Pseudoentropy

PhD Dissertation Talk

University of Warsaw

May 23, 2023

This talk

✔️ Overviews the goals, resources, and deliverables of my PhD project.
✔️ Demonstrates/sketches interesting techniques used in the dissertation.
✔️ Defends my position on the dissertation form, in view of reviews received 🛡️.
❌ Avoids complex definitions and proofs for brevity's sake (see the papers) ⏱️.
❌ Does not assess my own academic KPIs (see the documentation) 😳.

Outline

Outline

Credits

I am particularly grateful:

- ❤️ for love, to my wife Aneta
- 💰 for funding, to my advisor Stefan Dziembowski
- 💡 for merit support, to my co-advisor Krzysztof Pietrzak
- 👏 for motivation and recognition, to dozens of people with whom I shared ideas: research collaborators, reviewers, audience 😊

Funding

My PhD research received support from numerous funding sources:

Ideas for Poland

WELCOME

TOCNeT

PRELUDIUM

+ several travel grants from various research institutions

Outline

About Pseudoentropy

Introduced in [ILL89, HILL99] as a **computational variant of information-theoretic entropy**.

Recognized as a **useful tool and convenient language** in research around cryptography, computational complexity and information theory. Examples:

Pseudorandom generators from one-way functions [HILL99]

Computational Dense Model Theorem [RTTV08, Zha11], improving upon the result of Green-Tao-Ziegler

Promising but messy: suffers from **contextual definitions** and **insufficiently developed foundations**.

Goals

My PhD project set these goals:

- 🧹 **improve understanding of foundational properties** of pseudoentropy notions
- ⚙️ **demonstrate further technical applications**
- 💎 optionally, identify **new inspirational application areas**

Contribution

Works presented under the scope of this PhD project:

- ✔ **obtained characterizations and manipulation rules** for pseudoentropy notions, using **convex analysis as a toolbox**
- ✔ **simplified some of existing technical proofs**, for instance of Dense Model Theorem and of Computational Simulators
- ✔ **developed machine-learning inspired framework** for proving computational indistinguishability

My self-assesment:

- 🥅 these works contributed to the goals 🧹, ⚙️ and 💎 respectively.
- 🏃 goals were set broadly, leaving still room for improvement

Outline

Outline

Background

- 🔑 Pseudoentropy at least k when the distribution behaves *nearly as well* as with information-theoretic (min)entropy k in *cryptographic games*.
- 🔑 Program-input games used in definitions
  (a) Distinguish: discriminate between two distributions based on a sample.
  (b) Predict: guess a sampled outcome
  (c) Compress: successfuly decode after decoding

Outline

Outline

- 📖 Indistinguishability quantifies how close are two distributions under a given class of computationally bounded tests.
- ❓ What is the geometrical meaning of indistinguishability?
- ✋ Computational indistinguishability can be **characterized by inseparability by a class of feasible hyperplanes**. The margin of separation can be analytically characterized too!

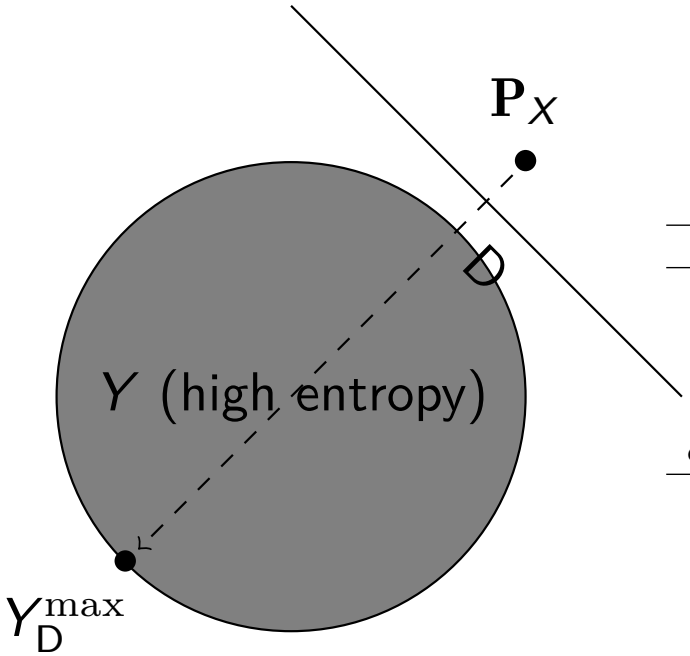
Contribution

The characterization has found the following applications:

- 🔑 Unifying unpredictability-based and indistinguishability-based pseudoentropy notions [SGP15]
- 🔑 Short proof of the Dense Model Theorem with optimal parameters [Sko15]
- 🔑 Further applications to key derivation [Sko17]
- 🔑 Simplifies other technical arguments [VZ12]

Technique (Sketch)

- In program-input indistinguishability games it makes sense to **characterize the optimal input player** Y against a given program player D.
- View D as a *separating hyperplane*, *maximize margin* with high-entropy Y
- The characterization may be complicated, as it depends on:
  (a) the class of feasible distnguishers
  (b) the entropy notion (min/Renyi, conditional/unconditional etc)
- Useful example: for min-entropy and boolean functions...



Symbol/Operator	Crypto	Geometry
X	candidate distribution	
$\mathbf{E}\mathrm{D}(Y)$	expectation	$\mathrm{D} \cdot \mathbf{P}_Y$ (dot-product)
D	distinguisher/program player	separating hyperplane
Y	input player	feasible point
$\epsilon = \mathbf{E}\mathrm{D}(Y) - \mathbf{E}\mathrm{D}(X)$	advantage	margin

Outline

Outline

📖 Applications of pseudoentropy use different notions, most commonly unpredictability-based and indistinguishability-based.

❓ Are unpredictability and indistinguishability entropies different?
Note: usually, distinguishing is easier than predicting[1]

✋ Surprisingly, **equivalent in high-entropy regimes**!

[1] Think of discriminating between dogs and cats versus predicting the breed.

Contribution

The following result was obtained [SGP15]:

- 🔑 **equivalence of unpredictability and indistinguishability** pseudoentropy definitions in **high-entropy regimes**, namely $n - O(\log n)$ for n-bit strings,
- 🔑 **geometric characterizations as a workhorse** of the proof.

Technique (Sketch)

The proof strategy is to *constructively convert a distinguisher into a predictor*:

(a) Indistinguishability fails: $\mathbf{E}\mathsf{D}(X) \geqslant \mathbf{E}\mathsf{D}(Y) + \epsilon$ for all Y of min-entropy k.

(b) $\mathbf{E}\mathsf{D}(X) \geqslant |\mathsf{D}|/2^k + \epsilon$ for boolean D, by geometrical characterizations (!)

(c) Sample A from the image of D, then $\mathbf{P}\{A = X\} > 2^{-k} + \frac{\epsilon}{\#\mathsf{D}}$.

(d) Approximate image sampling by rejection sampling ℓ times, then

$$\mathbf{P}\{A = X\} > \left(2^{-k} + \frac{\epsilon}{\#\mathsf{D}}\right) \cdot \left(1 - \frac{\#\mathsf{D}}{2^n}\right)^{\ell}.$$

(e) $\mathbf{P}\{A = X\} > 2^{-k}$ when $\ell \approx 2^{n-k}/\epsilon$ independently of $\#\mathsf{D}$!

⚠ More sophisticated rejection-sampling handles X with auxiliary input Z.

Outline

Outline

📖 Applications of pseudoentropy assume strength parameters that propagate through reduction proofs.

❓ Can we characterize what quality parameters are non-trivial?

✋ Yes, by time-advantage tradeoffs!

