Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 38 (2019)

Statistical Difference Beyond the Polarizing Regime

Itay Berman* Akshay Degwekar* Ron D. Rothblum†
Prashant Nalini Vasudevan‡

May 14, 2019

Abstract

The polarization lemma for statistical distance (SD), due to Sahai and Vadhan (JACM, 2003), is an efficient transformation taking as input a pair of circuits (C_0, C_1) and an integer k and outputting a new pair of circuits (D_0, D_1) such that if $\mathrm{SD}(C_0, C_1) \geq \alpha$ then $\mathrm{SD}(D_0, D_1) \geq 1-2^{-k}$ and if $\mathrm{SD}(C_0, C_1) \leq \beta$ then $\mathrm{SD}(D_0, D_1) \leq 2^{-k}$. The polarization lemma is known to hold for any constant values $\beta < \alpha^2$, but extending the lemma to the regime in which $\alpha^2 \leq \beta < \alpha$ has remained elusive. The focus of this work is in studying the latter regime of parameters. Our main results are:

1. Polarization lemmas for different notions of distance, such as *Triangular Discrimination* (TD) and *Jensen-Shannon Divergence* (JS), which enable polarization for some problems where the statistical distance satisfies $\alpha^2 < \beta < \alpha$. We also derive a polarization lemma for statistical distance with any inverse-polynomially small gap between α^2 and β (rather than a constant).
2. The average-case hardness of the statistical difference problem (i.e., determining whether the statistical distance between two given circuits is at least α or at most β), for any values of $\beta < \alpha$, implies the existence of one-way functions. Such a result was previously only known for $\beta < \alpha^2$.
3. A (direct) constant-round interactive proof for estimating the statistical distance between any two distributions (up to any inverse polynomial error) given circuits that generate them. Proofs of closely related statements have appeared in the literature but we give a new proof which we find to be cleaner and more direct.

*MIT. Emails: {itayberm, akshayd}@mit.edu. Research supported in part by NSF Grants CNS-1413920 and CNS-1350619, MIT-IBM Award, and by the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Research Office under Vinod Vaikuntanathan's DARPA Young Faculty Award and contracts W911NF-15-C-0226 and W911NF-15-C-0236.

†Technion. Email: rothblum@cs.technion.ac.il. This research was supported in part by the Israeli Science Foundation (Grant No. 1262/18).

‡UC Berkeley. Email: prashvas@berkeley.edu. Research supported in part from DARPA/ARL SAFEWARE Award W911NF15C0210, AFOSR Award FA9550-15-1-0274, AFOSR YIP Award, a Hellman Award and research grants by the Okawa Foundation, Visa Inc., and Center for LongTerm Cybersecurity (CLTC, UC Berkeley).

ISSN 1433-8092

Contents

1 Introduction **3**
- 1.1 Our Results . . . 4
- 1.2 Additional Related Works . . . 8
- 1.3 Organization . . . 9

2 Techniques **9**
- 2.1 One-Way Function From Statistical Difference with Any Noticeable Gap . . . 9
- 2.2 Interactive Proof for Statistical Distance Approximation . . . 12
- 2.3 TDP and JSP are SZK-Complete . . . 14

3 Preliminaries **17**
- 3.1 Information Theory Preliminaries . . . 18
- 3.2 Statistical Zero-Knowledge Interactive Proofs . . . 20

4 Complete Problems for SZK **24**
- 4.1 JSP is Complete for SZK . . . 24
- 4.2 TDP is Complete for SZK . . . 27

5 One Way Functions from SDP with Any Noticeable Gap **33**
- 5.1 Proving Theorem 5.5 . . . 34

6 Estimating Statistical Distance in AM ∩ coAM **39**
- 6.1 Proofs of Intermediates . . . 46

A Triangular Discrimination Inequalities **52**

1 Introduction

The Statistical Difference Problem, introduced by Sahai and Vadhan [SV03], is a central computational (promise) problem in complexity theory and cryptography, which is also intimately related to the study of statistical zero-knowledge (SZK). The input to this problem is a pair of circuits C_0 and C_1, specifying probability distributions (i.e., that are induced by feeding the circuits with a uniformly random string). YES instances are those in which the statistical distance[1] between the two distributions is at least 2/3 and NO instances are those in which the distance is at most 1/3. Input circuits that do not fall in one of these two cases are considered to be outside the promise (and so their value is left unspecified).

The choice of the constants 1/3 and 2/3 in the above definition is somewhat arbitrary (although not entirely arbitrary as will soon be discussed in detail). A more general family of problems can be obtained by considering a suitable parameterization. More specifically, let $0 \leq \beta < \alpha \leq 1$. The (α, β) parameterized version of the Statistical Difference Problem, denoted $\mathrm{SDP}^{\alpha,\beta}$, has as its YES inputs pairs of circuits that induce distributions that have distance at least α whereas the NO inputs correspond to circuits that induce distributions that have distance at most β.

Definition 1.1 (Statistical Difference Problem)**.** *Let $\alpha, \beta \colon \mathbb{N} \to [0,1]$ with $\alpha(n) > \beta(n)$ for every n. The* Statistical Difference Problem with promise (α, β)*, denoted $\mathrm{SDP}^{\alpha,\beta}$, is given by the sets*

$$\mathrm{SDP}^{\alpha,\beta}_Y = \{(C_0, C_1) \mid \mathrm{SD}(C_0, C_1) \geq \alpha(n)\} \text{ and}$$
$$\mathrm{SDP}^{\alpha,\beta}_N = \{(C_0, C_1) \mid \mathrm{SD}(C_0, C_1) \leq \beta(n)\},$$

where n is the output length of the circuits C_0 and C_1.[2]

(Here and below we abuse notation and use C_0 and C_1 to denote both the circuits and the respective distributions that they generate.)

The elegant *polarization lemma* of [SV03] shows how to polarize the statistical distance between two distributions. In more detail, for any constants α and β such that $\beta < \alpha^2$, the lemma gives a transformation that makes distributions that are at least α-far be extremely far and distributions that are β-close be extremely close. Beyond being of intrinsic interest, the polarization lemma is used to establish the SZK completeness of $\mathrm{SDP}^{\alpha,\beta}$, when $\alpha^2 > \beta$, and has other important applications in cryptography such as the amplification of weak public key encryption schemes to full fledged ones [DNR04, HR05].

Sahai and Vadhan left the question of polarization for parameters α and β that do not meet the requirements of their polarization lemma as an open question. We refer to this setting of α and β as the *non-polarizing* regime. We emphasize that by *non-polarizing* we merely mean that in this regime polarization is not currently known and not that it is impossible to achieve (although some barriers are known and will be discussed further below). The focus of this work is studying the Statistical Difference Problem in the non-polarizing regime.

[1] Recall that the statistical distance between two distributions P and Q over a set $\mathcal{Y}$ is defined as $\mathrm{SD}(P,Q) = \frac{1}{2}\sum_{y\in\mathcal{Y}}|P_y - Q_y|$, where P_y (resp., Q_y) is the probability mass that P (resp., Q) puts on $y \in \mathcal{Y}$.

[2] In prior works α and β were typically thought of as constants (and so their dependence on the input was not specified). In contrast, since we will want to think of them as parameters, we choose to let them depend on the output length of the circuit since this size seems most relevant to the distributions induced by the circuits. Other natural choices could have been the input length or the description size of the circuits. We remark that these different choices do not affect our results in a fundamental way.

1.1 Our Results

We proceed to describe our results.

1.1.1 Polarization and SZK Completeness for Other Notions of Distance

The statistical distance metric is one of the central information theoretic tools used in cryptography as it is very useful for capturing similarity between distributions. However, in information theory there are other central notions that measure similarity such as mutual information and KL divergence as well as others.

Loosely speaking, our first main result shows that polarization is possible even in *some* cases in which $\beta \geq \alpha^2$. However, this result actually stems from a more general study showing that polarization is possible for other notions of distance between distributions from information theory, which we find to be of independent interest.

When distributions are extremely similar or extremely dissimilar, these different notions of distance are often (but not always) closely related and hence interchangeable. This equivalence is particularly beneficial when considering applications of SZK—for some applications one distance measure may be easier to use than others. For example, showing that the average-case hardness of SZK implies one-way functions can be analyzed using statistical distance (e.g., [Vad99, Section 4.8]), but showing that every language in SZK has instance-dependent commitments is naturally analyzed using entropy (e.g., [OV08]).

However, as the gaps in the relevant distances get smaller (i.e., the distributions are only somewhat similar or dissimilar), the relation between different statistical properties becomes less clear (for example, the reduction from $\mathrm{SDP}^{\alpha,\beta}$ to the Entropy Difference Problem of [GV99] only works when roughly $\alpha^2 > \beta$). This motivates studying the computational complexity of problems defined using different notions of distance in this small gap regime. Studying this question can be (and, as we shall soon see, indeed is) beneficial in two aspects. First, providing a wider bag of statistical properties related to SZK, which can make certain applications easier to analyze. Second, the computational complexity of these distance notions might shed light on the computational complexity of problems involving existing distance notions (e.g., $\mathrm{SDP}^{\alpha,\beta}$ when $\alpha^2 < \beta$).

We focus here on two specific distance notions—the *triangular discrimination* and the *Jensen-Shannon divergence*, defined next.

Definition 1.2 (Triangular Discrimination)**.** *The* Triangular Discrimination *(a.k.a. Le Cam divergence) between two distributions P and Q is defined as*

$$\mathrm{TD}(P,Q) = \frac{1}{2}\sum_{y\in\mathcal{Y}} \frac{(P_y - Q_y)^2}{P_y + Q_y},$$

where $\mathcal{Y}$ is the union of the supports of P and Q.

The Triangular Discrimination Problem with promise (α,β)*, denoted* $\mathrm{TDP}^{\alpha,\beta}$*, is defined analogously to* $\mathrm{SDP}^{\alpha,\beta}$*, but with respect to* TD *rather than* SD*.*

The triangular discrimination is commonly used, among many other applications, in statistical learning theory for parameter estimation with quadratic loss, see [Cam86, P. 48] (in a similar manner to how statistical distance characterizes the 0-1 loss function in hypothesis testing). Jumping ahead, while the definition of triangular discrimination seems somewhat arbitrary at first glance, in Section 2 we will show that this distance notion characterizes some basic phenomena in the study of statistical zero-knowledge. Triangular discrimination has recently found usage in theoretical computer science, and even specifically in problems related to SZK. Yehudayoff [Yeh16] showed that

using TD yields a tighter analysis of the pointer chasing problem in communication complexity. The work of Komargodski and Yogev [KY18] uses triangular discrimination to show that the average-case hardness of SZK implies the existence of distributional collision resistant hash functions.

