A Uniform Min-Max Theorem and Characterizations of Computational Randomness

A Uniform Min-Max Theorem and Characterizations of Computational Randomness

A dissertation presented

by

Jia Zheng

to

The School of Engineering and Applied Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Computer Science

Harvard University

Cambridge, Massachusetts

July 2013

Thesis Advisor

Salil P. Vadhan

Author

Jia Zheng

A Uniform Min-Max Theorem and

Characterizations of Computational Randomness

Abstract

This thesis develops several tools and techniques using ideas from information theory, optimization, and online learning, and applies them to a number of highly related fundamental problems in complexity theory, pseudorandomness theory, and cryptography.

First, we give a new, more constructive proof of von Neumann's Min-Max Theorem for two-player zero-sum game, extending previous work of Freund and Schapire (Games and Economic Behavior '99). The resulting Uniform Min-Max Theorem enables a number of applications in cryptography and complexity theory, often yielding uniform security versions of results that were previously only proved for nonuniform security (due to use of the non-constructive Min-Max Theorem), and often with optimal parameters.

We then develop several applications of the Uniform Min-Max Theorem, including: Regularity Theorems that provide efficient simulation of distributions within any sufficiently nice convex set; an improved version of the Weak Regularity Lemma for graphs; a simple and more modular uniform version of the Hardcore Theorem for boolean circuits; Dense Model Theorems for uniform algorithms; and impossibility of constructing Succinct Non-Interactive Arguments (SNARGs) via black-box reductions under uniform hardness assumptions.

Next, we provide a new characterization of computational Shannon-entropy, in terms of the hardness of sampling a distribution. Given any joint distribution (X, B) where B takes values in a polynomial-sized set, we show that (X, B) is computationally indistinguishable to some joint distribution (X, C) with $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$, if and only if there is no poly-sized circuit S such that the KL divergence from B to $S(X)$ is smaller than δ. We then use this characterization to show that if f is a one-way function, then $(f(U_n), U_n)$ has "next-bit pseudoentropy" at least $n + \log n$, establishing a conjecture of Haitner, Reingold, and Vadhan (STOC '10). Plugging this into the construction of Haitner et al., this yields a simpler construction of pseudorandom generators from one-way functions. With an additional idea, we

also show how to improve the seed length of the pseudorandom generator to $\tilde{O}(n^3)$, compared to $\tilde{O}(n^4)$ in the construction of Haitner et al. In addition, this characterization establishes a connection to prediction markets based on market scoring rules.

We also provide a new characterization of pseudo-avg-min-entropy, generalizing the Hardcore Theorem to polynomial-sized (rather than binary) alphabets. The Uniform Min-Max Theorem is used to obtain uniform versions of both characterizations.

Contents

Citations to Previously Published Work

Chapter 2, 3, and 6 are based on the paper "A Uniform Min-Max Theorem with Applications in Cryptography" [VZ2] coauthored with Salil Vadhan, published in Advances in Cryptology – CRYPTO '13, with full version posted as ECCC Technical Report TR13-101 [VZ3].

Part of Chapter 4, and Chapter 5, are based on the paper "Characterizing Pseudoentropy and Simplifying Pseudorandom Generator Constructions" [VZ1] coauthored with Salil Vadhan, published in Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC '12), with full version posted as ECCC Technical Report TR11-141 [VZ4].

Acknowledgments

I am forever grateful to my PhD advisor Salil Vadhan, for his inspirations and influence, academically and personally, during my time at Harvard. He taught me, by repeated demonstrations, to relentlessly perfect every proof and presentation, to always seek the most fundamental understanding, and most importantly, to believe that simple and beautiful answers must exist. His perseverance, humility, passion (for life and research) and patience (for me) are everlasting inspirations on their own. I also learned tremendously from Salil about effective presentation and professionalism. Lastly, this thesis would not exist without the guidance and collaboration from Salil, the coauthor of the constituent papers.

I thank Michael Mitzenmacher and Leslie Valiant for serving on my committee. I am grateful to Les for being my advisor during my first year at Harvard.

Kai-Min Chung and Zhenming Liu taught me a lot about math, research, and life in general, have being good friends since I came to Harvard, and made all the long winter nights in MD 138 memorable. I am indebted to Kai-Min for several crucial discussions on topics in this thesis, and discussions on many other problems.

During the course of this research, I was fortunate to have discussed closely related topics with Luca Trevisan, Or Meir, Iftach Haitner, Thomas Holenstein, Bruce Kapron, and undoubtedly others that I have forgotten. I thank them for teaching me new tricks and showing me different perspectives, which helped shape my own. Many people have offered valuable comments and suggestions for this work: Kai-Min Chung, Jacob Fox, Oded Goldreich, Iftach Haitner, Thomas Holenstein, Moritz Hardt, Yuval Ishai, Bruce Kapron, Krzysztof Pietrzak, Omer Reingold, Leo Reyzin, Maciej Skorski, Jonathan Ullman, and Udi Wieder.

Special thanks to Madhu Sudan for his terrific Advanced Complexity Theory course at MIT, where my course project on the Hardcore Theorem turned out to be good preparation for my research. A significant portion of this work was done during my one year exchange

at Stanford (a much needed escape from New England weather); I thank Luca Trevisan for hosting, and the Stanford theory group for their great hospitality.

Before Harvard, I am deeply indebted to Hon Wai Leong, for introducing me to algorithms research and for his great advice on research and life; Sanjay Jain and Frank Stephen, for sparkling my interest in theoretical computer science; Hon Wai Leong, Tandy Warnow and Wing-Kin Sung, for their encouragement and guidance during my first few research experiences (in computational biology).

Friends are responsible for all my fond memories at Harvard and Stanford: Bingkan Xue, Andy Zheng, Bowen Zhou, Frank Zhang, Thomas Steinke, Varun Kanade, Jon Ullman, Justin Thaler, Anna Huang, Mark Bun, Scott Linderman, Yun-Ru Chen, Feng-Hao Liu, Huijia Lin, Ruichao Ma, Ella Chou, and the list goes on. I wish to also thank friends and teachers from National University of Singapore and Hangzhou Foreign Language School for the fun and fulfilling education experience.

I owe my deepest debt to my family.

献给父亲和母亲

Chapter 1

Introduction

Information theory and optimization theory have proved to be extremely useful in many fields of theoretical computer science. For example, much of modern cryptography and pseudorandomness theory was founded on (generalizations of) information-theoretic notions; machine learning has benefitted tremendously from tools in optimization. In this work, we bring together concepts and techniques from information theory, optimization, and online learning in the context of complexity theory, pseudorandomness, and cryptography.

1.1 A Uniform Min-Max Theorem and Applications

Von Neumann's Min-Max Theorem (equivalent to Linear Programming Duality and the finite-dimensional Hahn-Banach Theorem) is a fundamental result about zero-sum games between 2 players. It tells us that if for every Player 1 strategy (which can be randomized) Player 2 can respond accordingly to achieve a payoff of at least k, then Player 2 has a *universal* strategy (which is randomized) that guarantees such payoff *regardless* of Player 1's strategy. The Min-Max Theorem has proved to be an extremely useful tool in theoretical computer science. In cryptography and complexity theory, it gives rise to a number of

results such as Impagliazzo's Hardcore Theorem [Imp], equivalence of different notions of computational entropy [BSW], the Dense Model Theorem [RTTV], leakage-resilient cryptography [DP2, FR], efficient simulation of high entropy distributions [TTV], impossibility of constructing succinct non-interactive arguments (SNARGs) via black-box reductions [GW], cryptographic studies of forecast testing [FV, CLP1], and simpler construction of pseudorandom generators from one-way functions [VZ1].

A limitation of the Min-Max Theorem is that it is non-constructive; it only asserts the existence of such universal strategy for Player 2, but does not say how it can be found (algorithmically). In a typical result in cryptography or complexity, where the statement is of the form "hardness of A" implies "hardness of B," the proof is often a "reduction," constructing an adversary for A from an adversary for B. This is the case for most of the aforementioned cryptographic applications, where the Min-Max Theorem is applied during the construction of the adversary for A. Consequently, non-constructivity of the Min-Max Theorem means we can only *nonuniformly* construct an adversary for A. In other words, such results must make the stronger assumption that A is hard even for nonuniform boolean circuits, and not just uniform algorithms.

In Chapter 2 we give a new, more constructive proof of the Min-Max Theorem, using techniques from optimization and online learning [HW], and extending previous work of Freund and Schapire [FS]. The resulting Uniform Min-Max Theorem, when used in place of the Min-Max Theorem, can often yield a *uniform* construction of an adversary for B. With the Uniform Min-Max Theorem, we are thus able to prove uniform versions of some of the aforementioned results (where hardness is with respect to uniform algorithms), throughout Chapter 3, 4 and 6.

Remark. The statement of a uniform result is often incomparable to the statement of the corresponding nonuniform result, whose assumption and conclusion are both stronger

(i.e. hardness for boolean circuits). However, a uniform result often can also be stated as a uniform "reduction" that works for both uniform *and* nonuniform settings. In this sense, our uniform results in Chapter 3, 4 and 6 are stronger than the known nonuniform results, but for simplicity we choose not to state them in that way.

Regularity Theorems for Distributions Restricted to a Convex Set. In Chapter 2 we also apply the Uniform Min-Max Theorem to show a generalization and quantitative improvement to the "Regularity Theorem" of Trevisan, Tulsiani, and Vadhan [TTV], which (informally) says that any high min-entropy distribution X is indistinguishable from some high min-entropy, *low complexity* distribution Y. The result of [TTV] is itself a quantitative improvement of regularity and "decomposition" theorems in additive combinatorics [GT, TZ]. It is shown in [TTV] that such results can be used to deduce the Dense Model Theorem [TZ, RTTV, Gow], Impagliazzo's Hardcore Theorem [Imp], and other results, by replacing any unknown distribution X with an "equivalent" distribution Y that can be efficiently analyzed and manipulated. Among applications of our Regularity Theorems are an improved and optimal Weak Regularity Lemma for graphs of density $o(1)$, and a strengthening of a recent result of Jetchev and Pietrzak [JP].

Uniform Hardcore Theorem. Impagliazzo's Hardcore Theorem ([Imp] and later strengthened in [KS, Hol1, BHK]) is a fundamental result in complexity theory that says if a boolean function f is somewhat hard on average, then there must be a subset of inputs (the hardcore) on which f is extremely hard, and outside of which f is easy. There are two approaches to proving the theorem. One is constructive [Imp, KS, Hol1, BHK] and leads to a *Uniform Hardcore Theorem* where hardness of f is measured against uniform algorithms, rather than nonuniform boolean circuits, and has several important applications in cryptography [KS, Hol1, Hol2, HHR1, HRV]. However, the existing proofs either do not achieve all of the

optimal parameters simultaneously for a Uniform Hardcore Theorem, or when they do they tend to be somewhat ad hoc. Another approach due to Nisan [Imp] (and strengthened in [Hol1]) uses the (non-constructive) Min-Max Theorem and has the advantage of simplicity, but is restricted to the nonuniform measure of hardness.

In Chapter 3, Section 3.1, we show that by replacing the use of Min-Max Theorem in the proof of Nisan [Imp] or Holenstein [Hol1] with our Uniform Min-Max Theorem, we obtain a new proof of the Uniform Hardcore Theorem with the advantages of (i) optimal hardcore density; (ii) optimal complexity blow-up; and (iii) modularity and simplicity.

Uniform Dense Model Theorem. A celebrated result of Green and Tao [GT] shows that there exist arbitrarily long arithmetic progressions of prime numbers. A key new component of their proof is the Dense Model Theorem which, in the generalized form of Tao and Ziegler [TZ], says if X is a pseudorandom distribution and D is a distribution dense in X, then D is indistinguishable to a distribution M that is dense in the uniform distribution. Using the Min-Max Theorem, Reingold et al. [RTTV] provided another proof of Dense Model Theorem where the indistinguishability and complexity blow-ups are polynomial (rather than exponential); a similar proof was given by Gowers [Gow]. The polynomial blow-ups are crucial for applications in leakage-resilient cryptography [DP2, DP1, FOR], and for connections to computational differential privacy [MPRV]. In Chapter 3, Section 3.2, as another application of the Uniform Min-Max Theorem, we show how to obtain a Dense Model Theorem where the distinguishers are efficient (uniform) algorithms, with polynomial blow-ups in running time and indistinguishability.

Characterizations of Computational Entropy. In Chapter 4 we give new characterizations of notions of computational randomness, whose proof in the nonuniform setting involves the Min-Max Theorem. Our characterization of "computational average min-

entropy" (known as "pseudo-avg-min-entropy") is a generalization of the Hardcore Theorem to larger alphabets. Our characterization of computational conditional entropy (known as "conditional pseudoentropy") leads to simpler constructions of pseudorandom generators from arbitrary one-way functions in Chapter 5 (building on the work of Haitner, Reingold, and Vadhan [HRV]). Using the Uniform Min-Max Theorem, we proved our characterizations in the uniform setting. For computational conditional entropy, the uniform result gives rise to (simpler) pseudorandom generator from arbitrary one-way functions that are secure against uniform algorithms.

Impossibility of Black-Box Construction of Succinct Non-Interactive Argument. A result of Gentry and Wichs [GW] shows that there is no black-box construction of succinct non-interactive arguments (SNARGs) from any natural cryptographic assumption. Their result relies on the (mild) assumption that there exist *hard subset membership problems*, which is equivalent to the existence of subexponentially hard one-way functions. One limitation is that they need to assume nonuniformly secure one-way functions, in part due to their use of the non-constructive Min-Max theorem in Lemma 3.1 of [GW].

In Chapter 6, we show how to obtain the analogous result in the *uniform setting* by using the Uniform Min-Max Theorem. We show that, assuming that there exist subexponentially hard one-way functions that are secure against uniform algorithms, there is no construction of SNARGs whose security can be reduced in a black-box way to a cryptographic assumption against uniform algorithms (unless the assumption is already false).

Unlike some of the previous applications, Gentry and Wich's proof (for the nonuniform setting) relies on nonuniformity even outside the use of the Min-Max Theorem. A considerable amount of extra work is needed for us to remove these uses of nonuniformity.

1.2 Characterizing Computational Entropy and Applications

Computational analogues of information-theoretic notions have given rise to some of the most interesting phenomena in cryptography and pseudorandomness theory. For example, *indistinguishability* [GM2], which is the computational analogue of statistical distance, enabled bypassing Shannon's impossibility results on perfectly secure encryption [Sha], and provided the basis for the computational theory of pseudorandomness [BM, Yao2]. Informally, two distributions X and Y are said to be *indistinguishable* if for all efficient randomized algorithms P, the probabilities that $P(X) = 1$ and $P(Y) = 1$ differ negligibly (where the probability is over X, Y, and coins of P).

Computational analogues of *entropy* were introduced by Yao [Yao2] and Håstad, Impagliazzo, Levin, and Luby [HILL]. The Håstad et al. notions, known as *pseudoentropy* and *pseudo-min-entropy*, were key to their fundamental result establishing the equivalence of pseudorandom generators and one-way functions, and have also now become a basic concept in complexity theory and cryptography. A distribution X is said to have *pseudoentropy at least k* if there exists a distribution Y indistinguishable from X such that Y has entropy[1] at least k. Analogously, a distribution X is said to have *pseudo-min-entropy at least k* if there exists a distribution Y indistinguishable from X such that Y has min-entropy[2] at least k.

Conditional versions of the Håstad et al. notions are known as *conditional pseudoentropy* [HRV] and *pseudo-avg-min-entropy* [HLR], respectively. Pseudo-avg-min-entropy, in the special case involving only a binary alphabet, is equivalent to "hardcore distributions" introduced by Impagliazzo [Imp] (see the discussion on Impagliazzo's Hardcore Theorem above). Conditional pseudoentropy was introduced by Haitner, Reingold, and Vadhan [HRV] to give a simpler and more efficient construction of pseudorandom generators from one-way

[1] *The (Shannon) entropy of a distribution X is defined to be* $\mathrm{H}_{\mathrm{sh}}(X) = \mathbb{E}_{x \leftarrow X}\left[\log(1/\Pr[X = x])\right]$.

[2] *The min-entropy of a distribution X is defined to be* $\mathrm{H}_{\mathrm{sh}}(X) = \min_x \left[\log(1/\Pr[X = x])\right]$.

functions.

In Chapter 4, we give new characterizations of pseudo-avg-min-entropy, pseudoentropy, and conditional pseudoentropy in terms of certain (different) notions of "hardness" for distributions, using concepts and techniques from information theory, optimization, and pseudorandomness theory. Our characterizations of pseudo-avg-min-entropy and conditional pseudoentropy are as follows:

Theorem 1.1 (Characterizing pseudo-avg-min-entropy, informal)**.** *Let (X, B) be a polynomial-time samplable joint distribution where B takes values in a polynomial-sized set. Then B has pseudo-avg-min-entropy at least k given X if and only if there is no probabilistic polynomial-time algorithm S such that* $\Pr[S(X) = B] \geq 2^{-k}$.

Theorem 1.2 (Characterizing conditional pseudoentropy, informal)**.** *Let (X, B) be a joint distribution where B takes values in a polynomial-sized set. Then B has pseudoentropy at least* $\mathrm{H}_{\mathrm{sh}}(B|X)+\delta$[3] *given X if and only if there is no probabilistic polynomial-time algorithm S such that the KL divergence*[4] *from (X, B) to $(X, S(X))$ is at most δ.*

We note that the characterization of pseudo-avg-min-entropy is a generalization of the Hardcore Theorem to polynomial-sized (rather than binary) alphabets. Some of our techniques are rather generic, and can be used to deduce a meta theorem characterizing a class of computational entropy notions (in which pseudoentropy is a special case).

In addition, using the Uniform Min-Max Theorem, we obtain *uniform* versions of the these characterizations, namely with respect to probabilistic polynomial-time algorithms S.

In Chapter 5, we study how to further simplify and improve the construction of pseudorandom generators from arbitrary one-way functions, building on the recent work of

[3] $\mathrm{H}_{\mathsf{sh}}(B|X)$ denotes the conditional entropy of B given X.

[4] KL divergence is a common notion of "distance" between distributions.

Haitner, Reingold, and Vadhan [HRV]. In particular, we apply the characterization of conditional pseudoentropy to show that all one-way functions directly contain "next-bit pseudoentropy," establishing a conjecture of Haitner, Reingold, and Vadhan. This yields a further simplified construction of pseudorandom generators from arbitrary one-way functions. In addition, we will explore how to further improve the efficiency of the pseudorandom generator from the previous state-of-the-art $\tilde{O}(n^4)$ to $\tilde{O}(n^3)$.

In Chapter 7, we introduce the notion of algorithmic prediction markets based on market scoring rules, and prove lower bounds using our characterizations of conditional pseudoentropy.

1.3 Preliminaries

For more background on information theory, including the definitions and proofs of their basic properties stated in this section, see [CT].

1.3.1 Notations and Conventions

For a natural number n, $[n]$ denotes the set $\{1, \ldots, n\}$, U_n denotes the uniform distribution on binary strings of length n. For a finite set Σ, U_Σ denotes the uniform distribution on Σ. For a distribution X, $\mathrm{supp}(X)$ denotes the support of X, $x \leftarrow X$ and $x \sim X$ mean that x is a random sample drawn from distribution X. We write $\mathrm{Avg}_{a \le i \le b}$ as a shorthand for the average over all $i \in \{a, \ldots, b\}$. $\mathrm{Conv}(\cdot)$ denotes the convex hull. Where there is no ambiguity, a symbol may be used to denote both a function and a variable, e.g. $g = g(n)$; in such case we write $g(\cdot)$ to denote the function. All logs are base 2.

When we say *n is a security parameter and $K = K(n)$* (e.g. when talking about uniform algorithms), what we mean is that there is a sequence of objects (which can be numbers, distributions, etc) $K(1), K(2), \ldots$, and K is used as a shorthand for $K(n)$ wherever n has

been quantified (explicitly or implicitly).

For a joint distribution (X, C), we write $C(a|x)$ to denote the conditional probability $\Pr[C = a|X = x]$, whenever X is clear from the context. For a function $P : \Sigma \to \mathbb{R}_{\geq 0}$ where Σ is a finite set, we write Φ_P to denote the distribution C where $\Pr[C = a] = P(a)/\sum_{a\in\Sigma} P(a)$. For a distribution X that is clear from context and a function $P : \mathrm{supp}(X) \times \Sigma \to \mathbb{R}_{\geq 0}$, we write Φ_P to denote the distribution C jointly distributed with X, where $C(a|x) = P(x, a)/\sum_{a\in\Sigma} P(x, a)$ for each $x \in \mathrm{supp}(X)$. In particular, if P is $[0, 1]$-valued then P is called a *measure*:

Definition 1.3 (Measure and conditional measure)**.** For a distribution C, a function $P : \mathrm{supp}(C) \to [0, 1]$ is said to be a *measure for* C if $C = \Phi_P$.

For a joint distribution (X, C), a function $P : \mathrm{supp}(X) \times \mathrm{supp}(C) \to [0, 1]$ is a *conditional measure for* $C|X$ if $(X, C) = (X, \Phi_P)$.

1.3.2 Entropies, Divergences, and Projection

Definition 1.4 (Entropy)**.** For a distribution X, the *(Shannon) entropy* of X is

$$\mathrm{H}_{\mathrm{sh}}(X) = \mathop{\mathbb{E}}_{x\leftarrow X}\left[\log \frac{1}{\Pr[X = x]}\right].$$

For $\alpha = 2, 3, \ldots$, the *Renyi entropy of* X *of order* α is

$$\mathrm{H}_\alpha(X) = \log \frac{1}{\left(\sum_{x\in\mathrm{supp}(X)} \Pr[X = x]^\alpha\right)^{1/(\alpha-1)}},$$

and the *min-entropy* of X is

$$\mathrm{H}_\infty(X) = \lim_{\alpha\to+\infty} \mathrm{H}_\alpha(X) = \min_{x\in\mathrm{supp}(X)} \left(\log \frac{1}{\Pr[X = x]}\right).$$

Definition 1.5. For a joint distribution (X, B), the *conditional (Shannon) entropy of* B *given* X (or, *conditional (Shannon) entropy of* B when X is clear from the context) is

$$\mathrm{H}_{\mathrm{sh}}(B|X) = \mathop{\mathbb{E}}_{x\leftarrow X}\left[\mathrm{H}_{\mathrm{sh}}(B|_{X=x})\right].$$

Proposition 1.6 (Chain rule for Shannon entropy)**.** $\mathrm{H}_{\mathrm{sh}}(X, B) = \mathrm{H}_{\mathrm{sh}}(X) + \mathrm{H}_{\mathrm{sh}}(B|X)$.

Bregman divergence is a notion of distance between distributions:[5]

Definition 1.7 (Bregman divergence)**.** Let Σ be any finite set, and $H : \{\text{distributions on } \Sigma\} \to \mathbb{R}_{\geq 0}$ be any strictly concave function that is differentiable in the interior of the simplex in $|\Sigma|$-space. Let A and B be distributions on Σ. The *Bregman divergence associated with H from A to B* is defined to be

$$\mathrm{D}_H(A \parallel B) = H(B) - H(A) - \langle \nabla H(B), B - A \rangle$$

where $\nabla H(B)$ is the gradient vector, and we view $B - A$ as the difference between probability vectors.

For $\langle \nabla H(B), B - A \rangle$, if $H(B)$ is not differentiable w.r.t. $\Pr[B = a]$ for some $a \in \Sigma$, then by convention:

- If $\Pr[A = a] > \Pr[B = a] = 0$, then $\mathrm{D}_H(A \parallel B) = +\infty$;
- If $\Pr[A = a] = \Pr[B = a] = 0$, then it contributes zero to the inner product.

While Bregman divergence is *not* a metric (it is not symmetric and does not satisfy the triangle inequality), it does satisfy nonnegativity, and equals zero if and only if the distributions are identical:

Proposition 1.8 (Nonnegativity of Bregman divergence)**.** *For all distributions A and B,* $\mathrm{D}_H(A \parallel B) \geq 0$. *Moreover,* $\mathrm{D}_H(A \parallel B) = 0$ *if and only if* $A = B$.

A canonical example of Bregman divergence is the *KL divergence*, with H being the Shannon entropy function H_{sh}.

[5] In fact Bregman divergence can be defined between elements in an *arbitrary* convex subset of $\mathbb{R}^n$, rather than the unit simplex. However, the more restricted definition suffices for our purpose.

Definition 1.9 (KL divergence and conditional KL divergence)**.** The Bregman divergence associated with Shannon entropy is known as the *Kullback-Leibler (KL) divergence.* For distributions A and B, it is easily verified that the KL divergence from A to B equals

$$\mathrm{KL}(A \parallel B) = \mathrm{D}_H(A \parallel B) = \mathop{\mathbb{E}}_{a \leftarrow A}\left[\log \frac{\Pr[A=a]}{\Pr[B=a]}\right],$$

or conventionally $+\infty$ if $\mathrm{supp}(A) \not\subseteq \mathrm{supp}(B)$.

For joint distributions (X, A) and (Y, B), the *conditional KL divergence from $A|X$ to $B|Y$* is defined to be

$$\mathrm{KL}((A|X) \parallel (B|Y)) = \mathop{\mathbb{E}}_{(x,a)\leftarrow(X,A)}\left[\log \frac{\Pr[A=a|X=x]}{\Pr[B=a|Y=x]}\right] = \mathop{\mathbb{E}}_{x\leftarrow X}\left[\mathrm{KL}(A|_{X=x} \parallel B_{Y=x})\right].$$

Intuitively, the KL divergence from distribution A to distribution B measures how dense A is within B, on average (with zero divergence representing maximum density, i.e. $A = B$, and large divergence meaning that A is concentrated in a small portion of B). Like Shannon entropy, KL divergence has a chain rule:

Proposition 1.10 (Chain rule for KL divergence)**.** $\mathrm{KL}(X, A \parallel Y, B) = \mathrm{KL}(X \parallel Y) + \mathrm{KL}((A|X) \parallel (B|Y))$.

Like other distance measures between distributions, applying any (deterministic) function never increases the KL divergence:

Proposition 1.11 (Entropy-like property of KL divergence)**.** $\mathrm{KL}(g(A)||g(B)) \leq \mathrm{KL}(A||B)$ *for any function g.*[6]

Next we define *KL projection*, which can be seen as the analogue of Euclidean projection that minimizes KL divergence rather than Euclidean distance:

[6]This is in fact equivalent to the *log-sum inequality* [CT]. For a more direct proof, see [GV].

Definition 1.12 (KL projection)**.** Let X be a distribution on Σ, and $\mathcal{V}$ be a non-empty closed convex set of distributions on Σ. $Y^* \in \mathcal{V}$ is called a *KL projection of X on $\mathcal{V}$* if

$$Y^* = \arg\min_{Y \in \mathcal{V}} \mathrm{KL}(Y \parallel X).$$

A nice property of KL projection is the following geometric structure (see [CT], Chap 11, Section 6):

Theorem 1.13 (Pythagorean Theorem)**.** *Let $\mathcal{V}$ be a non-empty closed convex set of distributions on Σ. Let Y^* be a KL projection of X on $\mathcal{V}$. Then for all $Y \in \mathcal{V}$,*

$$\mathrm{KL}(Y \parallel Y^*) + \mathrm{KL}(Y^* \parallel X) \leq \mathrm{KL}(Y \parallel X).$$

In particular,

$$\mathrm{KL}(Y \parallel Y^*) \leq \mathrm{KL}(Y \parallel X).$$

Assuming $\mathrm{KL}(Y^* \parallel X)$ is finite, then the Pythagorean Theorem implies that the KL projection is unique:

Proposition 1.14. *The KL projection is unique.*

Proof. Suppose Y^* and Y are both KL projections of X on $\mathcal{V}$. Then by the Pythagorean Theorem (Theorem 1.13) $\mathrm{KL}(Y \parallel Y^*) = 0$, which implies $Y = Y^*$ by Proposition 1.8. □

Finding the exact KL projection is often computationally infeasible, so we consider *approximate KL projection*:

Definition 1.15 (Approximate KL projection)**.** We say Y^* is a *σ-approximate KL projection* of X on $\mathcal{V}$, if $Y^* \in \mathcal{V}$ and for all $Y \in \mathcal{V}$,

$$\mathrm{KL}(Y \parallel Y^*) \leq \mathrm{KL}(Y \parallel X) + \sigma.$$

In particular, by the Pythagorean Theorem (Theorem 1.13) Y' is a σ-approximate KL projection of X if $\mathrm{KL}(Y \parallel Y') \leq \mathrm{KL}(Y \parallel Y^*) + \sigma$ for all $Y \in \mathcal{V}$, where Y^* is the (exact) KL projection of X.

1.3.3 Indistinguishability

(Computational) indistinguishability is the computational analogue of two distributions being statistically close (i.e. small total variation distance), by considering a restricted class of statistical tests $\mathcal{W}$ rather than *all* statistical tests:

Definition 1.16 ((Computational) indistinguishability)**.** Let $\mathcal{W}$ be any set of functions $W : \Sigma \to [0, 1]$, for some finite set Σ. Two distributions X and Y on Σ are *ϵ-indistinguishable by $\mathcal{W}$* if for all $W \in \mathcal{W}$,

$$|\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]| < \epsilon.$$

$\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]$ is said to be the *distinguishing advantage* of W, and W is an *ϵ-distinguisher between X and Y* if its distinguishing advantage is at least ϵ.

When $\mathcal{W}$ is closed under "negation" i.e. $W \in \mathcal{W} \iff 1 - W \in \mathcal{W}$, ϵ-indistinguishability can be equivalently stated without taking absolute value:

$$\mathbb{E}[W(X)] - \mathbb{E}[W(Y)] < \epsilon.$$

Chapter 2

Uniform Min-Max Theorem and Regularity Theorems

Von Neumann's Min-Max Theorem (which is equivalent to Linear Programming Duality and the finite-dimensional Hahn-Banach Theorem) has proved to be an extremely useful tool in theoretical computer science, giving rise to a number of results in cryptography and complexity theory. In this chapter we give a new, more constructive proof of the Min-Max Theorem (extending previous work of Freund and Schapire [FS]), and use the resulting Uniform Min-Max Theorem to deduce new Regularity Theorems for distributions. These results will be applied throughout Chapter 3, 4 and 6 to obtain new "uniform" results in cryptography and complexity theory.

2.1 Introduction

Consider a zero-sum game between two players where for every mixed strategy V for Player 1 (as a distribution over his strategy space $\mathcal{V}$), Player 2 has a response $W \in \mathcal{W}$ that guarantees $\mathbb{E}[f(V,W)] \geq 0$, where f (payoff) can be an arbitrary function. The Min-Max

Theorem says that there must exist a Player 2's mixed strategy W^* (as a distribution over his strategy space $\mathcal{W}$) that guarantees $\mathbb{E}\left[f(V, W^*)\right] \geq 0$ for *all* strategies $V \in \mathcal{V}$ of Player 1.

The Min-Max Theorem gives rise to a number of results in cryptography and complexity theory such as Impagliazzo's Hardcore Theorem [Imp], equivalence of different notions of computational entropy [BSW], the Dense Model Theorem [RTTV], leakage-resilient cryptography [DP2, FR], efficient simulation of high entropy distributions [TTV], impossibility of constructing succinct non-interactive arguments (SNARGs) via black-box reductions [GW], cryptographic studies of forecast testing [FV, CLP1], and simple construction of pseudorandom generators from one-way functions [VZ1]. In a typical application like these, Player 1 chooses V from a convex set $\mathcal{V}$ of distributions over $\{0,1\}^n$, and Player 2 chooses W from a set $\mathcal{W}$ of (possibly randomized) boolean functions $\{0,1\}^n \to \{0,1\}$ and receives expected payoff $\mathbb{E}[f(V, W)]$ where $f(V, W) = \mathbb{E}[W(V)]$, i.e. the expected output of W when input is drawn from the distribution V. For example, $\mathcal{V}$ contains all high entropy distributions over $\{0,1\}^n$ and $\mathcal{W}$ contains all boolean functions of small circuit size.

A limitation of the Min-Max Theorem is that it is highly non-constructive; it only asserts the existence of the optimal strategy W^* but does not say how it can be found (algorithmically). Consequently, applications of the Min-Max Theorem only give rise to results about nonuniform boolean circuits, rather than uniform algorithms (e.g. we set cryptographic protocols based on nonuniform hardness rather than uniform hardness assumptions).

To overcome this, we consider the natural algorithmic task of constructing such an optimal strategy W^* for Player 2, assuming f is efficiently computable. When the sizes of strategy spaces $\mathcal{V}$ and $\mathcal{W}$ are small (e.g. polynomial) this can be done by linear programming, for which efficient algorithms are well-known. However, applications in cryptography and complexity theory such as ones just mentioned involve exponentially large strategy spaces,

and an optimal strategy W^* cannot be found in polynomial time in general. Thus we also require that, given any mixed strategy V for Player 1, not only does there exist a strategy $W \in \mathcal{W}$ for Player 2 with $\mathbb{E}[f(V,W)] \geq 0$, but such response W can be obtained efficiently by an oracle (or an efficient uniform algorithm).

Assuming such an oracle, Freund and Schapire [FS] show how to find an approximately optimal W^* for Player 2 in polynomial time and by making $O((\log|\mathcal{V}|)/\epsilon^2)$ adaptive oracle queries, using the idea of multiplicative weight updates. However, their algorithm still falls short in some of aforementioned applications where $\mathcal{V}$ is a set of distributions over $\{0,1\}^n$, and thus $\mathcal{V}$ can have doubly-exponentially many vertices. For example, consider the set of distributions on $\{0,1\}^n$ of min-entropy at least k; the vertices of $\mathcal{V}$ are uniform distributions on a subset of size 2^k, and there are $\binom{2^n}{2^k}$ such subsets.

In this chapter, we present a Uniform Min-Max Theorem that efficiently finds an approximately optimal strategy W^* for Player 2, given an oracle that for any of Player 1's mixed strategy $V \in \mathcal{V}$ returns some Player 2's strategy that guarantees reasonable payoff, even when $\mathcal{V}$ is a (sufficiently nice) set of distributions over $\{0,1\}^n$. Our theorem is inspired by the proof of Uniform Hardcore Theorem of Barak, Hardt, and Kale [BHK]. Like [BHK], the underlying algorithm uses "relative entropy (KL) projections" together with multiplicative weight updates (a technique originally due to Herbster and Warmuth [HW]). Our contribution is providing the right abstraction: formulating this algorithm as providing a Uniform Min-Max Theorem.

An advantage of our formulation of a Uniform Min-Max Theorem is that it is more modular, and not specific to the Hardcore Theorem. Consequently, it immediately enables a number of applications, including (but not limited to) deriving uniform versions of many of the aforementioned results, where we now deal with algorithms rather than nonuniform boolean circuits. Even for the Hardcore Theorem, where the uniform version was already

known [Hol1, BHK], there are advantages to deducing it using the Uniform Min-Max Theorem. Furthermore, even in nonuniform settings, replacing the use of standard Min-Max Theorem with the Uniform Min-Max Theorem can often yield improved, optimal parameters.

Regularity Theorems for Distributions Restricted to a Convex Set. We then apply the Uniform Min-Max Theorem to show a generalization and quantitative improvement to the "Regularity Theorem" of Trevisan, Tulsiani, and Vadhan [TTV] which (informally) says that any high min-entropy distribution X is indistinguishable from some high min-entropy, *low complexity* distribution Y. The result of [TTV] is itself a quantitative improvement of regularity and "decomposition" theorems in additive combinatorics [GT, TZ]. It is shown in [TTV] that such results can be used to deduce the Dense Model Theorem [TZ, RTTV, Gow], Impagliazzo's Hardcore Theorem [Imp], and other results, by replacing any unknown distribution X with an "equivalent" distribution Y that can be efficiently analyzed and manipulated.

Our result is more general than [TTV] in the sense that we are no longer restricted to distributions of high min-entropy. We show that for any sufficiently nice convex set of distributions $\mathcal{V}$, every distribution $X \in \mathcal{V}$ is indistinguishable from some distribution $Y \in \mathcal{V}$ where Y has "low complexity", for various notions of complexity and indistinguishability. In the case of min-entropy distributions, we obtain a high min-entropy Y with lower complexity than [TTV]. This also yields an improved and optimal Weak Regularity Lemma for graphs of density $o(1)$ (Section 2.3.2).

Average-case versions of our Regularity Theorems can be used to deduce "low complexity" versions of a technical lemma of [GW]. We note that our average-case Regularity Theorem for circuit complexity is a strengthening of a recent result of Jetchev and Pietrzak

[JP], with a simpler proof. The low circuit complexity version of the [GW] lemma (with slightly weaker parameters) was initially proved by Jetchev and Pietrzak [JP], and an interactive extension was proved by Chung, Lui, and Pass [CLP2] for applications in the context of distributional zero-knowledge.

2.2 A Uniform Min-Max Theorem

Consider a zero-sum game between two players, where the space of pure strategies for Player 1 is a $\mathcal{V} = [N]$, the space of pure strategies for Player 2 is $\mathcal{W}$, and the payoff to Player 2 is defined to be $f(V, W)$ for some function $f : \mathcal{V} \times \mathcal{W} \to [0, 1]$. Von Neumann's Min-Max Theorem says that

$$\min_{V \in \mathrm{Conv}(\mathcal{V})} \max_{W \in \mathcal{W}} \mathbb{E}[f(V, W)] = \max_{W \in \mathrm{Conv}(\mathcal{W})} \min_{V \in \mathcal{V}} \mathbb{E}[f(V, W)].$$

Equivalently, if for every mixed strategy $V \in \mathrm{Conv}(\mathcal{V})$ for Player 1, Player 2 has a response $W \in \mathcal{W}$ that guarantees $\mathbb{E}\left[f(V, W)\right] \geq p$, then there must exist a Player 2's mixed strategy $W^* \in \mathrm{Conv}(\mathcal{W})$ that guarantees $\mathbb{E}\left[f(V, W^*)\right] \geq p$ for *all* strategies $V \in \mathcal{V}$ of Player 1:

Theorem 2.1 (Min-Max Theorem)**.** *Consider a two-player zero-sum game where $\mathcal{V}$ and $\mathcal{W}$ are the finite sets of pure strategies for Player 1 and Player 2 (resp.), and the payoff to Player 2 is defined to be $f(V, W)$ for some function $f : \mathcal{V} \times \mathcal{W} \to [0, 1]$.*

Suppose for all Player 1's mixed strategies $V \in \mathrm{Conv}(\mathcal{V})$ there exists a Player 2 "response" strategy $W \in \mathcal{W}$ with expected payoff $\mathbb{E}[f(V, W)] \geq p$. Then there exists some $W^* \in \mathrm{Conv}(\mathcal{W})$ such that $\mathbb{E}[f(V, W^*)] \geq p$ for all *$V \in \mathcal{V}$.*

A natural algorithmic task is to find such optimal mixed strategy W^*. This is easy (and well-known) in the nonuniform setting, where we want to compute W^* by a small circuit, assuming Player 2 responses can be computed by a small circuit:

Theorem 2.2 (Nonuniform Min-Max Theorem)**.** *Consider a two-player zero-sum game where the sets of pure strategies for Player 1 and Player 2 are $\mathcal{V} = [N]$ and $\mathcal{W}$, and the payoff to Player 2 is defined to be $f(V, W)$ for some function $f : \mathcal{V} \times \mathcal{W} \to [0, 1]$.*

Suppose for all Player 1's mixed strategies $V \in \mathrm{Conv}(\mathcal{V})$ there exists a Player 2 "response" strategy $W \in \mathcal{W}$ with expected payoff $\mathbb{E}[f(V, W)] \geq p$. Then for every $\epsilon > 0$, there exists some $W^* \in \mathrm{Conv}(\mathcal{W})$ such that $\mathbb{E}[f(V, W^*)] \geq p - \epsilon$ for all *$V \in \mathcal{V}$, and W^* is the uniform distribution over a multiset of $S = O(\log N/\epsilon^2)$ elements of $\mathcal{W}$.*

Proof. By the Min-Max Theorem there is a mixed strategy $W \in \mathrm{Conv}(\mathcal{W})$ with an expected payoff of $\mathbb{E}[f(V, W)] \geq p$ for all $V \in \mathcal{V}$. Take S random samples from W and let W^* be uniformly distributed over these S samples. By a Chernoff bound, for each $V \in \mathcal{V}$ w.p. at least $1 - 2^{-\Omega(S \cdot \epsilon^2)}$ we have $\mathbb{E}[f(V, W^*)] \geq \mathbb{E}[f(V, W)] - \epsilon \geq p - \epsilon$. The result follows by a union bound. □

Note that W^* has small circuit size because W^* can be computed by picking a random element of the small S element multiset; if $\mathcal{W}$ contain circuits of size at most t, then W^* has circuit size $O(S \cdot t)$. Also note that it implies the standard Min-Max Theorem by taking $\epsilon \to 0$.

In many applications (including the Hardcore Theorem), the game must be set up such that the set of pure strategies for Player 1 is a *convex set* $\mathcal{V}$ of distributions over $[N]$, with the expected payoff still defined to be $\mathbb{E}\left[f(V, W)\right]$ for some function $f : [N] \times \mathcal{W} \to [0, 1]$. For example, $\mathcal{V}$ contains all the high entropy distributions over $[N]$. The Min-Max Theorem still holds for such generalized settings. And the Nonuniform Min-Max Theorem holds as well (with the same proof as before):

Theorem 2.3 (Nonuniform Min-Max Theorem (generalized))**.** *Consider a two-player zero-sum game where the sets of pure strategies for Player 1 and Player 2 are $\mathcal{V} \subseteq \{$distributions over $[N]\}$*

and $\mathcal{W}$, and the expected payoff to Player 2 is defined to be $\mathbb{E}[f(V,W)]$ for some function $f : [N] \times \mathcal{W} \to [0,1]$.

Suppose for all Player 1's mixed strategies $V \in \mathrm{Conv}(\mathcal{V})$ there exists a Player 2 "response" strategy $W \in \mathcal{W}$ with expected payoff $\mathbb{E}[f(V,W)] \geq p$. Then for every $\epsilon > 0$, there exists some $W^* \in \mathrm{Conv}(\mathcal{W})$ such that $\mathbb{E}[f(V,W^*)] \geq p - \epsilon$, and W^* is the uniform distribution over a multiset of $S = O(\log N/\epsilon^2)$ elements of $\mathcal{W}$.

This version of the Nonuniform Min-Max Theorem is implicit in Nisan's proof of the Hardcore Theorem [Imp], and has been used often since then.

A more ambitious goal to find such optimal mixed strategy W^* by a *uniform* algorithm. We now present a Uniform Min-Max Theorem that efficiently finds an approximately optimal strategy $W^* \in \mathrm{Conv}(\mathcal{W})$ for Player 2, given an oracle which, when fed any of Player 1's mixed strategies $V \in \mathrm{Conv}(\mathcal{V})$, returns a strategy for Player 2 that guarantees good expected payoff. Our algorithm is inspired by the proof of Uniform Hardcore Theorem of Barak, Hardt, and Kale [BHK]. Like [BHK], our algorithm uses "relative entropy (KL) projections" together with multiplicative weight updates (a technique originally due to Herbster and Warmuth [HW]).

Theorem 2.4 (A Uniform Min-Max Theorem)**.** *Consider a two-player zero-sum game where the sets of pure strategies for Player 1 and Player 2 are $\mathcal{V} \subseteq \{$distributions over $[N]\}$ and $\mathcal{W}$, and the expected payoff to Player 2 is defined to be $\mathbb{E}[f(V,W)]$ for some function $f : [N] \times \mathcal{W} \to [0,1]$. Then for every $0 < \epsilon \leq 1$ and S, Algorithm 2.1 (Finding Universal Strategy) always outputs a mixed strategy W^* for Player 2 such that*

$$\mathbb{E}[f(V,W^*)] \geq \operatorname*{Avg}_{1 \leq i \leq S} \mathbb{E}[f(V^{(i)},W^{(i)})] - O(\epsilon)$$

for all Player 1 strategies $V \in \mathcal{V}$ where $\mathrm{KL}(V \parallel V_1) \leq S \cdot \epsilon^2$. (This holds regardless of the arbitrary choice of $W^{(i)}$ and $V^{(i+1)}$ in the algorithm.)

In particular, taking $S \geq (\log N - \min_{V \in \mathcal{V}} \mathrm{H}_{\mathrm{sh}}(V))/\epsilon^2$ *where we set* $V^{(1)} = U_{[N]} \in \mathrm{Conv}(\mathcal{V})$ *yields that for all* $V \in \mathcal{V}$*,*

$$\mathbb{E}[f(V, W^*)] \geq \operatorname*{Avg}_{1 \leq i \leq S} \mathbb{E}[f(V^{(i)}, W^{(i)})] - O(\epsilon).$$

Arbitrarily choose an initial strategy $V^{(1)} \in \mathrm{Conv}(\mathcal{V})$ for Player 1

for $i \leftarrow 1$ **to** S **do**

Obtain an arbitrary strategy $W^{(i)} \in \mathcal{W}$ for Player 2, in response to $V^{(i)}$

Weight Update:

Let $V^{(i)'}$ be such that $\Pr[V^{(i)'} = x] \propto e^{-\epsilon \cdot f(x, W^{(i)})/2k} \cdot \Pr[V^{(i)} = x]$

Projection:

$V^{(i+1)} \leftarrow$ an arbitrary ϵ^2-approximate KL projection of $V^{(i)'}$ on $\mathrm{Conv}(\mathcal{V})$

end

Let W^* be the mixed strategy for Player 2 uniform over $W^{(1)}, \ldots, W^{(S)}$

return W^*

Algorithm 2.1: Finding Universal Strategy

By taking each $W^{(i)}$ to be a response to $V^{(i)}$ s.t. $\mathbb{E}[f(V^{(i)}, W^{(i)})] \geq p$, Theorem 2.4 implies the Nonuniform Min-Max Theorem (Theorem 2.3) with an improved $S = (\log N - \min_{V \in \mathcal{V}} \mathrm{H}_{\mathrm{sh}}(V))/\epsilon^2$. Such setting of S is shown to be tight in N and ϵ when $\mathcal{V}$ is set of all distributions [FS], and when $\mathcal{V}$ is the set of all δ-dense distributions, for any δ [LTW, Zha]. Even in nonuniform settings, it is often better to use the Uniform Min-Max Theorem (where the multiset $W^{(1)}, \ldots, W^{(S)}$ is constructed adaptively) rather than Theorem 2.3 (where the multiset $W^{(1)}, \ldots, W^{(S)}$ is constructed probabilistically); see Section 3.1 and 3.2 for discussions in more concrete settings.

Note that the number of iterations is at most $\log N/\epsilon^2$, so we can hope for running time $\mathrm{poly}(\log N, 1/\epsilon)$. However, Algorithm 2.1 is only an "algorithm template." To implement it

efficiently in particular applications, we need to specify:

1. An *compact* representation of the mixed strategies $V^{(i)}$ (the full pmf consists of N numbers, whereas we want running time $\text{poly}(\log N, 1/\epsilon)$).

2. An efficient algorithm to obtain a good response $W^{(i)} \in \mathcal{W}$ in response to a (compactly described) mixed strategy $V^{(i)}$. Typically this comes from our assumption/hypothesis in a given application.

3. An efficient algorithm to perform weight update and projection onto $\text{Conv}(\mathcal{V})$ for the compact representation of mixed strategies.

We do not present an abstract formulation of these requirements, since it will be rather complex to capture all of the applications. For example, in the application in Section 2.3 we are able to obtain a good response $W^{(i)}$ for $V^{(i)}$ only when $V^{(i)}$ is constructed by an efficient uniform algorithm.

Proof of Theorem 2.4. Consider any $V \in \mathcal{V}$ such that $\text{KL}(V \parallel V_1) \leq S \cdot \epsilon^2$. We show in Lemma A.1 that

$$\text{KL}(V \parallel V^{(i)}) - \text{KL}(V \parallel V^{(i)'}) \geq (\log e)\epsilon \left(\mathbb{E}[f(V^{(i)}, W^{(i)})] - \mathbb{E}[f(V, W^{(i)})] - \epsilon \right).$$

Since $V^{(i+1)}$ is an ϵ^2-approximate KL projection of $V^{(i)'}$ on $\text{Conv}(\mathcal{V})$, by definition we have $\text{KL}(V \parallel V^{(i+1)}) \leq \text{KL}(V \parallel V^{(i)'}) + \epsilon^2$. Therefore

$$\text{KL}(V \parallel V^{(i)}) - \text{KL}(V \parallel V^{(i+1)}) \geq (\log e)\epsilon \left(\mathbb{E}[f(V^{(i)}, W^{(i)})] - \mathbb{E}[f(V, W^{(i)})] - \epsilon \right) - \epsilon^2.$$

Summing over $i = 1, \ldots, S$ and telescoping, we obtain

$$\begin{aligned}
&\mathrm{KL}(V \parallel V^{(1)}) - \mathrm{KL}(V \parallel V^{(S+1)}) \\
&\geq (\log e)\epsilon \sum_{i=1}^{S} \Big(\mathbb{E}[f(V^{(i)}, W^{(i)})] - \mathbb{E}[f(V, W^{(i)})] - \epsilon \Big) - S\epsilon^2 \\
&= (\log e) S\epsilon \left(\operatorname*{Avg}_{1 \leq i \leq S} \mathbb{E}[f(V^{(i)}, W^{(i)})] - \mathbb{E}[f(V, W^*)] - \epsilon \right) - S\epsilon^2.
\end{aligned}$$

Since $\mathrm{KL}(V \parallel V^{(S+1)}) \geq 0$ and $\mathrm{KL}(V \parallel V_1) \leq S \cdot \epsilon^2$, rearranging yields

$$\operatorname*{Avg}_{1 \leq i \leq S} \mathbb{E}[f(V^{(i)}, W^{(i)})] - \mathbb{E}[f(V, W^*)] \leq \frac{\mathrm{KL}(V \parallel V^{(1)}) + S\epsilon^2}{(\log e) S\epsilon} + \epsilon = O(\epsilon).$$

□

Next, we describe an average case variant, where the set $\mathcal{V}$ of strategies for Player 1 is a set of distributions of the form (X, C) where C may vary, but the marginal distribution of X is fixed. This is convenient for a number of applications (e.g. Chapter 4 and 6) that involve distinguishers on such joint distributions (X, C).

Theorem 2.5 (Uniform Min-Max Theorem – Average Case)**.** *Let $\mathcal{V}$ be a subset of distributions over $[N] \times [q]$ of the form (X, C) where C may vary, but the marginal distribution of X is fixed. That is, for every (X, C), $(X', C') \in \mathcal{V}$ and every $x \in [N]$ we have $\sum_c \Pr[(X, C) = (x, c)] = \sum_c \Pr[(X', C') = (x, c)]$.*

Consider a two-player zero-sum game where the sets of pure strategies for Player 1 and Player 2 are $\mathcal{V}$ and $\mathcal{W}$, and the expected payoff to Player 2 is defined to be $\mathbb{E}\left[f((X, C), W)\right]$ for some function $f : [N] \times [q] \times \mathcal{W} \to [0, 1]$. Then for every $0 < \epsilon \leq 1$ and S, Algorithm 2.2 (Finding Universal Strategy – Average Case) always outputs a mixed strategy W^* for Player 2 such that

$$\mathbb{E}[f((X, C), W^*)] \geq \operatorname*{Avg}_{1 \leq i \leq S} \mathbb{E}[f((X, C^{(i)}), W^{(i)})] - O(\epsilon)$$

for all Player 1 strategies $(X, C) \in \mathcal{V}$ *where* $\mathrm{KL}(X, C \parallel X, C^{(1)}) \leq S \cdot \epsilon^2$. *(This holds regardless of the arbitrary choice of* $W^{(i)}$ *and* $C^{(i+1)}$ *in the algorithm.)*

In particular, taking $S \geq \left(\log q - \min_{(X,C)\in\mathcal{V}} \mathrm{H_{sh}}(C|X)\right)/\epsilon^2$ *where we set* $(X, C^{(1)}) = (X, U_{[q]}) \in \mathrm{Conv}(\mathcal{V})$ *(*$U_{[q]}$ *being independent of* X*) yields that for all* $(X, C) \in \mathcal{V}$,

$$\mathbb{E}[f((X, C), W^*)] \geq \operatorname*{Avg}_{1\leq i\leq S} \mathbb{E}[f((X, C^{(i)}), W^{(i)})] - O(\epsilon).$$

Arbitrarily choose an initial strategy $(X, C^{(1)}) \in \mathrm{Conv}(\mathcal{V})$ for Player 1

for $i \leftarrow 1$ **to** S **do**

Obtain an arbitrary strategy $W^{(i)} \in \mathcal{W}$ for Player 2, in response to

$(X, C^{(i)})$

Weight Update:

Let $C^{(i)'}$ be such that $\forall x, a$,

$\Pr[C^{(i)'} = a | X = x] \propto e^{-\epsilon \cdot f(x,a,W^{(i)})/2k} \cdot \Pr[C^{(i)} = a | X = x]$

Projection:

$(X, C^{(i+1)}) \leftarrow$ an arbitrary ϵ^2-approximate KL projection of $(X, C^{(i)'})$

on $\mathrm{Conv}(\mathcal{V})$

end

Let W^* be the mixed strategy for Player 2 uniform over $W^{(1)}, \ldots, W^{(S)}$

return W^*

Algorithm 2.2: Finding Universal Strategy – Average Case

Proof. Note that Algorithm 2.2 is the same as Algorithm 2.1, except for the difference that here we update $C^{(i)}$ instead of $V^{(i)}$. We show that the combined effect of the update and KL projection steps is identical in the two algorithms. Note that we can write $V^{(i)'}$ as $(X^{(i)'}, g_i(X^{(i)'}))$ for the randomized function g_i where $\Pr[g_i(x) = a] \propto e^{\epsilon \cdot f(x,a,W^{(i)})/2k} \cdot \Pr[C^{(i)} = a | X = x]$ for every x and a. For the same function g_i, we have $(X, g_i(X)) = (X, C^{(i)'})$. Thus, we can apply the following lemma. □

Lemma 2.6. *Let X' be a distribution on $[N]$ with $\mathrm{supp}(X') \supseteq \mathrm{supp}(X')$, and let $g : [N] \to [q]$ be a randomized function. Then the KL projection of $(X', g(X'))$ on $\mathrm{Conv}(\mathcal{V})$ equals the KL projection of $(X, g(X))$ on $\mathrm{Conv}(\mathcal{V})$.*

Proof. Consider any $(X, C) \in \mathrm{Conv}(\mathcal{V})$. We have

$$\begin{aligned}
&\mathrm{KL}(X, C \parallel X', g(X')) \\
&= \mathrm{KL}(X \parallel X') + \mathrm{KL}((C|X) \parallel (g(X')|X')) && \text{(by the chain rule for KL divergence)} \\
&= \mathrm{KL}(X \parallel X') + \mathrm{KL}((C|X) \parallel (g(X)|X)) && \text{(by definition of conditional KL divergence)} \\
&= \mathrm{KL}(X \parallel X') + \mathrm{KL}(X, C \parallel X, g(X)). && \text{(by the chain rule for KL divergence)}
\end{aligned}$$

Thus the KL projections are the same. □

2.3 Regularity Theorems for Distributions Restricted to a Convex Set

Another application of the Uniform Min-Max Theorem is to give a generalization and quantitative improvement to the "Regularity Theorem" of Trevisan, Tulsiani, and Vadhan [TTV] which (informally) says that any high min-entropy distribution X is indistinguishable from some high min-entropy, *low complexity* distribution Y. The result of [TTV] is itself a quantitative improvement of regularity and "decomposition" theorems in additive combinatorics [GT, TZ]. It is shown in [TTV] that such results can be used to deduce the Dense Model Theorem [TZ, RTTV, Gow], Impagliazzo's Hardcore Theorem [Imp], and other results, by replacing any unknown distribution X with an "equivalent" distribution Y that can be efficiently analyzed and manipulated, thus translating the problem to a simpler one. It also implies the Weak Regularity Lemma in graph theory [FK], mostly by a translation of notation.

Our result is more general than [TTV] in the sense that we are no longer restricted to distributions of high min-entropy. We show that for any sufficiently nice convex set of distributions $\mathcal{V}$, every distribution $X \in \mathcal{V}$ is indistinguishable from some distribution $Y \in \mathcal{V}$ where Y has "low complexity". In the case of min-entropy distributions, we obtain a high min-entropy Y with lower complexity than [TTV]. This also yields an improved and optimal Weak Regularity Lemma for graphs of density $o(1)$ (Section 2.3.2).

