Matroids in information theory: Conditional Ingleton inequalities

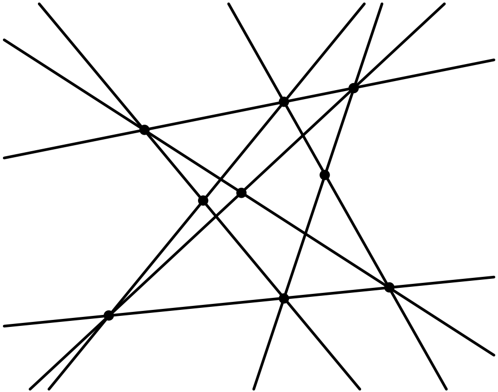
Tobias Boege

Aalto University
School of Science

Graduate Student Meeting on Applied Algebra and Combinatorics,
KTH Stockholm,
27 April 2023

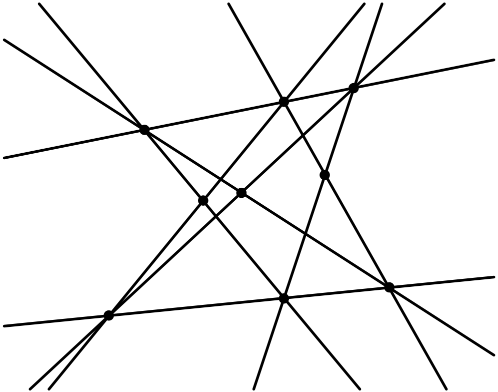
Matroids

- Matroids are combinatorial structures which model "special position" relations in geometry.



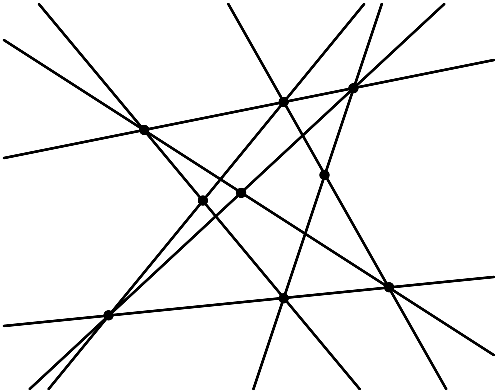
Matroids

- Matroids are combinatorial structures which model “special position” relations in geometry.
  - For example the matroid of a set of points in the projective plane records which triples of points lie on a line.

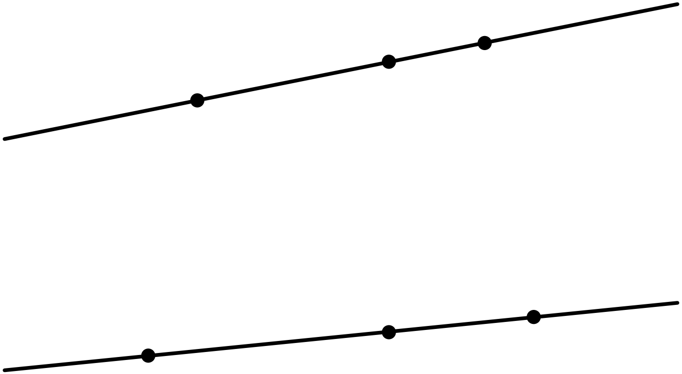


Matroids

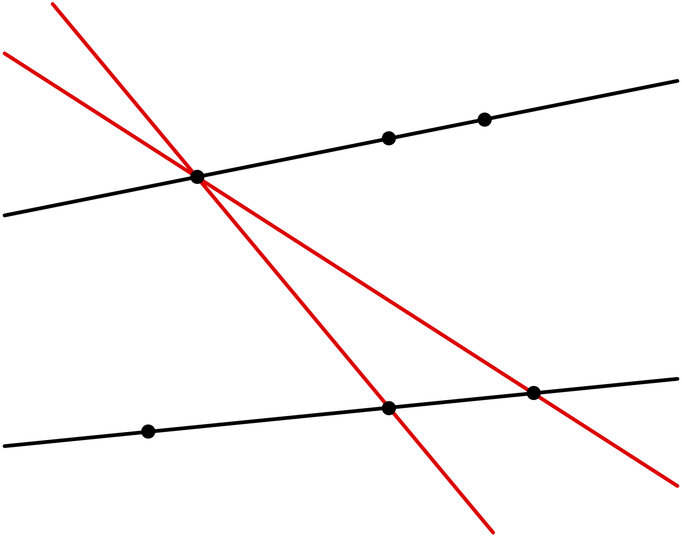
- Matroids are combinatorial structures which model "special position" relations in geometry.
  - For example the matroid of a set of points in the projective plane records which triples of points lie on a line.
- Non-realizability of matroids captures the (non-obvious) laws of geometry.



