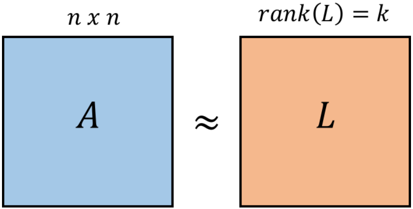
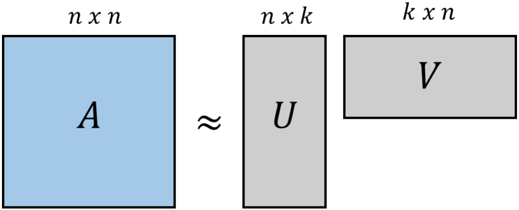
LOW-RANK APPROXIMATION FROM COMMUNICATION COMPLEXITY

Cameron Musco (University of Massachusetts Amherst)
Joint with Christopher Musco (NYU) and David Woodruff (CMU)

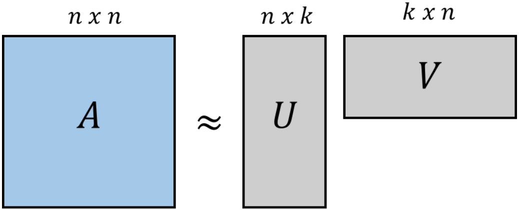
MASKED LOW-RANK APPROXIMATION



$$\underset{\{L:\text{rank}(L)=k\}}{\arg\min} \|A - L\|_F^2 = \sum_{i,j} (A - L)_{(i,j)}^2.$$

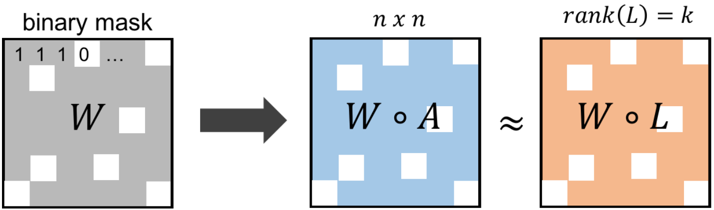


$$\underset{\{L:\text{rank}(L)=k\}}{\arg\min} \|A-L\|_F^2 = \sum_{i,j}(A-L)_{(i,j)}^2.$$



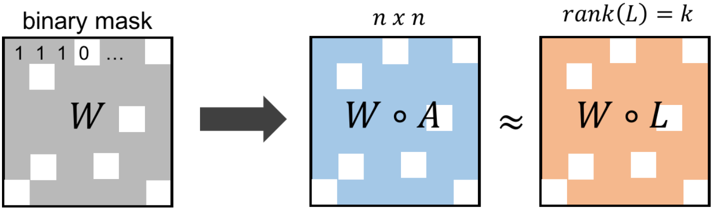
$$\arg\min_{\{L:\text{rank}(L)=k\}} \|A - L\|_F^2 = \sum_{i,j} (A - L)_{(i,j)}^2.$$

Often want to perform low-rank approximation when some entries in A are unknown or don't follow low-rank structure.



$$\arg\min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 = \sum_{(i,j):W_{i,j}=1} (A - L)_{i,j}^2.$$

Often want to perform low-rank approximation when some entries in A are unknown or don't follow low-rank structure.

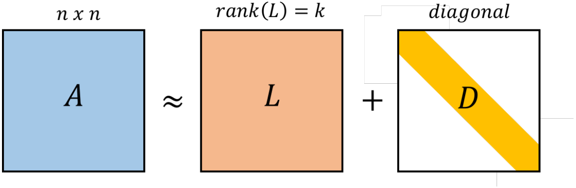


$$\arg\min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 = \sum_{(i,j):W_{i,j}=1} (A - L)_{i,j}^2.$$

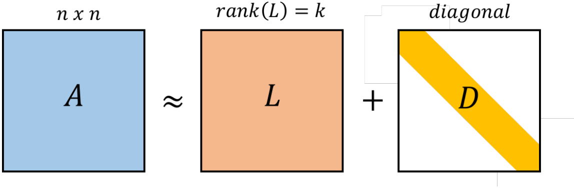
Often want to perform low-rank approximation when some entries in A are unknown or don't follow low-rank structure.

