Algorithmic Aspects of Information Theory (Dagstuhl Seminar 22301)

Phokion G. Kolaitis, Andrej E Romashchenko, Milan Studený, Dan Suciu, Tobias A. Boege

Report from Dagstuhl Seminar 22301

Algorithmic Aspects of Information Theory

Phokion G. Kolaitis[*1], Andrej E. Romashchenko[*2], Milan Studený[*3], Dan Suciu[*4], and Tobias A. Boege[†5]

1 **University of California – Santa Cruz, US & IBM Research, US.**
kolaitis@ucsc.edu
2 **University of Montpellier – LIRMM, FR & CNRS, FR.**
andrei.romashchenko@lirmm.fr
3 **The Czech Academy of Sciences – Prague, CZ.** studeny@utia.cas.cz
4 **University of Washington – Seattle, US.** suciu@cs.washington.edu
5 **MPI für Mathematik in den Naturwissenschaften – Leipzig, DE.**
post@taboege.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22301 "Algorithmic Aspects of Information Theory".

Constraints on entropies constitute the "laws of information theory". These constraints go well beyond Shannon's basic information inequalities, as they include not only information inequalities that cannot be derived from Shannon's basic inequalities, but also conditional inequalities and disjunctive inequalities that are valid for all entropic functions. There is an extensive body of research on constraints on entropies and their applications to different areas of mathematics and computer science. So far, however, little progress has been made on the algorithmic aspects of information theory. In fact, even fundamental questions about the decidability of information inequalities and their variants have remained open to date.

Recently, research in different applications has demonstrated a clear need for algorithmic solutions to questions in information theory. These applications include: finding tight upper bounds on the answer to a query on a relational database, the homomorphism domination problem and its uses in query optimization, the conditional independence implication problem, soft constraints in databases, group-theoretic inequalities, and lower bounds on the information ratio in secret sharing. Thus far, the information-theory community has had little interaction with the communities where these applications have been studied or with the computational complexity community. The main goal of this Dagstuhl Seminar was to bring together researchers from the aforementioned communities and to develop an agenda for studying algorithmic aspects of information theory, motivated from a rich set of diverse applications. By using the algorithmic lens to examine the common problems and by transferring techniques from one community to the other, we expected that bridges would be created and some tangible progress on open questions could be made.

Seminar July 24–29, 2022 – http://www.dagstuhl.de/22301

2012 ACM Subject Classification Mathematics of computing → Information theory; Information systems → Database design and models; Mathematics of computing → Discrete mathematics; Mathematics of computing → Probabilistic inference problems

Keywords and phrases Information theory, Information inequalities, Conditional independence structures, Database query evaluation and containment, Decision problems

Digital Object Identifier 10.4230/DagRep.12.7.180

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Algorithmic Aspects of Information Theory, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 180–204
Editors: Phokion G. Kolaitis, Andrej E. Romashchenko, Milan Studený, and Dan Suciu

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Phokion G. Kolaitis
Andrej E. Romashchenko
Milan Studený
Dan Suciu

The goal of this seminar was to bring together researchers from several communities who share an interest in the methods and the uses of information theory. Participants included experts in information theory, databases, secret sharing, algorithms, and combinatorics. There were four tutorials, two from the information theory community and two from the database community, that helped define a common language and a common set of problems. There were several contributed talks, from experts in all these fields. The proof of one of the major open problems in information theory was announced at the workshop by not one, but, by *two* researchers, namely Cheuk Ting Li and Geva Yashfe, who used quite different techniques to independently solve this open problem. Overall, the workshop was a success.

Organization of the Seminar

The seminar was held between July 25-29, 2022 (Monday to Friday), and had 25 on-site participants, and 8 remote participants. Since the participants represented quite diverse communities, we started the first day with an introduction of each participant. The four tutorials were scheduled during the first two days: two tutorials on information inequalities and conditional independence were given by László Csirmaz and Milan Studený, and two tutorials on different aspects of database theory were given by Marcelo Arenas and Hung Ngo. All four tutorials were very well received, with many questions and lively discussions during and after the tutorials. There were 18 contributed talks in total, spread over all 5 days of the seminar. We scheduled two sessions to discuss open problems: one on Tuesday afternoon, and one on Thursday afternoon. The seminar concluded with an hour-long discussion assessing the seminar and contemplating future directions. Our collector, Tobias Boege, recorded all open problems, and later typed them for inclusion in this report.

Outcomes of the Seminar

There are several major outcomes:

- Having participants with very diverse backgrounds enabled us to exchange interesting ideas and problems. Information theorists became inspired by problems that arise in database research, while database theoreticians learned tools and techniques from information theory; almost all talks raised algorithmic questions that inspired people from the algorithms community.
- We have assembled a list of open problems, which we included here, and we also plan to publish independently. We hope that this list will help define the community interested in the algorithmic aspects of information theory, and will also inspire young researchers to contribute to this emerging area.

- At the end of the workshop the participants expressed a lot of interest in continuing to have some organized forum for discussing problems in information theory. One of us (Andrei Romashchenko) is planning to organize regular talks, to be made publicly available online, via Zoom.
- Everyone was happily surprised that one major problem in information theory was essentially settled during this seminar. The problem asks whether the implication problem for conditional independence statements is decidable. This problem has been studied since at least the early 80's, and has resisted any prior attempts at settling it. Cheuk Ting Li announced a proof of the undecidability of this problem, and presented the high-level structure of the proof; he had posted on arXiv a paper describing the proof just a few weeks before the seminar. Geva Yashfe had solved a different open problem: he showed that it is undecidable whether a given $(2^n - 1)$-dimensional vector is an almost entropic vector. Through discussions at the seminar, he realized that his proof can be extended to also prove that the implication problem for conditional independences is undecidable. He gave a presentation of his proof on the blackboard, during the seminar.

Acknowledgements

We are grateful to the Scientific Directorate and to the staff of the Schloss Dagstuhl – Leibniz Center for Informatics for their support of this seminar. We also wish to express our sincere thanks to Dr. Tobias Boege for collecting the abstracts of the talks and compiling the list of open problems.

2 Table of Contents

3 Overview of Talks

3.1 Open Problems on Information-Theoretic bounds for Database Query Answers

Mahmoud Abo Khamis (relationalAI – Berkeley, US)

Joint work of Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu
Main reference Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu: "What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?", in Proc. of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, 2017, pp. 429–444, ACM, 2017.
URL https://doi.org/10.1145/3034786.3056105

Information theory has been used to derive tighter bounds on the output sizes of database queries as well as to devise matching algorithms that can answer these queries within the derived bounds. Such bounds are typically derived by translating database statistics into constraints over (conditional) entropies. The query output size is then bounded by the maximum joint entropy subject to these constraints as well as Shannon inequalities. The most general form of this class of bounds is called the polymatroid bound [1].

In this talk, we present two open problems related to the polymatroid bound.

1. The first problem asks whether we can make the polymatroid bound stronger by adding conditional independence constraints that can be inferred from the structure of the database query.
2. The second problem asks whether we can utilize the query structure to infer constraints on the multivariate mutual information, in the same way we utilize (conditional) independence to infer a zero-constraint on the (conditional) mutual information between two sets of variables.