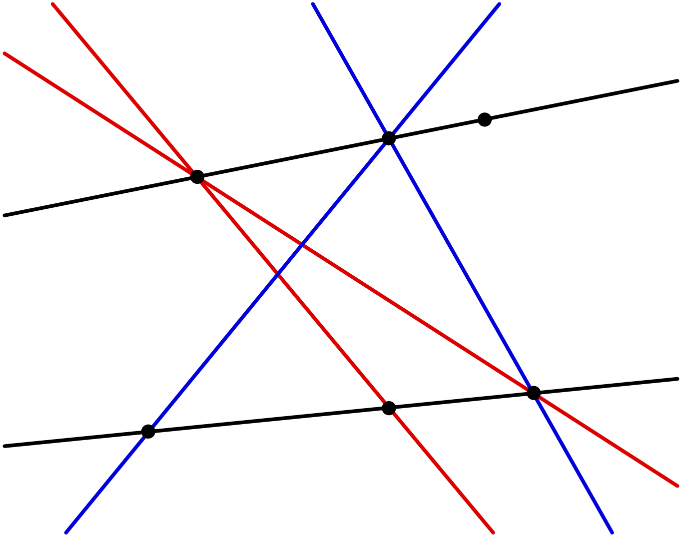
Laws of geometry



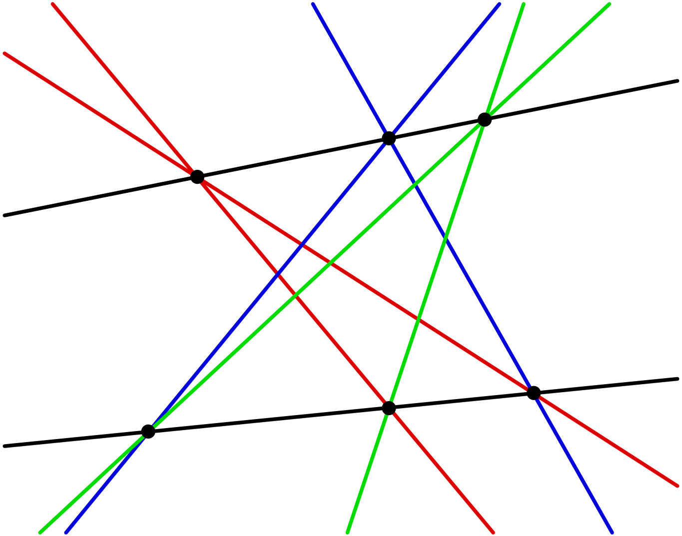
Laws of geometry



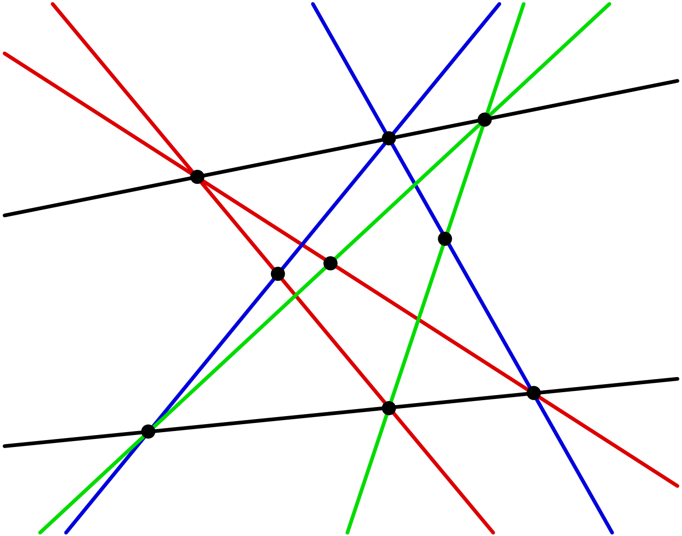
Laws of geometry



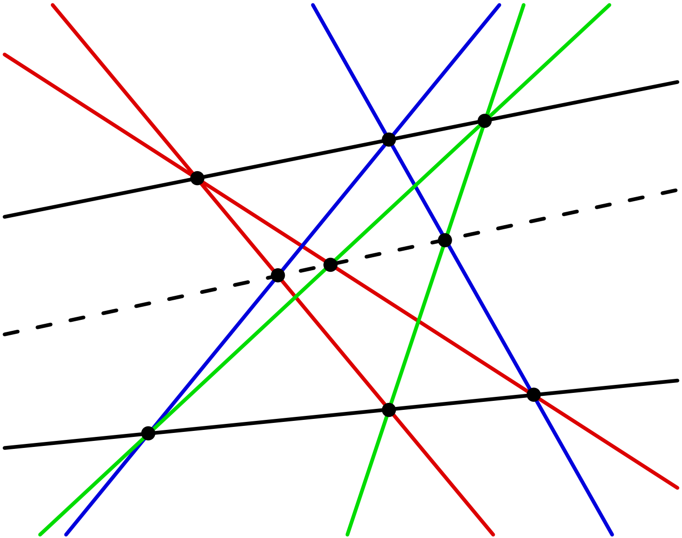
Laws of geometry



Laws of geometry



Laws of geometry



Entropy

Let X be a random variable taking finitely many values $\{1, \ldots, d\}$ with positive probabilities. Its *Shannon entropy* is

$$H(X) \coloneqq \sum_{i=1}^{d} p(X = i) \log {}^{1}/_{p(X = i)}.$$

- H is continuous on $\Delta(d)$ and analytic on the interior.
- A random vector $X \in \Delta(d_1, \ldots, d_n)$ is a random variable in $\Delta(\prod_{i=1}^{n} d_i)$, so the definition of H extends to vectors.
- For a random vector $X = (X_1, \ldots, X_n)$ we have 2^n marginals and we collect their entropies in an entropy vector $h_X : 2^{[n]} \to \mathbb{R}$.
  - For example (X, Y) has entropy vector $(0, H(X), H(Y), H(X, Y)) \in \mathbb{R}^4$.

Entropy as information

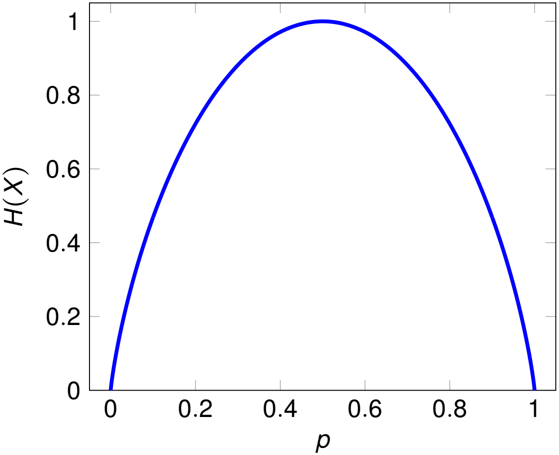


Figure: Entropy of a binary random variable X as a function of $p = p(X = \text{heads})$.

Dictionary matroid theory — information theory

Information-theoretical “special position” properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

Dictionary matroid theory — information theory

Information-theoretical “special position” properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank $\to$ entropy

Dictionary matroid theory — information theory

Information-theoretical “special position” properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank $\to$ entropy
- $h(x) = 0$: loop $\to$ constant random variable

Dictionary matroid theory — information theory

Information-theoretical "special position" properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank → entropy
- $h(x) = 0$: loop → constant random variable
- $h(x, y) = h(x)$: closure operator → functional dependence

Dictionary matroid theory — information theory

Information-theoretical "special position" properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank → entropy
- $h(x) = 0$: loop → constant random variable
- $h(x, y) = h(x)$: closure operator → functional dependence
- $h(x, y) = h(x) = h(y)$: parallel → functional equivalence

Dictionary matroid theory — information theory

Information-theoretical “special position” properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank → entropy
- $h(x) = 0$: loop → constant random variable
- $h(x, y) = h(x)$: closure operator → functional dependence
- $h(x, y) = h(x) = h(y)$: parallel → functional equivalence
- $h(x, y) = h(x) + h(y)$: independence → independence