A special case of weighted low-rank approximation. Depending on W, captures many problems.

Low-Rank + Diagonal Approximation: Many matrices can be well approximated by a low-rank component plus a diagonal matrix.

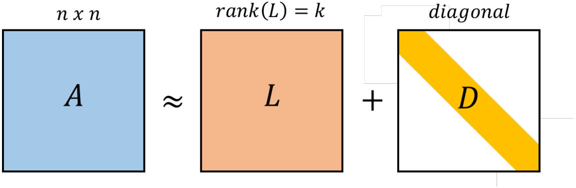


Low-Rank + Diagonal Approximation: Many matrices can be well approximated by a low-rank component plus a diagonal matrix.



Factor analysis (PCA with different noise variance in each dimension), kernel matrix approximation, source separation, ...

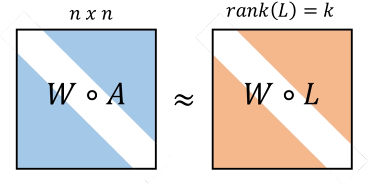
Low-Rank + Diagonal Approximation: Many matrices can be well approximated by a low-rank component plus a diagonal matrix.



Factor analysis (PCA with different noise variance in each dimension), kernel matrix approximation, source separation, ...

- Given L, optimal D is just $diag(A) - diag(L)$.
- $L^\star = \arg\min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2$ where W is zero on the diagonal and ones everywhere else.

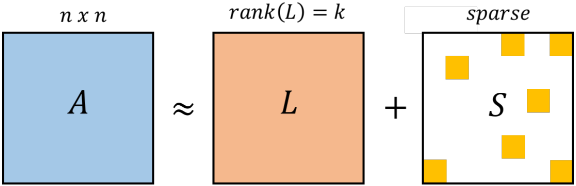
Low-Rank + Diagonal Approximation: Many matrices can be well approximated by a low-rank component plus a diagonal matrix.



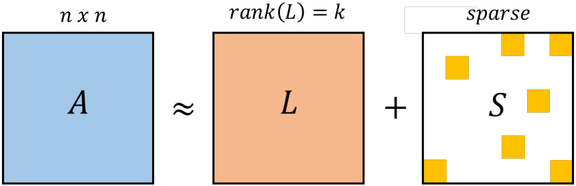
Factor analysis (PCA with different noise variance in each dimension), kernel matrix approximation, source separation, ...

- Given L, optimal D is just $diag(A) - diag(L)$.
- $L^\star = \arg\min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2$ where W is zero on the diagonal and ones everywhere else.

Low-Rank + Sparse Approximation/Robust PCA:



Low-Rank + Sparse Approximation/Robust PCA:



Significant attention since proposed by Candès et al. in 2009. Approximation of matrices with a few corrupted entries or non-low-rank components. E.g., background separation.

Low-Rank + Sparse Approximation/Robust PCA:

Significant attention since proposed by Candès et al. in 2009. Approximation of matrices with a few corrupted entries or non-low-rank components. E.g., background separation.

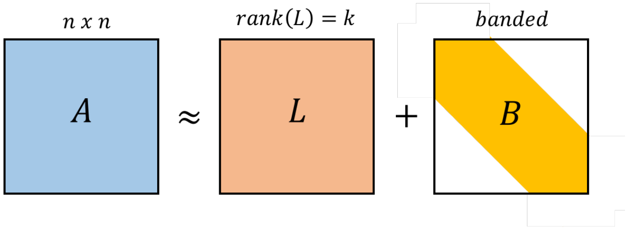
Low-Rank + Sparse Approximation/Robust PCA:

Significant attention since proposed by Candès et al. in 2009. Approximation of matrices with a few corrupted entries or non-low-rank components. E.g., background separation.

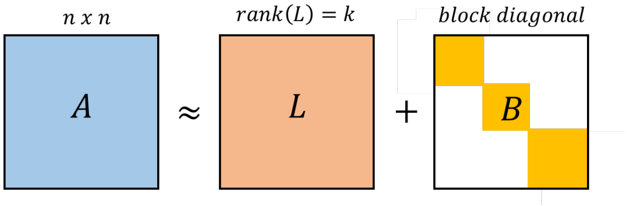
- $L^\star = \arg\min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2$ where W is zero at the corrupted locations. Optimal S is just $(1 - W) \circ (A - L^\star)$.
- Assume locations of corrupted entries are known.