References

1 Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu: What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another? In Emanuel Sallinger, Jan van den Bussche, Floris Geerts (editors): Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, pp. 429–444 (2017)

3.2 Tutorial: a brief introduction to database theory

Marcelo Arenas (PUC – Santiago de Chile, CL)

Main reference Marcelo Arenas, Leonid Libkin: "An information-theoretic approach to normal forms for relational and XML data", J. ACM, Vol. 52(2), pp. 246–283, 2005.
URL https://doi.org/10.1145/1059513.1059519

In this talk, we will give an overview of some fundamental concepts in database theory: relational schemas, queries, data dependencies and normal forms. Besides, we will present a connection between normalization theory and information theory.

3.3 Approximate Implication for PGMs and Relational DBs

Batya Kenig (Technion – Haifa, IL)

Joint work of Batya Kenig, Dan Suciu
Main reference Batya Kenig, Dan Suciu: "Integrity Constraints Revisited: From Exact to Approximate Implication", Log. Methods Comput. Sci., Vol. 18(1), 2022.
URL https://doi.org/10.46298/lmcs-18(1:5)2022
Main reference Batya Kenig: "Approximate implication with d-separation", in Proc. of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI 2021, Virtual Event, 27-30 July 2021, Proceedings of Machine Learning Research, Vol. 161, pp. 301–311, AUAI Press, 2021.
URL https://proceedings.mlr.press/v161/kenig21a.html

The implication problem studies whether a set of conditional independence (CI) statements (antecedents) implies another CI (consequent), and has been extensively studied under the assumption that all CIs hold exactly. In this work, we drop this assumption, and define an approximate implication as a linear inequality between the degree of satisfaction of the antecedents and consequent. More precisely, we ask what guarantee can be provided on the inferred CI when the set of CIs that entailed it hold only approximately. We use information theory to define the degree of satisfaction, and prove several results. In the general case, no such guarantee can be provided. We prove that such a guarantee exists for the set of CIs inferred in directed graphical models, making the d-separation algorithm a sound and complete system for inferring approximate CIs. We also prove an approximation guarantee for independence relations derived from marginal and saturated CIs.

3.4 Recent advances in secret sharing

Amos Beimel (Ben Gurion University – Beer Sheva, IL)

Main reference Benny Applebaum, Amos Beimel, Oded Nir, Naty Peter: "Better secret sharing via robust conditional disclosure of secrets", in Proc. of the Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020, pp. 280–293, ACM, 2020.
URL https://doi.org/10.1145/3357713.3384293

A secret-sharing scheme allows to distribute a secret s among n parties such that only some predefined "authorized" sets of parties can reconstruct the secret s, and all other "unauthorized" sets learn nothing about s. For over 30 years, it was known that any (monotone) collection of authorized sets can be realized by a secret-sharing scheme whose shares are of size $2^{n-o(n)}$ and until recently no better scheme was known. In a recent breakthrough, Liu and Vaikuntanathan [1] have reduced the share size to $2^{0.994n+o(n)}$, and this was further improved by several follow-ups accumulating in an upper bound of $1.5^{n+o(n)}$ [2]. In this talk will survey the known results on secret-sharing schemes and present some ideas of the new constructions.

References

1 Tianren Liu, Vinod Vaikuntanathan: Breaking the circuit-size barrier in secret sharing. STOC 2018: 699-708
2 Benny Applebaum, Oded Nir: Upslices, Downslices, and Secret-Sharing with Complexity of $1.5n$. CRYPTO (3) 2021: 627-655

3.5 Universality of Gaussian conditional independence models

Tobias Andreas Boege (MPI für Mathematik in den Naturwissenschaften, DE)

Main reference T. Boege: "The Gaussian conditional independence inference problem", PhD thesis, Otto-von-Guericke-Universität Magdeburg, 2022.
URL https://dx.doi.org/10.25673/86275

We study statistical models of jointly normal random variables defined by conditional independence (CI) constraints. These models are semialgebraic sets. In this talk I present a number of so-called "universality theorems" for these models:

1. all real algebraic numbers are necessary to witness that a given set of conditional independence constraints is consistent;
2. the problem of deciding consistency (or, equivalently, solving the conditional independence implication problem for Gaussians) is complete for the existential theory of the reals; and
3. all homotopy types of semialgebraic sets are attained by oriented Gaussian CI models.

These results parallel the celebrated universality theorems in matroid theory due to MacLane, Mnëv and Sturmfels.

3.6 Tutorial on information inequalities

László Csirmaz (Alfréd Rényi Institute of Mathematics – Budapest, HU)

The information content of the marginals of (finitely many) jointly distributed random variables reveals many important properties of the distribution, such as functional dependency or (conditional) independence. The information content is measured by the entropy, and information inequalities compare these marginal entropies. The entropy region is the collection of the entropy vectors of these distributions indexed by the non-empty subsets of the variables. The main focus of the talk is on discrete distributions, but many of the methods and concepts apply, with some modification, for linear, continuous, Gaussian, or quantum distributions. Points in the entropy region satisfy the basic Shannon inequalities [3], and the entropy region is bounded by collection of these inequalities, known as the Shannon bound. Points within the Shannon-bound are the also called polymatroids. The first non-Shannon entropy inequality was discovered by Zhang and Yeung [4]. The method obtaining this inequality was formalized and generalized by [1]. The general idea is to find an operation which preserves entropic points, but does not preserve polymatroids in general. Typical operations are restricting, factoring, conditioning, tightening, etc. Unfortunately they preserve both entropic points [2] and polymatroids, but might help in reducing the computational complexity of obtaining bounds on (or parts of) the entropy region. The two-step process of an adequate operation, dubbed as Copy Lemma, can be phrased as follows: first extend the distribution by adding identical copies of some of the variables (this step does not work in the quantum setting because of the no-cloning theorem), and then redefine the distribution so that the new and old variables become conditionally independent given the remaining variables. A known set of inequalities for the larger set of variables (e.g., the basic Shannon inequalities) might imply additional constraints on the old variables. The Copy Lemma can be considered to be

a special case of the Maximal Entropy Method. Fix some marginal distributions on $M > N$ random variables to be identical with certain marginal distributions on N random variables, and take the distribution on M with the maximal possible entropy. This distribution will have strong structural properties: depending on the fixed marginals several conditional independences hold. Harvesting them might yield new constrains on the entropies of the original distribution. The presentation concludes with several computational challenges.

1. Obtaining additional information inequalities requires embedding a distribution on N variables to a distribution on $M > N$ variables, and then applying known information inequalities on M variables. For M larger than 15 even working with the Shannon bounds is prohibitively expensive. Devise a method which simplifies this treatment.
2. Look systematically for limitations of the above methods with small number of added variables.
3. Study how the above technique can be applied for quantum information, and obtain new quantum-information inequalities.
4. A repository of (discrete) distributions with four variables whose convex combination approximates the complete entropic region might be extremely useful in answering practical / theoretical questions.

