Jointly Clustering Rows and Columns of Binary Matrices: Algorithms and Trade-offs

Jiaming Xu

Joint work with Rui Wu, Kai Zhu, Bruce Hajek, R. Srikant, and Lei Ying

University of Illinois, Urbana-Champaign

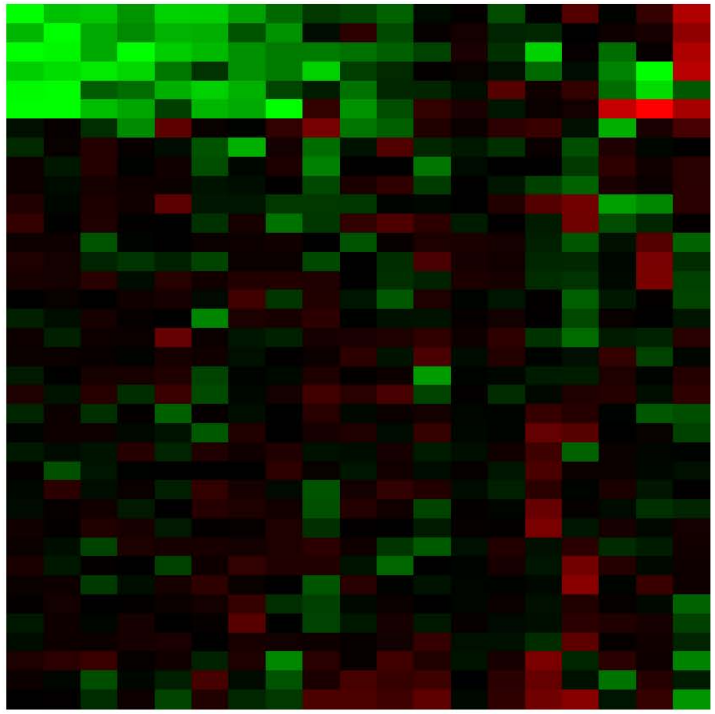
June 17, 2014

Motivation

Data matrices with both row and column cluster structure arise in many applications

	7.5			0.5	1.0	5.0		5.5	
	6.5	3.0	3.5	4.0			8.5	9.0	0.5
	6.0			3.0			0.5	9.5	10.0
			10.0		9.5				1.5
	8.0	10.0	6.0	4.5	8.5	5.0			
		9.0				9.5		10.0	
	5.0	6.0	4.0		9.0	5.0	7.5	2.0	
			2.5	7.0	7.5		3.0	6.5	
	1.0		6.5			7.5	7.5	7.0	8.0

User rating matrix



Gene expression matrix

Goal: Cluster rows and columns based on a noisy, partially observed matrix

Simple model

1. like: $+1$; dislike: -1
2. n users (movies) form r clusters of equal size K
3. users in the same cluster give the same rating to movies in the same cluster
4. block rating is $+1$ or -1 with equal prob.

+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	+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

Ground truth Y^*: binary block-constant matrix

Simple model

1. like: $+1$; dislike: -1
2. n users (movies) form r clusters of equal size K
3. users in the same cluster give the same rating to movies in the same cluster
4. block rating is $+1$ or -1 with equal prob.

+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	+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

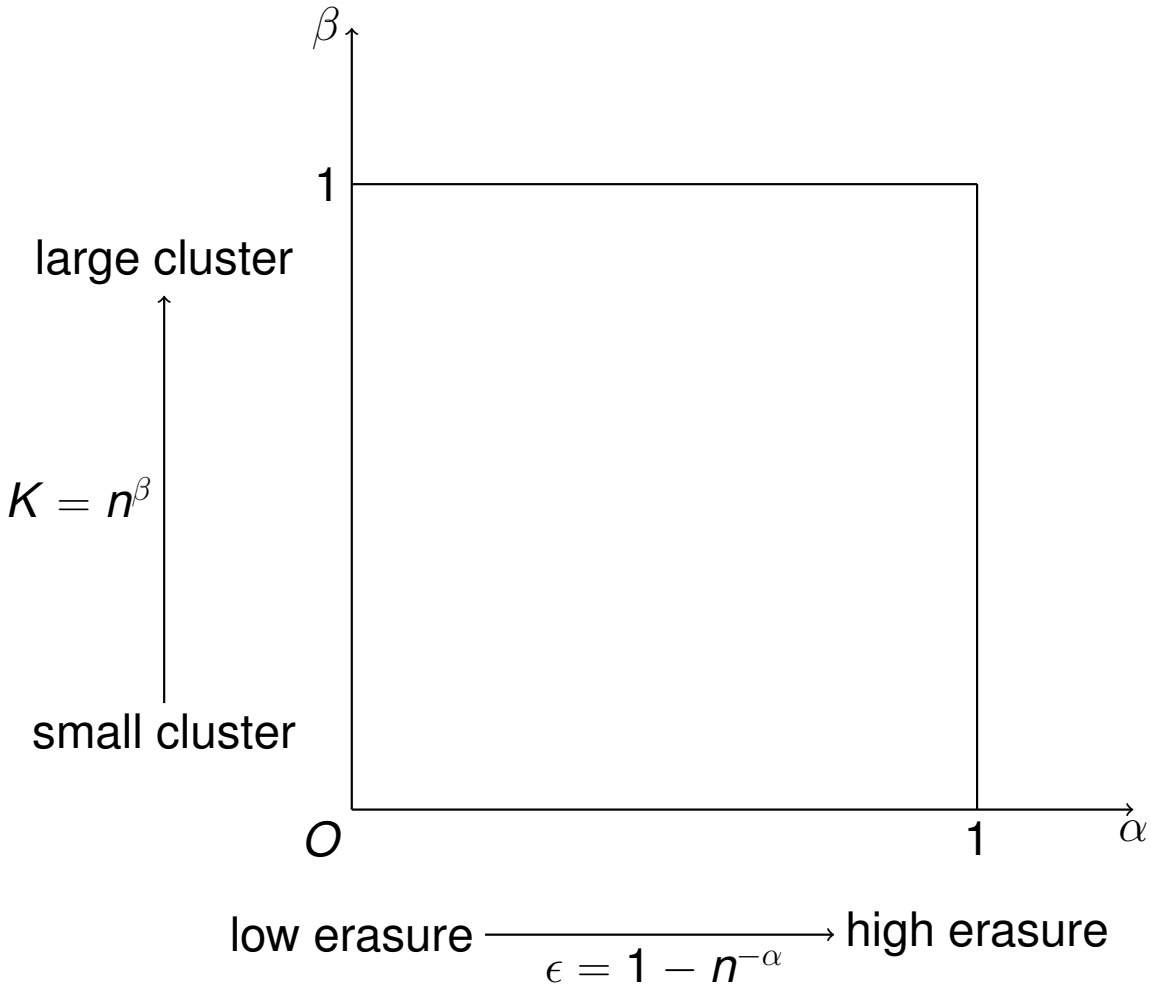
Ground truth Y^*: binary block-constant matrix

+1	*	*	*	+1	*	*	−1	*
+1	*	*	*	*	*	*	*	*
*	+1	+1	*	*	*	*	*	*
*	*	−1	+1	*	*	−1	−1	*
*	*	*	−1	*	*	*	*	*
*	*	−1	−1	*	*	*	+1	*
*	*	*	*	+1	*	*	−1	*
+1	*	+1	*	*	*	*	*	*
*	−1	*	*	*	*	*	−1	*

