1   Definition

Today we talked about Earth Mover Distance (EMD) which we mentioned in the previous lecture. Given two sets A and B of points in a metric space, EMD(A,B) is the cost of the minimum cost bipartite matching between points of A and B, i.e. sum of the cost of corresponding distances. We shall focus on 2-dimensional A and B. We can embed this into l_1 with some distortion.

2   Embedding EMD into l_1

In metric embedding, we generally try to embed harder space into easier space. We shall not prove this, but we state that EMD is at least as hard as l_1. We can embed $\{0,1\}^d$ into EMD with distortion 1. We pick l_1 over l_2 as we already have sketches and techniques for l_1 and l_1 is richer than l_2. We will be focusing on $[\Delta]^2$ integer grid.

Theorem 1 (Charikar'02, Indyk-Thaper'03)**.** *We can embed EMD over $[\Delta]^2$ into l_1 with distortion* $O(\log \Delta)$. *In fact, will construct a randomized* $f : 2^{[\Delta]^2} \to l_1$ *such that:*
for any $A, B \subset [\Delta]^2$: $EMD(A, B) \leq E[\| f(A) - f(B) \|_1] \leq O(\log \Delta).EMD(A, B)$
Time to embed a set of s points:$O(s \log \Delta)$

Here,$2^{[\Delta]^2}$ are all the subsets of the integer grid.Note that we can have duplicate points.
Consequences
Nearest Neighbor Search: $O(c \log \Delta)$ approximation with $O(sn^{1+1/c})$ space and $O(n^{1/c}s \log \Delta)$ query time. (i.e sublinear query time with $O(\log \Delta)$ approximation) where n is the size of data space.
Computation: $O(\log \Delta)$ approximation in $O(s \log \Delta)$ time. Best known is 1+ϵ approximation in O(s) time. [AS'12]

2.1   What if $|A| \neq |B|$?

We consider the case when the sizes of the two sets are unequal. Let B be the smaller set. So, $|A| = a$ and $|B| = b$ where $b < a$. We define

$$EMD_\Delta(A, B) = \Delta(a - b) + min_{A', \pi} \sum_{a \in A'} d(a, \pi(a))$$

where the last term denotes the minimum cost matching where minimum is taken over A' which ranges over all subsets of A of size b and π is a bijection from $A' \to B$ which ranges over all 1-to-1 mappings.For optimal A' we call all $a \in A/A'$ unmatched.
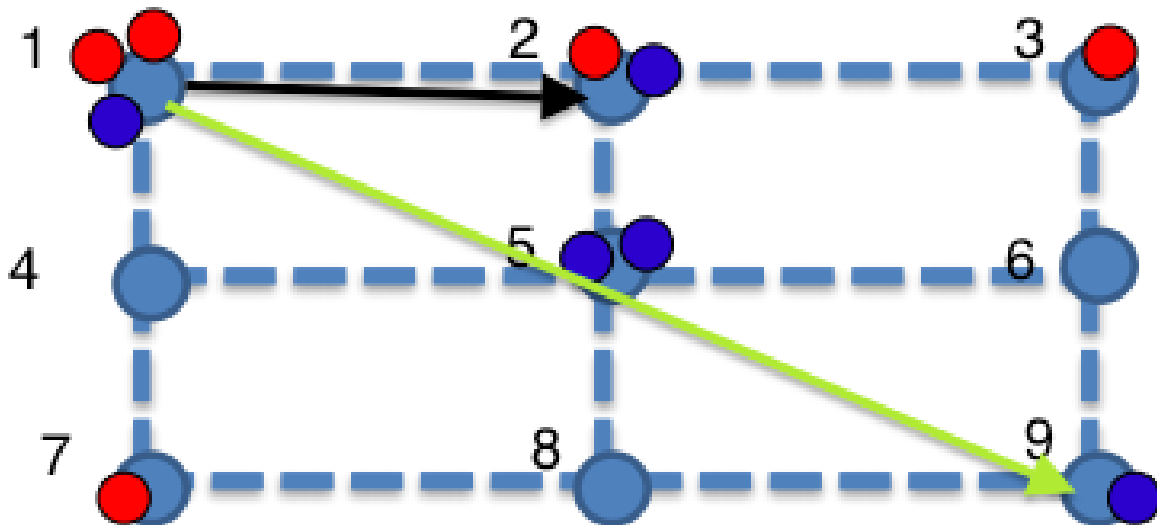
3 Embedding EMD over small grid

Suppose $\Delta = 3$

$f(A)$ has nine coordinates counting number of points in each integer point.

