Imperfectly Shared Randomness in Communication

Madhu Sudan
Microsoft Research

Joint work with Clément Canonne (Columbia), Venkatesan Guruswami (CMU) and Raghu Meka (?).

Dedicate to our SVC colleagues!

You are the best!

Context in Communication

- Context = Central element of Communication.
  - Shared between sender and receiver
  - Implicit. (Doesn't participate in n)
- Examples:
  - Meaning of bits (what action to take given rec'd message).
  - Shannon theory: Distribution of source, Channel behavior, Codes used.
  - Communication Complexity: Function being computed, Randomness being shared etc.
  - Human communication: Language, Grammar …

Uncertainty in sharing of context

- Whenever "large" amounts on information is "shared", there must be some imperfection.
  - Online Forms: Example – my bank:
    - "Please enter your PIN now"
      - But I have an ATM PIN, a phone PIN, a transaction PIN.
  - Compression: Do sender and receiver agree perfectly on the prior? [Juba,Kalai,Khanna,S.'11], [Haramaty,S.'14]
  - This talk: Shared Randomness in Communication Complexity.

Shared Randomness in CC

- Canonical example: Equality testing.
  - Alice has $x \in \{0,1\}^n$; Bob has $y \in \{0,1\}^n$;
  - Want to know if $x = y$?
  - Deterministically: Communicate $\Omega(n)$ bits
  - With private randomness: $\Theta(\log n)$ bits
    - Idea: Alice encodes $x \mapsto E(x)$; Picks $i \in [N]$; sends $(i, E(x)_i)$
  - With shared randomness: $O(1)$ bits
    - Just send $E(x)_i$
- Upshot: Randomness very helpful!

Compression with Uncertain Priors

- [JKKS'11]:
- Alice has $P = (P_1, \dots, P_N); m \leftarrow_P [N]$;
- Bob has $\mathrm{Q} = (Q_1, \dots, Q_N); P \approx_\Delta Q$;
- Both want to know m.
- State of affairs:
  - $P = Q$: Expected comm. = $H(P)$. [Huffman]
  - $P \approx_\Delta Q$ + shared randomness: $H(P) + 2\Delta$ [JKKS]
  - $P \approx_\Delta Q$ deterministically: $\mathrm{O}(H(P) + \Delta + \log\log \mathrm{N})$ [Haramaty+S.]

Uncertain Compression (thoughts)

- Is entropy the right measure of compressibility?
  - With uncertainty?
    - Deterministically … may be not (the $\log \log n$)
    - Randomized: Perfect sharing inconsistent with uncertainty!
    - Unless … randomness is shared imperfectly!

- Motivates: Imperfectly shared randomness in CC.

- "Independently" raised and studied by [Bavarian,Gaminsky,Ito'14].

Our Model

- General communication complexity with imperfectly shared randomness.
- Alice $\leftarrow r$; and Bob $\leftarrow s$ where $(r, s) =$ i.i.d. sequence of correlated pairs $(r_i, s_i)_i$; $r_i, s_i \in \{-1, +1\}; E[r_i] = E[s_i] = 0; E[r_i s_i] = \rho$.
- Notation:
  - $isr_\rho(f)$ = cc of f with ρ-correlated bits.
  - $psr(f)$: perfectly shared randomness cc.
  - $priv(f)$: cc with private randomness
- Starting point: for Boolean functions f
  - $psr(f) \leq isr_\rho(f) \leq priv(f) \leq psr(f) + \log n$

Results

- [Bavarian et al.]: Focus on simultaneous message model; more general correlations.
- Our focus:
  - One-way communication: Alice → Bob; Bob outputs f.
  - Problems where difference of $\log n$ significant.
- Results:
  - Uncertain Compression: $O_\rho(H(P) + \Delta)$
  - Equality testing: $O_\rho(1)$ (also [Bavarian et al.])
  - More generally: $psr(f) \leq k \Rightarrow ow\text{-}isr(f) \leq 2^k$
  - Converse: $\exists f$ with $ow\text{-}psr(f) \leq k$ & $ow\text{-}isr(f) \geq 2^k$

