Communication Complexity (Spring 2026) Spring 2026

Information Complexity

Instructor: *Toniann Pitassi* Presenter: *Hans Encarnacion*

1 Overview

The main goal of these notes is to prove that any randomized protocol that solves DISJ_n (set disjointness) requires $\Omega(n)$ bits of communication. We prove this using Information Complexity rather than Communication Complexity.

The proof proceeds in three steps:

1. Define entropy, mutual information, and two key properties: the chain rule and subadditivity.
2. Define a distribution ζ, reduce DISJ_n to AND_2, and show that every correct AND_2 protocol must reveal $\Omega(1)$ bits of information.
3. Conclude an $\Omega(n)$ lower bound for randomized protocols solving DISJ_n.

2 Entropy

Definition 1 (Entropy)**.** Let Ω be a finite set and P a probability distribution over Ω. For a random variable X taking values in Ω according to P, the *entropy* of X is

$$H(X) = \sum_{x \in \Omega} P(x) \log \frac{1}{P(x)}$$

All logarithms are base 2. We use the convention $0 \log(1/0) = 0$.

Intuitively, $H(X)$ measures uncertainty about X: $H(X) = 0$ means the outcome is determined; larger values mean more uncertainty.

Example 1 (Fair coin)**.** $\Omega = \{Heads, Tails\}$, $P(Heads) = P(Tails) = 1/2$

$$H(X) = \tfrac{1}{2} \log(2) + \tfrac{1}{2} \log(2) = 1 \text{ bit}$$

This makes sense – a fair coin flip carries 1 bit of uncertainty.

Example 2 (Biased coin)**.** $P(Heads) = 99/100$, $P(Tails) = 1/100$

$$H(X) = \tfrac{99}{100} \log(\tfrac{100}{99}) + \tfrac{1}{100} \log(100) \approx 0.0808 \text{ bits}$$

Low entropy reflects little uncertainty.

Definition 2 (Conditional entropy)**.**

$$H(X \mid Y) = \mathbb{E}_y\big[H(X \mid Y = y)\big], \quad \text{where } H(X \mid Y = y) = \sum_x P(x \mid y) \log \frac{1}{P(x \mid y)}$$

$H(X \mid Y)$ is the average uncertainty about X after observing Y.

Definition 3 (Joint entropy)**.**

$$H(X, Y) = H(X) + H(Y \mid X).$$

$H(X, Y)$ is the uncertainty about X plus the average uncertainty about Y once you already know X.

Example 3 (Independent coin flips)**.** X and Y are independent fair coins.

$$H(X) = 1, \quad H(Y \mid X) = 1, \quad H(X, Y) = 2 \text{ bits.}$$

Knowing X tells you nothing about Y.

Example 4 (Perfectly correlated coins)**.** $X = Y$ (both heads or both tails), each uniform.

$$H(X) = 1, \quad H(Y \mid X) = 0, \quad H(X, Y) = 1 \text{ bit.}$$

Knowing X determines Y exactly.

3 Mutual Information

Definition 4 (Mutual information)**.**

$$I(X\,;\,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X).$$

$I(X\,;\,Y)$ measures how much knowing Y reduces uncertainty about X. If Y reveals nothing about X then $I(X\,;\,Y) = 0$; if Y determines X completely then $I(X\,;\,Y) = H(X)$.

Continuing the examples above:

- Two independent fair coins: $I(X\,;\,Y) = 1 - 1 = 0$ bits.
- Two perfectly correlated coins $X = Y$: $I(X\,;\,Y) = 1 - 0 = 1$ bit.

4 Two Key Properties

Fact 1 (Chain rule for entropy)**.** $H(XY) = H(Y) + H(X \mid Y)$.

Fact 2 (Chain rule for mutual information)**.** $I(XY\,;\,Z) = I(X\,;\,Z) + I(Y\,;\,Z \mid X)$.

Fact 3 (Subadditivity of entropy)**.** $H(XY) \leq H(X) + H(Y)$,

Fact 4 (Conditioning never increases entropy)**.** $H(X \mid Y) \leq H(X)$

Notation: We write $H(XY)$ for $H(X, Y)$; Please note that XY denotes the pair, not a product of X and Y.

5 Information Cost of a Randomized Protocol

In a randomized communication protocol P, Alice receives input x and private random bits r_A. Bob receives input y and private random bits r_B. Each input pair (x, y) therefore produces a distribution of transcripts $\Pi(x, y)$ with one transcript per choice of (r_A, r_B).