$f(A) = (2, 1, 1, 0, 0, 0, 1, 0, 0)$

$f(B) = (1, 1, 0, 0, 2, 0, 0, 0, 1)$

Red and blue points in the same integer point can be matched between themselves. So, after matching all those points, we have each integer point in the grid with either blue or red points. For example, here A contributes one free red point in the first integer point in the grid (as 2-1=1). We claim that there is a $2\sqrt{2}$ distortion embedding. We shall see why.
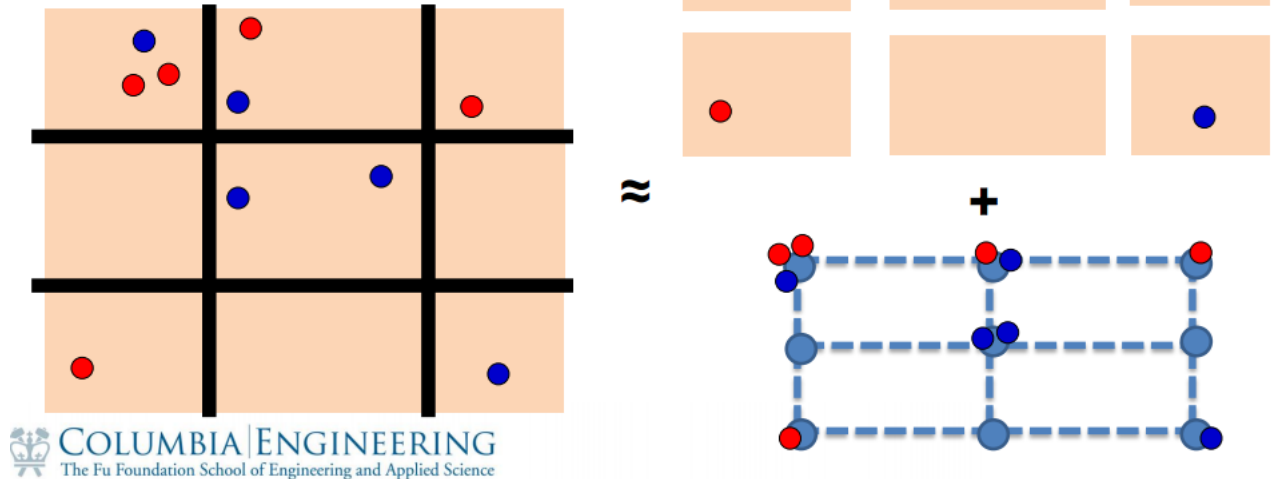
To match the free red point here, it has to pay at least 1 unit cost. This happens when it is matched to one of the blue points in its neighbouring integer points (tracing the black arrow in the figure above) and so distance between them is 1 unit. But in the worst case, it may be mapped to a blue point across the diagonal in the last integer point(tracing the green arrow in the figure above). Now, distance between them is $2\sqrt{2}$ units. So it will end up paying $2\sqrt{2}$ unit costs. Hence, the optimal lies between $(1, 2\sqrt{2})$. Distortion can be seen as the ratio of the largest distance between points to the smallest distance. If we have to scale this up for a general Δ, Distance of this embedding would be $\Delta\sqrt{2}$.

4 High level embedding

This is set in a $[\Delta]^2$ box. For embedding of A, we take a quad-tree/ grid of cell size $\Delta/3$. We randomly shift it. We keep on subpartitioning each cell into 3X3 till we get unit sized cells, Note that the grid is randomly shifted only once, all the following subpartitions are deterministic. We write down $f(A)_c$ which gives us the number of points in cell c.