Partial and noisy observation R: erasure prob. ϵ flipping prob. p

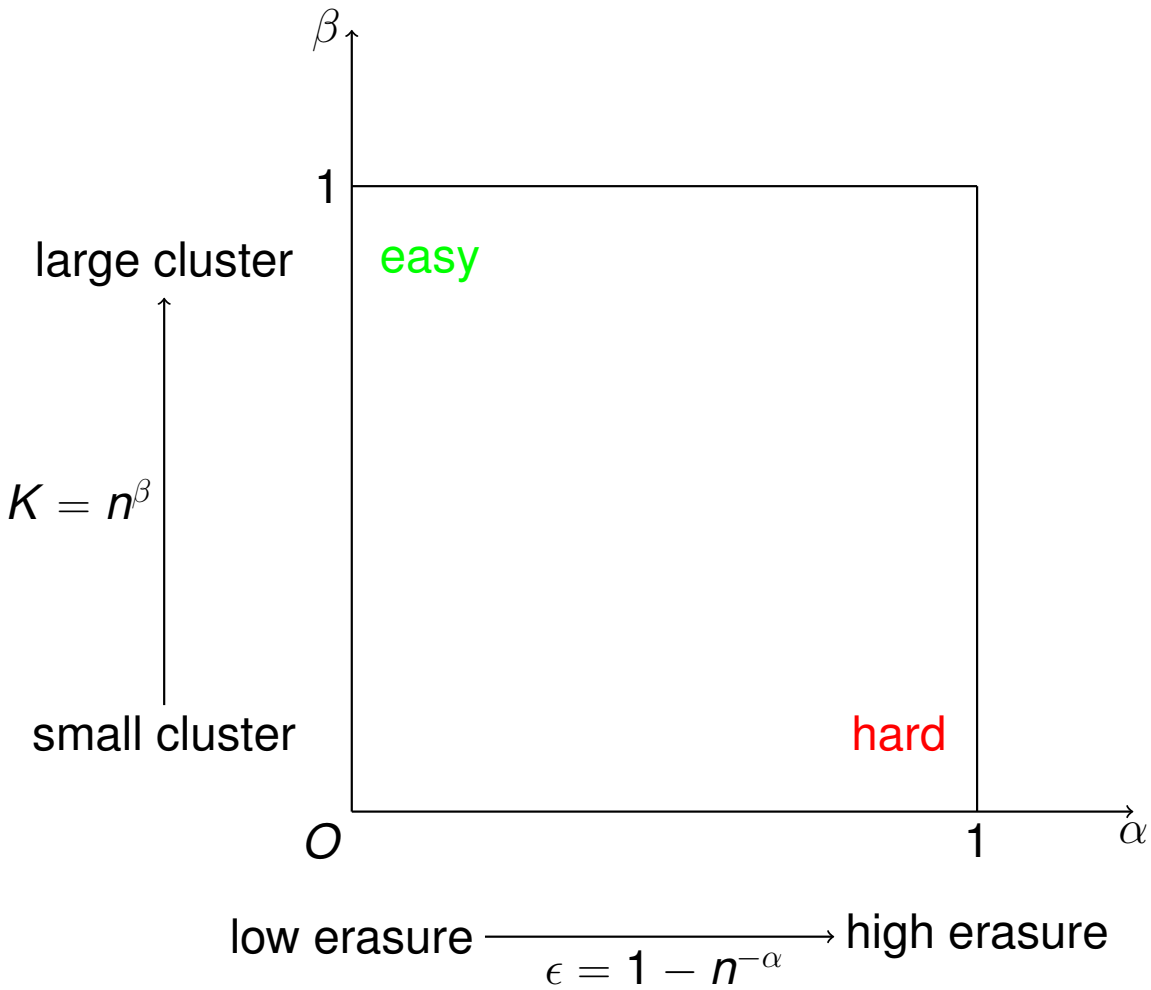
When cluster recovery is possible (impossible)?

Assume that $0 \leq p < 1/2$ is a constant. Our results apply to the general setting allowing any K, ϵ.



When cluster recovery is possible (impossible)?

Assume that $0 \leq p < 1/2$ is a constant. Our results apply to the general setting allowing any K, ϵ.



Outline of the remainder

1. Impossible regime
2. Nearest-neighbor clustering
3. Spectral method
4. Convex method
5. Maximum likelihood estimation (MLE)

Impossible regime

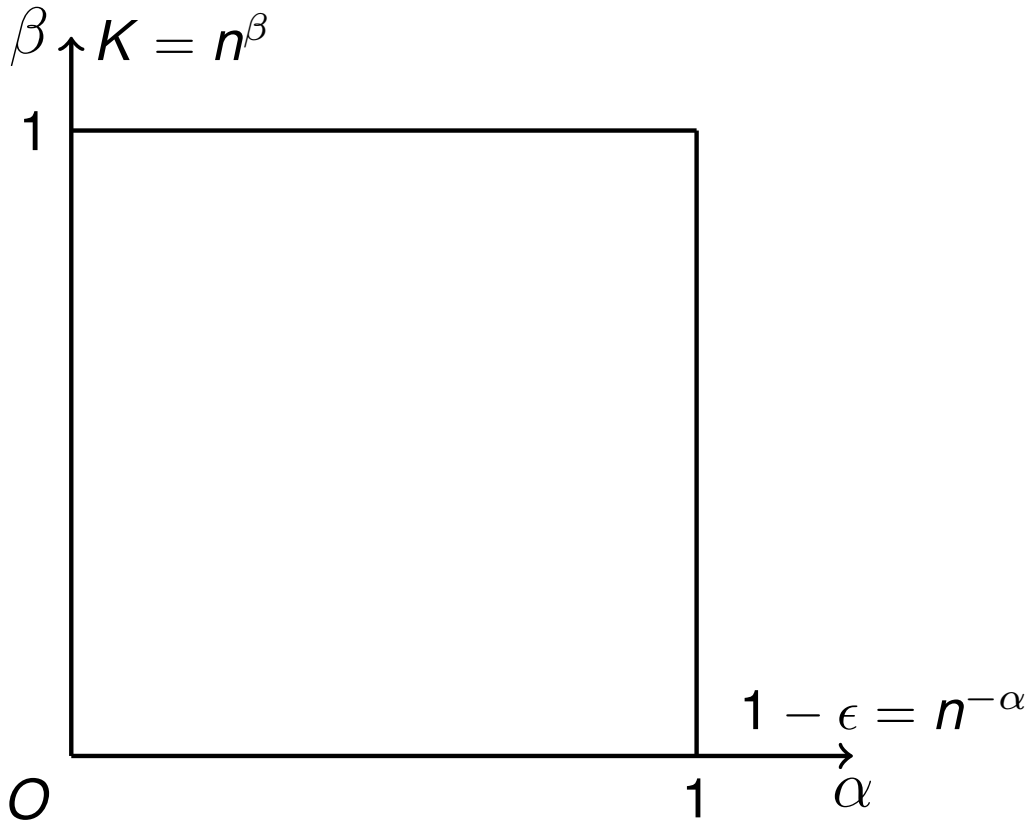
- Genie-aided with the set of flipped entries revealed

Impossible regime

- Genie-aided with the set of flipped entries revealed
- Construct a new user clustering by swapping two rows in two different row clusters

