Clustering, Hamming Embedding, Generalized LSH and the Max Norm

Behnam Neyshabur Yury Makarychev Nathan Srebro

Toyota Technological Institute at Chicago

Clustering

S={  }

sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

Clustering

S={  }

sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

Clustering

S={

}

$$\max_{K\in\{\pm1\}^{N\times N}} \sum_{x,y} \langle sim(x,y), K(x,y) \rangle$$

$$\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

s.t. K is a valid cluster matrix

sim =

$$\begin{bmatrix} 1 & 0.9 & -0.2 & -0.2 & -0.2 & -0.1 & 0 \\ 0.9 & 1 & -0.1 & -0.1 & -0.2 & -0.1 & 0 \\ -0.2 & -0.1 & 1 & 0.9 & 0.5 & 0.6 & 0.3 \\ -0.2 & -0.1 & 0.9 & 1 & 0.5 & 0.3 & 0.4 \\ -0.2 & -0.2 & 0.5 & 0.5 & 1 & 0.7 & 0.7 \\ -0.1 & -0.1 & 0.6 & 0.4 & 0.7 & 1 & 0.9 \\ 0 & 0 & 0.3 & 0.4 & 0.7 & 0.9 & 1 \end{bmatrix}$$

Clustering

S={  }

$$\max_{K\in\{\pm1\}^{N\times N}} \sum_{x,y} \langle sim(x,y), K(x,y)\rangle$$

$$\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

s.t. K is a valid cluster matrix

non-convex constraint

sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

Clustering

S={ , , , , , , }

$$\max_{K\in\{\pm1\}^{N\times N}} \sum_{x,y} \langle sim(x,y), K(x,y) \rangle$$

$$\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

s.t. K is a valid cluster matrix

non-convex constraint

convex relaxations

- Trace-norm [Jalali et al. ICML 2011] and max-norm [Jalali et al. ICML 2012] relaxations

***sim* =**

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

Clustering

S={

}

$$\max_{K\in\{\pm1\}^{N\times N}} \sum_{x,y} \langle sim(x,y), K(x,y)\rangle$$

$$\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

s.t. K is a valid cluster matrix

non-convex constraint

convex relaxations

- Trace-norm [Jalali et al. ICML 2011] and max-norm [Jalali et al. ICML 2012] relaxations

$$sim = \begin{array}{ccccccc} 1 & 0.9 & -0.2 & -0.2 & -0.2 & -0.1 & 0 \\ 0.9 & 1 & -0.1 & -0.1 & -0.2 & -0.1 & 0 \\ -0.2 & -0.1 & 1 & 0.9 & 0.5 & 0.6 & 0.3 \\ -0.2 & -0.1 & 0.9 & 1 & 0.5 & 0.3 & 0.4 \\ -0.2 & -0.2 & 0.5 & 0.5 & 1 & 0.7 & 0.7 \\ -0.1 & -0.1 & 0.6 & 0.4 & 0.7 & 1 & 0.9 \\ 0 & 0 & 0.3 & 0.4 & 0.7 & 0.9 & 1 \end{array}$$

- How tight are these SDP relaxations?
- What is the best tractable relaxation?

Binary Hashing

1	1	0	0	1	1	1
1	1	0	1	1	0	0
1	1	0	0	0	0	0
1	1	1	1	0	0	0

$$sim(x,y) \approx 1-\frac{2}{d}\delta_{Ham}(b(x),b(y))$$

sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

Binary Hashing

1	1	0	0	1	1	1
1	1	0	1	1	0	0
1	1	0	0	0	0	0
1	1	1	1	0	0	0

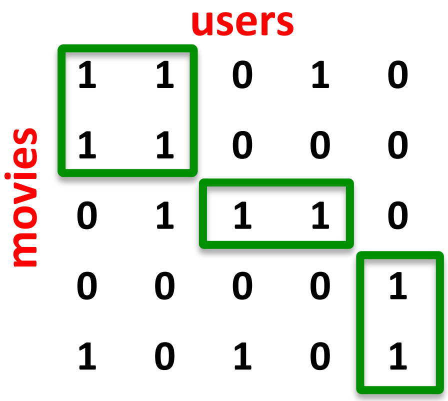
$$sim(x,y) \approx 1 - \frac{2}{d}\delta_{Ham}(b(x),b(y))$$

sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

- When is there a good binary hashing?
- What is the relationship between clustering and binary hashing?

Asymmetry



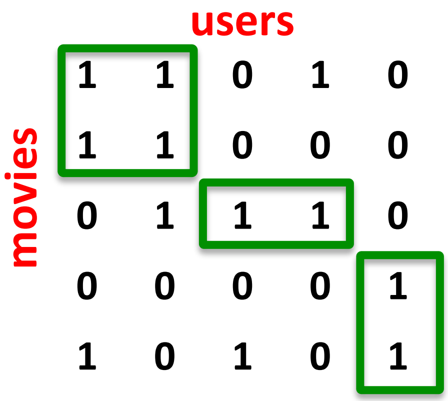
1	1	0	1	0
1	1	0	0	0
0	1	1	1	0
0	0	0	0	1
1	0	1	0	1

- Biclustering [Dhillon et al. SIGKDD 2003]
  - E.g. Netflix dataset (user-movie)
- Asymmetric Binary Hashing [Neyshabur et al. NIPS 2013]
  - Even if similarity matrix is symmetric the asymmetric hashes are more powerful

$$1 - \frac{2}{d}\delta_{\mathrm{Ham}}(b(x), \tilde{b}(y)) \approx \mathrm{sim}(x, y)$$

Asymmetry

- Biclustering [Dhillon et al. SIGKDD 2003]
  - E.g. Netflix dataset (user-movie)



1	1	0	1	0
1	1	0	0	0
0	1	1	1	0
0	0	0	0	1
1	0	1	0	1

- Asymmetric Binary Hashing [Neyshabur et al. NIPS 2013]
  - Even if similarity matrix is symmetric the asymmetric hashes are more powerful

$$1 - \frac{2}{d}\delta_{\mathrm{Ham}}(b(x), \tilde{b}(y)) \approx \mathrm{sim}(x, y)$$

- Can asymmetry help in clustering and LSH?
- Can we find asymmetric LSH for a given symmetric similarity function when there is no LSH?

Outline

- Clustering
- Binary Hashing
- LSH and ALSH
- Tight bounds on clustering, embedding and LSH

Clustering

- Set of objects: S
- Similarity function: sim:S × S → [-1,1]

