(Deterministic) Communication amid Uncertainty

Madhu Sudan

Microsoft, New England

Based on joint works with:

(1) Adam Kalai (MSR), Sanjeev Khanna (U.Penn), Brendan Juba (Harvard) and (2) Elad Haramaty (Technion)

Classical Communication

- The Shannon setting
  - Alice gets $m \in [N]$ chosen from distribution P
  - Sends some compression $y = E(m)$ to Bob.
  - Bob computes $\hat{m} = D(y)$
    - (with knowledge of $Q = P$).
  - Hope $m = \hat{m}$.

- Classical Uncertainty: $y \approx E(m)$

- Today's talk: Bob knows $Q \approx P$.

Outline

- Part 1: Motivation
- Part 2: Formalism
- Part 3: Randomized Solution
- Part 4: Issues with Randomized Solution
- Part 5: Deterministic Issues.

Motivation: Human Communication

- Human communication vs. Designed communication:
  - Human comm. dictated by languages, grammars …
    - Grammar: Rules, often violated.
    - Dictionary: $\exists$ multiple meanings to word.
    - Redundant: But $\neq$ error-correcting code.

- Theory for human communication?
  - Information theory?
  - Linguistics? (Universal grammars etc.)?

Behavioral aspects of natural communication

- (Vast) Implicit context.
- Sender sends increasingly long messages to receiver till receiver "gets" (the meaning of) the message.
  - Where do the options come from?
  - Comes from language/dictionary – but how/why?
- Sender may use feedback from receiver if available; or estimates receiver's knowledge if not.
  - How does estimation influence message.

Model:

- Reason to choose short messages: Compression.
  - Channel is still a scarce resource; still want to use optimally.

- Reason to choose long messages (when short ones are available): Reducing ambiguity.
  - Sender unsure of receiver's prior (context).
  - Sender wishes to ensure receiver gets the message, no matter what its prior (within reason).

Back to Problem

- Design encoding/decoding schemes (E/D) so that
  - Sender has distribution P on $[N]$
  - Receiver has distribution Q on $[N]$
  - Sender gets $m \in [N]$
  - Sends $E(P,m)$ to receiver.
  - Receiver receives $y = E(P,m)$
  - Decodes to $\hat{m} = D(Q,y)$

  - Want: $m = \hat{m}$ (provided P, Q close),
    - While minimizing $Exp_{m \leftarrow P} |E(P,m)|$

Contrast with some previous models

- Universal compression?
  - Doesn't apply: P,Q are not finitely specified.
  - Don't have a sequence of samples from P; just one!
- K-L divergence?
  - Measures inefficiency of compressing for Q if real distribution is P.
  - But assumes encoding/decoding according to same distribution Q.
- Semantic Communication:
  - Uncertainty of sender/receiver; but no special goal.

Closeness of distributions:

- P is Δ-close to Q if for all $m \in [N]$,

$$|\log P(m) - \log Q(m)| \leq \Delta$$

- P Δ-close to Q $\Rightarrow$ $D(P||Q), D(Q||P) \leq \Delta$

  (symmetrized, "worst-case" KL-divergence)

Dictionary = Shared Randomness?

- Modelling the dictionary: What should it be?
- Simplifying assumption – it is shared randomness, so …
- Assume sender and receiver have some shared randomness R and P, Q, m are independent of R.
  - $y = E(P, m, R)$
  - $\hat{m} = D(Q, y, R)$
- Want $\forall m, \ \Pr_R[\hat{m} = m] \geq 1 - \epsilon$

Solution (variant of Arith. Coding)

- Use R to define sequences
  - $R_1\ [1], R_1\ [2], R_1\ [3], \ldots$
  - $R_2\ [1], R_2\ [2], R_2\ [3], \ldots$
  - …
  - $R_N\ [1], R_N\ [2], R_N\ [3], \ldots.$
- $E_\Delta(P, m, R) = R_m[1 \ldots L],$ where L chosen s.t. $\forall z \neq m$

  Either $R_z[1 \ldots L] \neq R_m[1 \ldots L]$

  Or $\log P(z) < \log P(m) - 2\Delta$
- $D_\Delta(Q, y, R) = \hat{m}$ s.t. $\hat{m}$ max. $Q(\hat{m})$ among $\hat{m} \in \{z | R_z[1 \ldots L] = y\}$

Performance

- Obviously decoding always correct.

