

Rank Distribution of Sparse Random Linear Network Coding

Xiaolin LI, Wai Ho Mow, Fai Lung Tsang

Dept. of Electronic & Computer Engineering

Hong Kong University of Science and Technology

This work was supported by AoE Grant E-02/08 from HK UGC.

X. L. Li, HKUST

Outlines

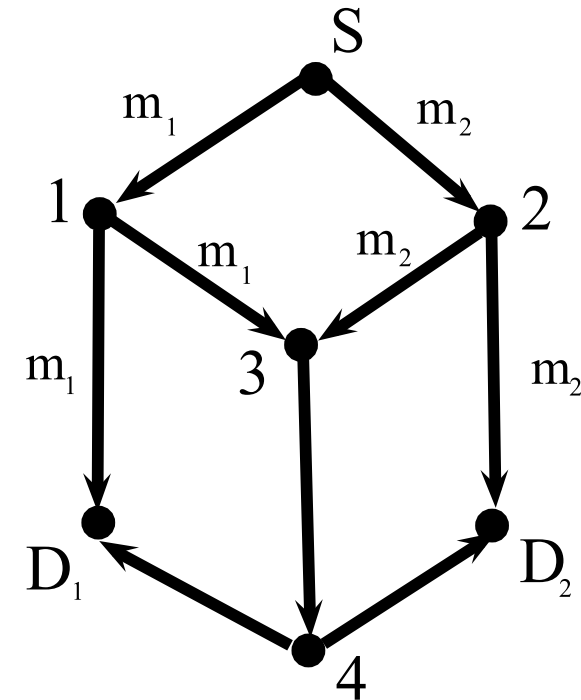
- Introduction and Problem Statement
- Zero Pattern and the Rank of Random Square Matrix
- Bounds for Rank Distribution
- Numerical Results
- Concluding Remarks

Outlines

- Introduction and Problem Statement
- Zero Pattern and the Rank of Random Square Matrix
- Bounds for Rank Distribution
- Numerical Results
- Concluding Remarks

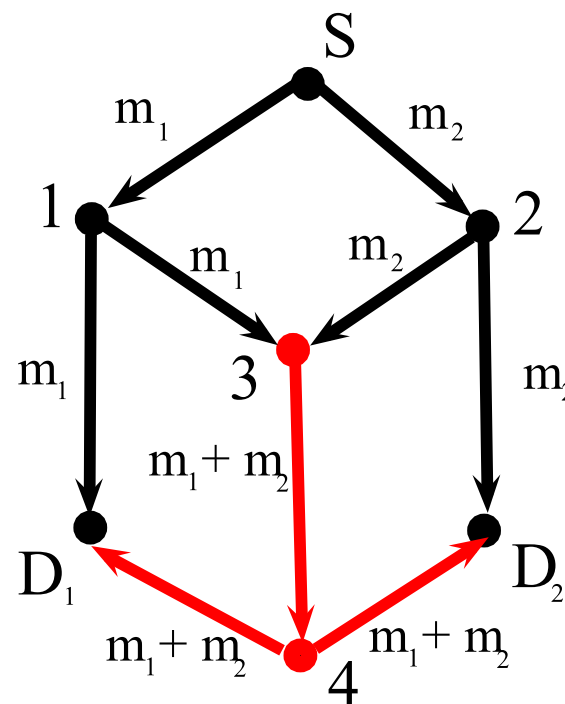
Network Coding

- Every edge has capacity 1.
- Source node S tends to send out m_1 and m_2 to D_1 and D_2 .



Network Coding

- Every edge has capacity 1.
- Source node S tends to send out m_1 and m_2 to D_1 and D_2 .
- Capacity of the network is increased by allowing intermediate nodes perform coding operations.

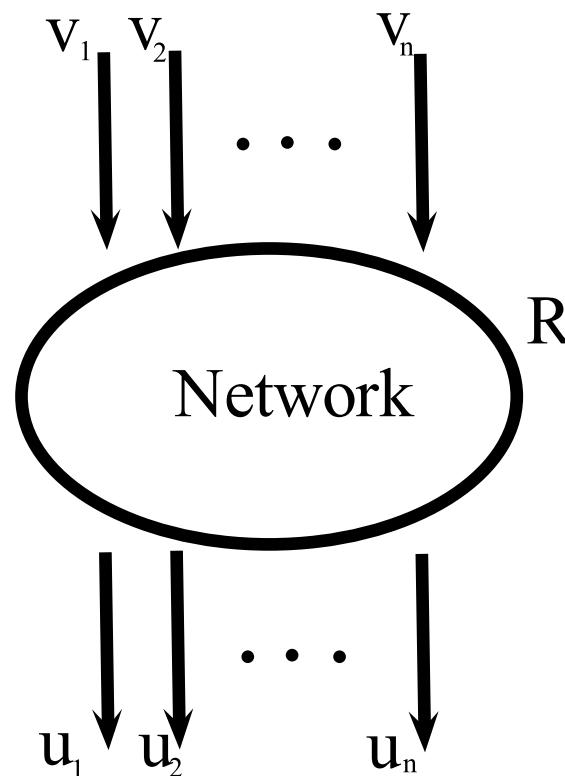


Noncoherent Network Coding [Kötter-Kschischang, IT-2008]

- Noncoherent: network topology is unknown.
- Codewords are represented by **vector spaces**.
-

$$[u_1, \dots, u_n]^T = \mathbf{R}[v_1, \dots, v_n]^T,$$

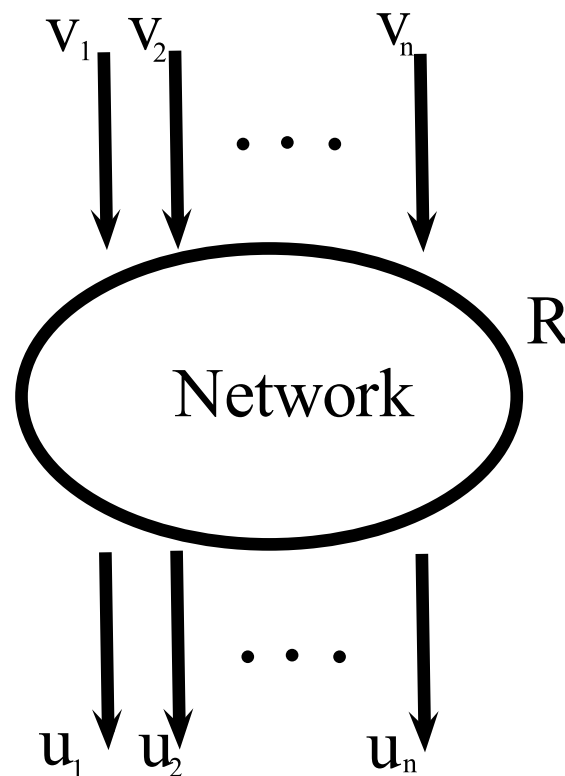
$\mathbf{R} = (r_{ij})$: $n \times n$ random matrix over \mathbb{F}_q .



