Modern notions of entropy in ergodic theory and representation theory

Tim Austin

Soon after Shannon introduced entropy for discrete random variables in his foundational work on information theory, it found striking applications to ergodic theory in work of Kolmogorov and Sinai. Many variants and other applications have appeared in pure mathematics since, connecting probability, combinatorics, dynamics and other areas.

I will survey a few recent developments in this story. I will focus largely on (i) Lewis Bowen's "sofic entropy", which helps us to study the dynamics of "large" groups such as free groups, and (ii) a cousin of sofic entropy in the world of unitary representations, which leads to new connections with certain random matrix models.

Large deviations and low complexity Gibbs measures

Amir Dembo, Stanford University

I will overview the emerging theory of large deviations for low complexity Gibbs measures, the naive mean field approximation of their partition functions and representing such measures as mixtures of not too many product measures. As time permits, we will consider certain applications, such as to the abundance of specific patterns in sparse random graphs, having many arithmetic progressions in a uniformly chosen random set and the universality of the Potts model on graphs of growing average degrees.

In the Footsteps of Giants

Michelle Effros

One of the many pleasures of working on information theoretic problems is that each of us, in pursuing our work, walks in the footsteps of giants like Katalin Marton. While Kati's initial period of information theory focus ended roughly as mine began, even early on I found her there, again and again, through her work. In this talk I will comment briefly on a few of the places where our paths crossed and talk about some recent work on random access source coding for asynchronous communication scenarios.

On Marton's Inner Bound to the Capacity Region of the Broadcast Channel

Amin Gohari

Abstract

A broadcast channel is a fundamental communication scenario where a single transmitter communicates with multiple receivers. Determining the capacity region of the broadcast channel has been a longstanding open problem in information theory. Marton's inner bound remains the best known inner bound to the capacity region of a general two-receiver broadcast channel.

In this talk, I will provide a high-level overview of the progress that has been made in understanding Marton's inner bound and the capacity region of the two-receiver broadcast channel over the past two decades. Additionally, I will discuss some conjectures and open questions in this area.

On Marton's conjecture

Let A be a subset of V, a vector space over the finite field with two elements. Suppose that $|A + A| \leq K|A|$ for some constant K, where $A + A$ denotes the sumset $\{a_1 + a_2 : a_1, a_2 \in A\}$. Katalin Marton conjectured that A is efficiently covered by cosets of subspaces of V; more precisely, there are cosets $H_1, \ldots, H_m$ of size at most $|A|$ whose union covers A, where m is bounded polynomially in terms of K. This conjecture was popularised by Imre Ruzsa and became a well-known open problem in additive combinatorics.

I will discuss the recent proof of Marton's conjecture, which was given in joint work with Tim Gowers, Freddie Manners and Terence Tao.

B. Green

Entropy, ortho-homomorphisms, and quantum computation

László Lovász

Kati Marton introduced an upper bound on the capacity of probabilistic graphs (defned earlier by Csiszár and Körner), using orthonormal representations, generalizing the bound given for the "ordinary" Shannon capacity by the speaker. Since then, these notions have interacted in a variety of circumstances, including statistical and quantum physics. We give a survey of some of these interesting (solved and unsolved) problems.

Scaling limit of SOS-type surfaces on a slope: From contracting Markov chains to the GFF

Eyal Lubetzky

The $(2+1)$D Solid-On-Solid (SOS) model famously exhibits a roughening transition: on an $N \times N$ torus with the height at the origin rooted at 0, the variance of $h(x)$, the height at a point x is $O(1)$ when the inverse-temperature β is large, vs $O(\log |x|)$ when β is small. The rigidity at large β is believed to fail once the surface is on a slope (tilted boundary conditions), which ought to destabilize it and induce the log-correlated behavior of the small β regime. The only rigorous result on this is by Sheffield (2005): if the slope θ is irrational, then $\mathrm{Var}(h(x))$ diverges with $|x|$ (with no known quantitative bound).