Next, we define the *Jensen-Shannon Divergence*. First, recall that the KL-divergence between two distributions P and Q is defined[3] as $\mathrm{KL}(P||Q) = \sum_{y \in \mathcal{Y}} P_y \log(P_y/Q_y)$. Also, given distributions P_0 and P_1 we define the distribution $\frac{1}{2}P_0 + \frac{1}{2}P_1$ as the distribution obtained by sampling a random coin $b \in \{0,1\}$ and outputting a sample y from P_b (indeed, this notation corresponds to arithmetic operations on the probability mass functions). The Jensen-Shannon divergence measures the mutual information between b and y.

Definition 1.3 (Jensen-Shannon Divergence)**.** *The* Jensen-Shannon divergence *between two distributions P and Q is defined as*

$$\mathrm{JS}(P,Q) = \frac{1}{2}\,\mathrm{KL}\left(P \middle\| \frac{P+Q}{2}\right) + \frac{1}{2}\,\mathrm{KL}\left(Q \middle\| \frac{P+Q}{2}\right).$$

The Jensen-Shannon Divergence Problem *with promise (α,β), denoted $\mathrm{JSP}^{\alpha,\beta}$, is defined analogously to $\mathrm{SDP}^{\alpha,\beta}$, but with respect to* JS *rather than* SD*.*

The Jensen-Shannon divergence enjoys a couple of important properties (in our context) that the KL-divergence lacks: it is symmetric and bounded. Both triangular discrimination and Jensen-Shannon divergence (as well as statistical distance and KL-divergence) are types of f-divergences, a central concept in information theory (see [PW17, Section 6] and references therein). They are both non-negative and bounded by one.[4] Finally, the Jensen-Shannon divergence is a metric, while the triangular discrimination is a square of a metric.

With these notions of distance and corresponding computational problems in hand, we are almost ready to state our first set of results. Before doing so, we introduce an additional useful technical definition.

Definition 1.4 (Separated functions)**.** *Let $g\colon \mathbb{N} \to [0,1]$. A pair of $\mathsf{poly}(n)$-time computable functions (α,β), where $\alpha = \alpha(n) \in [0,1]$ and $\beta = \beta(n) \in [0,1]$, is g-*separated *if $\alpha(n) \geq \beta(n) + g(n)$ for every $n \in \mathbb{N}$.*

*We denote by $(1/\mathsf{poly})$-*separated *the set of all pairs of functions that are $(1/p)$-separated for some polynomial p. Similarly, we denote by $(1/\mathsf{log})$-*separated *the set of all pairs of functions that are $(1/(c\log))$-separated for some constant $c > 0$.*

We can now state our first set of results: that both TDP and JSP, with a noticeable gap, are SZK complete.

Theorem 1.5. *Let (α,β) be $(1/\mathsf{poly})$-separated functions such that there exists a constant $\varepsilon \in (0,1/2)$ such that $2^{-n^{1/2-\varepsilon}} \leq \beta(n)$ and $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$, for every $n \in \mathbb{N}$. Then, $\mathrm{TDP}^{\alpha,\beta}$ is SZK complete.*

Theorem 1.6. *For (α,β) as in Theorem 1.5, the problem $\mathrm{JSP}^{\alpha,\beta}$ is SZK complete.*

The restriction on $2^{-n^{1/2-\varepsilon}} \leq \beta(n)$ and $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$ should be interpreted as a non-degeneracy requirement (which we did not attempt to optimize), where we note that some restriction

[3] To be more precise, in this definition we view $0 \cdot \log \frac{0}{0}$ as 0 and define the KL-divergence to be ∞ if the support of P is not contained in that of Q.

[4] In the literature these distances are sometimes defined to be twice as much as our definitions. In our context, it is natural to have the distances bounded by one.

seems inherent (see Remark 3.15 below). Moreover, we can actually decouple the assumptions in Theorems 1.5 and 1.6 as follows. To show that $\mathrm{TDP}^{\alpha,\beta}$ and $\mathrm{JSP}^{\alpha,\beta}$ are SZK-*hard*, only the non-degeneracy assumption (i.e., $2^{-n^{1/2-\varepsilon}} \leq \beta(n)$ and $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$) is needed. On the other hand, to show that these problems are in SZK we only require that (α, β) are $(1/\mathsf{poly})$-separated.

Note that in particular, Theorems 1.5 and 1.6 imply polarization lemmas for both TD and JS. For example, for triangular discrimination, since $\mathrm{TDP}^{\alpha,\beta} \in \mathsf{SZK}$ and $\mathrm{TDP}^{1-2^{-k},2^{-k}}$ is SZK-hard, one can reduce the former to the latter.

Beyond showing polarization for triangular discrimination, Theorem 1.5 has implications regarding the question of polarizing statistical distance, which was our original motivation. It is known that the triangular discrimination is sandwiched between the statistical distance and its square; namely, for every two distributions P and Q it holds that (see [Top00, Eq. (11)]):

$$\mathrm{SD}(P,Q)^2 \leq \mathrm{TD}(P,Q) \leq \mathrm{SD}(P,Q) \tag{1.1}$$

(for self containment we include a proof of this fact in Appendix A.)

Thus, the problem $\mathrm{SDP}^{\alpha,\beta}$ is immediately reducible to $\mathrm{TDP}^{\alpha^2,\beta}$, which Theorem 1.5 shows to be SZK-complete, as long as the gap between α^2 and β is noticeable. Specifically, we have the following corollary.

Corollary 1.7. *Let (α, β) be as in Theorem 1.5, with the exception that (α^2, β) are $(1/\mathsf{poly})$-separated (note that here α is squared). Then, the promise problem $\mathrm{SDP}^{\alpha,\beta}$ is SZK complete.*

We highlight two implications of Theorem 1.5 and Corollary 1.7 (which were also briefly mentioned above).

Polarization with Inverse Polynomial Gap. Observe that Corollary 1.7 implies polarization of statistical distance in a regime in which α and β are functions of n, the output length of the two circuits, and α^2 and β are only separated by an inverse polynomial. This is in contrast to most prior works which focus on α and β that are constants. In particular, Sahai and Vadhan's [SV03] proof of the polarization lemma focuses on constant α and β and can be extended to handle an inverse logarithmic gap, but does not seem to extend to an inverse polynomial gap.[5] Corollary 1.7 does yield such a result, by relying on a somewhat different approach.

Polarization Beyond $\alpha^2 > \beta$. Theorem 1.5 can sometimes go beyond the requirement that $\alpha^2 > \beta$ for polarizing statistical distance. Specifically, it shows that any problem with noticeable gap in the triangular discrimination can be polarized. Indeed, there are distributions (P, Q) and (P', Q') with $\mathrm{SD}(P,Q) > \mathrm{SD}(P',Q') > \mathrm{SD}(P,Q)^2$ but still $\mathrm{TD}(P,Q) > \mathrm{TD}(P',Q')$.[6] Circuits generating such distributions were until now not known to be in the polarizing regime, but can now be polarized by combining Theorem 1.5 and Eq. (1.1).

1.1.2 From Statistical Difference to One-way Functions

We continue our study of the Statistical Difference Problem, focusing on the regime where $\beta < \alpha$ (and in particular even when $\beta \geq \alpha^2$). We show that in this regime the $\mathrm{SDP}^{\alpha,\beta}$ problem

[5] Actually, it was claimed in [GV11] that the [SV03] proof does extend to the setting of an inverse polynomial gap between α^2 and β but this claim was later retracted, see http://www.wisdom.weizmann.ac.il/~/oded/entropy.html.

[6] For example, for a parameter $\gamma \in [0,1]$ consider the distributions R_0^γ and R_1^γ over $\{0,1,2\}$: R_b^γ puts γ mass on b and $1-\gamma$ mass on 2. It holds that $\mathrm{SD}(R_0^\gamma, R_1^\gamma) = \mathrm{TD}(R_0^\gamma, R_1^\gamma) = \gamma$. If, say, $(P,Q) = (R_0^{1/2}, R_1^{1/2})$ and $(P',Q') = (R_0^{1/3}, R_1^{1/3})$, then $\mathrm{SD}(P,Q) > \mathrm{SD}(P',Q') > \mathrm{SD}(P,Q)^2$ but $\mathrm{TD}(P,Q) > \mathrm{TD}(P',Q')$.

shares many important properties of SZK (although we fall short of actually showing that it lies in SZK—which is equivalent to polarization for any $\beta < \alpha$).

First, we show that similarly to SZK, the average-case hardness of $\mathrm{SDP}^{\alpha,\beta}$ implies the existence of one-way functions. The fact that average-case hardness of SZK (or equivalently $\mathrm{SDP}^{\alpha,\beta}$ for $\beta < \alpha^2$) implies the existence of one-way functions was shown by Ostrovsky [Ost91]. Indeed, our contribution is in showing that the weaker condition of $\beta < \alpha$ (rather than $\beta < \alpha^2$) suffices for this result.

Theorem 1.8. *Let (α, β) be $(1/\mathsf{poly})$-separated functions and suppose that $\mathrm{SDP}^{\alpha,\beta}$ is average-case hard. Then, there exists a one-way function.*

The question of constructing one-way functions from the (average-case) hardness of SDP is closely related to a result of Goldreich's [Gol90] showing that the existence of efficiently sampleable distributions that are statistically far but computationally indistinguishable implies the existence of one-way functions. Our proof of Theorem 1.8 allows us to re-derive the following strengthening of [Gol90], due to Naor and Rothblum [NR06, Theorem 4.1]: for any $(1/\mathsf{poly})$-separated (α, β), the existence of efficiently sampleable distributions whose statistical distance is α but no efficient algorithm can distinguish between them with advantage more than β, implies the existence of one-way functions. See further discussion in Remark 2.1.

1.1.3 Interactive Proof for Statistical Distance Approximation

As our last main result, we construct a new interactive protocol that lets a verifier estimate the statistical distance between two given circuits up to any noticeable precision.

Theorem 1.9. *There exists a constant-round public-coin interactive protocol between a prover and a verifier that, given as input a pair of circuits (C_0, C_1), a claim $\Delta \in [0,1]$ for their statistical distance, and a tolerance parameter $\delta \in [0,1]$, satisfies the following properties:*

- **Completeness:** *If $\mathrm{SD}(C_0, C_1) = \Delta$, then the verifier accepts with probability at least $2/3$ when interacting with the honest prover.*
- **Soundness:** *If $|\mathrm{SD}(C_0, C_1) - \Delta| \geq \delta$, then when interacting with any (possibly cheating) prover, the verifier accepts with probability at most $1/3$.*
- **Efficiency:** *The verifier runs in time $\mathsf{poly}(|C_0|, |C_1|, 1/\delta)$.*

(As usual the completeness and soundness errors can be reduced by applying parallel repetition. We can also achieve perfect completeness using a result from [FGM+89].)