Noncoherent Network Coding [Kötter-Kschischang, IT-2008]

- Noncoherent: network topology is unknown.
- Codewords are represented by **vector spaces**.
- $$[u_1, \dots, u_n]^T = \mathbf{R}[v_1, \dots, v_n]^T,$$

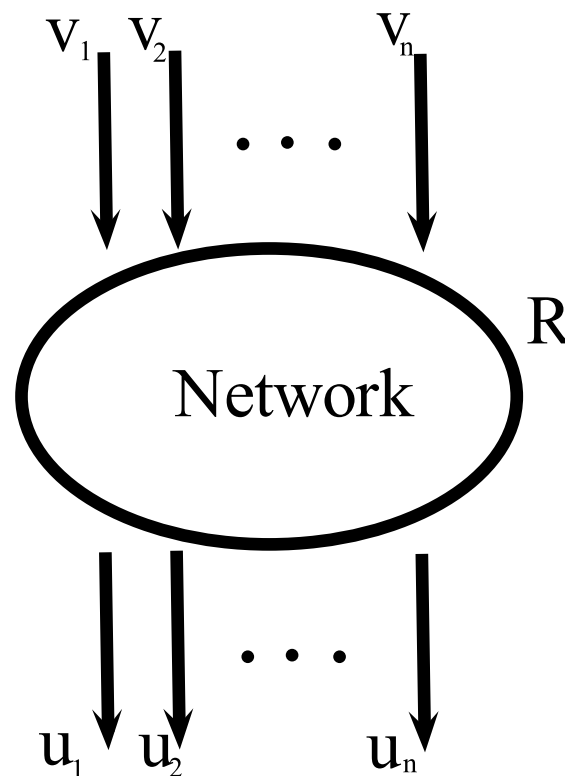
$$\mathbf{R} = (r_{ij}): n \times n \text{ random matrix over } \mathbb{F}_q.$$
- **Any problems?**



Noncoherent Network Coding [Kötter-Kschischang, IT-2008]

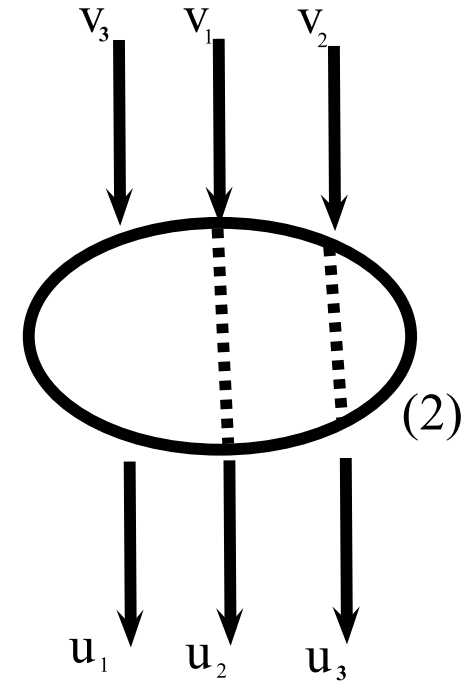
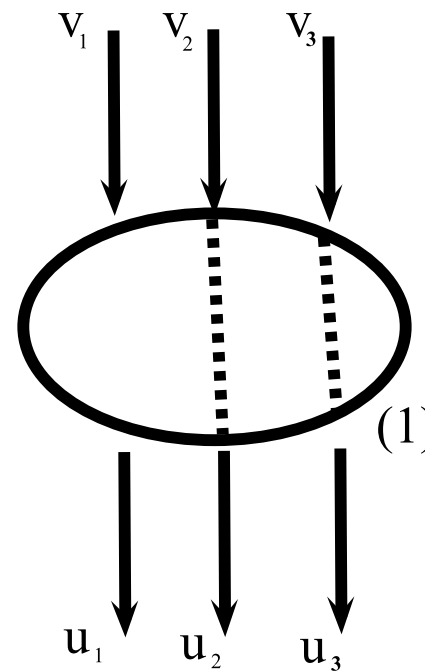
- Noncoherent: network topology is unknown.
- Codewords are represented by **vector spaces**.
- $$[u_1, \dots, u_n]^T = \mathbf{R}[v_1, \dots, v_n]^T,$$

$$\mathbf{R} = (r_{ij}): n \times n \text{ random matrix over } \mathbb{F}_q.$$
- **Any problems?**
- **What if \mathbf{R} is singular?**



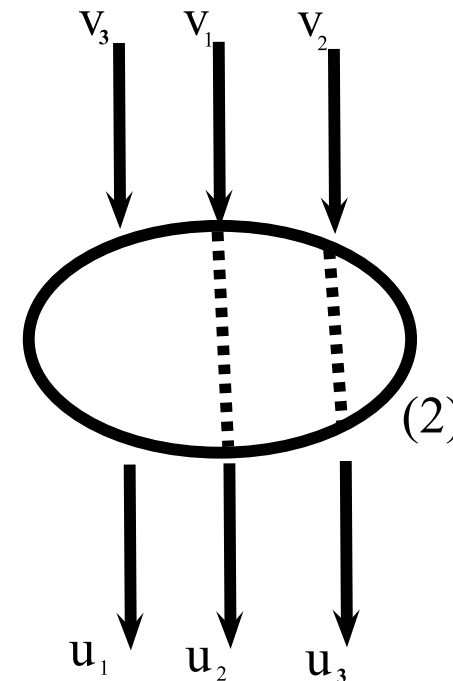
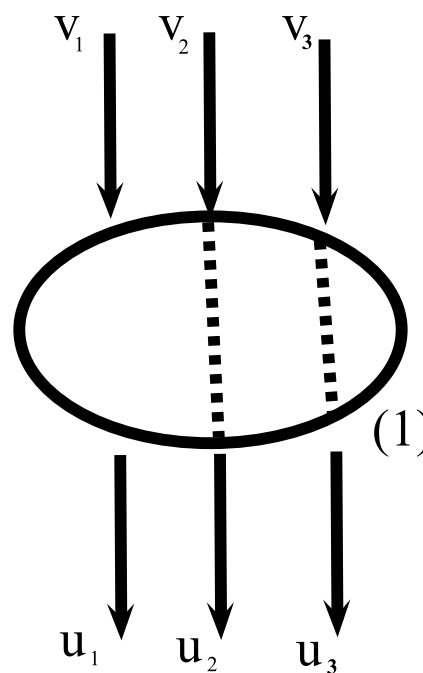
Singular transfer matrix

- In both cases, the same vector space is injected.
- In case (1),
 $\text{span}\{u_1, u_2, u_3\} = \text{span}\{v_2, v_3\}$;
- In case (2),
 $\text{span}\{u_1, u_2, u_3\} = \text{span}\{v_1, v_2\}$;