This section is divided into three parts, each proving results for a different notions of "complexity": Section 2.3.1 for information-theoretic notion of complexity, Section 2.3.3 for circuit complexity, and Section 2.3.4 for time complexity of uniform algorithms.

In addition, using the Uniform Min-Max Theorem – Average Case (Theorem 2.5) we obtain average-case variants, which can be used to deduce "low complexity" versions of a technical lemma of [GW]. We note that a special case of the average-case variant for circuits is a strengthening of a recent result of Jetchev and Pietrzak [JP], with a simpler proof. The low circuit complexity version of the [GW] lemma (with slightly weaker parameters) was initially proved by Jetchev and Pietrzak [JP], and an interactive extension was proved by Chung, Lui, and Pass [CLP2] for applications in the context of distributional zero-knowledge.

2.3.1 Regularity Theorems for Feature Complexity

Let $\mathcal{W}$ be an arbitrary class of functions $W : \Sigma \to [0,1]$ for some finite set Σ. Two distributions X and Y on Σ are *ϵ-indistinguishable by $\mathcal{W}$* if for every $W \in \mathcal{W}$, $|\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]| < \epsilon$. For starters, we shall consider the setting where the complexity of a distribution Y is purely information-theoretic: We say Y *has feature complexity at most m w.r.t. $\mathcal{W}$* if its mass function $x \mapsto \Pr[Y = x]$ is a function of $W_1(x), \ldots, W_m(x)$, for some $W_1, \ldots, W_m \in \mathcal{W}$. Notice that we can assume $\mathcal{W}$ to be closed under negation,

i.e. if $W \in \mathcal{W}$ then we can add $1 - W$ to $\mathcal{W}$ without affecting the meaning of complexity.

In order to obtain a low feature complexity approximation within a convex set $\mathcal{V}$ of distributions on Σ, we require $\mathcal{V}$ to be *permutation-invariant.* That is, for all permutations $\pi : \Sigma \to \Sigma$ we have $X \in \mathcal{V} \iff \pi(X) \in \mathcal{V}$. Permutation invariance is a natural condition; for example, the set of high entropy distributions should be permutation-invariant for any reasonable notion of entropy. However, for a fixed distribution X_0, $\{(X, C) : \mathrm{H}_{\mathrm{sh}}(C|X) \geq k, X = X_0\}$ is not permutation-invariant in general. We will use the following properties of a permutation-invariant convex set:

Lemma 2.7. *Let $\mathcal{V}$ be a permutation-invariant nonempty convex set of distributions on Σ. Then*

1. *$\mathcal{V}$ contains the uniform distribution on Σ.*
2. *Let X be a distribution on Σ having feature complexity at most m w.r.t. $\mathcal{W}$. Then the KL projection of X on $\mathcal{V}$ also has feature complexity at most m w.r.t. $\mathcal{W}$.*

Proof. 1. For any $Y \in \mathcal{V}$, the average of $\pi(Y)$ over all permutations π is still in $\mathcal{V}$ (by convexity and permutation-invariance), and is clearly the uniform distribution.

2. Let Y^* denote the KL projection of X on $\mathcal{V}$. For all $x_1, x_2 \in \Sigma$ where $\Pr[X = x_1] = \Pr[X = x_2]$, we must also have $\Pr[Y^* = x_1] = \Pr[Y^* = x_2]$; otherwise, swapping $\Pr[Y^* = x_1]$ and $\Pr[Y^* = x_2]$ yields some $\widehat{Y^*} \in \mathcal{V}$ (by permutation-invariance) that is also a KL projection of X, violating the uniqueness of KL projection (Lemma 1.14). Therefore $\Pr[Y^* = x]$ is a function of $\Pr[X = x]$, and Y^* has feature complexity at most that of X.

□

We show that every distribution $X \in \mathcal{V}$ is indistinguishable to some $Y \in \mathcal{V}$ of low feature complexity, as long as $\mathcal{V}$ is permutation-invariant:

Theorem 2.8 (A Regularity Theorem for feature complexity)**.** *Let Σ be a finite set, $\mathcal{W}$ be an arbitrary class of functions $W : \Sigma \to [0,1]$, $\mathcal{V}$ be a permutation-invariant convex set of distributions on Σ, and $\epsilon > 0$. Then for every distribution $X \in \mathcal{V}$ there exists $Y \in \mathcal{V}$ such that*

1. *X and Y are $O(\epsilon)$-indistinguishable by $\mathcal{W}$;*
2. *Y has feature complexity at most $S = (\log|\Sigma| - \mathrm{H}_{\mathrm{sh}}(X))/\epsilon^2$ w.r.t. $\mathcal{W}$. That is, there exist $W_1, \ldots, W_S \in \mathcal{W}$ and a function $\theta : [0,1]^S \to [0,1]$ such that $\forall x$,*

$$\Pr[Y = x] = \theta(W_1(x), \ldots, W_S(x)).$$

Remark. The main theorem of [TTV] (when considering feature complexity) is equivalent to Theorem 2.8 with $\mathcal{V}$ being fixed to be the set of distributions of min-entropy at least $\log|\Sigma| - \log(1/\delta)$, and has a worse bound on the feature complexity of Y. For a distribution X with $\mathrm{H}_\infty(X) = \log|\Sigma| - \log(1/\delta)$, [TTV] obtains a distribution Y with feature complexity at most $1/\epsilon^2\delta^2$ such that Y is $O(\epsilon)$-indistinguishable to X and $\mathrm{H}_\infty(Y) \geq \mathrm{H}_\infty(X)$, whereas Theorem 2.8 obtains such Y with feature complexity at most $\log(1/\delta)/\epsilon^2$.

Theorem 2.8 is interesting even if we do not require the low complexity Y to lie in $\mathcal{V}$. As mentioned in [TTV] (and pointed out by Elad Verbin), it easily follows from a Chernoff bound and a union bound that the uniform distribution over certain $O(\log|\mathcal{W}|/\epsilon^2)$ elements of Σ (which may not lie in $\mathcal{V}$) is ϵ-indistinguishable from X by $\mathcal{W}$. However, for large $\mathcal{W}$ the feature complexity of $O(\log|\mathcal{W}|/\epsilon^2)$ is potentially much higher than $S = (\log|\Sigma| - \mathrm{H}_{\mathrm{sh}}(X))/\epsilon^2$. Indeed, we do not use the fact that $Y \in \mathcal{V}$ when deducing the Weak Regularity Lemma of Frieze and Kannan [FK] from Theorem 2.8 (see Theorem 2.10 below); as shown in [TTV], the argument of Frieze and Kannan can be used to obtain a weaker variant of Theorem 2.8 where Y may not lie in $\mathcal{V}$, and the bound on S is worse.

Proof of Theorem 2.8. Suppose for contradiction that for every low feature complexity $Y \in \mathcal{V}$ there is some $W \in \mathcal{W}$ such that $\mathbb{E}[W(X)] - \mathbb{E}[W(Y)] \geq \epsilon'$ (recall that w.l.o.g. $\mathcal{W}$ is closed under negation), where $\epsilon' = c \cdot \epsilon$ for a sufficiently large constant c. Consider the zero-sum game where Player 1 selects some distribution $Y \in \mathcal{V}$, Player 2 selects some $W \in \mathcal{W}$ and receives (expected) payoff $\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]$. Consider Algorithm 2.1 (Finding Universal Strategy) where we set the initial strategy $V^{(1)}$ for Player 1 to be the uniform distribution on Σ (which lies in $\mathcal{V}$, by Lemma 2.7) and number of iterations to be S. Note that in each iteration the feature complexity of $V^{(i)}$ increases by at most one, due to the weight update using $W^{(i)}$, since KL projection on the permutation-invariant set $\mathcal{V}$ does not increase feature complexity (Lemma 2.7). Hence by assumption, in each iteration there exists $W^{(i)} \in \mathcal{W}$ such that $\mathbb{E}\left[W^{(i)}(X)\right] - \mathbb{E}\left[W^{(i)}(V^{(i)})\right] \geq \epsilon'$. By the Uniform Min-Max Theorem (Theorem 2.4), W^* (the uniform distribution over $W^{(1)}, \ldots, W^{(S)}$) satisfies

$$\mathbb{E}[W^*(X)] - \mathbb{E}[W^*(V)] \geq \epsilon' - O(\epsilon) > 0$$

for all Player 1 strategies $V \in \mathcal{V}$ such that $\mathrm{H}_{\mathrm{sh}}(V) \geq \mathrm{H}_{\mathrm{sh}}(X)$. Taking $V = X$ yields a contradiction. □

2.3.2 Improved Weak Regularity Lemma for Graphs of Density $o(1)$

An information-theoretic application of [TTV] is deducing the Weak Regularity Lemma of Frieze and Kannan [FK]. Our Theorem 2.8, with the improved bound, implies a Weak Regularity Lemma with parameters stronger than [FK] for graphs that are $o(1)$-dense. The Weak Regularity Lemma says that any graph $G = (V, E)$ is approximated within "cut-distance" σ by some edge-weighted graph G' on the vertices $\{1, \ldots, t\}$, where t depends only on σ (i.e. independent of the size of G), and each vertex i corresponds to a block $V_i \subseteq V$ in a partition $\{V_1, \ldots, V_t\}$ of V. The edge weight of (i, j) in the approximator G' is defined to be the *edge density* between V_i and V_j:

Definition 2.9 (Edge density)**.** The *density* of a directed graph $G = (V, E)$ equals $|E| / |V|^2$. The *edge density* between two sets of vertices V_1, V_2 of G equals

$$d_G(V_1, V_2) = \frac{|(V_1 \times V_2) \cap E|}{|V_1 \times V_2|}.$$

Theorem 2.10 (A Weak Regularity Lemma)**.** *For every directed graph $G = (V, E)$ of density $\delta = |E| / |V|^2 > 0$ and $\sigma > 0$, there is a partition of V into $t = \exp(O(\delta/\sigma)^2 \log(1/\delta))$ disjoint sets $V_1, \ldots, V_t$, such that for all A, $B \subseteq V$,*

$$\left| |(A \times B) \cap E| - \sum_{i,j} |A \cap V_i| \, |B \cap V_j| \cdot d_G(V_i, V_j) \right| < \sigma \cdot |V|^2 .$$

Note that the only interesting setting of parameters is $\delta > \sigma$, $\delta > 1/|V|^{O(1)}$ (i.e. G has average degree greater than 1), because if $\delta \leq \sigma$ then the trivial partition $V_1 = V$ would work, and if $\sigma < \delta \leq 1/|V|^{O(1)}$ we could take $t = |V|$ and use the trivial partition into single vertices. As pointed out to us by Jacob Fox, the number of partitions $\exp(O(\delta/\sigma)^2 \log(1/\delta))$ in Theorem 2.10 (as a function of δ and σ) is optimal up to a constant factor, which can be shown by adapting a lower bound argument in [CF].

Theorem 2.10 is stronger than Frieze and Kannan [FK] when G has density $\delta = o(1)$. For example, when $|V| = N$ and $\delta = 2\sigma = 1/\text{poly}(\log N)$, Theorem 2.10 produces a partition of size $\text{poly}(\log N)$, whereas [FK] only yields a trivial partition into more than N sets.

Proof of Theorem 2.10. We apply Theorem 2.8 with $\Sigma = V \times V$, $\mathcal{W} = \{\chi_{S \times T}, 1 - \chi_{S \times T} : S, T \subseteq V\}$ (where $\chi_{S \times T}$ denotes the characteristic function of $S \times T$), $\mathcal{V}$ being the set of all δ-dense distributions on Σ, $X = U_E \in \mathcal{V}$ (the uniform distribution on E), and $\epsilon = O(\sigma/\delta)$. By Theorem 2.8 there is some δ-dense distribution Y where:

1. Y has feature complexity at most $m = O((2 \log |V| - \mathrm{H_{sh}}(U_E))/\epsilon^2) = O((\delta/\sigma)^2 \log(1/\delta))$. That is, $\Pr[Y = e] = \phi(\chi_{S_1 \times T_1}(e), \ldots, \chi_{S_m \times T_m}(e))$ for a function ϕ and sets S_1,T_1, …, S_m, $T_m \subseteq V$.

2. U_E and Y are ϵ-indistinguishable for $\mathcal{W}$. That is, for every S, $T \subseteq V$,

$$|\mathbb{E}[\chi_{S\times T}(U_E)] - \mathbb{E}[\chi_{S\times T}(Y)]| < \epsilon.$$

The fact that Y has feature complexity at most m yields a partition $\{V_1, \ldots, V_t\}$, $t \leq 2^{2m}$, such that $\Pr[Y = e]$ has the same value for all $e \in V_i \times V_j$. (Specifically, the partition is the overlay of $S_1, T_1, \ldots, S_m, T_m$, i.e. formed by taking the intersection of, for each i, either S_i or $V - S_i$, and either T_i or $V - T_i$.)

Consider any A, $B \subseteq V$. Taking $S = A$, $T = B$ in Item 2 yields

$$\left|\frac{1}{|E|}|(A \times B) \cap E| - \mathbb{E}[\chi_{A\times B}(Y)]\right| = |\mathbb{E}[\chi_{A\times B}(U_E)] - \mathbb{E}[\chi_{A\times B}(Y)]| < \epsilon.$$

Thus, by triangle inequality it suffices to show that

$$\left|\frac{1}{|E|}\sum_{e\in A\times B}\text{weight}(e) - \mathbb{E}[\chi_{A\times B}(Y)]\right| < \epsilon.$$

To do so, we randomly generate a set $\widetilde{A}$ as follows: For each i, w.p. $|V_i \cap A| / |V_i|$ include all elements of V_i in $\widetilde{A}$, otherwise include none of the elements in $\widetilde{A}$. Similarly generate a random $\widetilde{B}$. Note that $\mathbb{E}_Y[\chi_{A\times B}(Y)] = \mathbb{E}_{\widetilde{A},\widetilde{B},Y}\left[\chi_{\widetilde{A}\times\widetilde{B}}(Y)\right]$ since within every $V_i \times V_j$, $\Pr[Y = e]$ is constant for all $e \in V_i \times V_j$, and

$$\frac{1}{|E|}\sum_{e\in A\times B}\text{weight}(e) = \frac{1}{|E|}\sum_{i,j}\frac{|V_i \cap A| \cdot |V_j \cap B| \cdot |(V_i \times V_j) \cap E|}{|V_i| \cdot |V_j|} = \mathop{\mathbb{E}}_{\widetilde{A},\widetilde{B},U_E}\left[\chi_{\widetilde{A}\times\widetilde{B}}(U_E)\right]$$

by linearity of expectation. Taking $S = \widetilde{A}$, $T = \widetilde{B}$ in Item 2 yields the required bound.

□

2.3.3 Regularity Theorems for Circuit Complexity

In this section, we extend the notion of complexity to be computational and consider (boolean) circuit complexity. Let $\mathcal{W}$ be the set of functions having low circuit complexity. Indeed, the highly constructive proof for Theorem 2.8 already provides a Y with low circuit

complexity, as long as there exist approximate KL projections computed by small circuits. Thus we require $\mathcal{V}$ to be *KL-projectable*:

Definition 2.11. Let $\mathcal{V}$ be a convex set of distributions on $\{0,1\}^n$. The *ϵ-neighborhood* of $\mathcal{V}$, denoted $\mathcal{V}^\epsilon$, is the set of all distributions X on $\{0,1\}^n$ such that for some $Y \in \mathcal{V}$ and for all $x \in \{0,1\}^n$,

$$\Pr[X = x] \in [e^{-2\epsilon}, e^{2\epsilon}] \cdot \Pr[Y = x].$$

$\mathcal{V}$ is said to be *KL-projectable* if for all $\epsilon > 0$, for every $X \in \mathcal{V}^\epsilon$ there exists some $Y \in \mathcal{V}$ such that

1. Y is an ϵ^2-approximate KL projection of X on $\mathcal{V}$;
2. If there is a size t circuit computing a measure M for X with outputs $M(x)$ of bit-length at most m, then there is a size $t + \text{poly}(m, \log(1/\epsilon))$ circuit M' computing a measure for Y with outputs $M'(x)$ of bit-length at most $m + \text{polylog}(1/\epsilon)$. (Recall that measures are $[0,1]$ bounded, unnormalized mass functions; see Definition 1.3.)

Many natural convex sets of distributions are KL-projectable. Examples include the set of distributions with min-entropy at least k (Theorem A.3) and the set of distributions with Shannon entropy at least k (Chapter 4 Theorem 4.52), for any $k > 0$.

We show that every distribution $X \in \mathcal{V}$ is indistinguishable, by all small circuits, to some $Y \in \mathcal{V}$ that has low circuit complexity, as long as $\mathcal{V}$ is permutation-invariant and KL-projectable:

Theorem 2.12 (A Regularity Theorem for circuit complexity)**.** *Let $\mathcal{V}$ be a KL-projectable convex set of distributions on $\{0,1\}^n$ that contains U_n, $t > 0$, and $\epsilon > 0$. Then for every distribution $X \in \mathcal{V}$ there exists $Y \in \mathcal{V}$ such that*

1. *X and Y are $O(\epsilon)$-indistinguishable by size t circuits;*

2. *Y has low complexity: Y has a measure of circuit size $t' = S \cdot t + \mathrm{poly}(S, \log(1/\epsilon))$, for $S = (n - \mathrm{H}_{\mathrm{sh}}(X))/\epsilon^2$.*

Proof. The proof is essentially the same as Theorem 2.8. Suppose for contradiction that for every low complexity $Y \in \mathcal{V}$ there is some size t circuit W such that $\mathbb{E}[W(X)] - \mathbb{E}[W(Y)] \geq \epsilon'$, where $\epsilon' = c \cdot \epsilon$ for a sufficiently large constant c. We will apply Theorem 2.4 (Uniform Min-Max Theorem), with

- $\mathcal{V} = \mathcal{V}$;
- $\mathcal{W} = \{\text{(deterministic) circuits of size } t\}$;
- $f(z, W) = \mathbb{E}[W(X)] - W(z)$.

This corresponds to the two-player zero-sum game where Player 1 chooses some distribution $Y \in \mathcal{V}$, and Player 2 chooses a t sized circuit W, with expected payoff $\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]$ for Player 2. We implement Algorithm 2.1 (Finding Universal Strategy) with KL projection on the set $\mathcal{V}$ as follows. Start with an initial distribution $V^{(1)}$ that is uniform on $\{0,1\}^n$ (which lies in $\mathcal{V}$). In each of the $S = (n - \mathrm{H}_{\mathrm{sh}}(X))/\epsilon^2$ iterations we represent the distribution $V^{(i)}$ by a circuit $M^{(i)}$ computing a measure for $V^{(i)}$, where $M^{(i)}(x)$ has bit-length at most $i \cdot \mathrm{polylog}(1/\epsilon)$. We implement the ith iteration as follows. For technical convenience we assume that $e^{-\epsilon}$ has bit-length at most $\log(1/\epsilon)$ (if not, we replace ϵ by some $\tilde{\epsilon} = O(\epsilon) > \epsilon$ such that $e^{-\tilde{\epsilon}}$ has bit-length at most $\log(1/\epsilon)$).

1. **Obtaining Player 2's Response $W^{(i)}$:** Suppose that we have constructed a $t_i \leq t'$ sized circuit $M^{(i)}$ computing a measure for $V^{(i)}$, and $M^{(i)}(x)$ has bit-length at most $i \cdot \mathrm{polylog}(1/\epsilon)$. By assumption, there is a size t circuit $W^{(i)}$ such that

$$\mathbb{E}\left[W^{(i)}(X)\right] - \mathbb{E}\left[W^{(i)}(V^{(i)})\right] \geq \epsilon'.$$

2. **Weight Update:** We represent the resulting distribution $V^{(i)'}$ by the circuit $M^{(i)'}(x) = \exp\left(-\epsilon \cdot (1 - W^{(i)}(x))\right) \cdot M^{(i)}(x)$ that computes a measure for $V^{(i)'}$. Since $W^{(i)}(x) \in \{0, 1\}$, $\exp\left(-\epsilon \cdot (1 - W^{(i)}(x))\right)$ has bit-length at most $\log(1/\epsilon)$. $M^{(i)}(x)$ has bit-length at most $i \cdot \text{polylog}(1/\epsilon)$, thus multiplication takes time $i \cdot \text{polylog}(1/\epsilon)$. Thus $M^{(i)'}$ has circuit size $t'_i = t_i + t + i \cdot \text{polylog}(1/\epsilon)$, and bit-length at most $i \cdot \text{polylog}(1/\epsilon) + \log(1/\epsilon)$.

3. **KL Projection:** By KL-projectability of $\mathcal{V}$ and the fact that $V^{(i)'} \in \mathcal{V}^\epsilon$, we have a circuit $M^{(i+1)}$ computing a measure for $V^{(i+1)}$ of size $t_{i+1} = t'_i + \text{poly}(i \cdot \text{polylog}(1/\epsilon), \log(1/\epsilon))$, and $M^{(i+1)}(x)$ has bit-length at most $i \cdot \text{polylog}(1/\epsilon) + \log(1/\epsilon) + \text{polylog}(1/\epsilon) = (i + 1) \cdot \text{polylog}(1/\epsilon)$.

Note that $t_1 = O(1)$ and $t_{i+1} = t_i + t + \text{poly}(i \cdot \text{polylog}(1/\epsilon), \log(1/\epsilon))$, thus $t_i \leq S \cdot t + \text{poly}(S, \log(1/\epsilon))$ and the assumption that $t_i \leq t'$ is satisfied for all $i \in [S]$. By Theorem 2.4, W^* (the uniform distribution over $W^{(1)}, \ldots, W^{(S)}$) satisfies

$$\mathbb{E}\left[W^*(X)\right] - \mathbb{E}\left[W^*(V)\right] \geq \epsilon' - O(\epsilon) > 0.$$

for all Player 1 strategies $V \in \mathcal{V}$ such that $\mathrm{H}_{\mathrm{sh}}(V) \geq \mathrm{H}_{\mathrm{sh}}(X)$. Taking $V = X$ yields a contradiction. □

Remark. Most of our results in this section (Theorem 2.12, 2.14) hold not just for small circuits, but for an arbitrary class of distinguishers $\mathcal{W}$ (like our Theorem 2.8) with a suitable definition of "complexity w.r.t. $\mathcal{W}$" (see [TTV] Theorem 1.1 for one such example). However, we avoid stating results in greater generality since the appropriate definition of "complexity" may vary depending on the choice of $\mathcal{V}$ (to account for the complexity of KL projections) and the application.

The above theorem also has an average-case variant. To express nonuniform complexity in the average-case setting, we consider conditional measures. Recall that for a joint

distribution (X, C), a function M is a conditional measure *for* $C|X$ if for all $x \in \text{supp}(X)$, the function $f(y) = M(x, y)$ is a measure for $C|_{X=x}$.

Definition 2.13. Let $\mathcal{V}$ be a convex set of joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ where X is fixed and C may vary. The *ϵ-neighborhood* of $\mathcal{V}$, denoted $\mathcal{V}^\epsilon$, is the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ such that for some $(X, B) \in \mathcal{V}$ and for all $(x, a) \in \{0,1\}^n \times \{0,1\}^\ell$,

$$\Pr[C = a|X = x] \in [e^{-2\epsilon}, e^{2\epsilon}] \cdot \Pr[B = a|X = x].$$

$\mathcal{V}$ is said to be *KL-projectable* if for all $\epsilon > 0$, for every $(X, C) \in \mathcal{V}^\epsilon$ there exists some $(X, B) \in \mathcal{V}$ such that

1. (X, B) is an ϵ^2-approximate KL projection of (X, C) on $\mathcal{V}$;
2. If there is a size t circuit computing a conditional measure M for $C|X$ with outputs $M(x, a)$ of bit-length at most m, then there is a size $t + \text{poly}(m, \log(1/\epsilon))$ circuit M' computing a conditional measure for $B|X$ with outputs $M'(x, a)$ of bit-length at most $m + \text{polylog}(1/\epsilon)$.

Many natural convex sets of such joint distributions are KL-projectable. Examples include the set of all distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$, the set of distributions (X, C) on $\{0,1\}^n \times \{0,1\}^{O(\log n)}$ with C having average min-entropy at least k given X (Chapter 4 Theorem 4.20), and the set of distributions (X, C) on $\{0,1\}^n \times \{0,1\}^{O(\log n)}$ with C having conditional Shannon entropy at least k given X (Chapter 4 Theorem 4.52), for any $k > 0$.

Theorem 2.14 (A Regularity Theorem for circuit complexity – average case)**.** *Let $\mathcal{V}$ be a KL-projectable set of joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ where X is fixed and C may vary, and $\mathcal{V}$ contains (X, U_ℓ). For every $t > 0$, $\epsilon > 0$, and joint distribution $(X, B) \in \mathcal{V}$, there is a joint distribution $(X, C) \in \mathcal{V}$ such that*

1. *(X, B) and (X, C) are $O(\epsilon)$-indistinguishable by size t circuits;*

2. *C has low complexity given X: $C|X$ has a conditional measure of circuit size $t' = S \cdot t + \mathrm{poly}(S, \log(1/\epsilon))$, for $S = (\ell - \mathrm{H}_{\mathrm{sh}}(C|X))/\epsilon^2$. Moreover, if $\mathcal{V}$ equals the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$, then $t' = S \cdot t + (S \cdot \log(1/\epsilon))^2$.*

Proof. The proof is identical to Theorem 2.12, except we use the Uniform Min-Max Theorem – Average Case (Theorem 2.5). If $\mathcal{V}$ equals the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$, the better bound on the complexity of $C|X$ follows from the fact that we do not need KL projections for such $\mathcal{V}$. □

With $\mathcal{V}$ being the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ (thus involving no KL projections), Theorem 2.14 gives a slight strengthening of a recent result of Jetchev and Pietrzak [JP], with a simpler proof. In [JP] they obtain an (X, C) where given x, $C|_{X=x}$ is samplable by a circuit of size $O\left(\left(2^\ell \cdot \ell \cdot (1/\epsilon)^2 \log(1/\epsilon)\right)^2 \cdot t\right)$. Theorem 2.14 provides an (X, C) where given x, $C|_{X=x}$ can be sampled by a circuit of size $O(2^\ell \cdot (\ell \cdot (1/\epsilon)^2 \cdot t + (\ell \cdot (1/\epsilon)^2 \log(1/\epsilon))^2))$ by computing $M(x, y)$ for all $y \in \{0,1\}^\ell$ and sampling $C|_{X=x}$ using its mass function.

Finally, we show an application of Theorem 2.14: deducing a technical lemma of [GW], which says if X and U are indistinguishable then for any short B jointly distributed with X, there is some C (the "auxilliary information") jointly distributed with U such that (X, B) and (U, C) are indistinguishable. Not only is our proof simpler, but also it guarantees that C has low circuit complexity given U. This "low complexity" version (with slightly weaker parameters) was initially proved by Jetchev and Pietrzak [JP], and an interactive extension was proved by Chung, Lui, and Pass [CLP2] for applications in the context of distributional zero-knowledge.

Lemma 2.15 (Low circuit complexity version of [GW] Lemma 3.1)**.** *Let X and U be distributions over $\{0,1\}^n$, and B be a distribution over $\{0,1\}^\ell$ jointly distributed with X. Suppose X and U are ϵ-indistinguishable by circuits of size t. Then there exists some $C \in \{0,1\}^\ell$ jointly distributed with U such that:*

1. *(X,B) and (U,C) are 2ϵ-indistinguishable by circuits of size $s = t/(2^\ell \cdot \ell \cdot (1/\epsilon)^2) - \ell \cdot ((1/\epsilon)\log(1/\epsilon))^2$.*
2. *C has low complexity given U: $C|U$ has a conditional measure of circuit size $\ell \cdot (1/\epsilon)^2 \cdot s + (\ell \cdot (1/\epsilon)^2 \log(1/\epsilon))^2$.*

Proof. We first apply Theorem 2.14 to obtain a distribution $(X, P(X))$ such that (X,B) and $(X,P(X))$ are ϵ-indistinguishable by size s circuits, where P is a randomized function, and there is a size $\ell \cdot (1/\epsilon)^2 \cdot s + (\ell \cdot (1/\epsilon)^2 \log(1/\epsilon))^2$ circuit M computing a conditional measure for $P(X)|X$. Thus $P(x)$ can be sampled in time $s' = O\left(2^\ell \cdot \left(\ell \cdot (1/\epsilon)^2 \cdot s + \left(\ell \cdot (1/\epsilon)^2 \log(1/\epsilon)\right)^2\right)\right)$ by computing $M(x,y)$ for all $y \in \{0,1\}^\ell$ and sampling $P(x)$ from its mass function.

Let $C = P(U)$. Since P is efficient, indistinguishability of X and U implies that $(X,P(X))$ and $(U,P(U))$ are ϵ-indistinguishable by circuits of size s. (Otherwise, given an s-sized ϵ-distinguisher D for $(X,P(X))$ and $(U,P(U))$ we get an ϵ-distinguisher $T(x) = D(x,P(x))$ for X and U, of circuit size $O(s+s') \leq t$.) By triangle inequality, (X,B) and $(U,P(U)) = (U,C)$ must be 2ϵ-indistinguishable by circuits of size s. □

2.3.4 Regularity Theorems for Time Complexity

In this section, we use the full strength of the Uniform Min-Max Theorem to obtain a low complexity approximation where complexity is measured using *uniform* algorithms. We prove a uniform analogue of Theorem 2.14 (i.e. the average-case setting), but for simplicity we take $\mathcal{V}$ to be the set of all joint distributions (X,C) on $\{0,1\}^n \times \{0,1\}^\ell$, thus

no KL-projection is needed. As an immediate corollary, we provide a "sampling" version of it (Theorem 2.17), which is cleaner and convenient for several applications, but involves exponential dependence on ℓ.

Theorem 2.16 (A Regularity Theorem for time complexity – average case)**.** *Let n be a security parameter, $\ell = \ell(n)$, $t = t(n) \geq n$, $\epsilon = \epsilon(n) > 0$ all computable in* $\mathrm{poly}(n)$ *time. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. Let A be a t-time randomized oracle algorithm. Then there is a $t' = \mathrm{poly}(t, 1/\epsilon)$-time randomized algorithm R such that w.p. $\Omega(\epsilon^2/\ell)$ over $M \leftarrow R(1^n)$ and $W \leftarrow A^M(1^n)$, if we interpret M as a deterministic circuit computing a conditional measure for $C|X$ and W as a randomized circuit $W : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$, we have:*

$$\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)] < \epsilon.$$

Proof. We will let R be an implementation of Algorithm 2.2 (Finding Universal Strategy – Average Case), using A as a subroutine. We then show R satisfies the desired properties by applying Theorem 2.5 (Uniform Min-Max Theorem – Average Case), with

- $\mathcal{V}$ being the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ (where the marginal distribution of X is fixed, and C may vary);
- $\mathcal{W} = \{\text{randomized circuits of size } t\}$;
- $f((x, y), W) = \mathbb{E}[W(X, B)] - \mathbb{E}[W(x, y)]$.

This corresponds to the two-player zero-sum game where Player 1 selects a distribution $(X, C) \in \mathcal{V}$, Player 2 selects a size t circuit W and receives expected payoff $\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]$.

Our implementation of Algorithm 2.2 using A is as follows. We set the ϵ in Algorithm 2.2 to be $\epsilon' = \epsilon/c$ for a sufficiently large constant c, and start with an initial distribution

$(X, C^{(1)}) = (X, U_\ell)$ (where U_ℓ is independent of X). In each of the $S = O(\ell/\epsilon^2)$ iterations we represent $C^{(i)}$ by a circuit $M^{(i)}$ computing a conditional measure for $C^{(i)}|X$, i.e. $M^{(i)}(x, y) \propto \Pr[C^{(i)} = y|X = x]$. So we can take $M^{(1)}(x, y) = 1$ for all x, y. We implement the ith iteration as follows, with $\gamma = 1/3S$:

1. **Obtaining Player 2's Response $W^{(i)}$:** Suppose that we have constructed a t_i-size circuit $M^{(i)}$ where $M^{(i)}(x, y)$ has bit-length $i \cdot \text{polylog}(1/\epsilon)$. There are two steps.

   (a) We run $A^{M^{(i)}}(1^n)$ to obtain a t-size randomized circuit $\widehat{W}^{(i)}$, and convert it into a $O(tm)$-size deterministic circuit $\widetilde{W}^{(i)}$ by hardwiring $m = O((1/\epsilon^2)\log(1/\gamma))$ samples of the coins of $\widehat{W}^{(i)}$, so that w.p. at least $1 - \gamma$,

   $$\mathbb{E}\left[\widetilde{W}^{(i)}(X, B)\right] - \mathbb{E}\left[\widetilde{W}^{(i)}(X, C^{(i)})\right] \geq \mathbb{E}\left[\widehat{W}^{(i)}(X, B)\right] - \mathbb{E}\left[\widehat{W}^{(i)}(X, C^{(i)})\right] - \epsilon'.$$

   (b) Our choice of $W^{(i)}$ is the following approximation to $\widetilde{W}^{(i)}$, so that $\exp(-\epsilon' \cdot (1 - W^{(i)}(x, y)))$ can be computed precisely and efficiently. First, we use Newton's method to compute a $\text{polylog}(1/\epsilon)$-bit approximation $E(x, y) \in (0, 1]$ of $\exp(-\epsilon' \cdot (1 - \widetilde{W}^{(i)}(x, y)))$ within $\pm\epsilon'^2$ error, in time $O(tm) + \text{polylog}(1/\epsilon)$. We define $W^{(i)}$ to be such that $\exp(-\epsilon' \cdot (1 - W^{(i)}(x, y))) = E(x, y)$. Thus $\left|W^{(i)}(x, y) - \widetilde{W}^{(i)}(x, y)\right| \leq \epsilon'$, and

   $$\mathbb{E}[W^{(i)}(X, B)] - \mathbb{E}[W^{(i)}(X, C^{(i)})] \geq \mathbb{E}[\widetilde{W}^{(i)}(X, B)] - \mathbb{E}[\widetilde{W}^{(i)}(X, C^{(i)})] - 2\epsilon'.$$

2. **Weight Update:** We represent the resulting distribution $C^{(i+1)}$ by the circuit

   $$M^{(i+1)}(x, y) = \exp\left(-\epsilon' \cdot (1 - W^{(i)}(x, y))\right) \cdot M^{(i)}(x, y)$$

   computing a conditional measure for $C^{(i+1)}|X$. Since $\exp\left(-\epsilon' \cdot (1 - W^{(i)}(x, y))\right) = E(x, y)$ has bit-length $\text{polylog}(1/\epsilon)$ and $M^{(i)}(x, y)$ has bit-length $i \cdot \text{polylog}(1/\epsilon)$, multiplication takes time $i \cdot \text{polylog}(1/\epsilon)$. Thus $M^{(i+1)}$ has circuit size $t_{i+1} = t_i + O(tm) +$

$i \cdot \text{polylog}(1/\epsilon)$ and bit-length $(i+1) \cdot \text{polylog}(1/\epsilon)$, and can be constructed in similar time.

3. **KL projection:** Do nothing as Player 1 strategies can be arbitrary conditional distributions $C^{(i)}|_{X=x}$.

Now let R be the algorithm that chooses a random $i \leftarrow [S]$, runs the above implementation of Algorithm 2.2 for $i-1$ iterations to construct and output $M^{(i)}$. Since $t_1 = O(1)$, we have $t_i = O(1) + S \cdot (O(tm) + S \cdot \text{polylog}(1/\epsilon))$ for all $i \in [S]$. Thus R runs in total time $\text{poly}(t, S, m, \log(1/\epsilon)) \leq t'$.

Suppose for contradiction that w.p. at least $1-\gamma$ over coins of R used to generate $M^{(i)}$ and A, $A^{M^{(i)}}(1^n)$ outputs a randomized circuit $\widehat{W}^{(i)}$ s.t. $\mathbb{E}[\widehat{W}^{(i)}(X,B)] - \mathbb{E}[\widehat{W}^{(i)}(X,C^{(i)})] \geq \epsilon$. By union bound w.p. at least $1 - 2\gamma \cdot S = 1/3$, in all iterations we have

$$\mathbb{E}[W^{(i)}(X,B)] - \mathbb{E}[W^{(i)}(X,C^{(i)})] \geq \mathbb{E}[\widehat{W}^{(i)}(X,B)] - \mathbb{E}[\widehat{W}^{(i)}(X,C^{(i)})] - 3\epsilon' \geq \epsilon - 3\epsilon'.$$

Let W^* be the uniform distribution over $W^{(1)}, \ldots, W^{(S)}$. By the Uniform Min-Max Theorem – Average Case (Theorem 2.5), w.p. at least $1/3$, W^* satisfies

$$\mathbb{E}\left[W^*(X,B)\right] - \mathbb{E}\left[W^*(X,C)\right] \geq \epsilon - 3\epsilon' - O(\epsilon') > 0$$

for all Player 1 strategies $(X,C) \in \mathcal{V}$. Taking $(X,C) = (X,B)$ yields a contradiction. □

As an immediate corollary, we obtain a "sampling" version, which is cleaner, and convenient for several applications. Recall that for a distribution Z, we denote by O_Z the sampling oracle of Z, i.e. on each query O_Z returns a random sample of Z.

Theorem 2.17 (A Regularity Theorem for time complexity – average case (sampling version))**.** *Let n be a security parameter, $\ell = \ell(n)$, $t = t(n) \geq n$, $\epsilon = \epsilon(n) > 0$ all computable in $\text{poly}(n)$ time. Let $(X,B) = (X,B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$, and*

$Q = Q(n)$ be any $\mathrm{poly}(n)$-time samplable distribution on $\{0,1\}^n$. Let A be a t-time randomized oracle algorithm. Then there is a $t' = \mathrm{poly}(2^\ell, t, 1/\epsilon)$-time randomized algorithm R that w.p. at least $\Omega(\epsilon^2/\ell)$ outputs a randomized circuit P of size at most t' satisfying:

$$\mathbb{E}[A^{O_{Q,P(Q)}}(X,B)] - \mathbb{E}[A^{O_{Q,P(Q)}}(X,P(X))] < \epsilon.$$

Proof. Given a t-time randomized oracle algorithm A, we define a $2^\ell \cdot \mathrm{poly}(t, 1/\epsilon)$-time randomized oracle algorithm A' to which we apply Theorem 2.16, as follows. First define the randomized function $\widehat{A}(x, y; a)$ to equal $A(x, y)$ where we fix the outputs of the sampling oracle to be $a \in \left(\{0,1\}^n \times \{0,1\}^\ell\right)^t$. For every $M : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$, let $A'^M(1^n)$ generate $a^{(1)}, \ldots, a^{(m)}$ as $m = O((1/\epsilon^2) \cdot \log(c\ell/\epsilon^2))$ random samples of $(Q, P_M(Q))^t$, where P_M is the randomized function such that M is a conditional measure for $P_M(Q)|Q$, and c is a constant to be determined later. Recall that Q is $\mathrm{poly}(n)$-time samplable by assumption, and we can construct from M a circuit that samples $(Q, P_M(Q))$ by computing $M(x, y)$ for all $y \in \{0,1\}^\ell$. We then let $A'^M(1^n)$ output a randomized circuit $W(x, y)$ computing the average of $\widehat{A}(x, y; a^{(i)})$ over all i. By a Chernoff bound, w.p. at least $\epsilon^2/c\ell$ over $W \leftarrow A'^M(1^n)$ we have

$$\mathbb{E}\left[W(X,B)\right] - \mathbb{E}\left[W(X,P_M(X))\right] \geq \mathbb{E}\left[A^{O_{Q,P_M(Q)}}(X,B)\right] - \mathbb{E}\left[A^{O_{Q,P_M(Q)}}(X,P_M(X))\right] - \epsilon/2.$$

By applying Theorem 2.16 to A', there is a $\mathrm{poly}(2^\ell, t, 1/\epsilon)$-algorithm R such that w.p. $\Omega(\epsilon^2/\ell)$ over $M \leftarrow R(1^n)$ and $W \leftarrow A'^M(1^n)$ we have

$$\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)] < \epsilon/2.$$

Thus w.p. at least $\Omega(\epsilon^2/\ell) - \epsilon^2/c\ell = \Omega(\epsilon^2/\ell)$ (for a sufficiently large c) over $M \leftarrow R(1^n)$,

$$\mathbb{E}[A^{O_{Q,P_M(Q)}}(X,B)] - \mathbb{E}[A^{O_{Q,P_M(Q)}}(X,P_M(X))] < \epsilon.$$

□

We now apply Theorem 2.17 to show Theorem 2.18, the uniform analogue of Theorem 2.15 (which in turn is the low circuit complexity version of [GW] Lemma 3.1). We do so mainly because it is convenient for applications, including (i) deriving a uniform Dense Model Theorem (see Section 3.2, Theorem 3.11); (ii) showing impossibility of constructing succinct non-interactive arguments (SNARGs) via black-box reductions under uniform hardness assumptions (see Section 6, Theorem 6.7).

Theorem 2.18 (Low time complexity version of Lemma 3.1 of [GW])**.** *Let n be a security parameter, $\ell = \ell(n)$, $s = s(n) \geq n$, $\epsilon = \epsilon(n) > 0$ all computable in* $\mathrm{poly}(n)$ *time. Let $X = X(n)$ and $U = U(n)$ be* $\mathrm{poly}(n)$*-time samplable distributions on $\{0,1\}^n$ that are ϵ-indistinguishable for s-time randomized algorithms. Let $B = B(n)$ be a distribution on $\{0,1\}^\ell$ jointly distributed with X, and let $Q = Q(n)$ be any* $\mathrm{poly}(n)$*-time samplable distribution on $\{0,1\}^n$. Let A be a t-time randomized oracle algorithm, for $t = s^{\Omega(1)}/\mathrm{poly}(2^\ell, 1/\epsilon)$. Then there is a $t' = \mathrm{poly}(2^\ell, t, 1/\epsilon)$-time randomized algorithm R such that w.p. at least $\Omega(\epsilon^2/\ell)$, R outputs a randomized circuit P satisfying*

$$\mathbb{E}[A^{O_{Q,P(Q)}}(X,B)] - \mathbb{E}[A^{O_{Q,P(Q)}}(U,P(U))] < 2\epsilon.$$

Proof. By Theorem 2.17, there is a t'-time algorithm R that w.p. at least $\gamma = \Omega(\epsilon^2/\ell)$ outputs a randomized circuit P satisfying

$$\mathbb{E}[A^{O_{Q,P(Q)}}(X,B)] - \mathbb{E}[A^{O_{Q,P(Q)}}(X,P(X))] < 0.9\epsilon.$$

Since P is efficient, ϵ-indistinguishability of X and U implies that with probability at least $1 - \gamma/2$ over P,

$$\mathbb{E}[A^{O_{Q,P(Q)}}(X,P(X))] - \mathbb{E}[A^{O_{Q,P(Q)}}(U,P(U))] < 1.1\epsilon.$$

Indeed, suppose that $A^{Q,P(Q)}$ achieves distinguishing advantage at least 1.1ϵ w.p. at least $\gamma/2$ over P, then we could obtain an ϵ-distinguisher for X and U by running R for

$O((1/\gamma)\log(1/\epsilon))$ times, each time testing the distinguisher $T(x) = A^{Q,P'(Q)}(x, P'(x))$ where P' is the randomized circuit output by R (by running on $O((1/\epsilon^2)\log(1/\epsilon))$ random samples of X, U and $(Q, P'(Q))$), and finally taking the best one. This yields an ϵ-distinguisher for X and U that runs in time $O((1/\epsilon^2)\log(1/\epsilon)) \cdot (\text{poly}(n) + (1/\gamma)\log(1/\epsilon) \cdot (t + \text{poly}(t'))) \leq s$, violating their indistinguishability.

Combining the two inequalities, we get with probability at least $\gamma/2$ over $P \leftarrow R$,

$$\mathbb{E}[A^{O_{Q,P(Q)}}(X, B)] - \mathbb{E}[A^{O_{Q,P(Q)}}(U, P(U))] < 2\epsilon.$$

□

Chapter 3

Uniform Hardcore Theorem and Dense Model Theorem

A fundamental result in complexity theory is Impagliazzo's Hardcore Theorem [Imp], which says that a boolean function f that is hard on average must contain a large subset of inputs on which f is indistinguishable to a random function.

Closely related is the Dense Model Theorem of Green and Tao [GT], Tao and Ziegler [TZ], various formulations of which are due to Reingold et al. and Gowers [RTTV, Gow]. Green and Tao used it as a key in their celebrated result of on arithmetic progressions of prime numbers. Roughly, the Dense Model Theorem says that a dense subset of a pseudorandom set must be indistinguishable to a dense set.

In this chapter, we apply the Uniform Min-Max Theorem and related results from Chapter 2 to obtain simpler proofs of Uniform Hardcore Theorem and Uniform Dense Model Theorem, where uniform means that indistinguishability is with respect to uniform polynomial-time algorithms.

3.1 Uniform Hardcore Theorem

Impagliazzo's Hardcore Theorem [Imp], in strengthened versions due to Klivans and Servedio [KS] and Holenstein [Hol1], says that every function $f : \{0,1\}^n \to \{0,1\}$ that is δ-hard for poly-sized boolean circuits (that is, every poly-sized circuit fails to compute f on at least δ fraction of inputs) must be *extremely* hard on a subset of inputs of density at least 2δ (the *hardcore set*) (and may be easy elsewhere). Following [Imp], we will deal with hardcore distributions instead of hardcore sets, which are equivalent up to a negligible additive difference in density, where density of a distribution is defined as follows:

Definition 3.1 (Density of distribution)**.** Let X and Y be distributions over some finite set Σ. We say X is δ-*dense in* Y if $\Pr[Y = x] \geq \delta \cdot \Pr[X = x]$ for all $x \in \Sigma$. We say X is δ-*dense* if it is δ-dense in U_Σ (equivalently, having min-entropy at least $\log|\Sigma| - \log(1/\delta)$). We denote by $\mathcal{C}_{m,\delta}$ the set of all δ-dense distributions on $\{0,1\}^m$.

The asymptotically optimal nonuniform Hardcore Theorem is due to [KS], using techniques from boosting and an idea of iteratively increasing hardcore size due to Wigderson, and can be stated as follows:

Theorem 3.2 (Hardcore Theorem [KS])**.** *Let* (X, B)[1] *be a joint distribution on* $\{0,1\}^n \times \{0,1\}$ *and* $\epsilon > 0$*. Let* B *be* (t,δ)*-hard given* X*, i.e. for every size* t *circuit* P *it holds that* $\Pr[P(X) = B] \leq 1 - \delta$*. Then there is a joint distribution* $(\hat{X}, \hat{B})$ *that is* 2δ*-dense in* (X, B)*, such that for every size* $t' = t/O(\log(1/\delta)/\epsilon^2)$ *circuit* A *it holds that* $\Pr[A(\hat{X}) = \hat{B}] \leq (1+\epsilon)/2$*.*

Theorem 3.2 is asymptotically optimal as it achieves optimal hardcore density 2δ, as

[1]The version we state is a slight generalization of the version in [KS], which only allows B to be a deterministic boolean function of X. However, the more general version follows readily from almost the same proof.

well as optimal complexity blow-up $O(\log(1/\delta)/\epsilon^2)$, where the lower bound of $\Omega(\log(1/\delta)/\epsilon^2)$ is due to Lu, Tsai, and Wu [LTW][2].

The original paper of Impagliazzo [Imp] contains both a non-trivial constructive proof, as well as a much simpler, yet non-constructive proof due to Nisan that uses the Min-Max Theorem. Nisan's proof has an appealing simplicity: Assume for contradiction that there is no hardcore distribution of high density. Then, by the Min-Max Theorem there is a *universal* predictor A^* such that for every $(\hat{X}, \hat{B})$ that is dense in (X, B) it holds that $\Pr\left[A^*(\hat{X}) = \hat{B}\right] > (1+\epsilon)/2$. A^* is a distribution over circuits of size t, and its prediction probability is taken over this distribution as well as $(\hat{X}, \hat{B})$. By subsampling we can assume that A^* is uniform over a multiset of $S = O((1/\epsilon^2)\log(1/\epsilon\delta))$ circuits of size t, while changing the advantage ϵ by at most a constant fraction. Given the universal predictor A^*, one can build a good predictor for B, contradicting the hardness of B given X, as formalized in Lemma 3.3:

Lemma 3.3 (From universal circuit to predictor [Imp])**.** *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$. Let A^* be the uniform distribution over a multiset of S circuits of size t. Suppose for every joint distribution $(\hat{X}, \hat{B})$ that is δ-dense in (X, B) it holds that $\Pr\left[A^*(\hat{X}) = \hat{B}\right] > (1+\epsilon)/2$. Then there is a circuit P of size $O(S \cdot t)$ such that $\Pr[P(X) = B] > 1 - \delta$.*

Specifically, we can let $P(x) = \text{majority}\{A(x) : A \in A^*\}$. Equivalently, $P(x)$ outputs 1 with probability

$$\frac{1}{2}\left(1 + \text{sign}\left(\Pr[A^*(x) = 1] - \frac{1}{2}\right)\right).$$

Unfortunately, both proofs in [Imp] yield a suboptimal hardcore density of δ. Following Nisan's proof using Min-Max Theorem, Holenstein [Hol1] proves the Hardcore Theorem with

[2] [LTW] showed a black-box lower bound on the number of t'-sized circuits that a black-box reduction needs to obtain to construct some P with $\Pr[P(X) = B] > 1 - \delta$.

optimal hardcore density of 2δ (Theorem 3.2), by strengthening the above lemma to Lemma 3.4 below (using a trick from Levin's proof of the XOR Lemma).

Lemma 3.4 (From universal circuit to *optimal* predictor [Hol1])**.** *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$. Let A^* be the uniform distribution over a multiset of S circuits of size t. Suppose for every joint distribution $(\hat{X}, \hat{B})$ that is 2δ-dense in (X, B) it holds that $\Pr\left[A^*(\hat{X}) = \hat{B}\right] > (1+\epsilon)/2$. Then there is a circuit P of size $O(S \cdot t)$ such that $\Pr[P(X) = B] > 1 - (1-\epsilon)\,\delta$.*

Specifically, we can let $P(x)$ output 1 with probability $p(x)$ truncated at 0 and 1 (i.e. $P(x) = \min\{\max\{p(x), 0\}, 1\}$), for

$$p(x) = \frac{1}{2}\left(1 + \frac{\Pr_{A^*}[A^*(x) = 1] - \frac{1}{2}}{\phi}\right)$$

where ϕ is the least number s.t. $\Pr_{X,B}\left[\Pr_{A^*}\left[A^*(X) = B\right] \leq 1/2 + \phi\right] \geq 2\delta$. (w.l.o.g. ϕ is a multiple of $1/S$.)

One drawback of proofs based on the standard Min-Max Theorem is the suboptimal complexity blow-up (due to suboptimal settings of S from the probabilistic construction of the multiset defining A^*). By replacing the use of Min-Max Theorem with the Uniform Min-Max Theorem, we immediately achieve optimal complexity blow-up (by replacing the probabilistic construction of the multiset with a smarter online learning/boosting algorithm).

Remark 3.5. In Chapter 4 we prove a generalization of the Hardcore Theorem where B, rather than being binary, can be $O(\log n)$ bits long. While the proof also begins with the Min-Max Theorem, it differs substantially thereafter. In particular, it achieves optimal hardcore density without explicitly relying on Lemma 3.4 (i.e. the trick from the XOR Lemma). Nonetheless, the complexity blow-up in that version is not known to be optimal (even after replacing its use of the Min-Max Theorem by the Uniform Min-Max Theorem).

Another drawback of proofs based on the standard Min-Max Theorem is that they are non-constructive. Indeed, a constructive proof such as the one by Impagliazzo [Imp] can be interpreted as a Hardcore Theorem for the *uniform* setting of hardness, where the hardness is with respect to efficient algorithms rather than small circuits. (See Theorem 3.6 below for the exact formulation). This *Uniform Hardcore Theorem* is needed for several important applications ([KS, Hol1, Hol2, HHR1, HRV]). Building on the constructive proof in [Imp], Holenstein [Hol1] also shows a *Uniform* Hardcore Theorem with optimal hardcore density, but is rather involved and fails to achieve the optimal complexity blow-up $O(\log(1/\delta)/\epsilon^2)$. Subsequently, Barak, Hardt, and Kale ([BHK]) gave an alternative proof of Uniform Hardcore Theorem achieving optimal complexity blow-up of $O(\log(1/\delta)/\epsilon^2)$ as well as optimal hardcore density 2δ (by using Lemma 3.4), based on ideas of multiplicative weights and Bregman projection.

As an application of the Uniform Min-Max Theorem (which itself is inspired by [BHK]), we offer a new proof of the Uniform Hardcore Theorem of [BHK] (with optimal hardcore density and complexity blow-up). The advantage of the new proof is that it is more modular: we simply replace the use of Min-Max Theorem in Holenstein's proof (of the nonuniform Hardcore Theorem, Theorem 3.2) with the Uniform Min-Max Theorem. In contrast, [BHK] adapt the analysis of multiplicative weights and Bregman projection (from [HW]) to the specific context of the Hardcore Theorem.

Notation. For a distribution Z, let O_Z denote the oracle that gives a random sample from Z when queried.

Theorem 3.6 (Uniform Hardcore Theorem)**.** *Let n be a security parameter, $m = m(n) =$ $\mathrm{poly}(n)$, $\delta = \delta(n)$, $\epsilon' = \epsilon'(n)$, $q = q(n)$ all computable in $\mathrm{poly}(n)$ time, and $(X, B) = g(U_m)$ be a joint distribution where $g : \{0,1\}^m \to \{0,1\}^n \times \{0,1\}$ is computable in $\mathrm{poly}(n)$ time. Suppose that (X, B) has no hardcore distribution of density at least 2δ, i.e. there is a t-time*

oracle algorithm A such that for infinitely many n and every $C \in \mathcal{C}_{m,2\delta}$,

$$\Pr_{(x,b)\leftarrow g(C)}\left[A^{O_C}(x) = b\right] > \frac{1}{2} + \epsilon'.$$

Then there is a $\mathrm{poly}(t, n, 1/\delta, 1/\epsilon')$-time randomized algorithm P such that for infinitely many n,

$$\Pr[P(X) = B] > 1 - \delta.$$

Moreover, P is constructed by making $O(\log(1/\delta)/\epsilon'^2)$ calls to A.

For the proof of Uniform Hardcore Theorem, we will need the notion of *measures*. Recall that measures are simply $[0, 1]$ bounded, unnormalized mass functions.

Definition 3.7 (Density of measure)**.** A measure $M : \mathcal{X} \to [0, 1]$ is *δ-dense* if its *density* $\mu(M) = \sum_{x\in\mathcal{X}} M(x)/\left|\mathcal{X}\right|$ is at least δ. We denote by $\mathcal{M}_{m,\delta}$ the set of all δ-dense measures defined on $\{0, 1\}^m$. One can verify that if $M \in \mathcal{M}_{m,\delta}$ then $\Phi_M \in \mathcal{C}_{m,\delta}$ (but not conversely).

Proof of Theorem 3.6. We will apply Chapter 2, Theorem 2.4 (Uniform Min-Max Theorem), with

- $\mathcal{V} = \mathcal{C}_{m,2\delta}$;
- $\mathcal{W} = \{$(deterministic) circuits of size $tm + \mathrm{poly}(t)\}$;
- $f(z, W) = I(W(x) = b)$, where $(x, b) = g(z)$ and $I(\cdot)$ is the indicator function.

This corresponds to the two-player zero-sum game where Player 1 chooses some distribution $C \in \mathcal{C}_{m,2\delta}$, and Player 2 chooses a $tm + \mathrm{poly}(t)$ sized circuit W, with expected payoff $\mathbb{E}[f(C, W)] = \Pr_{(x,b)\leftarrow g(C)}[W(x) = b]$ for Player 2. We will use A to show that Chapter 2, Algorithm 2.1 (Finding Universal Strategy) with KL projection on the set $\mathcal{V} = \mathcal{C}_{m,2\delta}$ can be implemented efficiently, such that for infinitely many n, in each iteration we obtain some W with good prediction probability. This gives us an efficient universal predictor A^* of B

given X, by the Uniform Min-Max Theorem. From the universal predictor, we then show how to obtain a $(1-\delta)$-predictor of B using Lemma 3.4.