Many other problems:

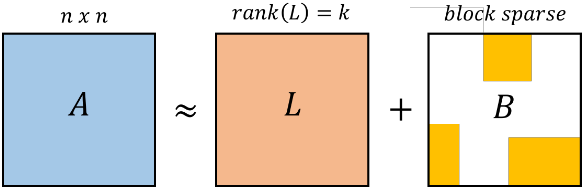
Many other problems: Low-rank plus banded.



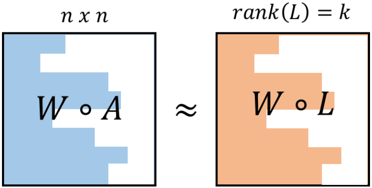
Many other problems: Low-rank plus block diagonal.



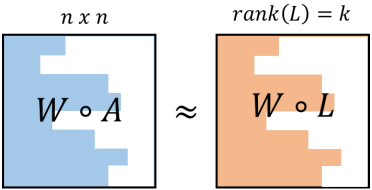
Many other problems: Low-rank plus block sparse.



Many other problems: Monotone missing data.



Many other problems: Monotone missing data.



Note: Masked low-rank approximation is closely related to matrix completion, but goal is different: approximate the unmasked entries, rather than recover the masked ones.

WHAT'S KNOWN

WHAT'S KNOWN

- NP-hard in general. For some special cases (e.g., low-rank plus diagonal) hardness is unresolved.

WHAT'S KNOWN

- NP-hard in general. For some special cases (e.g., low-rank plus diagonal) hardness is unresolved.
- Some can be solved provably via convex relaxation and alternating minimization, under various incoherence assumptions or random mask assumptions.
- See e.g., [Candès et al. 2009], [Chandrasekaran et al. 2011], [Netrapalli et al. 2014].

- NP-hard in general. For some special cases (e.g., low-rank plus diagonal) hardness is unresolved.
- Some can be solved provably via convex relaxation and alternating minimization, under various incoherence assumptions or random mask assumptions.
- See e.g., [Candès et al. 2009], [Chandrasekaran et al. 2011], [Netrapalli et al. 2014].
- Provable approximation algorithms for the more general weighted low-rank approximation problem are given in [Razenshteyn, Song, and Woodruff 2016] .
- Run in $\Omega(2^{\text{poly}(rk/\epsilon)} \cdot \text{poly}(n))$ time, where r is some measure of W's complexity (e.g., rank, number of distinct columns).

We give polynomial time bicriteria approximation algorithms.
Return L' with $rank(L') = k'$ satisfying:

$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon \|A\|_F^2.$$

OUR CONTRIBUTIONS

We give polynomial time bicriteria approximation algorithms.
Return L' with $rank(L') = k'$ satisfying:

$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2.$$

- $k' \geq k$ depends on the randomized communication complexity of $W \in \{0,1\}^{n \times n}$, when viewed as the two-player communication matrix of a Boolean function $f(a, b)$ with $a, b \in \{0,1\}^{\log n}$.

OUR CONTRIBUTIONS

We give polynomial time bicriteria approximation algorithms.
Return L' with $rank(L') = k'$ satisfying:

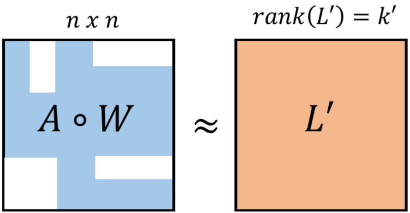
$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2.$$

- $k' \geq k$ depends on the randomized communication complexity of $W \in \{0,1\}^{n \times n}$, when viewed as the two-player communication matrix of a Boolean function $f(a, b)$ with $a, b \in \{0,1\}^{\log n}$.
- $k' = k \cdot \text{poly}(\log n/\epsilon)$ for all of the mentioned problems.