S={  , … }

…

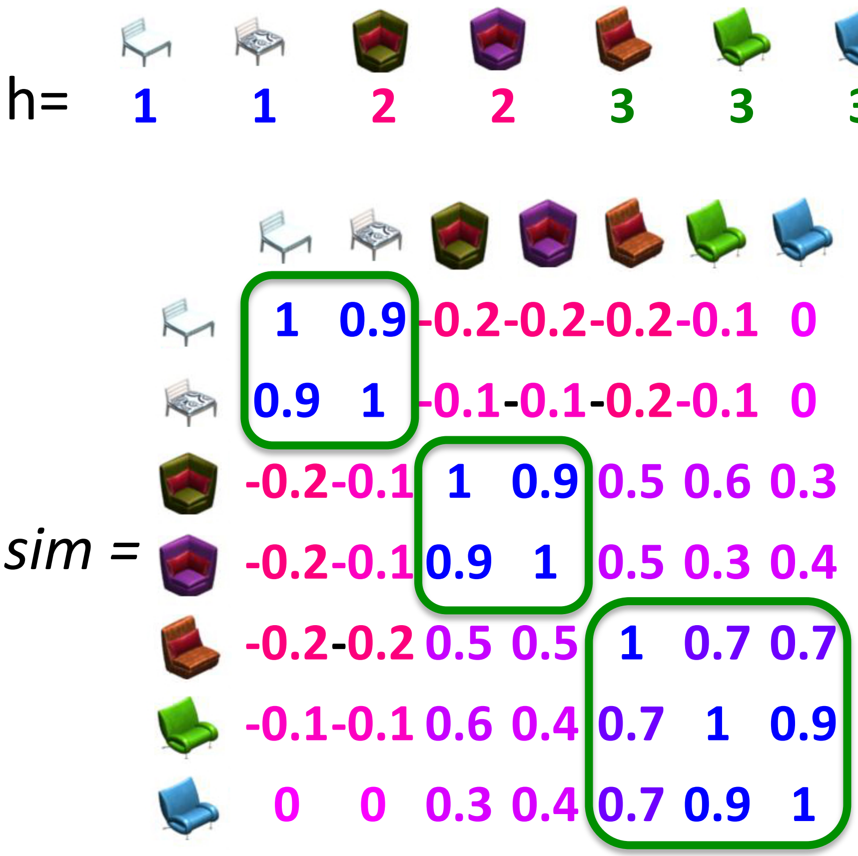
sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1

⋮

Clustering

- Mapping: h:S→Γ



h= 1 1 2 2 3 3 3

$$sim = \begin{pmatrix} 1 & 0.9 & -0.2 & -0.2 & -0.2 & -0.1 & 0 \\ 0.9 & 1 & -0.1 & -0.1 & -0.2 & -0.1 & 0 \\ -0.2 & -0.1 & 1 & 0.9 & 0.5 & 0.6 & 0.3 \\ -0.2 & -0.1 & 0.9 & 1 & 0.5 & 0.3 & 0.4 \\ -0.2 & -0.2 & 0.5 & 0.5 & 1 & 0.7 & 0.7 \\ -0.1 & -0.1 & 0.6 & 0.4 & 0.7 & 1 & 0.9 \\ 0 & 0 & 0.3 & 0.4 & 0.7 & 0.9 & 1 \end{pmatrix}$$

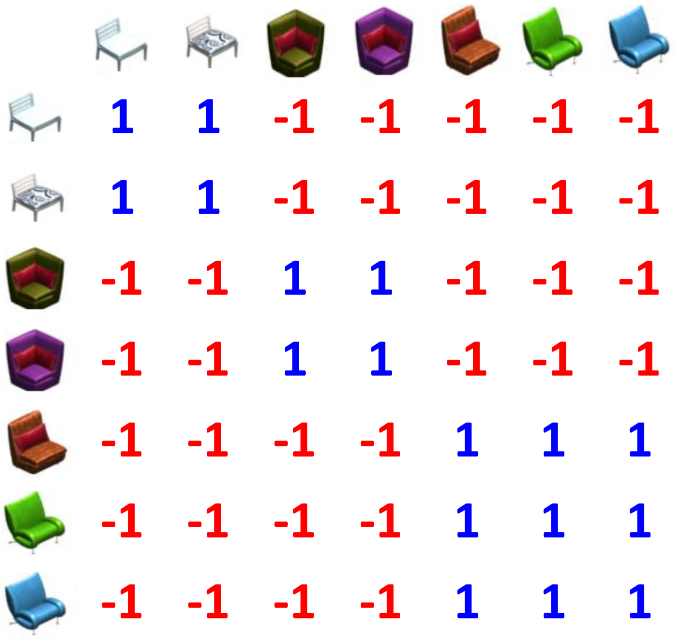
Clustering

- Mapping: $h: S \rightarrow \Gamma$
- Clustering function: $K_h(x,y)= \begin{cases} 1 & \text{if } h(x)=h(y) \\ -1 & \text{otherwise} \end{cases}$

h=	1	1	2	2	3	3	3

sim =

1	0.9	-0.2	-0.2	-0.2	-0.1	0
0.9	1	-0.1	-0.1	-0.2	-0.1	0
-0.2	-0.1	1	0.9	0.5	0.6	0.3
-0.2	-0.1	0.9	1	0.5	0.3	0.4
-0.2	-0.2	0.5	0.5	1	0.7	0.7
-0.1	-0.1	0.6	0.4	0.7	1	0.9
0	0	0.3	0.4	0.7	0.9	1



K_h =

1	1	-1	-1	-1	-1	-1
1	1	-1	-1	-1	-1	-1
-1	-1	1	1	-1	-1	-1
-1	-1	1	1	-1	-1	-1
-1	-1	-1	-1	1	1	1
-1	-1	-1	-1	1	1	1
-1	-1	-1	-1	1	1	1

Biclustering

- Two sets of objects S and T
- Similarity function sim:$S \times T \rightarrow [-1,1]$
- Mappings $f:S \rightarrow \Gamma$ and $g:T \rightarrow \Gamma$
- Clustering function $K_{f,g}(x,y) = \begin{cases} 1 & \text{if } f(x)=g(y) \\ -1 & \text{otherwise} \end{cases}$

$$K = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

Max-norm Relaxation

$$\min_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. K is a valid cluster matrix

$$\min_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $\|K\|_{\max} \leq b$

Max-norm:

$$\begin{aligned} \|Z\|_{\max} = \min_{UV^\top = Z} \quad & t \\ \text{s.t} \quad & \|U_i\|_2^2 \leq t \\ & \|V_i\|_2^2 \leq t \end{aligned}$$

Max-norm Relaxation

$$\min_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. K is a valid cluster matrix

$$\min_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $\|K\|_{\max} \leq b$

SDP representable

$$\min_{K,A,B} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $[\blacksquare A\&K@K\uparrow T\ \&B\ ]\succcurlyeq 0,$
$diag(A) \leq b$
$diag(B) \leq b$

Max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \ t \quad \text{s.t} \quad \|U_i\|_2^2 \leq t, \quad \|V_i\|_2^2 \leq t$$

Max-norm Relaxation

$$\min_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. K is a valid cluster matrix

$$\min_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $\|K\|_{\max} \leq b$

Max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} t \quad \text{s.t} \quad \|U_i\|_2^2 \leq t, \quad \|V_i\|_2^2 \leq t$$

$$K = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

$$U = \begin{array}{ccc} \text{Cluster 1} & \text{Cluster 2} & \text{Cluster 3} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array} \qquad V = \begin{array}{ccc} \text{Cluster 1} & \text{Cluster 2} & \text{Cluster 3} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}$$

