CiE 2010

Combinatorial Rectangles in Communication Complexity

(joint work)

University of Porto, Faculty of Sciencies, Department of Computer Science

Outline

Preliminaries

NP-completness of the 1-mcr problem

The largest monochromatic rectangles

Random “functions”: communication complexity lower bound

Complexity

Definition

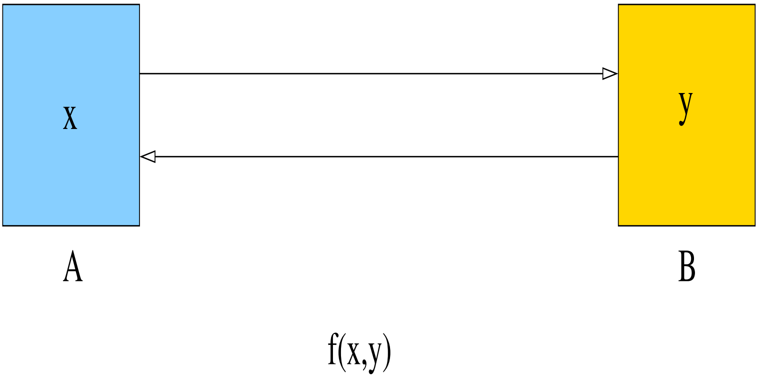
A decision problem A reduces polynomially to a decision problem B, $A \leq_p B$, if there is a function $f : \Sigma_A^\star \to \Sigma_B^\star$ (where Σ_A and Σ_B are the alphabets, $A \subseteq \Sigma_A^\star$, $B \subseteq \Sigma_B^\star$) computable in polynomial time such that, for every $x \in \Sigma^\star$, we have

$$x \in A \text{ if and only if } f(x) \in B$$

Definition

A decision problem A is NP - complete if $A \in$ NP and $B \leq_p A$ for every problem $B \in$ NP.

Communication complexity



Communication complexity

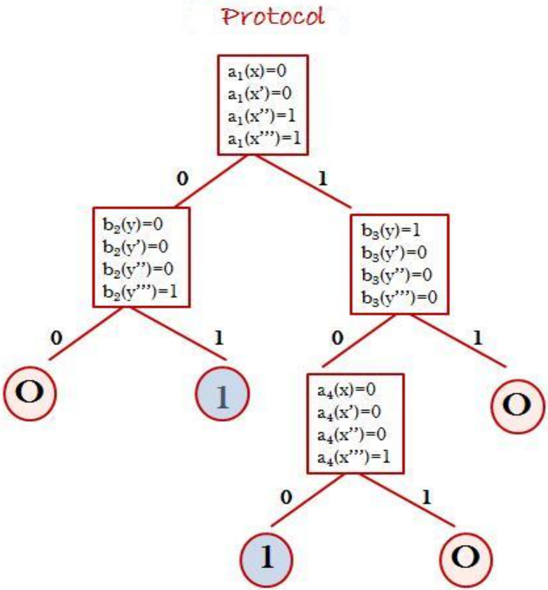
Definition

A protocol P over domain $X \times Y$ with range Z is a binary tree, where each internal node v is labeled either by a function $a_v : X \longrightarrow \Sigma$ or by a function $b_v : Y \longrightarrow \Sigma$, and each leaf is labeled with an element $z \in \{0, 1\}$.

Communication complexity

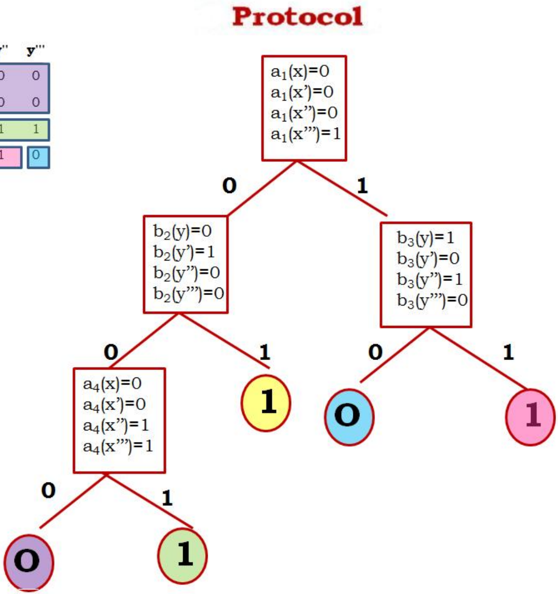
(similar to the example in Kushilevitz and Nisan, “Communication Complexity”)

