Metric Pseudoentropy: Characterizations and Applications

Maciej Skorski
maciej.skorski@gmail.com

Cryptology and Data Security Group, University of Warsaw

Abstract. Metric entropy is a computational variant of entropy, often used as a convenient substitute of HILL Entropy, slightly stronger and standard notion for entropy in cryptographic applications. In this paper we develop a general method to characterize metric-type computational variants of entropy, in a way depending only on properties of a chosen class of test functions (adversaries). As a consequence, we obtain a nice and elegant geometric interpretation of metric entropy. We apply these characterization to simplify and modularize proofs of some important results, in particular: (a) computational dense model theorem, (b) derivation of the improved version of Leftover Hash Lemma and (c) equivalence between unpredictability entropy and HILL entropy for short strings.

1 Introduction

1.1 Computational Entropy

Entropy. Entropy, as a measure of uncertainty or randomness, is a fundamental notion in information-theory. The most known metric of entropy is Shannon Entropy [Sha48]. For cryptographic applications such as *extracting randomness*, it is more convenient to work with so called min-entropy, which gives an upper bound on the probability that computationally unbounded adversary can guess a value sampled according to a given distribution. A slightly weaker but also very useful, especially in the context of *hashing*, is the notion of collision entropy which upperbounds the probability that two independent samples of a given distribution collide.

Defining computational variants of entropy. Computational analogues of entropy can be defined in different ways. In any case, we need to formalize that a distribution has, from a computational point of view, the same of almost the same properties like a distribution having "true" information-theoretic entropy. This might be based on hardness of compressing-decompressing, hardness of prediction or hardness of distinguishing. In this paper we follow the last approach, which is most widely used. A good survey of different entropy notions and their properties can be found in [BSW03] and [Rey11]. We stress that, contrarily to the information-theoretic case, for computational entropy it's not only the *amount* of entropy that matters but also its *quality* is important.

Computational Indistinguishability. Indistinguishability is a fundamental concept in computational complexity and cryptography. For two distributions X, Y taking values in the same space, a class $\mathcal{D}$ of $[0,1]$-valued functions (refereed to as the "attackers class") and a parameter ϵ (refereed to as the "distinguishing advantage"), we say that X and Y are $(\mathcal{D}, \epsilon)$-indistinguishable if for all $\mathrm{D} \in \mathcal{D}$ we have $|\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)| \leqslant \epsilon$. An attacker D can distinguish X and Y if $\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y) > 0$ or $\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y) < 0$, and the far from 0 this difference is, the better "advantage" he achieves. Sometimes we want to define indistinguishability between two sets $\mathbb{X}$ and $\mathbb{Y}$ of probability distributions. We can formalize this by saying that no single adversary D can achieve bigger than 0 advantage for *every* pair (X, Y) where X comes from $\mathbb{X}$ and Y comes from $\mathbb{Y}$. Since the expectation $\mathbf{E}\,\mathrm{D}(X)$ can be thought as the scalar product of vectors representing D and the distribution of X, the concept of indistinguishability is *exactly* the same concept as the idea of *separating hyperplanes*.

Computational Entropy. Having formalized the concept of "computational closeness", one can define the "computational" entropy, called also pseudoentropy, of a distribution X by one of the following ways:

(a) (stronger) X is computationally indistinguishable from a *single* distribution having required amount of information-theoretic entropy (min-entropy, Shannon Entropy etc.)
(b) (weaker) is computationally indistinguishable from a *set* of all distributions having required amount of information-theoretic entropy.

Both approaches turn out to be useful. Setting the underlying information-theoretic entropy measure to be the min-entropy, for case (a) we obtain the notion of HILL entropy [HILL99] which directly generalizes the notion of pseudorandomness, whereas for case (b) we get the notion of the so called Metric Entropy [BSW03]. Roughly speaking, with HILL entropy one generalizes most of information-theoretic facts about entropy, into the computational setting. Metric entropy is commonly thought as a less intuitive and understood notion than HILL entropy. Quite surprisingly it has been proven to be technically more convenient in many problems. The typical approach is to work with metric entropy and to convert it to HILL entropy (which is possible with some loss in quality [BSW03]). For example, the use of metric entropy simplifies and improves the proof of the computational variant of the dense model theorem [BL12], applicable in leakage-resilient cryptography [DP08]. Notions of pseudoentropy have found also important applications in general complexity theory, for example in [VZ12] a HILL-like variant of Shannon entropy is used to simplify the construction of a PRG from a one-way function. These two examples show also that the notion of pseudoentropy is a key ingredient of important or even breakthrough results and as such is worth of studying.

Worst Case Distributions. In problems which involve computational indistinguishability it is often convenient to know the distributions which makes the attacker's advantage maximal. This distribution is typically subjected to some entropy restrictions. In particular, one might ask the following question

> Given D and X, what is the best (minimal) attacker advantage $|\Delta^D| = |\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)|$ over all distributions Y of entropy as least k?

An answer to this question yields a bound on how (computationaly) close is X to the set of all distributions of entropy k. Such problems arises naturally where one uses HILL and Metric entropy, see for instance [BSW03, CKLR11, VZ12].

