Set disjointness with constant error

Yuval Filmus (joint with Yuval Dagan, Hamed Hatami, Yaqiao Li)

December 26, 2018

1 Introduction

Set disjointness is the following problem: Alice gets a subset $A \subseteq [n]$ and Bob gets a subset $B \subseteq [n]$. Their goal is to determine whether A and B intersect, using as little communication as possible in the worst case. This inocuous problem is behind many applications of communication complexity.

The trivial algorithm has A send her entire input to Bob, and then Bob sends the answer. In total, Alice and Bob communicate $n+1$ bits. A simple rank argument shows that this algorithm is optimal.

The picture becomes more interesting if we allow the protocol to be randomized; in this case, the parties should output the correct answer with some constant probability. Razborov gave a linear lower bound even for this setting.

In a tour-de-force, Braverman, Garg, Pankratov and Weinstein showed that there is a constant $\alpha \approx 0.4827$ such that the randomized communication complexity of set disjointness with error $o(1)$ is $\alpha n + o(n)$. They used tools of information complexity.

We show that the randomized communication complexity of set disjointness with fixed error ϵ is $(\alpha - \Theta(h(\epsilon))n$, by extending the arguments of Braverman et al.

2 Information complexity

In the one-way communication setting, the entropy of a random variable X is the amortized encoding length of samples of X: in order to encode n samples of X, we need $H(X)n$ bits on average (or with high probability).

Braverman defined an analog of entropy in the setting of communication complexity. Given a distribution μ on the two players' inputs X, Y, the information complexity of a protocol π is $\mathrm{IC}_\mu(\pi) = I(\Pi; X|Y) + I(\Pi; Y|X)$ (here Π is the transcript). Given a function f, we define $\mathrm{IC}_\mu(f, \epsilon)$ as the infimum of $\mathrm{IC}_\mu(\Pi)$ over all protocols Π computing f with error at most ϵ (on each input); the infimum is not always achieved. Braverman showed that the communication complexity of N copies of f with ϵ error per copy (with respect to μ) is $\mathrm{IC}_\mu(f, \epsilon)N + o(N)$.

Braverman et al. showed that the communication complexity of set disjointness with error $o(1)$ is $\mathrm{IC}^0(\mathrm{AND}, 0)n + o(n)$, where $\mathrm{IC}^0(\mathrm{AND}, 0)$ is the maximum information complexity of the AND function (with no error) over distributions on which the answer is always 0 (the relevant protocols, however, should be correct on *all* inputs). Intuitively, the hardest case is when one needs to check all n potential elements in the intersection.

We conjecture that the communication complexity of set disjointness with error ϵ is $\mathrm{IC}^0(\mathrm{AND}, \epsilon, 1 \to 0)n + o(n)$, where $\mathrm{IC}^0(\mathrm{AND}, \epsilon, 1 \to 0)$ is the maximum information complexity of AND over the same set of distributions as before, where protocols are allowed only a one-sided error (with probability ϵ). While we do not manage to prove this conjecture, we are able to prove the upper bound, that is, construct a protocol of set disjointness with that cost. Our lower bound is $\mathrm{IC}^0(\mathrm{AND}, \epsilon)n + o(n)$.

From now on we will forget about set disjointness and focus only on the AND function. Let us first dispense with the upper bound — it follows from taking the optimal protocol for the AND function, and modifying one of the input bits from 1 to 0 with probability ϵ. The hard part is proving the lower bound $\mathrm{IC}^0(\mathrm{AND}, \epsilon) \geq \mathrm{IC}^0(\mathrm{AND}, 0) - \Theta(h(\epsilon))$.

Before moving on, here is a simple argument showing that $\mathrm{IC}^0(\mathrm{AND}, \epsilon) \geq \mathrm{IC}^0(\mathrm{AND}, 0) - \Theta(h(\sqrt{\epsilon}))$. Consider any protocol for AND with error ϵ. At any leaf of the protocol, our uncertainty about the output is at most ϵ. This uncertainty can be split among the inputs of both players, but the worst case is when the uncertainty about each input is roughly $\sqrt{\epsilon}$. We "complete" the protocol to a zero-error protocol by having the two players exchange their bits, thus revealing $O(h(\sqrt{\epsilon}))$ more information. The resulting protocol has cost at most $\mathrm{IC}^0(\mathrm{AND}, 0)$.

3 Protocols as random walks, and concealed information

We can think of a protocol as a random walk in the space of distributions on $X \times Y$. The starting point is the distribution μ. Whenever Alice sends a bit, an outside observer can update their current belief on the distribution by scaling the rows (this uses Bayes' rule). Similarly, bits of Bob correspond to scaling the columns. Eventually, the players reach a point in which the answer is known.

Information complexity measures the amount on information revealed to each player about the other player's input. It will be more convenient to consider instead the *concealed information*:

$$\mathrm{CI}_\mu(\pi) = H(X|\Pi Y) + H(Y|\Pi X) = H(X|Y) + H(Y|X) - \mathrm{IC}_\mu(\pi).$$

Minimizing the information revealed is the same as maximizing the information concealed. It is not hard to check that the amount of information concealed depends only at the distribution of the leaves of the protocol.