We study this model at a large fixed β, on an $N \times N$ torus with a nonzero boundary condition slope θ, perturbed by a potential V of strength $\epsilon(\beta)$ per site (arbitrarily small). Our main result is (a) the measure on height gradients ∇h has a weak limit μ as $N \to \infty$; and (b) the scaling limit of a sample from μ converges to a full plane GFF. In particular, we recover the asymptotics of $\mathrm{Var}(h(x))$. To our knowledge, this is the first example of a random surface that approximates the 3D Ising model interface, or any perturbation of one, where the scaling limit is recovered at large finite β under tilted boundary conditions.

The proof looks at random monotone surfaces that approximate the SOS surface, and derives their GFF limit via combinatorial and probabilistic tools; we will elaborate on one of these, relying on contraction properties of appropriately constructed Markov chains.

Joint work with Benoît Laslier.

Estimating the mean of random vectors

Gábor Lugosi

One of the many highlights of Kati Marton's career is her brilliant contributions to concentration inequalities. Concentration inequalities appear naturally in statistics and in this talk we discuss the simplest statistical problem where such inequalities are relevant, which is mean estimation. This classical problem has recently attracted a lot of attention both in mathematical statistics and in theoretical computer science. We present some of the recent advances, focusing on high-dimensional aspects.

Bring in the noise

Muriel Médard
MIT

One of the apparent contradictions of information theory is the fact that most codes are good and yet finding codes is challenging. In fact, it has been co-designing error-correcting codes and, most importantly, their generally complex decoders that has proven difficult, particularly when these decoders must be amenable to implementation in efficient, dedicated and customized chips.

In this talk we describe "Guessing Random Additive Noise Decoding," or GRAND, by Duffy, Médard and their research groups, which renders universal, optimal, code-agnostic decoding possible for low to moderate redundancy settings. GRAND enables a new exploration of codes, in and of themselves, independently of tailored decoders, over a rich family of code designs, including random ones. Surprisingly, even the simplest code constructions, such as those used merely for error checking, match or markedly outperform state of the art codes when optimally decoded with GRAND.

GRAND relies on the fact that in most channels of interest noise entropy is far inferior to the entropy of the messages. By taking into account probabilistic information about the noise, for example conditional probabilities based on soft information, or noise correlation, GRAND can use the fact that conditioning or correlation reduce entropy to improve throughput. Simple constructions such as product codes, when component codes are decoded with GRAND, can outperform LDPCs when we consider codes with high redundancy.

Without the need for highly tailored codes and bespoke decoders, we can envisage using GRAND to avoid the issue of limited and sub-optimal code choices and instead have an open platform for coding and decoding. Moreover, recent work with Duffy, Médard, Yazicigil and their groups has demonstrated that such decoding can be implemented with extremely low latency and record-breaking low energy in silicon.

From a Single-letterization Argument to Capacity Regions and Information Inequalities

Chandra Nair
Chinese University of Hong Kong

Körner and Marton employed a single-letterization argument en route to their characterization of "Images of a set via two channels." We will show how this single-letterization "trick" led to determining the capacity region of the two receiver vector Gaussian broadcast channels via establishing the Gaussian optimality of a non-convex information functional. Motivated by this, we will explore how sub-additivity (or single-letterization) helps us establish a family of inequalities that combines the Entropy Power Inequality and the Brascamp-Lieb inequality. Finally, we will also show that these inequalities have an analogous statement in finite Abelian groups.

Set membership with a few classical and quantum probes

Jaikumar Radhakrishnan
International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Shivakote, Bengaluru 560089, Karnataka, India

We consider the following data structure problem. Given an n-element subset S of a universe of size m, represent S in memory as a bit string $x(S)$ so that membership queries of the form "Is x in S?" can be answered with a small number t of bit probes into $x(S)$. Let $s(m, n, t)$ be the minimum number of bits of memory, the length of $x(S)$, needed for this task. We will review the lower and upper bounds that are known for $s(m, n, t)$. We will then focus on the case $t = 2$, and present some recent upper and lower bounds for $s(m, n, t)$ in the classical and quantum settings. The arguments we use will be graph-theoretic, based on (i) dense graphs with large girth, and (ii) a theorem of Nash-Williams on covering the edges of a graph with forests. (The recent results were obtained in joint work with Shyam Dhamapurkar and Shubham Pawar.)