Our guarantees are achieved by a simple heuristic:

Our guarantees are achieved by a simple heuristic:

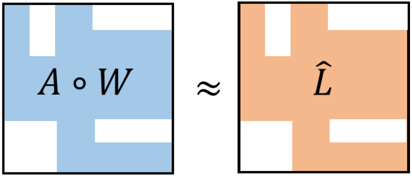
Set L' to the best rank-k' approximation of $A \circ W$.



Proof Sketch: Show there is a rank-k' approximation $\hat{L}$ that:

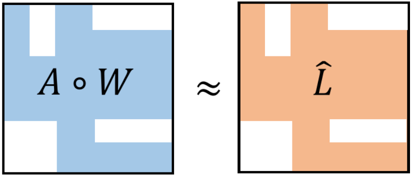
Proof Sketch: Show there is a rank-k' approximation $\hat{L}$ that:

1. Is exactly 0 wherever the mask W is zero.
2. Has $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2$.



Proof Sketch: Show there is a rank-k' approximation $\hat{L}$ that:

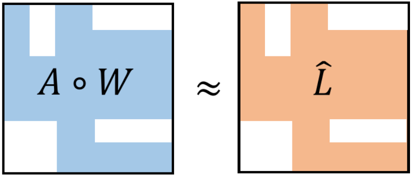
1. Is exactly 0 wherever the mask W is zero.
2. Has $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2$.



- $\hat{L}$ achieves 0 error in approximating the zeros in $A \circ W$.

Proof Sketch: Show there is a rank-k' approximation $\hat{L}$ that:

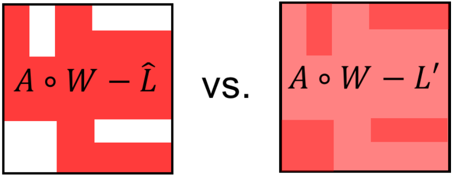
1. Is exactly 0 wherever the mask W is zero.
2. Has $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2$.



- $\hat{L}$ achieves 0 error in approximating the zeros in $A \circ W$.
- The best rank-k' approximation to $A \circ W$ (i.e., our output L') can only achieve worse error on these entries.

Proof Sketch: Show there is a rank-k' approximation $\hat{L}$ that:

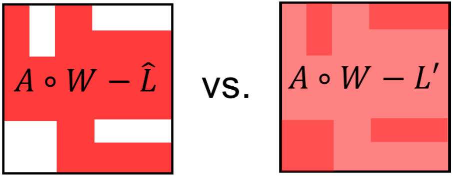
1. Is exactly 0 wherever the mask W is zero.
2. Has $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2$.



- $\hat{L}$ achieves 0 error in approximating the zeros in $A \circ W$.
- The best rank-k' approximation to $A \circ W$ (i.e., our output L') can only achieve worse error on these entries.

Proof Sketch: Show there is a rank-k' approximation $\hat{L}$ that:

1. Is exactly 0 wherever the mask W is zero.
2. Has $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2$.



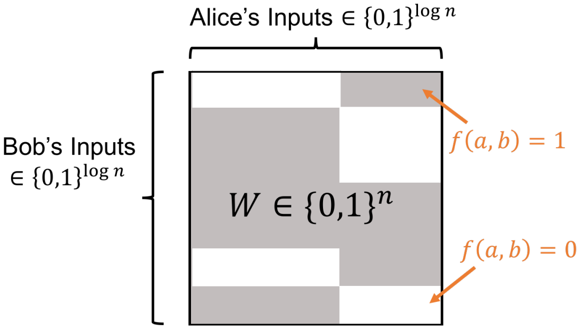
- $\hat{L}$ achieves 0 error in approximating the zeros in $A \circ W$.
- The best rank-k' approximation to $A \circ W$ (i.e., our output L') can only achieve worse error on these entries.
- So L' must achieve better error on the remaining entries. I.e.,

$$\|W \circ (A - L')\|_F^2 \leq \|W \circ (A - \hat{L})\|_F^2 \leq OPT + \epsilon\|A\|_F^2.$$

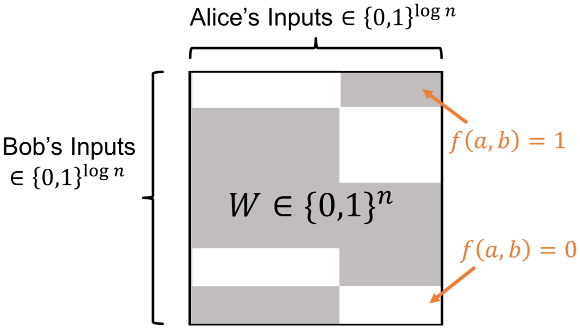
How do we construct a good low-rank approximation $\hat{L}$ that is 0 wherever W is 0?