Rest of the talk

- Uncertain Compression: $O_\rho(H(P) + \Delta)$
- Equality testing: $O_\rho(1)$ (also [Bavarian et al.])
- General upper bound: $psr(f) \leq k \Rightarrow ow\text{-}isr(f) \leq 2^k$
- Converse: $\exists f$ with ow-$psr(f) \leq k$ & $ow\text{-}isr(f) \geq 2^k$

Compression:

- [JKKS] psr solution: Let common randomness define "dictionary": arbitrarily long sequences r_m for every message m.
  - Alice sends "long enough" prefix of r_m
  - Bob does maximum likelihood decoding based on Q.
  - Analysis: Exercise
- Our isr solution:
  - Alice send longer prefix.
  - Bob does max. likelihood decoding among messages that are close enough to rec'd word.
- Moral: Protocols "natural" ⇒ Explains behavior?

Rest of the talk

- Uncertain Compression: $O_\rho(H(P) + \Delta)$
- Equality testing: $O_\rho(1)$ (also [Bavarian et al.])
- General upper bound: $psr(f) \leq k \Rightarrow ow\text{-}isr(f) \leq 2^k$
- Converse: $\exists f$ with ow-$psr(f) \leq k$ & $ow\text{-}isr(f) \geq 2^k$

Equality Testing

- Key idea: Think inner products.
  - Encode $x \mapsto X = E(x); y \mapsto Y = E(y); X, Y \in \{-1, +1\}^n$
    - $x = y \Rightarrow \langle X, Y \rangle = n$
    - $x \neq y \Rightarrow \langle X, Y \rangle \leq n/2$
- Estimating inner products:
  - Using ideas from low-distortion embeddings …
  - Alice: Picks Gaussian $G \in \mathbb{R}^n$, sends $\langle G, X \rangle$
  - Bob: compares $\langle G, X \rangle$ with $\langle G', Y \rangle$

  - (mod analysis): $O_\rho(1)$ bits suffice if $G \approx_\rho G'$
  - [Bavarian et al.] Alternate protocol.

Rest of the talk

- Uncertain Compression: $O_\rho(H(P) + \Delta)$
- Equality testing: $O_\rho(1)$ (also [Bavarian et al.])
- General upper bound: $psr(f) \le k \Rightarrow ow\text{-}isr(f) \le 2^k$
- Converse: $\exists f$ with ow-$psr(f) \le k$ & $ow\text{-}isr(f) \ge 2^k$

General Communication

- Idea: All communication $\leq$ Inner Products
- Example: One-way communication k bits:
  - For each random string R
    - Alice's message = $i_R \in [2^k]$
    - Bob's output = $f_R(i_R)$ where $f_R: [2^k] \rightarrow \{0,1\}$
    - W.p. $\geq \frac{2}{3}$ over R, $f_R(i_R)$ is the right answer.

General Communication

- For each random string R
  - Alice's message = $i_R \in [2^k]$
  - Bob's output = $f_R(i_R)$ where $f_R: [2^k] \to \{0,1\}$
  - W.p. $\geq \frac{2}{3}$, $f_R(i_R)$ is the right answer.
- Vector representation:
  - $i_R \mapsto x_R \in \{0,1\}^{2^k}$ (unit coordinate vector)
  - $f_R \mapsto y_R \in \{0,1\}^{2^k}$ (truth table of f_R).
  - $f_R(i_R) = \langle x_R, y_R \rangle$
- Gaussian protocol estimates inner products to within relative error ϵ with $O\left(\frac{1}{\epsilon^2}\right)$ communication.