1.2 Our Results

Summary of our contribution. As mentioned, the concept of characterizing the "worst case" distribution which optimizes the attacker advantage is very common, thought not always explicitly stated [BSW03, CKLR11, BL12, RTTV08]. In this paper we give a uniform treatment of this idea and use to obtain characterizations for pseudoentropy and other interesting corollaries.

Characterizing Metric Pseudoentropy via Optimizing Attacker's Advantage. Using standard constrained optimization techniques, we develop a general method to characterize metric-type pseudoentropy. A characterization is based on *explicitly* calculating the distribution which minimizes the attacker's advantage, subject to entropy constraints. These characterizations could be used in studying properties of variants of pseudoentropy based on entropy different than min-entropy. In particular, they could be applied in studying the problem of comparing the amount of metric pseudoentropy against *deterministic* and *randomized* adversaries, or verifying the so called "chain rule". We also unify the definitions of metric and HILL entropy in a nice geometric way.

Applications: the power of pseudoentropy characterizations. Our technique leads to interesting corollaries besides the basic properties of pseudoentropy. From the characterization of metric pseudo-entropy we immediately obtain the computational Dense Model Theorem [RTTV08, DP08, BL12] Extending our characterization into the conditional case when side information is available to the attacker, we reprove equivalence between unpredictability and indistinguishability based definition of pseudoentropy for short strings [VZ12]. Finally, from the characterization of collision-pseudoentropy we derive the improved Leftover Hash Lemma [BDK+11]. Our results show that metric entropy is a powerful tool which deserves the systematic study.

2 Preliminaries

Entropy notions. The min-entropy of a distribution X equals $\mathbf{H}_\infty(X) = -\log(\max_x \Pr[X = x])$. The collision entropy of X is $\mathbf{H}_2(X) = -\log(\sum_x \Pr[X = x]^2)$. If there is side information Z, we define the average conditional min-entropy [DORS08] of X given Z by $\widetilde{\mathbf{H}}_\infty(X|Z) = -\log(\mathbf{E}_{z\leftarrow Z} \max_x \Pr[X = x|Z = z])$.

Computational advantage. The advantage of an attacker D in distinugishing random variables X and Y, which take values in the same space, is defined to be $\Delta^{\mathrm{D}}(X;Y) = \mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)$.

Computational Entropy. There is many ways to define computational analogues of entropy. We follow the most popular approach, which is based on the concept of computational indistinguishability.

Definition 1 (HILL Pseudoentropy [HILL99]). *Let X be a distribution with the following property: there exists Y of min-entropy at least k such that for all circuits* D *of size at most s we have $|\Delta^{\mathrm{D}}(X;Y)| \leqslant \epsilon$. Then we say that X has k bits of HILL min-entropy of quality (s,ϵ) and denote by $\mathbf{H}_{\infty}^{\mathrm{HILL},(s,\epsilon)}(X) \geqslant k$.*

Remark 1 (HILL entropy against different circuits classes). It is known that for HILL entropy all kind of circuits: deterministic boolean, deterministic real valued and randomized boolean, are equivalent (for the same size s). That's why we can abbreviate the notation and omit declaring circuits type in Definition 1.

Definition 2 (Metric Pseudoentropy [BSW03]). *Let X be a distribution with the following property: for every deterministic boolean (respectively: deterministic real valued or boolean randomized) circuit* D *of size at most s there exists Y of min-entropy at least k such that $|\Delta^{\mathrm{D}}(X;Y)| \leqslant \epsilon$. Then we say that X has k bits of deterministic (respectively: deterministic real valued or boolean randomized) metric min-entropy of quality (s,ϵ) and denote by $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X)$ (respectively: $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X)$ and $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{rand}\{0,1\},(s,\epsilon)}(X)$).*

Definitions of HILL and metric entropy for entropy notions different than min-entropy, for instance collision entropy can be obtained by replacing min-entropy with collision entropy in Definition 1 and Definition 2.

Remark 2 (Metric Entropy against different circuits class). For metric min-entropy, it does not matter if the deterministic circuits are boolean or real valued (see [Rey11] and the errata of [BSW03]). However, this is not true for the conditional case and does not extend to other entropy notions.

Computational Entropy - Side information. Sometimes we assume that information Z correlated to X might be available to an adversary.

Definition 3 (Conditional HILL Pseudoentropy [HLR07]). *Let X, Z be a joint distribution with the following property: there exists Y of average conditional min-entropy at least k given Z such that for all circuits* D *of size at most s we have $|\Delta^{\mathrm{D}}(X,Z;Y,Z)| \leqslant \epsilon$. Then we say that X given Z has k bits of HILL min-entropy of quality (s,ϵ) and denote by $\mathbf{H}_{\infty}^{\mathrm{HILL},(s,\epsilon)}(X|Z) \geqslant k$.*

Remark 3 (HILL entropy against different circuits classes). Similarly to Remark 2, here all kinds of circuits: deterministic boolean, deterministic real valued and randomized boolean, are equivalent (for the same size s).