In Algorithm 2.1, we start with an initial distribution $V^{(1)}$ that is uniform on $\{0,1\}^m$. Let $\epsilon = \epsilon'/c$ for a sufficiently large constant c, and $\gamma = \epsilon/2S$. The number of iterations is

$$S = \left(m - \min_{C \in \mathcal{C}_{m,2\delta}} \mathrm{H}_{\mathrm{sh}}(C)\right)/\epsilon^2 = (m - (m - \log(1/2\delta)))/\epsilon^2 = (\log(1/\delta) - 1)/\epsilon^2.$$

In each iteration we represent the distribution $V^{(i)}$ (the current C) by a circuit $M^{(i)}$ computing a measure for $V^{(i)}$. So we can take $M^{(1)}(x) = 1$ for all x. We will need the following claim to implement an iteration.

Claim 3.8. There is a randomized algorithm that, given oracle access to a measure $M \in \mathcal{M}_{m,2\delta}$, w.p. at least $1-\gamma$ outputs a $tm + \mathrm{poly}(t)$ sized (deterministic) boolean circuit W such that $\Pr_{(x,b)\leftarrow g(\Phi_M)}[W(x) = b] > 1/2 + \epsilon' - 4\epsilon$. Moreover it runs in time $t + \mathrm{poly}(n, s, t, 1/\delta, 1/\epsilon', \log(1/\gamma))$ time where s is a bound on the bit length of $M(x)$.

Proof of Claim 3.8. Given oracle access to M, we can generate t random samples of Φ_M in time $t' = t \cdot O((1/\delta)\log(t/\epsilon)) \cdot (s+m) + \mathrm{poly}(n)$ and w.p. at least $1-\epsilon$, using rejection sampling (see Lemma A.2). Thus we can eliminate all A's oracle queries to O_{Φ_M} and obtain some t' time randomized algorithm A' such that $\Pr_{(x,b)\leftarrow g(\Phi_M)}[A'(x) = b] > 1/2 + \epsilon' - \epsilon$.

Write $A'(x) = A'(x; r)$ where r is the coin tosses of A' (which consists of coin tosses for A and at most t' random bits for the rejection sampling). For each r we compute an estimate $E(r)$ of $\Pr_{(x,b)\leftarrow g(\Phi_M)}[A'(x;r) = b]$ within $\pm\epsilon$ error with probability at least $\gamma/2q$, for $q = O((1/\epsilon)\log(1/\gamma))$. By a Chernoff bound, this can be done by testing $A'(\cdot; r)$ on $q' = O((1/\epsilon^2)\log(q/\gamma))$ random samples of $(x,b) \leftarrow g(\Phi_M)$ (which we generate with probability at least $1 - \Theta(\gamma/q)$, again using Lemma A.2). We repeat this for q randomly chosen r, and if $E(r) > 1/2 + \epsilon' - 3\epsilon$ output a circuit W computing $A'(\cdot; r)$.

By union bound with probability at least $1 - \gamma/2$, all q estimates $E(r)$ are within ϵ

error. By the Markov inequality, w.p. $1-(1-\Omega(\epsilon))^q \geq 1-\gamma/2$ at least one of the r's satisfies $\Pr_{(x,b)\leftarrow g(\Phi)}[A'(x;r)=b] > 1/2+\epsilon'-2\epsilon$ so $E(r) \geq 1/2+\epsilon'-3\epsilon$. Moreover we have $\Pr_{(x,b)\leftarrow g(C)}[A(x,r)=b] > 1/2+\epsilon'-4\epsilon$ whenever $E(r) > 1/2+\epsilon'-3\epsilon$. We conclude that w.p. at least $1-\gamma$ we output the desired circuit, all in time $q'q\cdot(t'+O((1/\delta)\log(qq'/\gamma))\cdot(s+m))+\text{poly}(n) = \text{poly}(n,s,t,1/\delta,1/\epsilon',\log(1/\gamma))$. Finally, the circuit W is of size $tm+\text{poly}(t)$ as it simply runs A using the t fixed samples of Φ_M (which can be stored as tm nonuniform bits). □

We now implement the ith iteration as follows. For technical convenience we assume that $e^{-\epsilon}$ has bit-length $\log(1/\epsilon)$ (if not, we replace ϵ by some $\tilde{\epsilon}=O(\epsilon)$ such that $e^{-\tilde{\epsilon}}$ has bit-length $\log(1/\epsilon)$).

1. **Obtaining Player 2's Response $W^{(i)}$:** Suppose that we have constructed a t_i sized circuit $M^{(i)}$ computing a measure for $V^{(i)}$, and outputs of $M^{(i)}$ have bit-length at most $O(i\cdot\log(1/\epsilon))$. Using Claim 3.8, we can obtain a (deterministic) circuit $W^{(i)}$ such that

$$\Pr_{(x,b)\leftarrow g(V^{(i)})}[W^{(i)}(x)=b] > \frac{1}{2}+\epsilon'-4\epsilon,$$

   in time $\text{poly}(t_i,n,t,1/\delta,1/\epsilon,\log(1/\gamma))$ and w.p. at least $1-\gamma$. Note, however, that the circuit size of $W^{(i)}$ is $tm+\text{poly}(t)$, independent of t_i.

2. **Weight Update:** We represent the resulting distribution $V^{(i)'}$ by the circuit $M^{(i)'}(z) = \exp\left(-\epsilon\cdot I\left(W^{(i)}(x)=b\right)\right)\cdot M^{(i)}(z)$, where $(x,b)=g(z)$, which computes a measure for $V^{(i)'}$. Since $I\left(W^{(i)}(x)=b\right)\in\{0,1\}$, $\exp\left(-\epsilon\cdot I\left(W^{(i)}(x)=b\right)\right)$ has bit-length $\log(1/\epsilon)$. $M^{(i)}(z)$ has bit-length $O(i\cdot\log(1/\epsilon))$, thus multiplication takes time $\text{poly}(i\cdot\log(1/\epsilon))$. Thus $M^{(i)'}$ has circuit size $t_i' = t_i+tm+\text{poly}(t)+i\cdot\text{polylog}(1/\epsilon)$, bit-length at most $O(i\cdot\log(1/\epsilon)+\log(1/\epsilon))$, and can be constructed in similar time.

3. **KL Projection:** It is shown in Lemma A.3 (approximating KL projection on high

min-entropy distributions, which is based on Lemma 2.3 of [BHK]) that given $M^{(i)'}$, w.p. $1-\gamma$ one can generate a $t_{i+1} = t'_i + \text{polylog}(1/\epsilon)$ sized circuit $M^{(i+1)}$ computing a measure for a distribution $V^{(i+1)}$ that is an ϵ^2-approximate KL projection of $V^{(i)'} = \Phi_{M^{(i)'}}$ on $\mathcal{C}_{m,2\delta}$. Furthermore, outputs of $M^{(i+1)}$ have bit-length at most $O((i+1)\log(1/\epsilon))$. This can be done in time $\text{poly}(n, 1/\epsilon, \log(1/\delta), \log(1/\gamma)) \cdot t'_i$.

By union bound w.p. at least $1 - 2\gamma S = 1 - \epsilon$ all S iterations complete successfully. Since $t_1 = O(1)$ and $t_{i+1} = t_i + tm + \text{poly}(t) + i \cdot \text{polylog}(1/\epsilon)$, we have $t_i = \text{poly}(n, t, 1/\epsilon, \log(1/\delta))$ for all $i \in [S]$. Let A^* be the uniform distribution over $W^{(1)}, \ldots, W^{(S)}$, thus A^* can be computed in total time $\text{poly}(n, t, 1/\delta, 1/\epsilon)$. By Chapter 2, Theorem 2.4 (Uniform Min-Max Theorem), for all Player 1 strategies $C \in \mathcal{C}_{m,2\delta}$,

$$\Pr_{(x,b)\leftarrow g(C)}[A^*(x) = b] > (1-\epsilon)\left(\frac{1}{2} + \epsilon' - 4\epsilon\right) - O(\epsilon) \geq \frac{1+\epsilon'}{2}.$$

Equivalently, for every joint distribution $(\hat{X}, \hat{B})$ that is 2δ-dense in $(X, B) = g(U_m)$ we have

$$\Pr[A^*(\hat{X}) = \hat{B}] > \frac{1+\epsilon'}{2}$$

(since $(\hat{X}, \hat{B})$ equals $g(C)$ for some $C \in \mathcal{C}_{m,2\delta}$).

From Universal Weak Predictor to $(1-\delta)$-Predictor. Now that we have a universal weak predictor A^* as the uniform distribution over $S = O(\log(1/\delta)/\epsilon'^2)$ circuits, applying Lemma 3.3 already proves a version of the Uniform Hardcore Theorem with suboptimal hardcore density.

To achieve optimal hardcore density, we apply Lemma 3.4 by guessing the value of $\phi \in [0, 1/2]$, which is a multiple of $1/S$. More concretely, for each $\lambda = 1/S, 2/S, \ldots, 1/2$, we compute some estimate E_λ of $\Pr[P_\lambda(X) = B]$, where P_λ denotes the predictor in Lemma 3.4 with ϕ set to λ. Our final (uniform) predictor P will run P_λ for the λ where the estimate E_λ is the highest.

We compute E_λ by taking $O((1/\epsilon'^2\delta^2)\log(1/\epsilon'\delta))$ samples of (X, B) and coins of P_λ, so that by a Chernoff bound, for each λ w.p. at least $1-\epsilon'\delta/4$ we have $|E_\lambda - \Pr[P_\lambda(X) = B]| \leq \epsilon'\delta/4$. The probability that either E_ϕ or the highest estimate is off by more than $\pm\epsilon'\delta/4$ is at most $\epsilon'\delta/2$. So it follows from Lemma 3.4 that

$$\Pr[P(X) = B] \geq \Pr[P_\phi(X) = B] - \epsilon'\delta/2 - \epsilon'\delta/2 > 1 - \left(1-\epsilon'\right)\delta - \epsilon'\delta = 1 - \delta$$

completing the proof. □

3.2 Uniform Dense Model Theorem

A celebrated result of Green and Tao [GT] shows that there exist arbitrarily long arithmetic progressions of prime numbers. A key new component of their proof is the Dense Model Theorem which, in the generalized form of Tao and Ziegler [TZ], says if X is a pseudorandom distribution and D is a distribution dense in X, then D is indistinguishable to a distribution M that is dense in the uniform distribution. Like our results in Chapter 2, Section 2.3.1, notions of indistinguishability and pseudorandomness in the Dense Model Theorem can be defined with respect to an arbitrary class of distinguishers $\mathcal{W}$, and are not restricted to classes of circuit distinguishers.

In the original proof, the indistinguishability (i.e. the bound on distinguishing probability) between D and M is exponentially larger than the indistinguishability between X and the uniform distribution, making it inapplicable for the typical complexity-theoretic or cryptographic settings of parameters. Using the Min-Max Theorem, Reingold et al. [RTTV] provided another proof where the indistinguishability and complexity blow-ups are only polynomial; a similar proof was given by Gowers [Gow]. These requirements are crucial for applications in leakage-resilient cryptography [DP2, DP1, FOR], and for connections to computational differential privacy [MPRV].

We now state a Dense Model Theorem due to Zhang [Zha], where the complexity blow-up $O((\delta/\epsilon)^2 \log(1/\delta))$ is asymptotically optimal.[3]

Recall from Definition 3.1 that for distributions X and Y on Σ, we say X is *δ-dense in Y* if $\Pr[Y = x] \geq \delta \cdot \Pr[X = x]$ for all $x \in \Sigma$, and say X is *δ-dense* if it is δ-dense in U_Σ. It will be convenient to denote by $\mathbf{Th}_t(x)$ the boolean threshold function i.e. $\mathbf{Th}_t(x) = 1$ if $x \geq t$ and $\mathbf{Th}_t(x) = 0$ if $x < t$.

Theorem 3.9 (Dense Model Theorem [Zha])**.** *Let Σ be a finite set, $\mathcal{W}$ be an arbitrary class of functions $W : \Sigma \to [0,1]$, $\epsilon > 0$, $\delta > 0$. Then the following holds for some $S = O((\delta/\epsilon)^2 \log(1/\delta))$.*

Let $\mathcal{W}'$ be the set of all functions $W' : \Sigma \to \{0,1\}$ defined by

$$W'(x) = \mathbf{Th}_t\left(\sum_{i=1}^{S} W_i(x)/S\right)$$

for some $W_1, \ldots, W_S \in \mathcal{W}$ and $t \in [0,1]$. Let X be a distribution on Σ that is ϵ-indistinguishable from U_Σ by $\mathcal{W}'$. Let D be a distribution δ-dense in X. Then there is a δ-dense distribution M such that D and M are $O(\epsilon/\delta)$-indistinguishable by $\mathcal{W}$.

A Min-Max Theorem based proof with a suboptimal blow-up of $S = O((\delta/\epsilon)^2 \log(1/\epsilon))$ proceeds as follows. (Note that we may assume $\delta > \epsilon$, else the conclusion of $O(\epsilon/\delta)$-indistinguishability is trivial.) Assume for contradiction that for every δ-dense M there is a distinguisher $W \in \mathcal{W}$. By the Min-Max Theorem there is a *universal* distinguisher W^* such that $\mathbb{E}[W^*(D)] - \mathbb{E}[W^*(M)] \geq O(\epsilon/\delta)$ for every δ-dense M. By subsampling we can assume that W^* is the average over a multiset of $O((\delta/\epsilon)^2 \log(1/\epsilon))$ elements of $\mathcal{W}$, while changing the distinguishing advantage by at most a constant fraction. Given such universal distinguisher W^* we can construct an ϵ-distinguisher in $\mathcal{W}'$ between X and U_Σ, as formalized in Lemma 3.10:

[3] Zhang [Zha] shows optimality by proving a black-box lower bound on the number of elements of $\mathcal{W}$ that a black-box reduction needs to obtain to construct a distinguisher between X and the uniform distribution.

Lemma 3.10 (Implicit in [RTTV])**.** *Let Σ be a finite set, $\epsilon > 0$, $\delta > 0$. Let X, D be distributions on Σ, and D is δ-dense in X. Let $W^* : \Sigma \to [0,1]$ be a function such that for every δ-dense distribution M we have*

$$\mathbb{E}\left[W^*(D)\right] - \mathbb{E}\left[W^*(M)\right] \geq O(\epsilon/\delta).$$

Then for some t as a multiple of $O(\epsilon/\delta)$, we have

$$\mathbb{E}\left[\mathbf{Th}_t(W^*(X))\right] - \mathbb{E}\left[\mathbf{Th}_t(W^*(U_\Sigma))\right] \geq \epsilon.$$

This proves a Dense Model Theorem, but with a suboptimal complexity blow-up of $O((\delta/\epsilon)^2 \log(1/\epsilon))$ (due to the probabilistic construction of the multiset defining W^*). Zhang [Zha] achieved optimal blow-up in Theorem 3.9 by adapting the technique of multiplicative weights with KL projection from Barak, Hardt, and Kale [BHK].

Replacing the use of the Min-Max Theorem in the above argument by our Uniform Min-Max Theorem (Chapter 2, Theorem 2.4), we immediate obtain a simple proof of Theorem 3.9, with an optimal complexity blow-up that comes from the setting of

$$S = \frac{(\log|\Sigma| - \min_{M\in\mathcal{V}} \mathrm{H}_{\mathrm{sh}}(M))}{\Omega(\epsilon/\delta)^2} = O\left(\frac{\log(1/\delta)}{(\epsilon/\delta)^2}\right)$$

in Theorem 2.4, with $\mathcal{V}$ being the set of δ-dense distribution on Σ. Compared to [Zha], the proof using the Uniform Min-Max Theorem is more modular, and avoids adapting the analysis of [HW] and [BHK] to the specific setting of the Dense Model Theorem.

In the rest of the section, we prove a Uniform Dense Model Theorem where the distinguishers are (uniform) algorithms rather than (nonuniform) $[0,1]$-valued functions. Rather than directly applying the Uniform Min-Max Theorem and using Lemma 3.10, we follow [TTV] and deduce the Dense Model Theorem from a Regularity Theorem. Specifically, [TTV] shows how to deduce the nonuniform Dense Model Theorem from a Nonuniform

Regularity Theorem analogous to Chapter 2, Theorem 2.12; we prove our Uniform Dense Model Theorem using a Uniform Regularity Theorem (Chapter 2, Theorem 2.18).

We begin with an overview of the proof of the nonuniform Dense Model Theorem in [TTV]. The distribution D being δ-dense in X means that there is a (possibly inefficient) binary random variable B jointly distributed with X such that $D = X|_{B=1}$, and $\Pr[B = 1] \geq \delta$. By a Regularity Theorem, there is an efficient randomized function P such that (X, B) and $(X, P(X))$ are indistinguishable. Since P is efficient, indistinguishability of X and U_n implies that $(X, P(X))$ and $(U_n, P(U_n))$ are also indistinguishable. So we can take $M = U_n|_{P(U_n)=1}$. M is δ-dense because $\Pr[P(U_n) = 1] \approx \Pr[P(X) = 1]$, again by indistinguishability of X and U_n. (Note that we use indistinguishability of X and U_n twice. In the uniform setting, the uniform distinguisher will have to determine which case to use, by testing whether $\Pr[P(U_n) = 1] \approx \Pr[P(X) = 1]$ or not.)

Theorem 3.11 (Uniform Dense Model Theorem)**.** *Let n be a security parameter, $\epsilon = \epsilon(n)$, $\delta = \delta(n)$, $s = s(n) \geq n$ all computable in* $\mathrm{poly}(n)$ *time. Let $X = X(n)$ and $U = U(n)$ be poly-time samplable distributions on $\{0,1\}^n$ such that X and U are ϵ-indistinguishable for s-time randomized algorithms. Let $D = D(n)$ be a distribution that is δ-dense in X. Then for some $t = s^{\Omega(1)}/\mathrm{poly}(1/\epsilon, 1/\delta)$ and all t-time randomized oracle algorithms A, there is a distribution $M = M(n)$ that is $(\delta - O(\epsilon))$-dense in U such that for all n,*

$$\mathbb{E}\left[A^{O_M}(D)\right] - \mathbb{E}\left[A^{O_M}(M)\right] \leq O(\epsilon/\delta).$$

Moreover, M is constructive: $M = U|_{P(U)=1}$ for some randomized circuit P such that some $\mathrm{poly}(t, 1/\epsilon)$*-time randomized algorithm R outputs P w.p. at least $\Omega(1/\epsilon^2)$.*

Proof. w.l.o.g. we assume $1 > \delta > c\epsilon$ for a sufficiently large constant c. D being δ-dense in X means that there is a (possibly inefficient) binary random variable B jointly distributed with X such that $D = X|_{B=1}$, and $\delta_B = \Pr[B = 1] = \delta$. Consider any t-time randomized oracle

algorithm A. Let A' be the randomized oracle algorithm where for every joint distribution (U, C) over $\{0,1\}^n \times \{0,1\}$, $A'^{O_{U,C}}$ on input (x, y) does the following:

1. Compute an estimate $\hat{\delta}_C$ of $\delta_C = \Pr[C = 1]$ such that $\left|\hat{\delta}_C - \delta_C\right| \leq \epsilon$ w.p. at least $1 - \epsilon$. To do so we take $O((1/\epsilon^2)\log(1/\epsilon))$ random samples of (U, C) and let $\hat{\delta}_C$ be the fraction on which C equals 1.
2. If $\hat{\delta}_C < \delta_B - 5\epsilon$ then return y; if $\hat{\delta}_C > \delta_B + 5\epsilon$ then return $1 - y$.
3. Otherwise, $\left|\hat{\delta}_C - \delta_B\right| \leq 5\epsilon$, and
   (a) If $y = 0$ then return zero.
   (b) If $y = 1$ then simulate $A^{O_N}(x)$ for the distribution $N = U|_{C=1}$, and return the output. To simulate $A^{O_N}(x)$, we obtain t random samples of N w.p. at least $1 - \epsilon$, where each sample is generated using rejection sampling from $O_{U,C}$ for $O((1/\delta_C)\log(t/\epsilon))$ times, where $\delta_C \geq \delta_B - 4\epsilon \geq \delta/2$.

A' runs in time $t' = t + O((1/\epsilon^2)\log(1/\epsilon)) \cdot \mathrm{poly}(n) + O((1/\delta)\log(t/\epsilon)) \cdot t \cdot \mathrm{poly}(n)$. By Chapter 2, Theorem 2.18, there is a $\mathrm{poly}(t', 1/\epsilon)$-time randomized algorithm R that w.p. at least $\Omega(\epsilon^2)$ outputs a randomized circuit P satisfying

$$\begin{aligned} 2\epsilon &> \mathbb{E}\left[A'^{O_{U,P(U)}}(X, B)\right] - \mathbb{E}\left[A'^{O_{U,P(U)}}(U, P(U))\right] \\ &\geq \Pr_{\hat{\delta}_C}\left[\hat{\delta}_C - \delta_B > 5\epsilon\right] \cdot (\delta_C - \delta_B) + \Pr_{\hat{\delta}_C}\left[\hat{\delta}_C - \delta_B < -5\epsilon\right] \cdot (\delta_B - \delta_C) \\ &\quad + \Pr_{\hat{\delta}_C}\left[\left|\hat{\delta}_C - \delta_B\right| \leq 5\epsilon\right] \cdot \left(\delta_B \cdot \mathbb{E}[A^{O_M}(D)] - \delta_C \cdot \mathbb{E}[A^{O_M}(M)] - \epsilon\right). \end{aligned} \tag{3.1}$$

Take $(U, C) = (U, P(U))$ and $M = U|_{C=1}$. We claim that $\delta_C \geq \delta_B - 6\epsilon$, i.e. M is $(\delta - O(\epsilon))$-dense in U. Indeed, if $\delta_C < \delta_B - 6\epsilon$ then a Chernoff bound implies

$$\Pr_{\hat{\delta}_C}\left[\hat{\delta}_C - \delta_B > 5\epsilon\right] \cdot (\delta_C - \delta_B) + \Pr_{\hat{\delta}_C}\left[\hat{\delta}_C - \delta_B < -5\epsilon\right] \cdot (\delta_B - \delta_C) > 5\epsilon$$

violating Eq. 3.1. By symmetry, we must have $\delta_C \in [\delta_B - 6\epsilon, \delta_B + 6\epsilon]$.

We now show that D and M are indistinguishable by A^{O_M}. Suppose that $\delta_C \in [\delta_B, \delta_B + 6\epsilon]$ (the case $\delta_C \in [\delta_B - 6\epsilon, \delta_B]$ is similar). Then $\Pr_{\hat{\delta}_C}\left[\hat{\delta}_C - \delta_B < -5\epsilon\right] \leq \epsilon$ and Eq. 3.1 implies

$$\begin{aligned} 2\epsilon \geq & \left(1 - \epsilon - \Pr_{\hat{\delta}_C}\left[\left|\hat{\delta}_C - \delta_B\right| \leq 5\epsilon\right]\right) \cdot (\delta_C - \delta_B) - \epsilon \\ & + \Pr_{\hat{\delta}_C}\left[\left|\hat{\delta}_C - \delta_B\right| \leq 5\epsilon\right] \cdot \left(\delta_B \left(\mathbb{E}[A^{O_M}(D)] - \mathbb{E}[A^{O_M}(M)]\right) - (\delta_C - \delta_B) - \epsilon\right) \end{aligned}$$

which simplifies to

$$\delta_B \cdot \left(\mathbb{E}[A^{O_M}(D)] - \mathbb{E}[A^{O_M}(M)]\right) < \frac{3\epsilon - (1-\epsilon)(\delta_C - \delta_B)}{\Pr\left[\left|\hat{\delta}_C - \delta_B\right| \leq 5\epsilon\right]} + 2(\delta_C - \delta_B) + \epsilon. \tag{3.2}$$

Consider these cases:

- If $0 \leq \delta_C - \delta_B < 4\epsilon$, then $\Pr\left[\left|\hat{\delta}_C - \delta_B\right| \leq 5\epsilon\right] \geq 1 - \epsilon$ hence RHS of Eq. 3.2 is at most $3\epsilon/(1-\epsilon) + (2 - (1-\epsilon)/(1-\epsilon))(\delta_C - \delta_B) + \epsilon \leq O(\epsilon)$.

- If $4\epsilon \leq \delta_C - \delta_B \leq 6\epsilon$, then RHS of Eq. 3.2 is at most $2(\delta_C - \delta_B) + \epsilon \leq O(\epsilon)$.

Thus we conclude that $\mathbb{E}[A^{O_M}(D)] - \mathbb{E}[A^{O_M}(M)] \leq O(\epsilon)/\delta_B \leq O(\epsilon/\delta)$. □

Chapter 4

Characterizations of Computational Entropies

Computational analogues of information-theoretic notions have given rise to some of the most interesting phenomena in cryptography and pseudorandomness theory. For example, *(computational) indistinguishability* [GM2], which is the computational analogue of statistical distance, enabled bypassing Shannon's impossibility results on perfectly secure encryption [Sha], and provided the basis for the computational theory of pseudorandomness [BM, Yao2].

Computational analogues of *entropy* were introduced by Yao [Yao2] and Håstad, Impagliazzo, Levin, and Luby [HILL]. The Håstad et al. notions, known as *pseudo-min-entropy* and *pseudoentropy*[1], were key to their fundamental result establishing the equivalence of pseudorandom generators and one-way functions, and have also now become a basic concept in complexity theory and cryptography.

Average-case variants of the Håstad et al. notions are known as *pseudo-avg-min-entropy*

[1]Håstad et al. uses a somewhat different terminology, e.g. pseudoentropy is called "computational entropy."

[HLR] and *conditional pseudoentropy* [HRV], respectively. Pseudo-avg-min-entropy, in the special case involving only a binary alphabet, is equivalent to dense "hardcore distributions" introduced by Impagliazzo [Imp] (see Chapter 3 for discussions on Impagliazzo's Hardcore Theorem). Conditional pseudoentropy was recently introduced by Haitner, Reingold, and Vadhan [HRV] to give a simpler and more efficient construction of pseudorandom generators from one-way functions.

In this chapter, we establish new characterizations of pseudo-avg-min-entropy, pseudoentropy, and conditional pseudoentropy, in terms of certain (different) measures of "hardness" for distributions.

4.1 Introduction

4.1.1 Characterizing Pseudo-Avg-Min-Entropy

Håstad et al. introduced the following computational analogue of min-entropy:

Definition 4.1 (Pseudo-min-entropy [HILL], informal)**.** A distribution X has *pseudo-min-entropy at least k* if there exists a distribution Y such that:

1. X is indistinguishable from Y.
2. $\mathrm{H}_\infty(Y) \geq k$, where $\mathrm{H}_\infty(\cdot)$ denotes min-entropy.

Pseudo-min-entropy is interesting because a distribution can have much higher pseudo-min-entropy than its min-entropy. Indeed, if $G : \{0,1\}^n \to \{0,1\}^m$ is a pseudorandom generator, then $G(U_n)$ has min-entropy at most n, but is indistinguishable from U_m (by definition) and hence has pseudo-min-entropy $m > n$.

A conditional version is known as *pseudo-avg-min-entropy*:

Definition 4.2 (Pseudo-avg-min-entropy [HLR], informal)**.** Let (X, B) be a joint distribution. We say that B has *pseudo-avg-min-entropy at least k given X* if there exists a distribution C jointly distributed with X such that

1. (X, B) is indistinguishable from (X, C).
2. $\widetilde{\mathrm{H}}_\infty(C \mid X)$, the *average min-entropy* of C given X, is at least k, where

$$\widetilde{\mathrm{H}}_\infty(C \mid X) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}\left[\frac{1}{2^{\mathrm{H}_\infty(C|_{X=x})}}\right]}\right) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}\left[\max_a \Pr[C=a|X=x]\right]}\right).$$

It can be shown that $\widetilde{\mathrm{H}}_\infty(C \mid X) \geq k$ iff for every (computationally unbounded) randomized predictor S, $\Pr[S(X) = C] \leq 2^{-k}$. Our result is a computational analogue of this equivalence:

Theorem 4.3 (Characterizing pseudo-avg-min-entropy, informal)**.** *Let (X, B) be a joint distribution where B takes values in a polynomial-sized set Σ. Then B has pseudo-avg-min-entropy at least k given X if and only if there is no probabilistic polynomial-time algorithm S such that $\Pr[S(X) = B] \geq 2^{-k} \pm n^{-\omega(1)}$.*

In other words, we show that pseudo-avg-min-entropy coincides with unpredictability entropy [HLR] for polynomial-sized alphabets. To provide some more intuition, we compare two previous results relating forms of computational randomness and unpredictability, both as special cases of Theorem 4.3:

1. Yao [Yao1] showed that if B is a single bit, then (X, B) is indistinguishable from (X, U_1) (i.e. B has pseudo-avg-min-entropy 1 given X) iff B cannot be predicted from X with probability noticeably more than $1/2$. This can be generalized to B taking values in a polynomial-sized alphabet Σ: $B \in \Sigma$ has pseudo-avg-min-entropy $k = \log|\Sigma|$ given X iff B cannot be predicted with probability noticeably greater

than $2^{-k} = 1/|\Sigma|$. Thus, Theorem 4.3 has been known to hold in the extreme case of maximal entropy ($k = \log|\Sigma|$).

2. The Hardcore Theorem of Impagliazzo [Imp] (and subsequent strengthenings [KS, Hol1, BHK]) can be interpreted (as done in [STV]) as saying that when B is a single bit, B cannot be predicted from X with probability greater than $1 - \delta$ iff "B is indistinguishable from a random bit on a 2δ fraction of the probability space (X, B)" (this fraction of the probability space is typically called the "hardcore measure"). The latter condition is equivalent to saying that (X, B) has pseudo-avg-min-entropy at least $\log(1/(1-\delta))$ given X (see discussions in Section 4.2 for details). Thus, Theorem 4.3 can be viewed as a generalization of the Hardcore Theorem to larger alphabets. We refer to Chapter 3 for more discussions on the Hardcore Theorem.

We note that the constraint that B takes values in a polynomial-sized set is essential for Theorem 4.3. If $f : \{0,1\}^n \to \{0,1\}^n$ is a one-way permutation and X is a uniformly random output, then it is very hard to predict $f^{-1}(X)$ given X, but the pseudo-avg-min-entropy of $f^{-1}(X)$ given X is negligible (since we can efficiently recognize $f^{-1}(X)$ given X).

For B that takes values exponentially large set, Goldreich and Levin [GL] showed that if B is very hard to predict from X (i.e. cannot be predicted with nonnegligible probability), then we can choose a random hash function H whose range is a polynomial-sized set Σ and it will hold that $H(B) \in \Sigma$ has pseudo-avg-min-entropy $\log|\Sigma|$ given X and H. While this is very useful and has many applications, it does not characterize the pseudo-avg-min-entropy of B itself (but rather a hash of it), requires a hash function that supports "local list-decoding," and again only talks about maximal entropy ($\log|\Sigma|$).

4.1.2 Characterizing (Conditional) Pseudoentropy

Håstad et al. also introduced the following computational analogue of Shannon entropy:

Definition 4.4 (Pseudoentropy [HILL], informal)**.** A distribution X has *pseudoentropy at least k* if there exists a distribution Y such that:

1. X is indistinguishable from Y.
2. $\mathrm{H}_{\mathrm{sh}}(Y) \geq k$, where $\mathrm{H}_{\mathrm{sh}}(\cdot)$ denotes Shannon entropy.

Pseudoentropy is interesting because a distribution can have much higher pseudoentropy than its min-entropy. As in the case of pseudo-min-entropy, a canonical example is the output distribution of a pseudorandom generator.

A useful, average-case generalization is the notion of *conditional pseudoentropy*:

Definition 4.5 (Conditional pseudoentropy [HRV], informal)**.** Let (X, B) be a joint distribution. We say that B has *(conditional) pseudoentropy at least k given X* if there exists a distribution C jointly distributed with X such that

1. (X, B) is indistinguishable from (X, C).
2. $\mathrm{H}_{\mathrm{sh}}(C|X) \geq k$.

Note that if B has pseudoentropy at least k given X, then (X, B) has pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(X) + k$, but the converse is false (consider X that has pseudoentropy $\mathrm{H}_{\mathrm{sh}}(X) + k$ on its own, with a B that has no pseudoentropy).

Conditional pseudoentropy is useful because it captures the pseudoentropy from the perspective of an adversary who first sees X and later B, instead of both X and B at once. Thus, the sum of the pseudoentropy of X and the pseudoentropy of B given X can be larger than the pseudoentropy of the joint distribution (X, B).

We give an exact characterization of (conditional) pseudoentropy, which bears a lot of similarity to the characterization of pseudo-avg-min-entropy (Theorem 4.3). Unlike Theorem 4.3, our result here refers to "hardness of sampling" rather than unpredictability:

Theorem 4.6 (Characterizing conditional pseudoentropy, informal)**.** *Let (X, B) be a joint distribution where B takes values in a polynomial-sized set. Then B has pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X)+\delta$ given X if and only if there is no probabilistic polynomial-time algorithm S such that the KL divergence from (X, B) to $(X, S(X))$ is at most δ.*

A nice feature of Theorem 4.6 compared to Theorem 4.3 is that it focuses on the *computational* hardness in B given X, as measured by the pseudoentropy gap δ. For example, suppose that B is a uniform random bit, independent of X. Then B has 1 bit of pseudo-avg-min-entropy given X and cannot be predicted from X with probability better than randomly guessing, but these are not for computational reasons (i.e. they also hold for computationally unbounded algorithms). It is often desirable to focus solely on the *computational* randomness in B. For pseudoentropy, we can do this by subtracting $\mathrm{H}_{\mathrm{sh}}(B|X)$ (from the pseudoentropy of B). For unpredictability, we can do this by considering the feasibility of *sampling* the distribution $B|_{X=x}$ given a sample $x \leftarrow X$. Thus, in the example that B is a random bit independent of X, this sampling is easy to do (in contrast to the task of predicting B from X).

The constraint that B takes values in a polynomial-sized set is essential here, for the same reason as Theorem 4.3. However, we do have an alternative version of our result that holds for B taking values in an exponentially large range. In that version, we replace the task of sampling a distribution $S(X)$ from X with that of computing a "measure" that, when normalized to be a distribution, has small KL divergence from (X, B). In particular, this alternative formulation is interesting even when X is empty and gives a characterization of pseudoentropy of an arbitrary distribution B:

Theorem 4.7 (Characterizing pseudoentropy, informal (nonuniform setting only))**.** *Let (X, B) be a joint distribution. Then B has pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X if and only if there is no polynomial-sized circuit P that computes a conditional measure for a joint distribution (X, C) such that the KL divergence from (X, B) to (X, C) is at most δ.*

In Chapter 5, we use Theorem 4.6 to obtain simplified and more efficient constructions of pseudorandom generators from one-way functions.

We also establish *uniform* versions of Theorem 4.3 and 4.6, namely with respect to probabilistic polynomial-time algorithms S.

4.1.3 Our Techniques

We give a high-level view of proof techniques for the more interesting direction of Theorem 4.3 and 4.6: unpredictability (or hardness of sampling) implies pseudo-avg-min-entropy (or conditional pseudoentropy). First, let us assume the nonuniform model of computation (i.e. boolean circuits).

1. In light of relation between Theorem 4.3 and the Hardcore Theorem, it is natural that our proofs also begin by applying the Min-Max Theorem, following Nisan and Holenstein's proofs of the Hardcore Theorem [Imp, Hol1]. Suppose for contradiction that B does not have high pseudo-avg-min-entropy (or pseudoentropy) given X. That is, for every joint distribution (X, C) where C has high pseudo-avg-min-entropy (or pseudoentropy) given X, there is a poly-sized boolean circuit W that achieves distinguishing advantage $\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)] \geq \epsilon$, for $\epsilon = 1/\mathrm{poly}(n)$. By the Nonuniform Min-Max Theorem (Chapter 2 Theorem 2.3), there is a single poly-sized boolean circuit W^* that achieves $\mathbb{E}[W^*(X, B)] - \mathbb{E}[W^*(X, C)] \geq \Omega(\epsilon)$ for *all* (X, C) where C has high pseudo-avg-min-entropy (or pseudoentropy).

2. Let (X, C^*) be the distribution that maximizes $\mathbb{E}[W^*(X, C)]$ among all C that have high pseudo-avg-min-entropy (or pseudoentropy) given X. Thus of all such (X, C), (X, C^*) is the hardest to distinguish from (X, B) for W^*. We show that C^* can be "represented" efficiently using the poly-sized circuit for W^*, where the meaning of "represent" differs for pseudo-avg-min-entropy and conditional pseudoentropy.

3. From W^* we construct some circuit S such that S's performance, i.e. $\Pr[S(X) = B]$ (or $\mathrm{KL}(X, B \parallel X, S(X))$), is expressed in terms of the distinguishing advantage $\mathbb{E}[W^*(X, B)] - \mathbb{E}[W^*(X, C^*)]$. We then plug in the fact $\mathbb{E}[W^*(X, B)] - \mathbb{E}[W^*(X, C^*)] \geq \Omega(\epsilon)$ to conclude $\Pr[S(X) = B] \geq 2^{-k}$ (or $\mathrm{KL}(X, B \parallel X, S(X)) \leq \delta$). Thus we violate the assumption that B is unpredictable (or infeasible to sample within δ KL divergence) given X.

For pseudo-avg-min-entropy, the S we construct in Step 3 is a generalization of the predictor constructed in Holenstein's proofs of the Hardcore Theorem to larger alphabets. Nonetheless, even when B is a single bit, our proof of Theorem 4.3 differs notably from existing approaches to the Hardcore Theorem. In particular, by looking at the hardest-to-distinguish (X, C^*) in Step 2, we achieve optimal "hardcore density" without relying on the technical lemma of Holenstein (see Chapter 3, Remark 3.5).

For conditional pseudoentropy, the S we construct in Step 3 is simply so that $S(X) = C^*$. To prove such S achieves $\mathrm{KL}(X, B \parallel X, S(X)) \leq \delta$, we develop a generic framework that potentially applies to *any* "sufficiently nice", concave function H, not just the Shannon entropy function. The generic framework relates "pseudo-H" (e.g. pseudoentropy when H is Shannon entropy) of a distribution B to the infeasibility of sampling a distribution close to B where closeness is measured by the "Bregman divergence" associated with H (cf. Definition 1.7). This generic framework, combined with Step 1 and 2, gives rise to a

meta characterization theorem:

Informal Theorem 4.8 (Meta characterization theorem)**.** *Let (X, B) be a joint distribution where B takes value in a polynomial-sized set. For all "sufficiently nice," strictly concave functions H (e.g. H is Shannon entropy), the following are equivalent:*

1. *There exists a joint distribution (X, C) such that (X, C) and (X, B) are indistinguishable for poly-sized circuits, and $H(C|X) \geq H(B|X) + \delta - 1/n^{\omega(1)}$;*
2. *For all polynomial-sized circuits S, we have $\mathrm{D}_H(X, B \parallel X, S(X)) > \delta - 1/n^{\omega(1)}$.*

Theorem 4.6 is an instantiation of the meta theorem, since if $H = \mathrm{H}_{\mathrm{sh}}$ then $\mathrm{D}_H = \mathrm{KL}$ (see Definition 1.9). We do not prove the meta theorem, but rather describe more concretely in Section 4.3.3 how it follows as an immediate corollary of a few underlying lemmas, where the exact requirement of "suffciently nice" is made clear.

Uniform Settings. In the uniform model of computation, we replace the use of Nonuniform Min-Max Theorem in Step 1 by the Uniform Min-Max Theorem in Chapter 2. In order to apply the Uniform Min-Max Theorem, we develop efficient algorithms to approximately compute KL projections on the set of all distributions with high average min-entropy, and on the set of all distributions with high conditional Shannon-entropy.

4.1.4 Relation to Inaccessible Entropy

A variety of computational notions of entropy have been studied in the cryptography and complexity literature, e.g. [Yao1, HILL, BSW, HLR, HRVW, HRV, HHR$^+$3, FR, Rey]. In addition to the notions discussed above, our work was also inspired by the works on *inaccessible entropy* [HRVW, HHR$^+$3].

Like our characterization of conditional pseudoentropy, inaccessible entropy refers to a difficulty of sampling a distribution B from a jointly distributed X. However, there are

important differences. In our characterization (Theorem 4.6), the sample of X is generated externally and fed to the adversary, who tries then to sample the conditional distribution $B|X$. In the [HHR$^+$3] notion of inaccessible entropy, the adversary is also given the random coins used to generate X, and we compare its output distribution conditioned on those coins to $B|X$. And in the original notion of inaccessible entropy, from [HRVW], the adversary is the one who generates X (or some approximation to it). These three notions are analogous to the security conditions for one-way functions, target collision-resistant hash functions (i.e. UOWHFs), and collision-resistant hash functions, respectively (thinking of $X = f(B)$ for $B \in_R \{0,1\}^n$). We note that the hardness of sampling we consider also differs from inaccessible entropy in the way it measures how well an adversary approximates the conditional distribution $B|X$. Roughly speaking, in our notion (measuring the KL divergence from $B|X$ to the adversary's output), the adversary's goal is to produce an output distribution that *contains* $B|X$ as tightly as possible. In the notions of inaccessible entropy, the adversary's goal is to produce an output distribution that is *contained within* $B|X$ as tightly as possible.

There is also significant similarity between how one-way functions can be used to generate inaccessible entropy [HRVW], and conditional pseudoentropy (see Chapter 5, Section 5.3). In [HRVW], it is shown that if f is a one-way function, then $(f(U_n), U_n)$ is a next-bit inaccessible entropy generator, just like we show that it is a next-bit pseudoentropy generator in Chapter 5, Theorem 5.5. However, for inaccessible entropy, it is only necessary to break $f(U_n)$ into bits (U_n can be treated as a single block), and for pseudoentropy it is only necessary to break U_n into bits ($f(U_n)$ can be treated as a single block). Nevertheless, there are enough similarities to suggest that there may be a deeper connection between inaccessible entropy and pseudoentropy; trying to formalize this connection is an interesting question for future work.

4.2 Characterizing Pseudo-Avg-Min-Entropy

4.2.1 Definitions

The conditional version of min-entropy we consider is defined as follows:

Definition 4.9 (Average min-entropy [DORS])**.** For every joint distribution (X, B), the *average min-entropy of B given X* is defined to be

$$\widetilde{\mathrm{H}}_\infty(B \mid X) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}\left[\frac{1}{2^{\mathrm{H}_\infty(B|_{X=x})}}\right]}\right) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}\left[\max_{a\in\mathrm{supp}(B)} B(a|x)\right]}\right).$$

We remark that there are other ways to define conditional entropies, but the above definition has turned out to be the most convenient, and has an unpredictability interpretation. That is, $\widetilde{\mathrm{H}}_\infty(B \mid X) \geq k$ if and only if it is impossible to predict B from X with probability more than 2^{-k}:

Proposition 4.10. *For every joint distribution (X, B),*

$$\mathrm{H}_\infty(B|X) \geq k \iff \forall \textit{(randomized) } S, \Pr[S(X) = B] \leq 2^{-k}.$$

Proof. To maximize $\Pr[S(X) = B]$, S should output the most probable value of $B|_{X=x}$, thus achieves $\Pr[S(X) = B] = \mathbb{E}_{x\sim X}\left[\max_a \Pr[B = a|X = x]\right] = 2^{-\mathrm{H}_\infty(B|X)}$. □

A computational analogue of average min-entropy is *pseudo-avg-min-entropy*, introduced by Hsiao, Lu, and Reyzin [HLR] (for the nonuniform setting). We begin with the nonuniform definition because it is simpler:

Definition 4.11 (Pseudo-avg-min-entropy, nonuniform setting)**.** Let (X, B) be a joint distribution. We say B has (T, ϵ) *nonuniform pseudo-avg-min-entropy at least k given X* if there exists a random variable C jointly distributed with X such that the following holds:

- $\widetilde{\mathrm{H}}_\infty(C|X) \geq k$;

- (X, B) and (X, C) are ϵ-indistinguishable by all size T circuits.

If $(X, B) = (X, B)(n)$ for a security parameter n, we say B has *pseudo-avg-min-entropy at least $k = k(n)$ given X* if for every constant c, B has $(n^c, 1/n^c)$ pseudo-avg-min-entropy at least k given X for all sufficiently large n.

In the uniform setting, where we consider randomized algorithms instead of circuits as the distinguishers, the right definitions are more subtle. It turns out that we must require indistinguishability even against algorithms equipped with an sampling oracle. (See remark below for more discussion.)

Notation. For a distribution Z, let O_Z denote the oracle that gives a random sample from Z when queried.

Definition 4.12 (Pseudo-avg-min-entropy, uniform setting)**.** Let n be a security parameter, $T = T(n)$, $\epsilon = \epsilon(n)$, $k = k(n)$, $\ell = \ell(n)$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B has (T, ϵ) *uniform pseudo-avg-min-entropy at least k given X* if for every randomized oracle algorithm A computable in time T, there is a distribution C jointly distributed with X such that the following holds for all sufficiently large n:

- $\widetilde{\mathrm{H}}_\infty(C|X) \geq k$;
- (X, B) and (X, C) are indistinguishable by $A^{O_{X,B,C}}$:

$$|\Pr[A^{O_{X,B,C}}(X, B) = 1] - \Pr[A^{O_{X,B,C}}(X, C) = 1]| < \epsilon.$$

We say B has *uniform pseudo-avg-min-entropy at least $k = k(n)$ given X* if for every constant c, B has $(n^c, 1/n^c)$ uniform pseudo-avg-min-entropy at least k given X.

The reason to give the distinguishers oracle access to $O_{X,B,C}$ is to ensure that the definition composes: if (X_1, B_1) and (X_2, B_2) are iid copies of (X, B), we would like to

say that (B_1, B_2) has pseudo-avg-min-entropy at least $2k$ given (X_1, X_2). Indeed we would want to say that (X_1, B_1, X_2, B_2) is indistinguishable from (X_1, C_1, X_2, C_2) where C_1, C_2 are iid copies of C. However, indistinguishability against uniform algorithms is not preserved under taking multiple independent samples in general [GM1]. Requiring indistinguishability against distinguishers with oracle access to $O_{X,B,C}$ ensures that indistinguishability will be preserved under taking multiple independent samples.

However, a consequence of our results is that the definition with oracle $O_{X,B,C}$ is equivalent to the definition with oracle $O_{X,B}$ provided B comes from a polynomial-sized alphabet. In particular, if (X, B) is also polynomial-time samplable, the definition is equivalent to one without oracle $O_{X,B,C}$. (See Corollary 4.22.)

In the definition of pseudo-avg-min-entropy, a question asked by Leo Reyzin is whether allowing changing *both* X and B (rather than changing (X, B) to (X, C), with X fixed) makes any difference. Another consequence of our results is that this is equivalent to the above definition. (See Corollary 4.22.)

For a joint distribution (X, B), it is a basic complexity-theoretic question how well B can be efficiently predicted given X:

Definition 4.13 (Hardness of prediction, nonuniform setting)**.** Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B is *nonuniformly (t, δ)-hard to predict given X* if for all size t circuits S it holds that $\Pr[S(X) = B] < 1 - \delta$.

We say B is *nonuniformly δ-hard to predict given X* if for every constant c, B is nonuniformly $(n^c, \delta - 1/n^c)$-hard to predict given X for all sufficiently large n.

Note that the (nonuniform) hardness of prediction generalizes the average-case hardness of a function $f : \{0,1\}^n \to \{0,1\}^\ell$ by taking $X = U_n$, $B = f(X)$, and is equivalent to *unpredictability entropy* studied by Hsiao, Lu, and Reyzin [HLR].

We can also define hardness of prediction with respect to uniform algorithms. Note that we give the predictor oracle access to the sampling oracle $O_{X,B}$ (which is redundant in case (X, B) is efficiently samplable):

Definition 4.14 (Hardness of prediction, uniform setting)**.** Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $\ell = \ell(n)$. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B is *uniformly (t, δ)-hard to predict given X* if for all time t randomized oracle algorithms S and all sufficiently large n, $\Pr[S^{O_{X,B}}(X) = B] < 1 - \delta$.

We say B is *uniformly δ-hard to predict given X* if for every constant c, B is uniformly $(n^c, \delta - 1/n^c)$-*hard to predict* given X.

4.2.2 Main Results

For a joint distribution (X, B) on $\{0,1\}^n \times \{0,1\}^\ell$ where $\ell = O(\log n)$, we give a characterization of the *pseudo-avg-min-entropy of B given X*, in terms of the hardness of predicting B given X:

Theorem 4.15 (Characterizing pseudo-avg-min-entropy)**.** *Let n be a security parameter, $\ell = \ell(n) = O(\log n)$, $r = r(n) \leq \ell(n)$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. Then B has (non)uniform pseudo-avg-min-entropy at least r given X if and only if B is (non)uniformly $(1 - 2^{-r})$-hard to predict given X.*

Note that this is an computational analogue of Proposition 4.10, which proves an equivalence between average min-entropy and unpredictability for computationally unbounded algorithms.

Relation to the Hardcore Theorem (Theorem 3.2 and 3.6). Theorem 4.15 can be viewed as a generalization of Impagliazzo's Hardcore Theorem [Imp] for functions that are not necessarily binary. Indeed, versions of the Hardcore Theorem [KS, Hol1] that

achieve optimal "hardcore size" are equivalent to Theorem 4.15 with $\ell = 1$. (For formal statement and discussions of the Hardcore Theorem, we refer to Chapter 3.) We consider the nonuniform setting for simplicity.

The proof that the Hardcore Theorem implies Theorem 4.15 with $\ell = 1$, $X = U_n$, and $B = f(X)$, follows an argument of [STV]. Suppose B is δ-hard to predict given X. By the Hardcore Theorem there is a 2δ-dense hardcore distribution on $\{0,1\}^n$, hence a hardcore set $H \subseteq \{0,1\}^n$ of size roughly $2\delta \cdot 2^n$, which can be formed by taking random samples from the hardcore distribution. Let (X, C) be the joint distribution such that given $X = x$, either C is a uniform random bit if $x \in H$, or $C = f(x)$ if $x \notin H$. Then C has average min-entropy at least $\log(1/(1-\delta))$ given X, and (X, C) is indistinguishable from (X, B) both inside H (since C is a uniform random bit and B is extremely hard [Yao1]) and outside H (since C equals B). Thus B has pseudo-avg-min-entropy at least $\log(1/(1-\delta))$.

To see that Theorem 4.15 implies the Hardcore Theorem, take $X = U_n$ and $B = f(X)$ where f is the δ-hard function (or, for the slightly generalized version of Chapter 3 Theorem 3.2, take (X, B) to be (X, B)). By Theorem 4.15 there is some joint distribution (X, C) indistinguishable from (X, B) such that C has average min-entropy at least $\log(1/(1-\delta))$ given X. Moreover, we can w.l.o.g. assume that $(x, a) \mapsto C(a|x)$ is computable by a poly-sized circuit, by the Regularity Theorem for circuit complexity — average case (Chapter 2 Theorem 2.12) setting $\mathcal{V}$ to be the set of all joint distributions (X, C') on $\{0,1\}^n \times \{0,1\}$ where C' has average min-entropy at least $\log(1/(1-\delta))$ given X (such $\mathcal{V}$ is KL-projectable; see Section 4.2.4.1).

Define a probabilistic function

$$T(x,a) = \begin{cases} 1, & \text{w.p. } \frac{\min\{C(0|x), C(1|x)\}}{C(a|x)} \\ 0, & \text{otherwise} \end{cases}.$$

Note that T is defined such that given $T(X,C)=1$, C is a uniform random bit, and

$$\begin{aligned}\mathbb{E}[T(X,C)] &= \mathop{\mathbb{E}}_{x\sim X}\left[\sum_{a\in\{0,1\}} C(a|x)\frac{\min\{C(0|x),C(1|x)\}}{C(a|x)}\right] \\ &= 2\cdot \mathop{\mathbb{E}}_{x\sim X}[\min\{C(0|x),C(1|x)\}] \\ &\geq 2\delta. \qquad \text{(by average min-entropy of } C\text{)}\end{aligned}$$

We claim that $(X,B)|_{T(X,B)=1}$ is a 2δ-dense hardcore distribution. Note that $(X,B)|_{T(X,B)=1}$ is 2δ-dense in (X,B) because $\mathbb{E}[T(X,B)] \approx \mathbb{E}[T(X,C)] \geq 2\delta$, by efficiency of T and indistinguishability. For hardcore-ness, consider any poly-sized circuit P. Define W such that $W(x,a)$ outputs 1 iff $P(x)=a$ and $T(x,a)=1$. Thus

$$\begin{aligned}\mathbb{E}\left[W(X,B)\right] &= \Pr[P(X)=B \wedge T(X,B)=1] \\ &= \Pr\left[T(X,B)=1\right]\cdot\Pr[P(X)=B|T(X,B)=1] \\ &\approx \Pr\left[T(X,C)=1\right]\cdot\Pr[P(X)=B|T(X,B)=1]\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[W(X,C)\right] &= \Pr[P(X)=C \wedge T(X,C)=1] \\ &= \Pr\left[T(X,C)=1\right]\cdot\Pr[P(X)=C|T(X,C)=1] \\ &= \Pr\left[T(X,C)=1\right]\cdot\frac{1}{2}\end{aligned}$$

where the last equality is because given $T(X,C)=1$, C is a uniform random bit. By efficiency of W and indistinguishability, $\mathbb{E}[W(X,B)]-\mathbb{E}[W(X,C)]<\epsilon$, which implies

$$\Pr[P(X)=B|T(X,B)=1] < \frac{1}{2}+\frac{\epsilon}{\mathbb{E}\left[T(X,C)\right]} < \frac{1+\epsilon/\delta}{2}.$$

Thus $(X,B)|_{T(X,B)=1}$ is a 2δ-dense hardcore distribution.

4.2.3 Hardness of Prediction Implies Pseudo-Avg-Min-Entropy, Nonuniform Setting

We begin with an outline of the proof for the nonuniform setting. Suppose for contradiction that B does not have high pseudo-avg-min-entropy given X. In other words, for every (X, C) where C has high average min-entropy given X, there is a small circuit W distinguishing (X, B) from (X, C). By the Nonuniform Min-Max Theorem (Chapter 2, Theorem 2.3), there exists a universal distinguisher W^* of small circuit size, i.e. W^* distinguishes (X, B) from *all* (X, C) where C has high average min-entropy given X. It turns out that, from such W^* we can construct a predictor P where $\Pr[P(X) = B] \geq 2^{-r} - \epsilon$, contradicting the hardness of predicting B from X.

This last step is stated as Lemma 4.16 below. Note that Lemma 4.16 even provides an efficient algorithm N for converting W^* to the predictor P; such uniformity is an overkill in the nonuniform setting, but will be needed for the uniform setting (Section 4.2.4).

The following notations are used throughout this and the next section.

Notation. For a function $W : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$, we let $W(x, \#i)$ denote the ith largest element of the multiset $\{W(x,a) : a \in \{0,1\}^\ell\}$, i.e. $W(x, \#1) \geq \cdots \geq W(x, \#2^\ell)$, breaking ties arbitrarily. For $\kappa = 1, \ldots, 2^\ell$ let $\Delta_W(x, \kappa) = \sum_{i=1}^{\kappa} (W(x, \#i) - W(x, \#\kappa))$, thus $0 = \Delta(x, 1) \leq \cdots \leq \Delta(x, 2^\ell)$. We denote by $C(\#i|x)$ the probability of the ith heaviest element of $C|_{X=x}$, breaking ties arbitrarily, so that $C(\#1|x) \geq C(\#2|x) \geq \ldots$.

Lemma 4.16. *There exists a randomized oracle algorithm N such that the following holds. Let $\epsilon = \epsilon(n) > 0$, $\gamma = \gamma(n) > 0$, $\ell = \ell(n)$, and $0 < r = r(n) \leq \ell$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$ such that $\Pr[X = x] \leq \epsilon$ for all x. Let $W : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ be a function whose output has bit length $\tau = \tau(n)$, and such that $\Delta_W(x, \kappa)$ is distinct for each pair of $x \in \{0,1\}^n$ and $\kappa \in [2^\ell]$. Then w.p. at least $1 - \gamma$,*

$N^{O_X,W}(2^{-r},\epsilon,\gamma)$ outputs a randomized oracle circuit P satisfying

$$\Pr[P^W(X)=B] \geq 2^{-r} - \epsilon + \frac{\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)]}{\lambda}$$

where $\lambda \in (0, 2^\ell]$ and (X,C) is a joint distribution with $\widetilde{\mathrm{H}}_\infty(C|X) \geq r$. Moreover, *$N$ runs in time* $\mathrm{poly}(n, 2^\ell, \tau, 1/\epsilon, \log(1/\gamma))$*, and P is of size* $2^\ell \cdot n \cdot \log(1/\epsilon) \cdot \mathrm{poly}(\tau,\ell)$ *making* 2^ℓ *oracle queries.*

Proof. Define $\kappa_x(\lambda) = \max\{\kappa : \Delta_W(x,\kappa) \leq \lambda\} \in [2^\ell]$, thus

$$\Delta_W(x,\kappa_x(\lambda)) \leq \lambda < \Delta_W(x,\kappa_x(\lambda)+1) \tag{4.1}$$

where for the second inequality we assume $\kappa_x(\lambda)+1 \leq 2^\ell$. Note that $\kappa_x(\lambda)$ is increasing in λ.