ENTROPY AND SUMSETS

IMRE Z. RUZSA

Katalin Marton may be the first to observe connections between information theory and structure of sumsets. Unfortunately she did not live to see the recent proof of her conjecture by Gowers, Green, Manners and Tao.

I intend to outline the story of this development which benefitted both subjects and whose apex but probably not the end is the above proof.

Alfréd Rényi Institute of Mathematics, Budapest, Pf. 127, H-1364 Hungary

Email address: ruzsa@renyi.hu

Linearization of optimal rates in zero error source and channel coding problems

Nicolas Charpenay[*], Maël Le Treust[†] and Aline Roumy[‡]

31th March 2024

In vanishing error regime, separate encoding of independent problems is optimal. A natural illustration of this property is the *linearization* of optimal rates, which holds for both source and channel coding problems. This property does not hold in the zero error regime. Specifically, in the context of channel coding, Haemers (1979) shows a counterexample based on the Schläfli graph, where joint encoding of independent problems results in higher zero-error capacity C_0 than separate encoding. Determining when separate encoding is optimal, or equivalently when the linearization holds, remains an open problem in the zero error regime. Recently, Schrijver (2023) and (Wigderson and Zuiddam, 2023, Theorem 4.1) show that for all graphs G and G',

$$C_0(G) + C_0(G') = C_0(G \wedge G') \iff \log\left(2^{C_0(G)} + 2^{C_0(G')}\right) = C_0(G \sqcup G'), \tag{1}$$

where $\wedge$ denotes the AND product, and $\sqcup$ denotes the disjoint union of graphs. In Charpenay et al. (2023), we show that the equivalence (1) also holds for the complementary graph entropy $\bar{H}$, which according to Alon and Orlitsky (1996) and Koulgi et al. (2003), determines the optimal rate in the zero error Slepian and Wolf source coding problem.

If these two results appear similar, they differ in that for C_0 the linearization property only depends on characteristic graphs, whereas for $\bar{H}$ it may also depend on the probability distribution on the vertices of the graphs.

In this work, we establish the equivalence between *the linearization of* C_0 *and* $\bar{H}$ when the probability distribution $P_{VV'}$ maximizes a third quantity: *the zero error capacity relative to a distribution* $C(G \wedge G', P_{VV'})$. This crucial notion, introduced in Csiszár and Körner (1981), is related to C_0 via the result of (Gargano et al., 1994, Theorem 2), see also (Simonyi, 2001, Theorem 13.68), and is related to $\bar{H}$ via the result of (Marton, 1993, Lemma 1). We show that for such a distribution on the product of vertices, the linearization properties of C_0, $\bar{H}$ and $C(G, P)$ with respect to the AND product, and to the disjoint union of graphs, are all equivalent.

References

Alon, N. and Orlitsky, A. (1996). Source coding and graph entropies. *IEEE Transactions on Information Theory*, 42(5):1329–1339.

Charpenay, N., Le Treust, M., and Roumy, A. (2023). Complementary graph entropy, and product, and disjoint union of graphs. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 2446–2451.

Csiszár, I. and Körner, J. (1981). On the capacity of the arbitrarily varying channel for maximum probability of error. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(1):87–101.

Gargano, L., Körner, J., and Vaccaro, U. (1994). Capacities: from information theory to extremal set theory. *Journal of Combinatorial Theory, Series A*, 68(2):296–316.

Haemers, W. (1979). On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(2):231–232.

Koulgi, P., Tuncel, E., Regunathan, S. L., and Rose, K. (2003). On zero-error source coding with decoder side information. *IEEE Transactions on Information Theory*, 49(1):99–111.