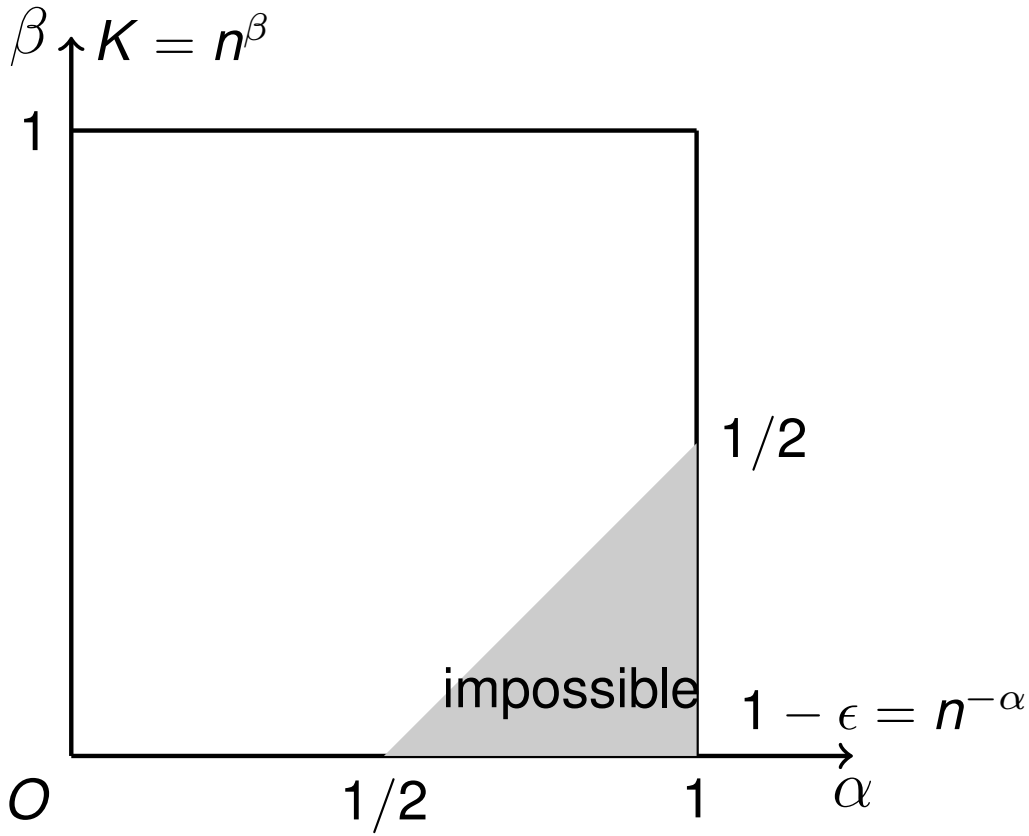
Impossible regime

- Genie-aided with the set of flipped entries revealed
- Construct a new user clustering by swapping two rows in two different row clusters



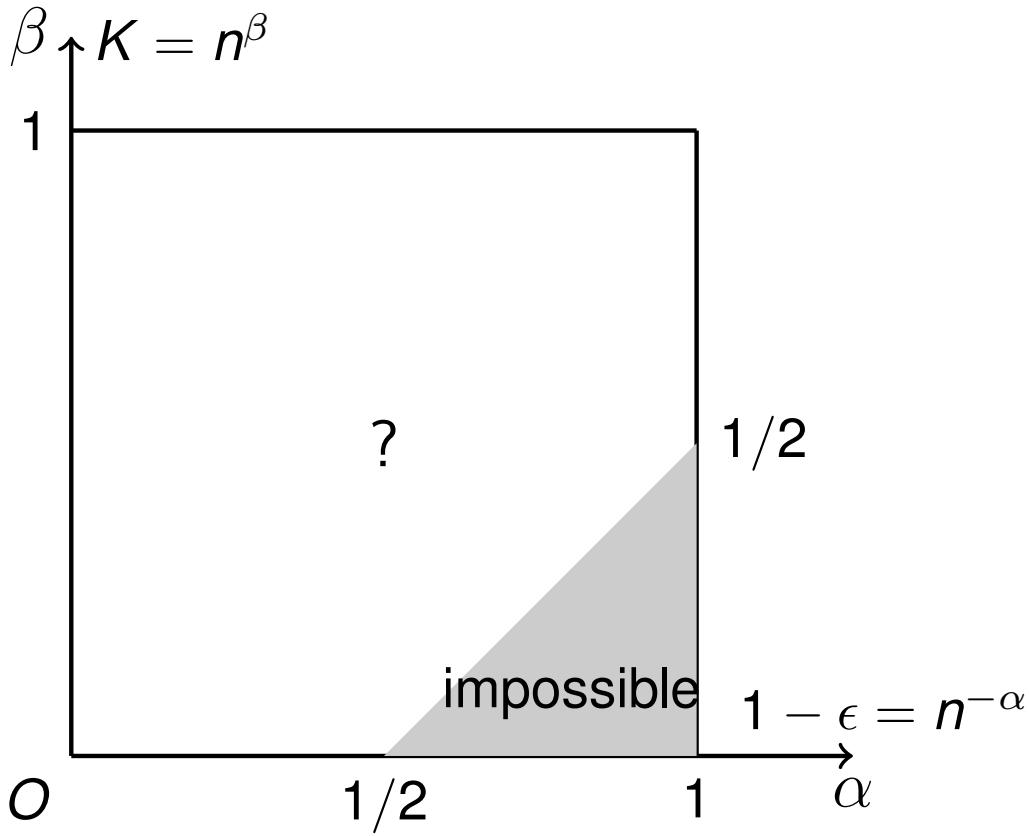
Impossible regime

- Genie-aided with the set of flipped entries revealed
- Construct a new user clustering by swapping two rows in two different row clusters



Impossible regime

- Genie-aided with the set of flipped entries revealed
- Construct a new user clustering by swapping two rows in two different row clusters



Nearest-neighbor clustering

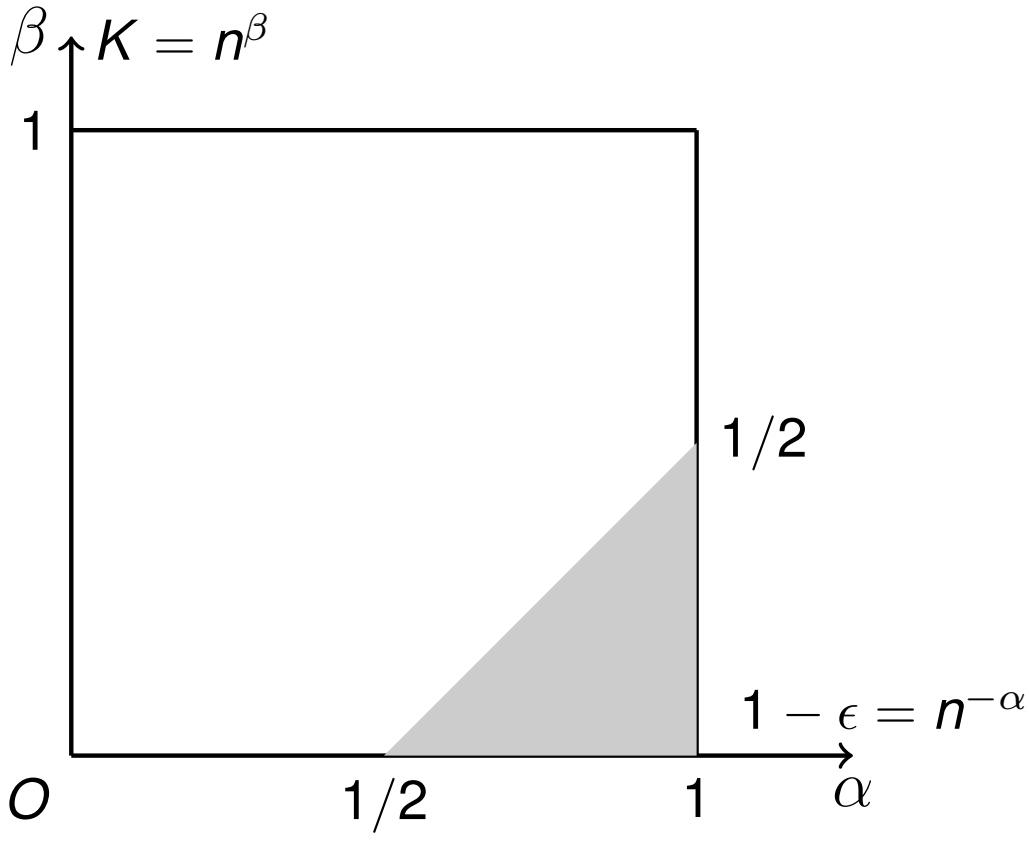
Similarity between two users: The number of movies with the same observed rating [Dabeer et al. '12]

Algorithm: Each user finds the $K - 1$ most similar users

Nearest-neighbor clustering

Similarity between two users: The number of movies with the same observed rating [Dabeer et al. '12]