How do we construct a good low-rank approximation $\hat{L}$ that is 0 wherever W is 0? Use a communication protocol for W!

How do we construct a good low-rank approximation $\hat{L}$ that is 0 wherever W is 0? Use a communication protocol for W!

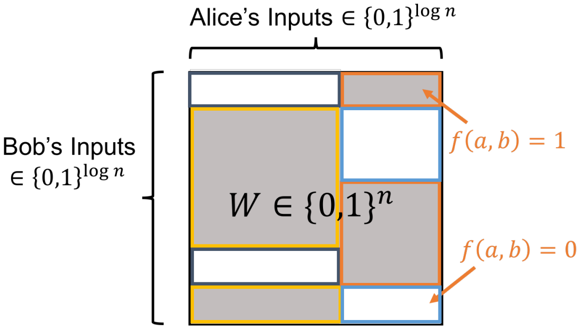


How do we construct a good low-rank approximation $\hat{L}$ that is 0 wherever W is 0? Use a communication protocol for W!



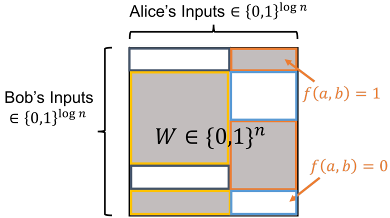
- Deterministic communication complexity $D(f)$ implies that W can be partitioned into $2^{D(f)}$ monochromatic combinatorial rectangles.

How do we construct a good low-rank approximation $\hat{L}$ that is 0 wherever W is 0? Use a communication protocol for W!

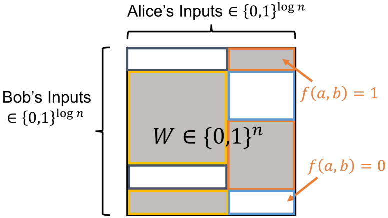


- Deterministic communication complexity $D(f)$ implies that W can be partitioned into $2^{D(f)}$ monochromatic combinatorial rectangles. Tight if log-rank conjecture holds.

Deterministic communication complexity $D(f) \implies W$ can be partitioned into $2^{D(f)}$ monochromatic rectangles.

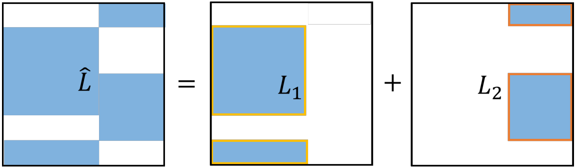


Deterministic communication complexity $D(f) \implies W$ can be partitioned into $2^{D(f)}$ monochromatic rectangles.



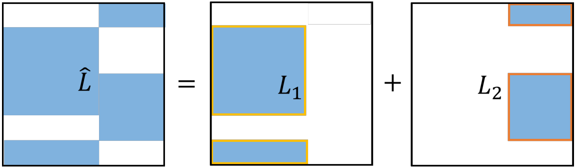
Let L_i be the optimal rank-k approximation on the i^{th} 1-rectangle and 0 elsewhere. Let $\hat{L} = \sum L_i$.

Deterministic communication complexity $D(f) \implies W$ can be partitioned into $2^{D(f)}$ monochromatic rectangles.



Let L_i be the optimal rank-k approximation on the i^{th} 1-rectangle and 0 elsewhere. Let $\hat{L} = \sum L_i$.

Deterministic communication complexity $D(f) \implies W$ can be partitioned into $2^{D(f)}$ monochromatic rectangles.