References

1 R. Dougherty, C. Freiling, K. Zeger, Non-Shannon information inequalities in four random variables, *ArXiv:1104.3602* (2011)
2 F. Matúš, Two constructions on limits of entropy functions, *IEEE Trans. Inform. Theory*, vol 53(1) (2007) pp. 320–330
3 R. W. Yeung, *A first course in information theory*. Kluwer Academic Publishers, New York, 2002
4 Z. Zhang, R. W. Yeung, On characterization of entropy function via information inequalities, *Proc IEEE Trans. Inform. Theory*, vol 44(4) (1998) pp. 1440–1452

3.7 Linear Programming Technique in the Search for Lower Bounds in Secret Sharing

Oriol Farràs (Universitat Rovira i Virgili – Tarragona, ES)

Main reference Oriol Farràs, Tarik Kaced, Sebastià Martín Molleví, Carles Padró: "Improving the Linear Programming Technique in the Search for Lower Bounds in Secret Sharing", in Proc. of the Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part I, Lecture Notes in Computer Science, Vol. 10820, pp. 597–621, Springer, 2018.
URL https://doi.org/10.1007/978-3-319-78381-9_22

Secret sharing scheme is a method by which a dealer distributes shares to parties such that only authorized subsets of parties can reconstruct the secret. The information ratio is an indicator of the efficiency of a secret sharing scheme; it is the size in bits of the largest share of the scheme divided by the size of the secret.

This talk is focused on the following optimization problem: Given a family of subsets of parties F, find the infimum of the information ratio of all secret sharing schemes whose authorized subsets are the ones in F. Lower bounds on this optimal value can be computed by solving linear programming problems involving information inequalities.

We present improvements of this linear programming technique that use the Ahlswede-Körner lemma and the common information of random variables, avoiding the use of explicit non-Shannon information inequalities. Moreover, we show results of the application of this technique to the classification of representable matroids. The results presented in this talk were published in references.

References

1 Michael Bamiloshin, Aner Ben-Efraim, Oriol Farràs, Carles Padró: Common information, matroid representation, and secret sharing for matroid ports. Des. Codes Cryptogr. 89(1): 143-166 (2021).
2 Oriol Farràs, Tarik Kaced, Sebastià Martín Molleví, Carles Padró: Improving the Linear Programming Technique in the Search for Lower Bounds in Secret Sharing. IEEE Trans. Inf. Theory 66(11): 7088-7100 (2020).

3.8 Information Complexity

Yuval Filmus (Technion – Haifa, IL)

Main reference Mark Braverman, Ankit Garg, Denis Pankratov, Omri Weinstein: "From information to exact communication", in Proc. of the Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013, pp. 151–160, ACM, 2013.
URL https://doi.org/10.1145/2488608.2488628

This talk surveys work by IMU Abacus Medal winner Mark Braverman and others.

The basic question asked by information theory is how many bits of communication are needed to transmit information from A to B. In contrast, communication complexity studies the amount of communication needed between two parties, A and B, who want to compute a joint function of their inputs. Information complexity is an approach to communication complexity using information theory, specifically the notion of Information Complexity which is analogous to entropy. Using this approach, it is possible to compute asymptotically tight bounds on communication complexity. For example, Braverman, Garg, Pankratov and Weinstein [1] considered the fundamental problem of Set Disjointness, in which A and B each hold a subset of $\{1, \ldots, n\}$, and their goal is to decide whether the two subsets are disjoint. They showed that the exact communication complexity of this function (with vanishing error) is roughly $0.4825\ldots n$, where 0.4825... is an explicitly computable constant.

References

1 Mark Braverman, Ankit Garg, Denis Pankratov, Omri Weinstein: From information to exact communication. STOC 2013: 151-160

3.9 Dependencies in team semantics

Miika Hannula (University of Helsinki, FI)

Joint work of Miika Hannula, Jonni Virtema
Main reference Miika Hannula, Jonni Virtema: "Tractability frontiers in probabilistic team semantics and existential second-order logic over the reals", Ann. Pure Appl. Log., Vol. 173(10), p. 103108, 2022.
URL https://doi.org/10.1016/j.apal.2022.103108

According to the traditional Tarski's truth definition the semantics of first-order logic is defined with respect to an assignment of values to the free variables. In team semantics, truth is defined with respect to a set (or a probability distribution) of such assignments. This allows modeling concepts that inherently arise only in the presence of multitudes. Examples of concepts available in team semantics, but not in the Tarski semantics, include concepts of dependence and independence. In this talk we will take a brief look at how in team semantics one can analyze the relationships between relational and probabilistic dependencies as well as their interplay with logical operations.

3.10 Entropy Inequalities, Lattices and Groups

Peter Harremoës (Niels Brock Copenhagen Business College, DK)

Main reference Peter Harremoës: "Entropy Inequalities for Lattices", Entropy, Vol. 20(10), p. 784, 2018.
URL https://doi.org/10.3390/e20100784

The notion of entropy inequalities of Shannon and non-Shannon type have mostly been studied for formal power sets of random variables. Such power sets form Boolean lattices with inclusion as ordering. For applications in database theory and the study of Bayesian networks and similar graphical models of independence it is also relevant to study other lattices than the Boolean lattices. In general an element in a lattice should correspond to a set of variables, and one element in the lattice dominates another point if and only if the first corresponding set of variables determine the corresponding second set of variables. For any lattice one may ask which entropy inequalities that will hold for variables that are related in a way determined by the lattice. For certain classes of lattices all entropy inequalities are of the Shannon type and one goal of this research is to identify these lattices. There is also a link between entropy inequalities and inequalities for subgroups of a group. It turns out that this is related to the question of whether a specific lattice can be represented as the lattice of subgroups of a given group. This relation can be used see how certain codes used in cryptography and channel coding can be realized by the algebraic structure of certain groups.

3.11 On the undecidability of conditional independence implication

Cheuk Ting Li (The Chinese University of Hong Kong, HK)

Main reference Cheuk Ting Li: "The Undecidability of Conditional Affine Information Inequalities and Conditional Independence Implication with a Binary Constraint", in Proc. of the IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021, pp. 1–6, IEEE, 2021.
URL https://doi.org/10.1109/ITW48936.2021.9611489

The conditional independence implication problem is to decide whether several statements on the conditional independence among random variables implies another such statement. In this talk, we show that this problem is undecidable if we also allow imposing cardinality constraints (e.g., "X is a binary random variable"). This is proved via a reduction from the domino problem about tiling the plane with a set of tiles. We will also briefly discuss a recent preprint which establishes the undecidability of the original conditional independence implication problem (without cardinality constraints) and related results, e.g., the undecidability of conditional information inequalities and network coding.

3.12 Tutorial on an Information Theoretic Approach to Estimating Query Size Bounds

Hung Ngo (relationalAI – Berkeley, US)

Joint work of Mahmoud Abo Khamis, Sungjin Im, Hossein Keshavarz, Phokion Kolaitis, Ben Moseley, Long Nguyen, Hung Ngo, Kirk Pruhs, Dan Suciu, Alireza Samadian Zakaria
Main reference Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu: "What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?", in Proc. of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, 2017, pp. 429–444, ACM, 2017.
URL https://doi.org/10.1145/3034786.3056105

Cardinality estimation is one of the most important problems in database management. One aspect of cardinality estimation is to derive a good upper bound on the output size of a query, given a statistical profile of the inputs. In recent years, a promising information-theoretic approach was devised to address this problem, leading to robust cardinality estimators which are used in practice.

The information theoretic approach led to many interesting open questions surrounding optimizing a linear function on the almost-entropic or polymatroidal cones. This talk briefly introduces the problem, the approach, summarizes some known results, and lists open questions.