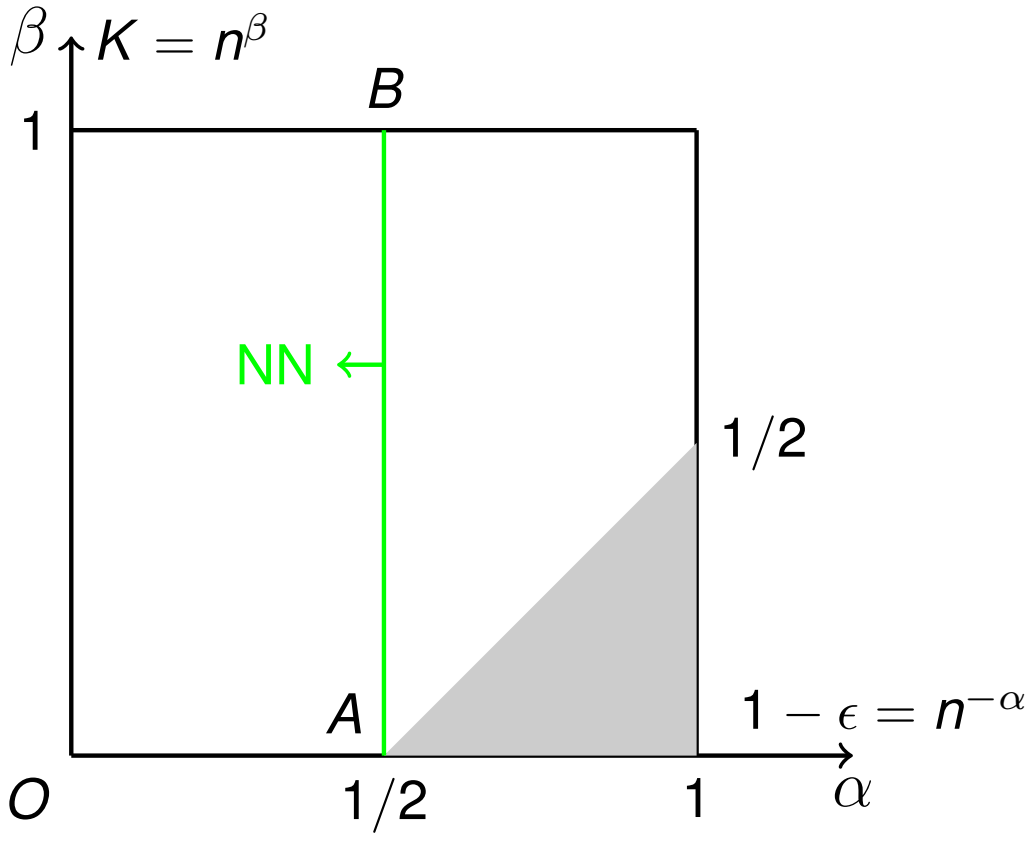
Algorithm: Each user finds the $K-1$ most similar users



Nearest-neighbor clustering

Similarity between two users: The number of movies with the same observed rating [Dabeer et al. '12]

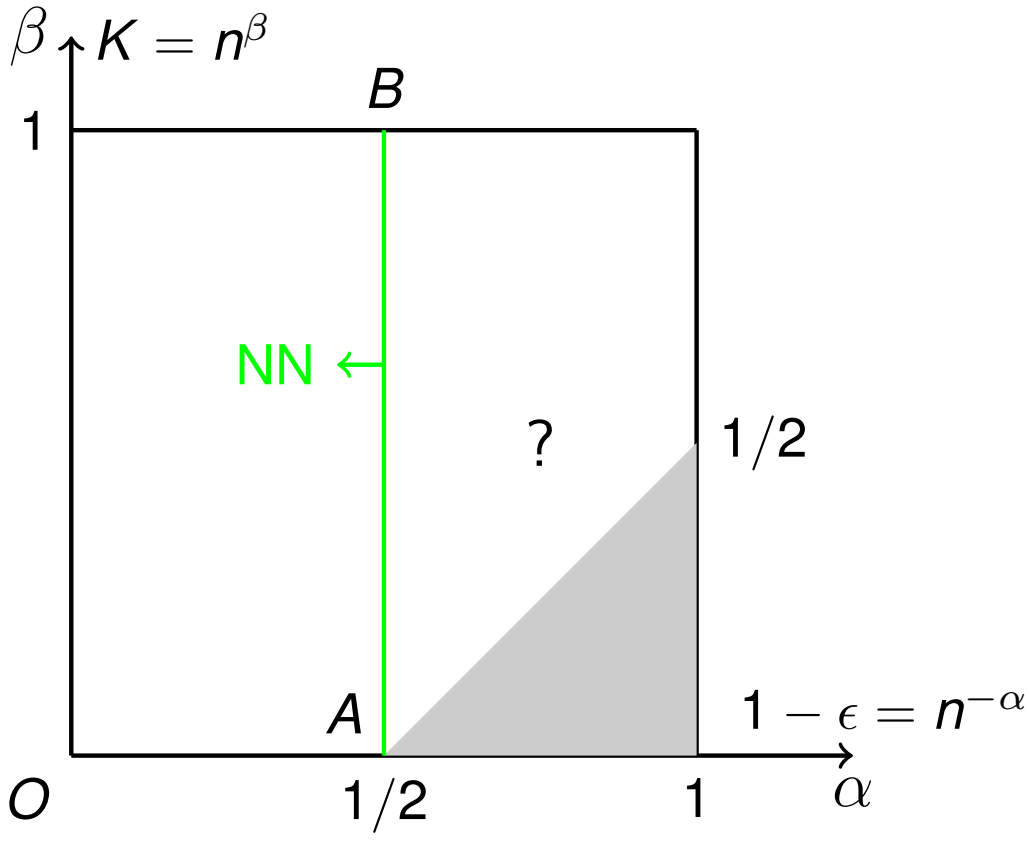
Algorithm: Each user finds the $K-1$ most similar users



Nearest-neighbor clustering

Similarity between two users: The number of movies with the same observed rating [Dabeer et al. '12]

Algorithm: Each user finds the $K-1$ most similar users

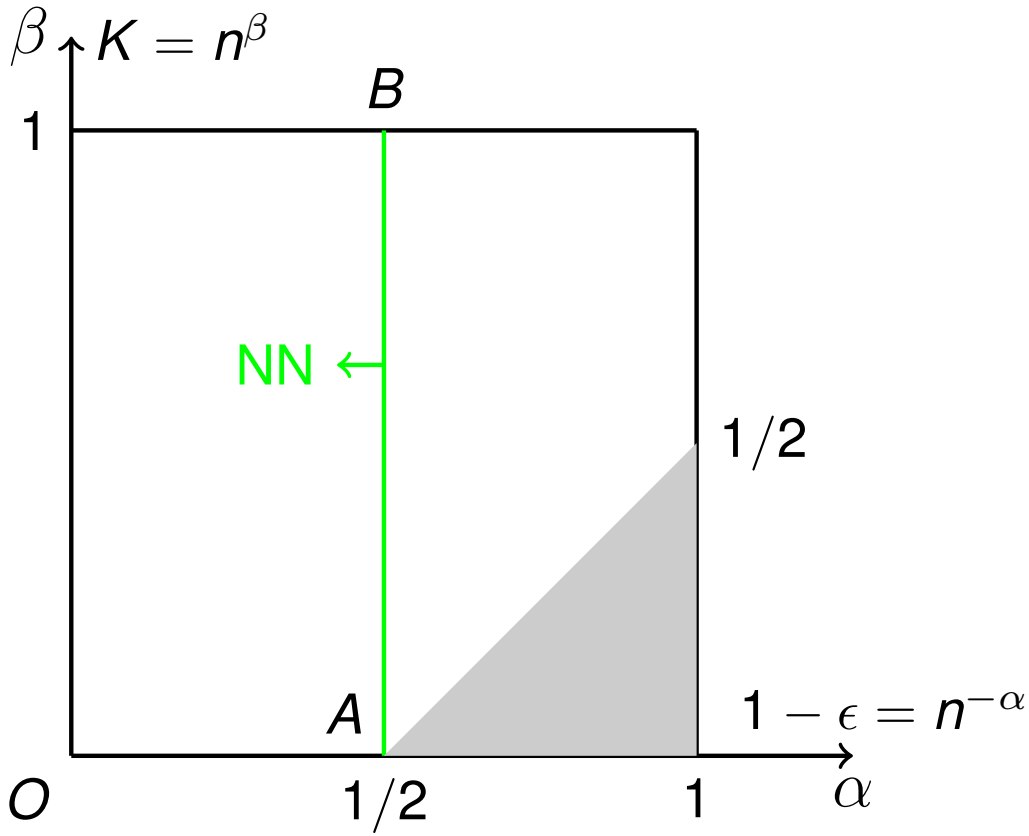


Spectral method

1. Approximately clustering rows and columns of the best rank r approximation $P_r(R)$
2. Majority voting within each block of R
3. Reclustering by assigning rows and columns to nearest centers