Theorem 1.9 is actually equivalent to the following statement.

Theorem 1.10 ([BL13, Theorem 6],[BBF16, Theorem 2])**.** *For any $(1/\mathsf{poly})$-separated (α, β), it holds that $\mathrm{SDP}^{\alpha,\beta} \in \mathsf{AM} \cap \mathsf{coAM}$.*[7]

It is believed that $\mathsf{AM} \cap \mathsf{coAM}$ lies just above SZK, and if we could show that $\mathrm{SDP}^{\alpha,\beta}$ is in SZK, that would imply SD polarization for such α and β.

Since Theorem 1.9 can be derived from existing results in the literature, we view our main contribution to be the proof which is via a single protocol that we find to be cleaner and more direct than alternate approaches.

[7] Recall that AM is the class of problems that have constant-round public-coin interactive proofs. coAM is simply the complement of AM.

Going into a bit more detail, [BL13, BBF16]'s proofs are in fact a combination of two separate constant-round protocols. The first protocol is meant to show that $\mathrm{SDP}^{\alpha,\beta} \in \mathsf{AM}$ and follows directly by taking the interactive proof for SDP presented by Sahai and Vadhan (which has completeness error $(1-\alpha)/2$ and soundness error $(1+\beta)/2$), and applying parallel repetition (and the private-coin to public-coin transformation of [GS89]).

The second protocol is meant to show that $\mathrm{SDP}^{\alpha,\beta} \in \mathsf{coAM}$, and is based on a protocol by Bhatnagar, Bogdanov, and Mossel [BBM11]. Another approach for proving that $\mathrm{SDP}^{\alpha,\beta} \in \mathsf{coAM}$ is by combining results of [GVW02] and [SV03]. Goldreich, Vadhan and Wigderson [GVW02] showed that problems with laconic interactive proofs, that is proofs where the communication from the prover to the verifier is small, have coAM proofs. Sahai and Vadhan [SV03], as described earlier, showed that $\mathrm{SDP}^{\alpha,\beta}$, and SZK in general, has an interactive proof where the prover communicates a single bit. Combining these results immediately gives a coAM protocol for $\mathrm{SDP}^{\alpha,\beta}$ *when* (α,β) *are* $\Omega(1)$*-separated.* As for (α,β) that are only $(1/\mathsf{poly})$-separated, while the [GVW02] result as-stated does not suffice, it seems that their protocol can be adapted to handle this case as well.[8]

As mentioned above, we give a different, and direct, proof of Theorem 1.9 that we find to be simpler and more natural than the above approach. In particular, our proof utilizes the techniques developed for our other results, which enable us to give a single and more general protocol—one that approximates the statistical difference (as in Theorem 1.9), rather than just deciding if that distance is large or small.

At a very high level, our protocol may be viewed as an application of the set-lower-bound-based techniques of Akavia et al [AGGM06] or Bogdanov and Brzuska [BB15] to our construction of a one-way function from the average-case hardness of SDP (i.e., Theorem 1.8), though there are technical differences in our setting. Both these papers show how to construct a coAM protocol for any language that can be reduced, to inverting a *size-verifiable* one-way function.[9] While we do not know how to reduce solving SDP in the worst-case to inverting any specific function, we make use of the fact that associated with each instance of SDP, there is an *instance-dependent* function [OW93], that is size-verifiable on the average.

1.2 Additional Related Works

Barriers to Improved Polarization. Holenstein and Renner [HR05] show that in a limited model dubbed "oblivious polarization", the condition $\alpha^2 > \beta$ on the statistical distance is necessary for polarizing statistical distance.[10] All the past polarization reductions fit in this framework and so do ours. Specifically, Holenstein and Renner show distributions where $\alpha^2 < \beta$ and cannot be polarized in this model. We show a condition that suffices for polarization, even for distributions where $\alpha^2 \leq \beta$. This does not contradict the [HR05] result because their distributions do not satisfy this condition.

In a more general model, [LZ17, CGVZ18] showed lower bounds for SZK-related distribution

[8] In more detail, the [GVW02] result is stated for protocols in which the gap between completeness and soundness is constant (specifically 1/3). In case α and β are only $1/\mathsf{poly}$-separated, the [SV03] protocol only has a $1/\mathsf{poly}$ gap (and we cannot afford repetition since it will increase the communication). Nevertheless, by inspecting the [GVW02] proof, it seems as though it can be adapted to cover any noticeable gap.

[9] Informally, a function f is size-verifiable if given an output $y = f(x)$, there exists an AM protocol to estimate $|f^{-1}(y)|$.

[10] Roughly speaking, an oblivious polarization is a randomized procedure to polarize without invoking the circuits; it takes as input a bit σ and an integer k, and outputs a sequence of bits $(b_1^\sigma, \ldots, b_\ell^\sigma)$ and a string r^σ. Given a pair of circuits (C_0, C_1), such a procedure defines a pair of circuits (D_0, D_1) as follows: D_σ samples $(b_1^\sigma, \ldots, b_\ell^\sigma)$ and r^σ and outputs $(C_{b_1^\sigma}, \ldots, C_{b_\ell^\sigma}, r^\sigma)$. We are guaranteed that if $\mathrm{SD}(C_0, C_1) \geq \alpha$, then $\mathrm{SD}(D_0, D_1) \geq 1 - 2^{-k}$, and if $\mathrm{SD}(C_0, C_1) \leq \beta$, then $\mathrm{SD}(D_0, D_1) \leq 2^{-k}$.

manipulation tasks. The model they consider allows the reduction arbitrary oracle access to the circuits that sample the distributions, as opposed to the more restricted model of oblivious polarization. In this model, Lovett and Zhang [LZ17] show that efficient entropy reversal is impossible[11], and Chen, Göös, Vadhan and Zhang [CGVZ18] showed that entropy flattening requires $\Omega(n^2)$ invocations to the underlying circuit. Showing lower bounds for polarization in this more general model remains an interesting open question.

Polarization for other Notions of Distance. Toward characterizing zero-knowledge in the help model, Ben-Or and Gutfreund [BG03] and Chailloux et al. [CCKV08] gave a polarization procedure that considers two different distances for every $(1/\mathsf{log})$-separated $\alpha > \beta$: if the statistical distance is at most β, then it decreases to 2^{-k}; and if the *mutual disjointness*[12] is at least α, then it increases to $1 - 2^{-k}$. Fehr and Vaudenay [FV17] raise the question of polarization for the fidelity measure[13] but leave resolving it as an open problem (see Section 2.3.3 for details).

SDP **and Cryptography.** We show that average-case hardness of $\mathrm{SDP}^{\alpha,\beta}$ implies one-way functions. In the reverse direction, Bitansky et al. [BDV17] show that one-way functions do not imply even worst-case hardness of $\mathrm{SDP}^{\alpha,\beta}$ in a black-box manner for any $(1/\mathsf{poly})$-separated α, β.[14]

1.3 Organization

In Section 2 we give an overview of the techniques that we use to prove our main results. Section 3 contains preliminaries. In Section 4 we prove that TDP and JSP are SZK complete. In Section 5 we construct a one-way function from average-case hardness of SDP. Lastly, in Section 6 we construct an interactive proof for estimating statistical distance.

2 Techniques

We begin in Section 2.1 by describing how to construct a one-way function from the average-case hardness of SD with any noticeable gap (Theorem 1.8). The techniques used there are also central in our interactive protocol for SD estimation (Theorem 1.9), which is described in Section 2.2, as well as in our proof that triangular discrimination and Jensen-Shannon divergence are SZK complete (Theorems 1.5 and 1.6), which are outlined in Section 2.3 below.

2.1 One-Way Function From Statistical Difference with Any Noticeable Gap

We first show the existence of *distributionally* one-way functions. Namely, an efficiently computable function f for which it is hard to sample a uniformly random pre-image for a random output y (rather than an arbitrary pre-image as in a standard one-way function). This suffices since Impagliazzo and Luby [IL89] showed how to convert a distributionally one-way function into a standard one.

Assume that we are given a distribution over a pair of circuits (C_0, C_1) such that it is hard to distinguish between the cases $\mathrm{SD}(C_0, C_1) \geq \alpha$ or $\mathrm{SD}(C_0, C_1) \leq \beta$, for some $\alpha > \beta + 1/\mathsf{poly}$. A

[11] Entropy reversal refers to the task of given circuit C and parameter t output (C', t') such that when $\mathrm{H}(C) > t$, then $\mathrm{H}(C') < t' - 1$ and if $\mathrm{H}(C) < t - 1$, then $\mathrm{H}(C') > t'$.

[12] For an ordered pair of distributions P and Q, their disjointness is $\mathrm{Disj}(P, Q) = \Pr_{y \sim P}[y \notin \mathrm{Supp}(Q)]$, and their mutual disjointness is $\mathrm{MutDisj}(P, Q) = \min(\mathrm{Disj}(P, Q), \mathrm{Disj}(Q, P))$.

[13] For two distributions P and Q, their fidelity is defined as $\mathrm{Fidelity}(P, Q) = \sum_y \sqrt{P_y \cdot Q_y}$.

[14] While [BDV17] state the result for constant α, β, the construction and analysis extend to our setting.

natural candidate for a one-way function is the (efficiently computable) function

$$f_{C_0,C_1}(b,x) = C_b(x). \tag{2.1}$$

Namely, f is parameterized by the circuits (C_0, C_1) (which are to be sampled according to the hard distribution), and the bit b chooses which of the two circuits would be evaluated on the string x. This function appears throughout the SZK literature (e.g., it corresponds to the verifier's message in the SDP protocol of [SV03]).

Assume that f is not distributionally one-way, and let A be an algorithm that given (C_0, C_1) and a random input y—sampled by first drawing a uniformly random bit b and a string x and then computing $y = C_b(x)$—outputs a uniformly random element (b', x') from the set $f^{-1}_{C_0,C_1}(y) = \{(b,x) \colon C_b(x) = y\}$. For simplicity, we assume that A is a perfect distributional inverter, that is for *every fixed* (C_0, C_1, y) it outputs uniformly random elements of $f^{-1}_{C_0,C_1}(y)$.