- Easy exercise:
  - $\text{Exp}_m \ [E(P,m)] = H(P) + 2\,\Delta$
    - ($H(P) \equiv \sum_m P(m) \log_2 \frac{1}{P(m)}$ "binary entropy")

- Limits:
  - No scheme can achieve $(1-\epsilon) \cdot [H(P) + \Delta]$
  - Can reduce randomness needed.

Implications

- Reflects the tension between ambiguity resolution and compression.
  - Larger the Δ ((estimated) gap in context), larger the encoding length.
- Coding scheme reflects the nature of human process (extend messages till they feel unambiguous).
- The "shared randomness'' is a convenient starting point for discussion
  - Dictionaries do have more structure.
  - But have plenty of entropy too.
  - Still … should try to do without it.

Deterministic Compression?

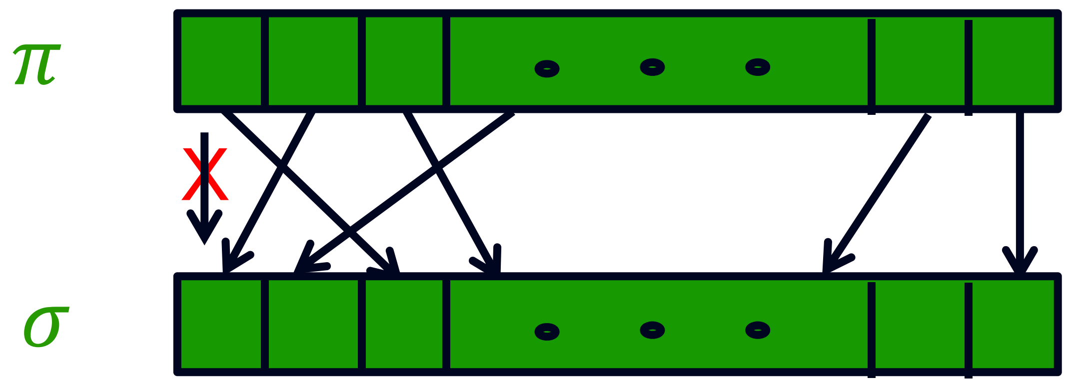
- Randomness fundamental to solution.
  - Needs R independent of P, Q to work.
- Can there be a deterministic solution?
  - Technically: Hard to come up with single scheme that compresses consistently for all (P, Q).
  - Conceptually: Nicer to know “dictionary” and context can be interdependent.

Challenging special case

- Alice has permutation π on $[N]$
  - i.e., π 1-1 function mapping $[N] \to [N]$
- Bob has permutation σ
- Know both are close:
  - $\forall\, m \in [N], |\pi^{-1}(m) - \sigma^{-1}(m)| \leq \ell$ (say $\ell = 2$)
- Alice and Bob know i (say $i = 1$).
  - Alice wishes to communicate $m = \pi(i)$ to Bob.
- Can we do this with few bits?
  - Say $O(1)$ bits if $i = 1$, $\ell = 2$.

Model as a graph coloring problem

- Consider family of graphs $U_{N,\ell}$:
  - Vertices = permutations on $[N]$
  - Edges = close permutations with distinct messages. (two potential Alices).



- Central question: What is $\chi(U_{N,\ell})$?

