MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

CTIC 交互计算

Elad Verbin
Aarhus University

Danmarks Grundforskningsfond
Danish National Research Foundation

Communication Complexity

Lower bounds for number-in-hand multiparty communication complexity, made easy

with Jeff Phillips, Qin Zhang, SODA '12

The Model

- *k* players, each has as input a length-*n* bitstring
- Want to communicate and compute some function of their inputs
- e.g. coordinate-wise AND, coordinate-wise XOR, set disjointess, etc
- **Note:** Number-in-hand, not Number-on-forehead
- Usually studied in the context of streaming lower bounds, as a promise problem. In our case, the problem is not a promise problem
- **Motivation:** Tracking/Monitoring; communication with a central server; fundamental communication problem

Our Results

Message-Passing Model:

- coordinate-wise XOR: $\Theta(n \cdot k)$
- coordinate-wise AND: $\Theta(n \cdot k)$

Blackboard Model:

- coordinate-wise XOR: $\Theta(nk)$
- coordinate-wise AND: $\Theta(n \cdot \log k)$

(only for some ranges of dependence of *n* on *k*)

New Technique: Symmetrization

- Non-trivial problem. Surprisingly difficult to get lower bound without using symmetrization. (Evidence: see strange hard distribution for coordinate-wise AND)

Followup Work: By Woodruff and Zhang.

Message Passing Model

- Reduction from *k*-player game to 2-player game
- 2-player game: Alice simulates a randomly-chosen player, Bob simulates all other *k*-1 players
- Distributional setting. Distribution called *symmetric* if it is invariant under renaming of players
- **For symmetric distributions:** If exists communication protocol for *k*-player game with communication ***C*** ⇒ exists protocol for 2-player game with expected communication ***C/k***.
By linearity of expectation
- Easy observation, strong consequences!

Particulars

- Hard distribution for XOR: everything uniform i.i.d. obviously symmetric.
- 2-player game: Alice gets *n*-bit input, Bob gets *n*-bit input, need coordinate-wise XOR. Obviously $\Omega(n)$.
- Hard distribution for AND: Random half of the coordinates are filled by random i.i.d bits ("confusion part"). The other half only has one 0 somewhere, in a random location. the rest are 1s.
- **Intuition:** if no confusion part and all is i.i.d, then players can communicate in a chain and send only $O(n \cdot \log k)$ bits in total. So need confusion part. If confusion part is all-1s, have a Slepian-Wolf type protocol, again with communication $O(n \cdot \log k)$. But if confusion part is i.i.d. random, players are indeed confused. The 2-player problem is similar to set-disjointness (Some technical complications arise)

Blackboard Model

- Reduction from k-player game to 2-player game
- 2-player game: Alice simulates a randomly-chosen player, Bob simulates *another* randomly-chosen player, the rest are simulated by both players, via shared randomness
- Distributional setting. Distribution called *product distribution* if the input of each player is chosen independently of the inputs of the other players
- **For symmetric product distributions:** If exists ***C***-communication protocol for *k*-player game ⇒ exists protocol for 2-player game with expected communication **2*C/k***. By linearity of expectation

Particulars

- Hard distribution for XOR: everything uniform i.i.d. obviously symmetric product distribution. 2-player game is to compute XOR of two uniformly random *n*-bit vectors
- Hard distribution for AND: each bit is 0 w.p. $1/k$, all are i.i.d. 2-player game: coordinate-wise AND on two such vectors, in ≈ $1/e$-fraction of the coordinates Complexity $\Omega(n \cdot \log k/k)$

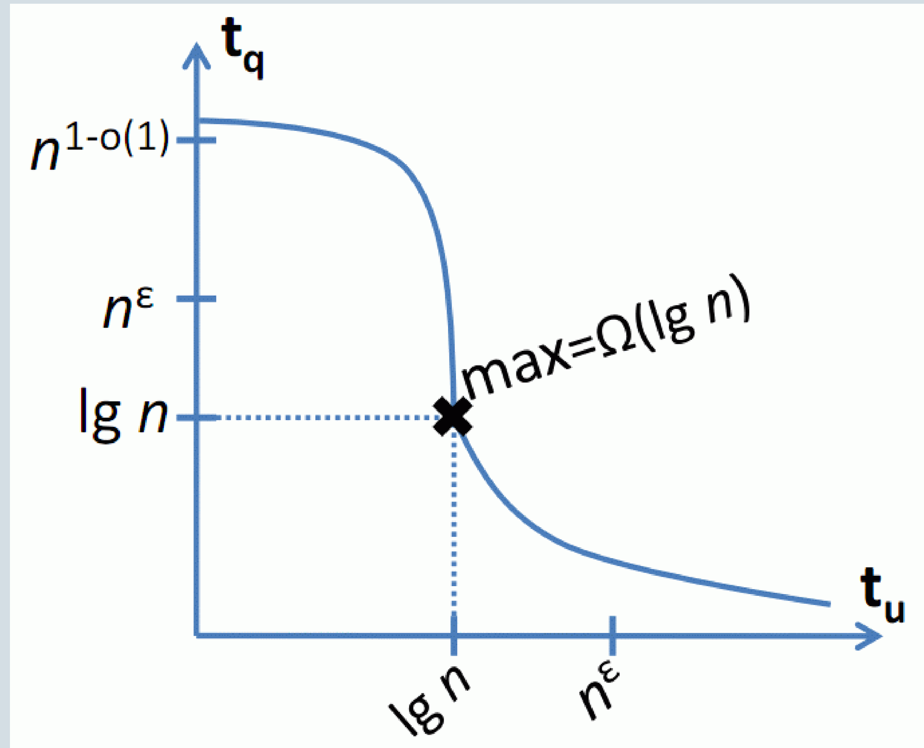
Data Structures

Work in progress

with Joshua Brody, Kasper Green Larsen

Succinct vs. Dynamic

- **Dynamic Data Structures:** support updates; support queries
- **Static Data Structures:** store little extra information; support queries
- **Theorem:** Dynamic DS lower bounds ⇔ Succinct DS lower bounds
- Dynamic lower bounds state of the art:



(Patrascu, '10)

Dynamic vs. Circuit Complexity

- Wish to prove polynomial lower bounds for an explicit dynamic data structure problem
- **Progress:** Proved lower bounds for non-adaptive queries and updates
- **Techniques:** Proof uses circuit complexity lower bound techniques by Jukna
- Alternative proof methods also exist
- **Next Step:** try to strengthen to get lower bounds for the adaptive case (hard!)

MADALGO – Center for Massive Data Algorithmics, a Center of the Danish National Research Foundation