Rest of the talk

- Uncertain Compression: $O_\rho(H(P) + \Delta)$
- Equality testing: $O_\rho(1)$ (also [Bavarian et al.])
- General upper bound: $psr(f) \leq k \Rightarrow ow\text{-}isr(f) \leq 2^k$
- Converse: $\exists f$ with $\text{ow-}psr(f) \leq k$ & $ow\text{-}isr(f) \geq 2^k$

Main Technical Result: Matching lower bound

- There exists promise problem f s.t.
  - $ow\text{-}psr(f) \leq k$
  - $ow\text{-}isr_\rho(f) \geq \exp(k)$

- The Problem:
  - Gap Sparse Inner Product (G-Sparse-IP).
  - Alice gets sparse $x \in \{0,1\}^n$; $\mathrm{wt}(x) \approx 2^{-k} \cdot n$
  - Bob gets $y \in \{-1,+1\}^n$
  - Promise: $\langle x, y \rangle \geq \left(\frac{1}{3}\right) 2^{-k} \cdot n$ or $\langle x, y \rangle \leq 0$.
  - Decide which.

psr Protocol for G-Sparse-IP

- Idea: $x_i \neq 0 \;\Rightarrow\; y_i$ correlated with answer.
- Use (perfectly) shared randomness to find random index i s.t. $x_i \neq 0$.
- Shared randomness: $i_1, i_2, i_3, \ldots$ uniform over $[n]$
- Alice $\rightarrow$ Bob: smallest index j s.t. $x_{i_j} \neq 0$.
- Bob: Accept if $y_{i_j} = 1$
- Expect $j \approx 2^k; psr \leq k$.

ISR lower bounds

- Challenge: Usual CC lower bounds give a distribution and prove lower bound against it.
- G-Sparse-IP has a low-complexity protocol for every input, with shared randomness.
- Thus for every distribution, there exists a deterministic low-complexity protocol!
- So usual method can't work …

- Need to fix strategy first and then "tailor-make" a hard distribution for the strategy …

ISR lower bound for GSIP: Overview

- Strategies: Alice $f_r(x) \in [\ell]$; Bob $g_s(y) \in \{0,1\}^\ell$;
- Two possibilities:
  - Case 1: Alice's strategy and Bob's strategy have common highly "influential coordinate":
    - (i s.t. flipping x_i changes Alice's message etc.)
    - Leads to protocol for "agreement distillation" [We prove this is impossible.]
  - Case 2: Strategies have no common influential variable:
    - Invariance Principle ⇒ Solves some Gaussian problem
    - Lower bound for Gaussian problem. (Details shortly)

Case 1: Agreement Distillation

- Problem: Charlie $\leftarrow r$; Dana $\leftarrow s$; (r, s) ρ-correlated
- Goal: Charlie outputs u; Dana outputs v;

$$H_\infty(u), H_\infty(v) \geq k; \qquad \Pr[u = v] \geq \gamma$$

- Lemma: With zero communication $\gamma = 2^{-\Omega(k)}$;
- Proof: “Small-set expansion of noisy hypercube”
  - See, e.g., [Analysis of Boolean functions, O’Donnell]

- Corollary: For c bits of communication,

$$c \geq \epsilon \cdot k + \log \gamma$$

Completing Case 1

- Bad $\triangleq$ $\{i \mid \Pr_r[\mathrm{Inf}_i(f_r) \geq \mathrm{high}] \geq \mathrm{large}\}$
  $\cup \{i \mid \Pr_s[\mathrm{Inf}_i(g_s) \geq \mathrm{high}] \geq \mathrm{large}\}$
- Fact: (for our defn of influence) any function has bounded number of high influence variables.
- (By Fact + Markov) Can assume $|\mathrm{Bad}| \leq \epsilon \cdot n$.
- Distributions on Yes and No instances:
  - No: x random sparse $\in \{0,1\}^n$; $y \leftarrow_U \{-1,1\}^n$
  - Yes: Same as No on Bad coordinates.
    - On rest, y_i is more likely to be +1 if $x_i = 1$.

Completing Case 1 (contd.)

- Agreement strategy for Charlie + Dana:
  - Charlie: $i \in [n] - \text{Bad}$ s.t. $\text{Inf}_i(f_r)$ high.
  - Dana: $j \in [n] - \text{Bad}$ s.t. $\text{Inf}_j(g_s)$ high.
- Analysis:
  - $H_\infty(i), H_\infty(j)$ large since $i, j \notin \text{Bad}$.
  - $i = j$?: Case 1 assumption.