Note that $W(x,\#1)$, ..., $W(x,\#2^\ell)$ are distinct, by distinctness of $\Delta_W(x,\cdot)$. We define the following function P_λ parameterized by $\lambda \in (0,2^\ell]$:

$$\Pr[P_\lambda(x)=i] = \begin{cases} \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\left(W(x,\#i) - \frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right), & 1 \leq i \leq \kappa_x(\lambda) \\ 0, & i \geq \kappa_x(\lambda)+1 \end{cases}.$$

We let $\hat{P}_\lambda$ be the predictor that computes $i \leftarrow P_\lambda(x)$, and outputs the string $a \in \{0,1\}^\ell$ such that $W(x,a) = W(x,\#i)$.

We check that P_λ is well-defined, i.e. satisfies (i) $\sum_{i=1}^{2^\ell} \Pr[P_\lambda(x)=i] = 1$; (ii) $\Pr[P_\lambda(x)=i] \geq 0$. We verify (i) by inspection. For (ii), note that for all $1 \leq i \leq \kappa_x(\lambda)$,

$$\begin{aligned}
&\Pr[P_\lambda(x)=i] \\
&\geq \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\left(W(x,\#\kappa_x(\lambda)) - \frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right) \quad (\text{since } W(x,\#\kappa_x(\lambda)) \leq W(x,\#i)) \\
&= \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\cdot\frac{-\Delta_W(x,\kappa_x(\lambda))}{\kappa_x(\lambda)} \\
&\geq \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\cdot\frac{-\lambda}{\kappa_x(\lambda)} = 0 \qquad (\text{by } 4.1)
\end{aligned}$$

We next show that $\hat{P}_\lambda$'s prediction probability $\hat{P}_\lambda(X) = B$ can be expressed in terms of W's distinguishing advantage $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)]$, for some C with high average min-entropy (that depends on λ).

Claim 4.17. For every $\lambda \in (0, 2^\ell]$ there exists a joint distribution (X,C) satisfying $\widetilde{\mathrm{H}}_\infty(C|X) \geq \log\left(1/\mathbb{E}_{x\sim X}\left[1/\kappa_x(\lambda)\right]\right)$ and

$$\Pr[\hat{P}_\lambda(X) = B] \geq \mathop{\mathbb{E}}_{x\sim X}\left[1/\kappa_x(\lambda)\right] + \frac{\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)]}{\lambda}.$$

Proof of Claim. Consider any x. For all $i \geq \kappa_x(\lambda) + 1$, we have

$$\begin{aligned}
&\Pr[P_\lambda(x) = i] = 0 \\
&= \frac{1}{\kappa_x(\lambda)} + \frac{-\lambda}{\lambda \cdot \kappa_x(\lambda)} \\
&\geq \frac{1}{\kappa_x(\lambda)} + \frac{-\Delta_W(x, \kappa_x(\lambda)+1)}{\lambda \cdot \kappa_x(\lambda)} && \text{(by 4.1)} \\
&= \frac{1}{\kappa_x(\lambda)} + \frac{\sum_{j=1}^{\kappa_x(\lambda)+1}\left(W(x,\#\kappa_x(\lambda)+1) - W(x,\#j)\right)}{\lambda \cdot \kappa_x(\lambda)} \\
&\geq \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda} \cdot \frac{\sum_{j=1}^{\kappa_x(\lambda)}\left(W(x,\#i) - W(x,\#j)\right)}{\kappa_x(\lambda)} && \text{(as } W(x,\#i) \leq W(x,\#\kappa_x(\lambda)+1)\text{)} \\
&= \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\left(W(x,\#i) - \frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right).
\end{aligned}$$

Thus for all i,

$$\Pr[P_\lambda(x) = i] \geq \frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\left(W(x,\#i) - \frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right) \tag{4.2}$$

(in the case of $1 \le i \le \kappa_x(\lambda)$, it follows from the definition of P_λ). Therefore,

$$\begin{aligned}
&\Pr[\hat{P}_\lambda(X) = B] \\
&= \mathop{\mathbb{E}}_{x\sim X}\left[\sum_i B(\#i|x)\Pr[P_\lambda(x) = i]\right] \\
&\ge \mathop{\mathbb{E}}_{x\sim X}\left[\frac{1}{\kappa_x(\lambda)} + \frac{1}{\lambda}\left(\sum_i B(\#i|x)W(x,\#i) - \frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right)\right] \qquad \text{(by 4.2)} \\
&= \mathop{\mathbb{E}}_{x\sim X}\left[\frac{1}{\kappa_x(\lambda)}\right] + \frac{\mathbb{E}[W(X,B)] - \mathbb{E}_{x\sim X}\left[\frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right]}{\lambda}.
\end{aligned}$$

To complete the proof, note that $\mathbb{E}_{x\sim X}\left[\frac{\sum_{j=1}^{\kappa_x(\lambda)} W(x,\#j)}{\kappa_x(\lambda)}\right] = \mathbb{E}[W(X,C)]$ for the following distribution C:

$$C(a|x) = \begin{cases} \frac{1}{\kappa_x(\lambda)}, & 1 \le j \le \kappa_x(\lambda) \\ 0, & j \ge \kappa_x(\lambda)+1 \end{cases}$$

where j is the number such that $W(x,a) = W(x,\#j)$. □

The algorithm. Given Claim 4.17, our algorithm works as follows:

1. Take a multiset T of $m = O(\left(1/\epsilon^2\right)(\log(\tau+\ell) + \log(1/\gamma)))$ random samples of $x \sim X$.
2. Search for the least $\lambda \in (0, 2^\ell]$ such that $E_\lambda \le 2^{-r} - .1\epsilon$ and λ is a multiple of $2^{-\tau}$, where $E_\lambda = \mathbb{E}_{x\in_R T}[1/\kappa_x(\lambda)]$ is an estimate of $\mathbb{E}_{x\sim X}[1/\kappa_x(\lambda)]$. This can be done by a $(\ell+\tau)$-round binary search, since $\mathbb{E}_{x\sim X}[1/\kappa_x(\lambda)]$ is decreasing in λ.
3. Let λ^* be the λ found in Step 2. Output an oracle circuit P^W computing $\hat{P}_{\lambda^*}$. (Specifically, P^W first computes $\Pr[\hat{P}_{\lambda^*}(x) = a]$ for each a, then samples $\hat{P}_{\lambda^*}(x)$ w.p. at least $1 - .1\epsilon$.)

The running time is $\mathrm{poly}(n, 2^\ell, \tau, 1/\epsilon, \log(1/\gamma))$, and the circuit P is of size $2^\ell \cdot n \cdot \log(1/\epsilon) \cdot \mathrm{poly}(\tau, \ell)$ with at most 2^ℓ queries.

Correctness. We assume that for all the $(\tau + \ell)$ values of λ examined during Step 2's binary search,

$$\left| E_\lambda - \mathop{\mathbb{E}}_{x \sim X} \left[\frac{1}{\kappa_x(\lambda)} \right] \right| \leq .1\epsilon.$$

This holds with all but at most $(\tau + \ell) \cdot 2^{-\Omega(m\epsilon^2)} \leq \gamma/2$ probability, by a Chernoff bound and union bound. According to Step 2 of the algorithm, this implies

$$\mathop{\mathbb{E}}_{x \sim X} \left[\frac{1}{\kappa_x(\lambda^*)} \right] \leq E_{\lambda^*} + .1\epsilon \leq 2^{-r},$$

as well as

$$\mathop{\mathbb{E}}_{x \sim X} \left[\frac{1}{\kappa_x(\lambda^* - 2^{-\tau})} \right] \geq E_{\lambda^* - 2^{-\tau}} - .1\epsilon \geq 2^{-r} - .2\epsilon.$$

Furthermore, since $\Delta_W(\cdot, \cdot)$ are all distinct multiples of $2^{-\tau}$, $\kappa_x(\lambda^*)$ and $\kappa_x(\lambda^* - 2^{-\tau})$ must be identical for all x except for at most one value $x = z$ where $\kappa_z(\lambda^*) - 1 = \kappa_z(\lambda^* - 2^{-\tau})$. That is,

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \sim X} \left[\frac{1}{\kappa_x(\lambda^*)} \right] &\geq \mathop{\mathbb{E}}_{x \sim X} \left[\frac{1}{\kappa_x(\lambda^* - 2^{-\tau})} \right] - \Pr[X = z] \cdot \left(\frac{1}{\kappa_z(\lambda^*) - 1} - \frac{1}{\kappa_z(\lambda^*)} \right) \\
&\geq 2^{-r} - .2\epsilon - \Pr[X = z] \cdot \frac{1}{2} \\
&\geq 2^{-r} - .7\epsilon
\end{aligned}$$

where we use the assumption that $\Pr[X = x] \leq \epsilon$ for all x.

Applying Claim 4.17 with $\lambda = \lambda^*$ yields

$$\begin{aligned}
\Pr[\hat{P}_{\lambda^*}(X) = B] &\geq \mathop{\mathbb{E}}_{x \sim X} \left[\frac{1}{\kappa_x(\lambda^*)} \right] + \frac{\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]}{\lambda^*} \\
&\geq 2^{-r} - .7\epsilon + \frac{\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]}{\lambda^*}
\end{aligned}$$

for some joint distribution (X, C) satisfying $\widetilde{\mathrm{H}}_\infty(C|X) \geq \log\left(1/\mathop{\mathbb{E}}_{x \sim X}\left[1/\kappa_x(\lambda^*)\right]\right) \geq r$. Therefore

$$\Pr[P^W(X) = B] \geq \Pr[\hat{P}_{\lambda^*}(X) = B] - .1\epsilon \geq 2^{-r} - \epsilon + \frac{\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]}{\lambda^*}$$

where the $.1\epsilon$ additional loss is due to the sampling by P. □

Theorem 4.18 (Hardness of prediction $\Longrightarrow$ pseudo-avg-min-entropy, nonuniform setting)**.** *Let $\epsilon > 0$, and (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$ such that B is nonuniformly $(t, 1-2^{-r})$-hard to predict given X. Then B has nonuniform (t', ϵ) pseudo-avg-min-entropy at least r given X, for $t' = t/\left(2^\ell \cdot \mathrm{poly}(n, \ell, 1/\epsilon)\right)$.*

Proof. Suppose for contradiction that B does not have nonuniform (t', ϵ) pseudo-avg-min-entropy at least r given X. That is, for every joint distribution (X, C) with $\widetilde{\mathrm{H}}_\infty(C|X) \geq r$ there is a size t' deterministic circuit W with $\mathbb{E}\left[W(X,B)\right] - \mathbb{E}\left[W(X,C)\right] \geq \epsilon$. We will construct a size t randomized circuit P such that $\Pr[P(X) = B] \geq 2^{-r}$.

Consider the following two player zero-sum game. Player 1 picks a distribution (X, C) with $\widetilde{\mathrm{H}}_\infty(C|X) \geq r$. Player 2 picks a size t' deterministic circuit $W : \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}$, and receives expected payoff $\mathbb{E}\left[W(X,B)\right] - \mathbb{E}\left[W(X,C)\right]$. Thus, our assumption says that for every mixed strategy for Player 1, there is a strategy for Player 2 that achieves payoff at least ϵ. So, by the Nonuniform Min-Max Theorem (Chapter 2, Theorem 2.3), Player 2 has a mixed strategy, uniformly distributed over $S = O(n/\epsilon^2)$ size t' circuits, that achieves expected at least than $3\epsilon/4$ regardless of Player 1's move. Rephrasing, there is a size $O(St')$ deterministic circuit $W^* : \{0,1\}^n \times \{0,1\} \to [0,1]$ (which computes an average) such that $\mathbb{E}\left[W^*(X,B)\right] - \mathbb{E}\left[W^*(X,C)\right] \geq \epsilon/2$ for all (X, C) that satisfies $\widetilde{\mathrm{H}}_\infty(C|X) \geq r$. This implies $\mathbb{E}\left[W^*(X,B)\right] - \mathbb{E}\left[W^*(X,C)\right] \geq \epsilon/4$ for all (X, C) such that $\widetilde{\mathrm{H}}_\infty(C|X) \geq r'$ where $2^{-r'} = 2^{-r} + \epsilon/4$.

In order to apply Lemma 4.16 we make the following assumptions:

1. $\Pr[X = x] \leq \epsilon/4$ for all x. This is w.l.o.g. because we can pad X with $\log(1/\epsilon) + 2$ random bits, i.e. replace (X, B) by the new joint distribution $((X, U_{\log(1/\epsilon)+2}), B)$.

2. We "perturb" W^* into $\widetilde{W^*}$ so that

   (a) $\Delta_{\widetilde{W^*}}(x, \kappa)$ is distinct for each pair of $x \in \{0,1\}^n$ and $\kappa \in [2^\ell]$;

(b) $\mathbb{E}\left[\widetilde{W^*}(X,B)\right] - \mathbb{E}\left[\widetilde{W^*}(X,C)\right] \geq \mathbb{E}\left[W^*(X,B)\right] - \mathbb{E}\left[W^*(X,C)\right] - \epsilon/4$ for all (X,C).

The can be done by adding a negligible $\min\{S^{-1}, .1\epsilon\} \cdot \left(x \cdot i/2^{n+\ell}\right)$ to $W^*(x,i)$, where we interpret x as a natural number. The resulting $\widetilde{W^*}(\cdot,\cdot)$ has bit length at most $\tau = n + O(\log n) + \ell + O(1/\epsilon)$, and is of size $t'' = O(St') + O(\tau + \log(1/S)) = t' \cdot \mathrm{poly}(n, \ell, 1/\epsilon)$.

By Lemma 4.16, there is a randomized circuit P of size $2^\ell \cdot (n \cdot \log(1/\epsilon) \cdot \mathrm{poly}(\tau, \ell) + t'') = t' \cdot 2^\ell \cdot \mathrm{poly}(n, \ell, 1/\epsilon) \leq t$ such that

$$\begin{aligned}\Pr[P(X) = B] &\geq 2^{-r'} - \frac{\epsilon}{4} + \frac{\mathbb{E}[\widetilde{W^*}(X,B)] - \mathbb{E}[\widetilde{W^*}(X,C)]}{\lambda} \\ &\geq 2^{-r'} - \frac{\epsilon}{4} + \frac{\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] - \epsilon/4}{\lambda}\end{aligned}$$

where $\lambda \in (0, 2^\ell]$ and (X,C) is a joint distribution with $\widetilde{\mathrm{H}}_\infty(C|X) \geq r'$. Since $\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] \geq \epsilon/4$, we get

$$\Pr[P(X) = B] \geq 2^{-r'} - \frac{\epsilon}{4} \geq 2^{-r}$$

contradicting the hardness of predicting B given X. □

4.2.4 Hardness of Prediction Implies Pseudo-Avg-Min-Entropy, Uniform setting

Our proof in the uniform setting only differs in the use of the Uniform Min-Max Theorem (Chapter 2, Theorem 2.5). Since Player 1's strategies are the set of all joint distributions (X,C) where C has high pseudo-avg-min-entropy, we must be able to compute (approximate) KL projections on the set, to instantiate the underlying algorithm of the Uniform Min-Max Theorem.

4.2.4.1 Approximating KL Projection on High Average Min-Entropy Distributions

Notation. We denote by $\widetilde{\mathcal{V}}_r(X)$ the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ (where C may vary and X is fixed) such that $\widetilde{\mathrm{H}}_\infty(C|X) \geq r$.

We begin by characterizing the (exact) KL projection of an joint distribution $(X, C) \notin \widetilde{\mathcal{V}}_r(X)$ on the set $\widetilde{\mathcal{V}}_r(X)$ (Lemma 4.19). Using this characterization, we then show in Theorem 4.20 how to efficiently compute an approximate KL projection (X, C'), where (X, C) is represented by a circuit computing the probability vectors of $C|_{X=x}$, likewise for (X, C').

Lemma 4.19 (KL projection on high average min-entropy distributions)**.** *Let (X, C) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$ such that $\Pr[C = b|X = x] > 0$ for all $x \in \mathrm{supp}(X)$ and $b \in \{0,1\}^\ell$. Let $(X, C^*_{\{\delta_x\}})$ be the distribution parameterized by the constant $\delta_x \in [2^{-\ell}, C(\#1|x)]$ for each x, defined as follows: For all $x \in \mathrm{supp}(X)$ and $a \in \{0,1\}^\ell$,*

$$C^*_{\{\delta_x\}}(a|x) = \min\{\delta_x, \rho_x \cdot C(a|x)\}$$

where $\rho_x \geq 1$ is a scaling factor such that $\sum_a C^*_{\{\delta_x\}}(a|x) = 1$. Note that given any δ_x, there is a unique $\rho_x = \rho_x(\delta_x)$ that ensures $\sum_a C^*_{\{\delta_x\}}(a|x) = 1$. Then for every $r \leq \ell$, the KL projection of (X, C) on $\widetilde{\mathcal{V}}_r(X)$ is the distribution $(X, C^*_{\{\delta_x\}}(a|x))$ where the values $\{\delta_x\}$ are determined as follows.

Let $\kappa_x(\beta) \in \{1, \ldots, 2^\ell\}$ denote the largest number i such that $\rho_x(\delta) \cdot C(\#i|x) \geq \delta$, i.e. the number of elements that would be capped if we set $\delta_x = \delta$. Let $g_x : (2^{-\ell}, C(\#1|x)] \to (-\infty, 0]$ be the function

$$g_x(\delta) = \sum_{i=1}^{\kappa_x(\delta)} \log\left(\frac{\delta}{\rho_x(\delta) \cdot C(\#i|x)}\right).$$

Then for all $x \in \mathrm{supp}(X)$,

1. *g_x is strictly increasing with range*

$$\mathrm{Range}(g_x) = \left(\sum_{i=1}^{2^\ell} \log \frac{C(\#2^\ell|x)}{C(\#i|x)}, 0 \right].$$

2.

$$\delta_x = \delta_x(\lambda) = \begin{cases} g_x^{-1}(\lambda), & \lambda \in \mathrm{Range}(g_x) \\ 2^{-\ell}, & \text{otherwise} \end{cases}$$

where $\lambda \in (-\infty, 0]$ is a constant chosen such that $\mathbb{E}_{x\sim X}\left[\delta_x(\lambda)\right] = 2^{-r}$.

Proof. Assume w.l.o.g. that $\mathrm{supp}(X) = \{0,1\}^n$. Note that for every x,

$$\begin{aligned} g_x(\beta) &= \sum_{i=1}^{\kappa_x(\delta)} \log\left(\frac{\delta}{\rho_x(\delta)\cdot C(\#i|x)}\right) \\ &= \frac{1}{\delta}\cdot\left(\sum_{i=1}^{\kappa_x(\delta)} \delta\cdot\log\left(\frac{\delta}{\rho_x(\delta)\cdot C(\#i|x)}\right) + \sum_{i=\kappa_x(\delta)+1}^{2^\ell} C^*_{\{\delta\}}(\#i|x)\cdot\log\left(\frac{\rho_x(\delta)\cdot C(\#i|x)}{\rho_x(\delta)\cdot C(\#i|x)}\right)\right) \\ &= \frac{1}{\delta}\cdot\sum_{i=1}^{2^\ell} C^*_{\{\delta\}}(\#i|x)\cdot\log\left(\frac{C^*_{\{\delta\}}(\#i|x)}{\rho_x(\delta)\cdot C(\#i|x)}\right) \\ &= \frac{1}{\delta}\cdot\left(\mathrm{KL}(C^*_{\{\delta\}}|_{X=x} \parallel C|_{X=x}) - \log\rho_x(\delta)\right). \end{aligned}$$

Thus g_x is continuous, by the continuity of $\mathrm{KL}(\cdot \parallel C|_{X=x})$, ρ_x, and the function $\delta \to C^*_{\{\delta\}}(a|x)$ for every a.

Next we show that g_x is strictly increasing. We can expand g_x to be

$$g_x(\delta) = \sum_{i=1}^{\kappa_x(\delta)} \log\left(\frac{\delta}{\rho_x(\delta)\cdot C(\#i|x)}\right) = \sum_{i=1}^{\kappa_x(\delta)} \log\frac{\left(1-\sum_{i=1}^{\kappa(\delta)} C(\#i|x)\right)\cdot\delta}{(1-\kappa(\delta)\cdot\delta)\cdot C(\#i|x)}. \qquad (*)$$

We partition $(2^{-\ell}, C(\#1|x)]$ into intervals $I_x^{2^\ell}, \ldots, I_x^1$ according to the value of $\kappa_x(\cdot)$, i.e. $\delta \in I_x^\kappa \Rightarrow \kappa_x(\delta) = \kappa$ (note that we listed I_x^κ in reverse order of κ, as $\kappa_x(\cdot)$ is monotone decreasing). Since g_x is continuous it suffices to show that g_x is strictly increasing within each interval I_x^κ. Indeed, for $\delta, \delta' \in I_x^\kappa$, $\delta < \delta'$, using expression $(*)$ we have

$$g_x(\delta') - g_x(\delta) = \sum_{i=1}^{\kappa} \log\frac{\delta'\cdot(1-\kappa\cdot\delta)}{\delta\cdot(1-\kappa\cdot\delta')} > 0.$$

Given monotonicity, the range of g_x is specified by its values at (or towards) endpoints of $(2^{-\ell}, C(\#1|x)]$. Using $(*)$ we calculate

$$\begin{aligned}\lim_{\delta\to 2^{-\ell}+} g_x(\delta) &= \lim_{\delta\to 2^{-\ell}+} \sum_{i=1}^{\kappa_x(\delta)} \log\left(\frac{\left(1-\sum_{i=1}^{\kappa(\delta)} C(\#i|x)\right)\cdot\delta}{(1-\kappa(\delta)\cdot\delta)\cdot C(\#i|x)}\right)\\ &= \sum_{i=1}^{2^\ell} \log\frac{C(\#2^\ell|x)}{C(\#i|x)}\end{aligned}$$

where we use $\kappa_x(\delta) = 2^\ell - 1$ (for $\delta \to 2^{-\ell}+$); as well as

$$g_x(C(\#1|x)) = \log\frac{C(\#1|x)\cdot(1-C(\#1|x))}{1-C(\#1|x)} + \log\frac{1}{C(\#1|x)} = 0$$

where we use $\kappa_x(C(\#1|x)) = 1$. Thus

$$\mathrm{Range}(g_x) = \left(\sum_{i=1}^{2^\ell} \log\frac{C(\#2^\ell|x)}{C(\#i|x)}, 0\right].$$

We can now prove that $(X, C^*_{\{\delta_x\}})$ is the KL projection. We assume $r < \ell$; otherwise the result holds trivially. First note that for any fixed x and any $\delta \in [2^{-\ell}, C(\#1|x)]$, subject to the constraint $\forall a, C^*(a|x) \le \delta$, $\mathrm{KL}(C^*|_{X=x} \parallel C|_{X=x})$ is minimized by setting $C^*|_{X=x}$ to equal $C^*_{\{\delta\}}|_{X=x}$. See Lemma 2.3 of Barak et al. [BHK]. We view $\mathrm{KL}(X, C^*_{\{\delta_x\}} \parallel X, C)$ as a function f defined on the 2^n variables $\{\delta_x : x \in \{0,1\}^n\}$. Thus, to minimize $\mathrm{KL}(X, C^* \parallel X, C)$ subject to $\widetilde{\mathrm{H}}_\infty(C|X) \ge r$, it is equivalent to minimize f subject to the constraints $\mathbb{E}_{x\sim X}[\delta_x] \le 2^{-r}$ and $2^{-\ell} \le \delta_x \le C(\#1|x)$. In fact, we have $\mathbb{E}_{x\sim X}[\delta_x] = 2^{-r}$ as the KL projection must be on the boundary (Lemma A.4). We now use the KKT condition to find such minimum.

It can be verified that, in the interior of every nonempty interval I^κ_x, f has a partial derivative w.r.t. δ_x of

$$\begin{aligned}\frac{\partial}{\partial\delta_x}\mathrm{KL}(X, C^*_{\{\delta_x\}} \parallel X, C) &= \Pr[X=x]\cdot\frac{\partial}{\partial\delta_x}\mathrm{KL}(C^*_{\{\delta_x\}}|_{X=x} \parallel C|_{X=x})\\ &= \Pr[X=x]\cdot g_x(\delta_x)\end{aligned}$$

As g_x is continuous, this is a continuous partial derivative of f within all of $(2^{-\ell}, C(\#1|x)]$. Also, the constraint $\mathbb{E}_{x\sim X}[\delta_x] = 2^{-r}$ is linear in δ_x with coefficient $\Pr[X = x]$. The KKT condition says that for the optimal $\{\delta_x\}$, there must exist a constant $\lambda \in (-\infty, 0]$ such that for all x, either

1. $\delta_x = 2^{-\ell}$; or
2. $\delta_x = C(\#1|x)$; or
3. $\delta_x \in (2^{-\ell}, C(\#1|x))$ and $\delta_x = g_x^{-1}(\lambda)$.

(Note that the KKT condition can be applied since all our constraints are linear. See e.g. [BV] Chapter 5 for details on the KKT condition.)

To complete the proof, it remains to show that Item 1 holds only if $\lambda \notin \mathrm{Range}(g_x)$ (the argument that Item 2 holds only if $\lambda = 0$ is similar). Suppose for contradiction that for some x', Item 1 holds namely $\delta_{x'} = 2^{-\ell}$, but $\lambda \in \mathrm{Range}(g_{x'})$. Let $\delta_{x'} \in (2^{-\ell}, C(\#1|x')]$ be such that $\delta_{x'} = g_{x'}^{-1}(\lambda)$. Since $\mathbb{E}_{x\sim X}[\delta_x] = 2^{-r} > 2^{-\ell}$, by averaging there must exist $x'' \neq x'$ such that $\delta_{x''} > 2^{-\ell}$. Imagine that we modify $C^*_{\{\delta_x\}}$ by increasing $\delta_{x'}$ from $2^{-\ell}$ by $\epsilon/\Pr[X = x']$ for some miniscule $\epsilon > 0$, and simultaneously decreasing $\delta_{x''} > 2^{-\ell}$ by $\epsilon/\Pr[X = x'']$. Since $\delta_{x''} > \delta_{x'}$ and the gradient $\Pr[X = x] \cdot g_x(\cdot)$ is strictly increasing for all x, the modified distribution will have a smaller KL divergence (with average min-entropy unchanged), contradicting the optimality of $(X, C^*_{\{\delta_x\}})$. □

Theorem 4.20 (Approximating KL projection on high average min-entropy distributions)**.** *Let n be a security parameter, $X = X(n)$ be a distribution on $\{0,1\}^n$. There exists a* $\mathrm{poly}(n, 2^\ell, \tau, 1/\sigma, \log(1/\gamma))$ *time randomized algorithm Π such that the following holds. Let W be a deterministic circuit of size s such that $W(x)$ outputs the probability vector of $C|_{X=x}$, for some distribution C on $\{0,1\}^{\ell=\ell(n)}$ jointly distributed with X where $C(\cdot|\cdot)$ has bit length*

at most $\tau = \tau(n)$. Then for all $\sigma > 0$ and $0 \le r \le \ell$, $\Pi^{O_X}(W, r, \sigma)$ outputs w.p. $1 - \gamma$ some deterministic circuit M such that

1. *$M(x)$ outputs the probability vector of $C'|_{X=x}$ for some distribution C' on $\{0,1\}^\ell$ jointly distributed with X, where (X, C') is a σ-approximate KL projection of (X, C) on $\widetilde{\mathcal{V}}_r$;*
2. *M is of size $s + 2^\ell \cdot \mathrm{poly}(n, \ell, \tau, \log(1/\sigma))$;*
3. *For all x, a, $C'(a|x)$ has bit length $\tau + O(\ell + \log(1/\sigma))$.*

Proof. On the high level, the algorithm Π performs a binary search to approximate the $\lambda \in (-\infty, 0]$ in Lemma 4.19 that achieves $\mathbb{E}_{x\sim X}[\delta_x(\lambda)] = 2^{-r}$, and then approximates $\delta_x(\lambda)$ to obtain an approximate KL projection. Binary search is possible because given that g_x is strictly increasing, $\mathbb{E}_{x\sim X}[\delta_x(\cdot)]$ is also strictly increasing.

We will assume all the notations of Lemma 4.19, which says that the (exact) KL projection equals $(X, C^*) = (X, C^*_{\{\delta_x(\lambda)\}})$. Given that g_x has range $(\sum_{i=1}^{2^\ell} \log \frac{C(\#2^\ell|x)}{C(\#i|x)}, 0]$ (by Lemma 4.19), we can w.l.o.g. assume that

$$\lambda \ge \sum_{i=1}^{2^\ell} \log \frac{C(\#2^\ell|x)}{C(\#i|x)} \ge -2^\ell \cdot \tau$$

where the last inequality holds because τ is an upper bound on the bit length of $C(\#2^\ell|x)$.

The Algorithm Π. We now describe the algorithm Π. Let $\epsilon = \sigma / \left(c \cdot 2^\ell\right)$ for a sufficiently large constant c. W.l.o.g. we assume τ, $1/\sigma$, and $1/\epsilon$ are all powers of 2. The algorithm Π proceeds in $t = \ell \cdot (\log \tau + \log(1/\epsilon))$ iterations. Initially the range of λ is $[\mathrm{low}_1, \mathrm{high}_1] = [-2^\ell \cdot \tau, 0]$. In the ith iteration, we reduce the range of λ from $[\mathrm{low}_i, \mathrm{high}_i]$ to either $[\mathrm{low}_i, \mathrm{mid}_i]$ or $[\mathrm{mid}_i, \mathrm{high}_i]$ where $\mathrm{mid}_i = (\mathrm{low}_i + \mathrm{high}_i)/2$, as follows:

1. (Deterministically) compute an approximation $\widetilde{\delta_x}(\mathrm{mid}_i)$ of $\delta_x(\mathrm{mid}_i)$, such that $\delta_x(\mathrm{mid}_i) - \widetilde{\delta_x}(\mathrm{mid}_i) \in [0, \epsilon]$ and $\widetilde{\delta_x}(\mathrm{mid}_i)$ is a multiple of $\epsilon/2$;

2. Compute an estimate E_i of $\mathbb{E}_{x\sim X}\left[\widetilde{\delta}_x(\mathrm{mid}_i)\right]$, by taking $m = O((1/\epsilon^2)\log(t/\gamma))$ independent samples of $x \sim X$;

3. If $E_i \le 2^{-r}$, we reduce the range to $[\mathrm{mid}_i, \mathrm{high}_i]$, otherwise we reduce the range to $[\mathrm{low}_i, \mathrm{mid}_i]$.

After all t iterations, we let $\widetilde{\lambda} = \mathrm{low}_{t+1} - 2\epsilon$ be the approximation for λ, and let $\left(X, C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}\right)$ be the desired σ-approximate KL projection.

Since we require a bounded bit length, we round down $C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(\#i|x)$ after $\max\{\tau + \log(1/\sigma) + 1, \ell + \log(1/\epsilon) + 1\}$ bits for all $i < 2^\ell$, and increase $C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(\#2^\ell|x)$ accordingly. (We keep at least $\tau + \log(1/\sigma) + 1$ bits to ensure that the approximation is within a factor of $1 + \sigma/2$ from $C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(\#i|x)$. We keep at least $\ell + \log(1/\epsilon) + 1$ bits to ensure that $C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(\#2^\ell|x)$ will never exceed $\widetilde{\delta_x}(\widetilde{\lambda})$ which is $\log(1/\epsilon) + 1$ bit long). Let $\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}$ be the resulting distribution; Π outputs a deterministic circuit M such that $M(x)$ outputs the probability vector of $\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}|_{X=x}$.

Implementation for Step 1. We now describe how to implement Step 1 efficiently. Since g_x is strictly increasing, we do an $O(\log(1/\epsilon))$-round binary search for the largest $\delta \in (2^{-\ell}, C(\#1|x)]$ such that $g_x(\delta) \le \mathrm{mid}_i$ and δ is a multiple of $\epsilon/2$. In fact, since $g_x(\delta)$ cannot be computed exactly, we will use an approximation $\widetilde{g_x}(\delta) \in [g_x(\delta) - \epsilon/2, g_x(\delta)]$ that can be computed efficiently. We let $\widetilde{\delta_x}(\mathrm{mid}_i) = \delta^*$ where δ^* is the outcome of the binary search.

Note that replacing $g_x(\delta)$ by $\widetilde{g_x}(\delta)$ is equivalent to replacing mid_i by some $\widetilde{\mathrm{mid}_i} \in [\mathrm{mid}_i - \epsilon/2, \mathrm{mid}_i]$. The $\epsilon/2$ granularity in binary search ensures that $\delta_x(\widetilde{\mathrm{mid}_i}) - \delta^* \in [0, \epsilon/2.]$ Moreover, the $\epsilon/2$ deviation from mid_i causes at most $\epsilon/2$ deviation from $\delta_x(\mathrm{mid}_i)$,

i.e. $\delta_x(\mathrm{mid}_i) - \delta_x(\widetilde{\mathrm{mid}_i}) \in [0, \epsilon/2]$, as shown in Claim 4.21 below. Therefore, we achieve

$$\delta_x(\mathrm{mid}_i) - \widetilde{\delta}_x(\mathrm{mid}_i) = \left(\delta_x(\mathrm{mid}_i) - \delta_x(\widetilde{\mathrm{mid}_i})\right) + \left(\delta_x(\widetilde{\mathrm{mid}_i}) - \delta^*\right) \in \left[0, \frac{\epsilon}{2} + \frac{\epsilon}{2}\right].$$

Finally we describe how to approximate $g_x(\delta)$ to $\tau' = 1 + \log(1/\epsilon)$ decimal places, for any $\delta \in (2^{-\ell}, C(\#1|x)]$ that is a multiple of $\epsilon/2$, using the formula

$$g_x(\delta) = \sum_{i=1}^{\kappa_x(\delta)} \log\left(\frac{\delta}{\rho_x(\delta) \cdot C(\#i|x)}\right) = \sum_{i=1}^{\kappa_x(\delta)} \log \frac{\left(1 - \sum_{i=1}^{\kappa(\delta)} C(\#i|x)\right) \cdot \delta}{(1 - \kappa(\delta) \cdot \delta) \cdot C(\#i|x)}.$$

Evaluating $\kappa_x(\delta)$ involves $O(2^\ell)$ comparisons of s-bit numbers in $[0,1]$. Approximating $g_x(\delta)$ involves: (i) computing $O(2^\ell)$ logarithms of $\max\{\tau,\tau'\}$-bit numbers in $[0,1]$, to $\tau' + \ell + O(1)$ decimal places; (ii) $O(2^\ell)$ additions on $\max\{\tau,\tau'\}$-bit numbers in $[0,1]$; (iii) $O(2^\ell)$ additions on $O(\tau + \tau' + \ell)$-bit numbers that each have $\tau' + \ell + O(1)$ decimal places.

Efficiency and Circuit Size. Approximating $g_x(\delta)$ requires $2^\ell \cdot (s + \mathrm{poly}(n, \ell, \tau, \log(1/\epsilon)))$ time. Thus the overall time of computing $\widetilde{\delta}_x(\cdot)$ is $2^\ell \cdot (s + \mathrm{poly}(n, \ell, \tau, \log(1/\epsilon)))$, and the overall running time of Π is $m \cdot \left(s + t \cdot 2^\ell \cdot \mathrm{poly}(n, \ell, \tau, \log(1/\epsilon))\right) = s \cdot \mathrm{poly}(n, 2^\ell, \tau, 1/\sigma, \log(1/\gamma))$. Moreover, Π constructs a circuit computing $\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}(a|x)$ of size $2^\ell \cdot (s + \mathrm{poly}(n, \ell, \tau, \log(1/\sigma)))$.

Correctness of Π. Recall that to prove $(X, \widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}})$ is a σ-approximate KL projection, by Pythagorean Theorem (Theorem 1.13), it suffices to show that $(X, \widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}) \in \widetilde{\mathcal{V}}_r(X)$ and $\mathrm{KL}(X, B \parallel X, \widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}) - \mathrm{KL}(X, B \parallel X, C^*_{\{\delta_x(\lambda_x)\}}) \le \sigma$ for all joint distributions $(X, B) \in \widetilde{\mathcal{V}}_r(X)$. Suppose $\delta_x(\lambda) - \widetilde{\delta}_x(\tilde{\lambda}) \in [0, 2^{-\ell-1}\sigma]$ for all x; then we are done, because that implies:

- $(X, \widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}) \in \widetilde{\mathcal{V}}_r(X)$, since

$$\widetilde{\mathrm{H}}_\infty(X, \widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}) \ge \log \frac{1}{\mathbb{E}_{x \sim X}\left[\widetilde{\delta}_x(\widetilde{\lambda})\right]} \ge \log \frac{1}{\mathbb{E}_{x \sim X}\left[\delta_x(\lambda)\right]} = r;$$

- $C^*_{\{\delta_x(\lambda)\}}(a|x) \le (1+\sigma/2) \cdot C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(a|x)$ for all pairs of (x,a). Recall that $\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}(a|x) \le$

$(1+\sigma/2)\cdot C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(a|x)$, thus for all joint distributions (X,B),

$$\begin{aligned}
&\mathrm{KL}(X,B \parallel X,\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}) - \mathrm{KL}(X,B \parallel X, C^*_{\{\delta_x(\lambda_x)\}}) \\
&= \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\log\frac{C^*_{\{\delta_x(\lambda)\}}(a|x)}{\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}(a|x)}\right] \\
&= \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\log\left(\frac{C^*_{\{\delta_x(\lambda)\}}(a|x)}{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(a|x)}\cdot\frac{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}(a|x)}{\widetilde{C^*_{\{\widetilde{\delta}_x(\tilde{\lambda})\}}}(a|x)}\right)\right] \\
&\le \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\log\left((1+\frac{\sigma}{2})\cdot(1+\frac{\sigma}{2})\right)\right] \\
&\le \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\cdot\sigma\right] = \sigma.
\end{aligned}$$

Hence for the rest of the proof we will show $\delta_x(\lambda)-\widetilde{\delta}_x(\tilde{\lambda})\in[0,2^{-\ell-1}\sigma]$, assuming that Step 2 of all iterations achieves an accurate estimation, i.e. $\left|E_i-\mathbb{E}_{x\sim X}\left[\widetilde{\delta}_x(\mathrm{mid}_i)\right]\right|\le\epsilon$. The latter assumption holds with all but $t\cdot 2^{-\Omega(m\epsilon^2)}\le\gamma$ probability, by a Chernoff bound and a union bound.

By triangle inequality,

$$\left|E_i-\mathop{\mathbb{E}}_{x\sim X}[\delta_x(\mathrm{mid}_i)]\right|\le\left|E_i-\mathop{\mathbb{E}}_{x\sim X}\left[\widetilde{\delta}_x(\mathrm{mid}_i)\right]\right|+\left|\mathop{\mathbb{E}}_{x\sim X}\left[\widetilde{\delta}_x(\mathrm{mid}_i)\right]-\mathop{\mathbb{E}}_{x\sim X}[\delta_x(\mathrm{mid}_i)]\right|\le 2\epsilon.$$

Thus Step 3 ensures $\lambda\in[\mathrm{low}_i-2\epsilon,\mathrm{high}_i+2\epsilon]$ for all i. In particular, since $\mathrm{high}_{t+1}-\mathrm{low}_{t+1}=\tau\cdot 2^\ell/2^t\le\epsilon$ and $\widetilde{\lambda}=\mathrm{low}_{t+1}-2\epsilon$, we have $\lambda-\widetilde{\lambda}\in[0,5\epsilon]$. By Claim 4.21 below, $\delta_x(\lambda)-\delta_x(\widetilde{\lambda})\in[0,5\epsilon]$, thus

$$\delta_x(\lambda)-\widetilde{\delta}_x(\tilde{\lambda})=\left(\delta_x(\lambda)-\delta_x(\widetilde{\lambda})\right)+\left(\delta_x(\widetilde{\lambda})-\widetilde{\delta}_x(\widetilde{\lambda})\right)\in[0,6\epsilon]=[0,2^{-\ell-1}\sigma]$$

completing the proof.

Claim 4.21. Let $\lambda_1<\lambda_2\le 0$. Then $\delta_x(\lambda_2)-\delta_x(\lambda_1)\in[0,\lambda_2-\lambda_1)$.

Proof. Note that, where differentiable, g_x has a gradient of

$$g_x'(\delta)=\frac{\log e}{\frac{\delta}{\kappa_x(\delta)}-\delta^2}>\frac{\log e}{4}$$

where we use the fact that $2^\ell < \delta < C(1|x)$ and $1 \leq \kappa_x(\delta) \leq 2^\ell - 1$ to obtain the bound. The claim follows because $\delta_x(\cdot)$ is continuous, and where differentiable, $\delta_x(\cdot)$ has a gradient of either $1/g'_x < 1$, or zero (when $\delta_x(\cdot) = 2^\ell$). $\delta_x(\cdot)$ is continuous because it equals the inverse of the strictly increasing continuous function g_x, except when outside the range of g_x, $\delta_x(\cdot)$ equals 2^ℓ, the left endpoint of the interval on which g_x is defined. □

□

4.2.4.2 Putting it Together

Theorem 4.22 (Hardness of prediction $\Longrightarrow$ pseudo-avg-min-entropy, uniform setting)**.** *Let n be a security parameter, $\ell = \ell(n)$, $t = t(n)$, $\epsilon = \epsilon(n)$, $r = r(n) \leq \ell(n)$, where t, ϵ, and 2^{-r} are computable in* poly(n) *time. Let $(X, B) = (X, B)(n)$ be a* poly(n) *time samplable joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$ such that B is uniformly $(t, 1 - 2^{-r})$-hard to predict given X. Then B has uniform (t', ϵ) pseudo-avg-min-entropy at least r given X, for $t' = \left(t \cdot 2^{-\ell}\right)^{\Omega(1)} / \text{poly}(n, \ell, 1/\epsilon)$.*

Proof. Suppose for contradiction that B does not have uniform (t', ϵ) pseudo-avg-min-entropy at least r given X. By definition, there is a time t' randomized oracle algorithm A such that for infinitely many n and every joint distribution (X, C) with $\widetilde{\mathrm{H}}_\infty(C|X) \geq r$, $A^{O_{X,B,C}}$ is an ϵ-distinguisher between (X, B) and (X, C). Using A we shall construct a time t oracle algorithm P such that for infinitely many n, $\Pr[P^{O_{X,B}}(X) = B] \geq 2^{-r}$.

Consider the two-player zero-sum game where Player 1 chooses some joint distribution $(X, C) \in \widetilde{\mathcal{V}}_r(X)$, Player 2 chooses a circuit W and receives expected payoff $\mathbb{E}[f((X, C), W)] = \mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]$. We will apply the Uniform Min-Max Theorem – Average Case (Chapter 2, Theorem 2.5) to this game, i.e. with

- $\mathcal{V} = \widetilde{V}_r$;

- $\mathcal{W} = \{\text{(deterministic) circuits}\}$;
- $f((x,a),W) = \mathbb{E}[W(X,B)] - W(x,a)$.

We run an instantiation of Algorithm 2.2 (Finding Universal Strategy – Average Case) for the game with KL projection on the set $\widetilde{\mathcal{V}}_r(X)$, which we describe below. Using the oracle algorithm $A^{(\cdot)}$, we will show that in each iteration we obtain some $W^{(i)}$ that distinguishes (X,B) and (X,C). Thus, by the Uniform Min-Max Theorem – Average Case, we obtain a universal distinguisher W^*. From this W^*, we then obtain an efficient predictor for B by applying Lemma 4.16, exactly like in the nonuniform setting.

Our instantiation of Algorithm 2.1 starts with an initial distribution $(X, C^{(1)})$ where $C^{(1)}$ is uniform on $\{0,1\}^\ell$ and independent of X. Let $\epsilon' = \epsilon/c$ for a sufficiently large constant c. The number of iterations is $S = O(\ell/\epsilon'^2)$, and we let $\gamma' = \epsilon/8S$. In each iteration we represent $C^{(i)}$ by a deterministic circuit $M^{(i)}$ such that $M^{(i)}(x)$ outputs the probability vector of $C^{(i)}|_{X=x}$. So we can take $M^{(1)}(x)$ to be the vector $(1/2^\ell, \ldots, 1/2^\ell)$ for every x. We show how to implement each of the S iterations of Algorithm 2.2 efficiently:

1. **Obtaining Player 2's Response $W^{(i)}$:** Suppose that we have constructed a t_i-size deterministic circuit $M^{(i)}$, and $C^{(i)}(\cdot|\cdot)$ has bit length τ_i. There are two steps:

   (a) Generate a deterministic circuit $\widetilde{W}^{(i)}$ such that

   $$\mathbb{E}[\widetilde{W}^{(i)}(X,B)] - \mathbb{E}[\widetilde{W}^{(i)}(X,C^{(i)})] \geq \mathbb{E}[A^{O_{X,B,C^{(i)}}}(X,B)] - \mathbb{E}[A^{O_{X,B,C^{(i)}}}(X,C^{(i)})] - \epsilon'.$$

   To do so, we first generate $m = O(\log(1/\gamma')/\epsilon'^2)$ random samples of $(X,B,C^{(i)})^{t'}$ and $U_{t'}$. This can be done in time $mt'2^\ell \cdot \mathrm{poly}(n,t_i)$, where we sample $C^{(i)}|_{X=x}$ from its probability vector $M^{(i)}(x)$. Now let $\widetilde{W}^{(i)}(x,a)$ runs $A^{(\cdot)}(x,a)$ for m times and returns the average of the m outputs; each time $A^{(\cdot)}(x,a)$ is run using one copy of $(X,B,C^{(i)})^{t'}$ to answer oracle queries of A, and one copy of $U_{t'}$ as

coin tosses of A. The m random samples are hardwired in $\widetilde{W}^{(i)}$, thus $\widetilde{W}^{(i)}$ is of size $t'' = O(t' \cdot m \cdot (n+\ell))$, which does not depend on the size of $M^{(i)}$ (but the size of $M^{(i+1)}$ will additively depend on t''). By a Chernoff bound, the above inequality holds w.p. at least $1-\gamma'$.

(b) Our choice of $W^{(i)}$ is the following approximation to $\widetilde{W}^{(i)}$, so that $\exp(-\epsilon' \cdot (1 - W^{(i)}(x,a)))$ can be computed precisely and efficiently. First, we use Newton's method to compute a $\mathrm{polylog}(1/\epsilon)$-bit approximation $E(x,a) \in (0,1]$ of $\exp(-\epsilon' \cdot (1-\widetilde{W}^{(i)}(x,a)))$ within $\pm\epsilon'^2$ error, in time $O(tm) + \mathrm{polylog}(1/\epsilon)$. We define $W^{(i)}$ to be such that $\exp(-\epsilon' \cdot (1-W^{(i)}(x,a))) = E(x,a)$. Thus $\left|W^{(i)}(x,a) - \widetilde{W}^{(i)}(x,a)\right| \leq \epsilon'$, and

$$\mathbb{E}[W^{(i)}(X,B)] - \mathbb{E}[W^{(i)}(X,C^{(i)})] \geq \mathbb{E}[\widetilde{W}^{(i)}(X,B)] - \mathbb{E}[\widetilde{W}^{(i)}(X,C^{(i)})] - 2\epsilon'.$$

2. **Weight Update:** We represent the resulting distribution $C^{(i)'}$ after weight update by the circuit $M^{(i)'}$ where $M^{(i)'}(x)$ outputs the probability vector of $C^{(i)'}|_{X=x}$. Since $E(x,a) = \exp\left(-\epsilon' \cdot (1 - W^{(i)}(x,a))\right)$ has bit length $\mathrm{polylog}(1/\epsilon)$ and $C(a|x)$ has bit length τ_i, multiplication takes time $\mathrm{polylog}(1/\epsilon') \cdot \tau_i$ for each pair of x, a. Thus, $M^{(i)'}$ has circuit size $t'_i = t_i + 2^\ell \cdot (t'' + \mathrm{polylog}(1/\epsilon') \cdot \tau_i)$ and can be constructed in similar time; $C^{(i)'}(\cdot|\cdot)$ has bit length $\tau'_i = \tau_i + \mathrm{polylog}(1/\epsilon')$.

3. **KL Projection:** We use Theorem 4.20 to efficiently obtain a circuit $M^{(i+1)}$ such that $M^{(i+1)}(x)$ outputs the probability vector of $C^{(i+1)}|_{X=x}$, where $(X, C^{(i+1)})$ is an ϵ'^2-approximate KL projection of $(X, C^{(i)'})$ on $\widetilde{\mathcal{V}}_r$. This can be done in time $\mathrm{poly}(t'_i, n, 2^\ell, \tau'_i, 1/\epsilon', \log(1/\gamma'))$ and w.p. at least $1-\gamma'$. Moreover, $M^{(i+1)}$ is of size $t_{i+1} = t'_i + 2^\ell \cdot \mathrm{poly}(n, \ell, \tau'_i, \log(1/\epsilon'))$, and $C^{(i+1)}(\cdot|\cdot)$ has bit length $\tau_{i+1} = \tau'_i + O(\ell + \log(1/\epsilon'))$.

We now prove efficiency of the algorithm. First, note that $\tau_1 = O(\ell)$, and $\tau_i = i \cdot$

$(\text{polylog}(1/\epsilon') + O(\ell))$ by induction. Also, $t_1 = O(\ell)$ and $t_{i+1} = t_i + 2^\ell \cdot \text{poly}(t', n, \ell, \tau_i, 1/\epsilon') = t_i + 2^\ell \cdot \text{poly}(t', n, \ell, 1/\epsilon')$. Thus the above algorithm runs in total time $2^\ell \cdot \text{poly}(t', n, \ell, 1/\epsilon)$.

Suppose that Step 1 (a) and Step 3 complete successfully in all iterations. By a union bound, this holds w.p. at least $1 - 2\gamma' S = 1 - \epsilon/4$. For all i, since $(X, C^{(i)}) \in \widetilde{\mathcal{V}}_r$,the pseudo-avg-min-entropy of B implies

$$\begin{aligned}
\mathbb{E}[W^{(i)}(X,B)] - \mathbb{E}[W^{(i)}(X,C^{(i)})] &\geq \mathbb{E}[\widetilde{W}^{(i)}(X,B)] - \mathbb{E}[\widetilde{W}^{(i)}(X,C^{(i)})] - 2\epsilon' \\
&\geq \mathbb{E}[A^{O_{X,B,C^{(i)}}}(X,B)] - \mathbb{E}[A^{O_{X,B,C^{(i)}}}(X,C^{(i)})] - \epsilon' - 2\epsilon' \\
&\geq \epsilon - 2\epsilon'.
\end{aligned}$$

Let W^* be the size $O(St'')$ deterministic circuit computing the average of $\widetilde{W}^{(1)}, \ldots, \widetilde{W}^{(S)}$. Note that W^* is at most $2\epsilon'$ apart from the average of $W^{(1)}, \ldots, W^{(S)}$. Hence by the Uniform Min-Max Theorem – Average Case (Theorem 2.5), for all Player 1 strategies $(X, C) \in \widetilde{\mathcal{V}}_r$,

$$\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] \geq \epsilon - 2\epsilon' - O(\epsilon') - 2\epsilon' \geq 3\epsilon/4.$$

This implies $\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] \geq 3\epsilon/4 - \epsilon/2 = \epsilon/4$ for all (X, C) such that $\widetilde{\mathrm{H}}_\infty(C|X) \geq r'$ where $2^{-r'} = 2^{-r} + \epsilon/2$. In other words, the algorithm constructs a universal $\epsilon/4$-distinguisher W^* between (X, B) and $\widetilde{\mathcal{V}}_{r'}(X)$ w.p. at least $1 - \epsilon/4$.

Given a universal distinguisher W^*, the remainder of the proof is similar to the nonuniform setting. In order to apply Lemma 4.16 we make the following assumptions:

1. $\Pr[X = x] \leq \epsilon/4$ for all x. This is w.l.o.g. because we can pad X with $\log(1/\epsilon) + 2$ random bits, i.e. replace (X, B) by the new joint distribution $((X, U_{\log(1/\epsilon)+2}), B)$.

2. We "perturb" W^* into $\widetilde{W^*}$ so that

   (a) $\Delta_{\widetilde{W^*}}(x, \kappa)$ is distinct for each pair of $x \in \{0,1\}^n$ and $\kappa \in [2^\ell]$;

   (b) $\mathbb{E}\left[\widetilde{W^*}(X,B)\right] - \mathbb{E}\left[\widetilde{W^*}(X,C)\right] \geq \mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] - \epsilon/4$ for all (X, C).

The can be done by adding a negligible $\min\{S^{-1}, .1\epsilon\} \cdot \left(x \cdot i/2^{n+\ell}\right)$ to $W^*(x,i)$, where we interpret x as a natural number. The resulting $\widetilde{W^*}(\cdot,\cdot)$ has bit length at most $\tau = n + O(\log n) + \ell + O(1/\epsilon)$, and is of size $t'' = O(St') + O(\tau + \log(1/S)) = t' \cdot \mathrm{poly}(n, \ell, 1/\epsilon)$.

We run the algorithm N in Lemma 4.16 to convert $\widetilde{W^*}$ into a predictor. By Lemma 4.16, this yields a time $2^\ell \cdot \mathrm{poly}(t', n, \ell, 1/\epsilon) \le t$ randomized oracle algorithm P such that

$$\begin{aligned}\Pr[P^{O_{X,B}}(X) = B] &\ge 2^{-r'} - \frac{\epsilon}{4} + \frac{\mathbb{E}[\widetilde{W^*}(X,B)] - \mathbb{E}[\widetilde{W^*}(X,C)]}{\lambda} \\ &\ge 2^{-r'} - \frac{\epsilon}{4} + \frac{\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] - \epsilon/4}{\lambda}\end{aligned}$$

where $\lambda \in (0, 2^\ell]$ and $(X,C) \in \widetilde{V}_{r'}(X)$. Recall that w.p. at least $1 - \epsilon/4$, we have $\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] \ge \epsilon/4$. Therefore,

$$\Pr[P^{O_{X,B}}(X) = B] \ge 2^{-r'} - \frac{\epsilon}{4} - \frac{\epsilon}{4} \ge 2^{-r}$$

contradicting the hardness of predicting B given X. □

4.2.5 Pseudo-Avg-Min-Entropy Implies Hardness of Prediction

This direction of the characterization is straightforward to show. In fact, we show that even a weak form of pseudo-avg-min-entropy suffices:

Definition 4.23 (Weak pseudo-avg-min-entropy, nonuniform setting)**.** Let (X,B) be a joint distribution. We say B has (T,ϵ) *weak nonuniform pseudo-avg-min-entropy at least k given X* if there exists a joint distribution (Y,C) such that the following holds:

- $\widetilde{\mathrm{H}}_\infty(C|Y) \ge k$;
- (X,B) and (Y,C) are ϵ-indistinguishable by all size T circuits.

If $(X, B) = (X, B)(n)$ for a security parameter n, we say B has *weak pseudo-avg-min-entropy at least $k = k(n)$ given X* if for every constant c, B has $(n^c, 1/n^c)$ weak pseudo-avg-min-entropy at least k given X for all sufficiently large n.

In the uniform setting, it suffices to assume an even weaker form of pseudo-avg-min-entropy, where we only require indistinguishability against distinguishers given oracle access to $O_{X,B}$ but not $O_{X,B,C}$:

Definition 4.24 (Weak pseudo-avg-min-entropy, uniform setting)**.** Let n be a security parameter, $T = T(n)$, $\epsilon = \epsilon(n)$, $k = k(n)$, $\ell = \ell(n)$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B has (T, ϵ) *weak uniform pseudo-avg-min-entropy at least k given X* if for every randomized oracle algorithm A computable in time T, there is a joint distribution (Y, C) such that the following holds for all sufficiently large n:

- $\widetilde{\mathrm{H}}_\infty(C|Y) \geq k$;
- (X, B) and (Y, C) are indistinguishable by $A^{O_{X,B}}$:

$$|\Pr[A^{O_{X,B}}(X, B) = 1] - \Pr[A^{O_{X,B}}(Y, C) = 1]| < \epsilon.$$

We say B has *weak uniform pseudo-avg-min-entropy at least $k = k(n)$ given X* if for every constant c, B has $(n^c, 1/n^c)$ weak uniform pseudo-avg-min-entropy at least k given X. Note that in this "polynomial" version, $O_{X,B}$ is redundant if (X, B) is polynomial-time samplable.