We want to prove that $E[\| f(A) - f(B) \|_1] \approx EMD(A, B)$, where E[$\| f(A) - f(B) \|_1$] is the expected l_1 distance.

The main intuition here is to see the same embedding in a different way.

We decompose EMD over $[\Delta]^2$ into EMDs over smaller grids. We recursively break down the problem till $\Delta = O(1)$.

So we have that the cost of the original problem is roughly equal to the sum of the cost of each of the smaller boxes (which represent the low level structure) plus the cost of the higher perspective. For each decomposition we pay a constant cost and we have roughly $\log \Delta$ such decomposition steps.

It is important to note here that we have to consider the cost of the higher perspective because if the points do not interact locally, their interaction is not captured in the low level structure. This is because the EMD inside a given cell does not know about anything else. Hence, if it has to match a red point but there is no blue point available locally, it does not know where the nearest blue point is located globally (outside the cell). Hence, it will not be able to contribute cost that is proportional to the distance between the red and blue point. So, we need the higher perspective to calculate it.
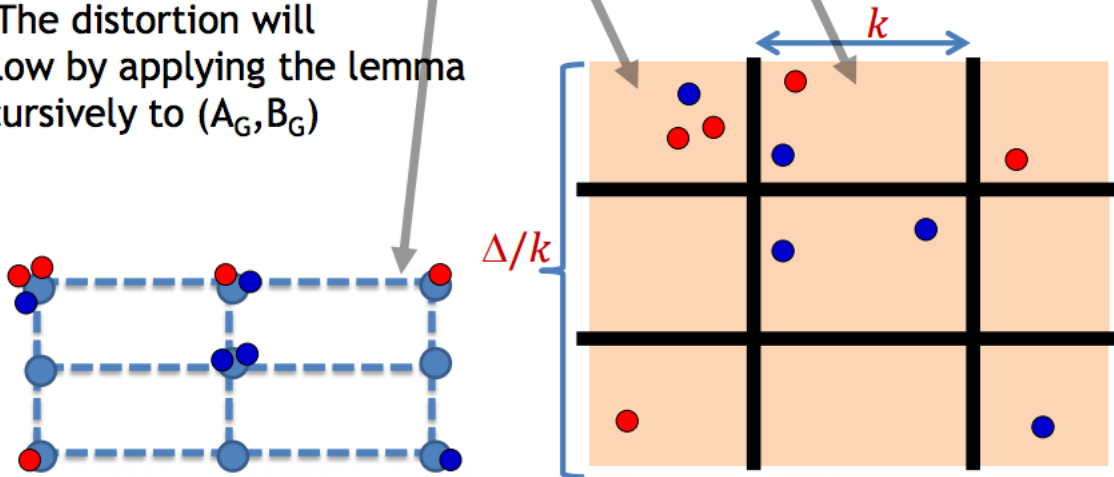
5   Decomposition Lemma

- For randomly-shifted cut-grid G of side length k, will prove:
  1) $EMD_\Delta(A, B) \leq \boxed{EMD_k(A_1, B_1)} + \boxed{EMD_k(A_2, B_2)} + \cdots$
     $+ k \cdot \boxed{EMD_{\Delta/k}(AG, BG)}$
  2) $EMD_\Delta(A, B) \geq \frac{1}{3} E[EMD_k(A_1, B_1) + EMD_k(A_2, B_2) + \cdots]$
  3) $EMD_\Delta(A, B) \geq E[k \cdot EMD_{\Delta/k}(A_G, B_G)]$
- The distortion will follow by applying the lemma recursively to (A_G, B_G)

Here k is a constant and Δ/k is the number of cells. Here, we are essentially trying to get an upper and lower bound. $EMD(A_i, B_i)$ corresponds to what happens inside cell i. Part 2 and 3 have randomization. Here, we note that a single randomization fixes all the following levels because of linearity of expectation.