Definition 4 (Conditional Metric Pseudoentropy [BL12]). *Let X, Z be a joint distribution with the following property: for every deterministic boolean (respectively: deterministic real valued or boolean randomized) circuit* D *of size at most s there exists Y of average conditional min entropy at least k given Z*

such that $|\Delta^{\mathrm{D}}(X,Z;Y,Z)| \leqslant \epsilon$. *Then we say that* X *given* Z *has* k *bits of deterministic (respectively: deterministic real valued or boolean randomized) metric min-entropy of quality* (s,ϵ) *and denote by* $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X|Z)$ *(respectively:* $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X|Z)$ *and* $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{rand}\{0,1\},(s,\epsilon)}(X|Z)$*).*

There is a variant of conditional pseudoentropy where (X,Z) is required to be computationally close to (Y,Z') but Z' is not necessarily the same as Z. This notion is called the "relaxed" HILL entropy [Rey11] and denoted by $\mathbf{H}^{\mathrm{HILL-rlx},(s,\epsilon)}(X)$ (for metric variants $\mathbf{H}^{\mathrm{M-rlx},\mathrm{det}\{0,1\},(s,\epsilon)}(X)$ and $\mathbf{H}^{\mathrm{M-rlx},\mathrm{det}[0,1],(s,\epsilon)}(X)$). Typically we want Z to be the same as Z'[1] but this relaxed notion is also useful [GW10, Rey11]. It satisfies the so called chain rule, a property desired in leakage-resilient cryptography, which doesn't hold for HILL entropy [KPW13].

Relations between HILL and Metric Pseudoentropy. For any "reasonable" notion of (information-theoretic) entropy, metric and HILL variants are equivalent up to some loss in quality parameters s,ϵ.

Lemma 1 (HILL vs Metric Pseudoentropy, [BSW03]). *Let* $\mathbf{H}$ *be an entropy notion which is concave*[2]*. Then for any* n*-bit random variable* X *we have*

$$\mathbf{H}^{\mathrm{HILL},(s',\epsilon')}(X) \geqslant \mathbf{H}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X)$$

where $\delta \in (0,1)$ *is arbitrary,* $s' = \mathcal{O}\left(s \cdot \delta^2/n\right)$ *and* $\epsilon' = \epsilon + \delta$*. The same is true for conditional pseudoentropy and relaxed pseudoentropy, with* $s' = \mathcal{O}\left(s \cdot \frac{\delta^2}{n+m}\right)$ *where* m *is the length of* Z*.*

3 Characterizing Metric Pseudoentropy

In what follows we assume that $\mathbf{H}$ is a concave entropy notion (like min-entropy or collision entropy), and that all distributions and distinguishers are over $\{0,1\}^n$.

3.1 Connections to separating hyperplanes

We start with the following simple observation, which gives a nice geometrical formulation of the definition of pseudo-entropy. We say that the sets $\mathbb{X}$ and $\mathbb{Y}$ of probability distributions are $(\mathcal{D},\epsilon)$-indistinguishable if there exists *no* adversary D such that $|\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)| \geqslant \epsilon$ for all $X \in \mathbb{X}$ and all $Y \in \mathbb{Y}$. It is easy to see that if $\mathbb{X}$ and $\mathbb{Y}$ are convex and if $\mathcal{D}$ is closed under complements (that is $\mathrm{D} \in \mathcal{D}$ implies $\mathbf{1} - \mathrm{D} \in \mathcal{D}$) then this is equivalent to

There is no $\mathrm{D} \in \mathcal{D}$ such that: $\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y) \geqslant \epsilon$ for all $X \in \mathbb{X}, Y \in \mathbb{Y}$.

[1] For instance, when Z represents information that adversary might have learned

[2] That is, a convex combination of distributions with entropy at least k is a distribution with entropy at least k. This assumption is fulfilled for most notions, for example for all Renyi entropies which include min-entropy and collision entropy

We can interpret the expectation $\mathbf{E}\,\mathrm{D}(X)$ as the scalar product $\langle \mathrm{D}, \mathbf{P}_X \rangle$ by identifying D and distributions of X with the vectors in $\mathbb{R}^{2^n}$. Hence we can write the above condition as

$$\text{There is no } \mathrm{D} \in \mathcal{D} \text{ such that: } \langle \mathrm{D}, \mathbf{P}_X - \mathbf{P}_Y \rangle \geqslant \epsilon \text{ for all } X \in \mathbb{X}, Y \in \mathbb{Y},$$

which means that the distinguisher D is precisely a *separating hyperplane*. If $\mathcal{D}$ is a circuit class, $\mathbb{X} = \{X\}$ and $\mathbb{Y} = \{Y : \mathbf{H}(Y) \geqslant k\}$ we obtain[3]

Corollary 1 (Alternative definitions of metric and HILL entropy). *Let X be an n-bit random variable and let $\mathbf{H}$ be a concave entropy notion. Then*

(a) $\mathbf{H}^{\mathrm{HILL},(s,\epsilon)}(X) \geqslant k$ *iff X is $(\mathcal{D}, \epsilon)$-indistinguishable from* some *Y of entropy $\mathbf{H}$ at least k, where $\mathcal{D}$ is the class of boolean circuits*[4] *of size s with n-inputs.*
(b) $\mathbf{H}^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$ *iff X is $(\mathcal{D}, \epsilon)$-indistinguishable from* the set *of* all *Y of entropy $\mathbf{H}$ at least k,*

where $\mathcal{D}$ is the class of all deterministic boolean circuits of size s with n-inputs (analogously for randomized and deterministic real valued circuits).

3.2 Reduction to constrained optimization

By the "geometric" view on pseudoentropy, given in Corollary 1, we obtain the following characterization of pseudoentropy.