Theorem 4.25 (Weak pseudo-avg-min-entropy $\Longrightarrow$ hardness of prediction, uniform and nonuniform settings)**.** *Let n be a security parameter, $\ell = \ell(n)$, $t = t(n)$, $\epsilon = \epsilon(n)$, $r = r(n) \leq \ell(n)$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$ such that B has weak (non)uniform (t, ϵ) pseudo-avg-min-entropy at least r given X. Then B is (non)uniformly $(t - O(1), 1 - 2^{-r} - \epsilon)$-hard to predict given X.*

Proof. We give a proof for the uniform setting; the proof for the nonuniform setting is essentially the same. Suppose for contradiction that there is a time $t - O(1)$ randomized oracle algorithm P such that $\Pr[P^{O_{X,B}}(X) = B] \geq 2^{-r} + \epsilon$. Define a distinguisher $W^{O_{X,B}}(x, a)$ that outputs 1 if $P^{O_{X,B}}(x) = a$, and 0 otherwise. Note that for all joint distributions (Y, C) we have an information-theoretic bound $\Pr[P^{O_{X,B}}(Y) = C] \leq \mathbb{E}_{x \sim Y}[\max_a C(a|x)]$. Thus all joint distributions (Y, C) on $\{0,1\}^n \times \{0,1\}^\ell$ with $\mathbb{E}_{x \sim Y}[\max_a C(a|x)] \leq 2^{-r}$, we have

$$\begin{aligned}\mathbb{E}[W^{O_{X,B}}(X, B)] - \mathbb{E}[W^{O_{X,B}}(Y, C)] &= \Pr[P^{O_{X,B}}(X) = B] - \Pr[P^{O_{X,B}}(Y) = C] \\ &\geq 2^{-r} + \epsilon - 2^{-r} = \epsilon,\end{aligned}$$

contradicting the fact that B has weak uniform (t, ϵ) pseudo-avg-min-entropy at least r. □

Since Theorem 4.25 only requires weak pseudo-avg-min-entropy, we now have the following equivalence:

Corollary 4.26. *Let n be a security parameter, and let $\delta = \delta(n) > 0$, and $\ell = \ell(n) = O(\log n)$ be computable in time* poly(n)*. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. Then the following are equivalent:*

1. *B is (non)uniformly $(1 - 2^{-k})$-hard to predict given X;*
2. *B has (non)uniform pseudo-avg-min-entropy at least k given X;*
3. *B has weak (non)uniform pseudo-avg-min-entropy at least k given X.*

Proof. $1 \implies 2$ by Theorem 4.18 and 4.22. $2 \implies 3$ by definition. $3 \implies 1$ by Theorem 4.25. □

4.3 Characterizing Pseudoentropy

4.3.1 Definitions

The computational analogue of Shannon entropy, *pseudoentropy*, was first introduced by Håstad et al. [HILL]. We begin with the nonuniform definition because it is simpler:

Definition 4.27 (Pseudoentropy, nonuniform setting)**.** We say that a distribution X has (T, ϵ) *nonuniform pseudoentropy at least* k if there exists a distribution Y with $\mathrm{H}_{\mathrm{sh}}(Y) \geq k$ such that X and Y are ϵ-indistinguishable by all size T circuits.

If $X = X(n)$ for a security parameter n, we say X has *nonuniform pseudoentropy at least* $k = k(n)$ if for every constant c, $X(n)$ has $(n^c, 1/n^c)$ nonuniform pseudoentropy at least $k(n) - 1/n^c$ for all sufficiently large n.

A natural generalization of *pseudoentropy* is the notion of *conditional* pseudoentropy:

Definition 4.28 (Conditional pseudoentropy, nonuniform setting)**.** Let (X, B) be a joint distribution. We say B has (T, ϵ) *nonuniform (conditional) pseudoentropy at least* k *given* X if there exists a distribution C jointly distributed with X such that the following holds:

- $\mathrm{H}_{\mathrm{sh}}(C|X) \geq k$;
- (X, B) and (X, C) are ϵ-indistinguishable by all size T circuits.

$k - \mathrm{H}_{\mathrm{sh}}(B|X)$ is known as the *pseudoentropy gap*.

If $(X, B) = (X, B)(n)$ for a security parameter n, we say B has *nonuniform (conditional) pseudoentropy at least* $k = k(n)$ *given* X if for every constant c, B has $(n^c, 1/n^c)$ nonuniform (conditional) pseudoentropy at least $k(n) - 1/n^c$ given X for all sufficiently large n.

Like pseudo-avg-min-entropy, both pseudoentropy and conditional pseudoentropy can be defined with respect to uniform observers:

Definition 4.29 (Pseudoentropy, uniform setting)**.** Let n be a security parameter, $T = T(n)$, $\epsilon = \epsilon(n)$, $k = k(n)$, $\ell = \ell(n)$. Let $X = X(n)$ be a distribution on $\{0,1\}^\ell$. We say X has (T, ϵ) *uniform pseudoentropy at least* k if for all time T randomized oracle algorithm A, there exists a distribution $Y = Y(n)$ such that the following holds for all sufficiently large n:

- $\mathrm{H}_{\mathrm{sh}}(Y) \geq k$;
- X, Y are ϵ-indistinguishable by $A^{O_{X,Y}}$:

$$\left|\Pr[A^{O_{X,Y}}(X) = 1] - \Pr[A^{O_{X,Y}}(Y) = 1]\right| < \epsilon.$$

We say X has *uniform pseudoentropy at least* $k = k(n)$ if for every constant c, $X(n)$ has $(n^c, 1/n^c)$ uniform pseudoentropy at least $k(n) - 1/n^c$.

Definition 4.30 (Conditional pseudoentropy, uniform setting)**.** Let n be a security parameter, $T = T(n)$, $\epsilon = \epsilon(n)$, $k = k(n)$, $\ell = \ell(n)$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B has (T, ϵ) *uniform (conditional) pseudoentropy at least* k *given* X if for every randomized oracle algorithm A computable in time T, there is a distribution C jointly distributed with X such that the following holds for all sufficiently large n:

- $\mathrm{H}_{\mathrm{sh}}(C|X) \geq k$;
- (X, B) and (X, C) are indistinguishable by $A^{O_{X,B,C}}$:

$$|\Pr[A^{O_{X,B,C}}(X, B) = 1] - \Pr[A^{O_{X,B,C}}(X, C) = 1]| < \epsilon.$$

We say B has *uniform (conditional) pseudoentropy at least* $k = k(n)$ *given* X if for every constant c, B has $(n^c, 1/n^c)$ uniform (conditional) pseudoentropy at least $k(n) - 1/n^c$ given X.

As in the definition of uniform pseudo-avg-min-entropy, we give the distinguishers oracle access to $O_{X,Y}$ (for pseudoentropy) and $O_{X,B,C}$ (for conditional pseudoentropy). For conditional pseudoentropy, however, a consequence of our results is that the definition with oracle $O_{X,B,C}$ is equivalent to the definition with oracle $O_{X,B}$ provided B comes from a polynomial-sized alphabet. In particular, if (X, B) is also polynomial-time samplable (which will be the case in our applications), the definition is equivalent to one without oracle $O_{X,B,C}$. (See Corollary 4.58.)

In the definition of conditional pseudoentropy, a question asked by Leo Reyzin is whether allowing changing *both* X and B (rather than changing (X, B) to (X, C), with X fixed) makes any difference. Another consequence of our results is that this is equivalent to the above definition. (See Corollary 4.58.)

To capture exactly the *computational* hardness in B given X, we consider the closest "distance" from (X, B) to any joint distribution (X, C) where the distribution C can be efficiently "represented" given X. We will use KL divergence as the "distance," and consider two ways to algorithmically represent C: (i) By a randomized algorithm or circuit S that samples C from X, i.e. $C = S(X)$; (ii) By an algorithm or circuit P that computes the (conditional) probability mass function (pmf) of C, i.e. $P(x, a) = \Pr[C = a|X = x]$. In general, having an efficient algorithm for one representation does not imply having an efficient algorithm for the other (under certain complexity assumptions) [KMR$^+$, Nao2]. But when C is short ($\ell = O(\log n)$), approximating the pmf of C given X (say to within $\pm\epsilon$) is equivalent to approximately sampling C given X (say to within statistical distance ϵ), up to a factor of $\text{poly}(2^\ell, 1/\epsilon)$ in running time. (See Lemma 4.36 and 4.37 below.) The sampler-based definition may appear more natural, as a closer parallel to Definition 4.13:

Definition 4.31 (KL-hard for sampling, nonuniform setting)**.** Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B is *nonuniformly* (t, δ) *KL-hard for sampling given* X if

for all size t randomized circuits $S : \{0,1\}^n \to \{0,1\}^\ell$ it holds that $\mathrm{KL}(X, B||X, S(X)) \geq \delta$.

Analogously to pseudoentropy, the nonuniform and uniform definitions differ in whether we need to give a sampling oracle to the adversary.

Definition 4.32 (KL-hard for sampling, uniform setting)**.** Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $\ell = \ell(n)$. Let (X, B) be a distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B is *uniformly* (t, δ) *KL-hard for sampling given* X if for all time t randomized oracle algorithms S, for all sufficiently large n, it holds that $\mathrm{KL}(X, B||X, S^{O_{X,B}}(X)) \geq \delta$.

In our characterization of pseudoentropy (Section 4.3), however, we adopt the pmf-based representation (rather than the sampling-based). This is because our techniques require finer manipulations of the distribution.

We will in fact use "measures" rather than pmfs, because it can be infeasible to maintain the normalization $\sum_a P(x, a) = 1$ while manipulating P if the alphabet size 2^ℓ is large. Recall that a function $P : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ is called a *conditional measure* for $C|X$ if:

$$C(a|x) = \frac{P(x,a)}{\sum_b P(x,b)},$$

and we denote (X, C) by (X, Φ_P). We generalize the pmf representation so that P only has to compute some conditional measure for $C|X$.

Definition 4.33 (KL predictors)**.** Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say that a conditional measure $P : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ is a δ*-KL predictor* of B given X if

$$\mathrm{KL}(X, B||X, \Phi_P) < \delta.$$

If P is randomized, we say that P is a δ*-KL predictor* of B given X if

$$\mathop{\mathbb{E}}_{p \sim P}[\mathrm{KL}(X, B||X, \Phi_p)] < \delta$$

where we view P as a distribution over functions $p : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$.

Definition 4.34 (KL-hard, nonuniform setting)**.** Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B is *nonuniformly* (t, δ) *KL-hard given* X if there is no circuit P of size t that is a δ-KL predictor of B given X.

We say B is *nonuniformly* δ *KL-hard given* X if for every constant c, B is nonuniformly $(n^c, \delta - 1/n^c)$ KL-hard given X for all sufficiently large n.

Definition 4.35 (KL-hard, uniform setting)**.** Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $\ell = \ell(n)$. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B is *uniformly* (t, δ) *KL-hard given* X if for all time t randomized oracle algorithms $P : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ and all sufficiently large n, $P^{O_{X,B}}$ is not a δ-KL predictor of B given X (where the randomness of $P^{O_{X,B}}$ consists both of its internal coin tosses and the samples it gets from the oracle $O_{X,B}$).

We say B is *uniformly* δ *KL-hard given* X if for every constant c, B is uniformly $(n^c, \delta - 1/n^c)$ KL-hard given X.

Note that by letting $P(x, a) = 1$, we already get $C = U_\ell$ i.e. $\mathrm{KL}(X, B||X, C) = \ell - \mathrm{H}_{\mathrm{sh}}(B|X) \leq \ell$. Thus it only makes sense to talk about KL-hardness for $\delta \leq \ell$.

Finally, we show that the two notions, KL-hard and KL-hard for sampling, are equivalent up to a polynomial factor in t, provided that ℓ is logarithmic in n:

Lemma 4.36. *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. If B is nonuniformly (t, δ) KL-hard for sampling given X, then B is nonuniformly $(\Omega(t/2^\ell), \delta)$ KL-hard given X. Conversely, if B is nonuniformly (t, δ) KL-hard given X, then B is nonuniformly $(t', \delta - \epsilon)$ KL-hard for sampling given X for $t' = t/\mathrm{poly}(n, 2^\ell, 1/\epsilon)$, for every $\epsilon > 0$.*

Proof. Suppose B is not nonuniformly (t', δ) KL-hard given X. That is, there exists a size t' circuit $P : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ such that $\mathrm{KL}(X, B||X, \Phi_P) \leq \delta$. Then we can sample

$S(x) = a$ w.p. $\Pr[\Phi_P = a|X = x]$ so that $\mathrm{KL}(X, B||X, S(X)) \leq \delta$. S has circuit size $O(2^\ell \cdot t')$. This contradicts the fact that B is nonuniformly (t, δ) KL-hard for sampling, for $t' = \Omega(t/q)$.

Conversely, suppose $\mathrm{KL}(X, B||X, S(X)) \leq \delta - \epsilon$ for some size t' circuit S. We will construct a size t *randomized* δ-KL predictor P (so that it will be useful for the uniform setting, Lemma 4.37, as well) as follows. We compute $E(x, a)$ such that w.p. at least $1 - \gamma$, $|\Pr[S(x) = a] - E(x, a)| \leq \epsilon^2/c^2 2^\ell$ for all x, a, where c is a large enough constant. This is done by taking $m = O\left(n + \ell + \log(1/\gamma)\right) \cdot 2^{2\ell}/\epsilon^4$ samples of the randomness of S. We then output $P(x, a) = \max\{E(x, a), \epsilon/c2^\ell\} \in (\epsilon/c2^\ell, 1]$.

We view P as a distribution over functions $p : \{0,1\}^n \times \{0,1\}^\ell \to (\epsilon/c2^\ell, 1]$. Consider any $p \in \mathrm{supp}(P)$ such that$|\Pr[S(x) = a] - E(x, a)| \leq \epsilon^2/c^2 2^\ell$ for all x, a. Notice that $\sum_b p(x, b) \leq 1 + 2^\ell \cdot \left(\epsilon/c2^\ell\right) = 1 + \epsilon/c$. If $\Pr[S(x) = a] > \epsilon/c2^\ell$, then

$$\log \frac{\Pr[S(x) = a]}{\Pr[\Phi_p = a|X = x]} \leq \log \frac{p(x, a) + \frac{\epsilon^2}{c^2}}{p(x, a)} + \log \sum_b p(x, b) \leq \log(1 + \frac{\epsilon}{c}) + \log(1 + \frac{\epsilon}{c}) \leq \frac{\epsilon}{2}.$$

If $\Pr[S(x) = a] \leq \epsilon/c2^\ell$, then

$$\log \frac{\Pr[S(x) = a]}{\Pr[\Phi_p = a|X = x]} = \log \frac{\Pr[S(x) = a]}{p(x, a)} + \log \sum_b p(x, b) \leq \log(1 + \epsilon/c) \leq \frac{\epsilon}{2}.$$

Thus we get

$$\begin{aligned}
&\mathrm{KL}(X, B||X, \Phi_p) \\
&= \mathrm{KL}(X, B||X, S(X)) + \mathop{\mathbb{E}}_{x \sim X}\left[\sum_a \Pr[B = a|X = x] \log \frac{\Pr[S(x) = a]}{\Pr[\Phi_p = a|X = x]}\right] \\
&\leq \delta - \epsilon + \frac{\epsilon}{2}.
\end{aligned}$$

On the other hand, for every $p : \{0,1\}^n \times \{0,1\}^\ell \to (\epsilon/c2^\ell, 1]$ it holds that

$$\begin{aligned}
\mathrm{KL}(X, B||X, \Phi_p) &= \mathbb{E}\left[\sum_a \Pr[B = a|X = x] \log\left(\Pr[B = a|X = x]/\Pr[\Phi_p = a|X = x]\right)\right] \\
&\leq \max_{x,a} \log\left(1/\Pr[\Phi_p = a|X = x]\right) = O\left(\ell + \log\frac{1}{\epsilon}\right).
\end{aligned}$$

Thus,

$$\mathop{\mathbb{E}}_{p\sim P}\left[\mathrm{KL}(X,B||X,\Phi_p)\right] \leq (1-\gamma)\cdot(\delta-\frac{\epsilon}{2})+\gamma\cdot O\left(\ell+\log\frac{1}{\epsilon}\right)\leq\delta,$$

for an appropriate choice of $\gamma = O(\epsilon/(\ell+\log(1/\epsilon)))$. Furthermore, P has circuit size $O(t'm) = t$. Thus B is not nonuniformly (t,δ) KL-hard given X. □

Lemma 4.37. *Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $p = p(n)$, $\epsilon = \epsilon(n) > 0$, $\ell = \ell(n)$ all computable in time* poly(n)*. Let (X,B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. If B is uniformly (t,δ) KL-hard for sampling given X, then B is uniformly $(\Omega(t/(2^\ell+n)),\delta)$ KL-hard given X. Conversely, if B is uniformly (t,δ) KL-hard given X, then B is uniformly $(t',\delta-\epsilon)$ KL-hard for sampling given X, for $t' = t/\mathrm{poly}(n,2^\ell,1/\epsilon)$.*

Proof. The proof for the second part is identical to Lemma 4.36. For the first part, suppose B is not uniformly (t',δ) KL-hard given X. That is, there is a time t' oracle algorithm P such that when $P^{O_{X,B}}$ is viewed as a distribution over functions $p : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$, for infinitely many n,

$$\mathbb{E}_{p\sim P^{O_{X,B}}}\left[\mathrm{KL}(X,B||X,\Phi_p)\right]\leq\delta.$$

Then we can sample $S(x) = a$ w.p. $\mathbb{E}_{p\sim P^{O_{X,B}}}\left[\Pr[\Phi_p = a|X = x]\right]$, where we first pick $p \sim P^{O_{X,B}}$ by fixing the internal coin tosses of P and samples from oracle $O_{X,B}$. By convexity of $\mathrm{KL}(X,B||X,\cdot)$,

$$\mathrm{KL}(X,B||X,S(X)) = \mathrm{KL}\left(X,B||X,\Phi_{P^{O_{X,B}}}\right)\leq\mathbb{E}_{p\sim P^{O_{X,B}}}\left[\mathrm{KL}\left(X,B||X,\Phi_p\right)\right]\leq\delta.$$

This contradicts the fact that B is uniformly (t,δ) KL-hard for sampling, for $t' = \Omega(t/(2^\ell+n))$. □

4.3.2 Main Results

We show that a distribution B having pseudoentropy given X, is equivalent to B being KL-hard given X (a notion which, as discussed above, captures the computational hardness

of sampling B given X in terms of KL divergence). We prove the equivalence in both nonuniform and uniform models of computation.

Theorem 4.38 (Characterizing pseudoentropy, nonuniform setting)**.** *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$, $\delta > 0$, $\epsilon > 0$.*

1. *If B is nonuniformly (t, δ) KL-hard given X, then for every $\epsilon > 0$, B has nonuniform (t', ϵ) pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$ given X, for $t' = t^{\Omega(1)}/\mathrm{poly}(n, \ell, 1/\epsilon)$.*

2. *Conversely, if B has nonuniform (t, ϵ) pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X, then for every $\sigma > 0$, B is nonuniformly (t', δ') KL-hard given X, for $t' = \min\{t^{\Omega(1)}/\mathrm{polylog}\,(1/\sigma), \Omega(\sigma/\epsilon)\}$ and $\delta' = \delta - \sigma$.*

Corollary 4.39. *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. Then B has nonuniform pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X if and only if B is nonuniformly δ KL-hard given X.*

By dropping X, the polynomial dependence on ℓ gives us a characterization of *nonuniform pseudoentropy* for an ℓ-bit distribution: (Note that without conditioning on X, the definition of KL-hard still makes sense, expressing the hardness of computing a measure that approximates the distribution B.)

Corollary 4.40. *A distribution B on $\{0,1\}^\ell$ has nonuniform pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B) + \delta$ if and only if B is nonuniformly δ KL-hard.*

We now state the uniform versions of our results, which are analogous to the nonuniform versions but have an *exponential* dependence on ℓ (we do not know whether it can be made polynomial like in Theorem 4.38, so we don't have a uniform analogue of Corollary 4.40.)

Theorem 4.41 (Characterizing pseudoentropy, uniform setting)**.** *Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $\epsilon = \epsilon(n) > 0$, $\ell = \ell(n)$, $\sigma = \sigma(n)$ all computable in time* poly(n)*. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$.*

1. *If B is uniformly (t, δ) KL-hard given X, then B has uniform (t', ϵ) pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$ given X, for $t' = t^{\Omega(1)}/\mathrm{poly}(n, 2^\ell, 1/\epsilon)$.*

2. *Conversely, if B has uniform (t, ϵ) pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X, then B is uniformly (t', δ') KL-hard given X, for $t' = \min\{t^{\Omega(1)}/\mathrm{poly}(n, \log(1/\sigma), \Omega(\sigma/\epsilon)\}$ and $\delta' = \delta - \sigma$.*

Corollary 4.42. *Let n be a security parameter, $\delta = \delta(n) > 0$, $\ell = \mathrm{polylog}(n)$ computable in time* poly(n)*. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. Then B has uniform pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X if and only if B is uniformly δ KL-hard given X.*

Note that we do not make any samplability assumption on X (in both nonuniform and uniform settings).

4.3.3 A Generic Framework

In this section, we provide a generic framework for proving statements such as the characterization of (conditional) pseudoentropy, yielding our meta characterization theorem (Theorem 4.8).

Throught the section, we consider an arbitrary strictly concave function

$$H : \{\text{distributions on } \Sigma\} \to \mathbb{R}_{\geq 0}$$

that is differentiable in the interior of the simplex in $|\Sigma|$-space (e.g. the Shannon entropy of a distribution on Σ). We also extend H to the conditional setting in the natural way: For

a joint distribution (X, C), define $H(C|X)$ to equal $\mathbb{E}_{x \leftarrow X}\left[H\left(C|_{X=x}\right)\right]$ (e.g. the conditional Shannon entropy of C given X).

A key to our framework is interpreting the Bregman divergence associated with H, $\mathrm{D}_H(X \parallel Y)$ (see Definition 1.7), in terms of the distinguishing advantage of a special distinguisher W:

Lemma 4.43. *Let X and Y be distributions on a finite set Σ. Let $W : \Sigma \to \mathbb{R}_{\geq 0}$ be the function*

$$W(x) = -\frac{\partial H(Y)}{\partial \Pr[Y = x]}$$

where the LHS denotes the derivative of H at Y w.r.t. $\Pr[Y = x]$. Then

$$\mathrm{D}_H(X \parallel Y) = H(Y) - H(X) - \left(\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]\right).$$

Moreover, $Y' = Y$ maximizes $\mathbb{E}[W(Y')]$ over all distributions Y' on Σ where $H(Y') \geq H(Y)$.

For some intuition, suppose W is a constant function and H is Shannon entropy. Then $C^* = U_\ell$ and Lemma 4.44 becomes the familiar identity

$$\mathrm{KL}(X \parallel U_\ell) = \ell - \mathrm{H}_{\mathrm{sh}}(X).$$

Proof of Lemma 4.43. By definition of Bregman divergence,

$$\begin{aligned}
\mathrm{D}_H(X \parallel Y) &= H(Y) - H(X) - \langle \nabla H(Y), Y - X \rangle \\
&= H(Y) - H(X) + \sum_x W(x) \cdot \Pr[Y = x] - \sum_x W(x) \cdot \Pr[X = x] \\
&= H(Y) - H(X) - \left(\mathbb{E}[W(X)] - \mathbb{E}[W(Y)]\right).
\end{aligned}$$

Nonnegativity of Bregman divergence (Proposition 1.8) and the above equality together imply if $\mathbb{E}[W(Y')] > \mathbb{E}[W(Y)]$ then $H(Y') < H(Y)$, i.e. $Y' = Y$ maximizes $\mathbb{E}[W(Y')]$ subject to $H(Y') \geq H(Y)$. □

In our applications it is often more convenient to consider an average-case version, which follows as an immediate corollary of Lemma 4.43:

Corollary 4.44. *Let (X, B) and (X, C) be joint distributions on $\{0,1\}^n \times \{0,1\}^\ell$. Let $W : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$ be the function*

$$W(x, a) = -\frac{\partial H(C|X)}{\partial \Pr[C = a|X = x]}$$

where the LHS denotes the derivative of $H(C|X)$ w.r.t. $\Pr[C = a|X = x]$. Then

$$\mathrm{D}_H(X, B \parallel X, C) = H(C|X) - H(B|X) - (\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]) .$$

Moreover, $(X, C') = (X, C)$ maximizes $\mathbb{E}[W(X, C')]$ subject to $H(C'|X) \geq H(C|X)$ (over all C' jointly distributed with X).

Now we show why the lemma is useful: If W achieves good distinguishing advantage $\mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]$ for *all* C where $H(C|X)$ is not too much larger than $H(B|X)$, then we can use C^* to approximate B within small Bregman divergence:

Lemma 4.45. *Let (X, B) be any joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$, $\epsilon > 0$, and $\delta > 0$. Let $W^* : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ be a function such that $\mathbb{E}[W^*(X, B)] - \mathbb{E}[W^*(X, C)] \geq \epsilon$ for all ℓ-bit random variables C jointly distributed with X with $H(C|X) \geq H(B|X) + \delta$. Suppose that there exists a joint distribution (X, C^*) such that $H(C^*|X) = H(B|X) + \delta$ and for some constant $\mu \geq 0$,*

$$\forall x, a, \quad \frac{\partial H(C^*|_{X=x})}{\partial \Pr[C^* = a|X = x]} = -\mu \cdot W^*(x, a).$$

Then

$$\mathrm{D}_H(X, B \parallel X, C^*) \leq \delta.$$

Proof. By Corollary 4.44,

$$\begin{aligned}\mathrm{D}_H(X,B \parallel X,C^*) &= H(C^*|X) - H(B|X) - \mu \cdot (\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C^*)]) \\ &\leq \delta - \mu \cdot \epsilon \leq \delta.\end{aligned}$$

□

Lemma 4.43 can also be used to show a "converse" to Lemma 4.45.

Lemma 4.46. *Let (X,B) and (Y,C^*) be joint distributions on $\{0,1\}^n \times \{0,1\}^\ell$, $\epsilon > 0$, $\delta > 0$. Suppose $\mathrm{D}_H(X,B \parallel Y,C^*) \leq \delta$. Then for the function W^* defined as*

$$W^*(x,a) = -\frac{\partial H(C^*|X)}{\partial \Pr[C^* = a|X = x]},$$

and for all joint distributions (Y,C) on $\{0,1\}^n \times \{0,1\}^\ell$ with $H(Y,C) \geq H(X,B) + \delta + \epsilon$, we have $\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(Y,C)] \geq \epsilon$.

Proof. Consider any joint distribution (Y,C) with $H(Y,C) \geq H(X,B) + \delta + \epsilon$. Applying Lemma 4.43, we obtain

$$H(Y,C^*) - H(X,B) - (\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(Y,C^*)]) = \mathrm{D}_H(X,B||Y,C^*) \leq \delta,$$

as well as

$$H(Y,C^*) - H(Y,C) - (\mathbb{E}[W^*(Y,C)] - \mathbb{E}[W^*(Y,C^*)]) = \mathrm{D}_H(Y,C||Y,C^*) \geq 0.$$

Together they yield

$$\begin{aligned}&\mathbb{E}[W^*(X,B) - W^*(Y,C)] \\ &= \mathbb{E}[W^*(X,B) - W^*(Y,C^*)] - \mathbb{E}[W^*(Y,C) - W^*(Y,C^*)] \\ &= H(Y,C^*) - H(X,B) - \mathrm{D}_H(X,B \parallel Y,C^*) - (H(Y,C^*) - H(Y,C) - \mathrm{D}_H(Y,C \parallel Y,C^*)) \\ &\geq H(Y,C) - H(X,B) - \delta \\ &\geq \epsilon.\end{aligned}$$

□

A consequence of Lemma 4.45 and 4.46 is the following meta characterization:

Informal Theorem 4.47 (Meta characterization theorem)**.** *Let n be a security parameter, $\delta = \delta(n)$, $\ell = \ell(n)$, and $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. For all "sufficiently nice" H, the following are equivalent:*

1. *There exists a joint distribution (X, C) where (X, C) and (X, B) are indistinguishable by poly-sized circuits, and $H(C|X) \geq H(B|X) + \delta - 1/n^{\omega(1)}$;*

2. *For all poly-sized circuits that "represent" a joint distribution (X, C^*), we have $\mathrm{D}_H(X, B \parallel X, C^*) > \delta - 1/n^{\omega(1)}$.*

We now outline how we would obtain an actual proof for such a theorem for a given H. We instantiate this approach for H being Shannon entropy in Section 4.3.4 below.

- Suppose that Item 1 of the meta theorem is false, i.e. for some constant $c > 0$, for every C where $H(C|X) \geq H(B|X) + \delta - 1/n^c$ there is a poly-sized $(1/n^c)$-distinguisher between (X, B) and (X, C). By the Nonuniform Min-Max Theorem (Chapter 2 Theorem 2.3), there is a poly-sized circuit W^* that $(1/n^c)$-distinguishes (X, B) from *all* (X, C) where $H(C|X) \geq H(B|X) + \delta - 1/n^c$. We assume H is "sufficiently nice" so that for some constant $\mu \geq 0$, there exists a joint distribution (X, C^*) satisfying

$$\forall x, a, \quad \frac{\partial H(C^*|_{X=x})}{\partial \Pr[C^* = a | X = x]} = -\mu \cdot W^*(x, a)$$

  and $H(C^*|X) \geq H(B|X) + \delta - 1/n^c$, and can be "represented" by a poly-sized circuit (which is often true, since W^* is a poly-sized circuit). Now Lemma 4.45 implies that $\mathrm{D}_H(X, B \parallel X, C^*) \leq \delta - 1/n^c$, and we conclude that Item 2 of the meta theorem is false.

- Suppose that Item 2 of the meta theorem is false, i.e. for some constant $c > 0$ and some (X, C^*) that can be "represented" by a poly-sized circuit, $\mathrm{D}_H(X, B \parallel X, C^*) \leq \delta - 1/n^c$. Assuming H is "sufficiently nice" so that (i) the W^* defined in Lemma 4.46 can be (approximately) computed by poly-sized circuit, and (ii) $|W^*(x, a)| \leq \mathrm{poly}(n)$, then W^* gives rise to a universal distinguisher between (X, B) and (X, C) for all (X, C) where $H(C|X) \geq H(B|X) + \delta - 1/n^c$, by Lemma 4.46. This concludes that Item 1 is false.

4.3.4 KL-hardness Implies Pseudoentropy, Nonuniform Setting

In this section, we prove one (the more interesting) direction of the characterization of nonuniform pseudoentropy (Theorem 4.38).

Consider the distribution (X, C^*) in Lemma 4.44 with $H = \mathrm{H}_{\mathrm{sh}}$. Given a function $W : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$, it is easy to verify that the hypothesis of Lemma 4.44 is satisfied by (X, C^*) defined as

$$C^*(a|x) = \frac{e^{W(x,a)}}{\sum_b e^{W(x,b)}}.$$

We denote such (X, C^*) by $(X, \mathbf{e}^W)$.

Proposition 4.48. *Let $W : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$. Then*

$$\forall x, a, \quad \frac{\partial H(\mathbf{e}^W|_{X=x})}{\partial \Pr[\mathbf{e}^W = a|X = x]} = -W(x, a).$$

Remark. $(X, \mathbf{e}^W)$ is a conditional version of the *Boltzmann distribution* (or *Gibbs distribution*; *canonical ensemble*) in statistical physics [LL], which is the unique distribution that achieves maximum entropy under a linear constraint on the pmf. We consider the conditional Boltzmann distribution in our context for a similar reason: for any distinguisher

W, $(X, C) = (X, \mathbf{e}^{kW})$ $(k \geq 0)$ minimizes $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)]$ among all C with $\mathrm{H}_{\mathrm{sh}}(C|X) \geq r = \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X)$. (The unconditional version is well known in statistical physics [LL].)

We apply Lemma 4.45 to show the following result about $(X, \mathbf{e}^{W})$, which captures the essence of why KL-hardness implies pseudoentropy:

Lemma 4.49. *Let (X,B) be any joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$, $\epsilon > 0$, and $0 < \delta \leq \ell - \mathrm{H}_{\mathrm{sh}}(B|X)$. Let $W : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ be a function such that $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)] \geq \epsilon$ for all ℓ-bit random strings C jointly distributed with X with $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$. Then there exists $k \in [0, \ell/\epsilon]$ such that $\mathrm{KL}(X, B \parallel X, \mathbf{e}^{kW}) \leq \delta$.*

Proof. Let $k_0 = \ell/\epsilon$. First we show that there exists $k \in [0, k_0]$ such that $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X) = \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$. By Lemma 4.44 and Proposition 4.48,

$$\begin{aligned}\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,\mathbf{e}^{k_0 W})] &= \frac{\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_0 W}|X) - \mathrm{H}_{\mathrm{sh}}(B|X) - \mathrm{KL}(X, B \parallel X, \mathbf{e}^{k_0 W})}{k_0} \\ &< \frac{\ell}{k_0} = \epsilon,\end{aligned}$$

where we use nonnegativity of H_{sh} and Bregman divergence. Thus, by assumption we must have $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_0 W}|X) < \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$. Now note that (i) $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{0 \cdot W}|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ since $\mathbf{e}^{0 \cdot W}$ is simply the uniform distribution; (ii) $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_0 W}|X) < \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$; (iii) $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X)$ is continuous as a function of $k \in [0, +\infty)$. By the Intermediate Value Theorem, there exists $k \in [0, k_0]$ such that $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X) = \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$.

The result now follows from Lemma 4.45 and Proposition 4.48. □

Theorem 4.50 (KL-hardness $\Longrightarrow$ pseudoentropy, nonuniform setting)**.** *Let (X,B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$, $\delta > 0$. If B is nonuniformly (t,δ) KL-hard given X, then for every $\epsilon > 0$, B has nonuniform (t', ϵ) pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$ given X for $t' = t^{\Omega(1)}/\mathrm{poly}\,(n, 1/\epsilon, \ell)$.*

Proof. Suppose for contradiction that B does not have nonuniform (t', ϵ) conditional pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$. By definition, for any ℓ-bit random string C jointly distributed with X where $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$, there is a size t' deterministic circuit W such that $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C))] \geq \epsilon$.

Consider the two-player zero-sum game where Player 1 selects some joint distribution (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ where $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$, Player 2 selects some deterministic circuit $W : \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}$ of size t', and receives expected payoff $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,C)]$. Note that by concavity of H_{sh}, any mixed strategy for Player 1 is still some joint distribution (X, C) with $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$. Since for all Player 1 mixed strategies (X, C) there exists a Player 2 strategy W with expected payoff at least ϵ, by the Nonuniform Min-Max Theorem (Theorem 2.3) there is some size $O(S \cdot t')$ randomized circuit W^*, for $S = O(\ell/\epsilon^2)$, such that

$$\mathbb{E}[W^*(X,B)] - \mathbb{E}[W^*(X,C)] \geq .9\epsilon$$

for all (X, C) with $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$. By Lemma 4.49, there exists $k \in [0, O(\ell/\epsilon)]$ such that $\mathrm{KL}(X, B||X, \mathbf{e}^{kW^*}) \leq \delta - \epsilon$. In other words, $P(x,a) = \exp(k \cdot W^*(x,a))$ is a $(\delta - \epsilon)$-KL-predictor. Thus it remains to show that $P(x,a) = \exp(k \cdot W^*(x,a))$ can be computed efficiently (within small error).

Efficiency. We approximate k by some rational $\tilde{k}$ to $\Theta(\epsilon/c)$ precision for a sufficiently large constant c, so that $\forall x,\ a,\ \left|\tilde{k} \cdot W^*(x,a) - k \cdot W^*(x,a)\right| \leq \epsilon/c$. Since $\tilde{k} \cdot W^*(x,a)$ is rational valued, we can use Newton's method to construct a circuit P approximating $\exp(\tilde{k} \cdot W^*(x,a))$. This can be done in such a way that

$$\mathrm{KL}\left(X, B||X, \Phi_P\right) \leq \mathrm{KL}(X, B||X, \mathbf{e}^{kW^*}) + \epsilon \leq \delta$$

and P has circuit size $t = \text{poly}(t', n, 1/\epsilon, \ell)$. See Lemma A.6 for details. This contradicts the hypothesis that B is nonuniformly (t, δ) KL-hard given X. □

4.3.5 KL-hardness Implies Pseudoentropy, Uniform Setting

To prove the uniform complexity version of Theorem 4.50, we replace the use of the Nonuniform Min-Max Theorem in the proof of Theorem 4.50 with the Uniform Min-Max Theorem – Average Case (Chapter 2, Theorem 2.5).

Notation. We denote by $\mathcal{V}_r(X)$ the set of all joint distributions (X, C) on $\{0,1\}^n \times \{0,1\}^\ell$ (where C may vary and X is fixed) such that $\mathrm{H}_{\mathrm{sh}}(C|X) \geq r$.

To implement Chapter 2, Algorithm 2.2 (Finding Universal Strategy – Average Case), we need to compute σ-approximate KL projections on the conditional entropy ball $\mathcal{V}_r(X)$.

4.3.5.1 Approximating KL Projection on High Conditional Entropy Distributions

In this section we describe how to efficiently find (X, C) as a σ-approximate KL projection of (X, C') on $\mathcal{V}_r(X)$. We first describe in Lemma 4.51 the exact KL projection of the joint distribution (X, C) on a conditional entropy ball $\mathcal{V}_r(X)$, then show how to approximate it.

By definition of KL projection we need to find some $(X, C') \in \mathcal{V}_r(X)$ minimizing $\mathrm{KL}(X, C' \parallel X, C)$. Recall that for a function $W : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$, $k \in \mathbb{R}$, we denote by $\mathbf{e}^{kW}$ the distribution jointly distributed with X such that:

$$\mathbf{e}^{kW}(a|x) = \frac{e^{kW(x,a)}}{\sum_b e^{kW(x,b)}}.$$

Now, if we write (X, C) as $(X, \mathbf{e}^W)$ for a function $W : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$, then by Lemma 4.44 and Proposition 4.48 we can minimize $\mathrm{KL}(X, C' \parallel X, \mathbf{e}^W)$ by maximizing

$\mathbb{E}[W(X,C')] - \mathbb{E}[W(X,\mathbf{e}^W)]$, assuming that the entropy difference is fixed. This is the idea of Lemma 4.51 below.

Lemma 4.51 (KL projection on a conditional entropy ball $\mathcal{V}_r(X)$)**.** *Let (X,C) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$ such that $\Pr[C=b|X=x] > 0$ for all x, b. Define*

$$W(x,a) = \log \frac{\Pr[C=a|X=x]}{\min_b \{\Pr[C=b|X=x]\}}$$

so that $(X,C) = (X,\mathbf{e}^W)$. Then for every $r \le \ell$, the KL projection of (X,C) on $\mathcal{V}_r(X)$ equals $(X,\mathbf{e}^{\alpha W})$ for some $\alpha \in (0,1]$ such that $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) \ge r$. (In fact $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) = r$ as long as $(X,C) \notin \mathcal{V}_r(X)$).

Proof. First, if $(X,C) \in \mathcal{V}_r(X)$ then the KL projection is $(X,C) = (X,\mathbf{e}^W)$ itself, i.e. $\alpha = 1$.

To find the KL projection for $(X,C) \notin \mathcal{V}_r(X)$, we first note there exists $\alpha \in (0,1)$ such that $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) = r$ (by the Intermediate Value Theorem, because $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^W|X) < r$, $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{0\cdot W}|X) = \ell \ge r$ and $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X)$ is continuous as a function of $k \in (0,1)$). By definition of KL projection, we want to minimize $\mathrm{KL}(X,C'||X,\mathbf{e}^W)$ over all C' where $\mathrm{H}_{\mathrm{sh}}(C'|X) = r$ (as KL projection is always on the boundary of $\mathcal{V}_r(X)$; see Lemma A.4). Now by Lemma 4.44 and Proposition 4.48,

$$\mathrm{KL}(X,C'||X,\mathbf{e}^W) = \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^W|X) - \mathrm{H}_{\mathrm{sh}}(C'|X) - \left(\mathbb{E}[W(X,C')] - \mathbb{E}\,W(X,\mathbf{e}^W)]\right).$$

So minimizing $\mathrm{KL}(X,C'||X,\mathbf{e}^W)$ is equivalent to maximizing $\mathbb{E}[W(X,C')] - \mathbb{E}\,W(X,\mathbf{e}^W)]$, and the result follows from 4.44. □

Lemma 4.52 (Approximating KL projection on a conditional entropy ball $\mathcal{V}_r(X)$)**.** *There exists a $\mathrm{poly}(\kappa, n, 2^\ell, 1/\sigma, \log(1/\gamma))$ time randomized algorithm Π such that given oracle access to a function $W : \{0,1\}^n \times \{0,1\}^\ell \to [0,\kappa]$ and O_X, for all $\sigma > 0$ and $0 \le r \le \ell - \sigma$, $\Pi^{W,O_X}(r,\sigma)$ outputs w.p. $1-\gamma$ some $\beta \in (0,1]$ such that $(X,\mathbf{e}^{\beta W})$ is a σ-approximate KL projection of $(X,\mathbf{e}^W)$ on $\mathcal{V}_r(X)$, and β has bit length $\log(\kappa/\sigma) + \log\ell + O(1)$.*

Proof. Our algorithm Π works as follows. Π computes an estimate $E_\beta \in [\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X) \pm \sigma/6]$ for every discrete β ranging from 0 to 1 in steps of $\sigma/(c\kappa\ell)$ for some sufficiently large constant c. This can be done in time $\mathrm{poly}(\kappa, n, 2^\ell, 1/\sigma, \log(1/\gamma))$, and w.p. $1-\gamma$ after a union bound over all $c\kappa\ell/\sigma$ values of β; see Lemma A.6 for details. Π then outputs any discrete β (as a multiple of $\sigma/(c\kappa\ell)$) satisfying $E_\beta \in [r+\sigma/6, r+5\sigma/6]$, and outputs 1 if no such β is found. We now argue the correctness in two cases.

Case 1: No discrete β satisfies $E_\beta \in [r+\sigma/6, r+5\sigma/6]$. We show that this happens only when $\left(X, \mathbf{e}^W\right) \in \mathcal{V}_r(X)$ i.e. the KL projection is itself, thus Π is correct in outputting 1. Indeed, suppose that $\left(X, \mathbf{e}^W\right) \notin \mathcal{V}_r(X)$. One can check that any $\sigma/(c\kappa\ell)$ variation in β causes at most $\sigma/3$ variation in $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X)$ (Lemma A.7). Since $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{0W}|X) = \ell \geq r+\sigma$ and $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{1W}|X) < r$, a discrete Intermediate Value Theorem says there exists a discrete $\beta \in [0,1]$ with $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X) \in [r+\sigma/3, r+2\sigma/3]$. In other words there exists β satisfying $E_\beta \in [r+\sigma/6, r+5\sigma/6]$.

Case 2: There exists some β satisfying $E_\beta \in [r+\sigma/6, r+5\sigma/6]$. Consider any such β. Closeness of E_β to both r and $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X)$ guarantees that

$$r \leq \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X) \leq r+\sigma.$$

Thus $(X, \mathbf{e}^{\beta W}) \in \mathcal{V}_r(X)$. Recall from Lemma 4.51 that the exact KL projection of $(X, \mathbf{e}^W)$ on $\mathcal{V}_r(X)$ equals $(X, \mathbf{e}^{\alpha W})$ where $\alpha = 1$ if $(X, \mathbf{e}^W) \in \mathcal{V}_r(X)$, or $0 < \alpha < 1$ and $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}) = r$ if $(X, \mathbf{e}^W) \notin \mathcal{V}_r(X)$. We need to show that $(X, \mathbf{e}^{\beta W})$ is a σ-approximate KL projection. By Pythagorean Theorem (Theorem 1.13) it suffices to show that for all $(X, C) \in \mathcal{V}_r(X)$,

$$\mathrm{KL}(X, C||X, \mathbf{e}^{\beta W}) - \mathrm{KL}(X, C||X, \mathbf{e}^{\alpha W}) \leq \sigma.$$

By Lemma 4.44,

$$\begin{aligned}
&\mathrm{KL}(X, C||X, \mathbf{e}^{\beta W}) - \mathrm{KL}(X, C||X, \mathbf{e}^{\alpha W}) \\
&= \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) \\
&\quad - \beta \left(\mathbb{E}[W(X, C)] - \mathbb{E}[W(X, \mathbf{e}^{\beta W})]\right) + \alpha \left(\mathbb{E}[W(X, C)] - \mathbb{E}[W(X, \mathbf{e}^{\alpha W})]\right) \\
&\leq (r + \sigma) - r - \beta \left(\mathbb{E}[W(X, C)] - \mathbb{E}[W(X, \mathbf{e}^{\beta W})]\right) + \alpha \left(\mathbb{E}[W(X, C)] - \mathbb{E}[W(X, \mathbf{e}^{\alpha W})]\right) \\
&= \sigma + (\alpha - \beta)\, \mathbb{E}\left[W(X, C)\right] + \beta \cdot \mathbb{E}[W(X, \mathbf{e}^{\beta W})] - \alpha \cdot \mathbb{E}\left[W(X, \mathbf{e}^{\alpha W})\right].
\end{aligned}$$

Note that $\alpha \geq \beta$, because either $\alpha = 1 \geq \beta$ (when $(X, \mathbf{e}^W) \in \mathcal{V}_r(X)$), or $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) = r \leq \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X)$ (when $(X, \mathbf{e}^W) \notin \mathcal{V}_r(X)$) and it follows from monotonicity of $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X)$ as a function of k in $[0, +\infty)$ (Lemma A.5). Thus by Item 2 of Lemma 4.44, $(\alpha - \beta)\, \mathbb{E}\left[W(X, C)\right] \leq (\alpha - \beta)\, \mathbb{E}\left[W(X, \mathbf{e}^{\alpha W})\right]$, and the above inequality becomes

$$\mathrm{KL}(X, C||X, \mathbf{e}^{\beta W}) - \mathrm{KL}(X, C||X, \mathbf{e}^{\alpha W}) \leq \sigma + \beta \left(\mathbb{E}\left[W(X, \mathbf{e}^{\beta W})\right] - \mathbb{E}\left[W(X, \mathbf{e}^{\alpha W})\right]\right).$$

Now applying Lemma 4.44 again yields

$$\begin{aligned}
&\alpha \left(\mathbb{E}[W(X, \mathbf{e}^{\beta W})] - \mathbb{E}[W(X, \mathbf{e}^{\alpha W})]\right) \\
&= \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X) - \mathrm{KL}(X, \mathbf{e}^{\beta W}||X, \mathbf{e}^{\alpha W}) \\
&\leq \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\alpha W}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{\beta W}|X) \leq 0,
\end{aligned}$$

where we used nonnegativity of KL divergence. Therefore

$$\mathrm{KL}(X, C||X, \mathbf{e}^{\beta W}) - \mathrm{KL}(X, C||X, \mathbf{e}^{\alpha W}) \leq \sigma.$$

□

4.3.5.2 Putting it Together

We now have all the tools ready to prove Theorem 4.41 (KL hardness implies pseudoentropy, uniform setting). We just will replace the use of the Min-Max Theorem in the

proof of Theorem 4.50 with the Uniform Min-Max Theorem for distinguishers (Chapter 2, Theorem 2.5), using Lemma 4.52 to implement the approximate KL projection. However, notice that $\mathrm{H}_{\mathrm{sh}}(B|X)$ hence the "radius" of the conditional entropy ball $\mathcal{V}_r(X)$ is unknown. We will simply try all radii (with quantization) and pick the distinguisher that results in the best KL predictor, which can be tested by sampling (X, B).

Theorem 4.53 (KL-hardness $\Longrightarrow$ pseudoentropy, uniform setting)**.** *Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $\epsilon = \epsilon(n) > 0$, $\ell = \ell(n)$ all computable in time* poly(n)*. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. If B is uniformly (t, δ) KL-hard given X, then B has uniform (t', ϵ) pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$ given X, for $t' = t^{\Omega(1)}/\mathrm{poly}(n, 2^\ell, 1/\epsilon)$.*

Proof. Suppose for contradiction that B does not have uniform (t', ϵ) conditional pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X)+\delta-\epsilon$. By definition, there is a time t' randomized oracle algorithm A such that for infinitely many n and every C with $\mathrm{H}_{\mathrm{sh}}(C|X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$, $A^{O_{X,B,C}}$ is an ϵ-distinguisher between (X, B) and (X, C).

Recall that in the nonuniform setting (Theorem 4.50), we begin by obtaining a universal distinguisher W^* using the Nonuniform Min-Max Theorem. Similarly, in the uniform setting, we first obtain a universal distinguisher *uniformly* using the Uniform Min-Max Theorem – Average Case (Chapter 2, Theorem 2.5), as captured in the following claim:

Claim 4.54. There is a randomized oracle algorithm Υ that, for any $r \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon/2$, $\Upsilon^{O_{X,B}}(r, n, \ell, t', \epsilon, \gamma)$ w.p. at least $1 - \gamma$ outputs some deterministic circuit W_r^* of size $\mathrm{poly}(t', n, \ell, 1/\epsilon, \log(1/\gamma))$ such that for all $(X, C) \in \mathcal{V}_r(X)$,

$$\mathbb{E}[W_r^*(X, B)] - \mathbb{E}[W_r^*(X, C)] \geq .9\epsilon.$$

Moreover, Υ runs in time $\mathrm{poly}(t', n, 2^\ell, 1/\epsilon, \log(1/\gamma))$.

Note that given Claim 4.54, the theorem almost follows by the same argument in the nonuniform setting (Theorem 4.50), by running $\Upsilon^{O_{X,B}}(r, n, \ell, t', \epsilon, \gamma)$ to obtain the desired universal distinguisher W^*, with $r = \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon/2$ and appropriate settings of γ. However, in general $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon$ is unknown (as $\mathrm{H}_{\mathrm{sh}}(B|X)$ may not be uniformly efficiently computable). To overcome this, we need to search for an appropriate value of r, which we settle after proving the claim.

Proof of Claim 4.54. Consider the two-player zero-sum game where Player 1 chooses some joint distribution $(X, C) \in \mathcal{V}_r(X)$, and Player 2 chooses a $\mathrm{poly}(t', n, \ell, 1/\epsilon, \log(1/\gamma))$ sized circuit W, with expected payoff $\mathbb{E}[f((X, C), W)] = \mathbb{E}[W(X, B)] - \mathbb{E}[W(X, C)]$ for Player 2. We will apply Theorem 2.5 (Uniform Min-Max Theorem – Average Case) to this game, i.e. with

- $\mathcal{V} = \mathcal{V}_r(X)$;
- $\mathcal{W} = \{\text{(deterministic) circuits of size } \mathrm{poly}(t', n, \ell, 1/\epsilon, \log(1/\gamma))\}$;
- $f((x, a), W) = \mathbb{E}[W(X, B)] - W(x, a)$.

We let the algorithm Υ be an instantiation of Chapter 2, Algorithm 2.2 (Finding Universal Strategy – Average Case) that we describe below for the game, with KL projection on the set $\mathcal{V} = \mathcal{V}_r(X)$. Then Υ outputs the deterministic circuit W_r^* that computes the average of $W^{(1)}, \ldots, W^{(S)}$. Using the oracle algorithm $A^{(\cdot)}$, we will show that in our instantiation of Algorithm 2.2, in each iteration we can obtain some $W^{(i)}$ that distinguishes (X, B) and (X, C). Thus, by the Uniform Min-Max Theorem – Average Case, W_r^* is indeed a universal distinguisher as desired.

Our instantiation of Algorithm 2.1 starts with an initial distribution $(X, C^{(1)})$ where $C^{(1)}$ is uniform on $\{0,1\}^\ell$ and independent of X. Let $\epsilon' = \epsilon/c$ for a sufficiently large constant

c. The number of iterations is $S = O(\ell/\epsilon'^2)$, and we let $\gamma' = \gamma/2S$. In each iteration we represent the joint distribution $(X, C^{(i)})$ by a circuit $W_i : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$ such that $(X, C^{(i)}) = (X, \mathbf{e}^{W_i})$. So we can take $W_1(x, a) = 0$ for all x, a. We show how to implement each of the S iterations of Algorithm 2.2 efficiently:

1. **Obtaining Player 2's Response $W^{(i)}$:** Suppose that we have constructed a t_i-size deterministic circuit $W_i : \{0,1\}^n \times \{0,1\}^\ell \to [0, \kappa_i]$. There are three steps:

   (a) Obtain a deterministic boolean circuit $M^{(i)}$ such that $(X, \Phi_{M^{(i)}})$ approximates $(X, C^{(i)}) = (X, \mathbf{e}^{W_i})$ in the following sense: (i) $\mathrm{H}(\Phi_{M^{(i)}}|X) \geq \mathrm{H}_{\mathrm{sh}}(C^{(i)}|X) - \epsilon'$; (ii) For every function $W' : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$,

   $$\begin{aligned} &\mathbb{E}[W'(X,B)] - \mathbb{E}[W'(X, C^{(i)})] \\ &\geq \mathbb{E}[W'(X,B)] - \mathbb{E}[W'(X, \Phi_{M^{(i)}})] - \epsilon'. \end{aligned}$$

   This can be done in time $\mathrm{poly}(t', t_i, n, \ell, \kappa_i, \log(1/\epsilon'))$ using Newton's method; see Lemma A.6 for details.

   (b) Generate $m = O(\log(1/\gamma')/\epsilon'^2)$ random samples of $(X, B, \Phi_{M^{(i)}})^{t'}$ and $U_{t'}$ (note that $\Phi_{M^{(i)}}$ is samplable using the circuit $M^{(i)}$). This can be done in time $\mathrm{poly}(t', t_i, n, 2^\ell, 1/\epsilon', \log(1/\gamma'))$.

   (c) Finally, let $W^{(i)}$ be the deterministic circuit that on input (x, a), runs $A^{(\cdot)}(x, a)$ for m times and returns the average of the m outputs. Each time, $A^{(\cdot)}(x, a)$ is run using one copy of $(X, B, \Phi_{M^{(i)}})^{t'}$ to answer oracle queries of A, and one copy of $U_{t'}$ as coin tosses of A. The m random samples are hardwired in $W^{(i)}$, thus $W^{(i)}$ is of size $t'' = O(t' \cdot m \cdot (n + \ell))$, which does not depend on the size of W_i (but the size of W_{i+1} will additively depend on t''). By a Chernoff bound, w.p. at

least $1-\gamma'$,

$$\begin{aligned}&\mathbb{E}[W^{(i)}(X,B)] - \mathbb{E}[W^{(i)}(X,\Phi_{M^{(i)}})]\\ &\geq \mathbb{E}[A^{O_{X,B,\Phi_{M^{(i)}}}}(X,B)] - \mathbb{E}[A^{O_{X,B,\Phi_{M^{(i)}}}}(X,\Phi_{M^{(i)}})] - \epsilon'.\end{aligned}$$

2. **Weight Update:** Note that $(X, C^{(i)'}) = (X, \mathbf{e}^{W_i+\epsilon'\cdot W^{(i)}})$, which is simply the consequence of multiplicative weight update. We represent $C^{(i)'}$ by the function $W_i' = W_i + \epsilon' \cdot W^{(i)}$.

3. **KL Projection:** We use Lemma 4.52 to efficiently obtain an ϵ'^2-approximate KL projection $(X, C^{(i+1)}) = (X, \mathbf{e}^{W_{i+1}})$ of $(X, C^{(i)'}) = (X, \mathbf{e}^{W_i'})$ on $\mathcal{V}_r(X)$, where $W_{i+1} = \beta_{i+1} \cdot W_i'$ for some $\beta_{i+1} \in (0,1]$ of bit length $O(\log(\ell\kappa_i/\epsilon'))$. This can be done in time $\mathrm{poly}(\kappa_i, t_i', n, 2^\ell, 1/\epsilon', \log(1/\gamma'))$ and w.p. at least $1-\gamma'$, where t_i' is the size of W_i'. Note that W_{i+1} is a $[0,\kappa_{i+1}]$-valued function with $\kappa_{i+1} = \kappa_i + \epsilon'$.