- Combined with lower bound for agreement distillation, implies Case 1 can't occur

Case 2: No common influential variable

- Key Lemma: Fix r, s; let $f = f_r$ and $g = g_s$.
  If ℓ small (2^{2^k}) and f, g distinguish Yes/No then f, g have common influential variable.
- Idea: Use "Invariance Principle":
  - Remarkable theorem: Mossel, O'Donnell, Oleskiewicz; Mossel++;
  - Informal form: f,g low-degree polynomials with no common influential variable ⇒ $\text{Exp}_{x,y}[f(x)g(y)] \approx \text{Exp}_{X,Y}[f(X)g(Y)]$
    - where x, y Boolean n-wise product dist.
    - and X, Y Gaussian n-wise product dist.

The Gaussian-IP Problem

- Suppose we can get the "perfect" invariance theorem for us …

- Would transform:

  Sol'n for G-Sparse-IP → Sol'n for G-Gaussian-IP
  - Alice, Bob get Gaussian vectors $X, Y \in \mathbb{R}^n$
  - Yes: $\langle X, Y \rangle \geq 2^{-k}$; No: $\langle X, Y \rangle \leq 0$

- Hope: Non-sparse ⇒ $\geq 2^k$ communication
  - Formally [Bar Yossef et al.]: Can reduce "indexing" to G-Gaussian-IP.

Invariance Principle + Challenges

- Informal Invariance Principle: f, g low-degree polynomials with no common influential variable $\Rightarrow \text{Exp}_{x,y}[f(x)g(y)] \approx \text{Exp}_{X,Y}[f(X)g(Y)]$
  - where x, y Boolean n-wise product dist.
  - and X, Y Gaussian n-wise product dist
- Challenges [+ Solutions]:
  - Our functions not low-degree [Smoothening]
  - Our functions not real-valued
    - $g: \{-1,1\}^n \to \{0,1\}^\ell$: [Truncate range to $[0,1]^\ell$]
    - $f: \{0,1\}^n \to [\ell]$: [???, [work with $\Delta(\ell)$]]

Invariance Principle + Challenges

- Informal Invariance Principle: f, g low-degree polynomials with no common influential variable $\Rightarrow \text{Exp}_{x,y}[f(x)g(y)] \approx \text{Exp}_{X,Y}[f(X)g(Y)]$
- Challenges
  - Our functions not low-degree [Smoothening]
  - Our functions not real-valued [Truncate]
  - Quantity of interest is not $f(x) \cdot g(y)$ …
    - [Can express quantity of interest as inner product.]
  - … (lots of grunge work …)
- Get a relevant invariance principle (next)

Invariance Principle for (one-way) CC

- Thm: ∃ transformations T_1, T_2 s.t.
  if $f: \{0,1\}^n \to \Delta(\ell)$ and $g: \{-1,1\}^n \to [0,1]^\ell$
  have no common influential variable, then
    $F = T_1 f: \mathbb{R}^n \to \Delta(\ell)$ and $G = T_2 g: \mathbb{R}^n \to [0,1]^\ell$ satisfy
    $\mathrm{Exp}_{x,y}[\langle f(x), g(y)\rangle] \approx \mathrm{Exp}_{X,Y}[\langle F(X), G(Y)\rangle]$

  - Main differences: f, g vector-valued.
  - Functions are transformed: $f \mapsto F; g \mapsto G$
  - Range is preserved exactly ($\Delta(\ell); [0,1]^\ell$)!
    - So F, G are still communication strategies!

Summarizing

- k bits of comm. with perfect sharing
  $\rightarrow 2^k$ bits with imperfect sharing.
- This is tight (for one-way communication)
  - Invariance principle for communication
  - Agreement distillation
  - Low-influence strategies

Conclusions

- Imperfect agreement of context important.
  - Dealing with new layer of uncertainty.
  - Notion of scale (context LARGE)
- Many open directions+questions:
  - Imperfectly shared randomness:
    - One-sided error?
    - Does interaction ever help?
    - How much randomness?
    - More general forms of correlation?

Thank You!