Lemma 2 (Characterization of metric pseudoentropy). *Let X and $\mathbf{H}$ be as in Corollary 1. Then $\mathbf{H}^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$, respectively $\mathbf{H}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$ if and only if for every boolean (respectively real valued) deterministic circuit* D *of size at most s we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) + \epsilon,$$

where Y^* is optimal to the following optimization problem

$$\begin{array}{ll} \underset{Y}{\text{maximize}} & \mathbf{E}\,\mathrm{D}(Y) \\ \text{s.t.} & \mathbf{H}(Y) \geqslant k \end{array}. \tag{1}$$

This results is useful if we can solve the optimization problem in Equation (1). In the next subsections we explain how to solve it in general and discuss the two concrete and simple cases: min-entropy and collision entropy.

[3] We can assume that the class circuits of size at most s is closed under complements because every complement is of size at most $s + 1$. Formally we need to start with size $s' = s + 1$ but we omit this negligible difference
[4] Randomized or deterministic- it makes no difference

3.3 Maximizing expectations under convex constraints

We can characterize optimal solutions of (1) in terms of Lagrange multipliers. Due to convexity, the characterization is both: necessary and sufficient.

Lemma 3 (Maximizing expectation under convex constraints). *Let f be a differentiable convex real-valued function on $\mathbb{R}^d$. Assume that a is a number such that $\min_p f(p) < a$ where the minimum is over all probability vectors, and consider the following optimization program*

$$\begin{aligned} \underset{(p_i)_i}{\text{maximize}} \quad & \sum_i \mathrm{D}_i p_i \\ \text{s.t.} \quad & \begin{cases} f(p) \leqslant a \\ -p_i \leqslant 0 \\ \sum_i p_i = 1 \end{cases}. \end{aligned} \tag{2}$$

Then a feasible point $p = p^*$ is optimal to (2) if and only if there exist $\lambda_1 \geqslant 0, \lambda_2 \geqslant 0$ and $\lambda_{3i} \in \mathbb{R}$ for $i = 1, \ldots, m$ such that the following relations hold

$$\mathrm{D}_i = \lambda_1 (\nabla f(p^*))_i - \lambda_{3i} + \lambda_2 \quad \text{for } i = 1, \ldots, m \tag{3}$$

and the following complementary condition is satisfied:

$$p_i \cdot \lambda_{3i} = 0 \tag{4}$$

Proof. The Slater Constraint Qualification holds, by the assumption on a, and we have strong duality. In other words, the first order Karush-Kuhn-Tucker condition is sufficient and necessary [BV04]. The numbers $\lambda_1, \lambda_2, \lambda_{3i}$ are exactly KKT multipliers for the convex program in Equation (2), and Equation (3) states that the gradient of the objective function is a combination of gradients of constraints. The condition in Equation (4) means that we take only active constraints into account. Finally, to the inequality constraints we assign non-negative multipliers which explains the requirement $\lambda_1 \geqslant 0$ and $\lambda_{3i} \geqslant 0$. □

Remark 4. If f is not differentiable, we replace the gradient of f in optimality conditions by the *subdifferential* of f, which always exists for a convex function.

3.4 Characterization of metric min entropy

For $\mathbf{H} = \mathbf{H}_\infty$ we obtain from Lemma 3 the following simple characterization of pseudoentropy based on min-entropy (see [BSW03] for a restricted variant)

Theorem 1 (Characterization of metric min-entropy). *Let X be an n-bit r.v.. Then $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$, respectively $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$ if and only if for every boolean (respectively real valued) deterministic circuit D of size at most s with n inputs we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) + \epsilon,$$

where Y^* is uniform over the set of 2^k values of x which correspond to the biggest values of $\mathrm{D}(x)$.

Extending Lemma 3 by adding additional constraints, to cover the case of side information, we obtain the characterization of conditional metric entropy

Theorem 2 (Characterization of conditional metric min-entropy). *Let X and Z be, respectively, n and m-bit random variables. Then $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$ (respectively $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$) iff for every boolean (respectively real valued) deterministic circuit D of size at most s on $\{0,1\}^{n+m}$ we have*

$$\mathbf{E}\,\mathrm{D}(X,Z) \leqslant \mathbf{E}\,\mathrm{D}(Y^*,Z) + \epsilon,$$

for Y^* such that $Y^*|Z=z$ is uniform over the set $\{\mathrm{D}(x,z) \geqslant t(z)\}$ for every z, where the thresholds $t(z)$ satisfy the following two conditions

$$\mathop{\mathbf{E}}_{x\leftarrow U_n} \mathbf{E}\max(\mathrm{D}(x,z) - t(z)) = \mathrm{const} \quad \text{for all } z$$

$$\mathop{\mathbf{E}}_{z\leftarrow Z}\left[1/\#\left\{x : \mathrm{D}(x,z) \geqslant t(z)\right\}\right] \leqslant 2^{-k} \leqslant \mathbf{E}\left[1/\#\left\{x : \mathrm{D}(x,z) > t(z)\right\}\right].$$

3.5 Characterization of metric collision entropy

The characterization of the worst-case collision entropy distribution is slightly different. It is *proportional* to a distinguisher, after taking a threshold.