Arguably, the most natural approach for distinguishing between the cases of high or low statistical distance given the two circuits and the inverter, is to choose x and b at random, invoke the inverter to obtain (b', x'), and check whether $b = b'$. Indeed, if $\mathrm{SD}(C_0, C_1) = 1$, then $\Pr[b = b'] = 1$, and if $\mathrm{SD}(C_0, C_1) = 0$, then $\Pr[b = b'] = \frac{1}{2}$. Thus, we can distinguish between the cases with constant advantage.

But what happens when the gap in the statistical distance is smaller? To analyze this case we want to better understand the quantity $\Pr[b = b']$. It turns out that this quantity is characterized by the triangular discrimination between the circuits. Let P_b denote the output distribution of C_b. Using elementary manipulations (and the fact that $\frac{1}{2}(P_0 + P_1)$ is a distribution), it holds that[15]

$$\begin{aligned}
\Pr\big[b = b'\big] &= \frac{1}{2} \Pr_{y \sim P_0}\big[b' = 0\big] + \frac{1}{2} \Pr_{y \sim P_1}\big[b' = 1\big] \\
&= \frac{1}{2} \sum_y \frac{P_0(y)^2 + P_1(y)^2}{P_0(y) + P_1(y)} \\
&= \frac{1}{4} \sum_y \frac{(P_0(y) + P_1(y))^2}{P_0(y) + P_1(y)} + \frac{1}{4} \sum_y \frac{(P_0(y) - P_1(y))^2}{P_0(y) + P_1(y)} \\
&= \frac{1}{2} + \frac{1}{4} \sum_y \frac{(P_0(y) - P_1(y))^2}{P_0(y) + P_1(y)} \\
&= \frac{1 + \mathrm{TD}(C_0, C_1)}{2}.
\end{aligned} \tag{2.2}$$

Based on the general bounds between triangular discrimination and statistical distance (Eq. (1.1)), which are known to be tight, all we are guaranteed is

$$\begin{aligned}
\mathrm{SD}(C_0, C_1) \geq \alpha \quad &\Longrightarrow \quad \Pr\big[b = b'\big] \geq \frac{1 + \alpha^2}{2} \\
\mathrm{SD}(C_0, C_1) \leq \beta \quad &\Longrightarrow \quad \Pr\big[b = b'\big] \leq \frac{1 + \beta}{2}.
\end{aligned}$$

So, this approach is limited to settings in which $\alpha^2 > \beta$.

To overcome this limitation we want to find a quantity that is more tightly characterized by the statistical distance of the circuits. This quantity, which we call *imbalance*, will be central in all

[15] In Section 1 we used P_y to denoted the probability mass a distribution P puts on an element y, while here we use $P(y)$. In the rest of this work we choose which notation to use based on readability and context.

of the proofs in this work. The imbalance measures how likely it is that an output string y was generated from C_1 versus C_0. Formally,

$$\theta_y \triangleq \Pr[b=1|y] - \Pr[b=0|y] = \frac{P_1(y) - P_0(y)}{P_1(y) + P_0(y)}. \tag{2.3}$$

Elementary manipulations yields that

$$\begin{aligned} \mathrm{SD}(C_0, C_1) &= \frac{1}{2}\sum_y |P_1(y) - P_0(y)| \\ &= \sum_y \frac{1}{2}(P_1(y) + P_0(y)) \cdot \frac{|P_1(y) - P_0(y)|}{P_1(y) + P_0(y)} \\ &= \mathop{\mathbb{E}}_{y \sim (\frac{1}{2}P_0 + \frac{1}{2}P_1)} [|\theta_y|]. \end{aligned} \tag{2.4}$$

(Recall that y is sampled by first drawing a uniform random bit b and a string x, and setting $y = C_b(x)$. Hence, using the notation that P_b denotes the output distributions of the circuit C_b, the marginal distribution of y is $\frac{1}{2}P_0 + \frac{1}{2}P_1$.)

Eq. (2.4) naturally gives rise to the following algorithm for approximating $\mathrm{SD}(C_0, C_1)$:

Algorithm to estimate $\mathrm{SD}(C_0, C_1)$ using the inverter A:

1. Sample polynomially many $y_1, \ldots, y_t$.
2. For every y_i:
   (a) Call $\mathsf{A}(y_i)$ polynomially many times to get $b'_1, \ldots, b'_k$.
   (b) Let m be the number of ones in $b'_1, \ldots, b'_k$.
   (c) Set $p_1 = m/k$, $p_0 = (k-m)/k$ and $\widehat{\theta}_i = p_1 - p_0$.
3. Return $\frac{1}{t}\sum_{i=1}^t |\widehat{\theta}_i|$.

The quantities p_1 and p_0 are in fact the empirical distribution of b condition on y, computed using k samples. By choosing large enough k, we get that $(p_1, p_0) \approx (\Pr[b=1|y], \Pr[b=0|y])$ and so $\widehat{\theta}_i \approx \theta_{y_i}$. By then choosing large enough t, we get that $\frac{1}{t}\sum_{i=1}^t |\widehat{\theta}_i| \approx \mathrm{SD}(C_0, C_1)$. Hence, we can distinguish between the cases $\mathrm{SD}(C_0, C_1) \geq \alpha$ or $\mathrm{SD}(C_0, C_1) \leq \beta$, for any $\alpha > \beta + 1/\mathsf{poly}$.

Essentially the same proof continues to work if A is not a perfect distributional inverter, but is close enough to being so—that is, on input y its output distribution is close to being uniform over $f^{-1}(y)$ for most (but not all) tuples C_0, C_1, y. We leave handling these details to the formal proof in Section 5

The above proof strategy also yields a new proof for Naor and Rothblum's [NR06] strengthening of [Gol90].[16] See Remark 2.1 below for a discussion about the differences between our techniques and those of [NR06].

Distributional Collision Resistant Hash Function. As a matter of fact, the above proof also shows that the average-case hardness of $\mathrm{SDP}^{\alpha,\beta}$ also implies that the function $f_{C_0,C_1}(b,x) = C_b(x)$ is a distributional k-multi-collision[17] resistant hash function, for $k = O\left(\frac{\log n}{(\alpha-\beta)^2}\right)$. That is, for a

[16] Namely, that for any $(1/\mathsf{poly})$-separated (α, β), the existence of efficiently sampleable distributions whose statistical distance is α but no efficient algorithm can distinguish between them with advantage more than β, implies the existence of one-way functions.

[17] Multi-collision hash functions, recently considered in several works [KNY17, KNY18, BKP18, BDRV18], are hash functions for which it is hard to find multiple inputs that all hash to the same output.

random output y of f, it is difficult to find k random preimages of y. This is because access to such a set of k random pre-images of random y_i's is all we use the inverter A for in the above reduction, and it could handily be replaced with a k-distributional multi-collision finder.

Remark 2.1 (Comparison to [NR06])**.** *Naor and Rothblum's proof implicitly attempts to approximate the maximal likelihood bit of y; that is, the bit b_{ml} such that $\Pr[b = b_{ml}|y] > \Pr[b = 1 - b_{ml}|y]$ (breaking ties arbitrarily). Indeed, the maximal likelihood bit, as shown by [SV03], is closely related to the statistical distance:*

$$\Pr[b = b_{ml}] = \frac{1 + \mathrm{SD}(C_0, C_1)}{2}. \tag{2.5}$$

To approximate b_{ml}, [NR06] make, like us, many calls to $\mathsf{A}(y)$, and take the majority of the answered bits. The idea is that when the statistical distance is large, the majority is likely to be b_{ml}, and when the statistical distance is small, the majority is equally likely to be b_{ml} or $1 - b_{ml}$.

To formally prove this intuition, it must hold that if $\mathrm{SD}(C_0, C_1)$ is large, then $\Pr[b = b_{ml}|y] - \Pr[b = 1 - b_{ml}|y]$ is sufficiently large; putting in our terminology and using Eq. (2.4), if $\mathbb{E}_y[|\theta_y|]$ is sufficiently large, then $|\theta_y|$ should be large for a random y (and the opposite should hold if $\mathrm{SD}(C_0, C_1)$ is small). While these statements are true, in order to prove them, [NR06]'s analysis involves some work which results in a more complicated analysis.

We manage to avoid such complications by using the imbalance θ_y and its characterization of statistical distance (Eq. (2.4)). Furthermore, [NR06]'s approach only attempts to distinguish between the cases when $\mathrm{SD}(C_0, C_1)$ is high or low, while our approach generalizes to approximate $\mathrm{SD}(C_0, C_1)$. Lastly, Naor and Rothblum do not construct one-way functions based on the average-case hardness of $\mathrm{SDP}^{\alpha,\beta}$ with any noticeable gap as we do. Using their technique to do so seems to require additional work—work that our analysis significantly simplifies.

2.2 Interactive Proof for Statistical Distance Approximation

We proceed to describe a constant-round public-coin protocol in which a computationally unbounded prover convinces a computationally bounded verifier that the statistical difference of a given pair of circuits is what the prover claims it to be, up to any inverse polynomial (additive) error. Such a protocol simultaneously establishes the inclusion of $\mathrm{SDP}^{\alpha,\beta}$ in both AM and coAM for any $\alpha > \beta + 1/\mathsf{poly}$.

Our starting point is the algorithm we described above that used a one-way function inverter to estimate the statistical distance. Specifically, that algorithm used the inverter to estimate θ_y for random y's, and then applied Eq. (2.4). We would like to use the prover, instead of the inverter, to achieve the same task.