Marton, K. (1993). On the Shannon capacity of probabilistic graphs. *Journal of Combinatorial Theory, Series B*, 57(2):183–195.

Schrijver, A. (2023). On the Shannon capacity of sums and products of graphs. *Indagationes Mathematicae*, 34(1):37–41.

Simonyi, G. (2001). Perfect graphs and graph entropy. An updated survey. In Reed, B. A. and Alfonsin, J. L. R., editors, *Perfect Graphs*, chapter 13, pages 293–328. John Wiley & Sons.

Wigderson, A. and Zuiddam, J. (2023). *Asymptotic spectra: Theory, applications and extensions.* manuscript.

[*]Univ. Rennes, CNRS, IRMAR UMR 6625, 35000 Rennes, France

[†]Univ. Rennes, CNRS, Inria, IRISA UMR 6074, 35000 Rennes, France

[‡]INRIA Rennes, Campus de Beaulieu, 35042 Rennes cedex, France

The concentration of information phenomenon for convex measures

Mokshay Madiman

The classical Shannon-McMillan-Breiman theorem expresses the fact that data from a stationary ergodic source lives with high probability in a "typical set" that is much smaller than the actual support of the process. In the last decade, this phenomenon has been found to be exhibited already in finite dimension if the probability measure in question lives on Euclidean space and has certain convexity properties. First, Bobkov and the speaker showed that the information content per coordinate of data from a log-concave distribution is highly concentrated; later this was generalized to a wider class of measures (allowing for heavy tails) by Fradelizi, Li and the speaker. This phenomenon turns out to have applications in information theory, probability, and convex geometry. If time permits, we will highlight some interesting open questions in discrete settings.

Exponential Strong Converse Theorems for Source and Channel Networks

Yasutada Oohama

Abstract

For two or multi terminal source or channel coding systems the converse coding theorems state that at any transmission rates exceeding the fundamental theoretical limit of the system the error probability of decoding *can not go to zero* when the block length n of the codes tends to infinity. On the other hand the strong converse theorems state that at any transmission rates exceeding the fundamental theoretical limit the error probability of decoding *must go to one* when n tends to infinity. Specifically, if the error probability of decoding tends to one or equivalently the correct probability of decoding tends to zero exponentially, we say that we have the exponential strong converse theorem. In this talk, we introduce the author's previous works on the exponential strong converse theorem source or channel networks. In those previous works the author propose a new unified method to prove strong converse theorems for several multitermial source or channel networks. The new method consists of two processes shown below:

1. Multi-letter expression on the lower bound of the optimal exponent function is expressed in the form of a max-min optimization on the moment generating function with respect to the information spectral quantity.
2. In the single-letterization of the multi-letter lower bound we introduce a new method called the recursive method.

The second process is the most important and original part in establishing the exponential strong converse theorems.

Y. Oohama is with University of Electro-Communications, 1-5-1 Chofugaoka Chofu-shi, Tokyo 182-8585, Japan.

Exponential strong converse revisited

Shun Watanabe

The strong converse for a coding theorem claims that the optimal asymptotic rate possible with vanishing error probability cannot be improved by allowing a fixed error probability. In order to prove such a claim for multi-user information theory problems, such as the source coding with coded side-information, Ahlswede, Gács, and Körner introduced the tool termed the blowing-up lemma. Later, Marton provided a simple proof of the blowing-up lemma, which is now known as an important methodology to prove measure concentration. The exponential strong converse further claims that, if a coding rate is beyond the asymptotic limit, the cerect decoding probability converges to zero exponentially. Even though the tight strong converse exponent for point-to-point problems have been identified, the strong converse exponent for multi-user problem have been unsolved until recently. In this talk, we present the recently obtained result by Takeuchi and Watanabe providing the tight exponential strong converse for the source coding with coded side-information. Instead of the blowing-up lemma, the proof is based on manipulations of information quantities as in the weak converse argument (called "change-of-measure argument"). We also discuss connection to Marton's work.