Theorem 3 (Characterization of metric collision entropy). *Let X be an n-bit r.v.. Then $\mathbf{H}_2^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$, respectively $\mathbf{H}_2^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$ if and only if for every boolean (respectively real valued) deterministic circuit D of size at most s with n inputs we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) + \epsilon,$$

where Y^* satisfies $\lambda \cdot \mathbf{P}_{Y^*}(x) = \max(\mathrm{D}(x) - t, 0)$ for some $t \in \mathbb{R}$ and $\lambda \geqslant 0$.

Remark 5. Note that t is a solution of $\mathbf{E}\,\mathrm{D}'(U)^2 = 2^{n-k}\left(\mathbf{E}\,\mathrm{D}'(U)\right)^2$ where $\mathrm{D}'(x) = \max(\mathrm{D}(x) - t, 0)$ and $\lambda = 2^n\,\mathbf{E}\,\mathrm{D}'(U)$. It follows that $\mathbf{E}\,\mathrm{D}'(Y^*) = 2^{n-k}\,\mathbf{E}\,\mathrm{D}'(U) = \mathbf{E}\,\mathrm{D}'(U) + \sqrt{\mathrm{Var}\mathrm{D}'(U)} \cdot \sqrt{2^{n-k}-1}$.

4 Applications

4.1 Computational Dense Model Theorem

We say that a distribution A is γ-dense in B if we have $\Pr[A = x] \leqslant \Pr[B = x]/\gamma$. The Dense Model Theorem is the statement of the following form: if X is (s,ϵ)-indistinguishable from the uniform distribution R and X' is γ-dense in X, then there exists a distribution R' which is γ-dense in R and is (s',ϵ')-indistinguishable from X', where s' and ϵ' depends as explicit functions on s

and ϵ. In this sense, R is a dense "model" for X'. The dense model theorem was proved first by Tao and Ziegler [TZ08]. It's efficient versions[5] have found important applications in complexity theory and cryptography [RTTV08, DP08, BL12], see also [TTV09]. Below we recall a version with improved parameters, stated in language of pseudoentropy and called the "leakage lemma":

Theorem 4 (Leakage Lemma [DP08, BL12]). *Let X be an n-bit random variable such that $\mathbf{H}_\infty^{\mathrm{HILL},(s,\epsilon)}(X) \geqslant k$ and let Z be correlated with X. Then we have $\mathbf{H}_\infty^{\mathrm{HILL},(s',\epsilon')}(X|_{Z=z}) \geqslant k'$ where $k' = k - \log(1/\Pr[Z=z])$, $s' = \mathcal{O}\left(s \cdot \delta^2/n\right)$ and $\epsilon' = \epsilon/\Pr[Z=z] + \delta$, for any $\delta \in (0,1)$.*

The lemma states that the amount of pseudoentropy due to leakage of t bits of information decreases roughly by t, hence its name. The original proof was simplified by the use of metric entropy [BL12]. We show how it can be simplified even further: just few lines using the basic facts about metric entropy!

Proof. If we can prove that

$$\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon/\Pr[Z=z]))}(X|_{Z=z}) \geqslant \mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) - \log(1/\Pr[Z=z])$$

then the result will follow by Lemma 1 and Remark 2. Note that by Theorem 1 for any X we have $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$ if and only if $\mathbf{E}\,\mathrm{D}(X) \leqslant \frac{|\mathrm{D}|}{2^k} + \epsilon$ for all boolean D of size at most s. From this we get

$$\mathbf{E}\,\mathrm{D}(X|_{Z=z}) \leqslant \mathbf{E}\,\mathrm{D}(X)/\Pr[Z=z] \leqslant |\mathrm{D}|/2^k \Pr[Z=z] + \epsilon/\Pr[Z=z]$$

for any D. Since the characterization is also sufficient, the results follows. □

4.2 Equivalence of HILL Entropy and Unpredictability Entropy for short strings

Unpredictability entropy. The notion of unpredictability entropy is based on the (assumed) hardness of guessing X given auxiliary information Z. More formally, we have $\mathbf{H}^{\mathrm{Unp},s}(X|Z) \geqslant k$ if and only if no adversary of size at most s can predict X given Z better than with probability 2^{-k}. For Z independent of X or of the relatively short length, this reduces to the min-entropy of X[6].

Seperation from HILL entropy. If f is a one-way function, U is the uniform distribution and $X = U, Z = f(U)$ then we see that $X|Z$ has large amount of unpredictability. It is also easy to see that $X|Z$ has almost no HILL entropy.

Equivalence for short strings. On the positive side, using metric entropy and the characterization in Theorem 2, we reprove the following result of Vadhan and Zheng who established the equivalence when X is short[7]

Theorem 5 ([VZ12]). *Suppose that X and Z are, respectively, n and m-bit random variables. Then $\mathbf{H}_\infty^{\mathrm{HILL},(s',\epsilon)}(X|Z) \gtrsim \mathbf{H}^{\mathrm{Unp},s}(X|Z)$ with $s' = \frac{s}{\mathrm{poly}(2^n, 1/\epsilon)}$.*

[5] With the loss at most $\mathrm{poly}(1/\delta)$ in s and ϵ. In the original proof the loss is $\exp(1/\delta)$

[6] Provided that $s > 2^m n$ so that the adversary can hardcore his best guess.