4 Optimal protocol for AND

The optimal protocol for AND depends on the probability distribution μ. Let us assume for simplicity that μ is symmetric. Alice and Bob pick uniformly random times $t_A, t_B \in [0, 1]$. A clock counts from 0 to 1 (continuously). When a player's time has been reached, if that player has 0 as an input, they press a buzzer and the protocol halts — the output is 0. If nobody pressed on the buzzer at time 1, the output is 1. (Equivalently, they choose t_A, t_B accordingly to an exponential distribution, and the protocol never halts.)

There is another way to view this protocol, as a random walk on $[0,1]^2$ manifold. Let us assume first that the initial distribution is a product distribution. Since the actions of Alice and Bob correspond to scaling rows and columns, at any given point in time the distribution is still a product distribution, and so we can parametrize the space of all reachable distributions as $[0,1]^2$. The semantics of the random walk ensures that it forms a martingale. By simple manipulations, we can guarantee that each bit sent is unbiased (as seen to an outside observer), and furthermore reveals an infinitesimal amount of information. In order to complete the definition of the protocol, it suffices to specify which player speaks at which state. The random walk starts at some distribution μ, and ends when the input is known: when reaching the left side, the bottom side, or the top-right corner. The concealed information is the expected value of a reward function which corresponds to the information concealed at each such point.

It turns out (this was also observed by Pankratov) that the optimal protocol is obtained by letting Alice speak if $\Pr[X = 1] \leq \Pr[Y = 1]$, and letting Bob speak otherwise. In other words, if we think of the two parameters of $[0,1]^2$ as $\Pr[X = 1]$ and $\Pr[Y = 1]$, then Alice speaks above the diagonal, and Bob speaks below the diagonal. If we discretize this continuous time random walk, then we do get the buzzer protocol.

When the initial distribution is not a product distribution, the random walk still takes place on a 2-dimensional manifold which is homeomorphic to $[0,1]^2$. In fact, we can squint and treat it as a random walk on $[0,1]^2$ with different reward.

5 Optimality of the buzzer protocol

Let us consider a particular initial distribution, and the manifold of distributions reachable from it. Suppose that f is the optimal concealed information for the distributions in the manifold. At any given point x, we can consider the following experiment: one of the player sends a bit, and from that point on we run the optimal protocol. Since f is the optimal concealed information, the corresponding protocol cannot conceal more information. When the bits being sent convey an infinitesimal amount of information, we obtain the following condition: f must be concave in both coordinates separately. Braverman et al. showed that this condition, together with some boundary conditions, characterizes the optimal concealed information function.

In order to show that the buzzer protocol is optimal, all one needs to do is to check the boundary conditions (which depend on the manifold) and to show that the function f is concave in both arguments separately. This argument is simpler (and more correct!) than the one given by Braverman et al.

6 Protocol completion for the buzzer protocol

The technique of protocol completion shows that $\mathrm{CI}_\mu(\mathrm{AND}, \epsilon) \leq \mathrm{CI}_\mu(\mathrm{AND}, 0) + O(h(\sqrt{\epsilon}))$. The idea is to take a protocol with error ϵ, and complete it to a zero-error protocol by revealing the players' inputs at the end. The cost is $O(h(\sqrt{\epsilon}))$ since an uncertainty of ϵ in the output corresponds (in the worst case) to uncertainties of $\sqrt{\epsilon}$ in the inputs. If we know that the uncertainties are skewed, say c and ϵ/c, then we could complete the

protocol using only $O(h(\epsilon))$. Our goal would be to show that this is typically the case for good protocols for the AND function.

We use the technique of *information wastage*, appearing in Braverman et al. in a slightly different context. The idea is that if a protocol differs from the optimal protocol in some specific way then the concavity constraint has a slack which guarantees that the protocol will conceal less information the the optimal protocol. The loss compared to the optimal protocol is the *information wastage*. In our case, the quantity which we want to compare to the information wastage is, roughly speaking, the time spent in the lower left corner of $[0,1]^2$ (more accurately, we look at $\mathbb{E}[(c-\ell)_+^2]$, where c is an appropriate constant and $\ell = \max(p,q)$ corresponds to the final leaf (p,q) of the protocol). We are able to lower bound the information wastage by our error parameter, thus showing that protocols close to the optimal concealed information typically are far from the lowr left corner. Consequently, protocol completion only incurs a loss of $O(h(\epsilon))$.

7 Optimal protocols for every function

Our work leaves several questions open, but perhaps the most interesting ones concern optimal protocols. Braverman et al. show that there are no information-optimal conventional protocols for the AND function. The buzzer protocol is a "generalized" protocol which is optimal. We conjecture that there always exists an information-optimal random walk protocol.

A related conjecture is about bounded round information-optimal protocols. Braverman et al. show that for the AND function, if we restrict the number of rounds to r then we lose $\Theta(1/r^2)$ in the information, and conjectured that the same upper bound holds for all functions. Some upper bound follows from the work of Braverman and Schneider on the computability of information complexity. One way to prove the conjecture is by discretizing random walk protocols. Another one (work in progress) is by considering a fixed-point approach for calculating the information complexity. In this approach, we iterate a "best response" operator, which determines the information-optimal protocol for successive number of rounds. We can prove the $O(1/r^2)$ conjecture by showing that this operator converges quickly to its fixed point.