In our protocol, the verifier draws polynomially many y's and sends them to the prover. The prover responds with values $\widehat{\theta}_i$'s, which it claims are the genuine θ_{y_i}'s. But how can the verifier trust that the prover sent the correct values? In the reduction in Section 2.1, we used k many samples of b conditioned on y to estimate b's true distribution. A standard concentration bound shows that as k grows, the number of ones out of $b_1, \ldots, b_k$, all sampled from $(b|y)$, is very close to $\Pr[b = 1|y] \cdot k$. Similarly, the number of zeros is very close to $\Pr[b = 0|y] \cdot k$. Consider the following *typical set* for any fixed y and arbitrary value θ:

$$\mathcal{T}_y^{k,\theta} = \left\{ (b_1, x_1, b_2, x_2, \ldots, b_k, x_k) \;\middle|\; C_{b_i}(x_i) = y \text{ for all } i, \text{ and } \frac{\sum_{i=1}^k b_i - \sum_{i=1}^k (1 - b_i)}{k} \approx \theta \right\}.$$

Namely, $\mathcal{T}_y^{k,\theta}$ contains every k-tuple of (b_i, x_i) such that all map to y, and each tuple can be used to estimate θ well—the difference between the number of ones and the number of zeros, normalized by k, is close to θ. Also consider the *pre-image* set of y: $\mathcal{I}_y = \{(b,x) \mid C_b(x) = y\}$. Since as k grows the estimation of θ_y improves, we expect that $\mathcal{T}_y^{k,\theta_y}$—the typical set of y with the value θ_y—to contain almost all tuples. Indeed, standard concentration bounds show that

$$\frac{\left|\mathcal{T}_y^{k,\theta_y}\right|}{|\mathcal{I}_y|^k} \geq 1 - e^{-\Omega(k)}. \tag{2.6}$$

On the other hand, the sets $\mathcal{T}_y^{k,\theta'}$, corresponding to values θ' that are far from θ_y, should be almost empty. Indeed, if $|\theta' - \theta_y| \geq \Omega(1)$, then,

$$\frac{\left|\mathcal{T}_y^{k,\theta'}\right|}{|\mathcal{I}_y|^k} \leq e^{-\Omega(k)}. \tag{2.7}$$

So, for the verifier to be convinced that the value $\widehat{\theta}$ sent by the prover is close to θ_y, the prover can prove that the typical set $\mathcal{T}_y^{k,\widehat{\theta}}$ is large. To do so, the parties will use the public-coin constant round protocol for set lower-bound of [GS89], which enables the prover to assert statements of the form "the size of the set $\mathcal{S}$ is at least s".

However, there is still one hurdle to overcome. The typical set $\mathcal{T}_y^{k,\theta_y}$ is only large *relative* to $|\mathcal{I}_y|^k$. Since we do not known how to compute $|\mathcal{I}_y|$ it is unclear what should be the size s that we run the set lower-bound protocol with. Our approach for bypassing this issue is as follows. First observe that the *expected value*, over a random y, of the logarithm of the size of $\mathcal{I}_y$ is the entropy[18] of (b,x) given y. Namely,

$$\mathop{\mathbb{E}}_{y}[\log|\mathcal{I}_y|] = \mathrm{H}(B, X|Y), \tag{2.8}$$

where the jointly distributed random variables (B, X, Y) take the values of randomly drawn (x, b, y). Thus, if we draw t independent elements $y_1, \ldots, y_t$, the average of $\log|\mathcal{I}_y|$ gets closer to $t \cdot \mathrm{H}(B, X|Y)$, as t grows. Specifically,

$$\Pr\left[\prod_{i=1}^{t} |\mathcal{I}_{y_i}| \approx 2^{t \cdot \mathrm{H}(B,X|Y)}\right] \geq 1 - e^{-\Omega(t/n^2)}, \tag{2.9}$$

where n denotes the output length of the given circuits. For large enough t, we can thus assume that the size of this product set is approximately $2^{t \cdot H(B,X|Y)}$, and run the set lower bound protocol for all the y_i's together. That is, we ask the prover to send t estimates $(\widehat{\theta}_1, \ldots, \widehat{\theta}_t)$ for the values $(\theta_{y_1}, \ldots, \theta_{y_t})$, and prove that the size of the product set $\mathcal{T}_{y_1}^{k,\widehat{\theta}_1} \times \cdots \times \mathcal{T}_{y_1}^{k,\widehat{\theta}_1}$ is almost $2^{t \cdot \mathrm{H}(B,X|Y)}$.

So far we have reduced knowing the size of $\mathcal{I}_y$ to knowing $\mathrm{H}(B, X|Y)$, but again it seems difficult for the verifier to compute this quantity on its own. Actually, standard entropy manipulations show that

$$\mathrm{H}(B, X|Y) = (m+1) - \mathrm{H}(Y),$$

where m denotes the input length of the given circuits. It thus suffices to approximate $\mathrm{H}(Y)$. Recall that y is the output of the circuit that maps (x, b) to $C_b(x)$, so Y is drawn according to an

[18] Recall that the entropy of a random variable X over $\mathcal{X}$ is defined as $(\mathrm{H}(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log(1/\Pr[X = x])$. The conditional entropy of X given Y is $\mathrm{H}(X|Y) = \mathbb{E}_{y \sim Y}[\mathrm{H}(X|Y = y)]$.

output distribution of a known circuit. Luckily, Goldreich, Sahai and Vadhan [GSV99] showed that approximating the output entropy of a given circuit is in NISZK, and thus has a constant-round public-coin protocol (since $\mathsf{NISZK} \subseteq \mathsf{AM} \cap \mathsf{coAM}$).

To conclude, we describe the entirety of our protocol, which proves Theorem 1.9.

Protocol to approximate $\mathrm{SD}(C_0, C_1)$, given the circuits (C_0, C_1) as input:

1. First, the prover sends the verifier a claim $\widehat{\mathrm{H}}$ of the value of $\mathrm{H}(Y)$.
2. The parties execute [GSV99]'s protocol to convince the verifier that this claim—that $\widehat{\mathrm{H}} \approx \mathrm{H}(Y)$—is correct.
3. The verifier uses $\widehat{\mathrm{H}}$ to compute $\widehat{\mathrm{H}}(B, X|Y)$ as $((m+1) - \widehat{\mathrm{H}})$.
4. The verifier samples $y_1, \ldots, y_t$ from $\frac{C_0+C_1}{2}$ and sends them to the prover.
5. The prover responds with $\widehat{\theta}_1, \ldots, \widehat{\theta}_t$ as claims for the values $\theta_{y_1}, \ldots, \theta_{y_t}$.
6. The parties run a set lower-bound protocol to prove that the set $\mathcal{T}_{y_1}^{\widehat{\theta}_1,k} \times \cdots \times \mathcal{T}_{y_t}^{\widehat{\theta}_t,k}$ is almost as large as $(\mathcal{I}_{y_1} \times \cdots \times \mathcal{I}_{y_t})^k$.
   - Here, they use $2^{tk\widehat{\mathrm{H}}(B,X|Y)}$ as a proxy for $(|\mathcal{I}_{y_1}| \cdot \cdots \cdot |\mathcal{I}_{y_t}|)^k$.
7. If the verifier has not rejected so far, it outputs $\frac{1}{t}\sum_{i=1}^{t} |\widehat{\theta}_i|$.

2.3 TDP and JSP are SZK-Complete

We show that both $\mathrm{TDP}^{\alpha,\beta}$ and $\mathrm{JSP}^{\alpha,\beta}$ with $\alpha > \beta + 1/\mathsf{poly}$ are SZK-complete. Since the proof of the former uses that of the latter we start by giving an outline that $\mathrm{JSP}^{\alpha,\beta}$ is SZK-complete.

2.3.1 Jensen-Shannon Divergence Problem is SZK-complete

We need to show that $\mathrm{JSP}^{\alpha,\beta}$ with $\alpha > \beta + 1/\mathsf{poly}$ is both in SZK and SZK-hard. In both parts we use the following characterization of the Jensen-Shannon divergence, which follows from its definition. Given a pair of circuits C_0 and C_1, consider the jointly distributed random variables (B, X, Y), where B is a uniformly random bit, X is a uniformly random string and $Y = C_B(X)$. Then, it follows from some elementary manipulations (see Proposition 4.1 below) that:

$$\mathrm{JS}(C_0, C_1) = 1 - \mathrm{H}(B|Y). \tag{2.10}$$

We use this characterization to tie Jensen-Shannon Divergence Problem to another SZK-complete problem—Entropy Difference Problem (EDP) with a gap function g. The input to EDP^g is also a pair of circuits C_0 and C_1. YES instances are those in which the entropy gap $\mathrm{H}(C_0) - \mathrm{H}(C_1)$ is at least $g(n)$ (where n is the output length of the circuits) and NO instances are those in which the gap is at most $-g(n)$. Goldreich and Vadhan [GV99] showed that EDP^g is SZK-complete for any noticeable function g. Our proof that $\mathrm{JSP}^{\alpha,\beta}$ is SZK-complete closely follows the reduction from the reverse problem of SDP (i.e., in which YES instances are distributions that are statistically *close*) to EDP [Vad99, Section 4.4].

$\mathrm{JSP}^{\alpha,\beta}$ is in SZK: We reduce $\mathrm{JSP}^{\alpha,\beta}$ to $\mathrm{ED}^{(\alpha-\beta)/2}$. Given C_0 and C_1, the reduction outputs a pair of circuits D_0 and D_1 such that D_1 outputs a sample from (B, Y) and D_0 outputs a sample from (B', Y), where B' is an independent random bit with $\mathrm{H}(B) = 1 - \frac{\alpha+\beta}{2}$. The chain rule

for entropy[19] implies that

$$\mathrm{H}(D_0) - \mathrm{H}(D_1) = 1 - \frac{\alpha+\beta}{2} - \mathrm{H}(B|Y) = \mathrm{JS}(C_0, C_1) - \frac{\alpha+\beta}{2},$$

where the second equality follows from Eq. (2.10). Thus, if $\mathrm{JS}(C_0, C_1) \geq \alpha$, then $\mathrm{H}(D_0) - \mathrm{H}(D_1) \geq \frac{\alpha-\beta}{2}$; and if $\mathrm{JS}(C_0, C_1) \leq \beta$, then $\mathrm{H}(D_0) - \mathrm{H}(D_1) \leq -\frac{\alpha-\beta}{2}$. And since $\mathrm{ED}^{(\alpha-\beta)/2} \in \mathsf{SZK}$, we get that $\mathrm{JSP}^{\alpha,\beta} \in \mathsf{SZK}$.

$\mathrm{JSP}^{\alpha,\beta}$ is SZK-hard: We reduce $\mathrm{SDP}^{1-2^{-k},2^{-k}}$ to $\mathrm{JSP}^{\alpha,\beta}$, for some large enough k. This suffices since $\mathrm{SDP}^{1-2^{-k},2^{-k}}$ is known to be SZK-hard [SV03].[20] In the presentation of related results in his thesis, Vadhan relates the statistical distance of the circuits to the entropy of B given Y [Vad99, Claim 4.4.2]. For example, if $\mathrm{SD}(C_0, C_1) = 0$ (i.e., the distributions are identical), then $B|Y$ is a uniformly random bit, and so $\mathrm{H}(B|Y) = 1$; and if $\mathrm{SD}(C_0, C_1) = 1$ (i.e., the distributions are disjoint), then B is completely determined by Y, and so $\mathrm{H}(B|Y) = 0$. More generally, Vadhan showed that if $\mathrm{SD}(C_0, C_1) = \delta$, then[21]

$$1 - \delta \leq \mathrm{H}(B|Y) \leq h\left(\frac{1+\delta}{2}\right). \tag{2.11}$$

By taking k to be large enough (as a function of α and β), and applying Eqs. (2.10) and (2.11), we have that if $\mathrm{SD}(C_0, C_1) \geq 1 - 2^{-k}$, then $\mathrm{JS}(C_0, C_1) \geq \alpha$; and if $\mathrm{SD}(C_0, C_1) \leq 2^{-k}$, then $\mathrm{JS}(C_0, C_1) \leq \beta$. Thus, the desired reduction is simply the identity function that outputs the input circuits.

