COMS E6998-9 F15

Lecture 16: Earth-Mover Distance

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Administrivia, Plan

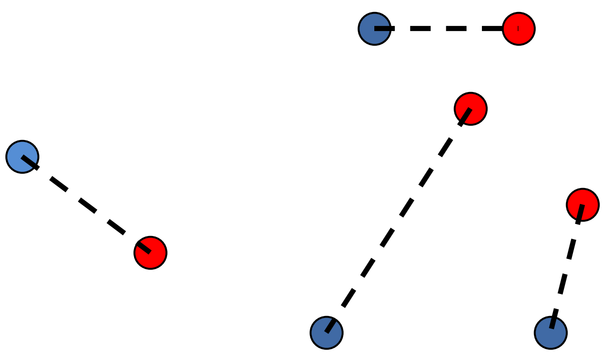
- Administrivia:
  - NO CLASS next Tuesday 11/3 (holiday)

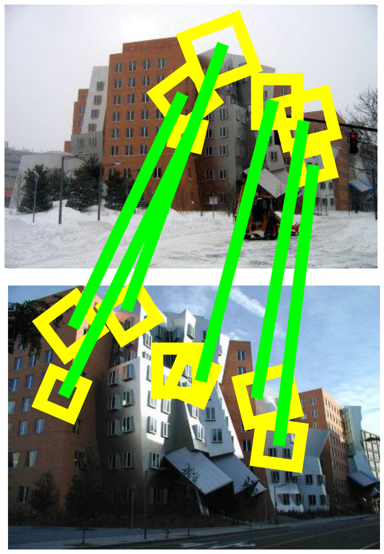
- Plan:
  - Earth-Mover Distance

- Scriber?

Earth-Mover Distance

- Definition:
  - Given two sets A, B of points in a metric space
  - $EMD(A, B)$ = min cost bipartite matching between A and B
- Which metric space?
  - Can be plane, ℓ_2, ℓ_1…
- Applications in image vision



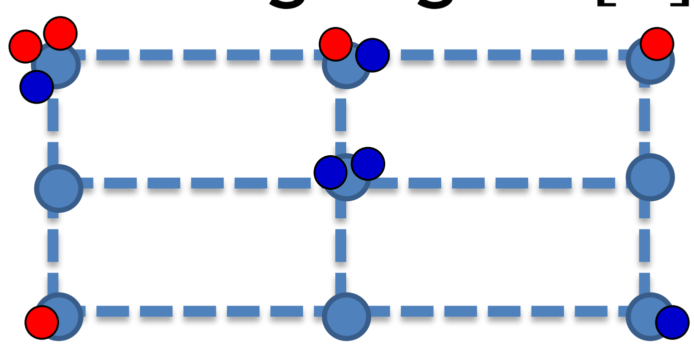


Images courtesy of Kristen Grauman

Embedding EMD into ℓ_1

- Why ℓ_1?
- At least as hard as ℓ_1
  - Can embed $\{0,1\}^d$ into EMD with distortion 1
- ℓ_1 is richer than ℓ_2

- Will focus on integer grid $[\Delta]^2$:



Embedding EMD into ℓ_1

[Charikar'02, Indyk-Thaper'03]

- **Theorem:** Can embed EMD over $[\Delta]^2$ into ℓ_1 with distortion $O(\log \Delta)$. In fact, will construct a randomized $f: 2^{[\Delta]^2} \rightarrow \ell_1$ such that:
  - for any $A, B \subset [\Delta]^2$:
    $EMD(A,B) \le \boldsymbol{E}[||f(A) - f(B)||_1] \le O(\log \Delta) \cdot EMD(A,B)$
  - time to embed a set of s points: $O(s \log \Delta)$.
- Consequences:
  - Nearest Neighbor Search: $O(c \log \Delta)$ approximation with $O(sn^{1+1/c})$ space, and $O(n^{1/c} \cdot s \log \Delta)$ query time.
  - Computation: $O(\log \Delta)$ approximation in $O(s \log \Delta)$ time
    - Best known: $1+\epsilon$ approximation in $\tilde{O}(s)$ time [AS'12]