Dictionary matroid theory — information theory

Information-theoretical "special position" properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank → entropy
- $h(x) = 0$: loop → constant random variable
- $h(x, y) = h(x)$: closure operator → functional dependence
- $h(x, y) = h(x) = h(y)$: parallel → functional equivalence
- $h(x, y) = h(x) + h(y)$: independence → independence
- $h(x, y, z) + h(z) = h(x, z) + h(y, z)$: modular pair → conditional independence

Dictionary matroid theory — information theory

Information-theoretical “special position” properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector:

- $h(x)$: rank → entropy
- $h(x) = 0$: loop → constant random variable
- $h(x, y) = h(x)$: closure operator → functional dependence
- $h(x, y) = h(x) = h(y)$: parallel → functional equivalence
- $h(x, y) = h(x) + h(y)$: independence → independence
- $h(x, y, z) + h(z) = h(x, z) + h(y, z)$: modular pair → conditional independence

Even though entropy is a transcendental function, many of these conditions are **polynomial** in the probabilities → algebraic statistics.

A glimpse at matroids in information theory

Entropy vectors are not matroids but polymatroids. Still, matroids and their combinatorial theory are central to the subject:

Theorem ([Mat92])

If a matroid h is linear over a finite field of size q, then $\log(q) \cdot h$ is entropic.

A glimpse at matroids in information theory

Entropy vectors are not matroids but polymatroids. Still, matroids and their combinatorial theory are central to the subject:

Theorem ([Mat92])

If a matroid h is linear over a finite field of size q, then $\log(q) \cdot h$ is entropic.

Theorem ([Mat17])

If h is algebraic, then it is almost-entropic.

A glimpse at matroids in information theory

Entropy vectors are not matroids but polymatroids. Still, matroids and their combinatorial theory are central to the subject:

Theorem ([Mat92])

If a matroid h is linear over a finite field of size q, then $\log(q) \cdot h$ is entropic.

Theorem ([Mat17])

If h is algebraic, then it is almost-entropic.

Theorem ([Mat07])

Every entropy vector can be approximated by scaled factors of entropic matroids.

Basic computational challenges

Problem

Find/Sample positive points from conditional independence varieties.

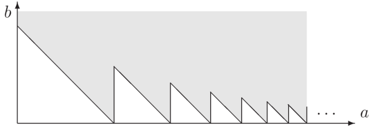
Problem

Optimize a holonomic function subject to polynomial constraints.

Let $\mathbf{H}_n^* \subseteq \mathbb{R}^{2^n}$ consist of all h_X where X is an n-variate discrete random vector. $\mathbf{H}_n^*$ is the image of $\bigcup_{d_1=1}^{\infty} \cdots \bigcup_{d_n=1}^{\infty} \Delta(d_1, \ldots, d_n)$ under the transcendental map $X \mapsto h_X$.

Problem

Find a description of the boundary of $\mathbf{H}_3^*$.



Entropy region and information inequalities

- $\mathbf{H}_n^*$ is a finite-dimensional space which captures special position information for all discrete random vectors of a fixed length (but unbounded state spaces).

Entropy region and information inequalities

- $\mathbf{H}_n^*$ is a finite-dimensional space which captures special position information for all discrete random vectors of a fixed length (but unbounded state spaces).
- Applications in cryptography, coding theory, engineering want to optimize linear functions over $\mathbf{H}_n^*$.

Entropy region and information inequalities

- $\mathbf{H}_n^*$ is a finite-dimensional space which captures special position information for all discrete random vectors of a fixed length (but unbounded state spaces).
- Applications in cryptography, coding theory, engineering want to optimize linear functions over $\mathbf{H}_n^*$.
- Elements of the dual cone (information inequalities) can give bounds for optimization problems.

Entropy region and information inequalities

- $\mathbf{H}_n^*$ is a finite-dimensional space which captures special position information for all discrete random vectors of a fixed length (but unbounded state spaces).
- Applications in cryptography, coding theory, engineering want to optimize linear functions over $\mathbf{H}_n^*$.
- Elements of the dual cone (information inequalities) can give bounds for optimization problems.

Theorem ([Mat07])

$\overline{\mathbf{H}_n^*}$ is a convex cone of dimension $2^n - 1$. Furthermore $\mathrm{relint}(\overline{\mathbf{H}_n^*}) \subseteq \mathbf{H}_n^*$.

- Information inequalities completely describe the topological closure of $\mathbf{H}_n^*$ which makes them powerful tools in optimization.