At last, Υ outputs the deterministic circuit W_r^* that computes the average of $W^{(1)}, \ldots, W^{(S)}$.

We argue that Υ runs in time $\mathrm{poly}(t', n, 2^\ell, 1/\epsilon, \log(1/\gamma))$ and outputs a circuit W_r^* of size $\mathrm{poly}(t', n, \ell, 1/\epsilon, \log(1/\gamma))$. Note that the way Weight Update and KL Projection are done guarantees that W_i' and W_{i+1} are always linear combinations of $W^{(1)}, \ldots, W^{(i)}$, with coefficients $\beta_1 \ldots \beta_j \epsilon'$ for some $0 \leq j \leq i$. Moreover, $W^{(1)}, \ldots, W^{(i)}$ have $\log m$ output bits (as the average of m boolean values), and each β_i ($i \leq S$) is of bit length $O(\log(\ell\kappa_i/\epsilon')) \leq O(\log(\ell i))$ (as $\kappa_i = (i-1)\epsilon'$). Thus, one can easily verify that W_{i+1} has circuit size $t_{i+1} \leq \mathrm{poly}(S\log(\ell S), \log m) + S\cdot t'' = \mathrm{poly}(t', n, \ell, 1/\epsilon', \log(1/\gamma'))$, and the same bound holds for t_i'. The total running time follows by plugging in t_{i+1} and t_i' in the running time of each step. Since W_r^* computes the average of $W^{(1)}, \ldots, W^{(S)}$ it has size $\mathrm{poly}(t', n, \ell, 1/\epsilon, \log(1/\gamma))$.

Now, suppose all S iterations complete successfully, which happens w.p. at least $1 - 2\gamma' S = 1-\gamma$, by a union bound. Since

$$\mathrm{H}(\Phi_{M^{(i)}}|X) \geq \mathrm{H}_{\mathrm{sh}}(C^{(i)}|X) - \epsilon' \geq (\mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon/2) - \epsilon' \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon,$$

the property of the distinguisher A guarantees that

$$\begin{aligned}\mathbb{E}[W^{(i)}(X,B)] - \mathbb{E}[W^{(i)}(X,C^{(i)})] &\geq \mathbb{E}[W^{(i)}(X,B)] - \mathbb{E}[W^{(i)}(X,\Phi_{M^{(i)}})] - \epsilon' \\ &\geq \mathbb{E}[A^{O_{X,B,\Phi_{M^{(i)}}}}(X,B)] - \mathbb{E}[A^{O_{X,B,\Phi_{M^{(i)}}}}(X,\Phi_{M^{(i)}})] - 2\epsilon' \\ &\geq \epsilon - 2\epsilon'.\end{aligned}$$

Hence by the Uniform Min-Max Theorem – Average Case (Theorem 2.5), for all Player 1 strategies $(X,C) \in \mathcal{V}_r(X)$,

$$\mathbb{E}[W_r^*(X,B)] - \mathbb{E}[W_r^*(X,C)] \geq \epsilon - 2\epsilon' - O(\epsilon') \geq .9\epsilon.$$

□

Given the algorithm Υ in Claim 4.54, we claim that the following time t randomized oracle algorithm P violates the hypothesis that B is uniformly (t,δ) KL-hard given X. We let $\gamma > 0$ be an error parameter to be fixed later, and c be a sufficiently large constant.

INPUT: $(x,a) \in \{0,1\}^n \times \{0,1\}^\ell$

ORACLE: $O_{X,B}$

for $r \leftarrow 0$ **to** ℓ *in steps of* ϵ/c **do**

  $W_r^* \leftarrow \Upsilon^{O_{X,B}}(r,n,\ell,t',\epsilon,\gamma)$

  for $k \leftarrow 0$ **to** ℓ/ϵ *in steps of* ϵ/c **do**

    $E_{r,k} \leftarrow$ an estimate of $\mathrm{KL}(X,B||X,\mathbf{e}^{k\cdot W_r^*}) + \mathrm{H}_{\mathrm{sh}}(B|X)$ within ϵ/c error

    (estimated using oracle $O_{X,B}$)

  end

end

Let r^*, k^* minimize $E_{r,k}$

Let $p(x,a) \in [0,1]$ be an approximation of $\exp(k^* \cdot W_{r^*}^*(x,a))/(2^\ell \cdot \exp(k^*))$

return $p(x,a)$

Algorithm 4.1: The oracle algorithm P

To prove correctness, first we claim that w.p. at least $1-\gamma$, in some iteration the variables r and k must satisfy

$$\mathrm{KL}(X,B||X,\mathbf{e}^{kW^*_r}) \le \delta - \epsilon/3 + \epsilon/c. \tag{$\star$}$$

Indeed, consider an iteration where $r \in [\mathrm{H}_{\mathrm{sh}}(B|X)+\delta-\epsilon/2, \mathrm{H}_{\mathrm{sh}}(B|X)+\delta-\epsilon/3]$. Suppose that $\mathbb{E}[W^*_r(X,B)] - \mathbb{E}[W^*_r(X,C)] \ge .9\epsilon$ for all C satisfying

$$\mathrm{H}_{\mathrm{sh}}(C|X) \ge \mathrm{H}_{\mathrm{sh}}(B|X) + \delta - \epsilon/3 \ge r.$$

This happens w.p. at least $1-\gamma$ by Claim 4.54 above. Recall that Lemma 4.49 says there exists $k' \in [0, \ell/\epsilon]$ such that $\mathrm{KL}(X,B||X,\mathbf{e}^{k'W^*_r}) \le \delta - \epsilon/3$. Now consider any inner iteration where $k \in [k'-\epsilon/c, k']$. Note that an ϵ/c difference between k and k' can introduce at most ϵ/c difference in $\mathrm{KL}(X,B||X,\mathbf{e}^{\cdot W^*_r})$ (see Lemma A.7 for details), thus we conclude

$$\mathrm{KL}(X,B||X,\mathbf{e}^{kW^*_r}) \le \mathrm{KL}(X,B||X,\mathbf{e}^{k'W^*_r}) + \epsilon/c \le \delta - \epsilon/3 + \epsilon/c.$$

It turns out that by sampling, for every pair of r and k, we can compute an estimate $E_{r,k}$ of $\mathrm{KL}(X,B||X,\mathbf{e}^{k\cdot W^*_r}) + \mathrm{H}_{\mathrm{sh}}(B|X)$ within ϵ/c error w.p. at least $1-\gamma/2$, in time $\mathrm{poly}\left(t', n, 1/\epsilon, 2^\ell, \log(1/\gamma)\right)$; see Lemma A.6 for details. Thus, it follows from $(\star)$ that w.p. at least $1-2\gamma$, the pair r^* and k^* that minimize $E_{r,k}$ must satisfy

$$\mathrm{KL}(X,B||X,\mathbf{e}^{k^*W^*_{r^*}}) \le \delta - \epsilon/3 + \epsilon/c + 2\epsilon/c. \tag{$\star\star$}$$

Finally, once k^* and $W^*_{r^*}$ are determined, the algorithm computes a (deterministic) approximation $p(x,a) \in [1, \exp(k^*)]$ of $\exp(k^* \cdot W^*_{r^*}(x,a))$. To make P $[0,1]$-valued (as required in the definition of KL-predictor), we normalize $p(x,a)$ to $[0,1]$ in the final step. Using Newton's method (see Lemma A.6 for details), such $p(x,a)$ can be computed in time $t = \mathrm{poly}\left(t', n, 1/\epsilon, 2^\ell\right)$ such that for the function $p : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$,

$$\mathrm{KL}\left(X,B||X,\Phi_p\right) \le \mathrm{KL}(X,B||X,\mathbf{e}^{k^*W^*_{r^*}}) + \epsilon/c. \tag{$\star\star\star$}$$

We view $P^{O_{X,B}}$ as a distribution on KL predictors $p : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$, where the randomness comes from k^* and $W^*_{r^*}$ (which in turn are generated from coins of P and $O_{X,B}$). By definition of KL-hardness, we need to show

$$\mathop{\mathbb{E}}_{p\sim P^{O_{X,B}}}[\mathrm{KL}(X,B||X,\Phi_p)] \le \delta.$$

We know from $(\star\star)$ and $(\star\star\star)$ that w.p. at least $1-2\gamma$ over $p \sim P^{O_{X,B}}$,

$$\mathrm{KL}\left(X,B||X,\Phi_p\right) \le \mathrm{KL}(X,B||X,\mathbf{e}^{k^*W^*_{r^*}}) + \epsilon/c \le \delta - \epsilon/3 + 4\epsilon/c \le \delta - \epsilon/4.$$

Meanwhile for every $p : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$,

$$\begin{aligned}\mathrm{KL}(X,B||X,\Phi_p) &= \mathbb{E}\left[\sum_a B(a|X)\log\left(B(a|X)/\Phi_p(a|X)\right)\right]\\ &\le \max_{x,a}\log\left(1/\Phi_p(a|x)\right) = O(\ell+1/\epsilon).\end{aligned}$$

Thus

$$\mathop{\mathbb{E}}_{p\sim P^{O_{X,B}}}[\mathrm{KL}(X,B||X,\Phi_p)] \le (1-2\gamma)\cdot(\delta-\epsilon/4) + (2\gamma)\cdot O(\ell+1/\epsilon) \le \delta$$

for an appropriate choice of $\gamma = \Omega(\epsilon/(\ell+1/\epsilon))$, completing the proof. □

4.3.6 Pseudoentropy Implies KL-hardness

We apply Theorem 4.46 to show that pseudoentropy implies KL-hardness. In fact, we show that even a weak form of pseudoentropy suffices:

Definition 4.55 (Weak conditional pseudoentropy, nonuniform setting)**.** Let (X,B) be a joint distribution. We say B has (T,ϵ) *weak nonuniform (conditional) pseudoentropy at least k given X* if there exists a joint distribution (Y,C) such that the following holds:

- $\mathrm{H}_{\mathrm{sh}}(C|Y) + \mathrm{H}_{\mathrm{sh}}(Y) - \mathrm{H}_{\mathrm{sh}}(X) \ge k$. In particular, $\mathrm{H}_{\mathrm{sh}}(C|Y) \ge k$ if $\mathrm{H}_{\mathrm{sh}}(Y) = \mathrm{H}_{\mathrm{sh}}(X)$;
- (X,B) and (Y,C) are ϵ-indistinguishable by all size T circuits.

If $(X, B) = (X, B)(n)$ for a security parameter n, we say B has *weak nonuniform (conditional) pseudoentropy at least $k = k(n)$ given X* if for every constant c, B has $(n^c, 1/n^c)$ weak nonuniform (conditional) pseudoentropy at least $k(n) - 1/n^c$ given X for all sufficiently large n.

In the uniform setting, it suffices to assume an even weaker form of pseudoentropy, where we only require indistinguishability against distinguishers given oracle access to $O_{X,B}$ but not $O_{X,B,C}$:

Definition 4.56 (Weak conditional pseudoentropy, uniform setting)**.** Let n be a security parameter, $T = T(n)$, $\epsilon = \epsilon(n)$, $k = k(n)$, $\ell = \ell(n)$. Let $(X, B) = (X, B)(n)$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. We say B has (T, ϵ) *weak uniform (conditional) pseudoentropy at least k given X* if for every randomized oracle algorithm A computable in time T, there is joint distribution (Y, C) such that the following holds for all sufficiently large n:

- $\mathrm{H}_{\mathrm{sh}}(C|Y) + \mathrm{H}_{\mathrm{sh}}(Y) - \mathrm{H}_{\mathrm{sh}}(X) \geq k$. In particular, $\mathrm{H}_{\mathrm{sh}}(C|Y) \geq k$ if $\mathrm{H}_{\mathrm{sh}}(Y) = \mathrm{H}_{\mathrm{sh}}(X)$;
- (X, B) and (Y, C) are ϵ-indistinguishable by $A^{O_{X,B}}$:

$$|\Pr[A^{O_{X,B}}(X, B) = 1] - \Pr[A^{O_{X,B}}(Y, C) = 1]| < \epsilon.$$

We say B has *weak uniform (conditional) pseudoentropy at least $k = k(n)$ given X* if for every constant c, B has $(n^c, 1/n^c)$ weak uniform (conditional) pseudoentropy at least $k(n) - 1/n^c$ given X. Note that in this "polynomial" version, $O_{X,B}$ is redundant if (X, B) is polynomial-time samplable.

Theorem 4.57 (Weak pseudoentropy $\Longrightarrow$ KL-hardness, nonuniform and uniform settings)**.** *Let n be a security parameter, $\delta = \delta(n) > 0$, $t = t(n) \in \mathbb{N}$, $\epsilon = \epsilon(n) > 0$, $\ell = \ell(n)$, $\sigma = \sigma(n)$*

all computable in time poly(n). *Let* (X, B) *be a joint distribution on* $\{0,1\}^n \times \{0,1\}^\ell$. *If* B *has weak (non)uniform* (t, ϵ) *pseudoentropy at least* $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ *given* X, *then* B *is (non)uniformly* (t', δ') *KL-hard given* X, *for* $t' = \min\{t^{\Omega(1)}/\mathrm{poly}(n, \log(1/\sigma)), \Omega(\sigma/\epsilon)\}$ *and* $\delta' = \delta - \sigma$.

Proof. We first give a proof for the nonuniform setting. The proof for the uniform setting will follow naturally.

Suppose for contradiction that B is not nonuniformly $(t', \delta - \sigma)$ KL-hard. Then there is size t' circuit $P : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$ such that $\mathrm{KL}(X, B \parallel X, \Phi_P) \leq \delta - \sigma$. We assume w.l.o.g. that $P(x, a) \neq 0$ (otherwise, the bounded KL divergence implies $(x, a) \notin \mathrm{supp}(X, B)$, i.e. we can set $P(x, a)$ to nonzero without affecting anything).

We first show that the function

$$W(x, a) = \frac{\log P(x, a)}{t'} + 1$$

satisfies $\mathbb{E}[W(X, B)] - \mathbb{E}[W(Y, C)] \geq \epsilon$ for every joint distribution (Y, C) with $\mathrm{H}_{\mathrm{sh}}(C|Y) + \mathrm{H}_{\mathrm{sh}}(Y) - \mathrm{H}_{\mathrm{sh}}(X) \geq \mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ i.e. $\mathrm{H}_{\mathrm{sh}}(Y, C) \geq \mathrm{H}_{\mathrm{sh}}(X, B) + \delta$. Note that W is indeed a $[0,1]$-valued function, because $2^{-t'} \leq P(x, a) \leq 1$.

Consider the function W^* defined as

$$\begin{aligned} W^*(x, a) &= -\frac{\partial \mathrm{H}_{\mathrm{sh}}(\Phi_P|_{X=x})}{\partial \Pr[\Phi_P = a | X = x]} \\ &= \log \frac{P(x, a)}{\sum_b P(x, b)} + \log e \\ &= t' \cdot W(x, a) - t' + \log e - \log \sum_b P(x, b). \end{aligned}$$

Since $\mathrm{KL}(X, B \parallel X, \Phi_P) \leq \delta - \sigma$, Lemma 4.46 (whose hypothesis is satisfied by W^*) implies that for every joint distribution (Y, C) with $\mathrm{H}_{\mathrm{sh}}(Y, C) \geq \mathrm{H}_{\mathrm{sh}}(X, B) + \delta$, we have $\mathbb{E}[W^*(X, B)] - \mathbb{E}[W^*(Y, C)] \geq \sigma$. Thus

$$\mathbb{E}[W(X, B)] - \mathbb{E}[W(Y, C)] = \frac{\mathbb{E}[W^*(X, B)] - \mathbb{E}[W^*(Y, C)]}{t'} \geq \frac{\sigma}{t'}.$$

Thus it remains to show that we can approximate W by a size t circuit. Let $\widetilde{W}$ be an approximation to W where $\log P(x,a)$ is computed to precision σ. Since $P(x,a)$ is represented as a rational p_1/p_2 where $p_1, p_2 \leq 2^{t'}$, the logarithm can be approximated to that precision in time $\text{poly}(t', \log(1/\sigma))$. Thus $\widetilde{W}$ has circuit size $\text{poly}(t', \log(1/\sigma)) \leq t$. Moreover, for all (Y,C) with $\mathrm{H}_{\mathrm{sh}}(Y,C) \geq \mathrm{H}_{\mathrm{sh}}(X,B) + \delta$, we have

$$\mathbb{E}[\widetilde{W}(X,B)] - \mathbb{E}[\widetilde{W}(Y,C)] \geq \mathbb{E}[W(X,B)] - \mathbb{E}[W(Y,C)] - \frac{1}{2t'} \cdot \sigma \geq \frac{\sigma}{2t'} \geq \epsilon,$$

contradicting the weak pseudoentropy of B given X.

Proof for the uniform setting follows quite naturally. Suppose for contradiction that we are given a t'-time randomized oracle algorithm P such that

$$\mathop{\mathbb{E}}_{p \sim P^{O_{X,B}}} [\mathrm{KL}(X,B||X,\Phi_p)] \leq \delta - \sigma$$

where $P^{O_{X,B}}$ is viewed as a distribution over functions $p : \{0,1\}^n \times \{0,1\}^\ell \to [0,1]$. Implicit in the nonuniform argument is an algorithm converting a λ-KL predictor to a universal $(\delta - \lambda - \sigma/2)/t'$-distinguisher, for all λ. Thus, we let A be the time $\text{poly}(n, t', \log(1/\sigma)) \leq t$ randomized oracle algorithm performing the above conversion from a λ-KL predictor to a universal $(\delta - \lambda - \sigma/2)/t'$-distinguisher, replacing the circuit output $P(x,a)$ with the output of simulating $P^{O_{X,B}}$ on (x,a) (using random coin tosses and $O_{X,B}$). Thus for every (Y,C) with $\mathrm{H}_{\mathrm{sh}}(Y,C) \geq \mathrm{H}_{\mathrm{sh}}(X,B) + \delta$,

$$\begin{aligned}\mathbb{E}\left[A^{O_{X,B}}(X,B)\right] - \mathbb{E}\left[A^{O_{X,B}}(Y,C)\right] &\geq \mathop{\mathbb{E}}_{p \sim P^{O_{X,B}}} \left[\frac{\delta - \mathrm{KL}(X,B||X,\Phi_p) - \sigma/2}{t'}\right] \\ &\geq \frac{\sigma}{2t'} \geq \epsilon,\end{aligned}$$

contradicting the weak pseudoentropy of B given X. □

Since Theorem 4.57 only requires weak conditional pseudoentropy, we obtain the following equivalence:

Corollary 4.58. *Let n be a security parameter, let $\delta = \delta(n) > 0$, and $\ell = \ell(n) = O(\log n)$ be computable in time* poly(n)*. Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}^\ell$. Then the following are equivalent:*

1. *B is (non)uniformly δ KL-hard given X;*
2. *B has (non)uniform pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X;*
3. *B has weak (non)uniform pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(B|X) + \delta$ given X.*

Proof. $1 \implies 2$ by Theorem 4.50 and 4.53. $2 \implies 3$ by definition. $3 \implies 1$ by Theorem 4.57. □

Chapter 5

Constructing Pseudorandom Generators from One-Way Functions

A centerpiece of the foundations of cryptography and pseudorandomness theory is the Håstad, Impagliazzo, Levin, and Luby [HILL] result that that pseudorandom generators can be constructed from arbitrary one-way functions. In this chapter, we simplify and improve the construction of pseudorandom generators from one-way functions, building on the previous state-of-the-art construction of Haitner, Reingold, and Vadhan [HRV].

The simplified construction uses our characterization of conditional pseudoentropy from Chapter 4 to obtain next-bit pseudoentropy from arbitrary one-way functions, proving a conjecture of [HRV]. In particular, the construction only performs hashing once, and only needs the hash functions that are randomness extractors (e.g. universal hash functions) rather than needing them to support "local list-decoding" (as in the Goldreich-Levin hard-core predicate [GL]). With an additional idea, we also show how to improve the efficiency

of the Haitner, Reingold, and Vadhan construction, reducing the seed length of the pseudorandom generator to $\tilde{O}(n^3)$ from $\tilde{O}(n^4)$ (which was already a significant improvement over Håstad et al.).

5.1 Introduction

5.1.1 Pseudorandom Generators and One-Way Functions

We begin with the definition of pseudorandom generators. A pseudorandom generator is an efficient deterministic algorithm G that stretches a short random string to a longer string that *looks* random:

Definition 5.1 (Pseudorandom generator (PRG) [BM, Yao2], informal)**.** A polynomial-time computable function $G : \{0,1\}^d \to \{0,1\}^\ell$, $d < \ell$, is a *pseudorandom generator* if $G(U_d)$ is computationally indistinguishable from U_ℓ.

This is a very strong definition of pseudorandom generators as it guarantees security (indistinguishability) against *all* efficient adversaries (distinguishers), even those that run in time greater than the time needed to compute the pseudorandom generator G itself (which is a fixed polynomial). Such kinds of pseudorandom generators are sometimes known as cryptographic pseudorandom generators (to distinguish from other kinds of pseudodorandom generators, used for derandomization), as they allow numerous other cryptographic primitives to be constructed, such as private-key cryptography [GGM, LR], bit-commitment schemes [Nao1], zero-knowledge proofs for NP [GMW], and identication schemes [FFS]. They are also the key assumptions for many complexity theoretic results, for example, hardness results in learning [Val] and the natural proofs barrier for circuit lower bounds [RR].

All of these applications beg the question whether pseudorandom generators exist at

all. However, it can be easily seen that pseudorandom generators imply $\mathbf{P} \neq \mathbf{NP}$; thus we do not expect to unconditionally establish their existence but rather hope to do so based on computational assumptions. Unfortunately, it seems out of reach to base their existence on complexity-theoretic statements such as $\mathbf{P} \neq \mathbf{NP}$. Nonetheless, it turns out that pseudorandom generators *can* be based on the very plausible assumptions of functions that are easy to compute but hard to invert:

Definition 5.2 (One-way function, informal)**.** A polynomial-time computable function $f : \{0,1\}^* \to \{0,1\}^*$ is *one-way* if no probabilistic polynomial-time algorithm A satisfies $\Pr_{y \sim f(U_n)}[f(A(y)) = y] \geq n^{-O(1)}$.

One-way functions are *essential* for complexity-based cryptography to exist and often considered the minimal cryptographic primitive, as all the cryptographic primitives we mentioned (and others) imply the existence of one-way functions, often via simple reductions [IL, IR]. In contrast to pseudorandomness, one-wayness is a much more "unstructured" property where hardness can be distributed arbitrarily across the n input bits. Nonetheless, Håstad, Impagliazzo, Levin, and Luby [HILL] showed that pseudorandom generators can be constructed from arbitrary one-way functions:

Theorem 5.3 (Håstad et al. [HILL])**.** *The following are equivalent:*

1. *One-way functions exist;*
2. *Pseudorandom generators $G : \{0,1\}^d \to \{0,1\}^\ell$ exist with $\ell = d + 1$;*
3. *For every constant c there exist pseudorandom generators with $\ell = d^c$.*

The Håstad et al. paper is one of the centerpieces of the foundations of cryptography and the theory of pseudorandomness. Not only does it tell us that all the cryptographic primitives we mentioned are in fact equivalent, it also introduced concepts and techniques

that now permeate the theory of pseudorandomness, such as pseudoentropy (see Chapter 4) and the Leftover Hash Lemma.

A drawback of the Håstad et al. construction, however, is that it is quite complicated. While it utilizes elegant notions and ideas, the actual construction has to integrate them in a rather ad hoc and indirect way (due to various technical issues). Moreover, the reduction showing the correctness of the construction is much more complex in the uniform setting. Aesthetic and pedagogical perspectives aside, the complexity of the construction also makes it highly inefficient. Specifically, the pseudorandom generator constructed from a one-way function $f : \{0,1\}^n \to \{0,1\}^n$ requires an input (known as the *seed*) length of $d = O(n^{10})$, unpractical even for very modest settings of parameters. (Håstad et al. also outlined a construction of seed length $O(n^8)$, which was later formalized and proved in [Hol2]).

Progress has been made to simplify the construction and improve its efficiency. By proving a Uniform Hardcore Theorem (cf. Chapter 3), Holenstein [Hol1, Hol2] substantially simplified and modularized the proof in the uniform setting. Haitner, Harnik, and Reingold [HHR2] reduced the seed length to $O(n^7)$. Holenstein [Hol2] generalized the Håstad et al. result to base on one-way functions of any "hardness." In particular, given a one-way function that is secure again exponential time ($2^{\Omega(n)}$) adversaries, the seed length was reduced to $O(n^4 \cdot \omega(\log n))$ (or $O(n^5)$ to obtain a PRG with exponential security), and subsequently improved by Haitner, Harnik, and Reingold [HHR1] to $O(n \cdot \omega(\log n))$ (or $O(n^2)$ to obtain a PRG with exponential security). All these constructions, however, still retain the overall structure of the Håstad et al. construction based on pseudoentropy, and thus retain some of the complex and ad hoc elements.

Recently, Haitner, Reingold, and Vadhan [HRV] provided a simpler and much more efficient construction, based on the more relaxed notion of *conditional pseudoentropy* (see Chapter 4) as embedded in their notion of *next-bit pseudoentropy*. The new construction

requires a much shorter seed length of $\tilde{O}(n^4)$. (If the one-way function is secure again exponential time adversaries, then the seed length matches the [HHR1] result.)

5.1.2 From One-Way Functions to Next-Bit Pseudoentropy

The Haitner, Reingold, and Vadhan construction proceeds in two stages: first, from a one-way function construct a *next-bit pseudoentropy* generator, then convert next-bit pseudoentropy to pseudorandomness. Next-bit pseudoentropy simply captures the total conditional pseudoentropy (see Chapter 4) across all the bits:

Definition 5.4 (Next-block pseudoentropy [HRV], informal)**.** A joint distribution $(X_1, \ldots, X_m)$ has *next-block pseudoentropy* at least k iff there exist a sequence of distributions $Y_1, \ldots, Y_m$, jointly distributed with $(X_1, \ldots, X_m)$ such that:

1. $(X_1, \ldots, X_{i-1}, X_i)$ is computationally indistinguishable from $(X_1, \ldots, X_{i-1}, Y_i)$, and
2. $\sum_i \mathrm{H}_{\mathrm{sh}}(Y_i | X_1, \ldots, X_{i-1}) \geq k$.

Equivalently, X_I has pseudoentropy at least k/m given $X_1, \ldots, X_{I-1}$, where I is uniformly distributed in $[m]$.

We say that a distribution X taking values in $\{0,1\}^m$ has *next-bit pseudoentropy* at least k iff when we break X into 1-bit blocks, then $X = (X_1, \ldots, X_m)$ has next-block pseudoentropy at least k.

Intuitively, next-bit pseudoentropy captures the pseudoentropy from the perspective of an adversary who gets the bits one at a time (from left to right), instead of all at once. Thus, the next-bit pseudoentropy of a distribution can be much larger than its pseudoentropy. For example, if $G : \{0,1\}^n \to \{0,1\}^m$ is a pseudorandom generator, then $(G(U_n), U_n)$ has next-bit pseudoentropy at least $m > n$, but does not have pseudoentropy larger than n.

Haitner, Reingold, and Vadhan [HRV] showed that if $f : \{0,1\}^n \to \{0,1\}^m$ is a one-way function, $X \in_R \{0,1\}^n$, and $H : \{0,1\}^n \to \{0,1\}^n$ is a random hash function from an appropriate family, then $(f(X), H, H(X))$ has next-bit pseudoentropy $n + r + \log n$, where r is the number of random bits used to describe the hash function H. The intuition for this is as follows: Condition on $f(X) = y$ for some $y \in \{0,1\}^n$. Given that $f(X) = y$, X is uniformly distributed in a set of size $|f^{-1}(y)|$. Thus, by the Leftover Hash Lemma [HILL], the first $\approx \log |f^{-1}(y)|$ bits of $H(X)$ are statistically close to uniform given the prefix preceding them. In addition, it is still difficult to invert f and predict X given these bits (since a uniform random string can't help in inverting). Thus, by the Goldreich–Levin Theorem [GL], the next $\approx \log n$ bits of $H(X)$ are computationally indistinguishable from uniform given the preceding bits. Therefore the next-bit pseudoentropy of $(f(X), H, H(X))$ is at least

$$\begin{aligned}
&\mathrm{H}_{\mathrm{sh}}(f(X)) + r + \mathop{\mathbb{E}}_{y \leftarrow f(X)}[\log |f^{-1}(y)|] + \log n \\
&= \mathrm{H}_{\mathrm{sh}}(f(X)) + r + \mathrm{H}_{\mathrm{sh}}(X|f(X)) + \log n \\
&= n + r + \log n.
\end{aligned}$$

Haitner, Reingold, and Vadhan [HRV] conjectured that the hashing in the above construction is not necessary, and the hardness of inverting a one-way function directly provides (next-bit) pseudoentropy. We prove their conjecture, using our characterization of uniform conditional pseudoentropy from Chapter 4, Theorem 4.41:

Theorem 5.5 (One-way function $\Longrightarrow$ next-bit pseudoentropy)**.** *If $f : \{0,1\}^n \to \{0,1\}^m$ is a one-way function and $X \in_R \{0,1\}^n$, then $(f(X), X)$ has next-bit pseudoentropy at least $n + \log n$.*

We will prove the theorem in the *uniform* security setting (where the adversaries for one-way function and next-bit pseudoentropy are uniform algorithms), but the theorem

holds also in the *nonuniform* setting (since it is proved via a uniform reduction converting any adversary violating next-bit pseudoentropy to one violating the one-wayness of f).

The proof of this theorem starts by showing that the one-wayness of f implies that for every probabilistic polynomial-time algorithm A, the KL divergence from $(f(X), X)$ to $(f(X), A(f(X)))$ is at least $\log n$; otherwise A would invert f with nonnegligible probability. Then we show that the same holds also in a "next-bit" sense: if we break X into bits $X = X_1 \cdots X_n$ and choose $I \in_R [n]$, then for every probabilistic polynomial-time S, the KL divergence from $(f(X), X_1, \ldots, X_I)$ to $(f(X), X_1, \ldots, X_{I-1}, S(f(X), X_1, \ldots, X_{I-1}))$ is at least $(\log n)/n$. (Otherwise by iteratively applying S n times, we can obtain a probabilistic polynomial-time A such that $(f(X), A(f(X)))$ has KL divergence at most $\log n$ from $(f(X), X)$.) By Chapter 4, Theorem 4.41, we deduce that X_I has pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(X_I | f(X), X_1, \ldots, X_{I-1}) + (\log n)/n$ given $f(X), X_1, \ldots, X_{I-1}$. That is, on average, the individual bits of X have $(\log n)/n$ extra bits of pseudoentropy (in addition to Shannon entropy) given $f(X)$ and the previous bits of X. Summing over all n bits of X, the next-bit pseudoentropy is at least $\log n$ bits larger than the Shannon entropy of $(f(X), X)$, which is n.

5.1.3 From Next-Bit Pseudoentropy to Pseudorandomness

Given the next-bit pseudoentropy generator $(f(X), X) \in \{0,1\}^{m+n}$ of Theorem 5.5, we can apply the construction of Haitner et al. [HRV] to obtain a pseudorandom generator through the following three steps:

- **Entropy Equalization:** To spread the pseudoentropy out evenly among the bits, we concatenate $u = \tilde{\Theta}(n)$ independent random evaluations of $(f(X), X)$, then drop the first I bits and the last $m+n-I$ bits of the $u \cdot (n+m)$-bit long result, for $I \in_R [m+n]$.

- **Converting Shannon Entropy to Min-Entropy and Amplifying the Gap:** Next, we take $t = \tilde{\Theta}(n^2)$ copies of the above next-bit pseudoentropy generator (after entropy equalization), but concatenate them "vertically" to obtain blocks, each of which consists of t bits. It can be shown that each of the blocks is indistinguishable from having high min-entropy conditioned on the previous ones.

- **Randomness Extraction:** Finally, we use a single random universal hash function to extract the pseudo-min-entropy from each of the blocks, and concatenate the results to produce our output.

Thus, to obtain a pseudorandom generator from a one-way function f, we simply need to evaluate f on $u \cdot t = \tilde{O}(n^3)$ random inputs, arrange the input and output bits into a matrix consisting of $(u-1) \cdot (m+n)$ columns and t rows, and apply a universal hash function to each column. (The seed of the pseudorandom generator consists of the $u \cdot t$ inputs to f, the t random shifts used for entropy equalization, and the description of the universal hash function.) The construction is illustrated in Figure 5.1. Note that we only need to hash once in the construction, and the only property we need of our hash function is randomness extraction (e.g. via the Leftover Hash Lemma). In contrast, all previous constructions of pseudorandom generators from one-way functions (even from one-way permutations) required hash functions with "local list-decoding" properties (e.g. the Goldreich–Levin hard-core predicate) in addition to randomness extraction. As pointed out to us by Yuval Ishai, an advantage of using only universal hash functions is that they can be implemented by linear-size boolean circuits [IKOS], and thus we can obtain PRGs computable by circuits of size linear in their stretch (from one-way functions that are computable by linear-size circuits but exponentially hard to invert). Such PRGs have applications to "cryptography with constant computational overhead" [IKOS].

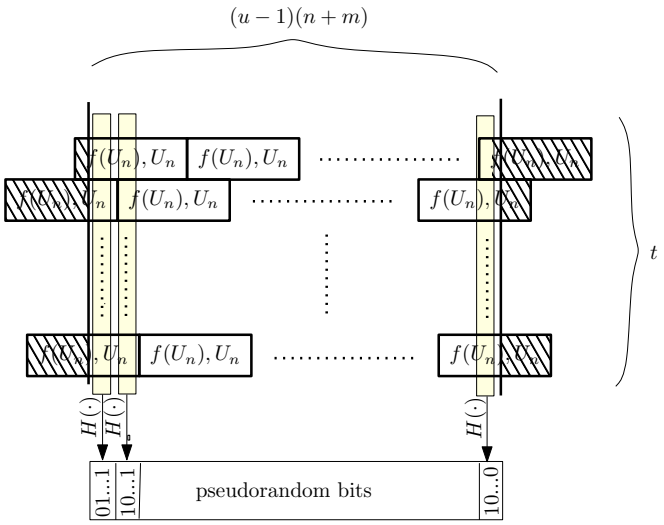

Figure 5.1: Simplified construction of PRG from one-way function $f : \{0,1\}^n \to \{0,1\}^m$. Each row contains iid copies of $(f(U_n), U_n)$, shifted by a random offset $I \in [n+m]$. To extract pseudorandom bits, an arbitrary universal hash function H (with a proper output length) is applied to all bits in the same column.

While simpler, the aforementioned construction achieves essentially the same parameters as [HRV]. Using an additional idea, we show how to save a factor of roughly $u = \tilde{\Theta}(n)$ in the seed length. The idea is that to extract the randomness from a column of the aforementioned matrix, we do not need to construct the entire matrix. We can use just enough seed to fill a single column, and then we can use randomness extracted from that column to help generate more columns, and iterate. (This idea is independent of our simplifications above, and can also be applied to the construction based on the [HRV] pseudoentropy generator.) Thus we show:

Theorem 5.6 (One-way function $\Longrightarrow$ pseudorandom generator, informal)**.** *Given a one-way function $f : \{0,1\}^n \to \{0,1\}^n$, we can construct a pseudorandom generator G with seed length $\tilde{O}(n^3)$.*

This theorem improves the seed length of $O(u \cdot t \cdot n) = \tilde{O}(n^4)$ from Haitner, Reingold, and Vadhan. This theorem generalizes to one-way functions of "any hardness," and both the

construction itself and the underlying security reduction are uniform (thus the construction also works for nonuniform security settings). We note that Haitner, Reingold, and Vadhan give a *nonuniform* construction of seed length $\tilde{O}(n^3)$, which requires poly(n) bits of nonuniform advice to compute the pseudorandom generator. (They do so by avoid Entropy Equalization, by nonuniformly hardwiring the amount of entropy contributed by each bit.) Also, our construction still requires evaluating the one-way function at least $u \cdot t = \tilde{\Theta}(n^3)$ times; we just no longer need these evaluations to be independent. Finally, like Haitner, Reingold, and Vadhan, the construction obtains $\Theta(\log n)$ bits of additive stretch per invocation of the one-way function, which is optimal [GGKT].

With Theorem 5.6, now the only blow-up in seed length in constructing pseudorandom generators from one-way functions is due to converting Shannon entropy to min-entropy. On the flip side, Holenstein and Sinha [HS] recently showed that any black-box construction of pseudorandom generator from arbitrary one-way functions requires $\Omega(n/\log n)$ calls to the underlying one-way function. It is an intriguing open problem whether our seed length blow-up of $\tilde{O}(n^2)$ and our complexity blow-up of $\tilde{O}(n^3)$ can be avoided, or shown to be necessary by strengthening the Holenstein and Sinha lower bound.

5.2 Definitions

Definition 5.7 (One-way functions)**.** A polynomial-time computable function $f : \{0,1\}^* \to \{0,1\}^*$ is (T, γ) *one-way* $T = T(n)$, $\gamma = \gamma(n)$ if for every time T randomized algorithm A, for all sufficiently large n, it holds that $\Pr_{y \sim f(U_n)}[f(A(y)) = y] < \gamma$. We say f is one-way if f is $(n^c, 1/n^c)$ one-way for every constant c.

Definition 5.8 (Pseudorandom)**.** Let n be a security parameter, $\ell = \ell(n)$. A distribution X on $\{0,1\}^\ell$ is (T, ϵ) *pseudorandom* for $T = T(n)$, $\epsilon = \epsilon(n)$ if for all time T randomized

algorithms A, $\Pr[A(X) = 1] - \Pr[A(U_\ell) = 1] \leq \epsilon$. A polynomial-time computable function $G : \{0,1\}^{d=d(n)} \to \{0,1\}^{\ell=\ell(n)}$ is a (T, ϵ) *pseudorandom generator* (PRG) if $G(U_d)$ is (T, ϵ) pseudorandom.

We say G is a *pseudorandom generator* if G is a $(n^c, 1/n^c)$ pseudorandom generator for every constant c. The input to a pseudorandom generator is called the *seed.* The number of extra bits, $\ell - d$, is called the *stretch*.

Note that nonuniform pseudorandomness and pseudorandom generators can be defined by replacing time T algorithms by size T boolean circuits.

It is useful to talk about the *total* conditional pseudoentropy of a sequence of jointly distributed strings, called the *next-block pseudoentropy*:

Definition 5.9 (Next-block pseudoentropy)**.** Let n be a security parameter, $k = k(n)$, and $B^{(i)}$ be a distribution for each $i = 1, \ldots, m = m(n)$. We say $\left(B^{(1)}, B^{(2)}, \ldots\right)$ has *(non)uniform next-block* (or *next-bit*, if each $B^{(i)}$ is a bit) *pseudoentropy at least* k if $B^{(I)}$ has (non)uniform pseudoentropy at least k/m given $B^{(1)} \ldots B^{(I-1)}$, for $I \in_R [m]$.

Note that next-bit pseudoentropy is a more relaxed notion than pseudoentropy, and to increase the next-block pseudoentropy, we would like "blocks" to be small i.e. as bits. Note that the next-bit pseudoentropy is sensitive to the order of the bits; for example, for any one-way function f, $(U_n, f(U_n))$ does not have next-bit pseudoentropy $n + 1$, but $(f(U_n), U_n)$ has next-bit pseudoentropy at least $n + \Omega(\log n)$ as we show in the next section.

5.3 From One-Way Functions to Next-Bit Pseudoentropy

In this section, we show how to obtain a next-bit pseudoentropy generator from an arbitrary one-way function f.

This section is structured as follows. Given a one-way function f, we first show that U_n is KL-hard for sampling given $f(U_n)$. By a chain rule for KL-hardness, we then argue it is KL-hard to sample the next bit of U_n given $f(U_n)$ and all previous bits of U_n. Finally, we use the equivalences between KL-hardness for sampling, KL-hardness, and conditional pseudoentropy (Chapter 4, Lemma 4.37 and Theorem 4.53) to derive that $(f(U_n), U_n)$ has a lot of total next-bit pseudoentropy.

Lemma 5.10 (One-way function $\Longrightarrow$ KL-hard for sampling)**.** *Let n be a security parameter, and $f : \{0,1\}^n \to \{0,1\}^n$ be (t, γ) one-way, for $t = t(n)$, $\gamma = \gamma(n)$. Then U_n is uniformly $(t', \log(1/\gamma))$ KL-hard for sampling given $f(U_n)$, for $t' = t/\mathrm{poly}(n)$.*

Proof. Suppose for contradiction that U_n is not uniformly $(t', \log(1/\gamma))$ KL-hard for sampling given $f(U_n)$, i.e. there exists a time t' randomized oracle algorithm S such that

$$\mathrm{KL}\left(f(U_n), U_n || f(U_n), S^{O_{f(U_n),U_n}}(f(U_n))\right) \leq \log \frac{1}{\gamma}.$$

Let $g(y, x)$ be the indicator function that $f(x) = y$. Since applying a (deterministic) function does not increase KL divergence (Lemma 1.11),

$$\mathrm{KL}\left(g\left(f(U_n), U_n\right) || g\left(f(U_n), S^{O_{f(U_n),U_n}}(f(U_n))\right)\right) \leq \log \frac{1}{\gamma}$$

where $g(f(U_n), U_n) \equiv 1$, and $g\left((f(U_n), S^{O_{f(U_n),U_n}}(f(U_n))\right)$ equals 1 with probability $p = \Pr[S^{O_{f(U_n),U_n}}(f(U_n)) = U_n]$. Since the KL divergence from Bernoulli(1) to Bernoulli(p) is $\log(1/p)$, we must have $p \geq \gamma$. That is,

$$\Pr[S^{O_{f(U_n),U_n}}(f(U_n)) = U_n] \geq \gamma.$$

Since $O_{f(U_n),U_n}$ can be simulated in time $\mathrm{poly}(n)$, this violates the fact that f is (t, γ) one-way for $t = t' \cdot \mathrm{poly}(n)$. □

Lemma 5.11 (Chain rule for KL-hardness)**.** *Let Y be a distribution over $\{0,1\}^n$, jointly distributed with Z. If Y is uniformly (t,δ) KL-hard for sampling given Z, then Y_I is uniformly $(t',\delta/n)$ KL-hard for sampling given $(Z,Y_1,\ldots,Y_{I-1})$, for $I \in_R [n]$, $t' = t/O(n)$.*

Proof. Suppose Y_I is not uniformly $(t',\delta/n)$ KL-hard for sampling given $(Z,Y_1,\ldots,Y_{I-1})$, that is there exists a time t' randomized oracle algorithm S such that

$$\mathrm{KL}\left(Z,Y_1,\ldots,Y_I||Z,Y_1,\ldots,Y_{I-1},S^{O_{Z,Y_1,\ldots,Y_I}}(Z,Y_1,\ldots,Y_{I-1})\right) \leq \frac{\delta}{n}.$$

Consider the time $O(nt') = t$ algorithm that samples $W_1,\ldots,W_n$ from Z using oracle $O_{Z,Y}$, where W_i is inductively defined to be $S^{O_{Z,Y_1,\ldots,Y_I}}(Z,W_1,\ldots,W_{i-1})$. By the chain rule for KL divergence (Proposition 1.10),

$$\begin{aligned}
&\mathrm{KL}(Z,Y_1,\ldots,Y_j||Z,W_1,\ldots,W_j) - \mathrm{KL}(Z,Y_1,\ldots,Y_{j-1}||Z,W_1,\ldots,W_{j-1}) \\
&= \mathrm{KL}((Y_j|Z,Y_1,\ldots,Y_{j-1})||(W_j|Z,W_1,\ldots,W_{j-1})) \\
&= \mathrm{KL}(Z,Y_1,\ldots,Y_j||Z,Y_1,\ldots,Y_{j-1},S^{O_{Z,Y_1,\ldots,Y_I}}(Z,Y_1,\ldots,Y_{j-1})),
\end{aligned}$$

where the last equality follows from definition of conditional KL divergence. Telescoping over $j = 1,\ldots,n$,

$$\begin{aligned}
&\mathrm{KL}(Z,Y||Z,W_1,\ldots,W_n) \\
&= \sum_{i=1}^{n} \mathrm{KL}(Z,Y_1,\ldots,Y_i||Z,Y_1,\ldots,Y_{i-1},S^{O_{Z,Y_1,\ldots,Y_I}}(Z,Y_1,\ldots,Y_{i-1})) \\
&= n \cdot \mathrm{KL}(Z,Y_1,\ldots,Y_I||Z,Y_1,\ldots,Y_{I-1},S^{O_{Z,Y_1,\ldots,Y_I}}(Z,Y_1,\ldots,Y_{I-1})) \\
&\leq n \cdot \frac{\delta}{n} = \delta.
\end{aligned}$$

This violates Y being uniformly (t,δ) KL-hard for sampling given Z.

□

Now the remainder of showing next-bit pseudoentropy of $(f(U_n),U_n)$ follows from (i)

KL-hard for sampling implies KL-hard (Chapter 4, Lemma 4.37); (ii) KL-hard implies conditional pseudoentropy (Chapter 4, Theorem 4.53). Formally,

Theorem 5.12 (One-way function $\Longrightarrow$ next-bit pseudoentropy)**.** *Let n be a security parameter, $t = t(n)$, $\gamma = \gamma(n)$, $\epsilon = \epsilon(n)$ all computable in polynomial time. Let $f : \{0,1\}^n \to \{0,1\}^n$ be (t, γ) one-way. Then $(f(U_n), U_n)$ has (t', ϵ) uniform next-bit pseudoentropy at least $n + \log(1/\gamma) - \epsilon$, for $t' = t^{\Omega(1)}/\mathrm{poly}(n, 1/\epsilon)$.*

Proof. Let $Z = f(U_n)$, $Y = U_n$ and $I \in_R [n]$. By Lemma 5.10 and 5.11, Y_I is uniformly $(t/\mathrm{poly}(n), \log(1/\gamma)/n)$ KL-hard for sampling given $(Z, Y_1, \ldots, Y_{I-1})$. By Chapter 4, Lemma 4.37, Y_I is uniformly $(t/\mathrm{poly}(n), \log(1/\gamma)/n)$ KL-hard given $(Z, Y_1, \ldots, Y_{I-1})$. By Chapter 4, Theorem 4.53, Y_I has (t', ϵ) uniform conditional pseudoentropy at least

$$\mathrm{H}_{\mathrm{sh}}(Y_I | Z, Y_1, \ldots, Y_{I-1}) + \log(1/\gamma)/n - \epsilon/n,$$

for $t' = t^{\Omega(1)}/\mathrm{poly}(n, 1/\epsilon)$. Equivalently, (Z, Y) has (t', ϵ) uniform next-bit pseudoentropy at least $\mathrm{H}_{\mathrm{sh}}(Y, Z) + \log(1/\gamma) - \epsilon = n + \log(1/\gamma) - \epsilon$. □

Remark 5.13. The argument in this section says that $(f(U_n), U_n)$ has a lot of next-bit pseudoentropy as long as U_n is KL-hard to sample from $f(U_n)$. The KL-hardness of sampling U_n from $f(U_n)$ is similar to the notion of a *distributional one-way function* [IL] which amounts to replacing KL divergence with statistical distance.

For U_n to be KL-hard to sample from $f(U_n)$, it is not necessary that f is one-way. For example, given any one-way function $h : \{0,1\}^n \to \{0,1\}^{n/2}$, define

$$f(x) = \begin{cases} x_{1,\ldots,n/2} & (x_{n/2+1,\ldots,n} = 0^{n/2}) \\ h(x) & (\text{otherwise}) \end{cases}.$$

Clearly f is not one-way, but U_n is still KL-hard to sample from $f(U_n)$. Thus, our construction of next-bit pseudoentropy generators (and later on, pseudorandom generators) can be

based on a larger class of functions.

5.4 From Next-Bit Pseudoentropy to Pseudorandomness

In this section, for brevity, we always assume the uniform setting whenever referring to one-way functions and computational notions of (conditional) entropy. Nonetheless, these results hold in the nonuniform setting too, with little or no change in the argument.

5.4.1 The Construction

Haitner et al. show a construction of a pseudorandom generator from any next-bit pseudoentropy generator G_{nb}. Their result can be stated as follows:

Theorem 5.14 (Pseudorandomness from next-bit pseudoentropy [HRV])**.** *Let n be a security parameter. Let $\Delta = \Delta(n) \in [1/\mathit{poly}(n), n]$, $m = m(n)$, $\kappa = \kappa(n) \in [n/2]$ be polynomial time computable. For every polynomial time computable $G_{nb} : \{0,1\}^n \to \{0,1\}^m$ such that $G_{nb}(U_n)$ has (T, ϵ) next-bit pseudoentropy at least $n + \Delta$, there exists a $(T - n^{O(1)}, n^{O(1)} \cdot (\epsilon + 2^{-\kappa}))$ pseudorandom generator $G : \{0,1\}^d \to \{0,1\}^{d \cdot (1+\Omega(\Delta/n))}$ with seed length*

$$d = O\left(\frac{m^2 n^2 \kappa \log^2 n}{\Delta^3}\right).$$

Moreover, G is computable in $\mathbf{NC}^1$ with $O(d/n)$ (uniformly random) oracle calls to G_{nb}.

By Theorem 5.12, we can simply use $U_n \to (f(U_n), U_n)$ as the next-bit pseudoentropy generator, and obtain the following construction of PRG G from one-way functions f (illustrated in Figure 5.1), by applying the construction in Theorem 5.14:

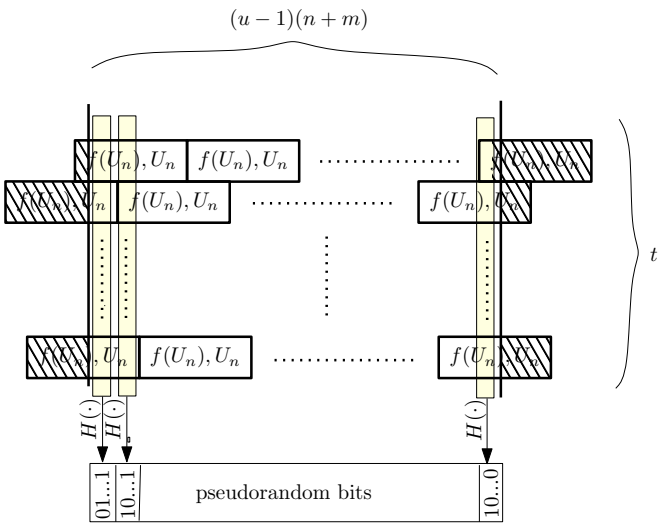

Figure 5.2: Simplified construction of PRG from one-way function $f : \{0,1\}^n \to \{0,1\}^m$. Each row contains $u = \Theta(n/\log n)$ iid copies of $(f(U_n), U_n)$, shifted by a random offset $I \in [n+m]$. To extract pseudorandom bits, an arbitrary universal hash function H (with a proper output length) is applied to all $t = \Theta(d/(u \cdot (n+m)))$ bits in the same column.

Construction 1. *Given input U_d, the pseudorandom generator output*

$$h, h(G_1^1 G_1^2 \dots G_1^t), h(G_2^1 G_2^2 \dots G_2^t), \dots$$

where h is a universal hash function, and for each $1 \le i \le t$, G^i consists of $u = \Theta(n/\Delta)$ iid copies of $(f(U_n), U_n)$, with the first I bits of the first copy and the last $m+n-I$ bits of the last copy discarded, for $I \in_R [n+m]$ (using a new copy of I for each G^i). We let $t = \Theta(d/(u \cdot n))$.

If $f : \{0,1\}^n \to \{0,1\}^n$ is one-way, then setting parameters $m = n$, $\Delta = \log n$ and $\kappa = \omega(\log n)$, G is a PRG with seed length any $d = \omega(n^4)$ and stretch $d \cdot \Omega((\log n)/n)$.

The following corollary was pointed out to us by Yuval Ishai: If f is a one-way function with exponential security and linear circuit size, by using universal hash functions that have linear circuit size as constructed in [IKOS], we can obtain a PRG whose circuit complexity is linear in its stretch. Such pseudorandom generators (with circuit complexity linear in their stretch) are useful for cryptography with constant computational overhead [IKOS].

Corollary 5.15 (Pseudorandom generators with constant overhead)**.** *Suppose that there is a function $f : \{0,1\}^n \to \{0,1\}^n$ computable by uniform circuits of size $O(n)$ and such that for some constant $\alpha > 0$ and every constant c, f is $(n^c, 2^{-\alpha n})$ one-way. Then there exists a pseudorandom generator $G : \{0,1\}^d \to \{0,1\}^{2d}$ computable by uniform boolean circuits of size $O(d)$, for $d = O(n \cdot \text{polylog}(n))$.*

Proof. By Theorem 5.12, $G_{nb}(U_n)$ has uniform next-bit pseudoentropy at least $(1+\alpha)n$. By Theorem 5.14, there exists a pseudorandom generator $G : \{0,1\}^d \to \{0,1\}^{d\cdot(1+\alpha)}$ with seed length $d = O(n \log^3 n)$. We see from the construction (Construction 1) that G (i) performs $O(d/n)$ evaluations of f, for a total circuit size of $O(d)$ since f has $O(n)$ circuit size; (ii) applies hashing on all $\Theta(n/\alpha)$ columns and a total of $O(d)$ bits, for a total circuit size of $O(d)$ using universal hash functions computable by uniform circuits of linear size [IKOS]. Thus G has circuit size $O(d)$. We then do iterative composition [Gol] $\lceil 1/\alpha \rceil$ times to increase the output length to $2d$; this increases the circuit size by a constant factor. □

This result does not follow from the [HRV] construction alone, since their next-bit pseudoentropy generator requires hash functions that support "local list-decoding" and are not known to be implementable in linear size.

5.4.2 Saving Seed Length

In this section, we show how to save the seed length of [HRV]'s construction of pseudorandom generators from next-bit pseudoentropy generators, by a factor of $\Theta(n)$.

There are three steps in the construction:

1. Entropy equalization — discarding the first I bits of the first copy and the last $m - I$ bits of the last copy of G_{nb}. Since G_{nb} is highly unstructured, nothing can be said about the conditional pseudoentropy in any fixed bit, yet by discarding a random

prefix, each position is now a random bit in G_{nb}. By taking many copies of G_{nb}, the amortized loss of next-bit pseudoentropy is small.

Lemma 5.16. [HRV] *Let n be a security parameter, $m = m(n) = poly(n)$ and $\ell = \ell(n) = poly(n)$ be* poly(n) *time computable integer functions, where $\ell(n) > 1$. Let X be a distribution on $\{0,1\}^m$ with (T, ϵ)-next-bit pseudoentropy at least k, for $T = T(n)$, $\epsilon = \epsilon(n)$ and $k = k(n)$. Let J be uniformly distributed over $[m]$ and let $\tilde{X} = X_J^{(1)}, \ldots, X_m^{(1)}, \ldots, X_1^{(\ell)}, \ldots, X_{J-1}^{(\ell)}$, where $X^{(i)}$'s are iid copies of X. Then every bit of $\tilde{X}$ has $(T - O(\ell \cdot m), \ell \cdot \epsilon)$ conditional pseudoentropy at least $(\ell - 1)k/(\ell m)$, conditioned on previous bits of $\tilde{X}$ and J.* [1]

2. Converting conditional Shannon entropy to conditional min-entropy — taking multiple (parallel) copies. This generalizes the standard procedure of converting Shannon entropy to min-entropy by taking sufficiently many copies. Conditional pseudo-min-entropy is defined analogously to conditional pseudoentropy; see [HRV].

Lemma 5.17. [HRV] *Let n be a security parameter, $m = m(n) = poly(n)$ and $t = t(n) = poly(n)$ be* poly(n) *time computable integer functions. Let X be a distribution on $\{0,1\}^m$ where every bit of X has (T, ϵ) conditional pseudoentropy at least α, for $T = T(n)$, $\epsilon = \epsilon(n)$, $\alpha = \alpha(n)$. Then for every $\kappa = \kappa(n) > 0$ it holds that every block of $\left(X_1^{(1)}, X_1^{(2)}, \ldots, X_1^{(t)}\right)$, ..., $\left(X_m^{(1)}, X_m^{(2)}, \ldots, X_m^{(t)}\right)$, conditioned on previous blocks, has (T', ϵ') conditional pseudo-*min*-entropy α', where $X^{(i)}$'s are iid copies of X, and*

- $T' = T'(n) = T - O(m \cdot t)$,
- $\epsilon' = \epsilon'(n) = t^2 \cdot (\epsilon + 2^{-\kappa} + 2^{-ct})$ *for a universal constant $c > 0$, and*
- $\alpha' = \alpha'(n) = t \cdot \alpha - \Gamma(t, \kappa)$, *for $\Gamma(t,k) \in O(\sqrt{t \cdot \kappa} \cdot \log t)$.*

[1] This is slightly stronger than the version in [HRV], which does not condition on J. However, it is easy to see from their proof that one can additionally condition on J.