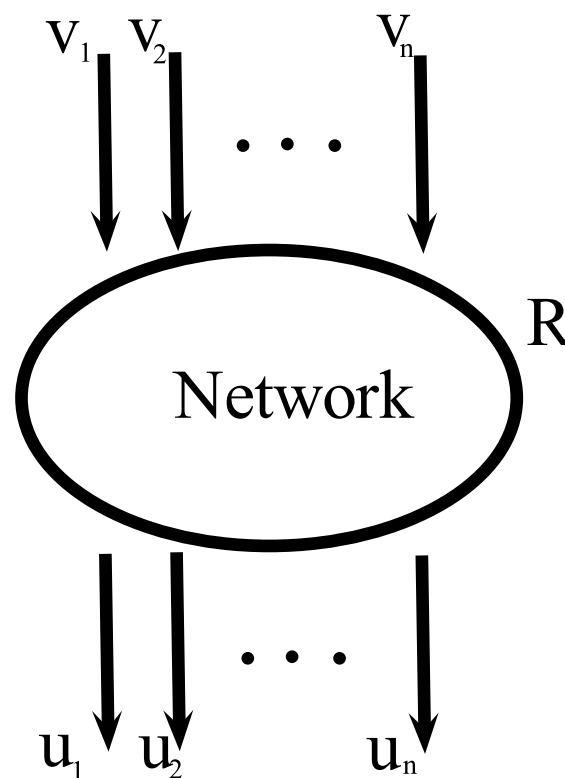
Singular transfer matrix

- In both cases, the same vector space is injected.
- In case (1),
 $\text{span}\{u_1, u_2, u_3\} = \text{span}\{v_2, v_3\}$;
- In case (2),
 $\text{span}\{u_1, u_2, u_3\} = \text{span}\{v_1, v_2\}$;
- **Conclusion:** output vector spaces are different even for the same input space.



Constant Dimension Codes

- Let \mathcal{V} be an N -dimensional vector space over \mathbb{F}_q , $\mathcal{P}(\mathcal{V})$ denote the set of all possible subspaces of \mathcal{V} .
- Code $\mathcal{W} = \{W_1, \dots, W_l\} \subseteq \mathcal{P}(\mathcal{V})$;
- If $\forall W_i, 1 \leq i \leq l, \dim(W_i) = n$, then \mathcal{W} is an n -dimensional **Constant Dimension Code**.



Constant Dimension Codes

- Subspace Metric:

$$d(U, V) \triangleq \dim(U \cup V) - \dim(U \cap V);$$

(Defined in [Kötter-Kschischang, IT-2008])

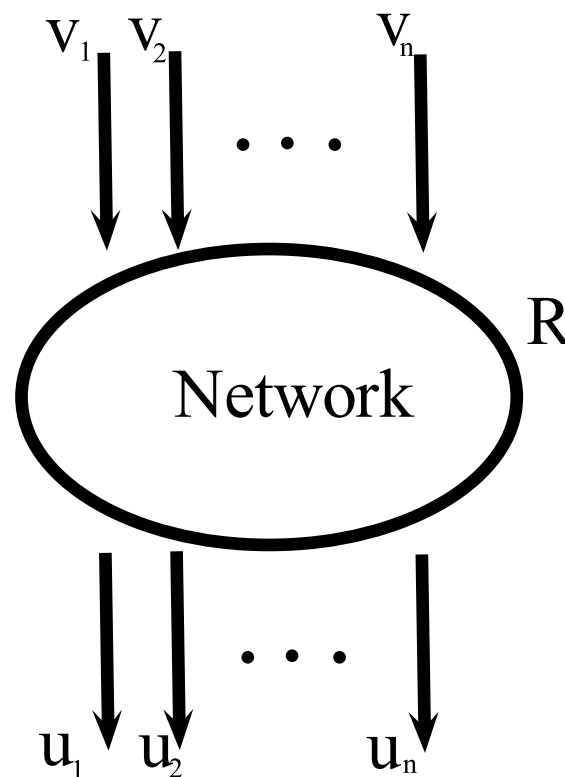
- Let $W_i = \text{span}\{v_1, \dots, v_n\}$, since

$$[u_1, \dots, u_n]^T = \mathbf{R}[v_1, \dots, v_n]^T,$$

Hence

$$d(U, V) = n - \text{rank}(\mathbf{R}).$$

- **Remark:** For Constant Dimension Code, the subspace distance is totally determined by the rank of the transfer matrix.



Constant Dimension Codes

- The probability of decoding failure is

$$P_e(W_i \in \mathcal{W}) = P\left(n - \text{rank}(\mathbf{R}) > \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor\right) = P\left(\text{rank}(\mathbf{R}) \leq n - \frac{d_{\min}}{2}\right)$$

(d_{\min} is always an even integer for Constant Dimension Codes.)

- **Conclusion:** The analysis for the **probability of decoding failure** can be turned into the **rank analysis** of the random transfer matrix.

Sparse Network Coding

- Implement the random matrix with sparse transfer matrix.
- Greatly reduce the implementation complexity.
- Increase the probability of decoding failure.

Sparse Network Coding

- Implement the random matrix with sparse transfer matrix.
- Greatly reduce the implementation complexity.
- Increase the probability of decoding failure.

Theoretical support: **rank analysis results** for large non-uniform random matrices over finite fields [Blömer-Karp-Welzl, RSA-1997], [Cooper, RSA-2000], [Xiaolin-Waiho-Failung, ICC-2011].

Problem Statement

Assumptions:

- The probability distribution of all elements in the matrix are mutually independent;
- The probability distribution of each element satisfies the following:

$$P(r_{ij} = k) = \begin{cases} p, & k = 0; \\ p_k(q), & \forall k \in \mathbb{F}_q \setminus \{0\}. \end{cases} \quad (1)$$

such that $\lim_{q \rightarrow \infty} p_k(q) = 0$.

Problem Statement:

For $\mathbf{R} \in \text{Matr}(n, \mathbb{F}_q)$, calculate the probability $P(\text{rank}(\mathbf{R}) \leq n - s)$.

Comparisons with previous work

All of the following assume the **special case** $p_k(q) = \frac{1-p}{q-1}$,

- In [Blömer-Karp-Welzl, RSA-1997], if $1 - p = \frac{\log n - c}{n}$, and $c > 0$,

$$E(\text{rank}(\mathbf{R})) = n - O(1)$$

- In [Cooper, RSA-2000], if $1 - p \geq \frac{\log n + d}{n}$, as $d(n) \geq -\log(\log n / 9q)$ tends to $+\infty$,

$$P(\text{rank}(\mathbf{R}) < n) \sim \prod_{i=1}^{\infty} (1 - q^{-i})$$

- In [Xiaolin-Waiho-Failung, ICC-2011], for large enough q , **any** $n \in \mathbb{Z}^+$, $0 \leq p \leq 1$, upper and lower bounds on $P(\text{rank}(\mathbf{R}) < n)$.

Comparisons with previous work

This paper generalizes [Xiaolin-Waiho-Failung, ICC-2011] in the following two aspects:

- The distribution of the elements over the finite field is more general;
- Bounds on the rank distribution $P(\text{rank}(\mathbf{R}) \leq n - s)$ is given (Previous work only deals with $s = 1$).