Main Results [w. Elad Haramaty]

- Claim: Compression length for toy problem

$$\in \left[\log \chi\left(U_{N,\ell}\right), \log \chi\left(U_{N,2\ell}\right)\right]$$

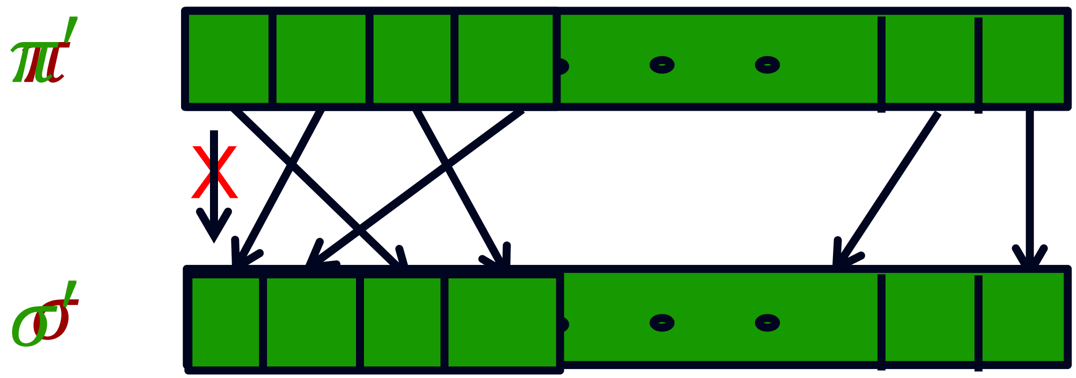
- Thm 1: $\chi\left(U_{N,\ell}\right) \leq \ell^{O(\ell \log^* N)}$
  - $\log^{(i)} N \equiv \log\log \ldots N$ (i times)
  - $\log^* N \equiv \min \{i \mid \log^{(i)} N \leq 1\}$.
- Thm 2: $\exists$ uncertain comm. schemes with
  1. $\mathrm{Exp}_m[\,|E(P,m)|] \leq O(H(P) + \Delta + \log\log N)$ (0-error).
  1. $\mathrm{Exp}_m[\,|E(P,m)|] \leq \ell^{O(\epsilon^{-1}(H(P)+\Delta+\log^* N))}$ (ϵ -error).

- Rest of the talk: Graph coloring

Restricted Uncertainty Graphs

- Will look at $U_{N,\ell,k}$
  - Vertices: restrictions of permutations to first k coordinates.
  - Edges: $\pi' \leftrightarrow \sigma'$

$$\Leftrightarrow \exists\, \pi \text{ extending } \pi' \text{ and } \sigma \text{ extending } \sigma' \text{ with } \pi \leftrightarrow \sigma$$



Homomorphisms

- G homomorphic to H ($G \to H$) if
  $\exists\ \phi: V(G) \to V(H)$ s.t. $u \leftrightarrow_G v \Rightarrow \phi(u) \leftrightarrow_H \phi(v)$
- Homorphisms?
  - G is k-colorable $\Leftrightarrow G \to K_k$
  - $G \to H$ and $H \to L \Rightarrow G \to L$
- Homomorphisms and Uncertainty graphs.
  - $U_{N,\ell} = U_{N,\ell,N} \to U_{N,\ell,N-1} \to \cdots \to U_{N,\ell,\ell+1}$
- Suffices to upper bound $\chi\left(U_{N,\ell,k}\right)$

Chromatic number of $U_{N,\ell,\ell+1}$

- For $f: [N] \to [2\ell]$, Let
  $\mathrm{I}_f = \{\pi \mid f(\pi_1) = 1,\ f(\pi_i) \neq 1,\ \forall\, i \in [2\ell] - \{1\}\}$
- Claim: $\forall f, I_f$ is an independent set of $U_{N,\ell,\ell+1}$
- Claim: $\forall \pi,\ \Pr_f \left[\, \pi \in I_f \,\right] \geq \frac{1}{4\ell}$
- Corollary: $\chi\left(U_{N,\ell,\ell+1}\right) \leq O(\ell^2 \log N)$