f(x,y)	y	y’	y’’	y’’’
x	0	0	0	1
x’	0	0	0	1
x’’	0	1	1	1
x’’’	0	0	0	0



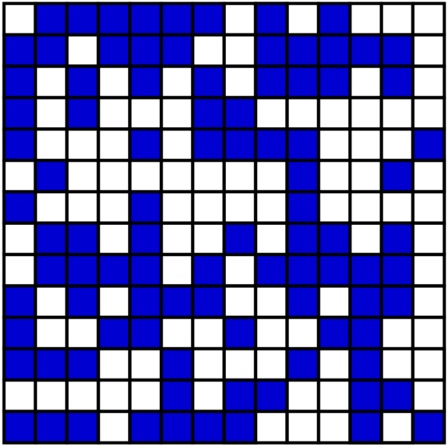
Communication complexity

f(x,y)	y	y'	y''	y'''
x	0	1	0	0
x'	0	1	0	0
x''	1	1	1	1
x'''	1	0	1	0



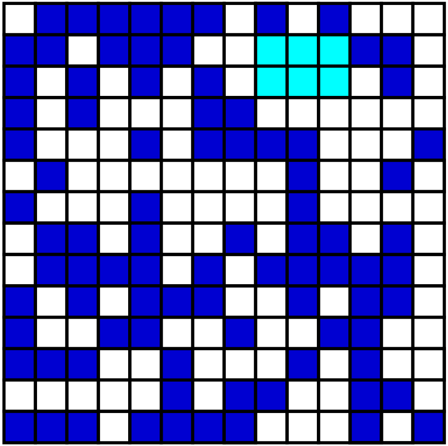
Communication complexity

"Random" square – the "functions" we will study…



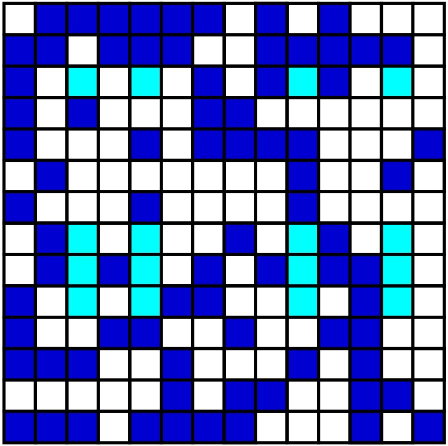
Communication complexity

A monochromatic geometrical rectangle with maximum area:



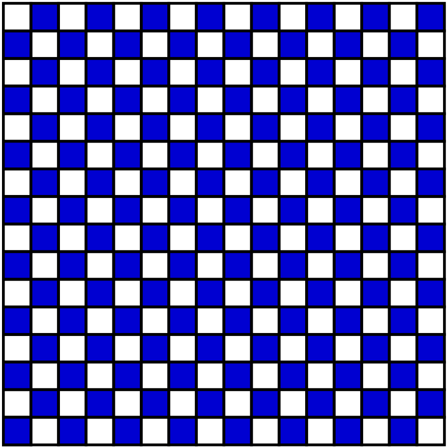
Communication complexity

A monochromatic combinatorial rectangle with maximum area



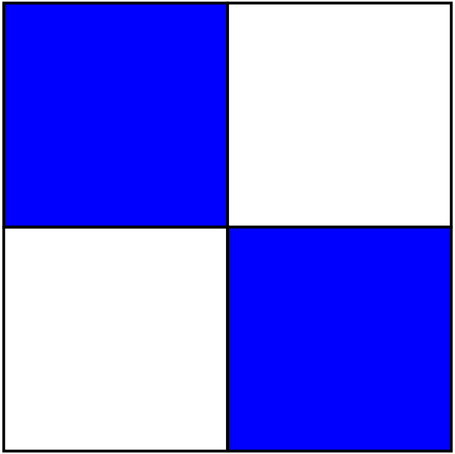
Communication complexity

What is the minimum number of monochromatic combinatorial rectangles needed in this case?