5.1   Lower bound

For a randomly-shifted cut-grid G of side-length k: $EMD_\Delta(A, B) \leq EMD_k(A_1, B_1) + EMD_k(A_2, B_2) + ...$
$+ k.EMD_{\Delta/k}(A_G, B_G)$

We want to solve this by doing some simpler operations $(EMD_k(A_i, B_i)$ and relating them.
We need to construct a matching. Note that the matching doesn't have to be optimal;it is enough to construct a matching that has cost less than this upper bound.
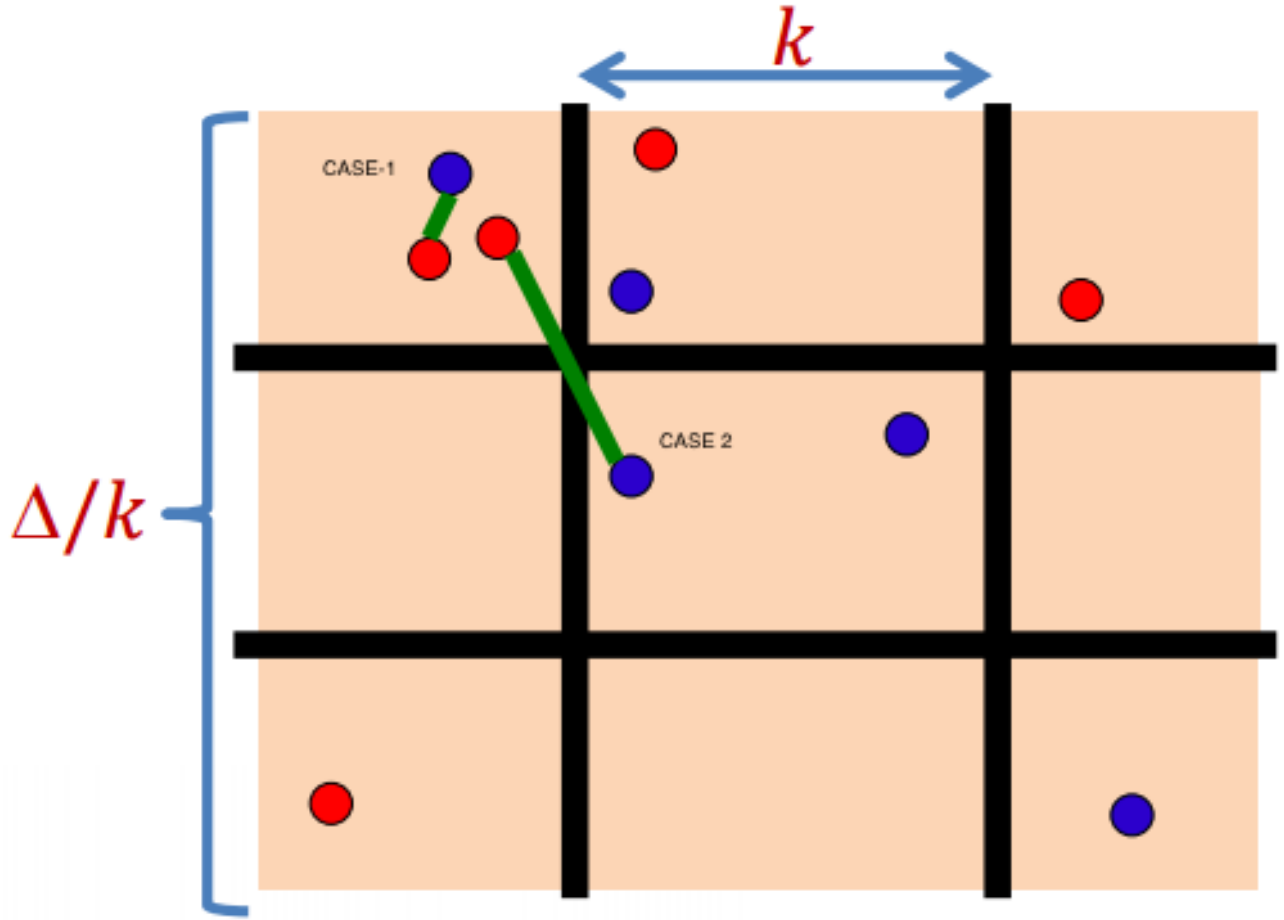
Construct a matching π on $EMD_\Delta(A, B)$ from the matchings on RHS as follows:

-For each $a \in A(suppose a \in A_i)$ it is either:

1). Matched in $EMD(A_i, B_i)$ to some $b \in B_i (if a \in A_i')$

2).Or $a \notin A_i'$) and so it is matched in $EMD(A_G, B_G)$ to some $b \in B_j (j \neq i)$ - then $\pi(a) = b$

- that is, if a belongs to some cell i then it will either be matched inside the cell (in which case we will retain this matching on LHS) or it will be matched outside the cell (which occurs when a is extra in A_i and so the set has extra points in the high level grid and has to be matched to a point in some other cell.).

For calculating the cost we do the following :

If it corresponds to case-1, the cost is paid by $EMD(A_i, B_i)$.

If it corresponds to case-2, we move a to the centre (Δ) (which we charge to $EMD(A_i, B_i)$) and then we move from cell i to j and charge k units to $EMD(A_G, B_G)$.

If size of A is greater than B, we have to pay for the extra unmatched points. So, we pay $k.(\Delta/k) = \Delta$ on the LHS and the RHS.

5.2 Upper bound

For a randomly-shifted cut-grid G of side-length k:
$EMD_\Delta(A, B) \geq \frac{1}{3}E[EMD_k(A_1, B_1) + EMD_k(A_2, B_2) + ...]$
$EMD_\Delta(A, B) \geq E[k.EMD_{\Delta/k}(A_G, B_G)]$

This is where we use the randomness. We fix a matching π which minimizes $EMD_\Delta(A, B)$ (optimal matching), which is represented by the green lines in the image. We will construct matchings for each EMD on the RHS. Place the grid randomly. We observe that some of the green edges are preserved uncut while some are cut. (If no green edges are cut, we get the equality). The uncut pairs are matched within their respective cell and hence remain unchanged. The cut pairs on the other hand, are unmatched in their respective mini-grid and hence have to be matched in (A_G, B_G) i.e the high level grid.
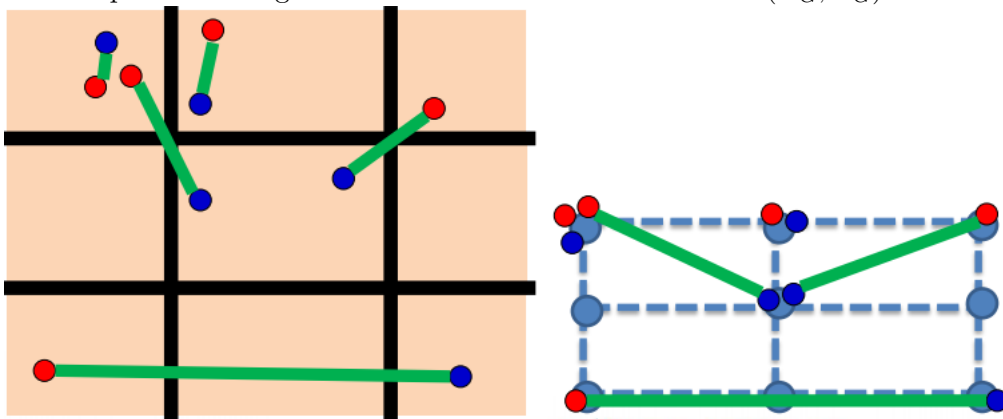
We claim that: $3.EMD_\Delta(A, B) \geq E[EMD_k(A_1, B_1) + EMD_k(A_2, B_2) + ...]$
As uncut pairs are matched in their respective cells, their total contribution to the cost is $\leq EMD_\Delta(A_G, B_G)]$ since we extract them from optimal matching.
For cut pairs we do the following analysis.
We observe that probability that a pair is cut decreases with distance. So, if a pair is very close, the probability that it is cut is very small.
We consider a cut pair (a,b) at a distance $a - b - (d_x, d_y)$. (a,b) contributes to the RHS as the two points may be unmatched in their mini-grids.
$Pr[(a,b)\text{is cut}] = 1 - (1 - \frac{d_x}{k})_+(1 - \frac{d_y}{k})_+ \leq \frac{d_x}{k} + \frac{d_y}{k} \leq \frac{1}{k} \parallel a - b \parallel_2$
Expected contribution of (a,b) to RHS:
$\leq Pr[(a,b)\text{is cut}].2k \leq 2 \parallel a - b \parallel_2$
(here we have an upperbound of 2k as a may be extra point in one cell and b may be extra point in another cell).
Hence, the total expected contribution to the cost on RHS $= 2.EMD_\Delta(A, B)]$
Hence, the total contribution from cut and uncut pairs is $3.EMD_\Delta(A, B)]$