Outlines

- Introduction and Problem Statement
- Zero Pattern and the Rank of Random Square Matrix
- Bounds for Rank Distribution
- Numerical Results
- Concluding Remarks

Zero Pattern of the Random Matrix

Some simple examples:

*	*	0
0	*	0
*	0	*

Proper Zero Pattern

*	*	0
0	0	*
0	0	*

Improper Zero Pattern

Define: $\mathcal{C}_n \triangleq$ the set of $n \times n$ *proper zero pattern*;
 If $C \notin \mathcal{C}_n$, the pattern is called *improper zero pattern*.

Zero Pattern of the Random Matrix

Relationship between zero pattern and the rank of random matrix:

Theorem 1: Define $\mathcal{C}_n = \{pat(\mathbf{R}) \mid \exists a_k \in N_k(\mathbf{R}), \forall k \in \{1, 2, \dots, n\} \text{ such that } \{a_1, a_2, \dots, a_n\} = \{1, 2, \dots, n\}\}$. Then for any $C_0 \in \mathcal{C}_n$,

$$\lim_{q \rightarrow \infty} P(\det(\mathbf{R}) \neq 0 \mid pat(\mathbf{R}) = C_0) = 1 .$$

Implication: as the field size tends to ∞ , the zero pattern will determine the rank of the random matrix with high probability (w.h.p.)

Zero Pattern of the Random Matrix

Theorem 1 is a direct consequence of the following lemma:

Lemma 1: If $\forall 1 \leq s \leq n$, $a_s, b_s \in \mathbb{F}_q$ are random variables, among which $a_s, \forall 1 \leq s \leq n$ follows the distribution of (1) and are independent, then

$$\lim_{q \rightarrow \infty} P\left\{ \sum_{1 \leq s \leq n} a_s b_s = 0 \mid a_s, b_s \in \mathbb{F}_q^*, \forall s \right\} = 0.$$

Note in the above lemma, the joint distribution of (b_1, b_2, \dots, b_n) is not specified, and they are not necessarily independent.

Proved by mathematical induction.

Zero Pattern of the Random Matrix

Back to the previous examples:

*	*	0
0	*	0
*	0	*

Proper Zero Pattern
(Full Rank)

*	*	0
0	0	*
0	0	*

Improper Zero Pattern
(Singular Rank)

Rank and Subpattern

- The corresponding zero pattern of the submatrix is called **subpattern**.
- **Simple Observation:**
The rank of the matrix is $n - s$. \Leftrightarrow The maximum size of the subpattern of the matrix is $(n - s) \times (n - s)$.

Example:

- The rank of the matrix for large q is 2.
- The maximum size of the subpattern is 2×2 .

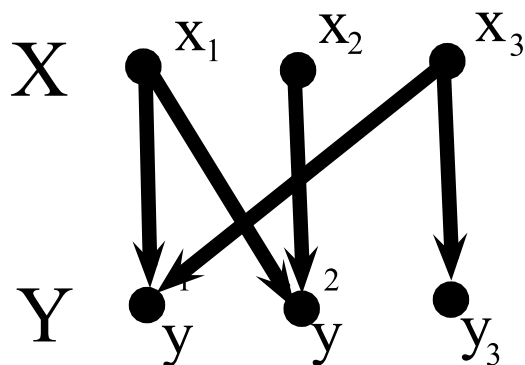
*	*	0
0	0	*
0	0	*

Bipartite Graph Representation

$\mathbf{R} \rightarrow G[X, Y]: r_{ij} \neq 0 \Leftrightarrow \exists$ connection between x_i and y_j .

*	*	0
0	*	0
*	0	*

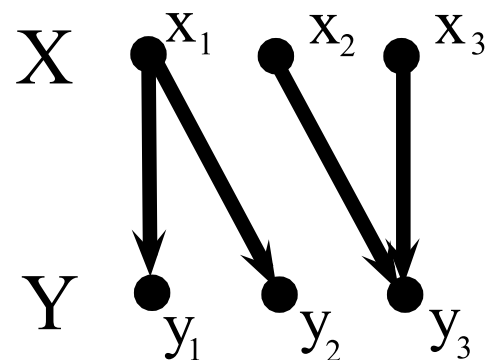
Proper Pattern



Perfect Matching

*	*	0
0	0	*
0	0	*

Improper Pattern



No perfect Matching

Simple observation: $pat(\mathbf{R}) \in \mathcal{C}_n \Leftrightarrow X \text{ \& } Y \text{ has a perfect matching.}$

Zero Rectangle

Definition: An $m \times l$ *zero rectangle* is an $m \times l$ all-0 submatrix after row and column permutations. If $m + l = n + s$, then we say the it is a *zero rectangle of type s* .

Some simple examples:

0	0	0
0	*	0
*	*	0

0	*	0
*	0	*
0	*	0

Maximal Subpattern and Zero Rectangle

$\forall A \subseteq X$, denote $K(A) \triangleq \{y \in Y \mid y \text{ is a neighbour of any } x \text{ in } A\}$.

Lemma 2 [Hall's Theorem]: $G[X, Y]$ with $|X| = |Y|$ has a perfect matching from X to Y if and only if $|A| \leq |K(A)|$ for any $A \subseteq X$.

Proper Pattern and Zero Rectangle

$\forall A \subseteq X$, denote $K(A) \triangleq \{y \in Y \mid y \text{ is a neighbour of any } x \text{ in } A\}$.

Lemma 2 [Hall's Theorem]: $G[X, Y]$ with $|X| = |Y|$ has a perfect matching from X to Y if and only if $|A| \leq |K(A)|$ for any $A \subseteq X$.

Lemma 3: If $G[X, Y]$ is a bipartite graph and $|X| = |Y| = n$, then for any given integer constant $0 \leq s \leq n$, the following conditions are equivalent:

- (a) $\forall A \subseteq X, |A| \leq |K(A)| + s$;
- (b) $\exists X' \subseteq X \& Y' \subseteq Y$ such that $|X'| = |Y'| = n - s$ and there exists a perfect matching between X' and Y' .

Proper Pattern and Zero Rectangle

The following lemma reveals the relationship between zero pattern and zero rectangle.

Lemma 4: The maximal proper subpattern for the random matrix \mathbf{R} is of size $(n - s) \times (n - s)$, if and only if the largest zero rectangle is of type s .

(Reminder: zero rectangle is of type s . \Leftrightarrow Zero rectangle is of size $m \times l$ with $m + l = n + s$.)

Theorem 2: As the finite field size q tends to $+\infty$, $\text{rank}(\mathbf{R}) = (n - s)$ if and only if the largest zero rectangle is of type s .

Outlines

- Introduction and Problem Statement
- Zero Pattern and the Rank of Random Square Matrix
- Bounds for Rank Distribution
- Numerical Results
- Concluding Remarks

Bounds for Rank Distribution

Let \mathcal{M} be the set containing all $n \times n$ zero patterns with the size of maximal zero rectangle of type s' with $s' \geq s$.