3.13 Term Coding

Søren Riis (Queen Mary University of London, GB)

In this presentation, I introduce Term Coding (TC), which can be seen as an interface between Universal Algebra and Coding Theory. The work grew out of research related to Information flows and Information bottlenecks and the relationship to multiuser information

theory [1, 2, 3, 4]. A large class of (finite) mathematical structures, e.g. universal algebras, can be defined by various equations. Traditionally, the main focus is on systems that satisfy the equations for all points in the domain. The concern in TC is finding structures (codes) that meet the defining equations for many but not necessarily all points. TC provide a general framework for (dynamic) network coding, index coding, and graph guessing games and is related to non-Shannon information inequalities and other advances in Information Theory [4].

References

1 Riis, S., 2007. Reversible and irreversible information networks. IEEE Transactions on Information Theory, 53(11), pp.4339-4349.
2 Gadouleau, M. and Riis, S., 2011. Graph-theoretical constructions for graph entropy and network coding based communications. IEEE Transactions on Information Theory, 57(10), pp.6703-6717.
3 Riis, S. and Gadouleau, M., 2011, July. A dispersion theorem for communication networks based on term sets. In 2011 IEEE International Symposium on Information Theory Proceedings (pp. 593-597). IEEE.
4 Baber, R., Christofides, D., Dang, A.N., Vaughan, E.R. and Riis, S., 2016. Graph guessing games and non-Shannon information inequalities. IEEE Transactions on Information Theory, 63(7), pp.4257-4267.
5 Riis, S. and Gadouleau, M., 2019. Max-flow min-cut theorems on dispersion and entropy measures for communication networks. Information and Computation, 267, pp.49-73.

3.14 A couple of unusual information inequalities and their applications

Andrej E. Romashchenko (University of Montpellier – LIRMM, FR & CNRS, FR)

Joint work of Emirhan Gürpinar, Andrej Romashchenko
Main reference Emirhan Gürpinar, Andrei E. Romashchenko: "Communication Complexity of the Secret Key Agreement in Algorithmic Information Theory", CoRR, Vol. abs/2004.13411, 2020.
URL https://arxiv.org/abs/2004.13411

It is known that the mutual information of a pair of objects x and y is equal to the size of the largest shared secret key that two parties (holding as their inputs x and y respectively) can establish via a communication protocol with interaction on a public channel. We discuss communication complexity of this problem and show show that a tight lower bound on the communication complexity can be proven with help of the expander mixing lemma combined with information inequalities.

References

1 Emirhan Gürpinar, Andrej Romashchenko. *Communication Complexity of the Secret Key Agreement in Algorithmic Information Theory*. arXiv:2004.13411

3.15 Conditional Ingleton inequalities

Milan Studený (The Czech Academy of Sciences – Prague, CZ)

Main reference Milan Studený: "Conditional Independence Structures Over Four Discrete Random Variables Revisited: Conditional Ingleton Inequalities", IEEE Transactions on Information Theory, Vol. 67(11), pp. 7030–7049, 2021.
URL https://doi.org/10.1109/TIT.2021.3104250

Linear information inequalities valid for entropy functions induced by discrete random variables play an important role in the task to characterize discrete conditional independence structures [3, 4, 5]. Specifically, the so-called conditional Ingleton inequalities in the case of 4 random variables are in the center of interest: these are valid under conditional independence assumptions on the inducing random variables. The four inequalities of this form were earlier revealed: by Yeung and Zhang in 1997 [7], by Matúš in 1999 [5] and by Kaced and Romashchenko in 2013 [2]. In a recent 2021 paper [6] the fifth inequality of this type was found. These five information inequalities can be used to characterize all conditional independence structures induced by four discrete random variables. One of open problems in that 2021 paper was whether the list of conditional Ingleton inequalities over 4 random variables is complete: the analysis can be completed by a recent finding of Boege [1] answering that question.

References

1 T. Boege: No eleventh conditional Ingleton inequality. A manuscript (2022) available at `https://arxiv.org/abs/2204.03971`.
2 T. Kaced, A. Romashchenko: Conditional information inequalities for entropic and almost entropic points. *IEEE Transactions on Information Theory 59* (2013), 7149–7167.
3 F. Matúš and M. Studený: Conditional independences among four random variables I. *Combinatorics, Probability and Computing 4* (1995), 269–278.
4 F. Matúš: Conditional independences among four random variables II. *Combinatorics, Probability and Computing 4* (1995), 407–417.
5 F. Matúš: Conditional independences among four random variables III: final conclusion. *Combinatorics, Probability and Computing 8* (1999), 269–276.
6 M. Studený: Conditional independence structures over four discrete random variables revisited: conditional Ingleton inequalities. *IEEE Transactions on Information Theory 67* (2021), 7030-7049.
7 Z. Zhang, R.W. Yeung: A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. on Inform. Theory 43* (1997), 1982–1986.

3.16 Tutorial on conditional independence implication problem

Milan Studený (The Czech Academy of Sciences – Prague, CZ)

URL https://dblp.dagstuhl.de/pid/34/5033.html

The beginning of the tutorial was a brief overview of classic results on conditional independence inference. Then basic concepts were introduced: discrete random vector and conditional independence concept. Basic observation about discrete structures is that they form a finite lattice and can be characterized in terms of the so-called Horn clauses. After that the concept of a semi-graphoid was introduced and relevant partial axiomatizability results recalled: this

concerns marginal and saturated independence [4], functional dependence [6] and relative two-antecedental completeness of semi-graphoidal inference [9]. The inspiration from the theory of relational databases for these results was explained [1, 8]. Basic information-theoretical measures were defined and their relation to the entropic function and the multiinformation functions recalled. A substantial part of the talk was devoted to algorithmic aspects of conditional independence inference, where the concept of a structural semi-graphoid plays the crucial role [7, 3]. The last part of the talk dealt with special conditional independence implications valid in case of Gaussian conditional independence structures [5, 2].

References

1 W. W. Armstrong (1974). Dependency structures of database relations. In Information Processing 74, North Holland, 580-583.
2 T. Boege, A. D'Alì, T. Kahle, B. Sturmfels (2019). The geometry of gaussoids. Foundations of Computational Mathematics 19 (4), 775-812.
3 R. Bouckaert, R. Hemmecke, S. Lindner, M. Studený (2010). Efficient algorithms for conditional independence inference. Journal of Machine Learning Research 11, 3453-3479.
4 D. Geiger, J. Pearl (1993). Logical and algorithmic properties of conditional independence and graphical models. Annals of Statistics 21 (4), 2001-2021.
5 R. Lněnička, F. Matúš (2007). On Gaussian conditional independence structures. Kybernetika 43 (3), 327-342.
6 F. Matúš (1991). Abstract functional dependency structures. Theoretical Computer Science 81 (1), 117-126.
7 M. Niepert (2009). Logical inference algorithms and matrix representations for probabilistic conditional independence. In 25th UAI conference, AUAI Press, 428-435.
8 Y. Sagiv and S. F. Walecka (1982). Subset dependencies and completeness result for a subclass of embedded multivalued dependencies. Journal of Association for Computing Machinery 29 (1), 103-117.
9 M. Studený (1997). Semigraphoids and structures of probabilistic conditional independence. Annals of Mathematics and Artificial Intelligence 21(1), 71-98.

