, it is possible to get down to $k = O(\frac{\log n}{\epsilon^2})$ dimensions.*

Furthermore, Alon (2001) showed that this result is almost tight:

Theorem 14.3.4 *With $D = (1+\epsilon)$, the uniform metric on n points requires at least $k = \Omega(\frac{\log n}{\epsilon^2 \log \frac{1}{\epsilon}})$ dimensions.*

In the next few lectures, we will see proofs of most of these theorems.

14.3.1 Introduction to the l_2 Dimension Reduction

Any set of n points in Euclidean space can be embedded in n dimensions. (Why?) So assume we have n points $x_1, \ldots, x_n \in \mathbf{R}^n$. Then we will show that there exists an embedding ϕ such that $\phi(x_i) = y_i \in \mathbf{R}^k$ where $k = O(\frac{\log(n)}{\epsilon^2})$. In fact, let us give two specific constructions for ϕ as follows: Let us define two maps using this matrix: let us populate the entries of a $k \times n$ matrix A with i.i.d. $N(0, 1)$ ("standard normal") random variables, and define

$$\phi(\overline{x}) = A\overline{x}.$$

For the other map, define a $k \times n$ matrix B with random $\{+1, -1\}$ variables, and then set

$$\phi_B(\overline{x}) = B\overline{x}.$$

The following result is a strengthening of Theorem 14.3.3, which was proved for the linear transformation $\phi : \mathbb{R}^n \to \mathbb{R}^k$ being given by projection onto a random k-dimensional subspace.

Theorem 14.3.5 (Indyk and Motwani (1998), Achlioptas (2000)) *Both ϕ and ϕ_B are distortion $(1 + \epsilon)$ maps with probability at least $1/2$.*

Note that this gives a randomized algorithm for dimension reduction. Recent results of Engebretsen, Indyk, and O'Donnell (2002) and Sivakumar (2002) show how to find such maps ϕ in deterministic polynomial time.

14.3.2 Some Preparation

Given a vector $\overline{v} = (v_1, \ldots, v_n) \in \mathbf{R}^n$, suppose we pick a random vector $\overline{R} = (R_1, \ldots, R_n)$ such that each $R_i \sim N(0,1)$. Then it is easy to see that $X = < \overline{v}, \overline{R} >$ is distributed $N(0, \ell^2)$ where $\ell = ||\overline{v}||_2$:

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n} v_i R_i\right] = \sum_{i=1}^{n} v_i \mathbf{E}[R_i] = 0, \text{ and} \tag{14.3.5}$$

$$\mathbf{Var}[X] = \sum_{i=1}^{n} \mathbf{Var}[v_i R_i] = \sum_{i=1}^{n} v_i^2 = \ell^2. \tag{14.3.6}$$

References

Achlioptas, D. (2000). Database friendly random projections. In *Proceedings of the Twenteenth ACM Symposium on Principles of Database Systems*.

Brinkman, B. and Charikar, M. S. (2003). On the impossibility of dimension reduction in ℓ_1. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 514–523.

Engebretsen, L., Indyk, P., and O'Donnell, R. (2002). Derandomized dimensionality reduction with applications. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 715–712.

Indyk, P. and Motwani, R. (1998). Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 604–613.

Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206.

Lee, J. R. and Naor, A. (to appear). Embedding the diamond graph in ℓ_p and dimension reduction in ℓ_1. In *Geometric and Functional Analysis (GAFA)*.

Sivakumar, D. (2002). Algorithmic derandomization via complexity theory. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 619–626. ACM Press.