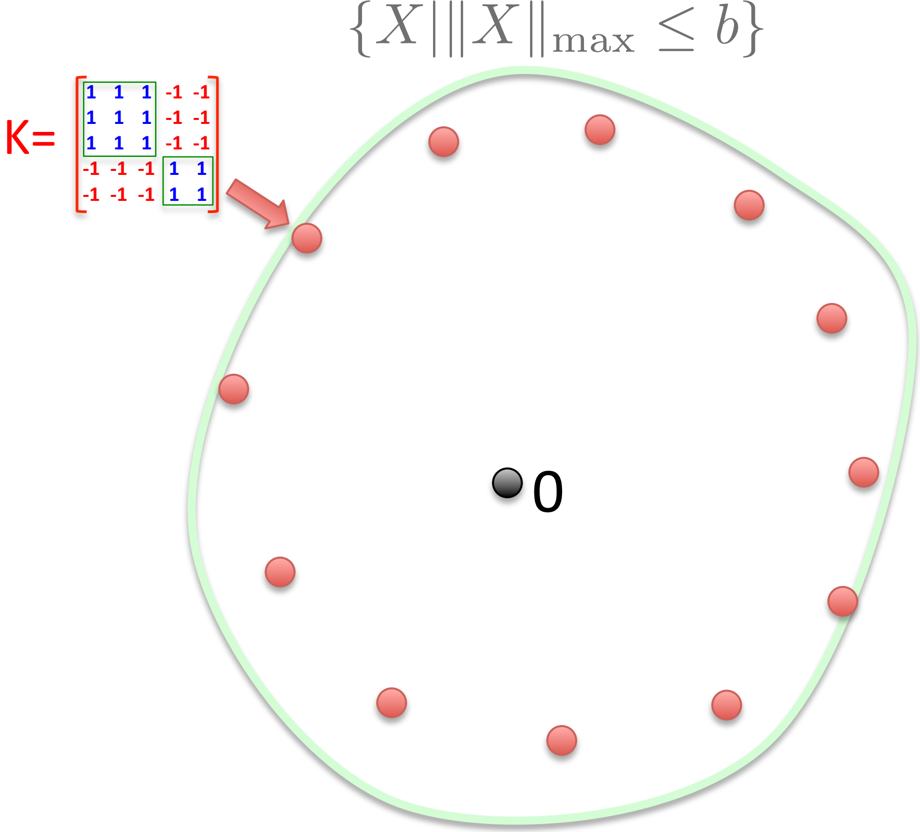
$$\begin{aligned} \|K\|_{\max} &= \|2UV^\top - 1\|_{\max} \\ &\leq 2\|UV^\top\|_{\max} + 1 \\ &= 3 \end{aligned}$$

Tightness of max-norm relaxation

M_k={cluster matrices with k partition}

$$\max_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $K \in M_k$



$$K=\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

$$\max_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $\|K\|_{\max} \leq b$