3.17 Max-Information Inequalities and the Domination Problem

Dan Suciu (University of Washington – Seattle, US)

Joint work of Dan Suciu, Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Ngo
Main reference Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, Dan Suciu: "Bag Query Containment and Information Theory", ACM Trans. Database Syst., Vol. 46(3), pp. 12:1–12:39, 2021.
URL https://doi.org/10.1145/3472391

A max-information inequality is an inequality involving linear expressions of entropic terms, and one occurrence of max. We consider the problem: given a max-information inequality, check if it holds for all entropic vectors. E.g. $\max(H(XY), H(YZ), H(ZX)) \geq {}^2/_3 H(XYZ)$ is valid. It is open whether this problem is decidable.

We say that a structure B "dominates" a structure A, if, for any other structure C, the number of homomorphisms $A \to C$ is less than or equal to the number of homomorphisms $B \to C$. We consider the problem: given two structures A, B, where B is acyclic, check whether B dominates A. The domination problem is equivalent to the "conjunctive query containment problem under bag semantics", which is of interest in database theory. It is open whether this problem is decidable.

We prove that these two problems are computationally equivalent. In particular, any progress on the decidability or undecidability of one of these problems will automatically carry over to the other problem.

3.18 A Conditional Information Inequality and Its Combinatorial Applications

Nikolay K. Vereshchagin (NRU Higher School of Economics – Moscow, RU)

Joint work of Tarik Kaced, Andrej Romashchenko, Nikolay K. Vereshchagin
Main reference Tarik Kaced, Andrei E. Romashchenko, Nikolai K. Vereshchagin: "A Conditional Information Inequality and Its Combinatorial Applications", IEEE Trans. Inf. Theory, Vol. 64(5), pp. 3610–3615, 2018.
URL https://doi.org/10.1109/TIT.2018.2806486

We show that the inequality $H(A \mid B, X) + H(A \mid B, Y) < H(A \mid B)$ for jointly distributed random variables A, B, X, Y, which does not hold in general case, holds under some natural condition on the support of the probability distribution of A, B, X, Y. This result generalizes a version of the conditional Ingleton inequality: if for some distribution $I(X : Y \mid A) = H(A \mid X, Y) = 0$, then $I(A : B) < I(A : B \mid X) + I(A : B \mid Y) + I(X : Y)$.

We present one application of this result. Assume that a family $\mathcal{F}$ of pair-wise disjoint "squares" $S \times S \subset U \times V$ is given (U, V are fixed finite sets). Assume that for each $u \in U$ there are at least L squares in $\mathcal{F}$, whose first projection covers u, and similarly, for each $v \in V$ there are at least R squares in F, whose second projection covers v. Then $|\mathcal{F}| \geq LR$.

3.19 When are Exhaustive Minimal Lists of Information Inequalities Scalable?

John MacLaren Walsh (Drexel University – Philadelphia, US)

Joint work of Yirui Liu, John MacLaren Walsh
Main reference Yirui Liu, Ph.D. Dissertation – Drexel University, October 2021.
Main reference Yirui Liu, John MacLaren Walsh: "Linear Complexity Entropy Regions", in Proc. of the IEEE International Symposium on Information Theory, ISIT 2021, Melbourne, Australia, July 12-20, 2021, pp. 1642–1647, IEEE, 2021.
URL https://doi.org/10.1109/ISIT45174.2021.9518030

Exhaustively determining the entropy region, and the information inequalities that describe it, form a tantalizingly fundamental problem in multi-terminal information theory. Among other equivalences, determining all information inequalities has been shown to be equivalent to determining the capacity regions of all networks under network coding, as well as determining all inequalities linking sizes of intersections of subgroups of a common group. Faces of the entropy region, in turn, dictate fundamental possible conditional independence relations among a series of random variables.

A key observation, however, is that many of these ultimate uses of the entropy region, both fundamental and applied, need only study the relationship between entropies of subsets that lie within a very small subfamily of the powerset of all subsets of the inscribed random variables. That is to say, were one able to exhaustively characterize the projection of the

entropy region onto only the (sub)family of exclusively subsets involved in a problem of interest, one may provide the fundamental limits in that problem while the harder problem of determining the entire entropy region on the powerset remains unsolved. A natural question one may then ask is, which types of systems of subsets enable one to provide an exhaustive characterization of the associated projection of the entropy region onto only these subsets? Furthermore, which types of such systems of subsets yield the closure of the projected entropy region to be polyhedral? Which of those systems of subsets yield polyhedra enable a description complexity, as measured by the total number of involved inequalities, which scales nicely, even linearly, in the size of the problem, measured as the number of random variables?

In this talk, we set about characterizing some of these families of subsets with scalable complexity through the idea of pasting the entropy regions of small, overlapping, subsets to obtain bounds on associate projections of the entropy region on their union. A straightforward argument shows that this pasting construction easily yields outer bounds, and as such, attention shifts to when these outer bounds are tight. Examples of infinite families of subsets where the pasted entropy regions exhaustively characterize the associated projection of the larger entropy region are detailed. Among these are included cases where the associated projection of the entropy region has a number of required minimal inequalities that scales linearly with the number of random variables. Moreover, a construction proves cases where such pasted outer bounds are guaranteed to be loose.

Bearing these cases where pasting small entropy regions together only yields a loose outer bound for the true entropy region in mind, attention then shifts to finding inner bound constructions whose inner bound property is preserved under pasting. Of particular interest is in the inner bound to the entropy region formed by the set of inequalities linking linear ranks which dictates the part of the entropy region reachable by time sharing linear codes, as well as those linked with quasi-uniform distributions. A first inner-bound preserving technique pastes together integral polyhedral quasi-uniform bound on a chain of sets in the overlap of their ground set. A second inner-bound preserving pasting technique based on requiring the existence of consistent common informations is then also provided for these types of inner bounds. These constructions, together with the constructions that correctly characterize the associated projections of the entropy region, form a substantial family of composable constructions that can be used to create inner bounds for the entropy region of controllable complexity.

3.20 Graph Information Ratio

Lele Wang (University of British Columbia – Vancouver, CA)

Joint work of Ofer Shayevitz, Lele Wang
Main reference Lele Wang, Ofer Shayevitz: "Graph Information Ratio", SIAM J. Discret. Math., Vol. 31(4), pp. 2703–2734, 2017.
URL https://doi.org/10.1137/16M1110066

We introduce the notion of information ratio $\mathrm{Ir}(H/G)$ between two (simple, undirected) graphs G and H, defined as the supremum of ratios k/n such that there exists a mapping between the strong products G^k to H^n that preserves non-adjacency. Operationally speaking, the information ratio is the maximal number of source symbols per channel use that can be reliably sent over a channel with a confusion graph H, where reliability is measured w.r.t. a

source confusion graph G. Various results are provided, including in particular lower and upper bounds on $\mathrm{Ir}(H/G)$ in terms of different graph properties, inequalities and identities for behavior under strong product and disjoint union, relations to graph cores, and notions of graph criticality. Informally speaking, $\mathrm{Ir}(H/G)$ can be interpreted as a measure of similarity between G and H. We make this notion precise by introducing the concept of information equivalence between graphs, a more quantitative version of homomorphic equivalence. We then describe a natural partial ordering over the space of information equivalence classes, and endow it with a suitable metric structure that is contractive under the strong product. Various examples and open problems are discussed.