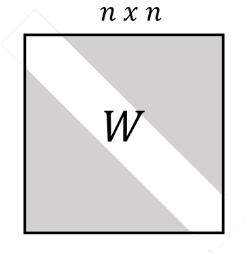
Let L_i be the optimal rank-k approximation on the i^{th} 1-rectangle and 0 elsewhere. Let $\hat{L} = \sum L_i$.

- $\hat{L}$ is exactly 0 wherever the mask W is zero.
- $\hat{L}$ has $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2$.
- $\hat{L}$ has rank $k' = \text{rank}\left(\sum L_i\right) \leq k \cdot 2^{D(f)}$.

Upshot: $L' = \arg\min_{\{L:\text{rank}(L)=k\}} \|A \circ W - L\|_F^2$ achieves $\|W \circ (A - L')\|_F^2 \leq OPT$ with rank $k' = k \cdot 2^{D(f)}$.

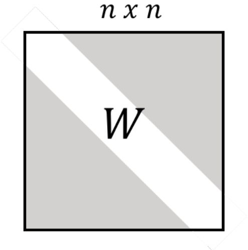
Upshot: $L' = \arg\min_{\{L:\text{rank}(L)=k\}} \|A \circ W - L\|_F^2$ achieves $\|W \circ (A - L')\|_F^2 \leq OPT$ with rank $k' = k \cdot 2^{D(f)}$.

- But, e.g., when W has zeros the diagonal, $D(f) = \log n$. Comm. complexity of (IN)EQUALITY on $\log n$ bit inputs.



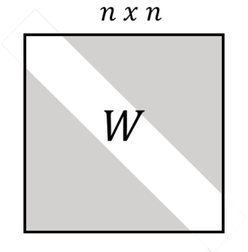
Upshot: $L' = \arg\min_{\{L:\text{rank}(L)=k\}} \|A \circ W - L\|_F^2$ achieves $\|W \circ (A - L')\|_F^2 \leq OPT$ with rank $k' = k \cdot 2^{D(f)}$.

- But, e.g., when W has zeros the diagonal, $D(f) = \log n$. Comm. complexity of (IN)EQUALITY on $\log n$ bit inputs.
- $k' = k \cdot 2^{\log n} = k \cdot n$, which is vacuous.



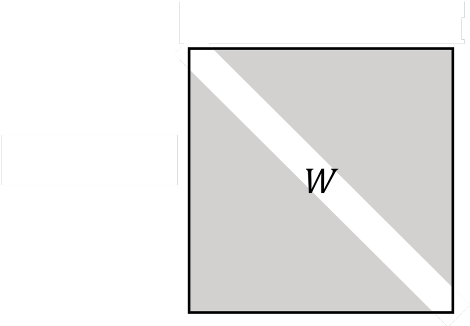
Upshot: $L' = \arg\min_{\{L:\text{rank}(L)=k\}} \|A \circ W - L\|_F^2$ achieves $\|W \circ (A - L')\|_F^2 \leq OPT$ with rank $k' = k \cdot 2^{D(f)}$.

- But, e.g., when W has zeros the diagonal, $D(f) = \log n$. Comm. complexity of (IN)EQUALITY on $\log n$ bit inputs.
- $k' = k \cdot 2^{\log n} = k \cdot n$, which is vacuous.



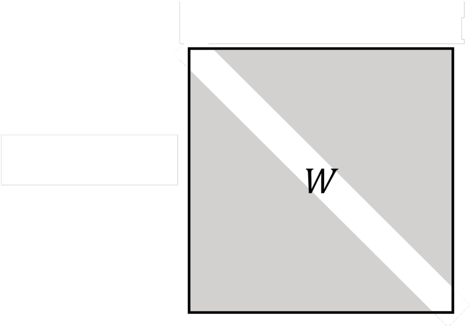
- Will get much better bounds with randomized communication complexity.

INEQUALITY admits a randomized communication protocol using $\log_2(1/\epsilon) + 5$ bits that outputs the incorrect answer on any positive instance (i.e., $a \neq b$) with probability $\leq \epsilon$.