Ingleton inequality

Let A, B, C, D be subspaces in a finite-dimensional vector space.
Then the Ingleton inequality holds for $h = \dim$:

$$\begin{aligned}\mathtt{I}(AB|CD) \coloneqq\, & h(A,C) + h(B,C) + h(A,D) + h(B,D) + h(C,D) - \\ & h(A,B) - h(C) - h(D) - h(A,C,D) - h(B,C,D) \quad \geq \quad 0.\end{aligned}$$

Ingleton inequality

Let A, B, C, D be subspaces in a finite-dimensional vector space.
Then the Ingleton inequality holds for $h = \dim$:

$$\begin{aligned}\mathtt{I}(AB|CD) \coloneqq & h(A,C) + h(B,C) + h(A,D) + h(B,D) + h(C,D) - \\ & h(A,B) - h(C) - h(D) - h(A,C,D) - h(B,C,D) \quad \geq \quad 0.\end{aligned}$$

The Ingleton inequality fails in general for $h = H$ but it has been discovered that certain special position assumptions make it true even in the entropic setting, e.g.,

Ingleton inequality

Let A, B, C, D be subspaces in a finite-dimensional vector space.
Then the Ingleton inequality holds for $h = \dim$:

$$\begin{aligned}\mathtt{I}(AB|CD) \coloneqq & h(A,C) + h(B,C) + h(A,D) + h(B,D) + h(C,D) - \\ & h(A,B) - h(C) - h(D) - h(A,C,D) - h(B,C,D) \quad \geq \quad 0.\end{aligned}$$

The Ingleton inequality fails in general for $h = H$ but it has been discovered that certain special position assumptions make it true even in the entropic setting, e.g.,

- If $C \perp\!\!\!\perp D$ then $\mathtt{I}(AB|CD) \geq 0$.

Ingleton inequality

Let A, B, C, D be subspaces in a finite-dimensional vector space.
Then the Ingleton inequality holds for $h = \dim$:

$$\begin{aligned}\mathtt{I}(AB|CD) \coloneqq\, & h(A,C) + h(B,C) + h(A,D) + h(B,D) + h(C,D) - \\ & h(A,B) - h(C) - h(D) - h(A,C,D) - h(B,C,D) \quad \geq \quad 0.\end{aligned}$$

The Ingleton inequality fails in general for $h = H$ but it has been discovered that certain special position assumptions make it true even in the entropic setting, e.g.,

- If $C \perp\!\!\!\perp D$ then $\mathtt{I}(AB|CD) \geq 0$.
- If $A \perp\!\!\!\perp C \mid D$ and $A \perp\!\!\!\perp D \mid C$ then $\mathtt{I}(AB|CD) \geq 0$.

Ingleton inequality

Let A, B, C, D be subspaces in a finite-dimensional vector space.
Then the Ingleton inequality holds for $h = \dim$:

$$\begin{aligned}\mathtt{I}(AB|CD) \coloneqq h(A,C) + h(B,C) + h(A,D) + h(B,D) + h(C,D) - \\ h(A,B) - h(C) - h(D) - h(A,C,D) - h(B,C,D) \quad \geq \quad 0.\end{aligned}$$

The Ingleton inequality fails in general for $h = H$ but it has been discovered that certain special position assumptions make it true even in the entropic setting, e.g.,

- If $C \perp\!\!\!\perp D$ then $\mathtt{I}(AB|CD) \geq 0$.
- If $A \perp\!\!\!\perp C \mid D$ and $A \perp\!\!\!\perp D \mid C$ then $\mathtt{I}(AB|CD) \geq 0$.
- ...

Ingleton inequality

Let A, B, C, D be subspaces in a finite-dimensional vector space.
Then the Ingleton inequality holds for $h = \dim$:

$$\begin{aligned}\mathtt{I}(AB|CD) \coloneqq\, & h(A,C) + h(B,C) + h(A,D) + h(B,D) + h(C,D) - \\ & h(A,B) - h(C) - h(D) - h(A,C,D) - h(B,C,D) \quad \geq \quad 0.\end{aligned}$$

The Ingleton inequality fails in general for $h = H$ but it has been discovered that certain special position assumptions make it true even in the entropic setting, e.g.,

- If $C \perp\!\!\!\perp D$ then $\mathtt{I}(AB|CD) \geq 0$.
- If $A \perp\!\!\!\perp C \mid D$ and $A \perp\!\!\!\perp D \mid C$ then $\mathtt{I}(AB|CD) \geq 0$.
- ...