[7] Logarithmically in the security parameter

The original proof is based on a result similar to Theorem 2 proved in a much more complicated way. We note that this part is a trivial consequence of KKT optimality conditions and also simplify the rest of the proof.

Proof (Sketch). We prove that $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}[0,1],(s',\epsilon)}(X|Z) < k$ implies $\mathbf{H}^{\mathrm{Unp},s}(X|Z) < k$. Suppose not, then we have $\mathbf{E}\,\mathrm{D}(X,Z) - \mathbf{E}\,\mathrm{D}(Y,Z) \geqslant \epsilon$ for all Y such that $\widetilde{\mathbf{H}}_\infty(X|Z) \geqslant k$. Let Y^* be the distribution which minimizes this expression, that is which maximizes $\mathbf{E}\,\mathrm{D}(Y,Z)$. Let $t(z)$ be as in Theorem 2 and denote $\mathrm{D}'(x,z) = \max(\mathrm{D}(x,z) - t(z), 0)$ and let $\lambda = \sum_x \mathrm{D}'(x,z)$ (according to Theorem 2 this sum does not depend on z). Consider the following predictor A:

$$\text{On input } z \text{ sample } x \text{ according to the probability } \Pr[A(z) = x] = \mathrm{D}'(x,z)/\lambda$$

Note that $Y^*|_{Z=z}$ is uniform over the set $\{x : \mathrm{D}'(x,z) > 0\}$. By Theorem 2 (the sufficiency part) it follows that Y^* is also maximal for D. For every z we have $\mathbf{E}\,\mathrm{D}'(Y^*|_{Z=z},z) = \mathbf{E}\,\mathrm{D}(Y^*|Z=z,z) - t(z)$. We have also $\mathbf{E}\,\mathrm{D}'(X|_{Z=z},z) \geqslant \mathbf{E}\,\mathrm{D}(X|_{Z=z},z) - t(z)$ by the definition of D'. This proves

$$\mathbf{E}\,\mathrm{D}'(X,Z) - \mathbf{E}\,\mathrm{D}'(Y,Z) \geqslant \epsilon \text{ for all } Y \text{ such that } \widetilde{\mathbf{H}}_\infty(X|Z) \geqslant k.$$

It is easy to observe that

$$\Pr_{z\leftarrow Z}[A(Z) = X] = \frac{\mathbf{E}\,\mathrm{D}'(X,Z)}{\lambda} > \mathop{\mathbf{E}}_{z\leftarrow}\left[\frac{\mathbf{E}\,\mathrm{D}'(Y|_{Z=z},z)}{\sum_x \mathrm{D}'(x,z)}\right] \geqslant \mathop{\mathbf{E}}_{z\leftarrow Z} 2^{-\mathbf{H}_\infty(Y^*|_{Z=z})}$$

which is at least 2^{-k}. The circuit $\mathrm{D}'(x,z)$ is of complexity $2^m \cdot \mathrm{size}(\mathrm{D})$, which is too big. However, if the domain of x is small, we can approximate the numbers $t(z)$ given λ from relations in Theorem 2 (and even λ, from the second relation, for the uniform setting). Indeed, knowing that $\mathbf{E}\max(\mathrm{D}(U,z) - t(z)) = \lambda$, we estimate $\mathbf{E}\max(\mathrm{D}(U,z) - t)$ for fixed t and then find a "right" value $t = t(z)$ by the binary search. This way for every z we can approximate $\mathrm{D}'(\cdot,z)$, and hence the distribution $\Pr[A(z) = x]$, up to a maximal error $\delta \ll 2^{-k}$ and with overwhelming probability $1 - \exp(-\mathrm{poly}(1/\delta))$, using $\mathrm{poly}(1/\delta)$ samples of D. On average over z we predict X with probability $2^{-k} - \delta \approx 2^{-k}$. □

4.3 Improved Leftover Hash Lemma for square-secure applications

In the key derivation problem we want to derive a secure m-bit key for some application P from an *imperfect* source of randomness X. The generic approach is to use a randomness extractor. However, as implied by the RT-bounds [RTS00], the min-entropy in X needs to be at least $m + 2\log(1/\epsilon)$ if we want the derived key to be ϵ-secure. Fortunately, as shown by Barak et. al [BDK+11], for many cryptographic applications, one can reduce this loss by half, that is to $L = \log(1/\epsilon)$. To this end, they introduce the class of *square-secure* applications, where the squared advantage, over the uniform choice of keys, of every

bounded attacker is small[8]. This class contains for example all unpredictability applications, stateless chosen plaintext attack secure encryption and weak pseudo-random functions. The reduction of entropy loss follows by combining universal hashing with the following lemma

Lemma 4 ([BDK+11]). *For a function* $\mathrm{D} : \{0,1\}^\ell \to [-1,1]$ *and* $X \in \{0,1\}^\ell$ *of collision entropy* k *we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(U_\ell) + \sqrt{\mathrm{Var}\mathrm{D}(U_\ell)} \cdot \sqrt{2^{\ell-k}-1}.$$