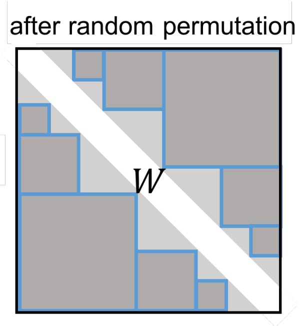
INEQUALITY admits a randomized communication protocol using $\log_2(1/\epsilon) + 5$ bits that outputs the incorrect answer on any positive instance (i.e., $a \neq b$) with probability $\leq \epsilon$.

- Randomly hash the inputs to $O(\log_2(1/\epsilon))$ bits then compare. Returns a false negative when two different inputs have the same hash.



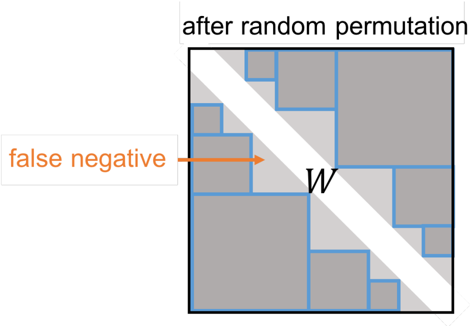
INEQUALITY admits a randomized communication protocol using $\log_2(1/\epsilon) + 5$ bits that outputs the incorrect answer on any positive instance (i.e., $a \neq b$) with probability $\leq \epsilon$.

- Randomly hash the inputs to $O(\log_2(1/\epsilon))$ bits then compare. Returns a false negative when two different inputs have the same hash.



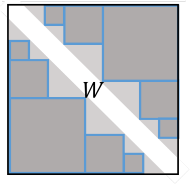
INEQUALITY admits a randomized communication protocol using $\log_2(1/\epsilon) + 5$ bits that outputs the incorrect answer on any positive instance (i.e., $a \neq b$) with probability $\leq \epsilon$.

- Randomly hash the inputs to $O(\log_2(1/\epsilon))$ bits then compare. Returns a false negative when two different inputs have the same hash.

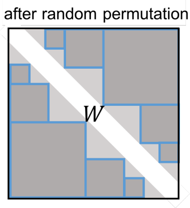


This randomized protocol partitions W into $2^{\log_2(1/\epsilon)+5} = \frac{32}{\epsilon}$ monochromatic rectangles.

after random permutation

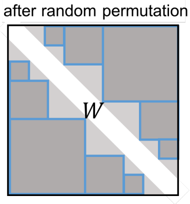


This randomized protocol partitions W into $2^{\log_2(1/\epsilon)+5} = \frac{32}{\epsilon}$ monochromatic rectangles.



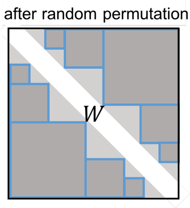
- In expectation, the 'uncovered entries' of $A \circ W$ have Frobenius norm $\epsilon \|A\|_F^2$.

This randomized protocol partitions W into $2^{\log_2(1/\epsilon)+5} = \frac{32}{\epsilon}$ monochromatic rectangles.



- In expectation, the 'uncovered entries' of $A \circ W$ have Frobenius norm $\epsilon\|A\|_F^2$.
- Pick any protocol achieving this expected error. Let $\hat{L}$ be sum of optimal rank-k approximations on each 1-rectangle.

This randomized protocol partitions W into $2^{\log_2(1/\epsilon)+5} = \frac{32}{\epsilon}$ monochromatic rectangles.



- In expectation, the 'uncovered entries' of $A \circ W$ have Frobenius norm $\epsilon \|A\|_F^2$.
- Pick any protocol achieving this expected error. Let $\hat{L}$ be sum of optimal rank-k approximations on each 1-rectangle.
  - $\|W \circ (A - \hat{L})\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon \|A\|_F^2$.
  - L has rank $k' = k \cdot 2^{\log_2(1/\epsilon)+5} = \frac{32k}{\epsilon}$.