2.3.2 Triangular Discrimination Problem is SZK-complete.

We need to show that $\mathrm{TDP}^{\alpha,\beta}$ with $\alpha > \beta + 1/\mathsf{poly}$ is both in SZK and SZK-hard. Showing the latter is very similar to showing that $\mathrm{JSP}^{\alpha,\beta}$ is SZK-hard, but using Eq. (1.1) to relate the triangular discrimination to statistical distance (instead of Eq. (2.11) that relates the Jensen-Shannon divergence to statistical distance). We leave the formal details to the body of this paper and focus here on showing that $\mathrm{TDP}^{\alpha,\beta}$ is in SZK.

A natural approach to show that $\mathrm{TDP}^{\alpha,\beta}$ is in SZK is to follow Sahai and Vadhan's proof that $\mathrm{SDP}^{2/3,1/3}$ is in SZK. Specifically, a main ingredient in that proof is to polarize the statistical distance of the circuits (to reduce the simulation error). Indeed, if we can reduce $\mathrm{TDP}^{\alpha,\beta}$ to, say, $\mathrm{TDP}^{0.9,0.1}$ by polarizing the triangular discrimination, then Eq. (1.1) would imply that we also reduce $\mathrm{TDP}^{\alpha,\beta}$ to $\mathrm{SDP}^{2/3,1/3}$, which we know is in SZK.

We are indeed able to show such a polarization lemma for triangular discrimination (using similar techniques to [SV03]'s polarization lemma). However, this lemma only works when the gap between α and β is roughly $1/\log$. Actually, the polarization lemma of [SV03] also suffers the same limitation with respect to the gap between α^2 and β.

Still, we would like to handle also the case that the gap between α and β is only $1/\mathsf{poly}$. To do so we take a slightly different approach. Specifically, we reduce $\mathrm{TDP}^{\alpha,\beta}$ to $\mathrm{JSP}^{\alpha',\beta'}$, where α' and β' are also noticeably separated.

[19] For a jointly distributed random variables X and Y, it holds that $\mathrm{H}(X, Y) = \mathrm{H}(X) + \mathrm{H}(Y|X)$.

[20] For the simplicity of presentation, we are ignoring subtle details about the relation of k to the output length of the circuits. See Section 4.1 for the formal proof.

[21] The function h is the binary entropy function. That is, $h(p) = -p\log(p) - (1-p)\log(1-p)$ is the entropy of a Bernoulli random variable with parameter p.

An important step toward showing this reduction is to characterize the triangular discrimination and the Jensen-Shannon divergence via the imbalance θ_y (see Eq. (2.3)), as we already did for statistical distance. Recall that given $Y = y$, the random variable B takes the value 1 with probability $\frac{1+\theta_y}{2}$, and 0 otherwise. Hence, Eq. (2.10) can also be written as

$$\mathrm{JS}(C_0, C_1) = 1 - \mathop{\mathbb{E}}_{y\sim Y}\left[h\left(\frac{1+\theta_y}{2}\right)\right]. \tag{2.12}$$

As for the triangular discrimination, it follows from the definition that

$$\mathrm{TD}(C_0, C_1) = \mathop{\mathbb{E}}_{y\sim Y}\left[\theta_y^2\right]. \tag{2.13}$$

Furthermore, by Taylor approximation, for small values of θ, it holds that

$$h\left(\frac{1+\theta}{2}\right) \approx 1 - \theta^2. \tag{2.14}$$

As we can see, the above equations imply that if all the θ_y's were small, a gap in the triangular discrimination would also imply a gap in the Jensen-Shannon divergence. Thus, we would like an operation that reduces all the θ_y.

The main technical tool we use to reduce θ_y is to consider the *convex combination* of the two input circuits. Given a pair of circuits C_0 and C_1, consider the pair of circuits D_0 and D_1 such that $D_b = \lambda \cdot C_b + (1-\lambda) \cdot \frac{C_0+C_1}{2}$.[22] Let Q_b denote the output distribution of D_b, and recall that P_b denotes the output distribution of C_b. We also let θ'_y be defined similarly to θ_y, but with respect to D_0 and D_1 (rather than C_0 and C_1). Using this notation, we have that $\theta_y = \frac{P_1(y)-P_0(y)}{P_1(y)+P_0(y)}$, and it may be seen that

$$\theta'_y = \frac{Q_1(y) - Q_0(y)}{Q_1(y) + Q_0(y)} = \lambda \cdot \theta_y. \tag{2.15}$$

So, our reduction chooses a sufficiently small λ, and outputs the circuits D_0 and D_1. Some care is needed when choosing λ. Eqs. (2.13) and (2.15) yield that $\mathrm{TD}(D_0, D_1) = \lambda^2 \cdot \mathrm{TD}(C_0, C_1)$. Hence, the convex combination also shrinks the gap in triangular discrimination. We show that by choosing $\lambda \approx \sqrt{\alpha - \beta}$, the approximation error in Eq. (2.14) is smaller than the aforementioned shrinkage, and the reduction goes through. The resulting gap in the Jensen-Shannon divergence is roughly $(\alpha - \beta)^2$, which is noticeable by the assumption that $\alpha > \beta + 1/\mathsf{poly}$.

This shows that $\mathrm{TDP}^{\alpha,\beta}$ is in SZK if $\alpha > \beta + 1/\mathsf{poly}$. By the relationship between TD and SD (Eq. (1.1)), this implies that $\mathrm{SDP}^{\alpha,\beta}$ is in SZK if $\alpha^2 > \beta + 1/\mathsf{poly}$. This, in turn, by the SZK-hardness of $\mathrm{SDP}^{2/3,1/3}$ and the known polarization lemma that applies for the same, implies polarization for statistical distance for any (α, β) such that $\alpha^2 > \beta + 1/\mathsf{poly}$.

2.3.3 Reflections and an Open Problem

Many f-divergences of interest can be expressed as an expectation, over $y \sim Y$, of a simple function of θ_y. That is, an expression of the form $\mathbb{E}_{y\sim Y}[g(\theta_y)]$, for some function $g : [-1, 1] \to [0, 1]$. For example:

- $\mathrm{SD}(C_0, C_1) = \mathbb{E}_{y\sim Y}[|\theta_y|]$ (i.e., $g(z) = |z|$, see Eq. (2.4));

[22] This definition of convex combination is more convenient to analyze than perhaps the more natural definition of $D_b = \lambda \cdot C_b + (1-\lambda) \cdot C_{1-b}$.

- $\mathrm{TD}(C_0, C_1) = \mathbb{E}_{y\sim Y}\left[\theta_y^2\right]$ (i.e., $g(z) = z^2$, see Eq. (2.13)); and
- $\mathrm{JS}(C_0, C_1) = \mathbb{E}_{y\sim Y}\left[1 - h\left(\frac{1+\theta_y}{2}\right)\right]$ (i.e., $g(z) = 1 - h\left(\frac{1+z}{2}\right)$, see Eq. (2.12)).

To reduce TDP to JSP, we took a convex combination of the two circuits and used the fact that $1 - h\left(\frac{1+\theta_y}{2}\right) \approx O(\theta_y^2)$ for small values of θ_y. While this worked for polarization of TD (which corresponds to $g(z) = z^2$), it seems unlikely to yield a polarization lemma for SD for an arbitrarily small (but noticeable) gap. The reason is that the function $g(z) = |z|$ — the g-function corresponding to SD — in not differentiable at 0 and in particular does not act like z^2 for small values of z. As we find this similarity between the different notions of distance striking, and indeed our proofs leverage the relations between them, we provide in Fig. 2.1 a plot comparing the different choices for the function g.

Another popular f-divergence that we have not discussed thusfar[23] is the *squared Hellinger distance*, defined as $\mathrm{H}^2(P, Q) = \frac{1}{2}\sum_y\left(\sqrt{P_y} - \sqrt{Q_y}\right)^2$. It can be shown that $\mathrm{H}^2(C_0, C_1) = \mathbb{E}_{y\sim Y}\left[1 - \sqrt{1-\theta_y^2}\right]$, and so also this distance falls within the above framework (i.e., by considering $g(z) = 1 - \sqrt{1-z^2}$).

Interestingly, the squared Hellinger distance also acts like JS (and TD) around 0; namely, $1 - \sqrt{1-\theta_y^2} \approx O(\theta_y^2)$ for small values of θ_y. However, unlike $\mathrm{TDP}^{\alpha,\beta}$, we do not know how to show that the Hellinger Difference Problem, denoted $\mathrm{HDP}^{\alpha,\beta}$ and defined analogously to $\mathrm{TDP}^{\alpha,\beta}$ (while replacing the distance TD with H^2), is in SZK for all $(1/\mathsf{poly})$-separated (α, β). We do mention that $\mathrm{H}^2(P, Q) \le \mathrm{TD}(P, Q) \le 2\,\mathrm{H}^2(P, Q)$, and thus $\mathrm{HDP}^{\alpha,\beta}$ is in SZK if α and $\beta/2$ are $(1/\mathsf{poly})$-separated. However, the proof described above does not go through if we try to apply it to the Hellinger distance—we cannot guarantee that the gap in the Hellinger distance after taking the convex combination is larger than the error in the Taylor approximation. Indeed, the question whether $\mathrm{HDP}^{\alpha,\beta}$ is in SZK for any $(1/\mathsf{poly})$-separated (α, β), first raised by Fehr and Vaudenay [FV17], remains open.

3 Preliminaries

We use calligraphic letters to denote sets, uppercase for random variables, lowercase for values and functions, and uppercase sans-serif (e.g., A) for algorithms (i.e., Turing Machines). All logarithms considered here are in base two.