Define

$$M(L) \triangleq |\{ pat(\mathbf{R}) \in \mathcal{M} \mid pat(\mathbf{R}) \text{ has exactly } L \text{ 0's} \}|$$

Graphical Interpretation of $M(L)$: Number of bipartite graphs $G[X, Y]$ with $(n^2 - L)$ edges and the size of maximal matching is at most $(n - s)$.

It follows from Theorem 2 that

$$\lim_{q \rightarrow \infty} P(rank(\mathbf{R}) \leq n - s) = P(\mathcal{M}).$$

Bounds for Rank Distribution

In the following we provide bounds for $M(L)$.

Upper Bounds

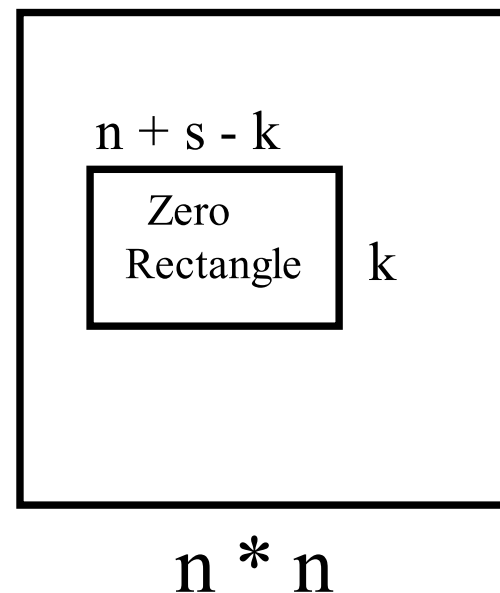
(1) *Counting zero rectangles*: If the pattern belongs to $M(L)$, it must contain at least one zero rectangle of size $k \times (n + s - k)$.

a. For $0 \leq L \leq ns$, $M_{u_1}(L) = 0$.

b. For $ns \leq L < n^2 - n + s$

$$M_{u_1}(L) = \sum_{s \leq k \leq n \& k(n+s-k) \leq L} \binom{n}{k} \binom{n}{n+s-k} \binom{n^2 - k(n+s-k)}{L - k(n+s-k)}$$

c. For $n^2 - n + s < L \leq n^2$, $M_{u_1}(L) = \binom{n^2}{L}$.



Bounds for Rank Distribution

(2) *Counting proper zero patterns*: Under-count the number of proper zero patterns.

a. For $0 \leq L \leq ns$, $M_{u_2}(L) = 0$.

b. For $ns < L < n^2 - n + s$

Denote $r = n - s + t$, where $1 \leq t \leq s$,

$M_{u_2}(L) = [\binom{n^2}{L} - \sum_{1 \leq t \leq s} N(t)]$, where $N(t)$ is given by:

(i) If $(r^2 - r)/2 + r(n - r) \geq n^2 - L - r > 0$, then

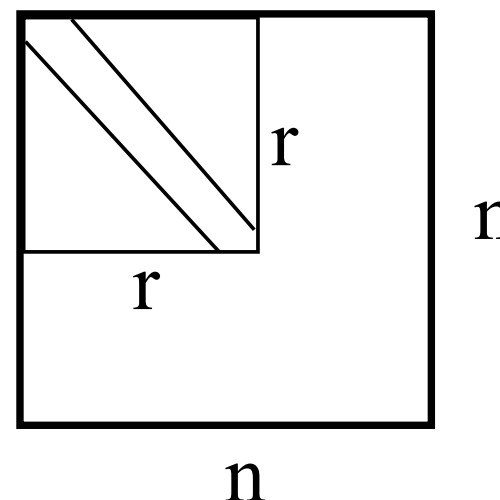
$$N(t) = 2 * r! \binom{n}{r}^2 \binom{(r^2 - r)/2 + r(n - r)}{n^2 - L - r};$$

(ii) If $(r^2 - r)/2 + r(n - r) \geq n^2 - L - r = 0$, then

$$N(t) = r! \binom{n}{r}^2;$$

(iii) If $(r^2 - r)/2 + r(n - r) < n^2 - L - r$, then $N(t) = 0$.

c. For $n^2 - n + s < L \leq n^2$, $M_{u_1}(L) = \binom{n^2}{L}$.



Bounds for Rank Distribution

Lower Bounds

(1) *Counting zero rectangles*: A bound can be obtained by deducting all those parts that had been over-counted by in part (1) of the upper bound.

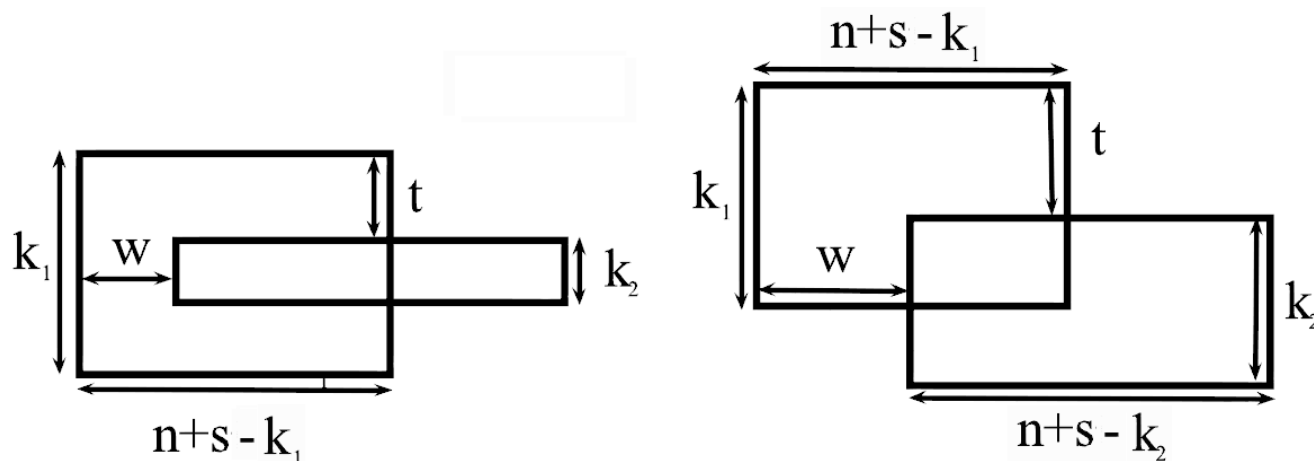


Figure 1: Combination of two zero rectangles.