Main Theorem: Given mask W with ϵ 1-sided error randomized communication $R_\epsilon(f)$, let $k' = k \cdot 2^{R_\epsilon(f)}$.
$L' = \arg\min_{\{L:\mathrm{rank}(L)=k'\}} \|A \circ W - L\|_F^2$ achieves:

$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\mathrm{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2.$$

Main Theorem: Given mask W with ϵ 1-sided error randomized communication $R_\epsilon(f)$, let $k' = k \cdot 2^{R_\epsilon(f)}$.
$L' = \arg\min_{\{L:\text{rank}(L)=k'\}} \|A \circ W - L\|_F^2$ achieves:

$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon \|A\|_F^2.$$

- $k' = O(k/\epsilon)$ for low-rank plus (block) diagonal (INEQUALITY)
- $k' = O(kt/\epsilon)$ for low-rank plus (block) t-sparse (variant of INEQUALITY)

Main Theorem: Given mask W with ϵ 1-sided error randomized communication $R_\epsilon(f)$, let $k' = k \cdot 2^{R_\epsilon(f)}$. $L' = \arg\min_{\{L:\text{rank}(L)=k'\}} \|A \circ W - L\|_F^2$ achieves:

$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon \|A\|_F^2.$$

- $k' = O(k/\epsilon)$ for low-rank plus (block) diagonal (INEQUALITY)
- $k' = O(kt/\epsilon)$ for low-rank plus (block) t-sparse (variant of INEQUALITY)
- Also give bounds that hold for 2-sided error protocols.

Main Theorem: Given mask W with ϵ 1-sided error randomized communication $R_\epsilon(f)$, let $k' = k \cdot 2^{R_\epsilon(f)}$.
$L' = \arg\min_{\{L:\text{rank}(L)=k'\}} \|A \circ W - L\|_F^2$ achieves:

$$\|W \circ (A - L')\|_F^2 \leq \min_{\{L:\text{rank}(L)=k\}} \|W \circ (A - L)\|_F^2 + \epsilon\|A\|_F^2.$$

- $k' = O(k/\epsilon)$ for low-rank plus (block) diagonal (INEQUALITY)
- $k' = O(kt/\epsilon)$ for low-rank plus (block) t-sparse (variant of INEQUALITY)
- Also give bounds that hold for 2-sided error protocols.
- $k' = k \cdot \text{poly}(\log n/\epsilon)$ for low-rank plus banded and monotone missing data (variants of GREATER-THAN).

Question 1: Is the connection between masked low-rank approximation and communication complexity tight?

Question 1: Is the connection between masked low-rank approximation and communication complexity tight?

Original goal: Show that no algorithm can give additive error $\epsilon \|A\|_F^2$ for $\epsilon = \Theta(1)$ with polynomial runtime and bicriteria rank $k' = 2^{o(R_\epsilon(f))}$.

Question 1: Is the connection between masked low-rank approximation and communication complexity tight?

Original goal: Show that no algorithm can give additive error $\epsilon\|A\|_F^2$ for $\epsilon = \Theta(1)$ with polynomial runtime and bicriteria rank $k' = 2^{o(R_\epsilon(f))}$.

Why it fails:

- For some instances, simple exact algorithms are known. E.g., $k = 1$ and $W = I$.
- Some matrices can be approximately partitioned into few monochromatic rectangles, even when their communication complexity is high. Recent refutation of the log approximate rank conjecture by Chattopadhyay, Mande, and Sherif.

Modified Question 1: Is the connection between masked low-rank approximation and communication complexity tight for at least a broad set of problems?

Modified Question 1: Is the connection between masked low-rank approximation and communication complexity tight for at least a broad set of problems?

- We show that when L is required to have a binary or non-negative factorization, no polynomial time algorithm can achieve rank $k' = 2^{o(D(f))}$ for a weight matrix based on a graph coloring problem.
- Holds in a regime ($k = 3$) where binary/non-negative low-rank approximation is polynomial time. Thus hardness comes from adding the mask.
- Assumes that we cannot n^{γ}-color a 3-colorable graph in polynomial time.

Question 2: Are there other interesting applications between communication complexity and linear algebraic problems?

We show:

- Multiparty communication complexity $\implies$ bicriteria masked tensor low-rank approximation.
- Nondeterministic communication complexity $\implies$ bicriteria masked Boolean low-rank approximation.
- 1-way communication complexity $\implies$ masked row-subset selection and tentatively some regression problems with missing/corrupted data.

Thanks! Questions?