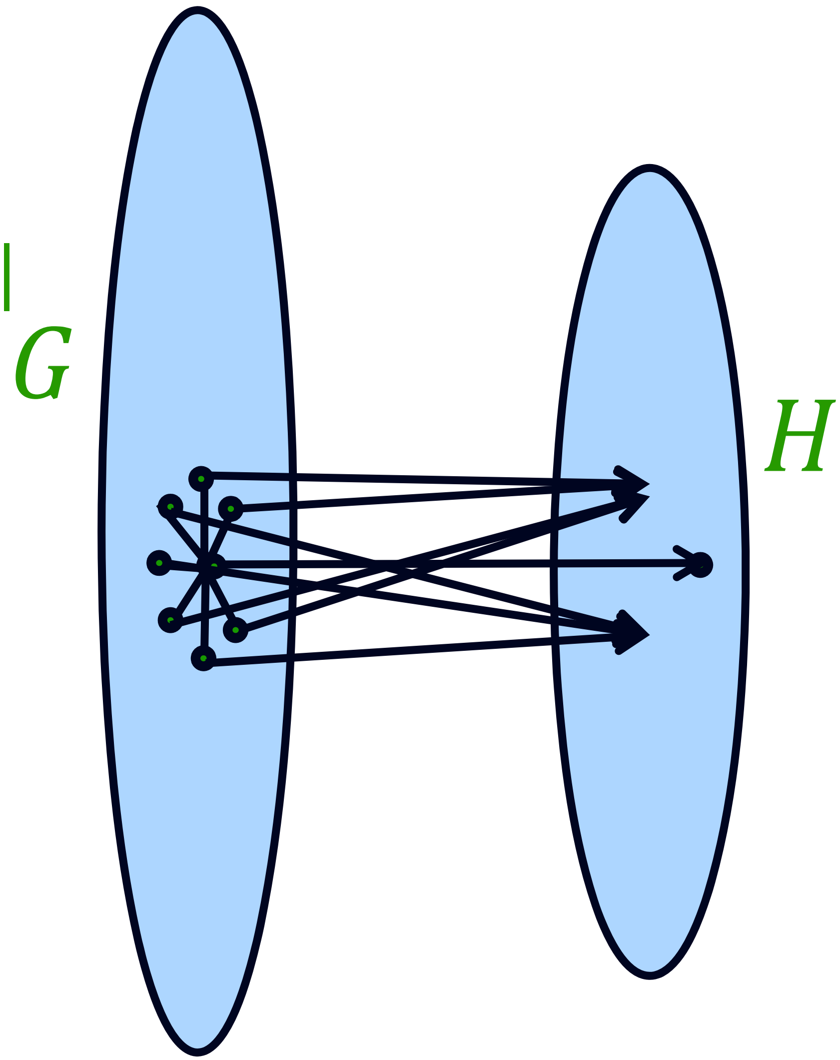
Better upper bounds:

- Say $\phi: G \to H$

- $d_\phi(u) \equiv |\{ \phi(v) \mid v \leftrightarrow_G u \}|$
  $$d_\phi \equiv \max_u \{d_\phi(u)\}$$

- Lemma:
  $$\chi(G) \leq O(d_\phi^2 \log \chi(H))$$

- For $\phi_k: U_{N,\ell,k} \to U_{N,\ell,k-\ell}$
  $$d_{\phi_k} = \ell^{O(k)}$$



Better upper bounds:

- $d_\phi \equiv \max_u \{ |\{\phi(v) | v \leftrightarrow_G u\}| \}$
- Lemma: $\chi(G) \leq O\left(d_\phi^2 \log \chi(H)\right)$
- For $\phi_k : U_{N,\ell,k} \to U_{N,\ell,k-\ell}, \quad d_{\phi_k} \leq \ell^{O(k)}$
- Corollary: $\chi\left(U_{N,\ell,k}\right) \leq \ell^{O(k)} \log^{\left(\frac{k}{\ell}\right)} N$

- Aside: Can show: $\chi\left(U_{N,\ell,k}\right) \geq \log^{\Omega\left(\frac{k}{\ell}\right)} N$
  - Implies can't expect simple derandomization of the randomized compression scheme.

Future work?

- Open Questions:
  - Is $\chi(U_{N,\ell}) = O_\ell(1)$?
  - Can we compress arbitrary distributions to $O(H(P) + \Delta)$? $O(H(P) + \Delta + \log^* N)$? or even $O(H(P) + \Delta + \log\log\log N)$?
- On conceptual side:
  - Better mathematical understanding of forces on language.
    - Information-theoretic
    - Computational
    - Evolutionary

Thank You