Given a random variable X, we write $x \sim X$ to indicate that x is selected according to the distribution of X. Similarly, given a finite set $\mathcal{S}$, we let $s \sim \mathcal{S}$ denote that s is selected according to the uniform distribution on $\mathcal{S}$. We denote the probability mass that a distribution P (over a finite set $\mathcal{Y}$) puts on an element $y \in \mathcal{Y}$ by either P_y or $P(y)$, where we choose which notation to use based on readability and context. The support of a distribution P over a finite set $\mathcal{Y}$, denoted $\mathrm{Supp}(P)$, is defined as $\{y \in \mathcal{Y} : P(y) > 0\}$. The support of a discrete random variable X is the support of its probability mass function, that is $\mathrm{Supp}(X) = \mathrm{Supp}(P_X)$, where P_X is the distribution of the random variable X.

The product distribution between two distributions P and Q is denoted by $P \otimes Q$. The product distribution of k independent copies of P is denoted by $P^{\otimes k}$. Given a boolean statement S (e.g., $X \ge 5$), let $\mathbb{1}\{S\}$ be the indicator function that outputs 1 if S is a true statement and 0 otherwise.

[23] Actually we will use the squared Hellinger distance in Section 4.2.2 to analyze triangular discrimination of direct product distributions. Also, the squared Hellinger distance is closely related to the Fidelity distance: $\mathrm{Fidelity}(P, Q) = 1 - \mathrm{H}^2(P, Q)$.

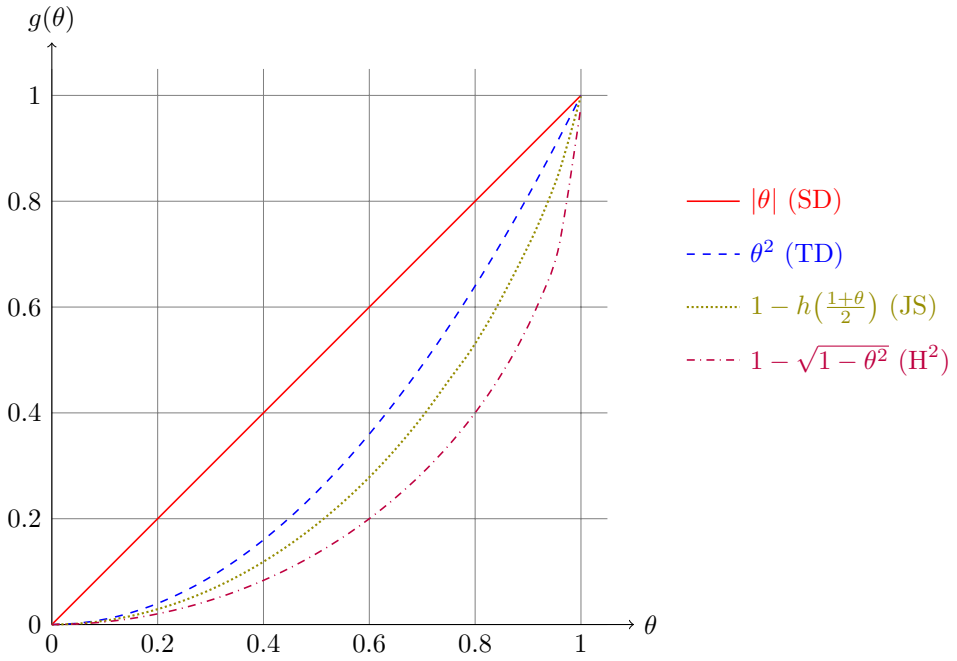


Figure 2.1: Comparison between the difference choices of the function g that were discussed. Since all functions are symmetric around 0, we restrict to the domain $[0, 1]$. Recall that $g_1(\theta) = |\theta|$ corresponds to SD, $g_2(\theta) = \theta^2$ to TD, $g_3(\theta) = 1 - h\left(\frac{1+\theta}{2}\right)$ to JS and $g_4(\theta) = 1 - \sqrt{1-\theta^2}$ to H^2.

We let poly denote the set all polynomials over the integers. Given a probabilistic polynomial-time algorithm A, we denote by $\mathsf{A}(x; r)$ the output of A given input x and randomness r. A function $\nu\colon \mathbb{N} \to [0, 1]$ is $\mathsf{negligible}$, denoted $\nu(n) = \mathsf{negl}(n)$, if $\nu(n) < 1/p(n)$ for every $p \in \mathsf{poly}$ and large enough n.

A function $f\colon \{0,1\}^* \to \{0,1\}^*$ is $\mathsf{efficiently\ computable}$ if there exists a probabilistic polynomial-time algorithm that on input $x \in \{0,1\}^*$ outputs $f(x)$.

3.1 Information Theory Preliminaries

3.1.1 Statistical Distance and Imbalance

The $\mathsf{statistical\ distance}$ between two distributions P and Q over a finite set $\mathcal{Y}$, is defined as $\mathrm{SD}(P, Q) \triangleq \max_{\mathcal{S}\subseteq\mathcal{Y}} P(\mathcal{S}) - Q(\mathcal{S}) = \frac{1}{2}\sum_{y\in\mathcal{Y}} |P_y - Q_y|$. The definition of statistical distance immediately implies the following property.

Proposition 3.1. *Let P be a distribution over a finite set $\mathcal{Y}$ and let U be the uniform distribution over $\mathcal{Y}$. Then* $\mathrm{SD}(P, U) \geq 1 - \frac{|\mathrm{Supp}(P)|}{|\mathcal{Y}|}$.

Proof. Let $\mathcal{S} = \mathrm{Supp}(P)$. Then

$$\mathrm{SD}(P, U) \geq P(\mathcal{S}) - U(\mathcal{S}) = 1 - \frac{|\mathrm{Supp}(P)|}{|\mathcal{Y}|}.$$

□

In this work, we will use the following view of statistical distance.

Definition 3.2 (The imbalance between P and Q)**.** *Let P and Q be two distributions over a finite set $\mathcal{Y}$. Let (B, Y) be the jointly distributed random variables defined as follows: $B \sim \{0,1\}$ and if $B = 1$, then $Y \sim P$ (that is, Y is a random variable drawn according to P), and if $B = 0$, then $Y \sim Q$. For every $y \in \mathrm{Supp}(Y)$ we define the imbalance $\theta_y^{P,Q} = \Pr[B = 1|Y = y] - \Pr[B = 0|Y = y]$.*

We will typically omit the distributions in the superscript from the notation (i.e., write θ_y instead of $\theta_y^{P,Q}$) when they are clear from the context.

Proposition 3.3. *Let P and Q be two distributions as in Definition 3.2. Then, $\mathrm{SD}(P, Q) = \mathbb{E}_{y\sim Y}[|\theta_y|]$.*

Proof. The proof follows from the definitions of θ_y and statistical distance, details follow.

It is easy to verify that for every $y \in \mathrm{Supp}(Y)$, it holds that $\Pr[Y = y] = \frac{P_y+Q_y}{2}$ and $\theta_y = \frac{P_y-Q_y}{P_y+Q_y}$. Thus,

$$\mathop{\mathbb{E}}_{y\sim Y}[|\theta_y|] = \sum_{y\in\mathrm{Supp}(Y)} \Pr[Y = y] \cdot |\theta_y| = \sum_{y\in\mathrm{Supp}(Y)} \frac{P_y + Q_y}{2} \cdot \frac{|P_y - Q_y|}{P_y + Q_y} = \mathrm{SD}(P, Q).$$

□

The following claim is in the heart of [SV03]'s protocol for SDP.

Proposition 3.4. *Let P and Q be two distributions and let B and Y be random variables as in Definition 3.2. Let B' be a random variable that depends only on Y. That is, we have the following Markov chain: $B \to Y \to B'$.*

Then,

$$\Pr\big[B = B'\big] \le \frac{1}{2} + \frac{\mathrm{SD}(P, Q)}{2},$$

where equality holds if $B'(y) = 1$ if $\theta_y \ge 0$ and $B'(y) = 0$ if $\theta_y < 0$ (that is, B' is the maximal likelihood estimator of B).

Proof. Let $\theta'_y = \Pr[B' = 1|Y = y] - \Pr[B' = 0|Y = y]$.

$$\begin{aligned}
\Pr\big[B = B'\big] &= \mathop{\mathbb{E}}_{y\sim Y}\Big[\Pr\big[B = B'|Y = y\big]\Big] \\
&= \mathop{\mathbb{E}}_{y\sim Y}\Big[\Pr[B = 1|Y = y] \cdot \Pr\big[B' = 1|Y = y\big] + \Pr[B = 0|Y = y] \cdot \Pr\big[B' = 0|Y = y\big]\Big] \\
&= \mathop{\mathbb{E}}_{y\sim Y}\left[\left(\frac{1+\theta_y}{2}\right) \cdot \left(\frac{1+\theta'_y}{2}\right) + \left(\frac{1-\theta_y}{2}\right) \cdot \left(\frac{1-\theta'_y}{2}\right)\right] \\
&= \frac{1}{2} + \frac{\mathbb{E}_{y\sim Y}\big[\theta_y \cdot \theta'_y\big]}{2} \\
&\le \frac{1}{2} + \frac{\mathbb{E}_{y\sim Y}[|\theta_y|]}{2} \\
&= \frac{1}{2} + \frac{\mathrm{SD}(P, Q)}{2},
\end{aligned}$$

where the second equlity follows since condition on $Y = y$, the random variables B and B are independent, and the inequality follows since $\theta'_y \in [-1, 1]$. Moreover, if $\theta'_y = \mathrm{sign}(\theta_y)$ then the inequality is actually an equality. B' satisfying the latter condition is exactly the maximal likelihood estimator for B. □

3.1.2 Entropy

Definition 3.5 (Entropy)**.** *The* entropy *of a discrete random variable X is defined as*

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

The binary entropy function $h\colon [0,1]\to[0,1]$ *is defined to be the entropy of $X\sim$ Bernoulli(p). That is, $h(p) = -p\log(p) - (1-p)\log(1-p)$, where we use the convention that $h(0) = h(1) = 0$.*

Definition 3.6 (Conditional entropy)**.** *Let X, Y be jointly distributed random variables. The* conditional entropy *of X given Y is defined as*

$$\mathrm{H}(X|Y) = \mathop{\mathbb{E}}_{y\sim Y}[\mathrm{H}(X|Y=y)] = \mathop{\mathbb{E}}_{(x,y)\sim(X,Y)}\left[\log\left(\frac{1}{\Pr[X=x|Y=y]}\right)\right].$$

Fact 3.7 (Chain rule for entropy)**.** *Let X, Y be jointly distributed random variables. Then*

$$\mathrm{H}(X,Y) = \mathrm{H}(X|Y) + \mathrm{H}(Y).$$

3.1.3 Concentration Bounds

We use the following well-known concentration bound.