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

- Suppose:
  - $|A| = a$
  - $|B| = b < a$
- Define

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

where

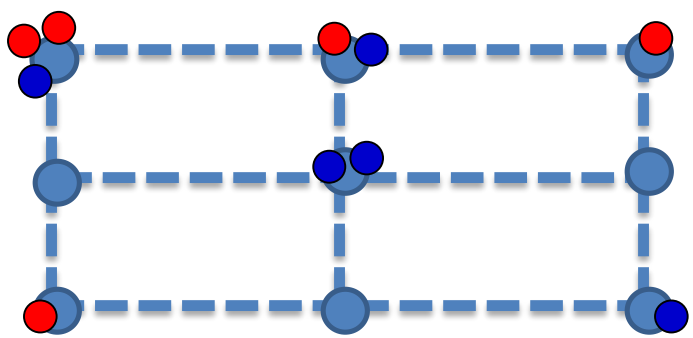
A' ranges over all subsets of A of size b

$\pi: A' \rightarrow B$ ranges over all 1-to-1 mappings

For optimal A', call $a \in A \backslash A'$ *unmatched*

Embedding EMD over small grid

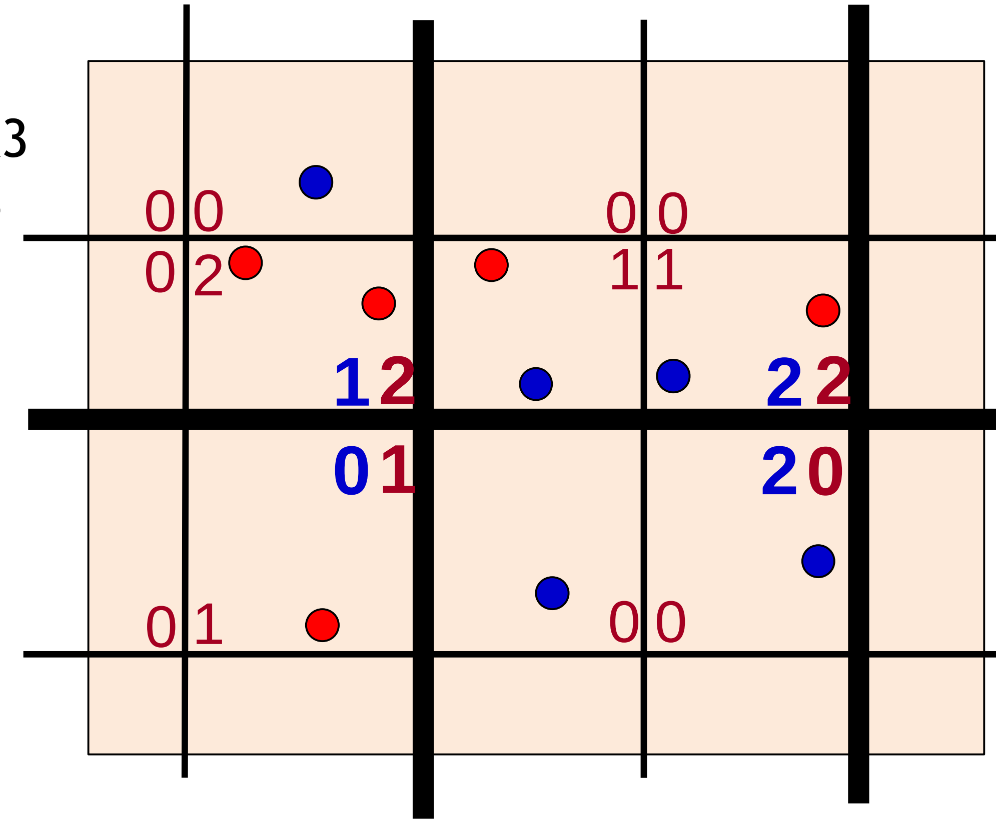
- Suppose $\Delta = 3$
- $f(A)$ has nine coordinates, counting # points in each integer point
  - $f(A) = (2,1,1,0,0,0,1,0,0)$
  - $f(B) = (1,1,0,0,2,0,0,0,1)$
- Claim: $2\sqrt{2}$ distortion embedding