3. Randomness extraction. This step is essentially a computational version of block source extraction. At the previous step, the amount of next-bit pseudo-min-entropy in each block is known. So we may choose hash functions of fixed output length to make the output pseudorandom.

   Lemma 5.18. [HRV] *Let n be a security parameter, $m = m(n) = poly(n)$, $t = t(n) = poly(n)$, $\alpha = \alpha(n) \in [t(n)]$ and $\kappa = \kappa(n) \in [\alpha(n)]$ be* poly(n) *time computable integer functions. Let $\{h_s : \{0,1\}^t \to \{0,1\}^{\alpha-\kappa}\}$ be some family of universal hash functions. Let $X_1, \ldots, X_m$ be distributions on $\{0,1\}^t$ such that every X_i conditioned on $X_1, \ldots, X_{i-1}$ has (T, ϵ) conditional pseudo-min-entropy α, for $T = T(n)$ and $\epsilon = \epsilon(n)$. Then $(h, h(X_1), \ldots, h(X_m))$ is $(T - m \cdot t^{O(1)}, m \cdot (\epsilon + 2^{-\kappa/2}))$ pseudorandom, where h is a random hash function from the family.*

We refer to [HRV] for the proofs and detailed explanation of intuition behind these steps.

The seed length blow up in [HRV] comes from Step 1 (Entropy Equalization) and Step 2 (Converting to conditional min-entropy), as each involves repeating the current generator on many independent seeds. We show how to save the blow up due to Entropy Equalization, by showing how randomness from a "few" copies of G_{nb} can be used to generate more copies of G_{nb}, and iteratively.

Specifically, we show that the [HRV] construction above, but taking only $\ell = 2$ copies in Entropy Equalization, gives rise to a "Z-seeded" PRG, one that given input distribution Z outputs some $(\tilde{Z}, \tilde{U}_\sigma)$ indistinguishable from (Z, U_σ). (If Z were uniformly distributed in $\{0,1\}^d$, this would be a standard PRG.) Then we apply iterative composition (just like iterative composition for standard PRGs [Gol]) to increase the number of pseudorandom bits (without changing the seed distribution Z).

We begin by describing the iterative composition of Z-seeded PRGs, illustrated in

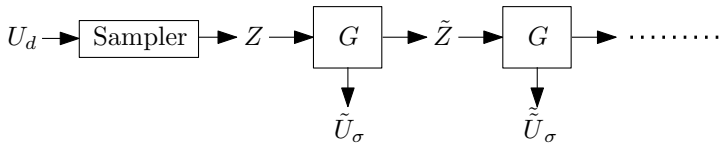

Figure 5.3: Iterative composition for Z-seeded PRG G

Figure 5.3.

Lemma 5.19 (Iterative composition of Z-seeded PRGs)**.** *Let n be a security parameter. Let $\sigma = \sigma(n), \ell = \ell(n) = \text{poly}(n)$ be $\text{poly}(n)$ time computable functions. Let $Z = Z(n)$ be a distribution samplable in $\text{poly}(n)$ time using $d = d(n)$ bits of randomness. Let G be a generator computable in $\text{poly}(n)$ time such that $G(Z) = (\tilde{Z}, \tilde{U}_\sigma)$ is (T, ϵ)-indistinguishable from (Z, U_σ), for $T = T(n), \epsilon = \epsilon(n)$. Then there is a $(T - \text{poly}(n), \ell\epsilon)$ pseudorandom generator $G' : \{0,1\}^d \to \{0,1\}^{\ell\sigma}$ computable in $\text{poly}(n)$ time.*

Proof. Consider the following algorithm $G_\ell(z)$: If $\ell = 0$ then output ϵ (the empty string). If $\ell \geq 0$ then let $(\tilde{z}, \tilde{u}) = G(z)$ and output $G_{\ell-1}(\tilde{z}) \circ \tilde{u}$.

We claim that $G_\ell(Z)$ is pseudorandom, so we obtain the desired PRG G' by composing G_ℓ with algorithm that samples Z given d random bits. Clearly G' runs in $\text{poly}(n)$ time. We show the pseudorandomness of $G_\ell(Z)$ by a hybrid argument.

Suppose for contradiction that $G_\ell(Z)$ is not $(T', \ell\epsilon)$-pseudorandom, i.e. there exists a T' time $\ell\epsilon$-distinguisher D between $G_\ell(Z)$ and $U_{\ell\sigma}$. For each $0 \leq i \leq \ell$ define a hybrid distribution $H_i = (G_i(Z), U_{(\ell-i)\sigma})$. Thus $H_0 = U_{\ell\sigma}$ and $H_\ell = G_\ell(Z)$. Let $I \in_R [\ell]$. Then

$$\mathbb{E}\left[D(H_I) - D(H_{I-1})\right] = \frac{1}{\ell}\sum_{k=1}^{\ell}\mathbb{E}\left[D(H_k) - D(H_{k-1})\right] = \frac{1}{\ell}\mathbb{E}[D(G_\ell(Z)) - D(U_{\ell\sigma})] > \epsilon.$$

We use this to break the pseudorandomness property of G. Denote $G(Z) = (\tilde{Z}, \tilde{U}_\sigma)$. We claim that $D'(z, u) = D(G_{I-1}(z) \circ u \circ U_{(\ell-I)\sigma})$, where $I \in_R [\ell]$ and $|u| = \sigma$, ϵ'-distinguishes (Z, U_σ) from $(\tilde{Z}, \tilde{U}_\sigma)$. Notice that given $(\tilde{z}, \tilde{u}) = G(z)$, we have $(G_{I-1}(\tilde{z}), \tilde{u}) = G_I(z)$ by

definition of G_ℓ. Thus, $D'(\tilde{Z}, \tilde{U}_\sigma) = D(G_I(Z) \circ U_{(\ell-I)\sigma}) = D(H_I)$ whereas $D'(Z, U_\sigma) = D(G_{I-1}(Z) \circ U_\sigma \circ U_{(\ell-I)\sigma}) = D(H_{I-1})$. It follows that

$$\mathbb{E}[D'(Z, U_\sigma) - D'(\tilde{Z}, \tilde{U}_\sigma)] = \mathbb{E}[D(H_I) - D(H_{I-1})] > \epsilon.$$

Moreover, D' is computable in $T' + \text{poly}(n)$ time. For an appropriate $T' = T - \text{poly}(n)$, this contradicts that (Z, U_σ) and $(\tilde{Z}, \tilde{U}_\sigma)$ are (T, ϵ) indistinguishable. Therefore, $G_\ell(Z)$ is $(T - \text{poly}(n), \ell\epsilon)$-pseudorandom. □

We now show how to construct a Z-seeded PRG G from any next-bit pseudoentropy generator G_{nb}, as demonstrated in Figure 5.4. By applying iterative composition, this gives rise to a seed-efficient construction of PRG from a pseudoentropy generator G_{nb} which should be compared to the original construction illustrated in Figure 5.1.

Theorem 5.20 (Next-bit pseudoentropy $\Longrightarrow$ Z-seeded PRG)**.** *Let n be a security parameter. Let $\Delta = \Delta(n) \in [1/\text{poly}(n), n]$, $m = m(n)$, $\kappa = \kappa(n) \in [n/2]$ be polynomial-time computable functions. For every polynomial-time computable $G_{nb} : \{0,1\}^n \to \{0,1\}^m$ such that $G_{nb}(U_n)$ has (T, ϵ) next-bit pseudoentropy at least $n + \Delta$ (for $T = T(n)$ and $\epsilon = \epsilon(n)$), there exists distribution $Z = Z(n)$ and generator G such that:*

1. *Z is samplable in polynomial time using*

$$d = O\left(\frac{m^2 n\kappa \log^2\left(\frac{n\kappa}{\Delta}\right)}{\Delta^2}\right)$$

   bits of randomness;

2. *G is computable in polynomial time and $G(Z)$ is $(T - n^{O(1)}, n^{O(1)} \cdot (\epsilon + 2^{-\kappa}))$ indistinguishable from (Z, U), U being uniformly random string of length $\Omega(d \cdot \Delta/n)$.*

Moreover, G is computable with $O(d/n)$ (uniform and independent) oracle calls to G_{nb}.

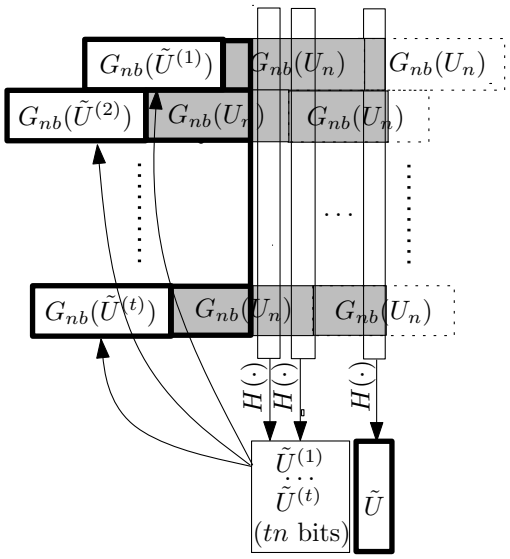

Figure 5.4: Construction of Z-seeded PRG G from any next-bit pseudoentropy generator G_{nb}. The shaded area represents input Z. The bold boxes are the output $G(Z) = (\tilde{Z}, \tilde{U})$. The ith row is shifted by a random offset $J^{(i)} \in [n+m]$. An arbitrary universal hash function H (with a proper output length) is then applied to all bits in the same column, producing pseudorandom bits $(\tilde{U}^{(1)}, \ldots, \tilde{U}^{(t)}, \tilde{U})$ where each $\tilde{U}^{(i)}$ is of length n. We then apply G_{nb} to each $\tilde{U}^{(i)}$. Together with unused bits of Z they form $\tilde{Z}$.
We ignore H, $J^{(1)}$, ..., $J^{(t)}$ in the figure since they are the same in the input and output of G.

Proof. Let t be a parameter to be fixed later. Let $J^{(1)}, \ldots, J^{(t)}$ be t iid copies of $J \in_R [m]$, and $H \in_R \{0,1\}^t$. Consider

$$Z = \left(H \circ J^{(1)} \ldots J^{(t)} \circ G_{nb}(U^{(1)})_{1,\ldots,J^{(1)}-1} \ldots G_{nb}(U^{(t)})_{1,\ldots,J^{(t)}-1} \circ G_{nb}(U^{(t+1)}) \ldots G_{nb}(U^{(2t)})\right)$$

where $U^{(i)}$'s are iid copies of U_n. Z is clearly samplable in polynomial time using $d = t + t \cdot (\log m + 2n) = O(tn)$ bits of randomness.

We now define G. Interpret G's input as

$$h \circ j^{(1)} \ldots j^{(t)} \circ G_{nb}(u^{(1)})_{1,\ldots,j^{(1)}-1} \ldots G_{nb}(u^{(t)})_{1,\ldots,j^{(t)}-1} \circ G_{nb}(u^{(t+1)}) \ldots G_{nb}(u^{(2t)})$$

where $h, j^{(i)}, u^{(i)}$ are strings of length t, $\log m$ and n respectively. G is defined as follows:

1. Entropy Equalization: For each $i \in [t]$ (that is, for each "row"), we set $y^{(i)} = \left(G_{nb}(u^{(t+i)})_{j^{(i)},\ldots,m} \circ G_{nb}(u^{(i)})_{1,\ldots,j^{(i)}-1}\right)$;

2. Apply a universal hash function $h : \{0,1\}^t \to \{0,1\}^{t'}$ where t' will be chosen later so that $t'm > tn$, on $y_j^{(1)} \circ \cdots \circ y_j^{(t)}$, for each $j \in [m]$ (that is, for each "column"). Thus m calls to h produce $t'm$ bits in total:

$$\tilde{u}^{(1)} \dots \tilde{u}^{(t)} \circ \tilde{u} = h(y_1^{(1)}, \dots, y_1^{(t)}) \circ h(y_2^{(1)}, \dots, y_2^{(t)}) \dots h(y_m^{(1)}, \dots, y_m^{(t)})$$

where $\tilde{u}^{(1)}, \dots, \tilde{u}^{(t)}$ are n-bit strings, and $\tilde{u}$ is the remaining $t'm - tn$ bits.

3. Output

$$h \circ j^{(1)} \dots j^{(t)} \circ G_{nb}(u^{(t+1)})_{1,\dots,j^{(1)}-1} \dots G_{nb}(u^{(2t)})_{1,\dots,j^{(t)}-1} \circ G_{nb}(\tilde{u}^{(1)}) \dots G_{nb}(\tilde{u}^{(t)}) \circ \tilde{u}.$$

We now prove that $G(Z)$ is computationally indistinguishable from $(Z \circ U)$ where $U = U_{t'm-tn}$ (i.e. a $t'm - tn$ bit random string). Suppose we run $G(Z)$ to obtain

$$G(Z) = \left(H \circ W \circ G_{nb}(\tilde{U}^{(1)}) \dots G_{nb}(\tilde{U}^{(t)}) \circ \tilde{U}\right)$$

where $\tilde{U}$ is of length $t'm - tn$, and

$$W = \left(J^{(1)} \dots J^{(t)} \circ G_{nb}(U^{(t+1)})_{1,\dots,J^{(1)}-1} \dots G_{nb}(U^{(2t)})_{1,\dots,J^{(t)}-1}\right).$$

In the following, we will show that $G(Z) = \left(H \circ W \circ G_{nb}(\tilde{U}^{(1)}) \dots G_{nb}(\tilde{U}^{(t)}) \circ \tilde{U}\right)$ is computationally indistinguishable from $(Z \circ U) = \left(H \circ W \circ G_{nb}^{(1)} \dots G_{nb}^{(t)} \circ U\right)$, where $G_{nb}^{(i)}$'s are iid copies of $G_{nb}(U_n)$. The proof is essentially the same 3-step analysis as in Haitner et. al, with the tweak that the conditional pseudoentropy and conditional pseudo-min-entropy are now additionally conditioned on W, and the final indistinguishablility holds for W taking any value.

In Step 1, we set $Y^{(i)} = G_{nb}(U^{(t+i)})_{J^{(i)},\dots,m} \circ G_{nb}(U^{(i)})_{1,\dots,J^{(i)}-1}$. Recall that $G_{nb}(U_n)$ has (T, ϵ) next-bit pseudoentropy at least $n + \Delta$. Applying Lemma 5.16 (Entropy Equalization) with $\ell = 2$, $X^{(1)} = G_{nb}(U^{(t+i)})$ and $X^{(2)} = G_{nb}(U^{(i)})$, we obtain that every bit of

$Y^{(i)}$ conditioned on previous bits of $Y^{(i)}$, $G_{nb}(U^{(t+i)})_{1,\ldots,J^{(i)}-1}$ and $J^{(i)}$, has $(T - O(m), 2\epsilon)$ conditional pseudoentropy at least $(\Delta + n)/m$.

Recall that $Y^{(1)}, \ldots, Y^{(t)}$ are t independent rows. By Lemma 5.17 (t-fold parallel repetition), $Y_j^{(1)}, \ldots, Y_j^{(t)}$ has $(T - O(mt), t^2 \cdot (2\epsilon + 2^{-\kappa} + 2^{-ct}))$ conditional pseudo-min-entropy at least $\alpha = t(\Delta + n)/m - O(\sqrt{t\kappa} \log t)$, conditioned on W and all $Y_k^{(1)}, \ldots, Y_k^{(t)}$ where $k < j$.

In Step 2, we apply hashing to each "column". By Lemma 5.18, if we set $t' = \alpha - 2\kappa$, then $(H \circ \tilde{U}^{(1)} \ldots \tilde{U}^{(t)} \circ \tilde{U})$ and $(H \circ U_{tn} \circ U)$ are $(T - O(mt) - mt^{O(1)}, mt^2 \cdot (2\epsilon + 2^{-\kappa} + 2^{-\Omega(t)}) + m \cdot 2^{-\kappa})$ indistinguishable, for W taking any value. Thus the same can be said about $(H \circ G_{nb}(\tilde{U}^{(1)}) \ldots G_{nb}(\tilde{U}^{(t)}) \circ \tilde{U})$ and $(H \circ G_{nb}^{(1)} \ldots G_{nb}^{(t)} \circ U)$. Thus we conclude that

$$G(Z) = \left(H \circ W \circ G_{nb}(\tilde{U}^{(1)}) \ldots G_{nb}(\tilde{U}^{(t)}) \circ \tilde{U}\right)$$

is $(T - O(mt) - mt^{O(1)}, mt^2 \cdot (2\epsilon + 2^{-\kappa} + 2^{-\Omega(t)}) + m \cdot 2^{-\kappa})$ indistinguishable from

$$\left(H \circ W \circ G_{nb}^{(1)} \ldots G_{nb}^{(t)} \circ U\right) = (Z \circ U)\,.$$

We are left to set the parameters. We need to guarantee

$$\Omega\left(\frac{\Delta}{n} d\right) \leq t'm - tn = \left(\frac{t(\Delta + n)}{m} - O(\sqrt{t\kappa} \log t) - 2\kappa\right) m - tn$$

where $d = O(tn)$. Assuming $\kappa \leq O(t)$, this can be simplified to

$$\frac{\sqrt{t}}{\log t} \geq O\left(\frac{m\sqrt{\kappa}}{\Delta}\right)$$

which is guaranteed for an appropriate choice of

$$t = O\left(\frac{m^2 \kappa \log^2\left(\frac{m\kappa}{\Delta}\right)}{\Delta^2}\right),$$

and consequently

$$d = O(tn) = O\left(\frac{m^2 n\kappa \log^2\left(\frac{m\kappa}{\Delta}\right)}{\Delta^2}\right) = O\left(\frac{m^2 n\kappa \log^2\left(\frac{n\kappa}{\Delta}\right)}{\Delta^2}\right).$$

So (Z, U) and $G(Z)$ are $(T - O(ts) - mt^{O(1)}, mt^2 \cdot (2\epsilon + 2^{-\kappa} + 2^{-\Omega(t)}) + m2^{-\kappa}) = (T - n^{O(1)}, n^{O(1)} \cdot (\epsilon + 2^{-\kappa}))$ indistinguishable. Moreover, G makes $O(d/n)$ uniformly random oracle calls to G_{nb}. □

Combining Lemma 5.19 and Theorem 5.20, we obtain a seed length efficient construction of pseudorandom generators:

Corollary 5.21 (Next-bit pseudoentropy $\Longrightarrow$ pseudorandomness)**.** *Let n be a security parameter. Let $\Delta = \Delta(n) \in [1/\text{poly}(n), n]$, $m = m(n)$, $\kappa = \kappa(n) \in [n/2]$, $\ell = \ell(n) = \text{poly}(n)$ be computable in time* poly(n)*. For every polynomial time computable $G_{nb} : \{0,1\}^n \to \{0,1\}^m$ such that $G_{nb}(U_n)$ has (T, ϵ) next-bit pseudoentropy at least $n + \Delta$ (for $T = T(n)$ and $\epsilon = \epsilon(n)$), there exists a polynomial-time computable $(T - n^{O(1)}, n^{O(1)} \cdot (\epsilon + 2^{-\kappa}))$ pseudorandom generator $G : \{0,1\}^d \to \{0,1\}^{d \cdot (\ell\Delta/n)}$ with seed length*

$$d = O\left(\frac{m^2 n\kappa \log^2\left(\frac{n\kappa}{\Delta}\right)}{\Delta^2}\right).$$

Moreover, G is computable with $O(\ell d/n)$ (uniformly random) oracle calls to G_{nb}.

Proof. By Theorem 5.20, there is a Z-seeded PRG G' where Z is samplable in polynomial time from U_d, and $G'(Z)$ is $(T - n^{O(1)}, n^{O(1)} \cdot (\epsilon + 2^{-\kappa}))$ indistinguishable from (Z, U). By Lemma 5.19 there exists a pseudorandom generator G with the above parameters. □

In particular, from a one-way function $f : \{0,1\}^n \to \{0,1\}^n$ and setting $m = n$, $\Delta = \log n$, $\kappa = \omega(\log n)$, $\ell = 2n/\Delta$ we can construct a pseudorandom generator of seed length any $d = \omega(n^3 \log n)$. Like [HRV], the construction obtains $\Theta(\log n)$ bits of additive stretch per invocation of the one-way function, which is optimal by [GGKT].

Chapter 6

Impossibility of Black-Box Construction of Succinct Non-Interactive Argument from Uniform Assumptions

A result of Gentry and Wichs [GW] shows that there is no black-box construction of succinct non-interactive arguments (SNARGs) from any natural cryptographic assumption (formally, they consider *falsifiable* cryptographic assumptions: ones that are defined by a polynomial-time security game). Their result relies on the (mild) assumption that there exist *hard subset membership problems*, which is equivalent to the existence of subexponentially hard one-way functions. One limitation is that they need to work in the non-uniform setting, in part due to their use of the Min-Max Theorem in Lemma 3.1 of [GW].

In this chapter, we show how to obtain the analogous result in the *uniform setting*, as an application of the Uniform Min-Max Theorem from Chapter 2. More precisely, we apply a

low time complexity version of Lemma 3.1 of [GW] (Chapter 2, Theorem 2.18) proved using the Uniform Min-Max Theorem. We show that, assuming that there exist subexponentially hard one-way functions that are secure against uniform algorithms, there is no black-box construction of SNARGs based on cryptographic assumptions where security is measured against uniform algorithms (unless the assumption is already false).

A succinct non-interactive argument (SNARG) is a non-interactive argument system where the proof size is bounded by a fixed polynomial, for all instances and witnesses whose size can be an arbitrarily large polynomial. Formally,

Definition 6.1 (SNARG)**.** Let L be an **NP** *language associated with relation* R. We say that a tuple (G, P, V) of probabilistic polynomial-time (PPT) algorithms is a *succinct non-interactive argument for* R if the following properties hold:

- **Completeness**: For all $(x, w) \in R$, if we choose $(\mathsf{CRS}, \mathsf{PRIV}) \leftarrow G(1^n), \Pi \leftarrow P(\mathsf{CRS}, x, w)$, then
$$\Pr\left[V\left(\mathsf{PRIV}, x, \Pi\right) = 0\right] = \mathrm{negl}(n).$$
- **Soundness**: For every PPT algorithm (efficient *adversary*) A, if we choose $(\mathsf{CRS}, \mathsf{PRIV}) \leftarrow G(1^n), (X, \Pi) \leftarrow A(1^n, \mathsf{CRS})$, then
$$\Pr\left[V\left(\mathsf{PRIV}, X, \Pi\right) = 1 \wedge X \notin L\right] = \mathrm{negl}(n).$$
- **Succinctness**: For all $(x, w) \in \mathrm{supp}(X, W)$ and $\mathsf{crs} \in \mathrm{supp}(\mathsf{CRS})$, the length of the proof $\pi = P(\mathsf{crs}, x, w)$ is $|\pi| = \mathrm{poly}(n)(|x| + |w|)^{o(1)}$. We also consider a weaker variant called *slightly succinct*, where we require the length of a proof to be $|\pi| = \mathrm{poly}(n)(|x| + |w|)^{\alpha} + o(|x| + |w|)$ for some constant $\alpha < 1$.[1]

[1]Earlier versions of [GW] contained a minor bug in the definition of slight succinctness. We use the corrected definition from the current version of their paper.

Our notion of a falsifiable cryptographic assumption is analogous to [GW], except that the adversary A is a *uniform* algorithm instead of circuit:

Definition 6.2 (Falsifiable assumption)**.** Given an interactive PPT algorithm Chal (the challenger), the *uniform falsifiable (cryptographic) assumption (associated with)* Chal states that for all (uniform) PPT algorithms H, the probability that $\text{Chal}(1^n)$ outputs a special symbol win after interacting with $H(1^n)$ is at most $\text{negl}(n)$ for all sufficiently large n.

For any randomized (possibly inefficient) function H, we let $\text{Break}_H(n)$ denote the above probability and say that H *breaks the assumption* if $\text{Break}_H(n) \geq 1/\text{poly}(n)$ for infinitely many n.

Remark. An alternative definition of falsifiable assumption allows specifying a constant β, and says that the probability $\text{Chal}(1^n)$ outputs win is at most $\beta + \text{negl}(n)$. However, it turns out that setting $\beta = 0$, i.e. our definition above, is without loss of generality [HH]. We adopt the simpler definition because it is convenient for our proof.

Next we define black-box reductions:

Definition 6.3 (Adversary and reduction)**.** For a randomized function A and a constant $c \in \mathbb{N}$, we say (A, c) is a (G, P, V)-*adversary* if $|A(1^n, \mathsf{crs})| \leq n^c$ and A violates the soundness condition infinitely often, i.e. if we choose $(\mathsf{CRS}, \mathsf{PRIV}) \leftarrow G(1^n), (X, \Pi) \leftarrow A(1^n, \mathsf{CRS})$, then

$$\Pr\left[V\left(\mathsf{PRIV}, X, \Pi\right) = 1 \wedge X \notin L\right] \geq n^{-c}$$

for infinitely many n. We say (A, c) is an *a.e.* (G, P, V)-*adversary* if A violates soundness for all sufficiently large n.

A *uniform black-box reduction showing the soundness of* (G, P, V) *based on a falsifiable assumption* Chal is a family of (uniform) probabilistic oracle algorithms $\{\text{Red}_c\}$ (one for each $c \in \mathbb{N}$) such that for every (G, P, V)-adversary (A, c), $\text{Red}_c^A(1^n)$ breaks the assumption and runs in time $\text{poly}_c(n)$ (i.e. a polynomial that depends on c).

For a probabilistic oracle algorithm Red, we say a query $(1^m, \mathsf{crs})$ of $\mathrm{Red}(1^n)$ has *length* m. In general, $\mathrm{Red}(1^n)$ may make queries of various lengths. We say Red is *length-mapping* if for all n, all queries of $\mathrm{Red}(1^n)$ are of the same length $m = m(n)$ and m is computable in time $\mathrm{poly}(n)$; denote this m by $\mathrm{query}_{\mathrm{Red}}(n)$. Most reductions in cryptography set $m = n$ i.e. preserve length; that is, the security parameter of (G, P, V) is equal to that of the assumption.

Following [GW], our results assume the existence of *hard subset membership problem.*

Definition 6.4 (Uniformly hard subset membership problem)**.** Let n be a security parameter, L be an **NP** language associated with relation R. We say $((X, W), U)$ is a *subset membership problem for* R if $(X, W) = (X, W)(n)$ is a $\mathrm{poly}(n)$-time samplable joint distribution whose support lies in R, and $U = U(n)$ a $\mathrm{poly}(n)$-time samplable distribution with $\Pr[U \notin L] \geq n^{-O(1)}$.

A subset membership problem $((X, W), U)$ is a *subexponentially hard* if X and U are $(2^{\Omega(n^\delta)}, 2^{-\Omega(n^\delta)})$-indistinguishable for a constant $\delta > 0$. We say it is *exponentially hard* if the above occurs and $|x| + |w| = O(n^\delta)$ for every $(x, w) \in \mathrm{supp}(X, W)$.

This is a relatively mild assumption; the existence of subexponentially hard subset membership problems is equivalent to the existence of subexponentially hard one-way functions.

Remark. Our definition of a hard subset membership problem is a variant of [GW] that is needed in the uniform setting, but also can be used in the nonuniform setting of [GW]. In [GW], they require that X is indistinguishable from a (not necessarily samplable) distribution U whose support is disjoint from L, whereas we require that U is samplable and allow it to hit L with probability up to $1 - n^{O(1)}$.

We now state the uniform analogue of the main result of [GW]. Compared to [GW],

our Theorem 6.5 makes the weaker assumption of subexponentially hard subset membership problem with respect to *uniform* algorithms, with the conclusion that a *uniform* falsifiable assumption cannot be broken also being weaker (unless the assumption is false).

Theorem 6.5 (Main theorem)**.** *Let L be an* **NP** *language associated with relation R that has* a *subexponentially hard subset membership problem, and (G, P, V) be an non-interactive proof system for R that satisfies the completeness and succinctness properties. Then for every uniform falsifiable assumption* Chal*, one of the following must hold:*

- *The assumption* Chal *is false, or*
- *There is no uniform black-box reduction showing the soundness of (G, P, V) based on* Chal*.*

The same conclusion also holds if we assume an exponentially *hard subset membership problem, and (G, P, V) is only* slightly *succinct.*

The same conclusion also holds if we require the uniform black-box reduction to work only for all (G, P, V)-adversary (A, c) where c is sufficiently large*.*

To prove it in the nonuniform setting, the main idea of [GW] is showing that any SNARG (G, P, V) has an inefficient adversary A that can be (efficiently) "simulated" i.e. there exists an efficient algorithm Sim (the simulator) such that $\mathrm{Red}^A(1^n) \approx \mathrm{Red}^{\mathsf{Sim}}(1^n)$ for all PPT oracle algorithms Red (cf. [GW] Lemma 4.1). Thus, if there were a black-box reduction Red showing the soundness of (G, P, V) based on a falsifiable assumption, then Red^A would break the falsifiable assumption (since A is an adversary) and so would $\mathrm{Red}^{\mathsf{Sim}}$ (since $\mathrm{Red}^A(1^n) \approx \mathrm{Red}^{\mathsf{Sim}}(1^n)$). In other words, the assumption would be false.

To prove it in the uniform setting, we use a similar approach with several necessary tweaks. We show that there is an *adversary simulator* Sim, which is a PPT algorithm that

with noticeable probability outputs a randomized circuit B_n that simulates some A_n, where A_n is an (inefficient) adversary on security parameter n:

Lemma 6.6 (Existence of adversary simulator)**.** *Let L be an* **NP** *language associated with relation R that has* a *subexponentially hard subset membership problem $((X,W),U)$, and (G,P,V) be a non-interactive proof system for R that satisfies the completeness and succinctness properties. Let n be a security parameter, $((X,W),U) = ((X,W),U)(n)$, $(\mathsf{PRIV},\mathsf{CRS}) = G(1^n)$, and $\Pi = P(\mathsf{CRS},X,W)$. Let $\ell = \ell(n) \geq n$ be a polynomial bound on the running time of $G(1^n)$ as well as the proof size $|\Pi|$, and c be a constant such that $|X| + |\Pi| \leq n^c$.*

Let Red *be any length-mapping PPT oracle algorithm where* $\mathrm{query}_{\mathrm{Red}}(k) = \omega(1)$*. Then there is a PPT algorithm* Sim *such that for all polynomials $q(\cdot)$, for all sufficiently large k, and for $n = \mathrm{query}_{\mathrm{Red}}(k)$, w.p. at least* $1/\mathrm{poly}(k)$*, $\mathsf{Sim}(1^k)$ outputs a randomized circuit B_n such that there is a randomized function A_n satisfying:*

- *(A_n, c) is a (G,P,V)-adversary on the security parameter n;*
- $\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{B_n}}(k) < 1/q(k)$*. (w.l.o.g. B_n only takes inputs $(1^n,\cdot)$.)*

The same conclusion also holds if we assume an exponentially *hard subset membership problem, and that (G,P,V) is only* slightly *succinct.*

Note that Lemma 6.6 is only stated for length-mapping reductions (unlike [GW]). We remove this restriction by a general technique when we prove the main theorem in Section 6.2.

6.1 Proof of Existence of Adversary Simulator

The proof is set up as follows. Given a subexponentially hard subset membership problem $((X,W),U)$, we can w.l.o.g. assume that X and U are $(2^{d\ell}, 2^{-d\ell})$-indistinguishable

for a sufficiently large constant d, where $\ell = \ell(n)$ is a bound on the length of the proof output by $P(\mathsf{crs}, x, w)$ for $(x, w) \in \mathrm{supp}(X, W)$ and $\mathsf{crs} \in \mathrm{supp}(\mathsf{CRS})$. (If X and U are only $(2^{n^\delta}, 2^{-n^\delta})$-indistinguishable for some $\delta > 0$, we simply re-index, replacing $X(n)$ with $X((d\ell)^{1/\delta}))$.) If $((X, W), U)$ is exponentially hard, we can also ensure that X and U are $(2^{d\ell}, 2^{-d\ell})$-indistinguishable by re-indexing so that $\ell \leq \mathrm{poly}(n) \cdot (|x| + |w|)^\alpha + o(|x| + |w|) = O(|x| + |w|)/d$ for all $(x, w) \in \mathrm{supp}(X, W)$ and $\mathsf{crs} \in \mathrm{supp}(\mathsf{CRS})$.

Overview of the Proof. Consider the joint distribution (CRS, X, Π) where $\mathsf{CRS} = \mathsf{CRS}(n)$ is the distribution of the common reference string, and $\Pi = \Pi(n)$ is the ℓ-bit proof produced by P for the instance/witness pair (X, W). Using the fact that Π is short (by succinctness), and X and U are ϵ-indistinguishable for $\epsilon = 2^{-O(\ell)}$, we can apply Chapter 2, Theorem 2.18 to conclude that, for every $2^{O(\ell)}$-time oracle algorithm D, there is a $\mathrm{poly}(2^\ell, 1/\epsilon)$-time randomized algorithm R that outputs a randomized circuit F_n such that with probability at least $\Omega(\epsilon^2/\ell)$ over F_n,

$$\mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))] < 2\epsilon \qquad (\star)$$

where Q can be any poly-time samplable distribution.

An adversary A_n can be defined to be $A_n(1^n, \mathsf{crs}) = (U, F_n(\mathsf{crs}, U))$ for *any* F_n where $(\star)$ holds, for an appropriate choice of D. (Note that F_n depends on our choice of D.) If we take D to be the verifier V, then we can show that such A_n breaks soundness on security parameter n. Indeed, V accepts (X, Π) with high probability, so by $(\star)$ it must also accept $(U, F_n(\mathsf{CRS}, U)) = A_n(1^n, \mathsf{CRS})$ with high probability. (Some extra work is needed to deal with the fact that V can access its private coins PRIV in addition to CRS.)

Thus we only need to argue that, for an appropriate choice of D, such A_n is simulated by some randomized circuit B_n generated by a PPT algorithm Sim; then combining the two choices of D will yield the desired adversary A_n. Our choice of B_n is the randomized circuit

such that $B_n(1^n, \mathsf{CRS}) = (X, \Pi)$. If we appropriately construct D from the reduction Red and challenger Chal, then using $(\star)$ we can show that

$$\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{B_n}}(k) \leq \mathrm{poly}(k) \cdot 2^{-O(\ell)},$$

where $\ell = \ell(n)$ for $n = \mathrm{query}_{\mathrm{Red}}(k)$. (If $\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{B_n}}(k) > \mathrm{poly}(k) \cdot 2^{-O(\ell)}$, then we could use Red and Chal to construct a $2^{-O(\ell)}$-distinguisher between $(\mathsf{CRS}, B_n(1^n, \mathsf{CRS})) = (\mathsf{CRS}, X, \Pi)$ and $(\mathsf{CRS}, A_n(1^n, \mathsf{CRS})) = (\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))$, violating $(\star)$.)

This completes the proof provided that $2^{-O(\ell)} \leq 1/\mathrm{poly}(k)$, which follows if Red does not make queries that are too short. If instead $2^{-O(\ell)} > 1/\mathrm{poly}(k)$, then we construct a simulator B_n differently — simply by letting B_n be such that $B_n(1^n, \mathsf{crs}) = (U, F_n(\mathsf{crs}, U)))$ where F_n is the random output of R. Then with probability at least $\Omega(\epsilon^2/\ell) \geq 1/\mathrm{poly}(k)$ over F_n, $(\star)$ holds for F_n, hence we can define the adversary A_n from F_n (defined to be $A_n(1^n, \mathsf{crs}) = (U, F_n(\mathsf{crs}, U))$, as explained above) to obtain a perfect simulator $B_n = A_n$. (Gentry and Wichs [GW] handle short queries using nonuniformity — by hardcoding the answers to all short queries.)

Lemma 6.7 (Existence of adversary simulator)**.** *Let L be an* **NP** *language associated with relation R that has* a *subset membership problem $((X, W), U)$, and (G, P, V) is a non-interactive proof system for R that satisfies the completeness property. Let n be a security parameter, $((X, W), U) = ((X, W), U)(n)$, $(\mathsf{PRIV}, \mathsf{CRS}) = G(1^n)$, $\Pi = P(\mathsf{CRS}, X, W)$. Let $\ell = \ell(n) \geq n$ be a polynomial bound on the running time of $G(1^n)$ as well as the proof size $|\Pi|$, and c be a constant such that $|X| + |\Pi| \leq n^c$.*

Suppose X and U are ϵ-indistinguishable for all t-time randomized algorithms, for appropriate $\epsilon = 2^{-O(\ell)}$ and $t = 2^{O(\ell)}$. Let Red *be any length-mapping PPT oracle algorithm where* $\mathrm{query}_{\mathrm{Red}}(k) = \omega(1)$*. Then there is a PPT algorithm* Sim *such that for all polynomials $q(\cdot)$, for all sufficiently large k, and for $n = \mathrm{query}_{\mathrm{Red}}(k)$, w.p. at least* $1/\mathrm{poly}(k)$*,* $\mathsf{Sim}(1^k)$

outputs a randomized circuit B_n such that there is a randomized function A_n satisfying:

- *(A_n, c) is a (G, P, V)-adversary on the security parameter n;*
- $\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{B_n}}(k) < 1/q(k)$. *(w.l.o.g. B_n only takes inputs $(1^n, \cdot)$.)*

Proof. Let S be the PPT algorithm that on input $(1^n, \mathsf{crs})$ samples $(x, w) \leftarrow (X, W)$ and outputs $(x, P(\mathsf{crs}, x, w))$, so that $S(1^n, \mathsf{CRS}) = (X, \Pi)$. For technical convenience we assume $|\mathsf{CRS}| = \ell/2$. To construct Sim, we shall apply Theorem Chapter 2, 2.18 to the following oracle algorithm D:

Claim 6.8. Let $Q = (U_{\ell/2}, U)$ (where $U_{\ell/2}$ is uniform on $\{0,1\}^{\ell/2}$ and independent from U). There is a $t' = 2^{O(\ell)} \cdot \mathrm{poly}(1/\epsilon)$-time oracle algorithm D such that the following holds for all polynomials $q(\cdot)$, all sufficiently large n, and all randomized functions $F_n : \mathrm{supp}(Q) \to \{0,1\}^\ell$ satisfying

$$\mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))] < \epsilon' = 2\epsilon.$$

Define

$$A_n(1^n, \mathsf{crs}) = \begin{cases} (U, F_n(\mathsf{crs}, U)), & \mathsf{crs} \in \mathrm{supp}(\mathsf{CRS}) \\ S(1^n, \mathsf{crs}), & \mathsf{crs} \notin \mathrm{supp}(\mathsf{CRS}) \end{cases}.$$

Then

- A_n break soundness of (G, P, V) on security parameter n; and
- For all $k \leq 2^\ell$ such that $\mathrm{query}_{\mathrm{Red}}(k) = n$,

$$\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{S}}(k) < 1/q(k).$$

Proof of Claim. We will prove the contrapositive. Suppose that either

Case 1. A_n does not break soundness of (G, P, V) on security parameter n, or

Case 2. For some $k \leq 2^\ell$ such that $\text{query}_{\text{Red}}(k) = n$,

$$\text{Break}_{\text{Red}^{A_n}}(k) - \text{Break}_{\text{Red}^{S}}(k) \geq 1/q(k).$$

We show how to construct a t'-time oracle algorithm D with

$$\mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))] \geq \epsilon'.$$

To do so, we will show how to construct D with distinguishing advantage at least $3\epsilon'$, both in Case 1 and in Case 2 *where we assume k is known.* This suffices, because then we can test the distinguisher in Case 1 as well as the distinguisher in Case 2 *for all* choices of $k = 1, \ldots, 2^\ell$, and output the best performing one. (More specifically, we run these $1 + 2^\ell$ distinguishers on $O((1/\epsilon'^2)\log(1/\epsilon'))$ independent samples of (CRS, X, Π) and $(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))$ as well as their coin tosses and oracle answers, and output the one with the highest average distinguishing advantage, and it follows from a Chernoff bound that this yields an ϵ'-distinguisher.)

Case 1. A_n does not break soundness on security parameter n. Recall that soundness says $\Pr[V(\mathsf{PRIV}, U, \Pi') = 1 \wedge U \notin L] \leq n^{-c}$ if we choose $(\mathsf{CRS}, \mathsf{PRIV}) \leftarrow G(1^n), (U, \Pi') \leftarrow A_n(1^n, \mathsf{CRS})$ (thus $\Pi' = F_n(\mathsf{CRS}, U)$). By union bound,

$$\Pr\left[V\left(\mathsf{PRIV}, U, \Pi'\right) = 1\right] \leq \left[V\left(\mathsf{PRIV}, U, \Pi'\right) = 1 \wedge U \notin L\right] + \Pr[U \in L] = 1 - n^{-O(1)}.$$

On the other hand, the completeness property says

$$\Pr[V(\mathsf{PRIV}, X, \Pi) = 1] = 1 - \text{negl}(n).$$

Thus V is an $n^{-O(1)}$-distinguisher between (PRIV, X, Π) and (PRIV, U, Π'). Note that conditioned on $\mathsf{CRS} = \mathsf{crs}$ for any crs, PRIV is independent of (X, Π), and that $\mathsf{PRIV}|_{\mathsf{CRS}=\mathsf{crs}}$ can

be sampled in $2^{O(\ell)}$ time given crs (by running $G(1^n; z)$ on all sequences $z \in \{0,1\}^\ell$ of coin tosses). Thus from V we also get a $2^{O(\ell)}$ time $n^{-O(1)}$-distinguisher D for (CRS, X, Π) and (CRS, U, Π'). Specifically, $D(\mathsf{crs}, x, \pi)$ samples $\mathsf{priv} \leftarrow \mathsf{PRIV}|_{\mathsf{CRS}=\mathsf{crs}}$ and outputs $V(\mathsf{priv}, x, \pi)$, so

$$\begin{aligned}
&\mathbb{E}[D(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))] \\
&= \mathbb{E}[D(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D(\mathsf{CRS}, U, \Pi')] \\
&= \Pr\left[V\left(\mathsf{PRIV}, X, \Pi\right) = 1\right] - \Pr\left[V\left(\mathsf{PRIV}, U, \Pi'\right) = 1\right] \\
&= n^{-O(1)} \geq 3\epsilon'.
\end{aligned}$$

Case 2. For some $k \leq 2^\ell$ such that $\mathrm{query}_{\mathrm{Red}}(k) = n$, we have

$$\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{S}}(k) \geq 1/q(k).$$

Assuming k is given, we use the hybrid argument to construct a distinguisher D between $(\mathsf{CRS}, X, \Pi) = (\mathsf{CRS}, S(1^n, \mathsf{CRS}))$ and $(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U)) = (\mathsf{CRS}, A_n(1^n, \mathsf{CRS}))$. Suppose $\mathrm{Red}(1^k)$ runs in time $p(k)$ for some polynomial p. Let H_i be the stateful oracle that behaves like A_n for the first i queries and S for all rest of the queries, so that $H_q = A_n$ and $H_0 = S$. By the hybrid argument, $\mathbb{E}\left[\mathrm{Red}^{H_{I+1}}(1^k)\right] - \mathbb{E}\left[\mathrm{Red}^{H_I}(1^k)\right] \geq 1/p(k)q(k)$ for a randomly chosen $I \in_R \{1, \ldots, p(k)\}$. This immediately gives us a distinguisher D' for $(Z, \mathsf{CRS}', S(1^n, \mathsf{CRS}'))$ and $(Z, \mathsf{CRS}', A_n(1^n, \mathsf{CRS}'))$ where Z is the internal state of the interaction $(\mathrm{Red}^{H_I}(1^k), \mathrm{Chal}(1^k))$ after $I \in_R \{1, \ldots, p(k)\}$ queries, and CRS' is the I-th query (which is determined by Z). Specifically: $D'(z, \mathsf{crs}, x, \pi)$ sets the internal state of $\mathrm{Red}(1^k)$ and $\mathrm{Chal}(1^k)$ to z, runs the interaction $(\mathrm{Red}^S(1^k), \mathrm{Chal}(1^k))$ starting from state z using (x, π) as the answer to the I-th query $(1^n, \mathsf{crs})$, and finally outputs 0 or 1 depending on

whether Chal outputs win. Thus

$$\begin{aligned}
&\mathbb{E}\left[D'(Z, \mathsf{CRS}', S(1^n, \mathsf{CRS}'))\right] - \mathbb{E}\left[D'(Z, \mathsf{CRS}', A_n(1^n, \mathsf{CRS}'))\right] \\
&\geq \mathbb{E}\left[\mathrm{Red}^{H_{I-1}}(1^k)\right] - \mathbb{E}\left[\mathrm{Red}^{H_I}(1^k)\right] \\
&\geq \frac{1}{p(k)q(k)} = 2^{-O(\ell)}.
\end{aligned}$$

To obtain a desired distinguisher D'' for $(\mathsf{CRS}, A_n(1^n, \mathsf{CRS}))$ and $(\mathsf{CRS}, S(1^n, \mathsf{CRS}))$, we simply let D'' sample $(z, \mathsf{crs}') \leftarrow (Z, \mathsf{CRS}')$ and output

$$D''(\mathsf{crs}, x, \pi) = \begin{cases} \frac{D'(z, \mathsf{crs}, x, \pi)}{2^\ell \cdot \Pr[\mathsf{CRS} = \mathsf{crs}]}, & (\mathsf{crs} = \mathsf{crs}') \\ 0, & (\mathsf{crs} \neq \mathsf{crs}') \end{cases}.$$

Note that D'' is $[0,1]$-bounded since CRS is sampled by G using ℓ coin tosses (so $\Pr[\mathsf{CRS} = \mathsf{crs}] \geq 2^{-\ell}$ for all $\mathsf{crs} \in \mathrm{supp}(\mathsf{CRS})$). We are dividing by $\Pr[\mathsf{CRS} = \mathsf{crs}]$ in order to "uniformize" CRS, so that

$$\begin{aligned}
&\mathbb{E}[D''(\mathsf{CRS}, A(1^n, \mathsf{CRS}))] \\
&= \sum_{\mathsf{crs} \in \mathrm{supp}(\mathsf{CRS})} \Pr[\mathsf{CRS} = \mathsf{crs}] \cdot \mathbb{E}\left[\frac{D'(Z, \mathsf{CRS}', A_n(1^n, \mathsf{CRS}'))}{2^\ell \cdot \Pr[\mathsf{CRS} = \mathsf{crs}]} \cdot I\left(\mathsf{CRS}' = \mathsf{crs}\right)\right] \\
&= \mathbb{E}\left[D'(Z, \mathsf{CRS}', A_n(1^n, \mathsf{CRS}')) \cdot I\left(\mathsf{CRS}' \in \mathrm{supp}(\mathsf{CRS})\right)\right] \cdot 2^{-\ell}
\end{aligned}$$

(where $I(\cdot)$ is the indicator function), and similarly for S. Thus D'' has distinguishing advantage

$$\begin{aligned}
&\mathbb{E}[D''(\mathsf{CRS}, S(1^n, \mathsf{CRS}))] - \mathbb{E}[D''(\mathsf{CRS}, A_n(1^n, \mathsf{CRS}))] \\
&= \mathbb{E}\left[D'(Z, \mathsf{CRS}', S(1^n, \mathsf{CRS}')) \cdot I\left(\mathsf{CRS}' \in \mathrm{supp}(\mathsf{CRS})\right)\right] \cdot 2^{-\ell} \\
&\quad - \mathbb{E}\left[D'(Z, \mathsf{CRS}', A_n(1^n, \mathsf{CRS}')) \cdot I\left(\mathsf{CRS}' \in \mathrm{supp}(\mathsf{CRS})\right)\right] \cdot 2^{-\ell} \\
&= \left(\mathbb{E}[D'(Z, \mathsf{CRS}', S(1^n, \mathsf{CRS}'))] - \mathbb{E}[D'(Z, \mathsf{CRS}', A_n(1^n, \mathsf{CRS}'))]\right) \cdot 2^{-\ell} \\
&\geq 2^{-O(\ell)} \cdot 2^{-\ell} = 4\epsilon',
\end{aligned}$$

where the second equality holds because $S(1^n, \mathsf{CRS}')$ and $A_n(1^n, \mathsf{CRS}')$ are identical whenever $\mathsf{CRS}' \notin \mathrm{supp}(\mathsf{CRS})$.

To conclude Case 2, it remains to show that D'' can be implemented in time $2^{O(\ell)} \cdot \mathrm{poly}(1/\epsilon)$. First, $\Pr[\mathsf{CRS} = \mathsf{crs}]$ can be computed in time $2^{O(\ell)}$ by enumerating coin tosses of $G(1^n)$. A query $(1^n, \mathsf{crs})$ to S can be answered in $\mathrm{poly}(n)$ time. A query $(1^n, \mathsf{crs})$ to A_n with $\mathsf{crs} \in \mathsf{CRS}$ can be answered by sampling $(Q, F_n(Q)) = (U_{\ell/2}, U, F_n(U_{\ell/2}, U))$ for up to $O(2^\ell \cdot \log(1/(\epsilon \cdot p(k))))$ times until $U_{\ell/2} = \mathsf{crs}$ (recall that we assume $|\mathsf{crs}| = \ell/2$ in the setup of Lemma 6.6). Thus we can sample $(z, \mathsf{crs}) \leftarrow (Z, \mathsf{CRS}')$ and run $D'(z, \mathsf{crs}, x, \pi)$ in $p(k) \cdot \max\left(\mathrm{poly}(p(k)), 2^{O(\ell)}\right) = 2^{O(\ell)}$ time. It follows from a union bound that

$$\mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))] \geq 3\epsilon'.$$

□

Given Claim 6.8, we now apply Chapter 2, Theorem 2.18 to the oracle algorithm D we constructed in Claim 6.8. Since X and U are ϵ-indistinguishable, Theorem 2.18 yields a $t'' = \mathrm{poly}(2^\ell, t', 1/\epsilon)$-time randomized algorithm $R(1^n)$ that w.p. at least $\Omega(\epsilon^2/\ell)$ outputs a randomized circuit F_n satisfying

$$\mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, X, \Pi)] - \mathbb{E}[D^{O_{Q,F_n(Q)}}(\mathsf{CRS}, U, F_n(\mathsf{CRS}, U))] < 2\epsilon. \tag{6.1}$$

We define the simulator $\mathsf{Sim}(1^k)$ to be the following algorithm:

1. Let $n = \mathrm{query}_{\mathrm{Red}}(k)$;
2. If $\ell(n) \geq \log k$, then output a circuit B_n where $B_n(1^n, \mathsf{crs})$ runs $S(1^n, \mathsf{crs})$;
3. Else, $\ell(n) < \log k$. We run $R(1^n)$ to obtain a randomized circuit F'_n, and output the

randomized circuit B_n where

$$B_n(1^n, \mathsf{crs}) = \begin{cases} (U, F'_n(\mathsf{crs}, U)), & \mathsf{crs} \in \mathrm{supp}(\mathsf{CRS}) \\ S(1^n, \mathsf{crs}), & \mathsf{crs} \notin \mathrm{supp}(\mathsf{CRS}) \end{cases}.$$

Note that Sim is a PPT algorithm since it runs in time $2^{O(\ell(n))} = \mathrm{poly}(k)$ if $\ell(n) < \log k$, and in time $\mathrm{poly}(k)$ if $\ell(n) \geq \log k$. To prove that Sim is indeed an adversary simulator, we define the adversary A_n (which depends on the coins of Sim) to be

$$A_n(1^n, \mathsf{crs}) = \begin{cases} (U, F^*_n(\mathsf{crs}, U)), & \mathsf{crs} \in \mathrm{supp}(\mathsf{CRS}) \\ S(1^n, \mathsf{crs}), & \mathsf{crs} \notin \mathrm{supp}(\mathsf{CRS}) \end{cases}$$

where F^*_n is defined as follows:

- If $\ell(n) \geq \log k$, we let F^*_n be any randomized circuit such that Eq. 6.1 holds for $F_n = F^*_n$;
- If $\ell(n) < \log k$, we let F^*_n be F'_n generated by R in Step 3 of Sim, so that Eq. 6.1 holds for $F_n = F^*_n$ w.p. at least $\Omega(\epsilon(n)^2/\ell(n)) = 2^{-O(\ell(n))} \geq 1/\mathrm{poly}(k)$ over the coins of R (hence coins of Sim).

We now apply Claim 6.8 to $F_n = F^*_n$. Note that Claim 6.8 holds "for all sufficiently large n", but since $\mathrm{query}_{\mathrm{Red}}(k) = \omega(1)$ it must also hold for all sufficiently large k and $n = \mathrm{query}_{\mathrm{Red}}(k)$. Thus Claim 6.8 implies that for all polynomials $q(\cdot)$, for all sufficiently large k and $n = \mathrm{query}_{\mathrm{Red}}(k)$, w.p. at least $1/\mathrm{poly}(k)$ over B_n, A_n satisfies

1. (A_n, c) is a (G, P, V)-adversary on the security parameter n; and
2. If $\ell(n) \geq \log k$, then $\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{S}}(k) < 1/q(k)$.

To conclude the proof it remains to show that

$$\mathrm{Break}_{\mathrm{Red}^{A_n}}(k) - \mathrm{Break}_{\mathrm{Red}^{B_n}}(k) < 1/q(k). \tag{6.2}$$

Indeed, if $\ell(n) \geq \log k$, then B_n runs S, so Eq. 6.2 follows from Item 2 above. If $\ell(n) < \log k$, then Eq. 6.2 holds because $A_n = B_n$ (since $F_n^* = F_n'$) thus $\text{Break}_{\text{Red}^{A_n}}(k) = \text{Break}_{\text{Red}^{B_n}}(k)$.

□

6.2 Proof of Main Theorem

The next two lemmas show that we can "convert" a generic black-box reduction into a length-mapping reduction, which in addition does not make very short queries. To do so, we first convert a generic black-box reduction into one that does not make very short queries (Lemma 6.9), by guessing "optimal" oracle answers for these very short queries. We then convert it to a length-mapping reduction (Lemma 6.10) by a "sparsification" trick, due to Chung, Mahmoody, and Pass [CMP]. As a consequence of "sparsification" the resulting length-mapping reduction no longer works with an arbitrary SNARG adversary. However, it still suffices for proving the main theorem using Lemma 6.6.

Lemma 6.9. *Let $c \in \mathbb{N}$ be a constant. Suppose there is a PPT oracle algorithm* Red *with the property that for every randomized function A where (A, c) is a (G, P, V)-adversary,* Red^A *breaks the falsifiable assumption. Then there is another PPT oracle algorithm* $\widehat{\text{Red}}$ *satisfying the same property, and in addition every query of* $\widehat{\text{Red}}$ *is of length at least* $s = s(n) = (\log \log n)^{\Omega(1)}$.

Proof. Suppose $G(1^m)$ outputs a crs of length m^d and let $s = s(n) = (\log \log n)^{1/(d+1)}$. We define $\widehat{\text{Red}}(1^n)$ as follows:

1. For each $m < s$, select a random function $B_m : \{0,1\}^{m^d} \to \{0,1\}^{m^c}$;
2. Run Red, using $B_m(\mathsf{crs})$ to answer every query $(1^m, \mathsf{crs})$ where $m < s$.