These are conditional information inequalities and they can tell apart honest boundary parts of $\mathbf{H}_n^*$ from fake boundary parts on $\overline{\mathbf{H}_n^*}$.

Conditional Ingleton inequalities

Theorem ([KR13] & [Stu21] & [Boe22])

Up to symmetry there are precisely ten minimal sets of conditional independence assumptions on four random variables which ensure $I \geq 0$.

Check out →`https://mathrepo.mis.mpg.de/ConditionalIngleton/`← for non-linear algebra and numerical optimization techniques used in part of the proof.

Conditional Ingleton inequalities

Theorem ([KR13] & [Stu21] & [Boe22])

Up to symmetry there are precisely ten minimal sets of conditional independence assumptions on four random variables which ensure $I \geq 0$.

Check out →`https://mathrepo.mis.mpg.de/ConditionalIngleton/`← for non-linear algebra and numerical optimization techniques used in part of the proof.

Corollary

On four discrete random variables there are precisely 18 478 realizable conditional independence structures. (Laws of information theory)

Conditional Ingleton inequalities

Theorem ([KR13] & [Stu21] & [Boe22])

Up to symmetry there are precisely ten minimal sets of conditional independence assumptions on four random variables which ensure $I \geq 0$.

Check out →`https://mathrepo.mis.mpg.de/ConditionalIngleton/`← for non-linear algebra and numerical optimization techniques used in part of the proof.

Corollary

On four discrete random variables there are precisely 18 478 realizable conditional independence structures. (Laws of information theory)

Problem

Extend this classification to functional dependence assumptions.

Computing the critical locus of $\mathtt{I}$ on $\Delta(2, 2, 2, 2)$

Problem

Find the critical points of the Ingleton functional for four binary random variables.

Computing the critical locus of $\mathtt{I}$ on $\Delta(2,2,2,2)$

Problem

Find the critical points of the Ingleton functional for four binary random variables.

- Computation for a subcase with 8 of the 16 variables using HC.jl [BT18]:

```
Numerical irreducible decomposition with 383 components
* 12 component(s) of dimension 5.
* 15 component(s) of dimension 3.
* 356 component(s) of dimension 1.

 degree table of components:
  dimension |       degrees of components
 -----------+-----------------------------------
      5     | (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...)
      3     | (1, 1, 1, 1, 1, 1, 2, 1, 1, 1, ...)
      1     | (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...)
```

References I

Tobias Boege. *No eleventh conditional Ingleton inequality*. 2022. arXiv: 2204.03971 [cs.IT].

Paul Breiding and Sascha Timme. "HomotopyContinuation.jl: a package for homotopy continuation in Julia". In: *Mathematical software – ICMS 2018. 6th international conference, South Bend, IN, USA, July 24–27, 2018. Proceedings*. Cham: Springer, 2018, pp. 458–465. ISBN: 978-3-319-96417-1; 978-3-319-96418-8. DOI: 10.1007/978-3-319-96418-8_54.

Tarik Kaced and Andrei Romashchenko. "Conditional information inequalities for entropic and almost entropic points". In: *IEEE Trans. Inf. Theory* 59.11 (2013), pp. 7149–7167. ISSN: 0018-9448. DOI: 10.1109/TIT.2013.2274614.

Frantisek Matús. "Piecewise linear conditional information inequality". In: *IEEE Trans. Inf. Theory* 52.1 (2006), pp. 236–238. ISSN: 0018-9448. DOI: 10.1109/TIT.2005.860438.

References II

František Matúš. “Ascending and descending conditional independence relations”. In: *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*. Vol. B. 1992, pp. 189–200.

František Matúš. “Two constructions on limits of entropy functions.”. In: *IEEE Trans. Inf. Theory* 53.1 (2007), pp. 320–330. ISSN: 0018-9448. DOI: 10.1109/TIT.2006.887090.

František Matúš. *Algebraic matroids are almost entropic*. Preprint, accepted to the Proceedings of the AMS. 2017.

Milan Studený. “Conditional independence structures over four discrete random variables revisited: conditional ingleton inequalities”. In: *IEEE Trans. Inf. Theory* 67.11 (2021), pp. 7030–7049. ISSN: 0018-9448. DOI: 10.1109/TIT.2021.3104250.