3.21 On entropic and almost-entropic representability of matroids

Geva Yashfe (The Hebrew University of Jerusalem, IL)

Joint work of Lukas Kühne, Geva Yashfe
Main reference Lukas Kühne, Geva Yashfe: "On entropic and almost multilinear representability of matroids", CoRR, Vol. abs/2206.03465, 2022.
URL https://doi.org/10.48550/arXiv.2206.03465

This talk discusses some recent results obtained jointly with Lukas Kühne and announces one new theorem. There is no algorithm which, given a matroid M,

1. Decides whether M is entropic.
2. Decides whether M is multilinear.
3. Decides whether M is almost-multilinear.
4. Decides whether M is almost-entropic.

(The last theorem was proved during the conference after an inspiring discussion with Janneke Bolt, Andrej Romashchenko, and Alexander Shen.)

Here a matroid M is a polymatroid with values in the natural numbers (including 0) and satisfying that every singleton has rank at most 1. It is entropic if, as a polymatroid, its ray intersects the entropic cone. It is almost-entropic if it is in the closure of the entropic cone. The multilinear variants are about the analogous cones of linear rank functions.

A corollary of these theorems is that the conditional independence problem and its "approximate" variant are undecidable. Closely related results in the non-approximate setting have been obtained by Cheuk-Ting Li (preceding ours by some weeks) and have also been presented at this conference.

3.22 Machine-Proving of Entropy Inequalities

Raymond W. Yeung (The Chinese University of Hong Kong, HK)

Joint work of Laigang Guo, Raymond W. Yeung
Main reference Laigang Guo, Raymond W. Yeung, Xiao-Shan Gao: "Proving Information Inequalities and Identities with Symbolic Computation", in Proc. of the IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022, pp. 772–777, IEEE, 2022.
URL https://doi.org/10.1109/ISIT50566.2022.9834774

The entropy function plays a central role in information theory. Constraints on the entropy function in the form of inequalities, viz. entropy inequalities (often conditional on certain Markov conditions imposed by the problem under consideration), are indispensable tools for proving converse coding theorems. In this talk, I will give an overview of the development of machine-proving of entropy inequalities for the past 25 years. To start with, I will present a geometrical framework for the entropy function, and explain how an entropy inequality can be formulated, with or without constraints on the entropy function. Among all entropy inequalities, Shannon-type inequalities, namely those implied by the nonnegativity of Shannon's information measures, are best understood. We will focus on the proving of Shannon-type inequalities, which in fact can be formulated as a linear programming problem. I will discuss ITIP, a software package originally developed for this purpose in the mid-1990s, as well as some of its later variants. In 2014, Tian successfully characterized the rate region of a class of exact-repair regenerating codes by means of a variant of ITIP. This is the first nontrivial converse coding theorem proved by a machine. At the end of the talk, I will discuss some recent progress in speeding up the proving of entropy inequalities.

4 Open problems

The following problems have been collected from discussions, talks, and open problem sessions. The problems have been grouped into several different a themes and are followed by references to the literature. The person posing each problem is indicated in square brackets after the statement of the problem.

Secret sharing and cryptography

(1.1) The share size of a perfect secret sharing scheme with n participants can be bounded by $\mathcal{O}(2^{0.525n}) \cap \Omega(n/\log n)$. Can the upper bound be improved $\mathcal{O}(2^{cn})$ with $c < 1/2$? [Amos Beimel]

(1.2) Is there a secret sharing scheme which can be realized by an abelian group but not by a field? [László Csirmaz]

(1.3) Two parties want to compute the `OR` of their random bits. What is the minimal amount of information either of them has to disclose about their bit? How does this number depend on the number of rounds? [Alexander Shen]

Peculiarities of the entropy region

(2.1) Consider the set $\mathbf{S}_{ij}$ inside of Γ_4^* defined in [11, Section VII]. Theorem 5 in the same paper shows that the infimal Ingleton score in Γ_4^* is attained in this region and Example 2 exhibits a polymatroid in it which refutes the 4-atom conjecture as formulated in [6]. By how far was the 4-atom conjecture off? That is, what is the infimal Ingleton score? Is it an algebraic number? Which distributions reach it? [László Csirmaz]

(2.2) According to the experiments by Csirmaz [4] and in the coordinate system chosen there, the face $\{\,\beta = 0\,\}$ of $\mathbf{S}_{ij}$ contains entropic points whose distributions have at most 8 states per variable but there is a gap in his visualization between the interior and the face. Can this face be approximated arbitrarily well at all from the interior with only a bounded alphabet size? [László Csirmaz]

(2.3) Let Γ_n^* denote the entropy region of discrete random variables with finite support and Γ_n^∞ the entropy region of discrete random variables with countable support. Clearly $\Gamma_n^* \subseteq \Gamma_n^\infty \subseteq \overline{\Gamma_n^*} = \overline{\Gamma_n^\infty}$. Are the containments strict? [Tobias Boege]

(2.4) Are there "holes" on the boundary of $\overline{\Gamma_n^*}$? More concretely, is there a ray on one of its faces and four numbers $x < y < y' < z$ such that on this ray the intervals (x, y) and (y', z) parametrize entropic points and the segment parametrized by (y, y') contains no entropic point? [László Csirmaz]

(2.5) Is there an extreme ray of $\overline{\Gamma_n^*}$ which contains no entropic point? [John MacLaren Walsh]

(2.6) Fix n discrete random variables and add another one. Which entropy profiles arise? Specifically, look at the triples $H(W)$, $H(A|W)$, $H(B|W)$ for fixed A, B and arbitrary W. [Alexander Shen]

(2.7) Are the interior of the entropy region and its complement effectively open sets? [Alexander Shen]

Information quantities

(3.1) Let A and B be jointly distributed and consider the optimization problem $\sup I(A : B|X)$ for X jointly distributed with (A, B). Is the supremum attained? How to compute it as a function of the joint probability table for A and B? What if $A \perp\!\!\!\perp B$? [Nikolay Vereshchagin]

(3.2) Suppose the joint distribution of A, B, C factors into $p(a, b, c) = p_1(a, b) \cdot p_2(a, c) \cdot p_3(b, c)$. What is the minimal value of $I(A : B : C) = I(A : B) - -I(A : B \mid C)$? Which assumptions on the distribution guarantee non-negativity? [Mahmoud Abo Khamis]

(3.3) Let $\Delta(A, B, C) \coloneqq I(A : B \mid C) + I(A : C \mid B) + I(B : C \mid A)$, $\Delta'(A, B) \coloneqq \inf_C \Delta(A, B, C)$ and $\Gamma(A, B) = \sup_{X,Y}[I(A : B) - -I(A : B|X) - -I(A : B|Y) - -I(X : Y)]$. Then $\Gamma(A, B) \le \Delta'(A, B)$. Can the inequality be strict? [Alexander Shen]

(3.4) Let S be a finite set of triples which is closed under rotation and $w : S \to \mathbb{R}$ a weight function. Consider rotation-invariant distributions of three discrete random variables A, B, C whose support are the triples in S and the following optimization problem:

$$\sup_{ABC} \left(\frac{1}{3}\Big(H(A) + H(B) + H(C)\Big) + H(ABC) - \sup_{XYZ \sim ABC} H(XYZ) + \mathbb{E}[w(A, B, C)] \right).$$

The constraint $XYZ \sim ABC$ above means all distributions X, Y, Z which have the same 1-marginals as A, B, C. How to compute the optimum? This problem comes up in fast matrix multiplication theory. See [10] for the background and a relaxation. [Yuval Filmus]

(3.5) Is there a zero-one law in information transmission? Suppose a transmission task with only one restricted link. Is there a rate threshold below which correct transmission is only possible with negligible probability and above which there exists an encoding that ensures correct transmission with high probability? [Alexander Shen]

Suppose that a pair of random variables A and B is given. Add a new, jointly distributed random variable X and record the conditional entropies $\big(H(A|X), H(B|X), H(AB|X)\big)$. This yields a point in $\mathbb{R}^3$. The collection of all such points over all X forms a closed convex subset of $\mathbb{R}^3$, the *extension profile* $\mathrm{Ext}_1(A, B)$. Adding k random variables $X_1, \ldots, X_k$ to the pair (A, B) and recording all the conditional entropies $\big(H(A|X_I), H(B|X_I), H(AB|X_I)$ for all non-empty subvectors X_I of $(X_1, \ldots, X_k)$ results in higher extension profiles $\mathrm{Ext}_k(A, B)$.