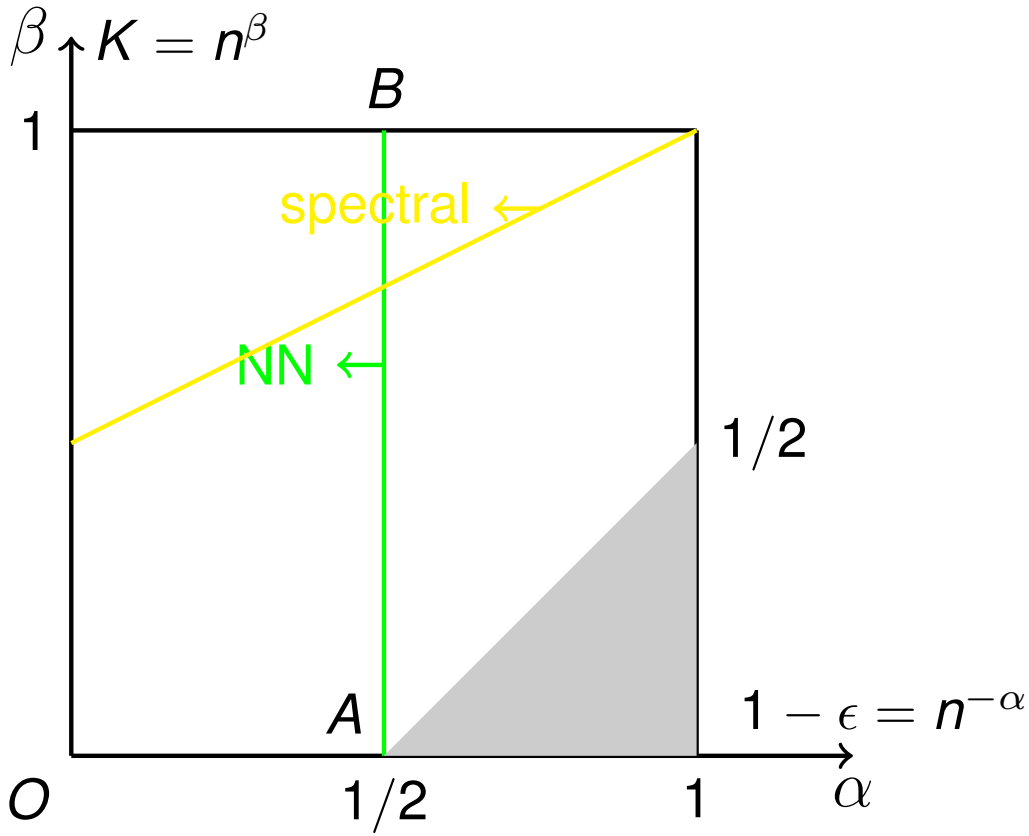
Spectral method

1. Approximately clustering rows and columns of the best rank r approximation $P_r(R)$
2. Majority voting within each block of R
3. Reclustering by assigning rows and columns to nearest centers



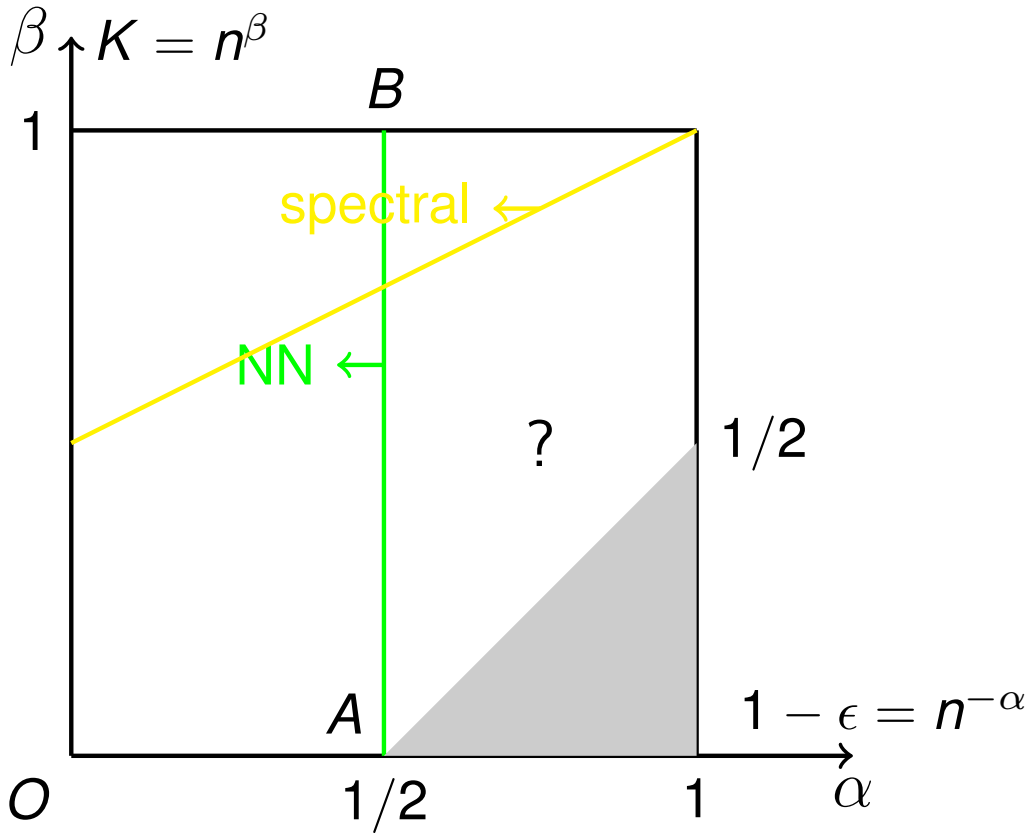
Spectral method

1. Approximately clustering rows and columns of the best rank r approximation $P_r(R)$
2. Majority voting within each block of R
3. Reclustering by assigning rows and columns to nearest centers



Spectral method

1. Approximately clustering rows and columns of the best rank r approximation $P_r(R)$
2. Majority voting within each block of R
3. Reclustering by assigning rows and columns to nearest centers



Convex method

Clustering by first recovering ground truth Y^*:

$$Y^* \longrightarrow R \longrightarrow Y$$

Convex method

Clustering by first recovering ground truth Y^*:

$$Y^* \longrightarrow R \longrightarrow Y$$

MLE is to find a block-constant binary matrix Y matching R as much as possible

Convex method

Clustering by first recovering ground truth Y^*:

$$Y^* \longrightarrow R \longrightarrow Y$$

MLE is to find a block-constant binary matrix Y matching R as much as possible
A convex relaxation of MLE:

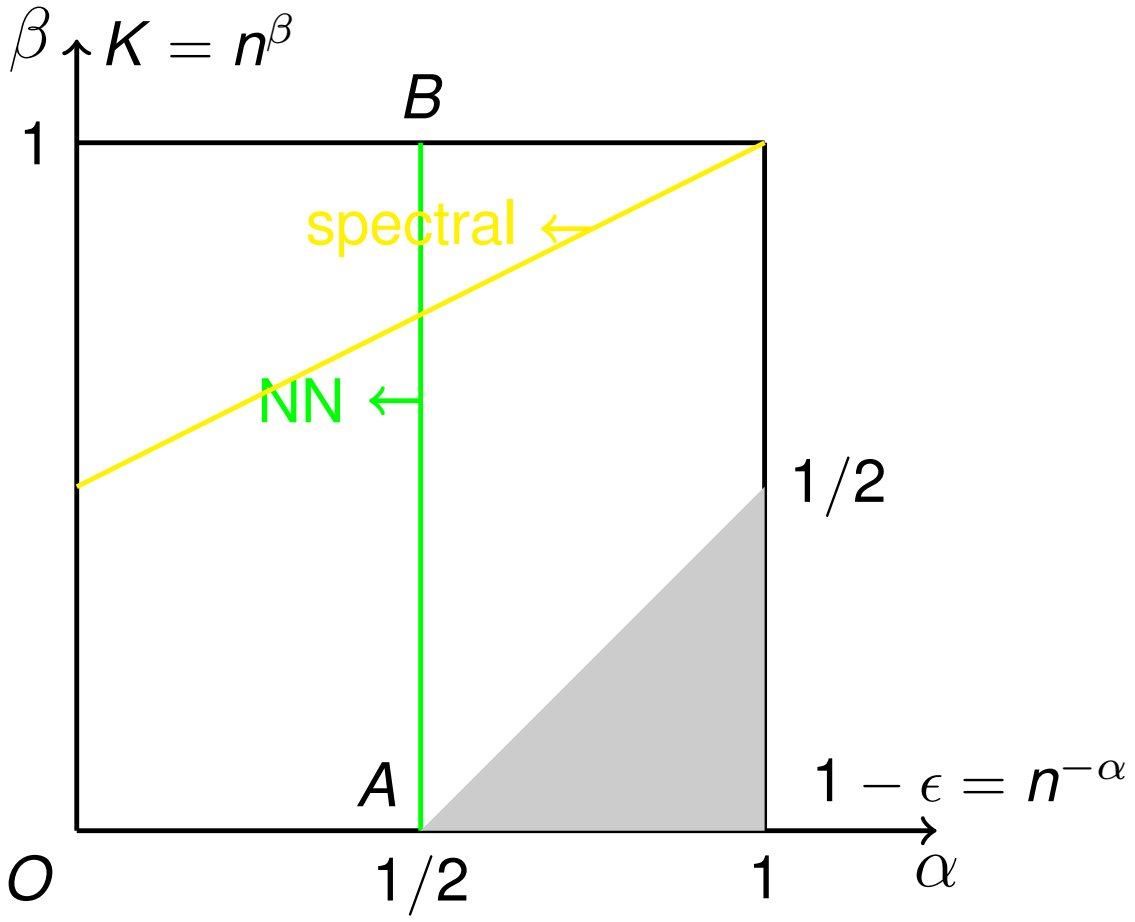
$$\max_{Y} \sum_{i,j} R_{ij} Y_{ij} - \lambda \| Y \|_*$$