High level embedding

- Set in $[\Delta]^2$ box
- Embedding of set A:
  - take a quad-tree
    - grid of cell size $\Delta/3$
    - partition each cell in 3x3
    - recurse until of size 3x3
  - randomly shift it
  - Each cell gives a coordinate: $f(A)_c$=#points in the cell c
- Want to prove

$$\boldsymbol{E}\left[\left|\left|f(A) - f(B)\right|\right|_1\right] \approx EMD(A,B)$$

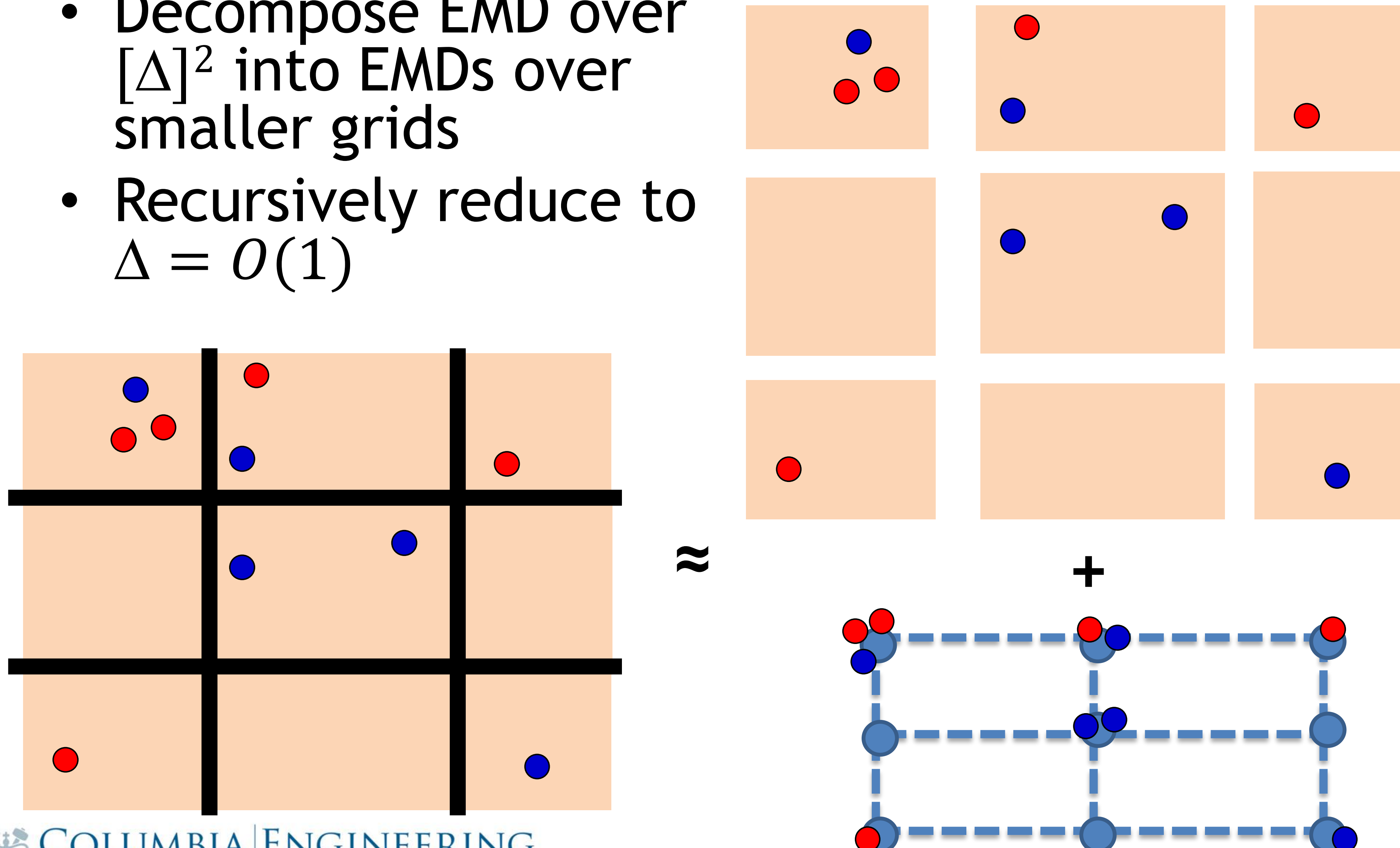