- a. For $0 \leq L \leq n - 1$, $M_{l_1}(L) = M_{u_1}(L) = 0$.
- b. For $n \leq L \leq n^2 - n$, $M_{l_1}(L) = M_{u_1}(L) - M_{d_1}(L) - M_{d_2}(L)$.
- c. For $n^2 - n < L \leq n^2$, $M_{l_1}(L) = \binom{n^2}{L}$.

where $M_{d_1}(L)$ and $M_{d_2}(L)$ are specified as the following:

$$M_{d_1}(L) = \sum_{w, k_1, k_2} \binom{n}{k_2} \binom{n - k_2}{k_1 - k_2} \binom{n}{n + s - k_1} \binom{k_1 - s}{k_1 - k_2 + w} \binom{n^2 - T_1}{L - T_1}$$

where the sum is computed over all values of k_1, k_2, w satisfying:

$$\begin{cases} n \geq k_1 \geq k_2 \geq s ; \\ 0 \leq w \leq \min\{k_2 - s, n + s - k_1\} ; \\ \text{If } k_1 = k_2, \text{ then } w \neq 0 ; \\ T_1 = k_2(n + s - k_2) + (k_1 - k_2)(n + s - k_1) + k_2w \leq L . \end{cases}$$

$$M_{d2}(L) = \sum_{w,t,k_1,k_2} \binom{n}{t} \binom{n-t}{k_1-t} \binom{n-k_1}{k_2-k_1+t} \binom{n}{w} \\ \binom{n-w}{n+s-k_1-w} \binom{k_1-s}{k_1-k_2+w} \binom{n^2-T_2}{L-T_2}$$

where the sum is computed over all values of k_1, k_2, t, w satisfying:

$$\left\{ \begin{array}{l} n \geq k_1 \geq k_2 \geq s ; \\ k_1 - k_2 < t \leq \min\{k_1, n - k_2\} ; \\ 0 \leq w \leq \min\{k_2 - s, n + s - k_1\} ; \\ T_2 = k_2(n + s - k_2) + t(n + s - k_1) + k_2w \leq L . \end{array} \right.$$

Bounds for Rank Distribution

(2) *Counting proper zero patterns*: Over-count the number of proper zero patterns.

a. For $0 \leq L \leq ns$, $M_{l_2}(L) = 0$.

b. For $ns < L < n^2 - n + s$

Denote $r = n - s + t$. Then

$M_{l_2}(L) = \left[\binom{n^2}{L} - \sum_{1 \leq t \leq s} N(t) \right]$, where $N(t)$ is defined below.

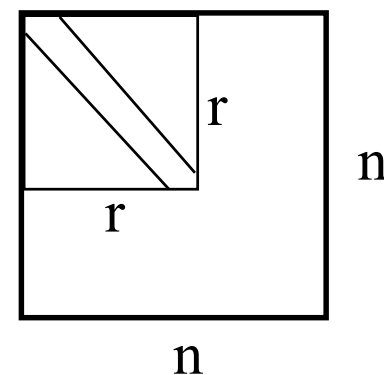
(i) If $nr - r \geq n^2 - L - r > 0$, then

$$N(t) = 2 * r! \binom{n}{r}^2 \binom{nr-r}{n^2-r-L};$$

(ii) If $nr - r \geq n^2 - L - r = 0$, then $N(t) = r! \binom{n}{r}^2$;

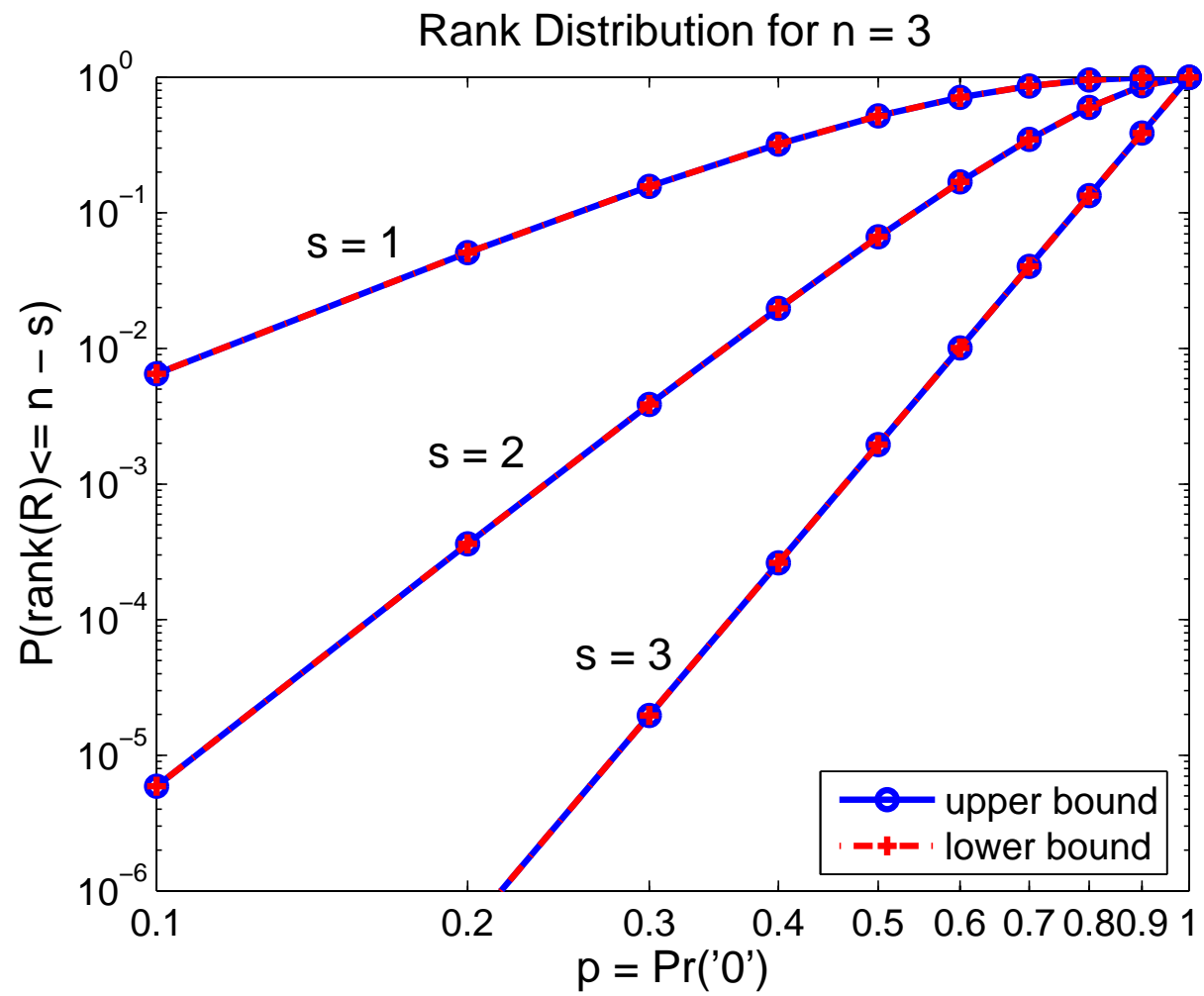
(iii) If $nr - r < n^2 - L - r$, then $N(t) = 0$.