$$\text{s.t.} \quad Y_{ij} \in [-1, 1],$$

$$\lambda = C\sqrt{(1-\epsilon)n}, C \geq 3$$

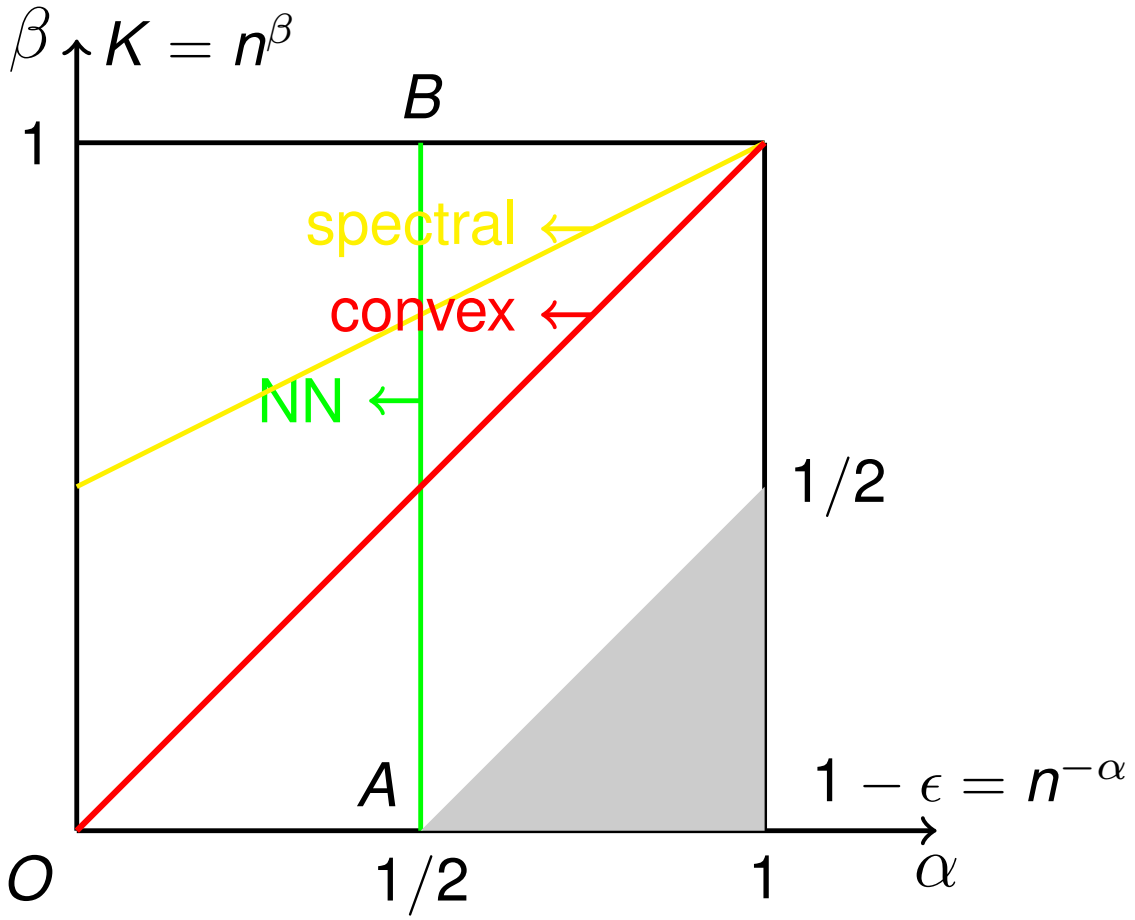
Performance of convex method

Assume a technical conjecture (come back later) holds



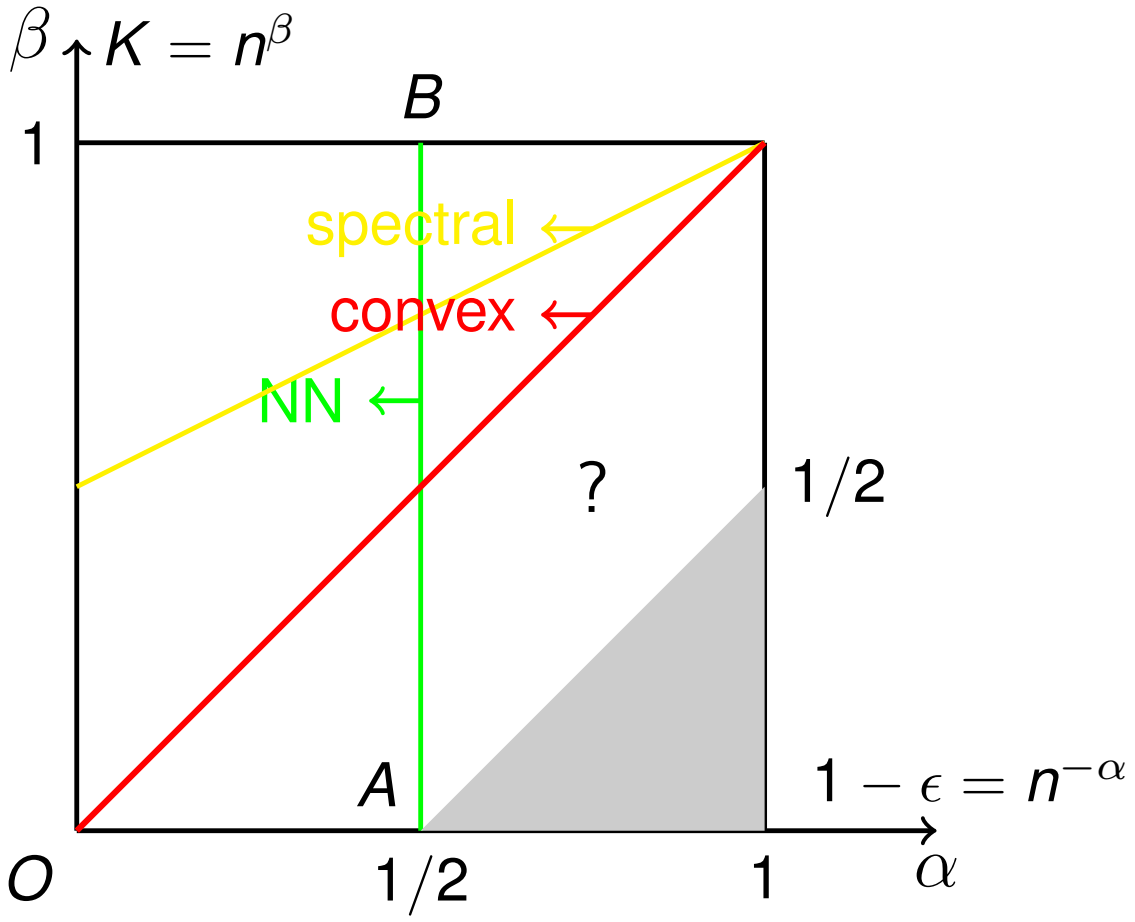
Performance of convex method

Assume a technical conjecture (come back later) holds

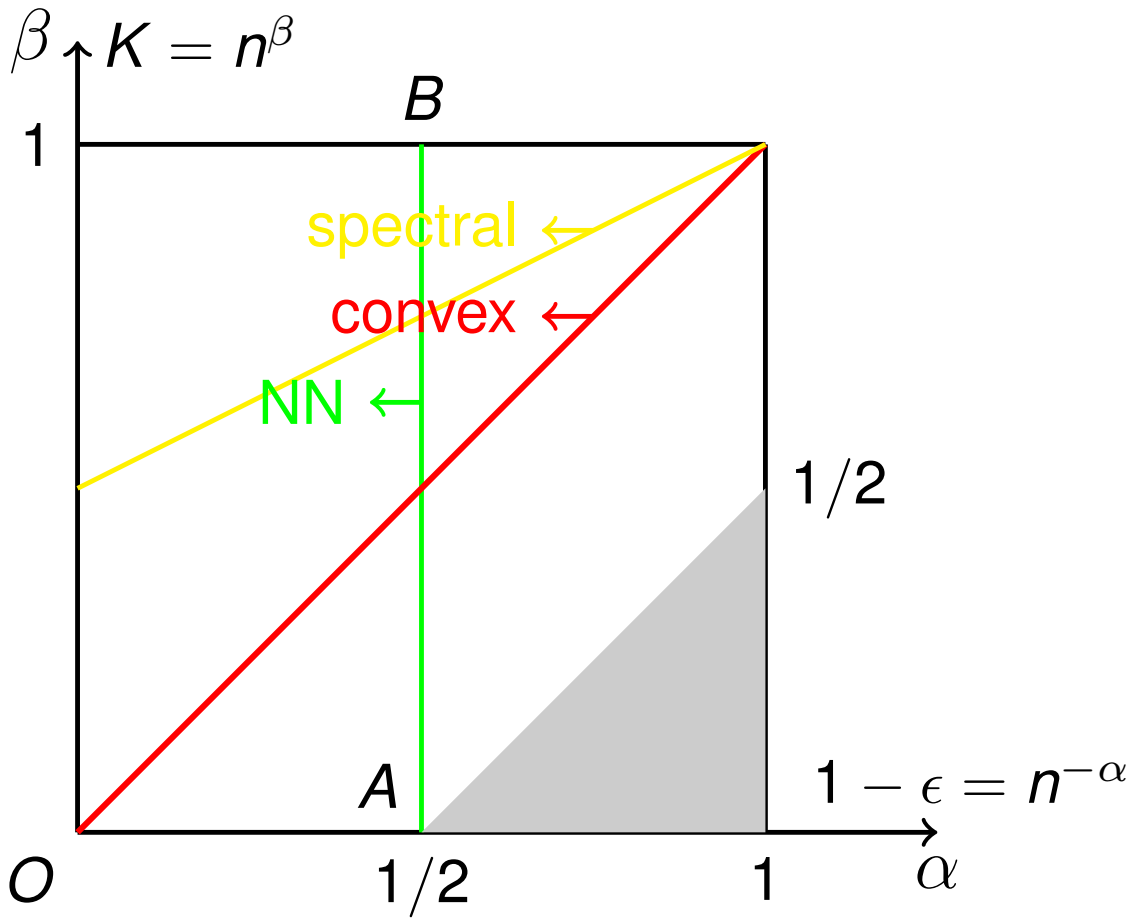


Performance of convex method

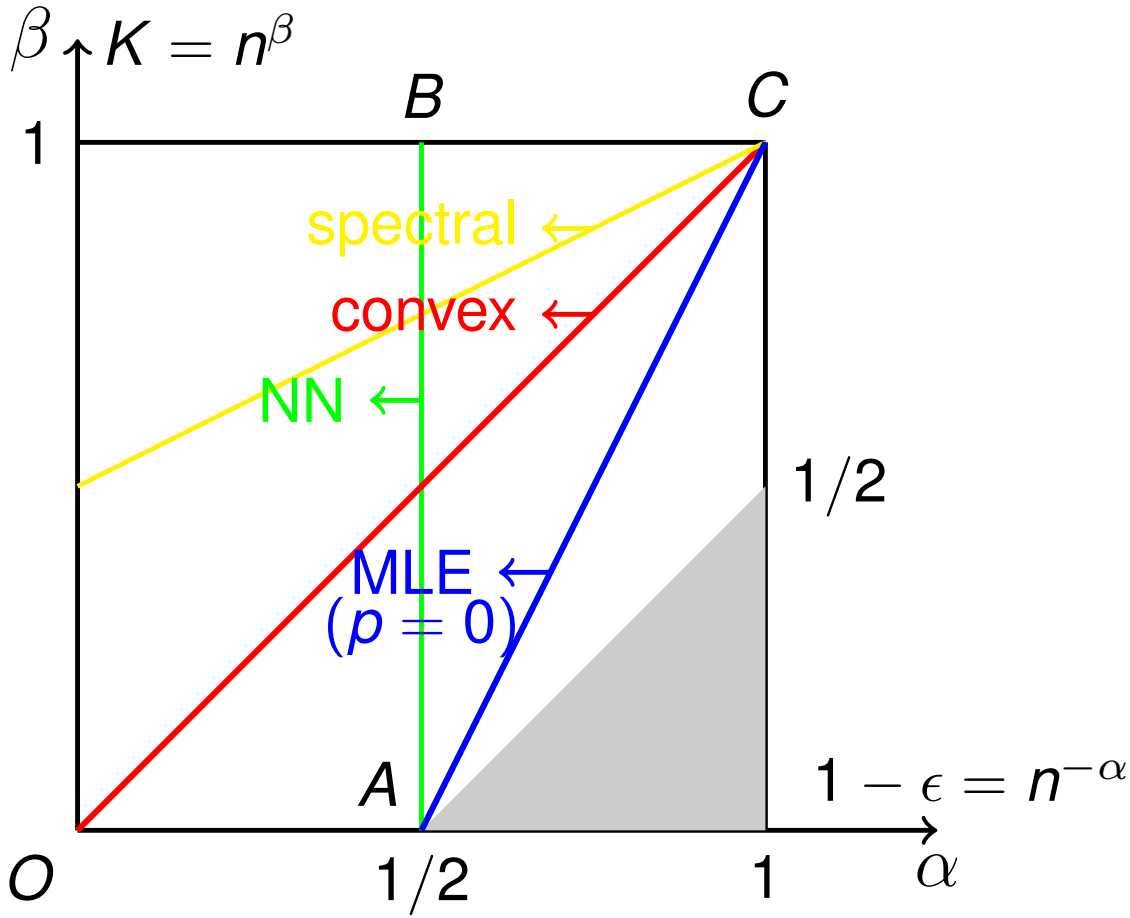
Assume a technical conjecture (come back later) holds