$$f(\boldsymbol{A}) = \ldots 2210 \ldots 0002 \ldots 0011 \ldots 0100 \ldots 0000 \ldots$$

$$f(\boldsymbol{B}) = \ldots 1202 \ldots 0100 \ldots 0011 \ldots 0000 \ldots 1100 \ldots$$

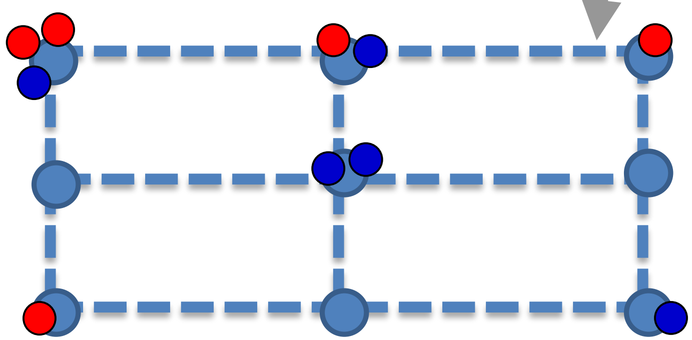
Main idea: intuition

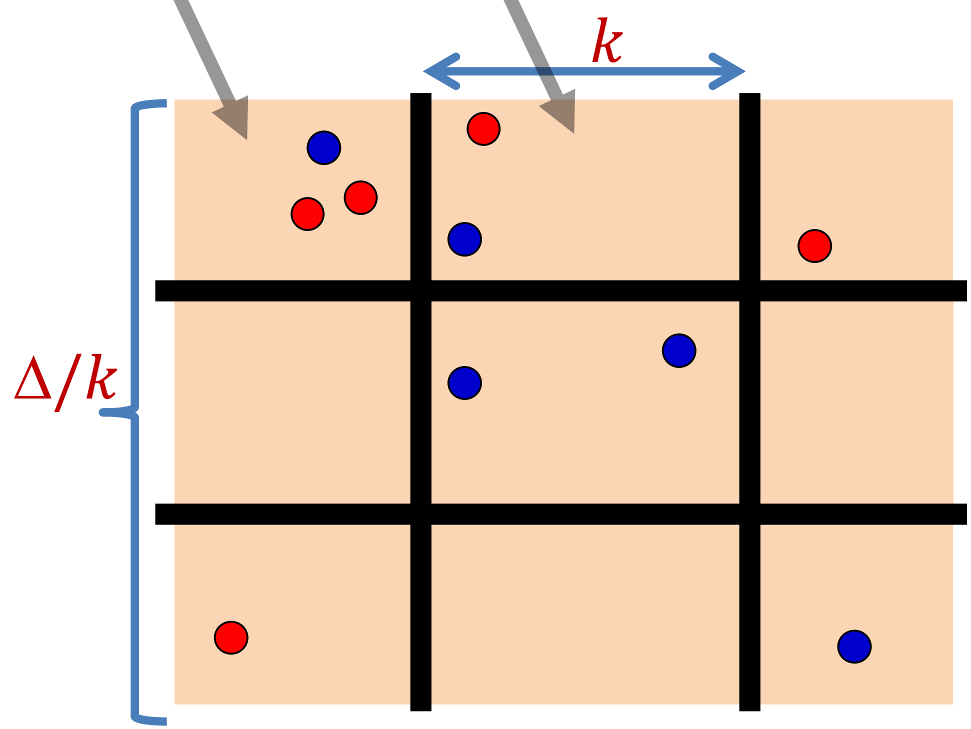
- Decompose EMD over $[\Delta]^2$ into EMDs over smaller grids
- Recursively reduce to $\Delta = O(1)$