(3.6) Does $\mathrm{Ext}_1(A, B)$ determine $\mathrm{Ext}_k(A, B)$ for all $k \geq 1$? This is, in spirit, similar to Grothendieck's reconstruction principle in geometry. [Rostislav Matveev]

Information inequalities

We consider various entropy-like regions Θ and the collections of linear inequalities which are satisfied by all points in Θ. These are contained in the dual cone $\Theta^\vee$.

(4.1) Is $(\Gamma_4^*)^\vee$ semialgebraic? The missing piece in an attempt in [7] to answer this question negatively is that the following Ingleton inequality is essentially conditional:

$$\begin{aligned} &I(A : C \mid D) = I(A : D \mid C) = I(B : C \mid D) = I(B : D \mid C) = 0 \\ \Rightarrow\ &I(C : D) \leq I(C : D \mid A) + I(C : D \mid B) + I(A : B). \end{aligned}$$

[Andrej Romashchenko]

(4.2) Is the fifth conditional Ingleton inequality of [14] essentially conditional? Is it valid for almost-entropic points? [Milan Studený]

(4.3) By [8, Theorem 7.1] the validity of a max-linear information inequality is equivalent to the validity of an unconditional linear information inequality with existentially quantified non-negative coefficients. Can these coefficients always be chosen rational? If yes, this would imply Turing-equivalence of the problems. [Dan Suciu]

(4.4) Are the cones of linear rank inequalities for $n \geq 6$ polyhedral? [Alexander Shen]

(4.5) Are the inequalities for Shannon entropies valid for *prefix complexity* with precision $\mathcal{O}(1)$? This is known with $\mathcal{O}(\log n)$ precision for Kolmogorov complexity; cf. [13, Chapter 10]. [Alexander Shen]

The Gaussian differential entropy region Υ_n^* is (up to a scaling and an additive term) made of all vectors $(\log\det\Sigma_K : K \subseteq [n])$ with Σ a positive definite $n \times n$ matrix and Σ_K the diagonal submatrix with rows and columns indexed by K. A rational point in $(\Upsilon_n^*)^\vee$ corresponds to a *determinantal inequality* for positive definite matrices under the logarithm and hence its validity is decidable in the existential theory of the reals. A study of the relation between $(\Gamma_n^*)^\vee$ and $(\Upsilon_n^*)^\vee$ was initiated in [2]. The convex conic closure of Υ_n^* is contained in that of Γ_n^* after adding certain functions ϕ_i to the latter, which are defined by

$$\phi_i(I) \coloneqq \begin{cases} -1 & \text{if } i \in I, \\ 0 & \text{otherwise.} \end{cases}$$

The functions ϕ_i have to be added because points in Υ_n^* can have negative entries. What if we instead consider the multiinformation functions of discrete and Gaussian random vectors? They are non-negative, increasing and supermodular. The multiinformation regions are linear images of the entropy regions and correspond to their tight parts.

(4.6) Is the multiinformation region of Gaussians contained in the one for discrete random variables (without ϕ_i's)? This would give a decidable subregion of the cone of linear information inequalities via determinantal inequalities for positive definite matrices. [Tobias Boege]

Let $\mathcal{C}$ be a set of finite groups closed under cartesian product and subgroups. Then given any $G \in \mathcal{C}$ and subgroups H_i, $i \in [n]$, define the vector $(\log[G : \bigcap_{i\in I} H_i] : I \subseteq [n])$, which is the entropy profile of a family of uniform distributions on cosets of $H_I = \bigcap_{i\in I} H_i$ in G. Let $\Gamma_n^*(\mathcal{C})$ be the region of all such vectors. The euclidean closure of each such region is a convex cone under the assumptions on $\mathcal{C}$. Note that the region for abelian groups satisfies the Ingleton inequality. The set of all finite groups yields the Shannon entropy region by [3].

(4.7) Study $\Gamma_n^*(\mathcal{C})$ for classes of finite groups with more structure theory, such as vector spaces over $\mathbb{F}_p$, abelian groups, nilpotent groups, solvable groups, ... Are the inclusions between their entropy regions all strict? [Rostislav Matveev]

Matroids in information theory

(5.1) By [9] it is undecidable if a matroid is entropic. Does this still hold for sparse paving matroids (which are conjectured to be almost all matroids)? [Geva Yashfe]

A $^k/_m$ *matroid approximation* of an integer polymatroid g is a restriction of the free expansion $E(m \cdot g)$ such that each factor X_i of the expansion retains at least a fraction of $^k/_m$ of its elements; see [15, Chapter 10] for the free expansion.

(5.2) If g is entropic, do there exist entropic matroid approximations with $^k/_m$ arbitrarily close to 1? A sufficiently good matroid approximation would violate the Ingleton inequality when g does and hence provide an example of a non-multilinear but entropic matroid. [Geva Yashfe]

Conditional independence

(6.1) How to define a conditional independence relation $\perp\!\!\!\perp$ on a general lattice so that the resulting CI structures fulfill the semigraphoid axioms? [Peter Harremoës]

(6.2) An information inequality of the form $I(X : Y \mid Z) \leq \sum c_i I(A_i : B_i \mid C_i)$ implies the conditional independence inference rule $\bigwedge_i [A_i \perp\!\!\!\perp B_i \mid C_i] \Rightarrow [X \perp\!\!\!\perp Y \mid Z]$. Consider an inference rule and all of its proofs from *Shannon-type* inequalities. What can be said about the size of the coefficients c_i of the "best" such proofs? [Batya Kenig]

(6.3) Is there a finite set of generalized conditional independence inference rules for structural semigraphoids in the sense of [1]? [Janneke Bolt]

Query size estimation

The talk by Hung Ngo introduced a number of bounds for query size estimation in databases. The following questions by Hung concern the relations of these bounds:

(7.1) Is the entropic bound computable? [All by Hung Ngo]

(7.2) What is the complexity of computing the polymatroid bound? Is it NP-hard?

(7.3) If so, what are instances with tractable parametrized complexity?

(7.4) Investigate the tightness of the various bounds, especially the gap between entropic and combinatorial bound.

(7.5) What are the worst-case lengths of proof sequences for Shannon flow inequalities?