Contribution

The following result was obtained:

- 🔑 generic attacks with time t succeed against psuedoentropy ammount k with advantage $\epsilon = O\left(\sqrt{t/2^k}\right)$
- 🔑 the result generalizes the famous time-advantage tradeoffs against pseudorandomness [DTT10]

Outline

Outline

📖 Applications of pseudoentropy **heavily rely on manipulation rules**, particularly chain rules and transformations [BSW03, FOR12]. Their use **weakens security guarantees**.

❓ Can we improve known manipulation rules?

✋ No, not by black-box reductions!

Contribution

The following results were obtained:

- 🔑 **impossibility of better proofs** by black-box reductions!
- 🔑 the **probabilistic construction of an oracle**, of independent interest, inspired by earlier work limitations of dense model theorems [Zha11]

Outline

Outline

📖 In security proofs it helps to model leakages as explicit functions of secrets.

❓ What leakages can be modelled, without substantial loss in security, as functions of secrets?

✋ Short leakages can be simulated!

Contribution

The following important results were obtained:

- 🔑 Construction of a simulator for m bits of leakage which makes only $2^{O(m)}\epsilon^{-2}$ calls to achieve ϵ-indistinguishability.
- 🏆 The reasoning, inspired by ML techniques, **builds on gradient descent** and was recognized with the *best student paper award at TCC*.

Outline

References I

Boaz Barak, Ronen Shaltiel, and Avi Wigderson.

Computational analogues of entropy.
In *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques, 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2003 and 7th International Workshop on Randomization and Approximation Techniques in Computer Science, RANDOM 2003, Princeton, NJ, USA, August 24-26, 2003, Proceedings*, pages 200–215, 2003.

Anindya De, Luca Trevisan, and Madhur Tulsiani.

Time space tradeoffs for attacks against one-way functions and prgs.
In *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, pages 649–665, 2010.

Benjamin Fuller, Adam O'Neill, and Leonid Reyzin.

A unified approach to deterministic encryption: New constructions and a connection to computational entropy.
In *Theory of Cryptography - 9th Theory of Cryptography Conference, TCC 2012, Taormina, Sicily, Italy, March 19-21, 2012. Proceedings*, pages 582–599, 2012.

Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby.

A pseudorandom generator from any one-way function.
SIAM J. Comput., 28(4):1364–1396, 1999.

R. Impagliazzo, L. A. Levin, and M. Luby.

Pseudo-random generation from one-way functions.
In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 12–24, New York, NY, USA, 1989. ACM.

Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan.

Dense subsets of pseudorandom sets.
In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 76–85, 2008.

References II

Maciej Skorski, Alexander Golovnev, and Krzysztof Pietrzak.

Condensed unpredictability.
In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 1046–1057, 2015.

Maciej Skorski.

Nonuniform indistinguishability and unpredictability hardcore lemmas: New proofs and applications to pseudoentropy.
In *Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings*, pages 123–140, 2015.

Maciej Skorski.

Lower bounds on key derivation for square-friendly applications, 2017.
To appear in the proceedings of STACS'17.

Salil P. Vadhan and Colin Jia Zheng.

Characterizing pseudoentropy and simplifying pseudorandom generator constructions.
In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 817–836, 2012.

Jiapeng Zhang.

On the query complexity for showing dense model.
Electronic Colloquium on Computational Complexity (ECCC), 18:38, 2011.

Outline

Addressing Reviewers Feedback

R: Editorial changes and reference requests.

M: Addressed, thanks for the feedback!

R: A book-style dissertation would be better than a mixture of conference works.

M: I discussed this form with senior researchers, but found *ineffective*:

- 🥕 Gain citations! 🤔 *Time-consuming, better to keep writing papers.*
- 🥕 Get your PhD distinguished. 🤔 *Prestigious conferences not enough?*
- 💦 Take your time to present it better! 🤔 *Why to work harder? We count conference works when granting junior/senior professorships!*

R: Parts of lengthy works might not have been fully reviewed at conferences.

M: Indeed, the same risk as in case of granted junior professorships 😔