Decomposition Lemma

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

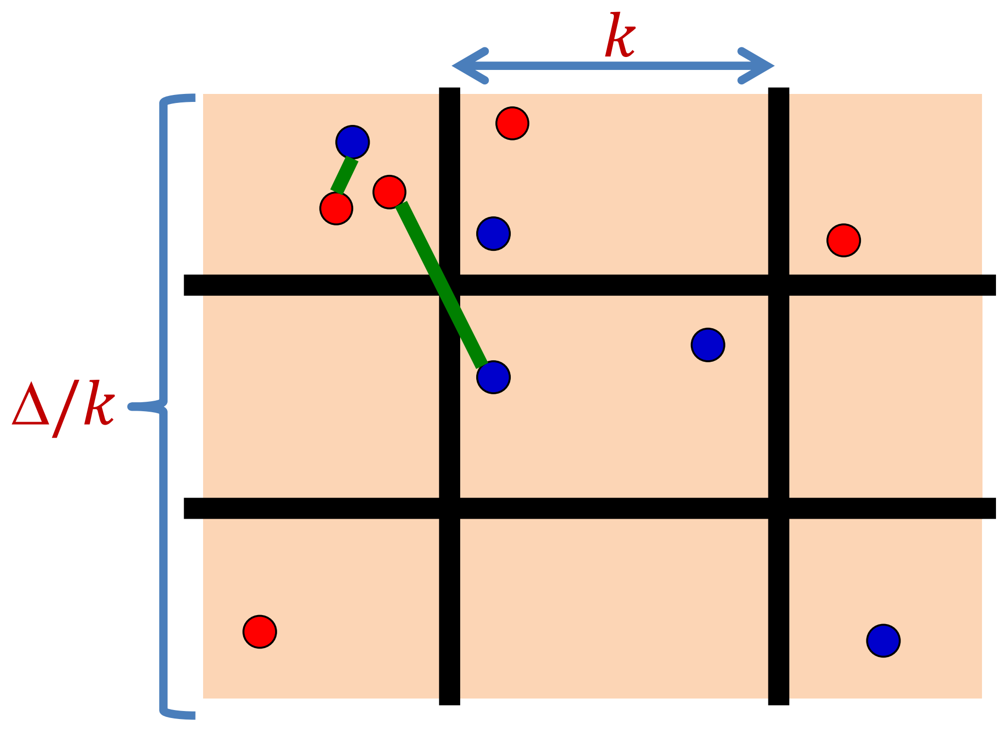




1 (lower bound)

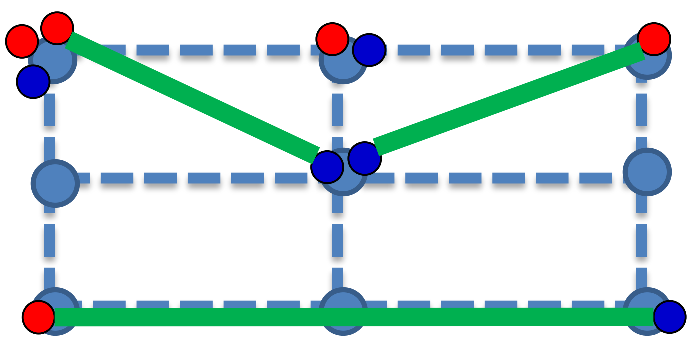
- Claim 1: for a randomly-shifted cut-grid G of side length k:
$$EMD_\Delta(A,B) \le EMD_k(A_1,B_1) + EMD_k(A_2,B_2) + \cdots + k \cdot EMD_{\Delta/k}(A_G,B_G)$$
- Construct a matching π for $EMD_\Delta(A,B)$ from the matchings on RHS as follows
- For each $a \in A$ (suppose $a \in A_i$) it is either:
  1) matched in $EMD(A_i,B_i)$ to some $b \in B_i$ (if $a \in A_i'$)
     - then $\pi(a) = b$
  2) or $a \notin A_i'$, and then it is matched in $EMD(A_G,B_G)$ to some $b \in B_j$ $(j \neq i)$
     - then $\pi(a) = b$
- Cost?
  1) paid by $EMD(A_i,B_i)$
  2) Move a to center (Δ)
     - Charge to $EMD(A_i,B_i)$