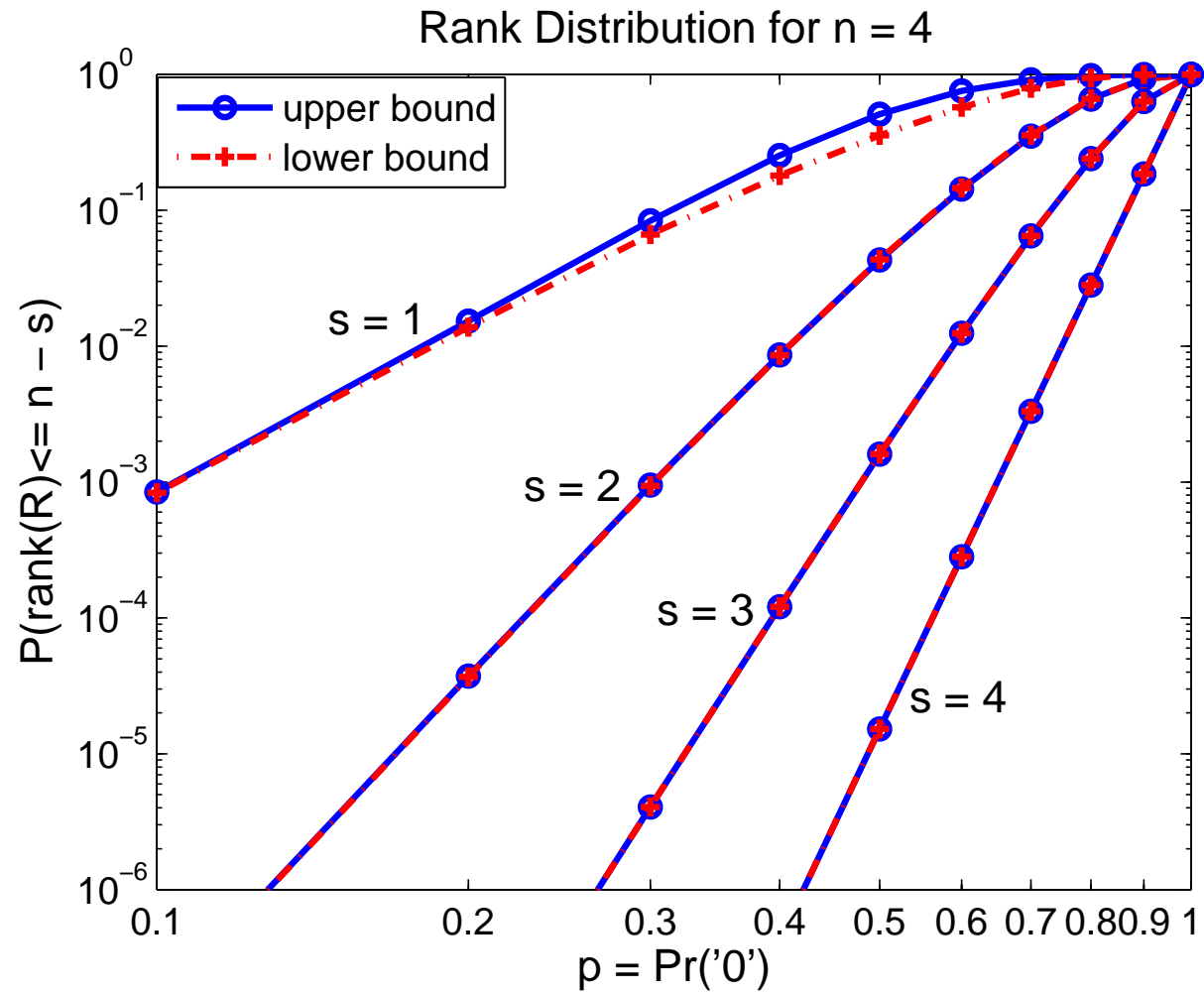
c. For $n^2 - n + s < L \leq n^2$, $M_{l_2}(L) = \binom{n^2}{L}$.