Communication complexity

The answer: 4



Communication complexity

Definition

For a function $f : X \times Y \to Z$, the (deterministic) communication complexity of f is:

$$D(f) = \min_P \{D_P(f) : P \text{ is a protocol for } f\}$$

Communication complexity

Protocol

A trivial protocol for communication complexity:

- Alice sends x to Bob;
- Bob computes $f(x,y)$ and sends the result to Alice.

Theorem

For every function $f : X \times Y \to Z$,

$$D(f) \leq \log|X| + 1$$

Communication complexity

Definition

A square $R \times R$ with a function $c : R \times R \longrightarrow \{0, 1\}$ is called by colored square.

Definition

Given a finite set R, a combinatorial rectangle is a set $A \times B$, where A and B are subsets of R.

Definition

We say that a combinatorial rectangle $A \times B$ is z-monochromatic if $c(a, b)$ has the same value z (0 or 1) for all $a \in A$ and $b \in B$.

Communication complexity

Lemma 1

Any protocol P for a function f induces a partition of $X \times Y$ into z-monochromatic rectangles. The number of rectangles is the number of leaves of P.

Communication complexity

Lemma 2

If any partition of $X \times Y$ into z-monochromatic rectangles requires at least t rectangles, then $D(f) \geq \lceil \log t \rceil$.

Lemma 3

If all rectangles have size not greater than k, then the number of rectangles is, at least, $\frac{|X||Y|}{k}$. Thus,

$$D(f) \geq \log |X| + \log |Y| - \log k$$

Outline

Preliminaries

NP-completeness of the 1-mcr problem

The largest monochromatic rectangles

Random “functions”: communication complexity lower bound

Result: The 1-mcr problem is NP-complete

We denote the problem of finding the largest 1-monochromatic rectangle of a given colored square by 1-mcr.
The corresponding decision problem to 1-mcr is:

1MMR, maximum area 1-mcr
INSTANCE: $\langle (Q, c), n, k \rangle$ *where* (Q, c) *is a colored square with side* n *and* $k \in \mathbb{Z}^+$.
QUESTION: *Does* Q *contain a 1-mcr with area at least* k*?*

The 1-mcr problem is NP*-complete*

Consider the following problems:

3SAT
INSTANCE: $\langle U, C \rangle$ *where U is a finite set of logical variables and C is a collection of clauses, where each clause has exactly 3 literals, no two of which are derived from the same variable.*
QUESTION: *Is there a truth assignment for U that satisfies all the clauses in C?*

MEiB, maximum edge independent biclique
INSTANCE: $\langle (V_1 \cup V_2, E), k \rangle$ *where* $(V_1 \cup V_2, E)$ *is a bipartite graph and k is a positive integer.*
QUESTION: *Does G contain an i-biclique* (A, B) *such that* $|A| \cdot |B| \geq k$*?*

The 1-mcr problem is NP*-complete*

It is known that 3SAT $\leq_p$ MEiB; it is not difficult to see that MEiB$\leq_p$1MMR:

$$\begin{array}{rcl} \text{MEiB} & \leq_p & \text{1MMR} \\ \langle (V_1 \cup V_2, E), k \rangle & \rightarrow & \langle Q, n, k' \rangle \end{array}$$

Assume that $|V_1| \geq |V_2|$ and let $V_1 = \{x_1, x_2, \cdots, x_m\}$ and $V_2 = \{y_1, y_2, \cdots, y_p\}$, with $m \geq p$. The reduction is defined by:

- $n = m = |V_1|$;
- $Q = \{1, 2, \cdots, n\} \times \{1, 2, \cdots, n\}$;
- $Q_{i,j} = 1$ iff $(i, j) \in E$ and $Q_{i,j} = 0$ for $p < j \leq n$;
- $k' = k$

Outline

Preliminaries

NP-completness of the 1-mcr problem

The largest monochromatic rectangles

Random "functions": communication complexity lower bound

Summary
Asymptotic results (not all proved in this presentation):

- Both sides $> 2\log n \;\Rightarrow$: impossible.
- Maximum area: about $n/2$
- Corresponding shapes: $(1 \times n/2)$, $(2 \times n/4)$, $(n/2 \times 1)$, $(n/4 \times 2)$.
- No "intermediate" large rectangles.
- Preferential shapes: $(2 \times n/4)$, $(n/4 \times 2)$.

Both sides $> 2\log n \;\Rightarrow\;$ asymptotically impossible!

Consider a random colored square (Q, c), where each $c(a, b)$ is a random variable z such that:

$$z = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Theorem

If a, b are such that

$$a, b > c \log n / \log(1/p)$$

with $c > 2$, then $\lim_{n\to\infty} E(a, b) = 0$. So, areas greater than $4\log^2 n/\log^2(1/p)$ do not exist in the limit.

The largest monochromatic rectangles

Both sides $> 2\log n\ldots$ particular case: $p = \frac{1}{2}$

Theorem

If $c > 2$ and a, b are such that $a, b > c\log n$, then the asymptotic probability of existing a monochromatic rectangle of area $a \times b$ is 0.

Proof

$E(a, b)$: expected number of 1mcrs of area $a \times b$. Depends of a, b and n.

$$E(a,b) \;=\; 2^{-ab}\binom{n}{a}\binom{n}{b} \;\leq\; 2^{-ab}n^{a+b} \;=\; 2^{-ab+(a+b)\log n}$$

Put $a = 2\log n + \alpha,\ b = 2\log n + \beta,\quad$ with $\alpha, \beta > 0$

Both sides $> 2\log n\ldots$ Particular case: $p = \frac{1}{2}$

Proof (Cont.)

The exponent of 2 may be written as:

$$
\begin{aligned}
& -ab + (a+b)\log n && (1)\\
&= -(2\log n + \alpha)(2\log n + \beta) + (2\log n + \alpha + 2\log n + \beta)\log n && (2)\\
&= -4\log^2 n - 2(\alpha+\beta)\log n - \alpha\beta + (4\log n + \alpha + \beta)\log n && (3)\\
&= -(\alpha+\beta)\log n - \alpha\beta && (4)\\
&\leq -(\alpha+\beta)\log n && (5)
\end{aligned}
$$

Thus, $E(a,b) \leq 2^{-(\alpha+\beta)\log n}$, so $\lim_{n\to\infty} E(a,b) = 0$.

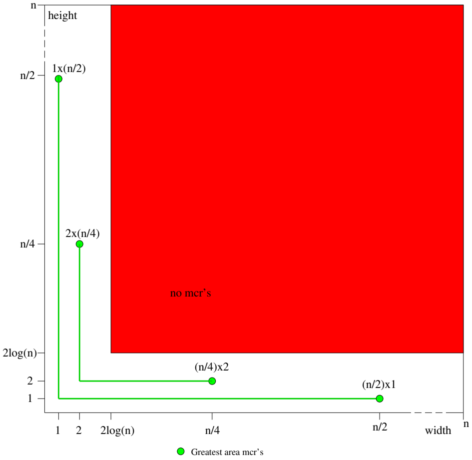
The largest monochromatic rectangles

Asymptotic lower upper bounds of 1-mcr areas

- Maximum area is slightly larger than $n/2$
- Corresponding asymptotic shapes: $(n/2 \times 1)$, $(n/4 \times 2)$, $(1 \times n/2)$, $(2 \times n/4)$.
- Example, random square 2000×2000:
  - Max. area of geometrical monochromatic rectangles: (about) 24, experimental.
  - Max. area of combinatorial monochromatic rectangles: (about) 1000 (1000×1 and 500×2)