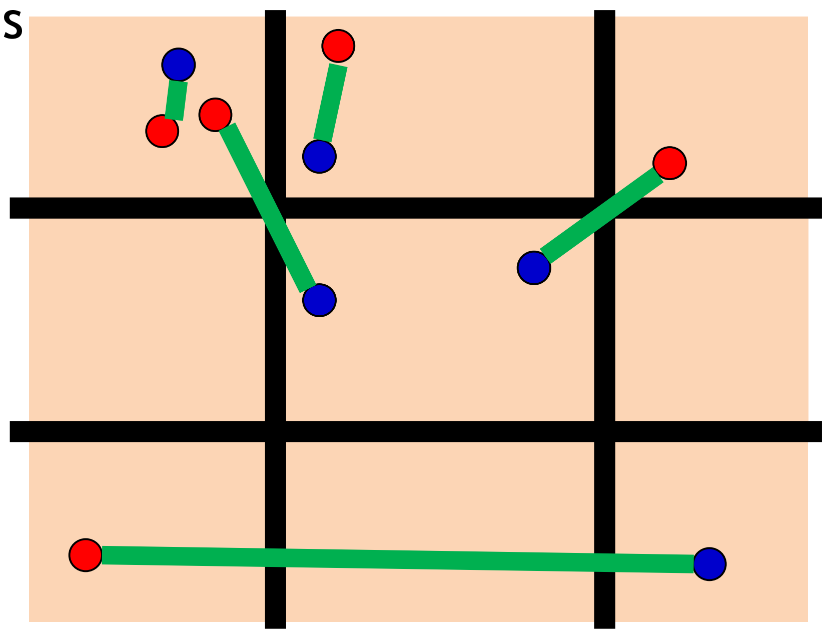
     Move from cell i to cell j
     - Charge k to $EMD(A_G,B_G)$
- If $|A| > |B|$, extra $|A| - |B|$ pay $k \cdot \frac{\Delta}{k} = \Delta$ on LHS & RHS



2 & 3 (upper bound)

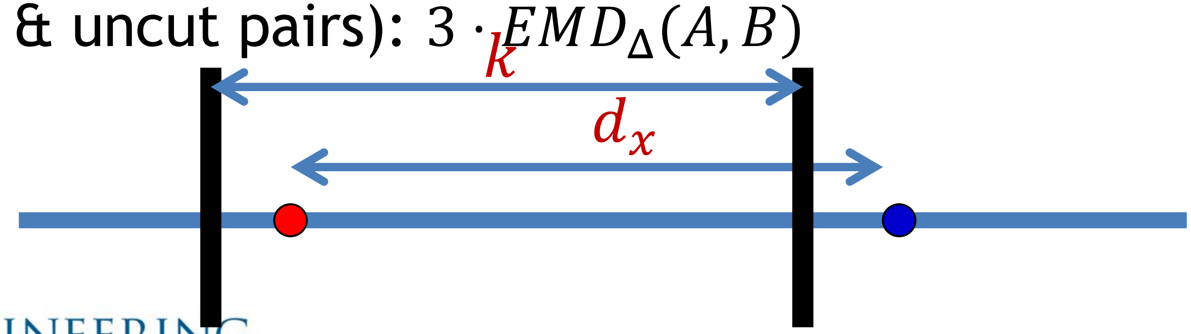
- Claims 2,3: for a randomly-shifted cut-grid G of side length k, we have:
  2) $EMD_{\Delta}(A,B) \geq \frac{1}{3}\boldsymbol{E}[EMD_k(A_1,B_1) + EMD_k(A_2,B_2) + \cdots]$
  3) $EMD_{\Delta}(A,B) \geq \boldsymbol{E}[k \cdot EMD_{\Delta/k}(A_G,B_G)]$
- Fix a matching π minimizing $EMD_{\Delta}(A,B)$
  – Will construct matchings for each EMD on RHS
- *Uncut* pairs $(a,b) \in \pi$ are matched in respective (A,B)
- *Cut* pairs $(a,b) \in \pi$:
  – are unmatched in their mini-grids
  – are matched in (A_G,B_G)