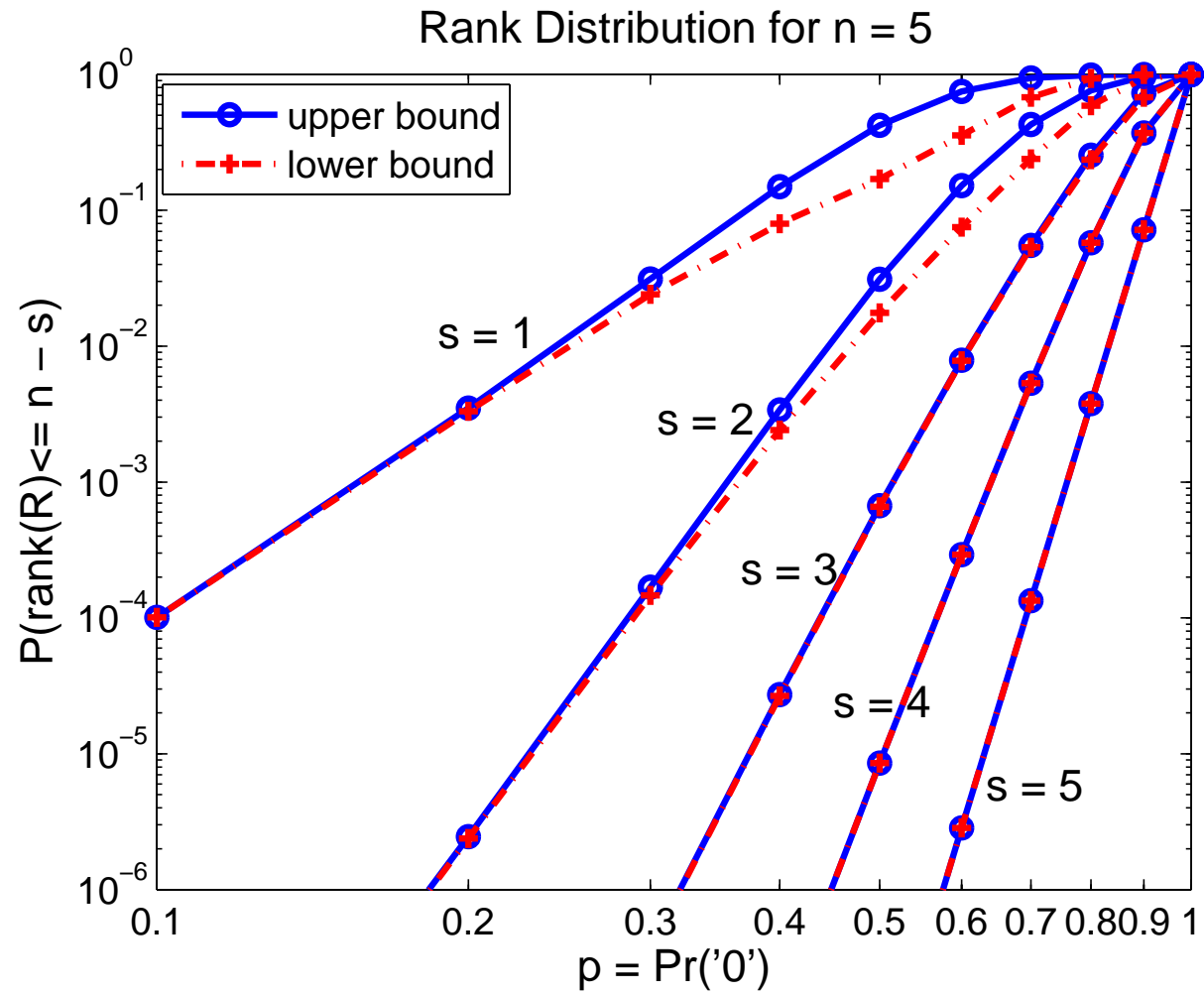


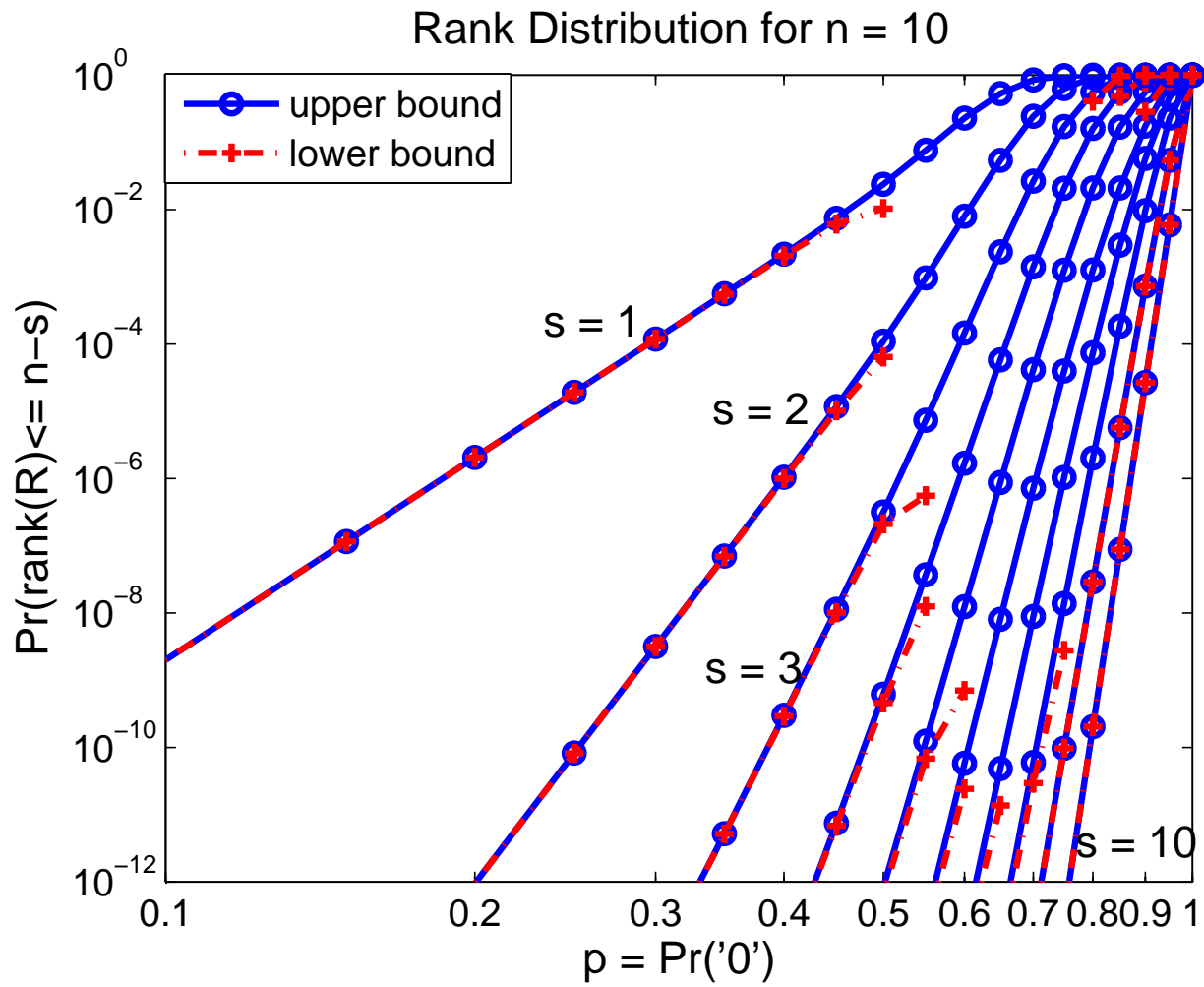
Outlines

- Introduction and Problem Statement
- Zero Pattern and the Rank of Random Square Matrix
- Bounds for Rank Distribution
- Numerical Results
- Concluding Remarks









Observations from Numerical Results

- The bounds are exact for $n = 3$ (and also for $n = 2$);
- For a given n , the maximum value of p , such that the difference between the upper and lower bounds is sufficiently small, is increasing as s increases;
- Even though in some regions the upper bounds cannot be guaranteed to be tight, they can still be used to guarantee the decoding failure probability.

Outlines

- Introduction and Problem Statement
- Zero Pattern and the Rank of Random Square Matrix
- Bounds for Rank Distribution
- Numerical Results
- Concluding Remarks

Concluding Remarks

- The problem of estimating the **decoding failure probability** for constant dimension codes arising from **sparse random linear network coding** has been addressed;
- Useful upper and lower bounds on the **rank distribution** for random transfer matrix have been derived and shown to be tight over a wide range of p ;
- The concept of **zero pattern** has been introduced to turn the problem into a combinatorial problem, which is of theoretical interest on its own.

References

- [1] [Blömer-Karp-Welzl, RSA-1997]J. Blömer, R. Karp and E. Welzl, "The Rank of Sparse Random Matrices Over Finite Fields," *Random Structures Algorithms*, vol. 10, 1997, p. 407–419.
- [2] [Cooper, RSA-2000]C. Cooper, "On the Distribution of Rank of a Random Matrix over a Finite Field," *Random Structures Algorithms*, vol. 36, 2000, pp. 197–212.
- [3] [Kötter-Kschischang, IT-2008]R. Kötter and F. R. Kschischang, "Coding for Errors and Erasures in Random Network Coding," *IEEE Trans. Inform. Theory*, vol. 54, 2008, pp. 3549–3591.
- [4] [Silva-Kschischang-Kötter, IT-2010]D. Silva, F. R. Kschischang, and R. Kötter, "Communication over Finite-Field Matrix Channels," *IEEE Trans. Inform. Theory*, vol. 56, March 2010, pp. 1296–1305.
- [5] [Xiaolin-Waiho-Failung, ICC-2011]X.L. Li, W.H. Mow, and F.L. Tsang, "Singularity Probability Analysis for Sparse Random Linear Network Coding," *Proc. ICC'11*, Jun 5–9, 2011, Kyoto, Japan.