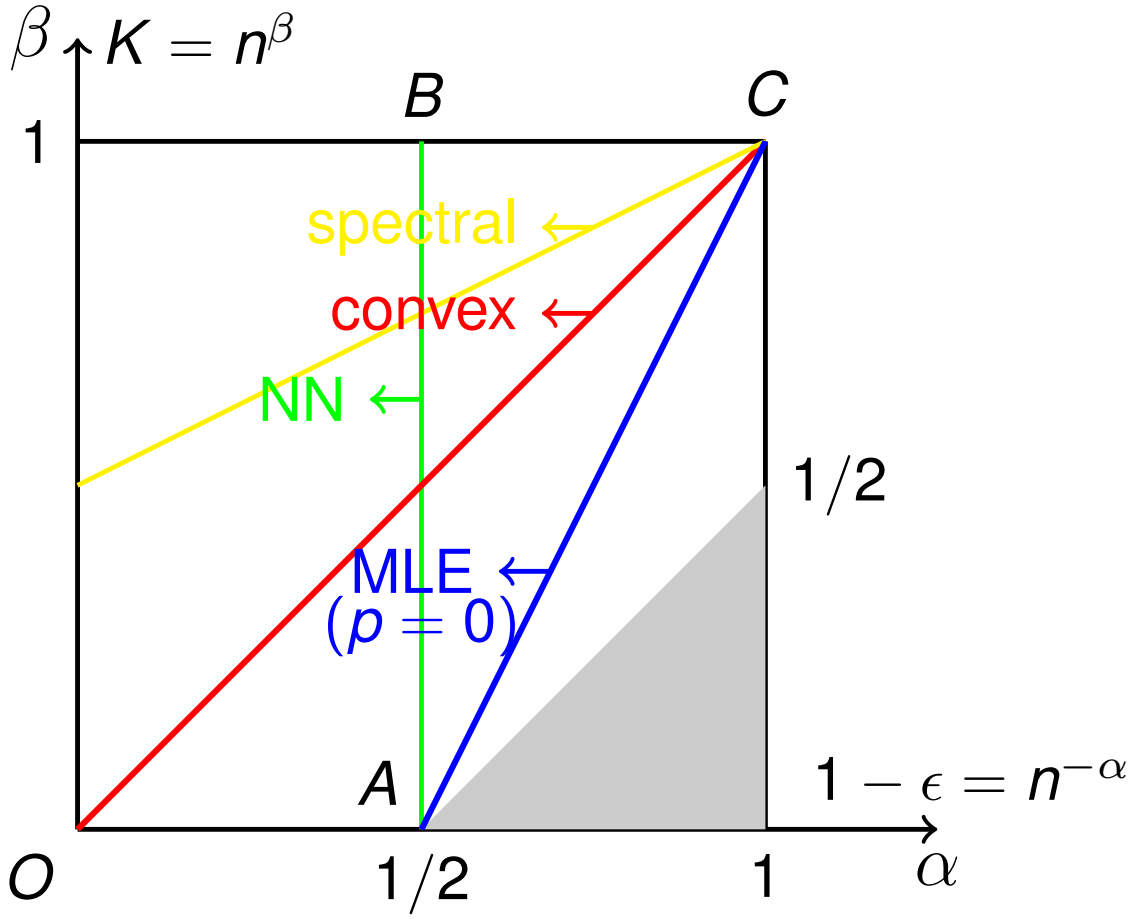
Performance of MLE



Performance of MLE



Performance of MLE



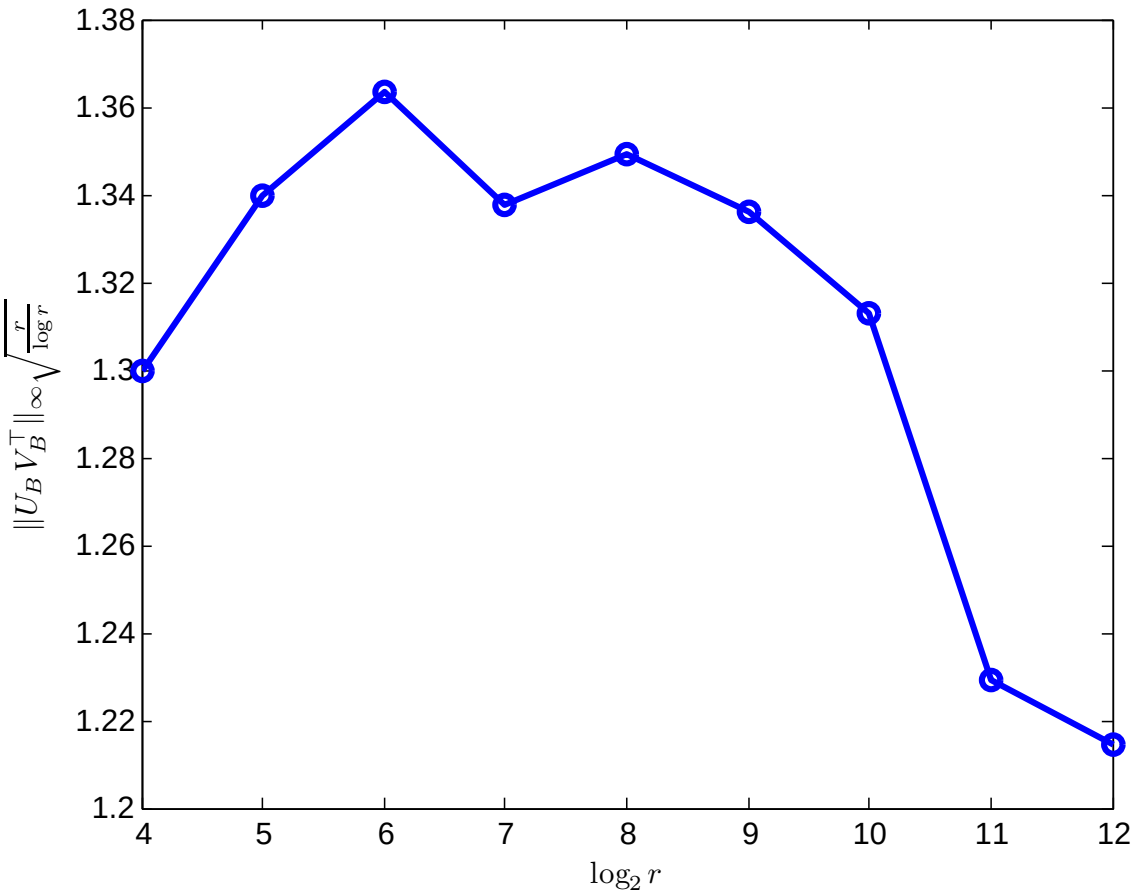
Conjecture: MLE succeeds all the way up to the gray region

Conjecture on convex method

Conjecture: For a $r \times r$ random sign matrix B with SVD $B = U\Sigma V^\top$, $\|UV^\top\|_\infty$ scales as $\sqrt{\frac{\log r}{r}}$.

Conjecture on convex method

Conjecture: For a $r \times r$ random sign matrix B with SVD $B = U\Sigma V^\top$, $\|UV^\top\|_\infty$ scales as $\sqrt{\frac{\log r}{r}}$.



Please check our paper for details
Thank you!

Questions?