Tightness of max-norm relaxation

M_k={cluster matrices with k partition}

$$\max_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

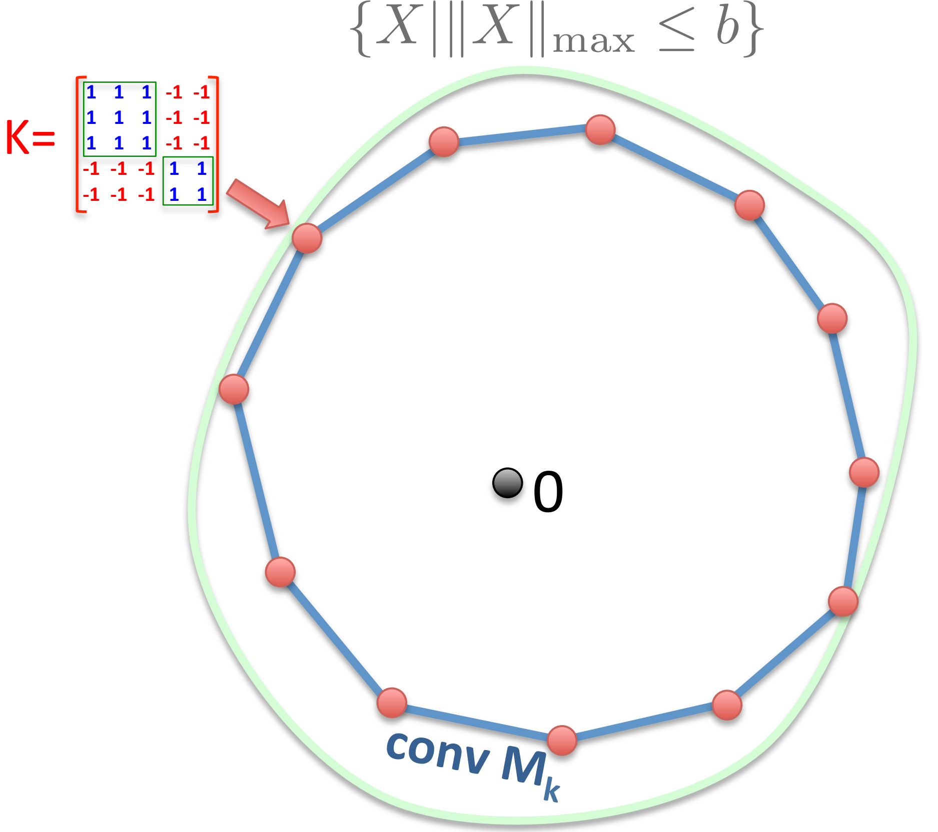
s.t. $K \in M_k$

$$\max_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

s.t. $K \in convM_k$

$$\max_{K} \sum_{x,y} Err(sim(x,y), K(x,y))$$

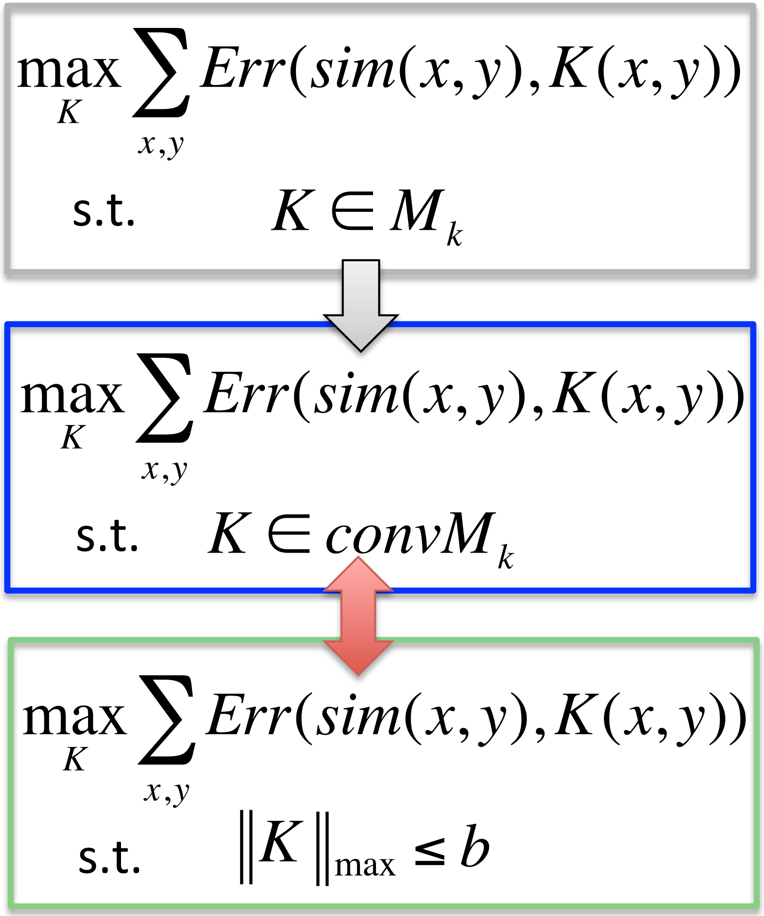
s.t. $\|K\|_{\max} \leq b$



$$K=\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

Tightness of max-norm relaxation

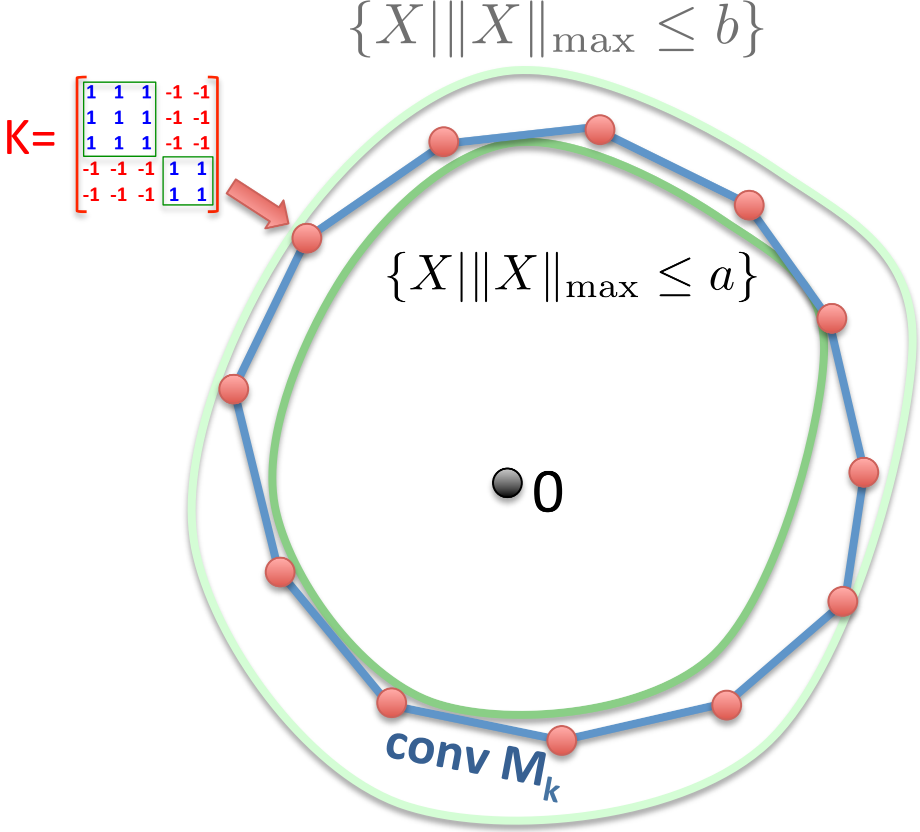
M_k={cluster matrices with k partition}



$$\max_K \sum_{x,y} Err(sim(x,y),K(x,y))$$
s.t. $K \in M_k$

$$\max_K \sum_{x,y} Err(sim(x,y),K(x,y))$$
s.t. $K \in convM_k$

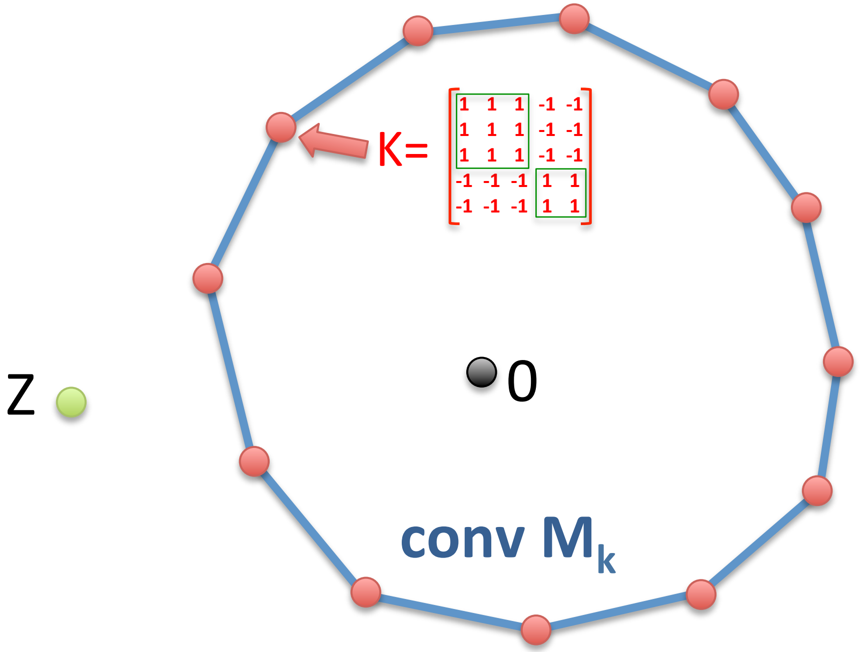
$$\max_K \sum_{x,y} Err(sim(x,y),K(x,y))$$
s.t. $\|K\|_{\max} \leq b$



How tight is the relaxation?
(Upper bound for b/a)