3: Cost

- Claim 2:
  - $3 \cdot EMD_\Delta(A,B) \geq \boldsymbol{E}[EMD_k(A_1,B_1) + EMD_k(A_2,B_2) + \cdots]$
- Uncut pairs (a,b) are matched in respective (A_i,B_i)
  - Total contribution from uncut pairs $\leq EMD_\Delta(A,B)$
- Consider a cut pair (a,b) at distance $a-b=(d_x,d_y)$
  - (a,b) can contribute to RHS as they may be *unmatched* in their own mini-grids
  - $\Pr[(a,b) \text{ cut}] = 1 - \left(1-\frac{d_x}{k}\right)_+ \left(1-\frac{d_y}{k}\right)_+ \leq \frac{d_x}{k} + \frac{d_y}{k} \leq \frac{1}{k}||a-b||_2$
  - Expected contribution of (a,b) to RHS:
    $\leq \Pr[(a,b) \text{ cut}] \cdot 2k \leq 2||a-b||_2$
  - Total expected cost contributed to RHS:
    $2 \cdot EMD_\Delta(A,B)$
- Total (cut & uncut pairs): $3 \cdot EMD_\Delta(A,B)$

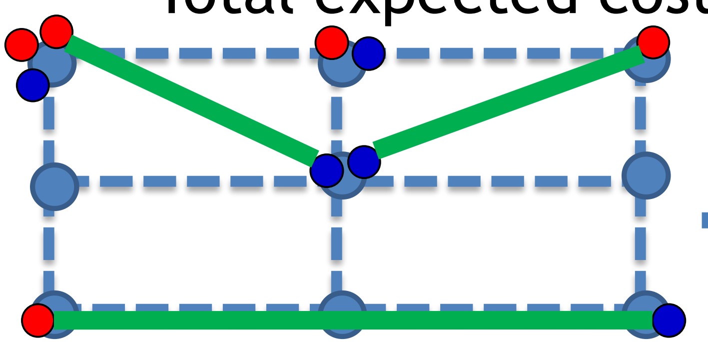


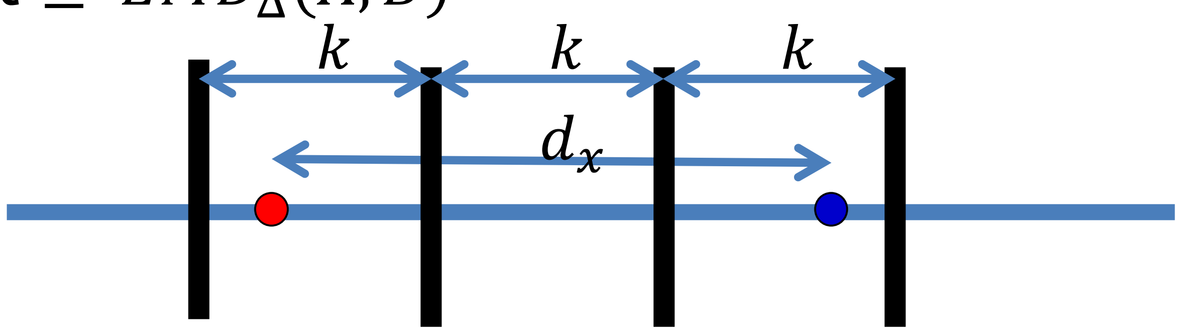
3: Cost

- Claim:
  - $EMD_\Delta(A,B) \geq \boldsymbol{E}[k \cdot EMD_{\Delta/k}(A_G, B_G)]$
- Uncut pairs: contribute zero to RHS!
- Cut pair: $(a,b) \in \pi$ with $a - b = (d_x, d_y)$
  - if $|d_x| = xk + r_k$, and $|d_y| = yk + r_y$, then
  - expected cost contribution to $k \cdot EMD_{\Delta/k}(A_G, B_G)$:

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

- Total expected cost $\leq EMD_\Delta(A,B)$





Recurse on decomposition

- For randomly-shifted cut-grid G of side length k, we have:
  1) $EMD_{\Delta}(A,B) \leq EMD_k(A_1,B_1) + EMD_k(A_2,B_2) + \cdots + k \cdot EMD_{\Delta/k}(AG,BG)$
  2) $EMD_{\Delta}(A,B) \geq \frac{1}{3}\boldsymbol{E}[EMD_k(A_1,B_1) + EMD_k(A_2,B_2) + \cdots]$
  3) $EMD_{\Delta}(A,B) \geq \boldsymbol{E}[k \cdot EMD_{\Delta/k}(A_G,B_G)]$
- We applying 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_2 on $[\Delta/3]^2$
  - Obtain more grids $[3]^2$, and another big grid $[\Delta/9]^2$
  - Then choose randomly-shifted cut-grid G_3 on $[\Delta/9]^2$
  - …
- Then, embed each of the small grids $[3]^2$ into ℓ_1, using $O(1)$ distortion embedding, and concatenate the embeddings
  - Each $[3]^2$ grid occupies 9 coordinates on ℓ_1 embedding

Proving recursion works

- Claim: embedding contracts distances by $O(1)$:
  $EMD_\Delta(A,B) \leq$
  $\leq \sum_i EMD_{k(A_i, Bi)} + k \cdot EMD_{\Delta/k}(A_{G_1}, B_{G_1})$
  $\leq \sum_i EMD_k(A_i, Bi) + k\sum_i EMD_{k(A_{G_1,i}, B_{G_1,i})}$
  $+k \cdot EMD_{\frac{\Delta}{k^2}}(A_{G_2}, B_{G_2})$
  $\leq$ …
  $\leq$ sum of EMD_3 costs of 3×3 instances
  $\leq \frac{1}{2\sqrt{2}} ||f(A) - f(B)||_1$
- Claim: embedding distorts distances by $O(\log \Delta)$ in expectation:
  $(3 \log_k \Delta) \cdot EMD_\Delta(A,B)$
  $\geq 3 \cdot EMD_\Delta(A,B) + \left(3 \log_k \frac{\Delta}{k}\right) \cdot EMD_\Delta(A,B)$
  $\geq \mathbf{E}[\sum_i EMD_k(Ai, Bi) + \left(3 \log_k \frac{\Delta}{k}\right) \cdot k \cdot EMD_{\Delta/k}(A_{G_1}, B_{G_1})]$
  $\geq \cdots$
  $\geq$ sum of EMD_3 costs of 3×3 instances
  $\geq ||f(A) - f(B)||_1$

Final theorem

- **Theorem:** can embed EMD over $[\Delta]^2$ into ℓ_1 with $O(\log \Delta)$ distortion in expectation.
- Notes:
  - Dimension required: $O(\Delta^2)$, but a set A of size s maps to a vector that has only $O(s \cdot \log \Delta)$ non-zero coordinates.
  - Time: can compute in $O(s \cdot \log \Delta)$
  - By Markov's, it's $O(\log \Delta)$ distortion with 90% probability
- Applications:
  - Can compute $EMD(A, B)$ in time $O(s \cdot \log \Delta)$
  - NNS: $O(c \cdot \log \Delta)$ approximation, with $O(n^{1+1/c} \cdot s)$ space, and $O(n^{1/c} \cdot s \cdot \log \Delta)$ query time.