Complexity and expressivity

(8.1) Is the validity of linear information inequalities (with rational coefficients), i.e., their containment in $(\Gamma_n^*)^\vee$, decidable? This is open even for $n = 4$. [Many participants]

(8.2) Are there valid inequalities for the linear rank region which cannot be proved by applying valid *entropic* inequalities to subspaces, their sums and intersections (i.e., common informations); cf. [5]? [Alexander Shen]

(8.3) How do linear rank inequalities depend on the field size? How do they depend on the characteristic? [Alexander Shen]

(8.4) The Copy lemma can be used to derive new inequalities from the Shannon inequalities via projection. How does the proof strength of this method increase with dimension or number of copies? [Alexander Shen]

(8.5) Is the entropy maximization principle strictly stronger than the Copy lemma? (The inequalities provable by a single use of the Copy lemma can also be proved by a single use of the MEP. Iterated applications of Copy lemma can be used to simulate a single use of MEP.) [László Csirmaz]

(8.6) Construct an explicit ternary function on a (large) set X that cannot be represented by a circuit of 10 arbitrary binary gates with inputs/ouputs on X. This may be seen as a discrete version of Hilbert's 13^{th} problem. [Alexander Shen]

Combinatorics and Kolmogorov complexity

(9.1) Given an n-bit string s of complexity m. What is the largest complexity obtainable from s by changing at most k bits? [Alexander Shen]

(9.2) Given an n-bit string of complexity m of which every bit flips with probability ε. Which complexity increase can be guaranteed with high probability? The analogue in probability is the increase in entropy of a vector of binary random variables subject to noise. [Alexander Shen]

(9.3) Is there a general procedure for obtaining parallel results in Shannon entropy and Kolmogorov complexity? For a negative result, see [12]. [Alexander Shen]

Let $S \subseteq \mathbb{N}^{\{a,b,c\}}$ be a finite set, $N_I(S)$, for $I \subseteq \{a, b, c\}$, the cardinality of the I-coordinate projection of S. Suppose that $V \cdot \ell \leq N_{ab}(S) \cdot N_{ac}(S)$ for some integers V and ℓ. Then S can be split into S_1 and S_2 such that $N_{abc}(S_1) \leq V$ and $N_a(S_2) \leq \ell$. This is a combinatorial analogue of the non-negativity of conditional mutual information; cf. [13, Section 10.7].

(9.4) Find such analogues for (non-Shannon) information inequalities. [Alexander Shen]

(9.5) How to interpret known *conditional* information inequalities for entropies combinatorially? [Alexander Shen]

References

1. Janneke H. Bolt and Linda C. van der Gaag. Generalized rules of probabilistic independence. In *Symbolic and quantitative approaches to reasoning with uncertainty. 16th European conference, ECSQARU 2021, Prague, Czech Republic, September 21–24, 2021. Proceedings*, pages 590–602. Springer, 2021.
2. Terence Chan, Dongning Guo, and Raymond W. Yeung. Entropy functions and determinant inequalities. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1251–1255, 2012.
3. Terence H. Chan and Raymond W. Yeung. On a relation between information inequalities and group theory. *IEEE Transactions on Information Theory*, 48(7):1992–1995, 2002.
4. László Csirmaz. Visualizing the entropy region. `https://github.com/lcsirmaz/entropy-rules/blob/089f64bb/visual/`.
5. Randall Dougherty, Chris Freiling, and Kenneth Zeger. Linear rank inequalities on five or more variables, 2009.
6. Randall Dougherty, Chris Freiling, and Kenneth Zeger. Non-shannon information inequalities in four random variables, 2011.
7. Arley Gomez, Carolina Mejia, and J. Andres Montoya. Defining the almost-entropic regions by algebraic inequalities. *Int. J. Inf. Coding Theory*, 4(1):1–18, 2017.
8. Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. Bag query containment and information theory. *ACM Trans. Database Syst.*, 46(3), 2021.
9. Lukas Kühne and Geva Yashfe. On entropic and almost multilinear representability of matroids, 2022.
10. François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation, ISSAC 2014, Kobe, Japan, July 23–25, 2014*, pages 296–303. Association for Computing Machinery (ACM), 2014.
11. František Matúš and László Csirmaz. Entropy region and convolution. *IEEE Trans. Inf. Theory*, 62(11):6007–6018, 2016.
12. Andrej A. Muchnik and Nikolay K. Vereshchagin. Shannon entropy vs. Kolmogorov complexity. In *Computer science – theory and applications. First international computer science symposium in Russia, CSR 2006, St. Petersburg, Russia, June 8–12, 2006. Proceedings.*, pages 281–291. Berlin: Springer, 2006.
13. Alexander Shen, Vladimir A. Uspensky, and Nikolay K. Vereshchagin. *Kolmogorov complexity and algorithmic randomness. Translated from Russian*, volume 220 of *Math. Surv. Monogr.* American Mathematical Society (AMS), 2017.
14. Milan Studený. Conditional independence structures over four discrete random variables revisited: conditional ingleton inequalities. *IEEE Trans. Inf. Theory*, 67(11):7030–7049, 2021.
15. Neil White, editor. *Theory of matroids*, volume 26 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2008.

Participants

- Marcelo Arenas
PUC – Santiago de Chile, CL
- Albert Atserias
UPC Barcelona Tech, ES
- Amos Beimel
Ben Gurion University – Beer Sheva, IL
- Tobias Andreas Boege
MPI für Mathematik in den Naturwissenschaften – Leipzig, DE
- Janneke Bolt
TU Eindhoven, NL
- Laszlo Csirmaz
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Kyle Deeds
University of Washington – Seattle, US
- Oriol Farras
Universitat Rovira i Virgili – Tarragona, ES
- Yuval Filmus
Technion – Haifa, IL
- Emirhan Gürpinar
University of Montpellier, LIRMM – Montpellier, FR
- Miika Hannula
University of Helsinki, FI
- Peter Harremoës
Niels Brock Copenhagen Business College, DK
- Batya Kenig
Technion – Haifa, IL
- Phokion G. Kolaitis
University of California – Santa Cruz, US & IBM Research, US
- Rostislav Matveev
MPI für Mathematik in den Naturwissenschaften – Lepzig, DE
- Fabio Mogavero
University of Naples, IT Hung Ngo, relationalAI – Berkeley, US
- Carles Padró
UPC Barcelona Tech, ES
- Andrei Romashchenko
University of Montpellier & CNRS, LIRMM – Montpellier FR
- Sudeepa Roy
Duke University – Durham, US
- Alexander Shen
University of Montpellier & CNRS – LIRMM, FR
- Milan Studený
The Czech Academy of Sciences – Prague, CZ
- Dan Suciu
University of Washington – Seattle, US
- John MacLaren Walsh
Drexel University – Philadelphia, US
- Lele Wang
University of British Columbia – Vancouver, CA
- Geva Yashfe
The Hebrew University of Jerusalem, IL

Remote Participants

- Mahmoud Abo Khamis
RelationalAI – Berkeley, US
- George Konstantinidis
University of Southampton, GB
- Cheuk Ting Li
The Chinese University of Hong Kong, HK
- Frederique Oggier
Nanyang TU – Singapore, SG
- Soren Riis
Queen Mary University of London, GB
- Yufei Tao
The Chinese University of Hong Kong, HK
- Nikolay K. Vereshchagin
NRU Higher School of Economics – Moscow, RU
- Raymond W. Yeung
The Chinese University of Hong Kong, HK