To see that $\widehat{\text{Red}}$ satisfies the same property as Red, consider any (G, P, V)-adversary (A, c). By averaging the coins of A, for each $m < s$ we can fix some (deterministic) function

$A_m : \{0,1\}^{m^d} \to \{0,1\}^{m^c}$ and define

$$\widehat{A}(1^m, \mathsf{crs}) = \begin{cases} A_m(\mathsf{crs}), & m < s \\ A(1^m, \mathsf{crs}) & m \geq s \end{cases},$$

such that $\mathrm{Red}^{\widehat{A}}$ breaks the falsifiable assumption. Note that $\{B_m : m < s\}$ can be encoded as an $s \cdot (s^c)^{s^d} = O(\log n)$ bit string. Thus w.p. at least $1/\mathrm{poly}(n)$, $\widehat{\mathrm{Red}}^A(1^n)$ sets $B_m = A_m$ for all $m < s$ and behaves identically to $\mathrm{Red}^{\widehat{A}}$. Therefore $\widehat{\mathrm{Red}}^A$ also breaks the falsifiable assumption. □

Lemma 6.10 (Chung, Mahmoody, and Pass [CMP])**.** *Let $c \in \mathbb{N}$ be a constant. Suppose there is a PPT oracle algorithm* Red *with the property that for every randomized function A where (A, c) is a (G, P, V)-adversary,* Red^A *breaks the falsifiable assumption, and every query of* Red *is of length at least $s(n) = (\log \log n)^{\Omega(1)}$. Then there is a* length-mapping *PPT oracle algorithm* $\widehat{\mathrm{Red}}$ *where* $\mathrm{query}_{\widehat{\mathrm{Red}}}(n) \geq s(n)$*, such that for infinitely many n and $m = \mathrm{query}_{\widehat{\mathrm{Red}}}(n)$, for every randomized function A_m where (A_m, c) is a (G, P, V)-adversary on security parameter m (of SNARG),* $\widehat{\mathrm{Red}}^{A_m}(1^n)$ *breaks the assumption on security parameter n (of the assumption).*

Proof. We construct $\widehat{\mathrm{Red}}$ from Red as follows. Fix a sparse sequence $h_1, h_2, \ldots$ where $h_1 = 1$ and $h_{m+1} = 2^{2^{2^{h_m}}}$ for $m \geq 1$. Note that the interval $[s(n), \mathrm{poly}(n)]$ contains at most one element of the sequence $h_1, h_2, \ldots$, for some n_c and all $n \geq n_c$. Let $\widehat{\mathrm{Red}}^A(1^n)$ run $\mathrm{Red}^A(1^n)$, where a query $(1^m, \mathsf{crs})$ is answered as follows:

1. If $n < n_c$ or $m \notin \{h_1, h_2, \ldots\}$, then answer the query with a special symbol $\perp$;
2. Otherwise, answer the query using oracle A.

$\widehat{\mathrm{Red}}$ is length-mapping, because every query of $\mathrm{Red}^A(1^n)$ has length in the interval $[s(n), \mathrm{poly}(n)]$ (since Red runs in time $\mathrm{poly}(n)$), and for all $n \geq n_c$, at most one of $h_1, h_2, \ldots$

lies in that interval.

Suppose for contradiction that the $\widehat{\text{Red}}$ we construct does not satisfy the desired properties. That is, for all sufficiently large n and $m = \text{query}_{\widehat{\text{Red}}}(n)$, there exists some randomized function A_m where (A_m, c) is a (G, P, V)-adversary on security parameter m, but $\widehat{\text{Red}}^{A_m}(1^n)$ does not break the assumption on security parameter n.

Let A be any randomized function such that $A(1^m, \mathsf{crs}) = A_m(1^m, \mathsf{crs})$ where $m = \text{query}_{\widehat{\text{Red}}}(n)$ and for all sufficiently large n. Thus (A, c) is an a.e. (G, P, V)-adversary. Let $\widehat{A}$ be a "sparsification" of A: $\widehat{A}(1^m, \mathsf{crs}) := A(1^m, \mathsf{crs})$ whenever $m \in \{h_1, h_2, \dots\}$ and $\widehat{A}(1^m, \mathsf{crs}) := \perp$ for all other m. Thus $(\widehat{A}, c)$ is a (G, P, V)-adversary.

Since $(\widehat{A}, c)$ is a (G, P, V)-adversary $\text{Red}^{\widehat{A}}$ breaks the falsifiable assumption. On the other hand, $\widehat{\text{Red}}^{A}(1^n)$ behaves like $\widehat{\text{Red}}^{A_m}$ for all sufficiently large n, hence does not break the assumption. This yields a contradiction, because by construction $\widehat{\text{Red}}^{A}(1^n) = \text{Red}^{\widehat{A}}(1^n)$ for all $n \geq n_c$. □

Finally, we use Lemma 6.6 to deduce Theorem 6.5. Note that the length-mapping reduction we obtain from Lemma 6.10 is slightly weaker, as it requires that the adversary break soundness on a *fixed* infinite sequence of security parameters (rather than *any* infinite sequence of security parameters). However, it suffices because Lemma 6.6 provides adversaries that break soundness on *almost all* security parameters.

Proof of Theorem 6.5 (Main Theorem). Suppose there is a generic uniform black-box reduction showing the soundness of (G, P, V) based on a uniform falsifiable assumption. We will show that the falsifiable assumption is already false, by constructing a PPT algorithm that breaks it.

Fix c to be the constant given by Lemma 6.6. By Lemma 6.9 and Lemma 6.10, there is a length-mapping PPT oracle algorithm Red where $\text{query}_{\text{Red}}(k) = \omega(1)$, and for infinitely

many k, for $n = \text{query}_{\text{Red}}(k)$, and for every randomized function A_n where (A_n, c) is a (G, P, V)-adversary on security parameter n, $\text{Red}^{A_n}(1^k)$ breaks the assumption on security parameter k.

We now apply Lemma 6.6 to Red to obtain a PPT algorithm Sim such that for all polynomials $q(\cdot)$, all sufficiently large k and $n = \text{query}_{\text{Red}}(k)$, w.p. at least $1/\text{poly}(k)$, $\mathsf{Sim}(1^k)$ outputs a randomized circuit B_n such that there is a randomized function A_n satisfying:

1. (A_n, c) is a (G, P, V)-adversary on the security parameter n;
2. $\text{Break}_{\text{Red}^{A_n}}(k) - \text{Break}_{\text{Red}^{B_n}}(k) < 1/q(k)$.

By the previous discussion, for infinitely many k, Item 1 implies that $\text{Red}^{A_n}(1^k)$ breaks the assumption on security parameter k. Thus by Item 2, for infinitely many k, $\text{Red}^{B_n}(1^k)$ also breaks the assumption on security parameter k. Hence we obtain a PPT algorithm breaking the assumption for infinitely many k: first generate the circuit B_n by running $\mathsf{Sim}(1^k)$, then run $\text{Red}^{B_n}(1^k)$. □

Chapter 7

Pseudoentropy and Algorithmic Prediction Markets

In economics, prediction markets are speculative markets designed to elicit people's beliefs. Prediction markets based on market scoring rules, introduced by Hanson [Han], have several important advantages. In particular, the automated market maker is able to elicit true beliefs from experts, by paying a worst-case cost proportional to the distribution's entropy.

In this chapter, we describe an application of the characterization of pseudoentropy (Chapter 4) in the context of such prediction markets generalized to a more algorithmic setting.

7.1 Introduction

Consider an unknown distribution B on some outcome space Ω; for example, the distribution of the winner of a horse race. An effective way to elicit experts' beliefs about B is asking them to wager. A *prediction market* is a mechanism where for every possible

outcome $\omega \in \Omega$, traders (the experts) can buy and sell shares of security (i.e. bet) on ω. At last, when the true outcome ω^* is drawn from B and revealed, each trader is rewarded $\$i$ for holding i shares of security on ω^*.

The de facto standard prediction markets with automated market makers are ones based on *market scoring rules* (MSR), introduced by Hanson [Han]. In Hanson's MSR-based prediction markets, there is an *automated* market maker who will accept *any* transaction, and all transactions happen between market maker and the trader(s). In the transaction, the price of buying or selling one share of security on $\omega \in \Omega$ is determined by the market scoring rule, and depends on the current number of outstanding shares of each security. These MSR-based prediction markets have several desirable properties, including infinite liquidity (traders can always trade), truth revelation (any risk-neutral trader's transaction reveals his belief about B), and bounded worst-case loss (market maker does not pay too much to elicit such truthful information even when B is adversarially chosen), assuming a good scoring rule. In practice, MSR-based prediction markets are used by a number of companies including Microsoft.

In reality, B is often jointly distributed with a set of random variables in a Bayes net, which we collectively encode as an n-bit binary string X. The standard prediction markets can only be used to elicit beliefs about (i) the marginal distribution of B (i.e. without involving X), or (ii) the conditional distribution $B|_{X=x}$ for the specific x that has been observed, by revealing x to traders in advance.

In light of this, we consider an extension called *algorithmic prediction markets* that are capable of eliciting experts' beliefs about the distribution $B|_{X=x}$ *for every* $x \in \{0,1\}^n$ (but without explicitly running a separate prediction market on $B|_{X=x}$ for each x). More concretely, we require a trader to present his transaction as a polynomial-sized Boolean circuit that on input x outputs a transaction as if he is trading in the prediction market for

$B|_{X=x}$. We show that such algorithmic prediction markets can be executely efficiently, and are "best effort" truth revealing when proper market scoring rules are used.

It is necessary to assume a polynomial-sized outcome space Ω in order to efficiently execute a prediction market. This is because the market maker needs $\Omega(|\Omega|)$ time and space simply to maintain how much has been wagered on each security.[1] With a polynomial-sized Ω, a trader in standard prediction market can always trade efficiently according to his beliefs, which can be an arbitrary distribution on Ω. However, this is no longer true for algorithmic prediction markets, because the mapping from x to the trader's belief about $B|_{X=x}$ can be computationally hard.

Assuming logarithmic scoring rules, Hanson shows that the market maker's worst-case loss is characterized by $\mathrm{H}_{\mathrm{sh}}(B)$. For algorithmic prediction markets, we show that the worst-case loss is characterized by the *pseudoentropy* of B given X. This is proved using our characterization of conditional pseudoentropy from Chapter 4.

7.2 Algorithmic Prediction Markets

Notations. We use boldface lower-case letters to denote vectors in $\mathbb{R}^q$. If $\mathbf{p}$ is a probability vector, we (abusing notation) let $\mathbf{p}$ also denote the corresponding probability distribution on $\{1, \ldots, q\}$.

7.2.1 Strictly Proper Scoring Rule

Consider outcome space $\Omega = \{1, \ldots, q\}$. Let $H : \{\text{distributions on } \Omega\} \to \mathbb{R}$ be a strictly concave, differentiable function. Let $\mathbf{p}$ be a probability vector of a distribution on Ω. The

[1] One workaround is to restrict the possible securities available to the traders to a small set of predicates; such prediction markets are known as *combinatorial prediction markets*, and their complexity has been studied by Chen et al. [CGP, CFL$^+$].

strictly proper scoring rule associated with H is the following function S_H:

$$S_H(\mathbf{p}, \omega) = -H(\mathbf{p}) + \langle \nabla H(\mathbf{p}), \mathbf{p} \rangle - \frac{\partial}{\partial \mathbf{p}_\omega} H(\mathbf{p})$$

where $\nabla H(\mathbf{p})$ denotes the gradient vector, and $\mathbf{p}_\omega$ denotes the element of $\mathbf{p}$ indexed by $\omega \in \Omega$. If H is set to $b\mathrm{H}_{\mathsf{sh}}$ where H_{sh} is the Shannon entropy function and $b > 0$, then S_H is called a *logarithmic scoring rule*.

The key property of a strictly proper scoring rule is that, as a utility function, S_H can be used to elicit true beliefs. Consider the setting where an expert is asked to report his belief about the distribution B, and is then rewarded $S_H(\mathbf{p}, \omega^*)$ dollars, where $\mathbf{p}$ is the distribution reported by the expert and ω^* is the outcome drawn from B. Thus, an expert who reports $\mathbf{p}$ while believing the actual distribution to be $\mathbf{r}$ is expecting a utility of $\mathbb{E}_{\omega \sim \mathbf{r}} [S_H(\mathbf{p}, \omega)]$. A strictly proper scoring rule S_H guarantees that this quantity is maximized by setting $\mathbf{p} = \mathbf{r}$ (thus an expert will always report his true belief $\mathbf{r}$):

Proposition 7.1. $\mathbb{E}_{\omega \sim \mathbf{r}} [S_H(\mathbf{r}, \omega)] - \mathbb{E}_{\omega \sim \mathbf{r}} [S_H(\mathbf{p}, \omega)] = \mathrm{D}_H(\mathbf{r} || \mathbf{p}) \geq 0$, where equality holds iff $\mathbf{p} = \mathbf{r}$.

Proof. By definition,

$$\mathop{\mathbb{E}}_{\omega \sim \mathbf{r}} [S_H(\mathbf{p}, \omega)] = -H(\mathbf{p}) + \langle \nabla H(\mathbf{p}), \mathbf{p} \rangle - \sum_{\omega \in \Omega} \mathbf{r}_\omega \cdot \frac{\partial}{\partial \mathbf{p}_\omega} H(\mathbf{p}) = -H(\mathbf{p}) + \langle \nabla H(\mathbf{p}), \mathbf{p} - \mathbf{r} \rangle .$$

Thus

$$\mathop{\mathbb{E}}_{\omega \sim \mathbf{r}} [S_H(\mathbf{r}, \omega)] - \mathop{\mathbb{E}}_{\omega \sim \mathbf{r}} [S_H(\mathbf{p}, \omega)] = H(\mathbf{p}) - H(\mathbf{r}) - \langle \nabla H(\mathbf{p}), \mathbf{p} - \mathbf{r} \rangle = \mathrm{D}_H(\mathbf{r} || \mathbf{p}) \geq 0$$

where we use nonnegativity of Bregman divergence (Proposition 1.8). □

7.2.2 Prediction Markets Based on Market Scoring Rules

Given an unknown target distribution B on $\Omega = \{1, \ldots, q\}$, the prediction market based on market scoring rule S_H works as follows. A security is offered for each $\omega \in \Omega$, and

every share of security ω pays off \$1 if ω happens, \$0 otherwise. The market starts with a vector $\mathbf{m}$ of initial *total outstanding shares* (that is, $\mathbf{m}_\omega$ is the total share of security ω held by all traders initially). There is a centralized market maker who, at any time, accepts to buy or sell any security with any trader, at a price determined below. The trader can query the market maker about the cost of a hypothetical transaction, as well as the *instantaneous price* of any $\omega \in \Omega$ (see below). Finally, an outcome ω^* is drawn from X and the market maker pays \$1 to each trader holding one share of security ω^*.

Consider a transaction that changes the vector of total outstanding shares from $\mathbf{p}$ to $\mathbf{q}$. The cost of the transaction to a trader is defined to be

$$C(\mathbf{q}) - C(\mathbf{p})$$

for some nonnegative cost function C, whose gradient $\nabla C(\mathbf{q})$ is called the vector of *instantaneous prices* at $\mathbf{q}$. The function C is chosen so that (i) the gradient $\nabla C(\mathbf{q})$ is a probability vector; (ii) the trader's net profit $C(\mathbf{p}) - C(\mathbf{q}) + \mathbf{p}_{\omega^*} - \mathbf{q}_{\omega^*}$ equals

$$S_H(\nabla C(\mathbf{q}), \omega^*) - S_H(\nabla C(\mathbf{p}), \omega^*).$$

Such cost function C can be defined for all proper scoring rules H [Han]. For example, with the logarithmic scoring rule $H = b\mathrm{H}_{\mathsf{sh}}$, the cost function C is defined to be

$$C(\mathbf{p}) = b \cdot \ln \sum_{\omega \in \Omega} \exp(\mathbf{p}_\omega / b),$$

and given current outstanding shares $\mathbf{p}$, the instantaneous price of ω equals

$$\left(\nabla C(\mathbf{p})\right)_\omega = \frac{\exp(\mathbf{p}_\omega / b)}{\sum_{\omega' \in \Omega} \exp(\mathbf{p}_{\omega'} / b)}.$$

Loss and Utility. From a transaction that changes total outstanding shares from $\mathbf{p}$ to $\mathbf{q}$, the trader who believes the target distribution is $\mathbf{r}$ expects a net utility of

$$\mathop{\mathbb{E}}_{\omega \sim \mathbf{r}} \left[C(\mathbf{p}) - C(\mathbf{q}) + \mathbf{p}_\omega - \mathbf{q}_\omega\right] = \mathop{\mathbb{E}}_{\omega \sim \mathbf{r}} \left[S_H(\nabla C(\mathbf{q}), \omega)\right] - \mathop{\mathbb{E}}_{\omega \sim \mathbf{r}} \left[S_H(\nabla C(\mathbf{p}), \omega)\right].$$

By Proposition 7.1, a risk-neutral trader always wants to make a transaction such that $\nabla C(\mathbf{q}) = \mathbf{r}$. Therefore, such prediction markets can elicit true beliefs from risk-neutral traders. Note that traders do not necessarily know the actual distribution B (that is, hold belief $\mathbf{r} = B$), and their beliefs may converge over time by observing the beliefs of other traders. However, traders's final transactions will still reflect their true beliefs. It is up to the market maker how to combine their beliefs to extract knowledge.

Let $\mathbf{s}$ denote the vector of outstanding shares after all transactions. Assuming a logarithmic scoring rule $H = b\mathrm{H}_{\mathrm{sh}}$, Proposition 7.1 says that the market maker's loss equals

$$\mathop{\mathbb{E}}_{\omega\sim B}[S_H(\nabla C(\mathbf{s}),\omega)] - \mathop{\mathbb{E}}_{\omega\sim B}[S_H(\nabla C(\mathbf{m}),\omega)] = b\cdot(\mathrm{KL}(B||\nabla C(\mathbf{m})) - \mathrm{KL}(B||\nabla C(\mathbf{s}))).$$

By setting $\nabla C(\mathbf{m})$ to be the uniform distribution (e.g. with zero initial outstanding shares), the market maker's loss at any time is at most $b\cdot\mathrm{KL}(B||U_\Omega) = b\cdot(\log|\Omega| - \mathrm{H}_{\mathrm{sh}}(B))$.

7.2.3 Algorithmic Prediction Markets

Let (X, B) be a joint distribution on $\{0,1\}^n \times \Omega, \quad = \{1,\ldots,q\}$ for $q = \mathrm{poly}(n)$. We are interested in learning from experts an "efficient rule" describing how B is distributed given X. For example, the formula that determines the distribution of an athlete's performance from a number of n binary factors.

A *algorithmic prediction market* works by running 2^n MSR-based prediction markets simultaneously; for each $x \in \{0,1\}^n$, we run a market for the unknown target distribution $B|_{X=x}$, which we shall call the *xth market*. These 2^n markets are run in parallel, and use the same set of traders (experts) and proper scoring rule S_H. At the end, an outcome (x^*, ω^*) is drawn from (X, B), and the market maker pays \$1 to each trader holding one share of security ω^* in the x^*th market.

As a trader simultaneously trades in all markets, it is infeasible to explicitly submit all 2^n transactions to the market maker. Instead, the market maker specifies two polynomials

$p(n)$ and $t(n)$ which are announced to the traders in advance. A trader's transaction is a (deterministic) $p(n)$-sized Boolean circuit A such that $A(x)$ outputs his transaction in the xth market, represented as a vector $\mathbf{t}$ indexed by Ω. The initial outstanding shares in the xth market is zero (for simplicity), and no more than $t(n)$ total transactions are allowed. Similar to the standard setting, the trader can query the market maker about the cost of a hypothetical transaction in the xth market, as well as instantaneous price of any $\omega \in \Omega$ in the xth market.

Note that we do not actually *commit* all 2^n transactions $\{A(x) : x \in \{0,1\}^n\}$ (e.g. physical exchange of cash and shares); the market maker merely keeps tracks of all transaction history as a list of circuits $A_1, A_2, \ldots, A_s$, $s \leq t(n)$. It is only after x^* is finally drawn from X that the market maker physically commits the polynomially many transactions $A_1(x^*)$, $A_2(x^*)$, ..., $A_s(x^*)$.

Efficiency. Such algorithmic prediction markets can be efficiently implemented as long as for each x, the xth market can be efficiently implemented, e.g. when logarithmic scoring rules are used. Specifically, for every $x \in \{0,1\}^n$, the instantaneous price $\nabla C(\cdot)$ and the cost function $C(\cdot)$ in the xth market can be computed from the vector of total outstanding shares in the xth market, which in turn can be computed by combining the (polynomially long) transaction history $A_1(x)$, $A_2(x)$, ... so far.

Definition 7.2 (Worst-Case Loss for Algorithmic Prediction Markets)**.** In an algorithmic prediction market, the *market maker has a worst-case loss of at most k* if all polynomials $p(n)$ and for all sequences of $p(n)$-sized circuits $A_1, \ldots, A_{p(n)}$, the transactions $A_1, \ldots, A_{p(n)}$ incur a total loss of at most $k - 1/p(n)$ to the market maker.

7.3 Characterizing Worst-Case Loss and Expected Utility

Consider an algorithmic prediction market with logarithmic scoring rule (i.e. $H = -b\mathrm{H}_{\mathrm{sh}}$). We give a characterization of the market maker's worst-case loss in terms of the pseudoentropy of B given X. Recall that in a standard prediction markets using the logarithmic scoring rule, the market maker's worst-case loss is characterized in terms of the entropy of B. Thus our result can be viewed as a computational generalization.

Theorem 7.3. *Let n be a security parameter, $\Omega = \Omega(n)$ with $|\Omega| = \mathrm{poly}(n)$, and $(X, B) = (X, B)(n)$ be a polynomial-time samplable joint distribution on $\{0,1\}^n \times \Omega$. Consider the algorithmic prediction market on (X, B) with logarithmic scoring rule S_H where $H = -b\mathrm{H}_{\mathrm{sh}}$. Then the market maker's has a worst-case loss of at most $b \cdot (\log|\Omega| - k)$ if and only if B has pseudoentropy at least k given X.*

Proof. Let $\mathbf{0}$ (the all zero vector) and $\mathbf{q}^x$ be the vectors of initial and final total outstanding shares in the xth market after transactions $A_1, \ldots, A_s$. It follows from Proposition 7.1 that the expected loss at this point equals

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(x,\omega)\sim(X,B)}[S_H(\nabla C(\mathbf{q}^x), \omega)] - \mathop{\mathbb{E}}_{(x,\omega)\sim(X,B)}[S_H(\nabla C(\mathbf{0}), \omega)] \\
&= b \cdot \mathop{\mathbb{E}}_{x\sim X}[\mathrm{KL}(B|_{X=x} \parallel \nabla C(\mathbf{0})) - \mathrm{KL}(B|_{X=x}||\nabla C(\mathbf{q}^x))] \\
&= b \cdot \left(\mathrm{KL}(X, B||X, U_\Omega) - \mathrm{KL}(X, B||X, \nabla C(\mathbf{q}^X))\right) \\
&= b \cdot \left(\log|\Omega| - \mathrm{H}_{\mathrm{sh}}(B|X) - \mathrm{KL}(X, B||X, \nabla C(\mathbf{q}^X))\right).
\end{aligned}$$

By Chapter 4, Theorem 4.38, it suffices to show the following are equivalent:

1. There exists a polynomial $p(n)$ and a sequence of transactions A_1, ..., $A_{p(n)}$ as $p(n)$-sized circuits such that $\mathrm{KL}(X, B||X, \nabla C(\mathbf{q}^X)) \leq k - \mathrm{H}_{\mathrm{sh}}(B|X) - 1/p(n)$;
2. There exists a polynomial-sized circuit P such that $\mathrm{KL}(X, B||X, \Phi_P) \leq k - \mathrm{H}_{\mathsf{sh}}(B|X) - 1/n^{O(1)}$.

First suppose that Item 1 holds. Recall that the final outstanding shares $\mathbf{q}^x$ equals $A_1(x) + \cdots + A_{p(n)}(x)$. Thus Item 2 holds for the polynomial-sized circuit P that computes $\mathbf{q}^x$ by running $A_1(x), \ldots, A_{p(n)}(x)$, and outputs the pmf of $\nabla C(\mathbf{q}^x)$. Since

$$(\nabla C(\mathbf{p}))_\omega = \frac{\exp(\mathbf{p}_\omega/b)}{\sum_{\omega' \in \Omega} \exp(\mathbf{p}_{\omega'}/b)},$$

$\nabla C(\mathbf{q}^x)$ can be efficiently approximated using standard numerical techniques (e.g. Newton's method) such that

$$\mathrm{KL}(X, B||X, \Phi_P) \leq k - \mathrm{H}_{\mathsf{sh}}(B|X) - 1/n^{O(1)}$$

(see Lemma A.6 for details; the same approximation is done in the proof of Chapter 4, Theorem 4.50).

Now suppose Item 2 holds. To show Item 1, we simply let there be a single transaction A_1, which is a polynomial-sized circuit that outputs the vector $(\nabla C)^{-1}(\Phi_P|_{X=x})$ where $(\nabla C)^{-1}$ denotes the inverse of the gradient of C. It can be checked that $(\nabla C)^{-1}$ can be approximated efficiently, so that $\mathrm{KL}(X, B||X, \nabla C(\mathbf{q}^X)) \leq \mathrm{KL}(X, B||X, \Phi_P) + n^{\omega(1)} \leq k - \mathrm{H}_{\mathrm{sh}}(B|X) - 1/p(n)$. □

"Best effort" truth-revealing. In an algorithmic prediction market, if the trader's belief (X, B') is such that $(x, \omega) \to \Pr[X = x|B' = \omega]$ can be computed by a polynomial-sized circuit, then he can and will be truth-revealing by making a transaction such that $(X, \nabla C(\mathbf{p}^X)) = (X, B')$. On the other hand, if the trader's belief B' has pseudoentropy noticeably more than $\mathrm{H}_{\mathrm{sh}}(B'|X)$ given X, then he is incentivized to minimize $\mathrm{KL}(X, B'||X, \nabla C(\mathbf{p}^X))$ (subject to being computationally bounded). We call the later behavior "best-effort" truth-revealing.

Chapter 8

Conclusion

In this work we developed several tools, such as the Uniform Min-Max Theorem and the Regularity Theorems, and provided a wide range of applications in cryptography and complexity theory. In addition, we developed techniques relating Bregman divergence to indistinguishability in pseudorandomness theory; using this (and other) techniques we show how to characterize different notions of computational entropies, which in turn have many applications in cryptography and complexity theory such as simplifying and improving the construction of pseudorandom generators. We hope that our tools and techniques would offer readers more insights into the applications we demonstrated, help further our understanding of fundamental problems in the areas of cryptography and complexity, and will be extended to more applications in broader contexts.

Bibliography

[BHK] Boaz Barak, Moritz Hardt, and Satyen Kale. The uniform hardcore lemma via approximate bregman projections. In *SODA '09: Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms*, pages 1193–1200, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[BM] Manuel Blum and Silvio Micali. How to generate cryptographically strong sequences of pseudo random bits. pages 112–117, 1982.

[BSW] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *RANDOM-APPROX*, pages 200–215, 2003.

[BV] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[CF] David Conlon and Jacob Fox. Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.*, 22(5):1191–1256, 2012.

[CFL$^+$] Yiling Chen, Lance Fortnow, Nicolas S. Lambert, David M. Pennock, and Jennifer Wortman. Complexity of combinatorial market makers. In *ACM Conference on Electronic Commerce*, pages 190–199, 2008.

[CGP] Yiling Chen, Sharad Goel, and David M. Pennock. Pricing combinatorial markets for tournaments. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 305–314, New York, NY, USA, 2008. ACM.

[CLP1] Kai-Min Chung, Edward Lui, and Rafael Pass. Can theories be tested?: a cryptographic treatment of forecast testing. In *ITCS*, pages 47–56, 2013.

[CLP2] Kai-Min Chung, Edward Lui, and Rafael Pass. From weak to strong zero-knowledge and applications. Cryptology ePrint Archive, Report 2013/260, 2013. `http://eprint.iacr.org/`.

[CMP] Kai-Min Chung, Mohammad Mahmoody, and Rafael Pass. Personal communication, 2012/12.

[CT] Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.

[DORS] 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(1):97–139, 2008.

[DP1] Yevgeniy Dodis and Krzysztof Pietrzak. Leakage-resilient pseudorandom functions and side-channel attacks on Feistel networks. In *Advances in cryptology—CRYPTO 2010*, volume 6223 of *Lecture Notes in Comput. Sci.*, pages 21–40. Springer, Berlin, 2010.

[DP2] Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *FOCS*, pages 293–302. IEEE Computer Society, 2008.

[FFS] Uriel Feige, Amos Fiat, and Adi Shamir. Zero-knowledge proofs of identity. *Journal of Cryptology*, 1(2):77–94, 1988.

[FK] Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[FOR] Benjamin Fuller, Adam O'Neill, and Leonid Reyzin. A unified approach to deterministic encryption: New constructions and a connection to computational entropy. In Ronald Cramer, editor, *TCC*, volume 7194 of *Lecture Notes in Computer Science*, pages 582–599. Springer, 2012.

[FR] Benjamin Fuller and Leonid Reyzin. Computational entropy and information leakage. 2011. (available at http://www.cs.bu.edu/fac/reyzin).

[FS] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

[FV] Lance Fortnow and Rakesh Vohra. The complexity of forecast testing: abstract. In *ACM Conference on Electronic Commerce*, page 139, 2008.

[GGKT] Rosario Gennaro, Yael Gertner, Jonathan Katz, and Luca Trevisan. Bounds on the efficiency of generic cryptographic constructions. *SIAM Journal on Computing*, 35(1):217–246, 2005.

[GGM] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, October 1986.

[GL] Oded Goldreich and Leonid A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing*, pages 25–32, Seattle, Washington, 15–17 May 1989.

[GM1] Oded Goldreich and Bernd Meyer. Computational indistinguishability: algorithms vs. circuits. *Theoretical Computer Science*, 191(1-2):215–218, 1998.

[GM2] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984.

[GMW] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *Journal of the ACM*, 38(1):691–729, 1991.

[Gol] Oded Goldreich. *Computational Complexity: A Conceptual Perspective*. 2006.

[Gow] W. T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.*, 42(4):573–606, 2010.

[GT] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[GV] Oded Goldreich and Salil Vadhan. Comparing entropies in statistical zero knowledge with applications to the structure of szk. In *In Proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity*, pages 54–73. IEEE Computer Society Press, 1998.

[GW] Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *STOC*, pages 99–108. ACM, 2011.

[Han] Robin Hanson. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets*, 1:2007, 2002.

[HH] Iftach Haitner and Thomas Holenstein. On the (im)possibility of key dependent encryption. In *TCC*, pages 202–219, 2009.

[HHR1] Iftach Haitner, Danny Harnik, and Omer Reingold. Efficient pseudorandom generators from exponentially hard one-way functions. In *Automata, Languages and Programming, 24th International Colloquium, ICALP*, 2006.

[HHR2] Iftach Haitner, Danny Harnik, and Omer Reingold. On the power of the randomized iterate. In *Advances in Cryptology – CRYPTO 2006*, 2006.

[HHR+3] Iftach Haitner, Thomas Holenstein, Omer Reingold, Salil P. Vadhan, and Hoeteck Wee. Universal one-way hash functions via inaccessible entropy. In *EUROCRYPT*, pages 616–637, 2010.

[HILL] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[HLR] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In *EUROCRYPT*, pages 169–186, 2007.

[Hol1] Thomas Holenstein. Key agreement from weak bit agreement. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 664–673, 2005.

[Hol2] Thomas Holenstein. Pseudorandom generators from one-way functions: A simple construction for any hardness. In *TCC*, pages 443–461, 2006.

[HRV] Iftach Haitner, Omer Reingold, and Salil Vadhan. Efficiency improvements in constructing pseudorandom generators from one-way functions. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 437–446, 2010.

[HRVW] Iftach Haitner, Omer Reingold, Salil Vadhan, and Hoeteck Wee. Inaccessible entropy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC '09)*, pages 611–620, 31 May–2 June 2009.

[HS] Thomas Holenstein and Makrand Sinha. Constructing a pseudorandom generator requires an almost linear number of calls. In *FOCS*, pages 698–707, 2012.

[HW] M. Herbster and M. Warmuth. Tracking the best linear predictor. *Journal of Machine Learning Research*, 1:281–309, 2001.

[IKOS] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Cryptography with constant computational overhead. In *STOC*, pages 433–442, 2008.

[IL] Russell Impagliazzo and Michael Luby. One-way functions are essential for complexity based cryptography. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 230–235, 1989.

[Imp] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 538–545, 1995.

[IR] Russell Impagliazzo and Steven Rudich. Limits on the provable consequences of one-way permutations. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC)*, pages 44–61. ACM Press, 1989.

[JP] Dimitar Jetchev and Krzysztof Pietrzak. How to fake auxiliary input. Cryptology ePrint Archive, Report 2013/869, 2013. `http://eprint.iacr.org/`.

[KMR$^+$] Michael J. Kearns, Yishay Mansour, Dana Ron, Ronitt Rubinfeld, Robert E. Schapire, and Linda Sellie. On the learnability of discrete distributions. In *STOC*, pages 273–282, 1994.

[KS] Adam R. Klivans and Rocco A. Servedio. Boosting and hard-core set construction. *Machine Learning*, 51(3):217–238, 2003.

[LL] L.D. Landau and E.M. Lifshitz. *Statistical physics*, volume 5 of *Statistical Physics*. Oxford: Pergamon Press, 1980.

[LR] Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM J. Comput.*, 17(2):373–386, 1988.

[LTW] Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu. Complexity of hard-core set proofs. *Computational Complexity*, 20(1):145–171, 2011.

[MPRV] Ilya Mironov, Omkant Pandey, Omer Reingold, and Salil Vadhan. Computational differential privacy. In *Advances in cryptology—CRYPTO 2009*, volume 5677 of *Lecture Notes in Comput. Sci.*, pages 126–142. Springer, Berlin, 2009.

[Nao1] Moni Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991.

[Nao2] Moni Naor. Evaluation may be easier than generation. In *STOC*, pages 74–83, 1996.

[Rey] Leonid Reyzin. Some notions of entropy for cryptography. In *ICITS*, pages 138–142, 2011.

[RR] Alexander A. Razborov and Steven Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, August 1997.

[RTTV] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS '08)*, pages 76–85. IEEE, 26–28 October 2008.

[Sha] Claude Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[STV] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. *Journal of Computer and System Sciences*, 62:236–266, 2001.

[TTV] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity (CCC '09)*, pages 126–136, 15–18 July 2009. Preliminary version posted as *ECCC* TR08-103.

[TZ] Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Math.*, 201(2):213–305, 2008.

[Val] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[VZ1] Salil Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC '12)*, pages 817–836, 19–22 May 2012.

[VZ2] Salil Vadhan and Colin Jia Zheng. A uniform min-max theorem with applications in cryptography. In *Advances in cryptology—CRYPTO 2013*, Lecture Notes in Comput. Sci. Springer, Berlin, 2013.

[VZ3] Salil Vadhan and Colin Jia Zheng. A uniform min-max theorem with applications in cryptography. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:101, 2013.

[VZ4] Salil P. Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:141, 2011.

[Yao1] A. C. Yao. Protocols for secure computations. pages 160–164, 1982.

[Yao2] Andrew C. Yao. Theory and applications of trapdoor functions. pages 80–91, 1982.

[Zha] Jiapeng Zhang. On the query complexity for showing dense model. *Electronic Colloquium on Computational Complexity (ECCC)*, 2011.

Appendix A

Missing Lemmas and Proofs

Lemma A.1 (Multiplicative weight update decreases KL)**.** *Let A, B be distributions on $[N]$, $f : [N] \to [0, 1]$ be a function, and $0 \le \epsilon \le 1$. Then the distribution A' defined to be*

$$\Pr[A' = x] \propto e^{\epsilon \cdot f(x)} \Pr[A = x]$$

satisfies $\mathrm{KL}(B \parallel A') \le \mathrm{KL}(B \parallel A) - (\log e)\epsilon \left(\mathbb{E}[f(B)] - \mathbb{E}[f(A)] - \epsilon\right)$.

Proof. By definition,

$$\begin{aligned}
\mathrm{KL}(B \parallel A) - \mathrm{KL}(B \parallel A') &= \sum_x \Pr[B = x] \left(\log \frac{\Pr[B = x]}{\Pr[A = x]} - \log \frac{\Pr[B = x]}{\Pr[A' = x]} \right) \\
&= \sum_x \Pr[B = x] \log \frac{\Pr[A' = x]}{\Pr[A = x]} \\
&= \sum_x \Pr[B = x] \left(\log \frac{e^{\epsilon f(x)}}{\sum_y e^{\epsilon f(y)} \Pr[A = y]} \right) \\
&= (\log e) \left(\epsilon \mathbb{E}[f(B)] - \ln \left(\sum_y e^{\epsilon f(y)} \Pr[A = y] \right) \right)
\end{aligned}$$

Applying the inequalities $1 + z \le e^z$, $e^z \le 1 + z + z^2$ for $0 \le z \le 1$, and using $0 \le f(x) \le 1$,

we have

$$\begin{aligned}\mathrm{KL}(B \parallel A) - \mathrm{KL}(B \parallel A') &\geq (\log e)\left(\epsilon\mathbb{E}[f(B)] - \ln\left(\sum_y \left(1+\epsilon f(y)+\epsilon^2\right)\Pr[A=y]\right)\right)\\ &= (\log e)\left(\epsilon\mathbb{E}[f(B)] - \ln\left(1+\epsilon\mathbb{E}[f(A)]+\epsilon^2\right)\right)\\ &\geq (\log e)\left(\epsilon\mathbb{E}[f(B)] - \left(\epsilon\mathbb{E}[f(A)]+\epsilon^2\right)\right)\\ &= (\log e)\epsilon\left(\mathbb{E}[f(B)] - \mathbb{E}[f(A)] - \epsilon\right)\end{aligned}$$

□

For the following lemma, recall from Chapter 3 that a measure $M : \mathcal{X} \to [0,1]$ is δ-*dense* if its *density* $\mu(M) = \sum_{x\in\mathcal{X}} M(x)/|\mathcal{X}|$ is at least δ. We denote by $\mathcal{M}_{m,\delta}$ the set of all δ-dense measures defined on $\{0,1\}^m$.

Lemma A.2 (Sampling from a high density measure)**.** *Let n be a security parameter, $\delta = \delta(n)$, $\sigma = \sigma(n)$. Then for $k = O((1/\delta)\log(1/\sigma))$, there is a randomized algorithm that, given k and oracle access to a measure $M \in \mathcal{M}_{n,\delta}$, w.p. at least $1-\sigma$ outputs a random sample of Φ_M. The algorithm runs in $O(k(s+n))$ time and makes k oracle queries, where s is a bound on the bit length of $M(x)$.*

Proof. Use rejection sampling. Select a random $z \in_R \{0,1\}^n$ and output z w.p. $M(z)$. Repeat up to $k = O((1/\delta)\log(1/\sigma))$ times until some z is outputted. Thus with all but $(1-\mathbb{E}_z[M(z)])^k = (1-\delta)^k \leq \sigma$ probability we output some $z \leftarrow \Phi_M$. □

Lemma A.3 (Approximating KL projection on high min-entropy distributions)**.** *Let $\mathcal{C}$ be the set of distributions over $\{0,1\}^n$ with min-entropy at least $n - \log(1/\delta)$. Then there is a probabilistic algorithm which, given any n, $\delta > 0$, $\epsilon > 0$, $\eta > 0$, achieves the following in $\mathrm{poly}(n, 1/\delta, 1/\epsilon, \log(1/\eta))$ time. Given oracle access to a measure N with $\Phi_N \in \mathcal{C}^\epsilon$ (where $\mathcal{C}^\epsilon$ denotes the ϵ-neighborhood of $\mathcal{C}$; see Definition 2.11), the algorithm w.p. at least $1-\eta$ computes a measure M where Φ_M is an ϵ^2-approximate KL projection of Φ_N on $\mathcal{C}$.*

Specifically, $M(x) = \min(1, c \cdot N(x))$ *for some constant* $c \in [1, 1 + e^{\epsilon}]$ *as a multiple of* $\Omega(\epsilon^2)$.

This follows immediately from Lemma 2.3 of Barak et al. [BHK], where they show how to approximate the KL projection on the set of high density *measures* (rather than high min-entropy distributions), which is equivalent to KL projection on high density distributions.

Proof. For measures M and N, we define the *KL divergence from* M *to* N to be

$$\mathrm{KL}\left(M \| N\right) = \sum_x \left(M(x) \log \frac{M(x)}{N(x)} - M(x) + N(x) \right).$$

Note that

$$\mathrm{KL}\left(\Phi_M \| \Phi_N\right) = \frac{\mathrm{KL}\left(M \| N\right)}{|M|} + 1 - \frac{|N|}{|M|} + \log \frac{|N|}{|M|}.$$

Barak et al. [BHK] show how to compute $\widetilde{M^*}$, a $\sigma \cdot (\delta 2^n)$-approximate KL projection of N on the set of high density measures $\mathcal{M}_\delta$. Let M^* be the KL projection of N on $\mathcal{M}_\delta$, with $|M^*| = \delta 2^m$ (w.l.o.g. the KL projection is always on the boundary; see Lemma A.4). Thus by the above equality, Φ_{M^*} is the (exact) KL projection of N on $\mathcal{C}_\delta$. Furthermore, for every $M \in \mathcal{M}_\delta$,

$$\begin{aligned}
&\mathrm{KL}\left(\Phi_M \| \Phi_{\widetilde{M^*}}\right) - \mathrm{KL}\left(\Phi_M \| \Phi_{M^*}\right) \\
&= \frac{\mathrm{KL}\left(M \| \widetilde{M^*}\right) - \mathrm{KL}\left(M \| M^*\right)}{|M|} - \left(\frac{\left|\widetilde{M^*}\right|}{|M|} - \frac{|M^*|}{|M|} \right) + \left(\log \frac{\left|\widetilde{M^*}\right|}{|M|} - \log \frac{|M^*|}{|M|} \right) \\
&\leq \frac{\mathrm{KL}\left(M \| \widetilde{M^*}\right) - \mathrm{KL}\left(M \| M^*\right)}{|M|}
\end{aligned}$$

where the inequality holds because $\left|\widetilde{M^*}\right| \geq |M^*|$. Thus $\Phi_{\widetilde{M^*}}$ is a σ-approximate KL projection of Φ_N on $\mathcal{C}_\delta$. The parameters follow from Lemma 2.3 of [BHK]. □

Lemma A.4. *The KL projection of any measure* N *on any convex set* $\mathcal{M} \not\ni N$ *must be on the boundary of* $\mathcal{M}$.

Proof. Follows since the KL projection minimizes the convex function $\mathrm{KL}(\cdot \parallel N)$. □

For the following lemmas, refer to Chapter 4, Section 4.3 for the definition of $\mathbf{e}^W$.

Lemma A.5. *For every function $W : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$, $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{kW}|X)$ is monotone decreasing in k for $k \in [0, +\infty)$.*

Proof. Consider any $k_2 \geq k_1 \geq 0$. Applying Lemma 4.43 Item 1 twice:

$$\begin{aligned}&\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_2 W}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_1 W}|X) - k_2 \cdot \mathbb{E}\left[W(X, \mathbf{e}^{k_1 W}) - W(X, \mathbf{e}^{k_2 W})\right] \\ &= \mathrm{KL}(X, \mathbf{e}^{k_1 W} || X, \mathbf{e}^{k_2 W}) \geq 0,\end{aligned}$$

$$\begin{aligned}&\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_1 W}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_2 W}|X) - k_1 \cdot \mathbb{E}\left[W(X, \mathbf{e}^{k_2 W}) - W(X, \mathbf{e}^{k_1 W})\right] \\ &= \mathrm{KL}(X, \mathbf{e}^{k_2 W} || X, \mathbf{e}^{k_1 W}) \geq 0,\end{aligned}$$

where we use nonnegativity of KL divergence. Scaling the inequalities by k_1 and k_2 resp. and taking the sum yields

$$(k_2 - k_1)\left(\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_1 W}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_2 W}|X)\right) \geq 0,$$

i.e. $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_1 W}|X) \geq \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{k_2 W}|X)$. □

Lemma A.6 (Approximations)**.**

1. *There is a time $\mathrm{poly}(t, n, \ell, \kappa, \log(1/\sigma))$ deterministic algorithm that, given a size t deterministic circuit $\tilde{W} : \{0,1\}^n \times \{0,1\}^\ell \to [0, \kappa]$, $\sigma > 0$, and $\kappa > 0$, outputs a deterministic circuit $P : \{0,1\}^n \times \{0,1\}^\ell \to [1, e^\kappa]$ satisfying the following. For all functions W where $\forall x, a$, $\left|W(x,a) - \tilde{W}(x,a)\right| \leq \sigma$ and for all functions $W' : \{0,1\}^n \times \{0,1\}^\ell \to [0, \kappa']$:*

$$\left|\mathbb{E}\left[W'(X, \Phi_P)\right] - \mathbb{E}\left[W'(X, \mathbf{e}^W)\right]\right| = \kappa' \cdot O(\sigma),$$

$$\left|\mathrm{KL}(X,B||X,\Phi_P) - \mathrm{KL}(X,B||X,\mathbf{e}^W)\right| = O(\sigma),$$

$$\left|\mathrm{H_{sh}}(\Phi_P|X) - \mathrm{H_{sh}}(\mathbf{e}^W|X)\right| = \left(\mathrm{H_{sh}}(\mathbf{e}^W|X)+1\right)\cdot O(\sigma).$$

2. *There is a* $\mathrm{poly}(t,n,2^\ell,\kappa,1/\epsilon,\log(1/\gamma))$ *time randomized algorithm that given a size* t *deterministic circuit* $W:\{0,1\}^n\times\{0,1\}^\ell\to[0,\kappa]$*,* $\epsilon>0$*,* $\kappa>0$*, and* $\gamma>0$*, with probability at least* $1-\gamma$ *(over its coins) estimates* $\mathrm{H_{sh}}(\mathbf{e}^W|X)$ *within* $O(\epsilon)$ *additive error.*

3. *There is a* $\mathrm{poly}(t,n,2^\ell,\kappa,1/\epsilon,\log(1/\gamma))$ *time randomized oracle algorithm such that the following holds for all joint distributions* (X,B) *on* $\{0,1\}^n\times\{0,1\}^\ell$*. Given oracle access to* $O_{X,B}$*, a size* t *deterministic circuit* $W:\{0,1\}^n\times\{0,1\}^\ell\to[0,\kappa]$*,* $\epsilon>0$*,* $\kappa>0$*, and* $\gamma>0$*, the algorithm w.p. at least* $1-\gamma$ *(over its coins) estimates both* $\mathbb{E}[W(X,B)]-\mathbb{E}[W(X,\mathbf{e}^W)]$ *and* $\mathrm{KL}(X,B||X,\mathbf{e}^W)+\mathrm{H_{sh}}(B|X)$ *within* $O(\epsilon)$ *additive error.*

Proof. For Item 1, we construct a circuit P such that for all x,a, $\left|e^{\tilde{W}(x,a)}-P(x,a)\right|\le\sigma$. To do so, we approximate $e^{\tilde{W}(x,a)}\in[1,\exp(\kappa)]$ to precision $\pm\sigma$ using Newton's method in time $\mathrm{poly}(n,\ell,t,\kappa,\log(1/\sigma))$.

We now prove the required bounds. First we claim $P(x,a)/2^{W(x,a)}\in\left[e^{-O(\sigma)},e^{O(\sigma)}\right]$, because

$$\begin{aligned}
&|\log P(x,a) - D(x,a)|\\
&\le \left|\tilde{W}(x,a)-W(x,a)\right| + \left|\log P(x,a)-\tilde{W}(x,a)\right|\\
&\le \sigma + \left|\log\left(1-\frac{e^{\tilde{W}(x,a)}-P(x,a)}{e^{\tilde{W}(x,a)}}\right)\right|\\
&\le \sigma + \left|\log\left(1\pm\frac{\sigma}{e^{\tilde{W}(x,a)}}\right)\right|\\
&\le \sigma + |\log(1\pm\sigma)| = O(\sigma),
\end{aligned}$$

where we use $e^{\tilde{W}(x,a)} \geq 1$ in the last inequality. With this, we can bound the following quantities:

$$\begin{aligned}
\left|\Phi_P(a|x) - \mathbf{e}^W(a|x)\right| &= \left|\frac{P(x,a)}{\sum_b P(x,b)} - \frac{e^{W(x,a)}}{\sum_b e^{W(x,b)}}\right| \\
&\leq \left|\frac{e^{W(x,a)} \cdot e^{\pm O(\sigma)}}{\sum_b e^{W(x,b)} \cdot e^{\pm O(\sigma)}} - \frac{e^{W(x,a)}}{\sum_b e^{W(x,b)}}\right| \\
&\leq \frac{e^{W(x,a)}}{\sum_b e^{W(x,b)}}\left(e^{O(\sigma)} - 1\right) \\
&= \mathbf{e}^W(a|x) \cdot O(\sigma)
\end{aligned} \tag{A.1}$$

and

$$\begin{aligned}
&\left|\log\frac{1}{\Phi_P(a|x)} - \log\frac{1}{\mathbf{e}^W(a|x)}\right| \\
&\leq \left|\log P(x,a) - W(x,a)\right| + \left|\log\left(\sum_b P(x,b)\right) - \log\left(\sum_b e^{W(x,b)}\right)\right| \\
&\leq O(\sigma) + \left|\log\frac{\sum_b e^{W(x,b)} \cdot e^{\pm O(\sigma)}}{\sum_b e^{W(x,b)}}\right| = O(\sigma).
\end{aligned} \tag{A.2}$$

Using (A.1) and (A.2), we show the required bounds in turn:

$$\begin{aligned}
\left|\mathbb{E}\left[W'(X,\Phi_P)\right] - \mathbb{E}\left[W'(X,\mathbf{e}^W)\right]\right| &\leq \mathop{\mathbb{E}}_{x\sim X}\left[\kappa' \sum_a \left|\Phi_P(a|x) - \mathbf{e}^W(a|x)\right|\right] \\
&\leq \mathop{\mathbb{E}}_{x\sim X}\left[\kappa'\left(\sum_a \mathbf{e}^W(a|x)\right) O(\sigma)\right] = \kappa' \cdot O(\sigma),
\end{aligned}$$

where the last inequality follows from (A.1).

$$\begin{aligned}
&\left|\mathrm{KL}(X,B||X,\Phi_P) - \mathrm{KL}(X,B||X,\mathbf{e}^W)\right| \\
&= \left|\mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\log\frac{B(a|x)}{\Phi_P(a|x)}\right] - \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\log\frac{B(a|x)}{\mathbf{e}^W(a|x)}\right]\right| \\
&\leq \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a B(a|x)\left|\log\frac{1}{\Phi_P(a|x)} - \log\frac{1}{\mathbf{e}^W(a|x)}\right|\right], \\
&\leq O(\sigma)
\end{aligned}$$

where the last inequality follows from (A.2).

$$
\begin{aligned}
&\left|\mathrm{H}_{\mathrm{sh}}(\Phi_P|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^W|X)\right| \\
&= \left| \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a \Phi_P(a|x)\log\frac{1}{\Phi_P(a|x)}\right] - \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a \mathbf{e}^W(a|x)\log\frac{1}{\mathbf{e}^W(a|x)}\right]\right| \\
&\le \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a \left|\Phi_P(a|x)\log\frac{1}{\Phi_P(a|x)} - \mathbf{e}^W(a|x)\log\frac{1}{\mathbf{e}^W(a|x)}\right|\right] \\
&= \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a \left|\Phi_P(a|x)\left(\log\frac{1}{\Phi_P(a|x)} - \log\frac{1}{\mathbf{e}^W(a|x)}\right) + \log\frac{1}{\mathbf{e}^W(a|x)}\left(\Phi_P(a|x) - \mathbf{e}^W(a|x)\right)\right|\right] \\
&\le \mathop{\mathbb{E}}_{x\sim X}\left[\sum_a \left(\Phi_P(a|x)\cdot O(\sigma) + \log\frac{1}{\mathbf{e}^W(a|x)}\cdot \mathbf{e}^W(a|x)\cdot O(\sigma)\right)\right] \\
&\le \left(\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^W|X) + 1\right)\cdot O(\sigma),
\end{aligned}
$$

where the second last inequality follows from (A.1) and (A.2).

For Item 2 and Item 3, first note that we only need to estimate $\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^W|X)$ and $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,\mathbf{e}^W)]$ within $O(\epsilon)$ error, since by Lemma 4.43 Item 1,

$$
\mathrm{KL}(X,B||X,\mathbf{e}^W) + \mathrm{H}_{\mathrm{sh}}(B|X) = \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^W|X) - \left(\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,\mathbf{e}^W)]\right).
$$

By Item 1 (where we set $\sigma = O(\epsilon/\kappa\ell)$), it suffices to estimate $\mathrm{H}_{\mathrm{sh}}(\Phi_P|X)$ and $\mathbb{E}[W(X,B)] - \mathbb{E}[W(X,\Phi_P)]$ within $O(\epsilon)$ error. Recall $\Phi_P(a|x)$ can be computed in time $\mathrm{poly}(t,n,2^\ell,\kappa,\log(1/\epsilon))$.

Consider $\mathrm{H}_{\mathrm{sh}}(\Phi_P|X) = \mathbb{E}_{x\sim X}\left[\mathrm{H}_{\mathrm{sh}}(\Phi_P|_{X=x})\right]$. For each x we can compute the value $\mathrm{H}_{\mathrm{sh}}(\Phi_P|_{X=x}) = -\sum_a \Phi_P(a|x)\log\Phi_P(a|x)$ within ϵ error in time $\mathrm{poly}(t,n,2^\ell,\kappa,\log(1/\epsilon))$, by approximating $\log\Phi_P(a|x)$ to precision ϵ using Taylor series. We can then estimate $\mathbb{E}_{x\sim X}[\mathrm{H}_{\mathrm{sh}}(\Phi_P|_{X=x})]$, where $\mathrm{H}_{\mathrm{sh}}(\Phi_P|_{X=x}) \in [0,\ell]$, from $O(\log(1/\gamma)(\ell/\epsilon)^2)$ random samples of x. By a Chernoff bound, w.p. at least $1-\gamma$ the estimate is within $O(\epsilon)$ error.

Similarly, we can estimate $\mathbb{E}\left[W(X,B)\right] - \mathbb{E}\left[W(X,\Phi_P)\right]$, where $W(x,a) \in [0,\kappa]$, from $O\left(\log(1/\gamma)(\kappa/\epsilon)^2\right)$ random samples of (X,B,Φ_P). Note that (X,B) can be sampled using $O_{X,B}$, and Φ_P can be sampled given X in time $\mathrm{poly}(t,n,2^\ell,\kappa,\log(1/\epsilon))$. By a Chernoff bound, w.p. at least $1-\gamma$ the estimate is within $O(\epsilon)$ error. □

Lemma A.7. *For any functions $W_1, W_2 : \{0,1\}^n \times \{0,1\}^\ell \to \mathbb{R}_{\geq 0}$ and any joint distribution (X, B) on $\{0,1\}^n \times \{0,1\}^\ell$, we have*

$$\left|\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_1}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_2}|X)\right| = \left(\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_2}|X) + 1\right) \cdot O\left(\max_{x,a} |W_1(x,a) - W_2(x,a)|\right),$$

$$\left|\mathrm{KL}(X, B||X, \mathbf{e}^{W_1}) - \mathrm{KL}(X, B||X, \mathbf{e}^{W_2})\right| = O\left(\max_{x,a} |W_1(x,a) - W_2(x,a)|\right).$$

Proof. Setting $\tilde{W} = W_1$, $W = W_2$ and $\sigma = \max_{x,a} |W_1(x,a) - W_2(x,a)|$ in Lemma A.6, we obtain

$$\begin{aligned}
&\left|\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_1}|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_2}|X)\right| \\
&\leq \left|\mathrm{H}_{\mathrm{sh}}(\Phi_P|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_1}|X)\right| + \left|\mathrm{H}_{\mathrm{sh}}(\Phi_P|X) - \mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_2}|X)\right| \\
&= \left(\mathrm{H}_{\mathrm{sh}}(\mathbf{e}^{W_2}|X) + 1\right) \cdot O\left(\max_{x,a} |W_1(x,a) - W_2(x,a)|\right).
\end{aligned}$$

Setting $\tilde{W} = W_1$, $W = W_2$ and $\sigma = \max_{x,a} |W_1(x,a) - W_2(x,a)|$ in Lemma A.6, we obtain

$$\begin{aligned}
&\left|\mathrm{KL}(X, B||X, \mathbf{e}^{W_1}) - \mathrm{KL}(X, B||X, \mathbf{e}^{W_2})\right| \\
&\leq \left|\mathrm{KL}(X, B||X, \Phi_P) - \mathrm{KL}(X, B||X, \mathbf{e}^{W_1})\right| + \left|\mathrm{KL}(X, B||X, \Phi_P) - \mathrm{KL}(X, B||X, \mathbf{e}^{W_2})\right| \\
&= O(\sigma) = O\left(\max_{x,a} |W_1(x,a) - W_2(x,a)|\right).
\end{aligned}$$

□