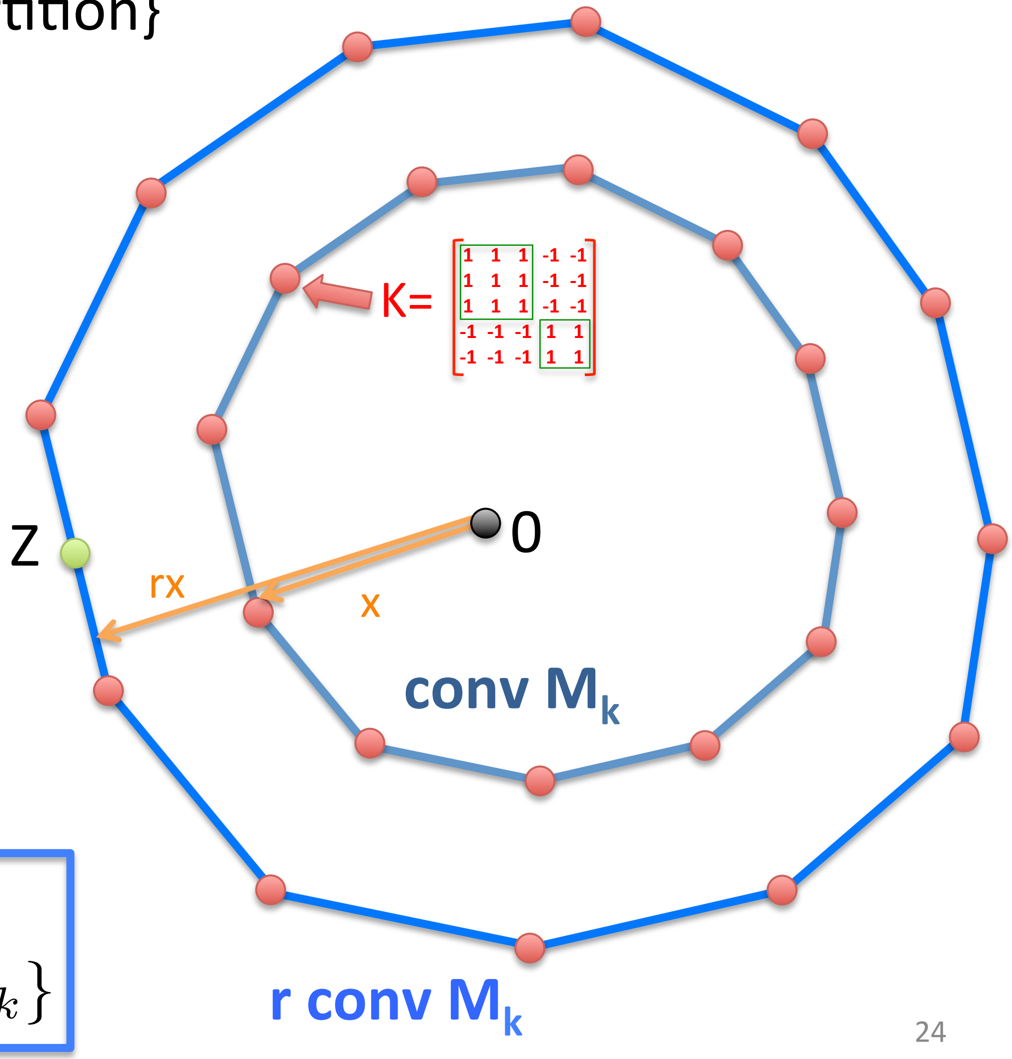
The Cluster Ratio

M_k={cluster matrices with k partition}



The Cluster Ratio

M_k={cluster matrices with k partition}

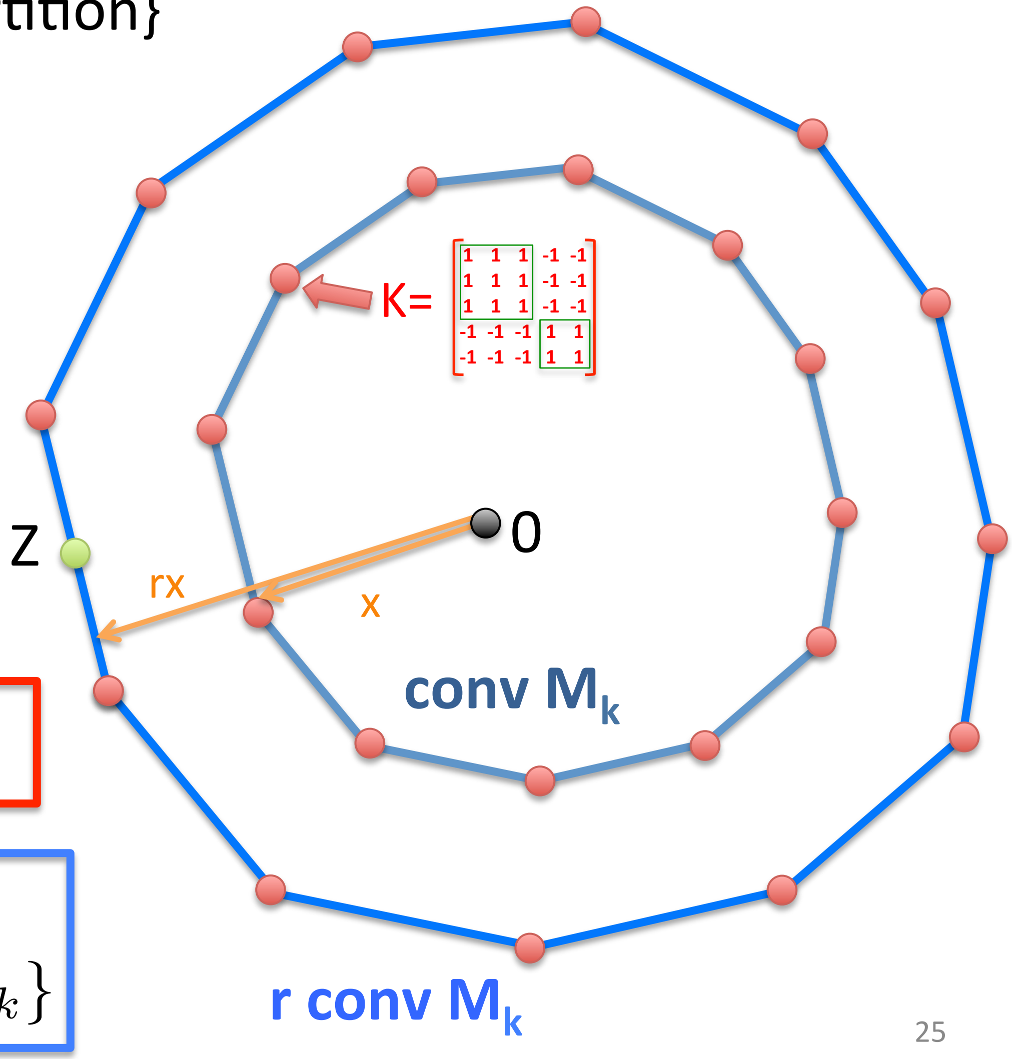


Cluster ratio function:

$$\rho_k(Z) = \min\{r | Z \in r\mathrm{conv} M_k\}$$

The Cluster Ratio

M_k={cluster matrices with k partition}



$$K=\begin{bmatrix}1&1&1&-1&-1\\1&1&1&-1&-1\\1&1&1&-1&-1\\-1&-1&-1&1&1\\-1&-1&-1&1&1\end{bmatrix}$$

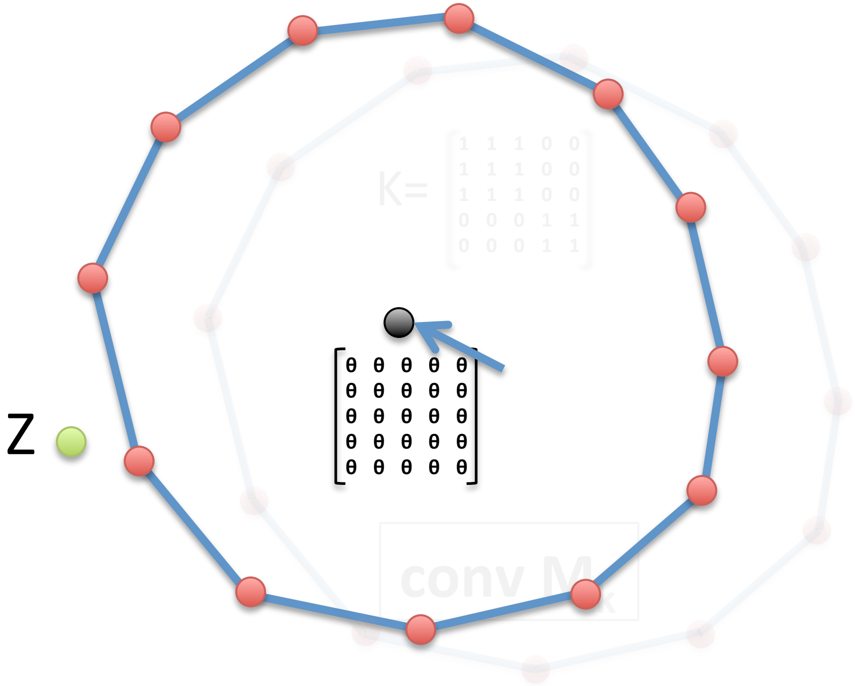
a||Z||↓max ≤ρ↓k (Z)≤b||Z||↓max

Cluster ratio function:

$$\rho_k(Z) = \min\{r|Z \in r\mathrm{conv}M_k\}$$

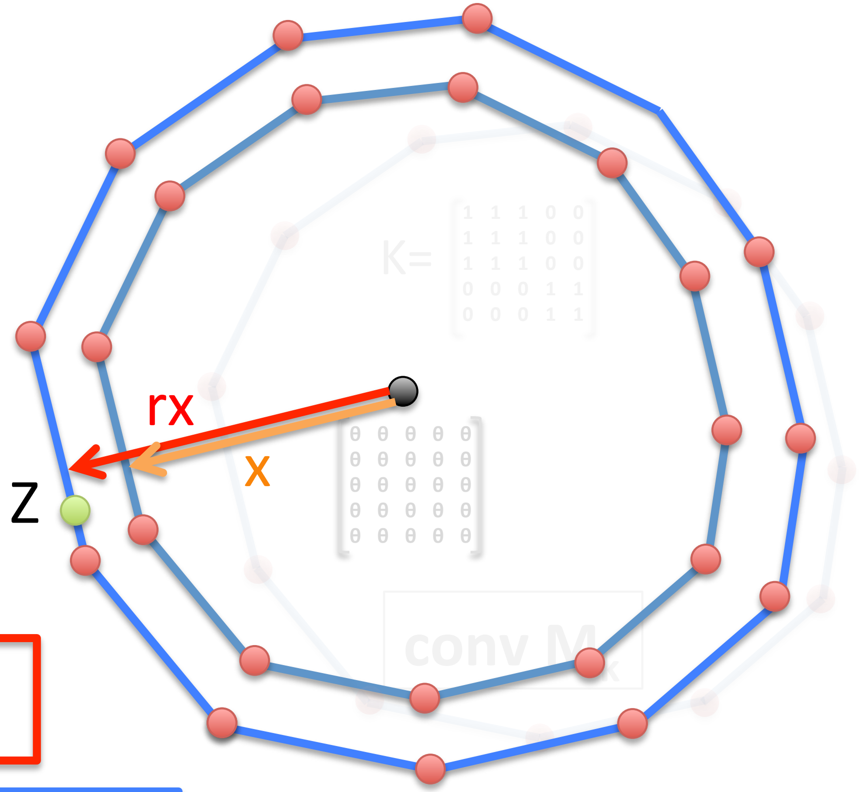
The Centralized Cluster Ratio

M_k={cluster matrices with k partition}



The Centralized Cluster Ratio

M_k={cluster matrices with k partition}



$a\|Z\|\downarrow\max \quad \le\rho \downarrow k\ (Z)\le b\|Z\|\downarrow\max$

Centralized cluster ratio function:

$$\rho_k(Z) = \min\{r|Z \in r\mathrm{conv}M_k + \theta\}$$

Binary Embedding

- Mapping to a binary string: $b : S \rightarrow \{\pm 1\}^d$
- Similarity is approximated by hamming distance:

$$sim(x,y) \approx 1 - \frac{2\delta_{Ham}(b(x),b(y))}{d}$$

b=

-1	**-1**	**1**	**1**	**1**	**1**	**1**
-1	**-1**	**-1**	**1**	**1**	**-1**	**-1**
-1	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**
-1	**-1**	**-1**	**-1**	**-1**	**1**	**1**

Binary Embedding

- Mapping to a binary string: $b : S \rightarrow \{\pm 1\}^d$
- Similarity is approximated by hamming distance:

$$sim(x,y) \approx 1 - \frac{2\delta_{Ham}(b(x),b(y))}{d} \boxed{= \frac{1}{d}\sum_{i=1}^{d} K_{b_i}(x,y)}$$

b=

	-1	**-1**	**1**	**1**	**1**	**1**	**1**
b_i=	**-1**	**-1**	**-1**	**1**	**1**	**-1**	**-1**
	-1	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**
	-1	**-1**	**-1**	**-1**	**-1**	**1**	**1**

Binary Embedding

- Mapping to a binary string: $b: S \to \{\pm 1\}^d$
- Similarity is approximated by hamming distance:

$$sim(x,y) \approx 1 - \frac{2\delta_{Ham}(b(x),b(y))}{d} \boxed{= \frac{1}{d}\sum_{i=1}^{d} K_{b_i}(x,y)}$$

b=

	-1	-1	1	1	1	1	1
b_i=	-1	-1	-1	1	1	-1	-1
	-1	-1	-1	-1	-1	-1	-1
	-1	-1	-1	-1	-1	1	1

Binary Embedding is the average of several clusterings.

Asymmetric Binary Embedding

- Mapping to a binary string: $b : S \rightarrow \{\pm 1\}^d,\ \tilde{b} : T \rightarrow \{\pm 1\}^d$
- Similarity is approximated by hamming distance:

$$sim(x,y) \approx 1 - \frac{2\delta_{Ham}(b(x),\tilde{b}(y))}{d} \boxed{= \frac{1}{d}\sum_{i=1}^{d} K_{b_i,\tilde{b}_i}(x,y)}$$

Asymmetric Binary Embedding is the average of several biclusterings.

Locality Sensitive Hashing (LSH)

- We talked about embedding as an average of clusterings:

$$\frac{1}{d}\sum_{i=1}^{d} K_{h_i}(x,y) \approx sim(x,y)$$

- LSH [Charikar, STOC 2002]: given function sim:S→[0,1], LSH is the probability distribution on family H of hash functions that:

$$P_{h\in H}[h(x)=h(y)] = sim(x,y)$$

- Equivalent definition when sim:S→[-1,1]: LSH is the probability distribution on the family H of clustering functions such that:

$$\mathrm{E}_{h\in H}[K_h(x,y)] = sim(x,y)$$

Locality Sensitive Hashing (LSH)

- We talked about embedding as an average of clusterings:

$$\frac{1}{d}\sum_{i=1}^{d} K_{h_i}(x,y) \approx sim(x,y)$$

- LSH [Charikar, STOC 2002]: given function sim:S→[0,1], LSH is the probability distribution on family H of hash functions that:

$$P_{h\in H}[h(x)=h(y)] = sim(x,y)$$

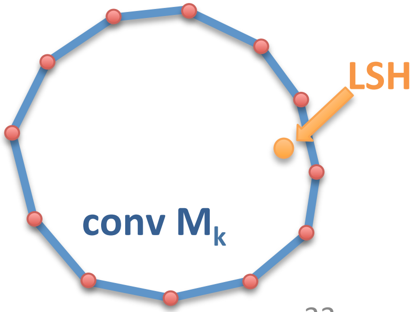
- Equivalent definition when sim:S→[-1,1]: LSH is the probability distribution on the family H of clustering functions such that:

$$\mathrm{E}_{h\in H}[K_h(x,y)] = sim(x,y)$$

LSH is a distribution over clusterings.

↕

LSH is a point on the convex hull of clusterings.



α-LSH

- α-LSH: relaxing LSH and tolerating affine relationship:

$$\alpha \mathrm{E}_{h\in H}[K_h(x,y)] - \theta = sim(x,y)$$

- The length d of LSH required to get accurate approximation scales quadratically with α:

$$Var\left[\frac{\alpha}{d}\sum K_h(x,y) - \theta\right] \leq \frac{\alpha^2}{d}$$

So we want an α-LSH with minimum possible α.