Now we consider the higher level grid.
We claim that:
$EMD_\Delta(A, B) \geq E[k.EMD_{\Delta/k}(A_G, B_G)]$
As the uncut pairs have already been matched in the lower level structure and paid the cost of matching,

their contribution in the higher level is 0. Hence, they contribute nothing to the RHS here.

Cut pair: $(a, b) \in \pi$ with $a - b = (d_x, d_y)$

if $|d_x| = xk + r_x$ and $|d_y| = yk + r_y$

then expected cost contribution to $k.EMD_{\Delta/k}(A_G, B_G)$]:

$\leq (x + \frac{r_x}{k}).k + (y + \frac{r_y}{k}).k = d_x + d_y = \| a - b \|_2$

Hence, total expected cost $\leq EMD_\Delta(A, B)$

For a randomly-shifted cut-grid G of side-length k: $1)EMD_\Delta(A, B) \leq EMD_k(A_1, B_1) + EMD_k(A_2, B_2) +$

...

$+ k.EMD_{\Delta/k}(A_G, B_G)$

$2)EMD_\Delta(A, B) \geq \frac{1}{3}E[EMD_k(A_1, B_1) + EMD_k(A_2, B_2) + ...]$

$3)EMD_\Delta(A, B) \geq E[k.EMD_{\Delta/k}(A_G, B_G)]$

We apply decomposition recursively for k=3.

Choose randomly shifted cut-grid G_1 on $[\Delta]^2$

Obtain many grids $[3]^2$ and a big grid $[\Delta/3]^2$

Then choose randomly shifted cut-grid G_1 on $[\Delta/3]^2$

Obtain more grids $[3]^2$ and a big grid $[\Delta/9]^2$

Keep repeating this process till bigger grid is $[3]^2$.

Embed each of the small $[3]^2$ grids into l_1 using O(1) distortion embedding and concantenate all the embeddings. So we have each $[3]^2$ grid occupying 9 coordinates on the l_1 embedding.

We claim that the embedding contracts distance by O(1).

$EMD_\Delta(A, B) \leq$

$\leq \sum EMD_k(A_i, B_i) + k.EMD_{\Delta/k}(A_{G1}, B_{G1})$

$\leq \sum EMD_k(A_i, B_i) + k. \sum EMD_{\Delta/k}(A_{G,i}, B_{G,i}) + k.EMD_{\Delta/k}(A_{G2}, B_{G2})$

$\leq ...$

$\leq EMD_3$costs of 3X3 instances

$\leq \frac{1}{2\sqrt{2}} \| f(A) - f(B) \|_1$

Here $2\sqrt{2}$ comes from distortion of 3X3.

We claim that the embedding distorts distance by $O(\log \Delta)$ in expectation.

$(3 \log_k \Delta)EMD_\Delta(A, B) \geq$

$\geq 3EMD_\Delta(A, B) + (3 \log_k \Delta/k)EMD_\Delta(A, B)$

$\geq E[\sum EMD_k(A_i, B_i) + (3 \log_k \Delta/k)k.EMD_{\Delta/k}(A_{G1}, B_{G1})]$

$\geq ..$

$\geq EMD_3$costs of 3X3 instances

$\geq \| f(A) - f(B) \|_1$

6 Final Theorem

We ran out of time here. So we just glanced through the slide.