To see this, let $\mathrm{Win}_{\mathsf{A}}(r,h)$, for arbitrary attacker $\mathsf{A} \in \mathcal{A}$, be the probability that A breaks the key r given in addition[9] h and let $\mathrm{D}_{\mathcal{A}}(r,h) = \mathrm{Win}_{\mathsf{A}}(r,h) - \frac{1}{2}$ be its advantage. Let X be any n-bit random variable of min-entropy $m + \log(1/\epsilon)$. We apply a randomly chosen universal hash function[10] H from n to m bits. It is easy to see that $H(X), H$ is a distribution with collision entropy $m + \log|\mathcal{H}| - \log(1+\epsilon)$. From the lemma it follows now that

$$\mathbf{E}\,\mathrm{D}_{\mathsf{A}}(H(X),H) \leqslant \mathbf{E}\,\mathrm{D}_{\mathsf{A}}(U,H) + \sqrt{\mathrm{Var}\mathrm{D}_{\mathsf{A}}(U,H)} \cdot \sqrt{\epsilon}$$

If we assume that $\max_h \mathbf{E}\,\mathrm{D}_{\mathsf{A}}(U,h) \leqslant \epsilon$ (which means ϵ-security against $\mathcal{A}$ with the uniform key) and that $\max_h \mathbf{E}\,\mathrm{D}_{\mathsf{A}}(U,h)^2 \leqslant \sigma$ with $\sigma = \mathcal{O}(\epsilon)$ (which means σ-square-security against $\mathcal{A}$ with the uniform key) then we achieve $\mathcal{O}(\epsilon)$ security for the *extracted* key, with entropy loss only $\log(1/\epsilon)$.

An alternative proof. We show that Theorem 3 implies Lemma 4. Indeed, set $k = \ell$ and $\epsilon = 0$ in Theorem 3. Let Y^* be the distribution of collision entropy at least $k = \ell$ which maximizes $\mathbf{E}\,\mathrm{D}(Y)$, and let t, λ and D' be as in the characterization. Denote $S = \{x : \mathrm{D}(x) \geqslant t\}$ and let $\mathrm{D}|_S$ be the restriction of D to the set S. Note that $Y^*|S \stackrel{d}{=} Y^*$ maximizes $\mathrm{D}|_S$ and $\mathrm{D}|_S(x) = \mathrm{D}'|_S(X) + t$ for every $x \in S$. By Remark 5 we get

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) = \mathbf{E}\,\mathrm{D}|_S(Y^*|S) = \mathbf{E}\,\mathrm{D}|_S(U_S) + \sqrt{\mathrm{Var}\mathrm{D}_S(U_S)} \cdot \sqrt{|S|2^{-k}-1}.$$

We show that one can replace S by the $\{0,1\}^\ell$ on the right hand side. This will follow by the following general lemma

Lemma 5. *Let* X *be a random variable,* $c > 1$ *be a constant and* S *be an event of probability* $\mathbf{P}(S) > c^{-1}$*. Then*

$$\mathbf{E}[X|S] + \sqrt{\mathrm{Var}[X|S]} \cdot \sqrt{c\mathbf{P}(S) - 1} \leqslant \mathbf{E}[X] + \sqrt{\mathrm{Var}[X]} \cdot \sqrt{c-1} \qquad (5)$$

The proof follows by a few algebraic manipulations and is given in Appendix A.

[8] Which essentially means that the probability that an attacker break the key is concentrated over keys

[9] For the uniformly chosen key this doesn't help the adversary, at least in the nonuniform model

[10] A family $\mathcal{H}$ functions from n to m bits is universal if $\Pr_{h \leftarrow \mathcal{H}}[h(x) = h(x')] = 2^{-m}$ for $x \neq x'$

4.4 Some further applications

Lower bounds on square security. Using the characterization of metric collision entropy Theorem 3 one can derive some non-trivial lower bounds on square-security needed for key derivation. We discuss this problem in a separate paper.

References

BDK+11. Boaz Barak, Yevgeniy Dodis, Hugo Krawczyk, Olivier Pereira, Krzysztof Pietrzak, Francois-Xavier Standaert, and Yu Yu, *Leftover hash lemma, revisited*, Cryptology ePrint Archive, Report 2011/088, 2011, http://eprint.iacr.org/.

BL12. Fuller Benjamin and Reyzin Leonid, *A unified approach to deterministic encryption: New constructions and a connection to computational entropy*, TCC 2012, volume 7194 of LNCS, Springer, 2012, pp. 582–599.

BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy.*, RANDOM-APPROX (Sanjeev Arora, Klaus Jansen, Jos D. P. Rolim, and Amit Sahai, eds.), Lecture Notes in Computer Science, vol. 2764, Springer, 2003, pp. 200–215.

BV04. Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, New York, NY, USA, 2004.

CKLR11. Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, *Memory delegation*, Cryptology ePrint Archive, Report 2011/273, 2011, http://eprint.iacr.org/.

DORS08. Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, SIAM J. Comput. **38** (2008), no. 1, 97–139.

DP08. Stefan Dziembowski and Krzysztof Pietrzak, *Leakage-resilient cryptography in the standard model*, IACR Cryptology ePrint Archive **2008** (2008), 240.

GW10. Craig Gentry and Daniel Wichs, *Separating succinct non-interactive arguments from all falsifiable assumptions*, Cryptology ePrint Archive, Report 2010/610, 2010, http://eprint.iacr.org/.

HILL99. Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *A pseudorandom generator from any one-way function*, SIAM J. Comput. **28** (1999), no. 4, 1364–1396.