Definition 5 (Information cost)**.** The information cost of a protocol P is:

$$\mathrm{IC}(P) = I\big((X, Y)\,;\,\Pi(X, Y)\big)$$

which is the mutual information between the distribution of inputs (X, Y) and the distribution of transcripts $\Pi(X, Y)$. In other words, the information cost of a protocol P tells us how much seeing the transcript reduces our uncertainty about the inputs.

Note, for proving lower bound of DISJ_n, (X, Y) are drawn from a correlated distribution (not a product distribution) which will be essential for our proof.

Fact 5 (Key inequality)**.** $\mathrm{IC}(P) \leq |\text{bits communicated by } P|$

A transcript is just a sequence of bits. Each bit communicated can reveal at most one bit of information. So information cost of a protocol P provides a lower bound for communication cost of P.

6 Set Disjointness

Definition 6 (Set disjointness)**.**

$$\mathrm{DISJ}_n(x, y) = \begin{cases} 1 & \text{if } x \text{ and } y \text{ share an element} \\ 0 & \text{if } x \text{ and } y \text{ share nothing (are disjoint)} \end{cases}$$

Sets are represented as bit vectors of length n; position i is 1 iff element i is in the set.

Example. $n = 6$, $S_A = \{1, 3, 5\} \Rightarrow x = (1, 0, 1, 0, 1, 0)$; $S_B = \{2, 3, 6\} \Rightarrow y = (0, 1, 1, 0, 0, 1)$. Both contain 3, so $\mathrm{DISJ}_6(x, y) = 1$.

6.1 Reduction from DISJ_n to AND_2

$\mathrm{DISJ}_n(x, y) = 1$ iff there exists a coordinate i with $x_i = 1$ and $y_i = 1$, i.e. $\mathrm{AND}_2(x_i, y_i) = 1$. Formally:

$$\mathrm{DISJ}_n(x, y) = \bigvee_{i=1}^{n} (x_i \wedge y_i)$$

Thus DISJ_n is equivalent to running AND_2 on n independent coordinate pairs.

7 The Distribution ζ

7.1 One-coordinate distribution

Definition 7 (Distribution ζ)**.** Sample one coordinate (A_i, B_i, D_i) from ζ over $\{0, 1\}^2 \times \{a, b\}$ as follows:

- Choose D_i uniformly from $\{a, b\}$.
- If $D_i = a$: set $A_i = 0$ and choose B_i uniformly from $\{0, 1\}$.
- If $D_i = b$: set $B_i = 0$ and choose A_i uniformly from $\{0, 1\}$.

Conditioned on D_i, the bits A_i and B_i are independent.

The marginal probabilities on (A_i, B_i) are:

$$\Pr[(A_i, B_i) = (0,0)] = \tfrac{1}{2}, \quad \Pr[(A_i, B_i) = (0,1)] = \tfrac{1}{4}, \quad \Pr[(A_i, B_i) = (1,0)] = \tfrac{1}{4}, \quad \Pr[(A_i, B_i) = (1,1)] = 0.$$

In particular, $A_i \wedge B_i = 0$ always under ζ.

7.2 Product distribution $\zeta^{\otimes n}$

Draw n independent copies: $(A_1, B_1, D_1), \ldots, (A_n, B_n, D_n)$ and concatenate:

$$A = (A_1, \ldots, A_n), \quad B = (B_1, \ldots, B_n), \quad \vec{D} = (D_1, \ldots, D_n).$$

Under $\zeta^{\otimes n}$, every possible (A, B) is disjoint, since $A_i \wedge B_i = 0$ for every i.

7.3 Why $\zeta^{\otimes n}$ is useful

A protocol that only needed to work on $\zeta^{\otimes n}$ could trivially output 0. But a correct protocol must work on *all* inputs, including those where some coordinate is $(1, 1)$. Hence even on disjoint inputs from $\zeta^{\otimes n}$, the transcript must reveal enough information to distinguish those from intersecting inputs. Moreover, after conditioning on $\vec{D}$, the coordinate pairs (A_i, B_i) become independent, enabling us to analyze these coordinates one by one for the proof for lower bound of DISJ_n.

8 Conditional Information Cost and Its Lower Bound

Let $\Pi(A, B)$ denote the transcript of the DISJ_n protocol on inputs A, B. The quantity we analyze is

$$I\Big(AB\,;\,\Pi(A,B) \mid \vec{D}\Big)$$

which measures how much the transcript reveals about Alice and Bob's inputs given $\vec{D}$. The chain of inequalities

$$I\Big(AB\,;\,\Pi(A,B) \mid \vec{D}\Big) \;\le\; H(\Pi(A,B)) \;\le\; |\text{bits communicated}|$$

shows that a large information lower bound implies a large communication lower bound. Thus, proving an information lower bound for DISJ_n proves a lower bound for bits communicated for any randomized protocol that solves DISJ_n.