α-LSH

- α-LSH: relaxing LSH and tolerating affine relationship:

$$\alpha \mathrm{E}_{h \in H}[K_h(x,y)] - \theta = sim(x,y)$$

- The length d of LSH required to get accurate approximation scales quadratically with α:

$$Var\left[\frac{\alpha}{d}\sum K_h(x,y) - \theta\right] \le \frac{\alpha^2}{d}$$

So we want an α-LSH with minimum possible α.

- Having **α-LSH** is equivalent to being **embeddable** to hamming space with no distortion.
- There is no α-LSH for **Euclidean inner product** and some other usefull similarity measures.

How to overcome this?

Asymmetric LSH (ALSH)

- Similarity function sim:S × T → [-1,1]
- Mappings f:S→Γ and g:T→Γ
- ALSH: the probability distribution on the families F and G of clustering functions such that

$$\mathrm{E}_{f\in F, g\in G}[K_{f,g}(x,y)] = sim(x,y)$$

- α-ALSH:the probability distribution on the families F and G of clustering functions such that

$$\alpha\mathrm{E}_{f\in F, g\in G}[K_{f,g}(x,y)] - \theta = sim(x,y)$$

Can we really gain from this asymmetry?

Symmetric LSH vs Asymmetric LSH

- Given any large enough set of low dimensional unit vectors, there is **no α-LSH** for the Euclidian inner product.

- There is **no α-LSH** for several important similarity measures such as:
  - The Euclidian inner product
  - Overlap coefficient
  - Dice's coefficient.
  - …

- Any similarity function has α-ALSH and there exist a constant factor approximation algorithm for that.

Symmetric LSH vs Asymmetric LSH

- Given any large enough set of low dimensional unit vectors, there is **no α-LSH** for the Euclidian inner product.
- There is **no α-LSH** for several important similarity measures such as:
  - The Euclidian inner product
  - Overlap coefficient
  - Dice's coefficient.
  - …

- Any similarity function has α-ALSH and there exist a constant factor approximation algorithm for that.

What is the best α we can get?

The Centralized Cluster Ratio

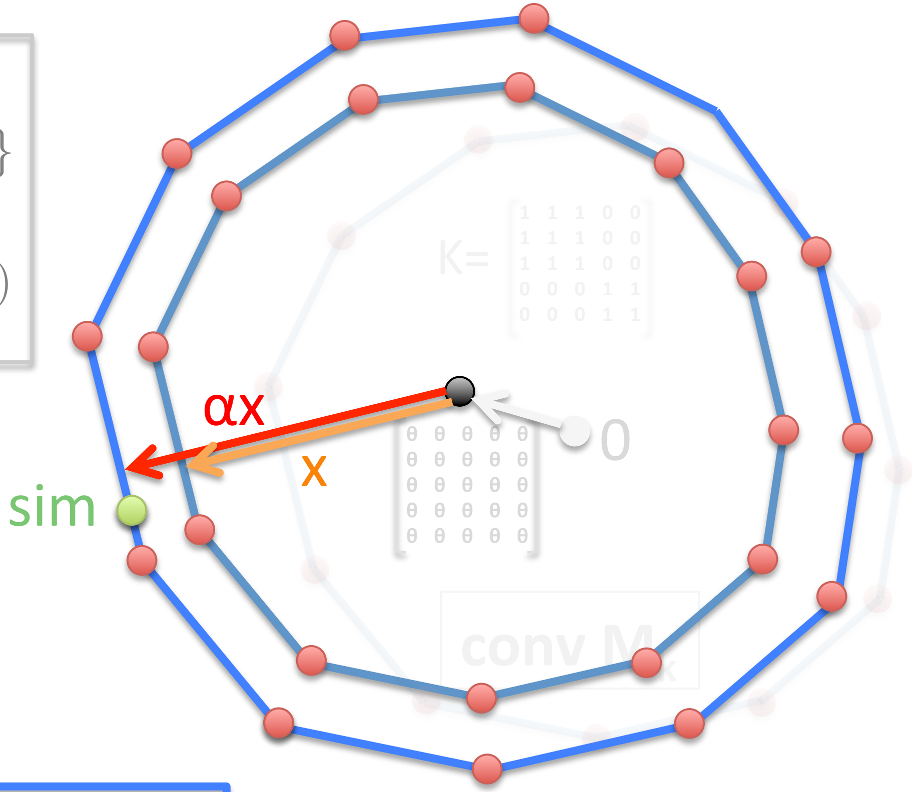
M_k={cluster matrices with k partition}

Centralized Cluster Ratio:

$$\rho_k(Z) = \min\{r | Z \in r\,\mathrm{conv} M_k + \theta\}$$

α-ALSH:

$$\alpha \mathbb{E}_{f\in\mathcal{F},g\in\mathcal{G}}[K_{f,g}(x,y)] - \theta = \mathrm{sim}(x,y)$$



The best α for α-ALSH is nothing but the centralized cluster ratio.

The Centralized Cluster Ratio

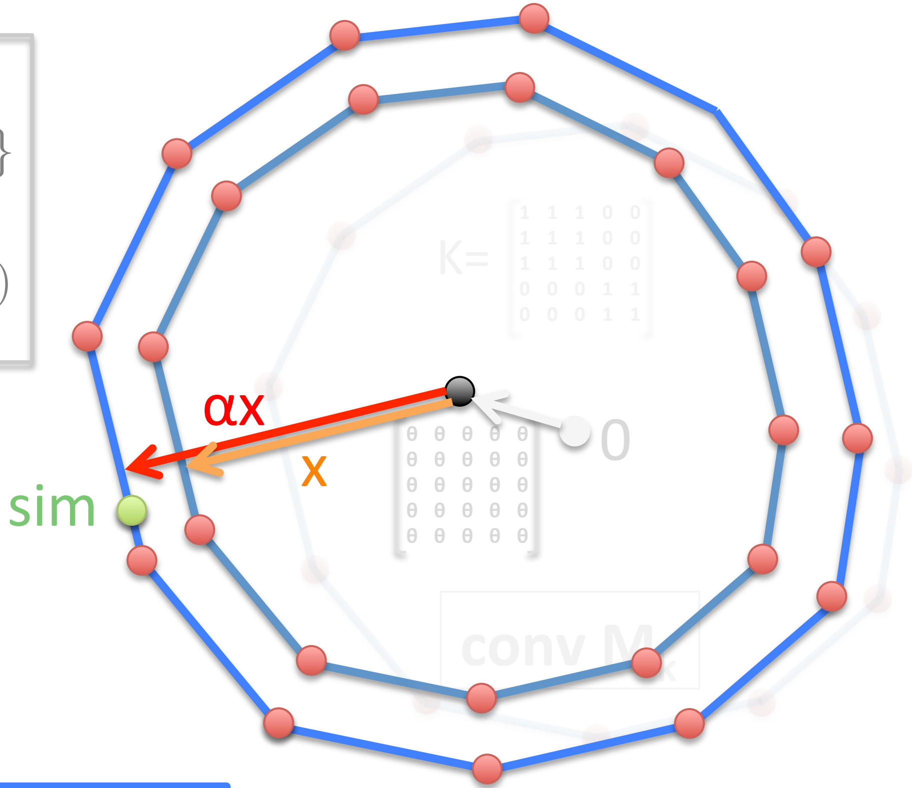
M_k={cluster matrices with k partition}

Centralized Cluster Ratio:

$$\rho_k(Z) = \min\{r | Z \in r\,\mathrm{conv} M_k + \theta\}$$

α-ALSH:

$$\alpha \mathbb{E}_{f\in\mathcal{F}, g\in\mathcal{G}}[K_{f,g}(x, y)] - \theta = \mathrm{sim}(x, y)$$



Everything comes down to the cluster ratio and centralized cluster ratio!