The largest monochromatic rectangles

Summary; the shape of the largest 1-mcr's for $p = 1/2$



The largest monochromatic rectangles

a: number of rows, x: number of columns, s: area

	$a=1$	$a=2$	$a=3$
$E(x)$	$n/2$	$n/4$	$n/8$
$E(s)$	$n/2$	$n/2$	$3n/8$
$\rho(s)$	$\sqrt{n/2}$	$\sqrt{3n}/2$	$3\sqrt{7n}/8$

Outline

Preliminaries

NP-completness of the 1-mcr problem

The largest monochromatic rectangles

Random "functions": communication complexity lower bound

A lower bound of the deterministic communication complexity

Idea (for $p = 1/2$):

$$\text{Maximum area} \leq n/2$$

$$\Rightarrow$$

$$\text{Number of mcr's} \geq n^2/(n/2) = 2n$$

$$\Rightarrow$$

$$D(f) \geq \lceil \log(2n) \rceil$$

A lower bound of the deterministic communication complexity

The random "function"

$$f_p(x,y) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

Theorem ($D(f)$ lower bound)

The asymptotic deterministic communication complexity of a random "function" $f_p(x,y)$ satisfies

$$D(f) \geq \left\lceil \log n + \log\left(p \ln \frac{1}{p} + (1-p)\ln\frac{1}{1-p}\right) + \log e \right\rceil$$

A lower bound of the deterministic communication complexity

Comparison

Let P the trivial protocol previously described. $D_P(f) \leq \lceil \log n \rceil + 1$.
Then the last theorem shows that for a fix p $(0 < p < 1)$, the protocol P is "almost" optimal, in the sense that

$$\begin{aligned} D_P(f) - D(f) &\leq \lceil \log n \rceil + 1 - \left\lceil \log n + \log\left(p \ln \tfrac{1}{p} + (1-p) \ln \tfrac{1}{1-p}\right) + \log e \right\rceil \\ &< 2 - \log e - \log\left(p \ln \tfrac{1}{p} + (1-p) \ln \tfrac{1}{1-p}\right) \end{aligned}$$

A lower bound of the deterministic communication complexity

Comparison

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$D_P(f) - D(f)$	2.18	1.56	1.27	1.13	1.09	1.13	1.27	1.56	2.18

Conclusion

We conclude that

- in the limit, **no protocol** for the random “function” is significantly better than the trivial protocol P.
- in this case, the lower bound of $D(f)$ based on the number of mcr’s is almost optimal.

Experiments...

Conclusion. Small relative error: only for relatively large values of n.

Monochromatic combinatorial rectangle; maximum areas (20 samples)

height=1		
n	MA	Dev
1	0.4	-20%
2	1.3	30%
5	3.7	48%
10	7.2	43%
20	14.2	42%
50	33.0	32%
100	63.0	26%
200	118.7	19%
500	273.9	10%
1000	548.1	10%
2000	1076.5	8%

height=2		
n	MA	Dev
1	0.0	-100%
2	1.1	10%
5	4.8	92%
10	10.6	112%
20	19.7	97%
50	45.1	80%
100	82.4	65%
200	150.9	51%
500	339.0	36%
1000	618.0	24%
2000	1192.4	19%

height=3			
n	MA	Dev	
1	0.0	-100%	
2	0.0	-100%	
5	4.4	132%	
10	12.1	223%	
20	23.4	212%	
50	50.0	166%	
100	84.8	126%	
200	147.8	97%	
500	321.0	71%	(1)
1000	567.0	51%	(2)
2000	1032.0	38%	(2)

MA- maximum area (average of 20 samples except [1] and [2])
(1) 5 samples
(2) 1 sample
Dev - percentual deviation to theoretical value

The end.
Thank you!