9 Information Cost Decomposition Lemma

Lemma 1 (Information cost decomposition)**.**

$$I\Big(AB\,;\,\Pi(A,B)\mid\vec{D}\Big)\;\geq\;\sum_{i=1}^{n}I\Big(A_iB_i\,;\,\Pi(A,B)\mid\vec{D}\Big)\,.$$

This means the conditional information revealed by the transcript about Alice and Bob's full n-bit inptus A and B is at least the sum of the conditional information revealed about each coordinate pair (A_i, B_i).

Let's prove this lemma below:

Proof. Abbreviate $\Pi = \Pi(A,B)$. By definition of mutual information:

$$I\Big(AB\,;\,\Pi\mid\vec{D}\Big)=H\Big(AB\mid\vec{D}\Big)-H\Big(AB\mid\Pi\vec{D}\Big)\,.$$

Since the coordinate pairs $(A_1,B_1),\ldots,(A_n,B_n)$ are independent given $\vec{D}$,

$$H\Big(AB\mid\vec{D}\Big)=\sum_{i=1}^{n}H\Big(A_iB_i\mid\vec{D}\Big)\,.$$

By subadditivity of entropy,

$$H\Big(AB\mid\Pi\vec{D}\Big)\leq\sum_{i=1}^{n}H\Big(A_iB_i\mid\Pi\vec{D}\Big)\,.$$

Combining:

$$I\Big(AB\,;\,\Pi\mid\vec{D}\Big)\geq\sum_{i=1}^{n}H\Big(A_iB_i\mid\vec{D}\Big)-\sum_{i=1}^{n}H\Big(A_iB_i\mid\Pi\vec{D}\Big)=\sum_{i=1}^{n}I\Big(A_iB_i\,;\,\Pi\mid\vec{D}\Big)\,.$$

Thus, we have proved the above lemma.

10 Reduction Lemma (DISJ_n to AND_2)

Lemma 2 (Reduction lemma)**.** Let Π compute DISJ_n with error at most ε on every input. Let $((A,B),\vec{D})\sim\zeta^{\otimes n}$ and $((U,V),D)\sim\zeta$. Then for every coordinate i:

$$I\Big(A_iB_i\,;\,\Pi(A,B)\mid\vec{D}\Big)\;\geq\;\inf_{P}I(UV\,;\,P(U,V)\mid D)\,,$$

where the infimum is over all protocols P that compute AND_2 with error at most ε.

Note: Infimum means the greatest lower bound. Here, it means we look over all correct AND_2 protocols and ask: what is the smallest information cost any of them can achieve? Then pick that smallest information cost.

Combining both lemmas (Information Cost Decomposition Lemma and Reduction Lemma) gives:

$$I\Big(AB\,;\,\Pi(A,B)\mid\vec{D}\Big)\;\geq\;n\cdot\inf_{P}I(UV\,;\,P(U,V)\mid D)\,.$$

It therefore suffices to prove a *constant* lower bound for AND_2.

11 Hellinger Distance

Hellinger distance provudes us a way to measure how different two probability distributions p and q are from each other.

Definition 8 (Squared Hellinger distance)**.** For distributions p and q over the same finite set,

$$h^2(p,q) = 1 - \sum_x \sqrt{p(x) \cdot q(x)}$$

$h^2(p,q) \in [0,1]$; it equals 0 iff $p = q$ and 1 iff p and q are completely different. We use three properties of Hellinger distance.

Fact 6 (Information $\geq$ Hellinger)**.** For any protocol P and single-bit input uniform $Z \in \{0,1\}$,
$I(Z\,;\,\Pi(Z)) \geq h^2(p_0, p_1)$, where p_z is the transcript distribution on input $Z = z$.

Fact 7 (Cut-and-paste)**.** For any protocol P and inputs x, x', y, y': $h(p_{x,y}, p_{x',y'}) = h(p_{x,y'}, p_{x',y})$.

Fact 8 (Distinguishing lemma)**.** If P computes a function f (like AND_2) with error at most ε and inputs (u,v) and (u',v') have different f-values, then $h^2(p_{u,v}, p_{u',v'}) \geq 1 - 2\sqrt{\varepsilon}$.

Fact 9 (Triangle Inequality for Hellinger Distance)**.** $h(p,r) \leq h(p,q) + h(q,r)$ where p, q, r are probability distributions.

