MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS


Elad Verbin
Aarhus University

Danmarks Grundforskningsfond
Danish National Research Foundation

Multiparty Communication Lower Bounds

Work in progress

with Jeff Phillips, Qin Zhang

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: Θ(nk)
- coordinate-wise AND: Θ(nk)

Blackboard Model:

- coordinate-wise XOR: Θ(nk)
- coordinate-wise AND: Θ(nlogk)

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

Open: Set disjointness. Other interesting functions? More applications?

New Technique: Symmetrization

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

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 Ω(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(nlogk) bits in total. So need confusion part. If confusion part is all-1s, have a Slepian-Wolf type protcol, again with communicaton O(nlogk). 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 **2C/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 Ω(nlogk/k).

Succinct Data Structures

Work in progress

with Pooya Davoodi, Kasper Larsen, Qin Zhang

The Problem

Range Minimum Problem:

- Preprocess a length-n array A of w-bit integers.
- Want to answer queries of the form: min[i..j] which returns the smallest element in the range A[i..j]
- 3 variants:
  - location: returns the location of the minimum
  - value: returns the value of the minimum)
  - location+value (returns both)
- What is the length, in bits, of the best encoding possible, if we don't care about query time? 2n±polylog, nw and nw, respectively.
- Suppose we want efficient operations. Suppose we are allowed space 2n+r (in the location variant) or wn+r (in the other variants). How fast can the query be?

Results

- Conjecture: More or less, $r=n/w^t$. Lower bound side should be the same expression as Patrascu-Viola result on rank-select, upper bound side should be same expression as Patrascu's "Succincter" data structure.
- Patrascu-Viola's techniques don't quite work. Need new Ideas.
- Right now can prove lower bound for "non-adaptive" data structures – those where any particular query always probes the same cells.
- Approach: Information Theory, plus probabilistic analysis of random cartesian trees.