The best α for α-ALSH is nothing but the centralized cluster ratio.

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \leq \frac{1}{2}\hat{\rho}_2(sim) \leq \hat{\rho}_\infty(sim) \leq \hat{\rho}_k(sim) \leq \hat{\rho}_2(sim) \leq \kappa_R\|sim\|_{\widehat{\max}}$$

$$\frac{1}{3}\|sim\|_{\max} \leq \rho_\infty(sim) \leq \rho_k(sim) \leq \rho_2(sim) \leq \kappa_R\|sim\|_{\max}$$

$1.67 \leq \kappa_R \leq 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \leq \frac{1}{2}\hat{\rho}_2(sim) \leq \hat{\rho}_\infty(sim) \leq \hat{\rho}_k(sim) \leq \hat{\rho}_2(sim) \leq \kappa_R\|sim\|_{\widehat{\max}}$$

$$\frac{1}{3}\|sim\|_{\max} \leq \rho_\infty(sim) \leq \rho_k(sim) \leq \rho_2(sim) \leq \kappa_R\|sim\|_{\max}$$

$1.67 \leq \kappa_R \leq 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

Interpretation 1

Tight (factor of less than 4) characterization of the smallest α for which we can obtain an α-ALSH (and an approximation algorithm for that.)

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \leq \frac{1}{2}\hat{\rho}_2(sim) \leq \hat{\rho}_\infty(sim) \leq \hat{\rho}_k(sim) \leq \hat{\rho}_2(sim) \leq \kappa_R\|sim\|_{\widehat{\max}}$$

$$\frac{1}{3}\|sim\|_{\max} \leq \rho_\infty(sim) \leq \rho_k(sim) \leq \rho_2(sim) \leq \kappa_R\|sim\|_{\max}$$

$1.67 \leq \kappa_R \leq 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

Interpretation 1

Tight (factor of less than 4) characterization of the smallest α for which we can obtain an α-ALSH (and an approximation algorithm for that.)

Interpretation 2

SDP relaxation given by max-norm provided a very tight (factor of less than 6) relaxation for co-clustering and asymmetric hamming embedding:

$$\{Z \mid \|Z\|_{\max} \leq 1/K\} \subseteq \mathrm{conv} M_k \subseteq \{Z \mid \|Z\|_{\max} \leq 3\}$$

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \le \frac{1}{2}\hat{\rho}_2(sim) \le \boxed{\hat{\rho}_\infty(sim) \le \hat{\rho}_k(sim) \le \hat{\rho}_2(sim)} \le \kappa_R\|sim\|_{\widehat{\max}}$$

$$\frac{1}{3}\|sim\|_{\max} \le \boxed{\rho_\infty(sim) \le \rho_k(sim) \le \rho_2(sim)} \le \kappa_R\|sim\|_{\max}$$

$1.67 \le \kappa_R \le 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

$$\forall k > 2, \quad \mathrm{conv}\, M_2 \subset \mathrm{conv}\, M_k \subset \mathrm{conv}\, M_\infty$$

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \leq \boxed{\frac{1}{2}\hat{\rho}_2(sim) \leq \hat{\rho}_\infty(sim)} \leq \hat{\rho}_k(sim) \leq \hat{\rho}_2(sim) \leq \kappa_R\|sim\|_{\widehat{\max}}$$

$$\frac{1}{3}\|sim\|_{\max} \leq \rho_\infty(sim) \leq \rho_k(sim) \leq \rho_2(sim) \leq \kappa_R\|sim\|_{\max}$$

$1.67 \leq \kappa_R \leq 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

Generalization of the following lemma to **α-ALSH:**

[Charikar, STOC 2002]
If sim(x,y) is LSH-able then (1+sim(x,y))/2 is LSH-able by binary hash functions.

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \le \frac{1}{2}\hat{\rho}_2(sim) \le \hat{\rho}_\infty(sim) \le \hat{\rho}_k(sim) \le \hat{\rho}_2(sim) \le \kappa_R\|sim\|_{\widehat{\max}}$$

$$\boxed{\frac{1}{3}\|sim\|_{\max} \le \rho_\infty(sim)} \le \rho_k(sim) \le \rho_2(sim) \le \kappa_R\|sim\|_{\max}$$

$1.67 \le \kappa_R \le 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

$$K = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

$$U = \begin{array}{c} \begin{array}{ccc} \text{Cluster 1} & \text{Cluster 2} & \text{Cluster 3} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \quad V = \begin{array}{c} \begin{array}{ccc} \text{Cluster 1} & \text{Cluster 2} & \text{Cluster 3} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$\begin{aligned} \|K\|_{\max} &= \|2UV^\top - 1\|_{\max} \\ &\le 2\|UV^\top\|_{\max} + 1 \\ &= 3 \end{aligned}$$

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\boxed{\frac{1}{2}\|sim\|_{\widehat{\max}} \le \frac{1}{2}\hat{\rho}_2(sim)} \le \hat{\rho}_\infty(sim) \le \hat{\rho}_k(sim) \le \hat{\rho}_2(sim) \le \kappa_R\|sim\|_{\widehat{\max}}$$

$$\frac{1}{3}\|sim\|_{\max} \le \rho_\infty(sim) \le \rho_k(sim) \le \rho_2(sim) \le \kappa_R\|sim\|_{\max}$$

$1.67 \le \kappa_R \le 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

$$K = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad U = \begin{array}{c} \text{Cluster 1} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \end{array} \quad V = \begin{array}{c} \text{Cluster 1} \\ \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{array}$$

Tight bounds on Cluster Ratio based on Max-norm relaxation

$$\frac{1}{2}\|sim\|_{\widehat{\max}} \le \frac{1}{2}\hat{\rho}_2(sim) \le \hat{\rho}_\infty(sim) \le \hat{\rho}_k(sim) \le \boxed{\hat{\rho}_2(sim) \le \kappa_R\|sim\|_{\widehat{\max}}}$$

$$\frac{1}{3}\|sim\|_{\max} \le \rho_\infty(sim) \le \rho_k(sim) \le \boxed{\rho_2(sim) \le \kappa_R\|sim\|_{\max}}$$

$1.67 \le \kappa_R \le 1.79$

max-norm:

$$\|Z\|_{\max} = \min_{UV^\top = Z} \max(\|U_i\|_2^2, \|V_i\|_2^2)$$

centralized max-norm:

$$\|Z\|_{\widehat{\max}} = \min_{\theta \in \mathbb{R}} \|Z - \theta\|_{\max}$$

Max-norm relaxation and then random projections.

Similar to [Alon et al. SIAM J. Comput. 2006]

Clustering, Hamming Embedding, Generalized LSH and the Max Norm

Behnam Neyshabur Yury Makarychev Nathan Srebro

- ✓ Tight (factor ≤ 6) max-norm based SDP relaxation for co-clustering and asymmetric hamming embedding
- ✓ Introducing ALSH, proving that α-ALSH exists for any similarity function while α-LSH does not exist for many functions
- ✓ Tight (factor ≤ 4) characterization of the smallest α for which we can obtain an α-ALSH (and an approximation algorithm for that)