12 Proving $\mathrm{IC}(\mathrm{AND}_2) = \Omega(1)$

We fix a protocol P computing AND_2 with error at most ε, and show

$$I(UV\,;\,P(U,V) \mid D) \;\geq\; \tfrac{1}{4}(1 - 2\sqrt{\varepsilon}).$$

Step 1: Define the distribution ζ

Under ζ with D uniform over $\{a,b\}$:

if $D = a$ then $U = 0$ and V is uniform over $\{0,1\}$

if $D = b$ then $V = 0$ and U is uniform over $\{0,1\}$

Let Z be uniform over $\{0,1\}$. Then:

$$I(UV\,;\,P(U,V) \mid D) = \tfrac{1}{2}\, I(Z\,;\,P(0,Z)) + \tfrac{1}{2}\, I(Z\,;\,P(Z,0)).$$

Step 2: Apply the Information $\geq$ Hellinger property.
Let p_{uv} denote the transcript distribution when P runs on input (u,v). By Fact 1 (Information $\geq$ Hellinger):

$$I(Z\,;\,P(0,Z)) \geq h^2(p_{00}, p_{01}), \qquad I(Z\,;\,P(Z,0)) \geq h^2(p_{00}, p_{10}).$$

Hence:

$$I(UV\,;\,P(U,V)\mid D)\;\geq\;\tfrac{1}{2}\,h^2(p_{00},p_{01})+\tfrac{1}{2}\,h^2(p_{00},p_{10}).$$

Step 3: Apply Cauchy-Schwarz and the triangle inequality.
By Cauchy-Schwarz inequality $\frac{a^2+b^2}{2}\geq\frac{(a+b)^2}{4}$:

$$\tfrac{1}{2}\,h^2(p_{00},p_{01})+\tfrac{1}{2}\,h^2(p_{00},p_{10})\;\geq\;\tfrac{1}{4}\big(h(p_{00},p_{01})+h(p_{00},p_{10})\big)^2.$$

By the triangle inequality for Hellinger distance $h(p_{01},p_{00})+h(p_{00},p_{10})\geq h(p_{01},p_{10})$:

$$I(UV\,;\,P(U,V)\mid D)\;\geq\;\tfrac{1}{4}\,h^2(p_{01},p_{10}).$$

Step 4: Apply the Cut-and-Paste lemma.
By Fact 7: $h(p_{01},p_{10})=h(p_{00},p_{11})$, so:

$$I(UV\,;\,P(U,V)\mid D)\;\geq\;\tfrac{1}{4}\,h^2(p_{00},p_{11}).$$

Step 5: Apply the Distinguishing Lemma.
Since $\mathrm{AND}_2(0,0)=0$ and $\mathrm{AND}_2(1,1)=1$, Fact 8 gives $h^2(p_{00},p_{11})\geq 1-2\sqrt{\varepsilon}$. Therefore:

$$\boxed{I(UV\,;\,P(U,V)\mid D)\;\geq\;\tfrac{1}{4}(1-2\sqrt{\varepsilon}).}$$

Since this holds for every correct AND_2 protocol with error at most ϵ. Therefore it holds for the protocol that achieves the minimum information cost and therefore:

$$\mathrm{IC}(\mathrm{AND}_2)=\inf_P I(UV\,;\,P(U,V)\mid D)=\Omega(1).$$

13 Conclusion: $R_\varepsilon(\mathrm{DISJ}_n)=\Omega(n)$

Theorem 10. Any randomized protocol that solves DISJ_n with error at most $\varepsilon<1/4$ requires $\Omega(n)$ bits of communication.

Proof. Combining the Information Cost Decomposition Lemma and the Reduction Lemma:

$$I\Big(AB\,;\,\Pi(A,B)\mid\vec{D}\Big)\;\geq\;n\cdot\inf_P I(UV\,;\,P(U,V)\mid D)\;\geq\;n\cdot\tfrac{1}{4}(1-2\sqrt{\varepsilon})\;=\;\Omega(n).$$

Since the number of bits communicated is at least the information revealed, $R_\varepsilon(\mathrm{DISJ}_n)=\Omega(n)$.

Variable glossary.

- $A=(A_1,\ldots,A_n)$: Alice's n-bit input to DISJ_n.
- $B=(B_1,\ldots,B_n)$: Bob's n-bit input to DISJ_n.
- $\Pi(A,B)$: transcript distribution of the DISJ_n protocol.
- $\vec{D}=(D_1,\ldots,D_n)$: helper labels; $D_i\in\{a,b\}$ indicates which player's bit is forced to 0 at coordinate i.

- U: Alice's one-bit input for AND_2.
- V: Bob's one-bit input for AND_2.
- D: the single helper label for the one-coordinate AND_2 problem.