COMS-E6998

Information Theory in Theoretical Computer Science

Wednesdays, 10:10 - 12:00 pm.

Lecturer: Omri Weinstein

omri@cs.columbia.edu
http://www.cs.columbia.edu/ omri/
CSB 523
Office Hours: Tuesdays, 3-4 pm , CSB 523.

TA: Zhenrui Liao

zhenrui.liao@columbia.edu

The following syllabus contains the main topics that will be covered in the course, and is subject to modifications depending on the progress we make .

Course Description: Information theory plays a key role in coding theory, compression, ML and digital communication, but it is also a powerful mathematical tool for analyzing computation in various models, and found a myriad of applications in combinatorics and theoretical computer science.
The primary goal of this course is to develop tools in information theory and communication complexity, and use them to prove upper and lower bounds on important problems in theoretical CS, such as linear programs, interactive compression, streaming algorithms, data structures and locally-decodable codes.
This is an advanced course geared towards CS and EE graduate students, though it is designed to be self contained. Evaluation is based on home works and a final project (reading, implementation, or research). The class satisfies the track electives for Theory (Foundations) track.

This is an advanced course geared towards CS and EE graduate students, though it is designed to be self contained. Evaluation is based on home works and a final project (reading, implementation, or research). The class satisfies the track electives for Theory (Foundations) track.

Prerequisite(s): There are no mandatory prerequisites other than familiarity with probability theory and linear algebra. Background in Complexity Theory (e.g., COMS 4236) and information theory is recommended but not a prerequisite. However, mathematical maturity is a must, and lectures are based on theoretical ideas and will be proof-heavy. Students are expected to be able to read and write formal mathematical proofs.

Credit Hours: 3

Text(s):

1) *Elements of Information Theory*, T.Cover.
2) *Communication Complexity*, A. Rao and A.Yehudayoff (link can be found in Anup Rao's website)

Tentative Syllabus :

1. **Lecture 1: Introduction and motivation for this course.**
   Complexity theory and the holy grail of unconditional lower bounds (von-Neumann information bottleneck). Shannon's information theory as: (i) a tool for analyzing communication problems, and (ii) the underlying theory of coding, ditributed storage and modern distributed system. Course structure (alternating between *techniques* (building the theory) and *applications*).
2. **Lecture 2: Entropy and source coding.**
   Introduction to Entropy, axiomatic definition and operational meaning (Shannon's noiseless-coding thm), prefix-free codes and Kraft's inequality, Shannon-Fano (one-shot) code and Huffman code ("one-shot" compression), entropy lower bound on *expected* encoding length, "on-the-fly" arithmetic codes for the amortized case.
3. **Lecture 3: Conditional entropy and Shearer's Lemma.**
   Joint entropy, conditional entropy and its properties (chain rule, subadditivity etc.). Brief discussion of the (amortized) Slepian-Wolf Thm. Bergman's Thm (counting perfect matchings in regular graphs). Shearer's lemma and applications to counting subgraph embeddings in general graphs.
4. **Lecture 4: Divergence, mutual information and basic inequalities.** Mutual Information, KL divergence, their relationship, operational interpretation and properties (chain rule, convexity, non-negativity etc.), applications to binomial tail and the Chernoff bound. KL vs Statistical distance, operational meaning (rejection sampling) and Pinsker's inequality. Fano's inequality, data-processing inequality and some simple applications.
5. **Lecture 5: Deterministic Communication Complexity.**
   Exposition of various two-party communication models, canonical problems (Equality, Disjointness, Greater-Than, IP), motivation from VLSI lower bounds (Thompson's theorem). Deterministic protocol trees, combinatorial rectangles, fooling sets, rank LB (with application to Disjointness), the log-rank conjecture. Time permitting: Nondeterministic communication complexity and application to LP lower bounds – an $n^{\Omega(\lg n)}$ LB on the size of the vertex-packing polytope via the Clique-vs-Independent-Set problem. Nonnegative Rank Factorization via Common Information (Wyner's Theorem).
6. **Lecture 6: Randomized Communication Complexity.**
   Randomized and distributional CC, private vs. public coin protocols, Equality as a motivating example, Yao's minimax theorem, Newman's Thm (tight example: small-set disjointness). Randomized LB techniques: Discrepancy LB (and application to IP_n), introduction to Information Complexity.
7. **Lecture 7: Interactive Compression and Direct Sums.**
   The direct sum (and product) conjectures, connection to one-shot interactive compression. Sketch of bounded-round compression [BR10] and an interactive analogue of Shannon's noiseless coding theorem. State-of-art interactive compression protocols.
8. **Lecture 8: Information Cost, Hellinger Distance and a Randomized LB for Disjointness.**

Information Complexity and information cost, subadditivity of IC. Hellinger distance, geometric interpretation and properties ("cut-and-paste" lemma, relation to statistical an KL distance). History of Disjointness, a sketch of [BFK]'s $\Omega(\sqrt{n})$ lower bound for *product* distributions. An $\Omega(n)$ lower bound via information complexity and Hellinger [BJKS].

9. **Lecture 9: Applications: Streaming Lower Bounds**
   Introduction to the turnstile streaming model, norm estimation as a motivating example (art-gallery problem and distributed functional monitoring). The low-space AMS algorithm for ℓ_2 estimation. A simple space LB for randomized *exact* ℓ_∞ computation via randomized Set Disjointness. An $\Omega(n/c^2)$ space LB for *c-approximating* ℓ_∞, via an information-complexity LB on Gap-ℓ_∞. Hellinger distance and the "Z Lemma" for protocols.
10. **Lecture 10: Intro to the Cell Probe model and Predecessor Search.**
    Motivation (Near-Neighbor, Range Counting), the Static and Dynamic cell-probe models. Holy grail LBs and existing (static) frontiers. Asymmetric communication complexity and round-elimination. Lower and upper bounds for Predecessor search (vEB tress, fusion trees + hashing, a round-elimination proof via information cost).
11. **Lecture 11: The Cell-Sampling Method, Dictionaries and Succinct Data Structures.**
    Polynomial Evaluation LB via cell-sampling FKS Dictionaries and sketch of [DPT]'s succinct DS. The sparse membership problem, linear tradeoff via arithmetic coding, Patrascu's exponential tradeoff. Patrascu-Viola matching LB. Locally-Decodable source coding for *correlated* files (dictionaries with non-product priors).
12. **Lecture 12: Lopsided Set Disjointness and the Cell-Sampling Method.**
    A canonical reduction from data structure protocols to Lopsided Set-Disjointness (time permitting: applications to Partial Match [Madhu's lectures] and Approximate Near-Neighbor.
13. **Lecture 13: Dynamic data structure lower bounds.**
    Existing dynamic LB frontiers. Dynamic partial sums $O(\lg^2 n)$ upper bound via dynamic "range-trees". A matching LB via the Chronogram method (an information-theoretic perspective on [Fredman-Saks] proof). The "Multiphase" conjecture.
14. **Lecture 14: Epilogue.**
    What we didn't cover: channel coding, multiparty communication complexity, MapReduce, algorithmic information theory. Applications to circuit LBs, Neural nets, economics (auctions)