HLR07. Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin, *Conditional computational entropy, or toward separating pseudoentropy from compressibility*, Proceedings of the 26th annual international conference on Advances in Cryptology (Berlin, Heidelberg), EUROCRYPT '07, Springer-Verlag, 2007, pp. 169–186.

KPW13. Stephan Krenn, Krzysztof Pietrzak, and Akshay Wadia, *A counterexample to the chain rule for conditional hill entropy*, Theory of Cryptography (Amit Sahai, ed.), Lecture Notes in Computer Science, vol. 7785, Springer Berlin Heidelberg, 2013, pp. 23–39.

Rey11. Leonid Reyzin, *Some notions of entropy for cryptography*, Information Theoretic Security (Serge Fehr, ed.), Lecture Notes in Computer Science, vol. 6673, Springer Berlin Heidelberg, 2011, pp. 138–142.

RTS00. Jaikumar Radhakrishnan and Amnon Ta-Shma, *Bounds for dispersers, extractors, and depth-two superconcentrators*, SIAM JOURNAL ON DISCRETE MATHEMATICS **13** (2000), 2000.

RTTV08. Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan, *Dense subsets of pseudorandom sets*, Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science (Washington, DC, USA), FOCS '08, IEEE Computer Society, 2008, pp. 76–85.

Sha48. C. E. Shannon, *A mathematical theory of communication*, Bell system technical journal **27** (1948).

TTV09. Luca Trevisan, Madhur Tulsiani, and Salil Vadhan, *Regularity, boosting, and efficiently simulating every high-entropy distribution*, Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity (Washington, DC, USA), CCC '09, IEEE Computer Society, 2009, pp. 126–136.

TZ08. Terence Tao and Tamar Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Mathematica **201** (2008), no. 2, 213–305 (English).

VZ12. Salil Vadhan and Colin Jia Zheng, *Characterizing pseudoentropy and simplifying pseudorandom generator constructions*, Proceedings of the 44th symposium on Theory of Computing (New York, NY, USA), STOC '12, ACM, 2012, pp. 817–836.

A Proof of Lemma 5

Proof (Proof of Lemma 5). Denote $p = \mathbf{P}(S), q = 1 - p$ and $a = \mathbf{E}[X|S], b = \mathbf{E}[X|S^c], v = \mathrm{Var}[X|S]$. Applying the Jensen's Inequality we obtain

$$\begin{aligned}\mathrm{Var}X &= \mathbf{E}\left(X - \mathbf{E}X\right)^2 \\ &= \mathbf{P}(S)\mathbf{E}\left[\left(X - \mathbf{E}X\right)^2 \middle| X \in S\right] + \mathbf{P}(S^c)\mathbf{E}\left[\left(X - \mathbf{E}X\right)^2 \middle| X \in S^c\right] \\ &\geqslant \mathbf{P}(S)\mathbf{E}\left[\left(X - \mathbf{E}X\right)^2 | X \in S\right] + \mathbf{P}(S^c)\left(\mathbf{E}[X|S^c] - \mathbf{E}X\right)^2\end{aligned}$$

Observe that

$$\begin{aligned}\mathbf{E}\left[\left(X - \mathbf{E}X\right)^2 | X \in S\right] &= \mathbf{E}\left[\left((X - \mathbf{E}[X|S]) + (\mathbf{E}[X|S] - \mathbf{E}X)\right)^2 | X \in S\right] \\ &= \mathbf{E}\left[\left(X - \mathbf{E}[X|S]\right)^2 \middle| X \in S\right] + \left(\mathbf{E}[X|S] - \mathbf{E}X\right)^2 \\ &= \mathrm{Var}[X|S] + \left(\mathbf{E}[X|S] - \mathbf{E}X\right)^2\end{aligned}$$

By the total probability law we obtain

$$\begin{aligned}\mathbf{E}[X|S] - \mathbf{E}X &= \mathbf{P}(S^c)(\mathbf{E}[X|S] - \mathbf{E}[X|S^c] \\ \mathbf{E}[X|S^c] - \mathbf{E}X &= \mathbf{P}(S)(\mathbf{E}[X|S^c] - \mathbf{E}[X|S].\end{aligned}$$

Putting this all together we see that it is enough to prove the following inequality

$$a + \sqrt{v} \cdot \sqrt{cp - 1} \leqslant pa + (1-p)b + \sqrt{c-1} \cdot \sqrt{\begin{aligned}&pv + p(1-p)^2(a-b)^2 + \\ &(1-p)^2 p(a-b)^2\end{aligned}}$$

which after introducing $u = a - b \in (-1, 1)$ becomes

$$(1-p)u + \sqrt{v} \cdot \sqrt{cp-1} \leqslant \sqrt{v + (1-p)u^2} \cdot \sqrt{cp-p}$$

Setting $A = v$, $B = cp - 1$, $C = u^2$ and $D = 1 - p$ we rewrite it as

$$D\sqrt{C} + \sqrt{AB} \leqslant \sqrt{A + CD} \cdot \sqrt{B + D}$$

(where we assume $A, B, C \in [0, 1]$ and $B \geqslant 0$). This inequality, by taking the squares of both side, is equivalent to $0 \leqslant (\sqrt{BC} - \sqrt{A})^2$, which finishes the proof. □