Fact 3.8 (Chernoff-Hoeffding bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $[a,b]$ and let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for every $\varepsilon > 0$ it holds that*

$$\Pr\left[\overline{X} - \mathbb{E}\left[\overline{X}\right] \geq \varepsilon\right] \leq e^{-2n\varepsilon^2/(b-a)^2}$$
$$\Pr\left[\overline{X} - \mathbb{E}\left[\overline{X}\right] \leq -\varepsilon\right] \leq e^{-2n\varepsilon^2/(b-a)^2}.$$

We also use the following fact, showing that computing the empirical distribution from large enough number of samples approximates the original distribution well.

Fact 3.9 (Folklore (see, e.g., [Gol17, Exercise 11.4]))**.** *Let P be a distribution over n elements and let $\hat{P}$ be the empirical distribution obtained from taking N samples $P_1, \ldots, P_N$ from P, namely, $\hat{P}(i) = |\{j\colon P_j = i\}|/N$. Then, if $N \geq \left\lceil \frac{n+\log(1/\delta)}{2\varepsilon^2}\right\rceil$, it holds that*

$$\Pr\left[\mathrm{SD}(P,\hat{P}) \geq \varepsilon\right] \leq \delta.$$

3.2 Statistical Zero-Knowledge Interactive Proofs

In this section we give the standard definitions for statistical zero-knowledge proofs and recall classical results regarding such proofs. We follow [Vad99].

Definition 3.10 (View of interactive protocol)**.** *Let (P,V) be an r-message interactive protocol. The* view *of V on a common input x is defined by $\mathsf{view}_{\mathsf{P},\mathsf{V}}(x) \triangleq (m_1, m_2, \ldots, m_r; \rho)$, where $m_1, m_2, \ldots, m_r$ are the messages sent by the parties in a random execution of the protocol, and ρ contains of all the random coins V used during this execution.*

We allow cheating verifiers to be non-uniform by giving them an auxiliary input. For an algorithm A and a string $z \in \{0,1\}^*$ (all auxiliary inputs will be binary strings, regardless of the properties' alphabet), let $\mathsf{A}_{[z]}$ be A when z was given as auxiliary input. Following [Vad99], we adopt the convention that the running time of A is independent of z, so if z is too long, A will not be able to access it in its entirety. We also allow probabilistic algorithms to fail by outputting $\perp$. An algorithm A is useful if $\Pr[\mathsf{A}(x) = \perp] \leq 1/2$ for every x, and let $\widetilde{\mathsf{A}}(x)$ denote the output distribution of $\mathsf{A}(x)$, conditioning on $\mathsf{A}(x) \neq \perp$.

Definition 3.11 (Statistical zero-knowledge interactive proofs)**.** *Let $\Pi = (\text{YES}, \text{NO})$ be a promise problem. A* statistical zero-knowledge interactive proof system *for Π is an interactive protocol (P, V) if the following holds:*

- **Efficiency***: The verifier V is a probabilistic polynomial-time algorithm.*
- **Completeness***: If $x \in \text{YES}$, then, when $\mathsf{V}(x)$ interacts with $\mathsf{P}(x)$, with probability $\frac{2}{3}$ it accepts.*
- **Soundness***: If $x \in \text{NO}$, then for every prover strategy P^*, when $\mathsf{V}(x)$ interacts with P^*, with probability $2/3$ it rejects.*
- **Zero-Knowledge***: For every probabilistic polynomial-time V^*, there exists a useful probabilistic polynomial-time S and a negligible function ν such that for all $x \in \text{YES}$ and $z \in \{0,1\}^*$, it holds that*

$$\text{SD}\left(\widetilde{\mathsf{S}}_{[z]}(x), \mathsf{view}_{\mathsf{P},\mathsf{V}^*_{[z]}}(x)\right) \leq \nu(|x|).$$

SZK *denotes the class of promise problems possessing statistical zero-knowledge interactive proof system.*

3.2.1 Complete Problems

Central in the study of statistical zero-knowledge are problems dealing with properties of distributions encoded by circuits.

Definition 3.12 (Distributions encoded by circuits)**.** *Let C be a Boolean circuit with m input gates and n output gates. The* distribution encoded by *C is the distribution induced on $\{0,1\}^n$ by evaluating C on a uniformly selected string from $\{0,1\}^m$. By abuse of notation, we also write C for the distribution defined by the circuit C.*

Two particularly interesting problems are the Statistical Difference Problem (see Definition 1.1) and the Entropy Difference Problem.

Definition 3.13 (Entropy Difference Problem)**.** *Let $g\colon \mathbb{N} \to \mathbb{R}^+$. The* Entropy Difference Problem with promise *g, denoted by EDP^g, is given by the sets*

$$\text{EDP}^g_Y = \{(C_0, C_1) \mid \text{H}(C_0) \geq \text{H}(C_1) + g(n)\},$$
$$\text{EDP}^g_N = \{(C_0, C_1) \mid \text{H}(C_1) \geq \text{H}(C_0) + g(n)\}.$$

where n is the output length of the circuits C_0 and C_1.

Both SDP and EDP are known to be complete for SZK, though for different setting of parameters.

Theorem 3.14 ([SV03, GSV98]). *Let α, β be such that (α^2, β) are $(1/\mathsf{poly})$-separated functions and that there exists a constant $\varepsilon \in (0, 1/2)$ with $2^{-n^{1/2-\varepsilon}} \leq \beta(n)$ and $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$ for every $n \in \mathbb{N}$. Then, the promise problem $\mathrm{SDP}^{\alpha,\beta}$ is SZK complete. Furthermore, the problem remains complete if it is restricted to length-preserving pair of circuits.*

Remark 3.15. *The restrictions placed on α and β in Theorem 3.14—that they are not closer than $2^{-n^{1/2}}$ to 1 or 0—are a result of the extent to which the polarization lemma of [SV03] can polarize. While, for some small constant c, it might be possible to push this to allow them to be 2^{-cn}-close to 1 or 0 using a more efficient polarization technique (see, for instance, the proof of Theorem 12 in [AARV17]), there are reasons to believe that this restriction cannot be weakened much more. For instance, if $\alpha = 1$, then $\mathrm{SDP}^{\alpha,\beta}$ is in PZK, and is thus unlikely to be complete for SZK (indicated by a known oracle separation between these classes [BCH+17]). Similarly, if $\alpha/\beta \geq 2^{n/2}$, then $\mathrm{SDP}^{\alpha,\beta}$ is contained in the class PP, which SZK is again oracle-separated from [BCH+17].*

With regard to the separation between α and β, in the extreme case that $|\alpha(n) - \beta(n)| \leq 2^{-n}$, we note that $\mathrm{SDP}^{\alpha,\beta}$ can be shown to be NP-hard by a reduction from Circuit-SAT, and thus is not contained in SZK unless the polynomial hierarchy collapses [BHZ87].

Since the furthermore part of Theorem 3.14 was not stated in prior works (but will be used by us for technical reasons), we include a proof sketch.

Proof Sketch of Theorem 3.14. While this theorem is not stated in this generality in [SV03], it extends to this by setting some parameters appropriately. We show how to do this using the proof of the equivalent theorem in [Vad99]. The approach is to first establish the SZK-hardness of $\mathrm{SDP}^{2/3,1/3}$, and then use the polarization lemma [Vad99, Lemma 3.1.12] to reduce this to $\mathrm{SDP}^{1-2^{-k},2^{-k}}$, to show that the former is in SZK. Here we sketch how, for every $\alpha > \beta$ such that (α^2, β) are $(1/\mathsf{poly})$-separated and any $\varepsilon \in (0, 1/2)$, the problem $\mathrm{SDP}^{\alpha,\beta}$ is reduced—via the polarization lemma—to $\mathrm{SDP}^{1-2^{-k},2^{-k}}$, for $k = n^{1/2-\varepsilon}$ and n being the input and output length of the circuit (since the furthermore clause requires us to produce length-preserving circuits).

The polarization lemma, for some α and β and an integer k, gives a way to take a pair of circuits (C_0, C_1) and efficiently produce another pair (D_0, D_1) such that if $\mathrm{SD}(C_0, C_1) \geq \alpha$, then $\mathrm{SD}(D_0, D_1) \geq 1 - 2^{-k}$, and if $\mathrm{SD}(C_0, C_1) \leq \beta$, then $\mathrm{SD}(D_0, D_1) \leq 2^{-k}$. While it can do this for any k, the output lengths of D_0 and D_1 grow as k grows, and we are interested in k as a function of these output lengths. In particular, we wish to show that, for any constant $\varepsilon \in (0, 1/2)$, we can end up with an output length n for D_0 and D_1 such that $k \geq n^{1/2-\varepsilon}$. Actually, since we want the resulting circuits to be length-preserving, we analyze how the polarization lemma also affects the input length.

The proof of [Vad99, Lemma 3.1.12] makes use of the following two transformations on pairs of circuits:

- ℓ-fold Repetition: the input circuits (C_0, C_1) are mapped to (D_0, D_1), where D_b is the concatenation of ℓ copies of C_b. If m and n are the input and output lengths of C_0 and C_1, then $m \cdot \ell$ and $n \cdot \ell$ are the input and output lengths of D_0 and D_1.
- r-fold XOR: the input circuits (C_0, C_1) are mapped to (D_0, D_1), where D_b first samples $(b_1, \ldots, b_r) \sim \{0,1\}^r$ conditioned on $b_1 \oplus \cdots \oplus b_r = b$ and outputs the concatenation of $C_{b_1}, \ldots, C_{b_r}$. Notice that in order to sample the desired $b_1, \ldots, b_r$, it suffices to have r uniformly random bits: if $b_1 \oplus \cdots \oplus b_r = b$, then use these bits; otherwise, flip the first bit b_1. The resulting sampled bits satisfy the desired distribution. Hence, if m and n are the input and output lengths of C_0 and C_1, then $(m \cdot r + r)$ and $n \cdot r$ are the respective input and output lengths of D_0 and D_1.

Initially, the polarization lemma sets $\lambda = \min(\alpha^2/\beta, 2)$ and $\ell = \lceil \log_\lambda 4k \rceil$, and proceeds as follows:

1. Repeat (C_0, C_1) for ℓ times to get circuits $C_0^{(1)}$ and $C_1^{(1)}$.
2. XOR $(C_0^{(1)}, C_1^{(1)})$ for $r = \lambda^\ell/(2\alpha^{2\ell})$ times to get circuits $C_0^{(2)}$ and $C_1^{(2)}$.
3. Repeat $(C_0^{(2)}, C_1^{(2)})$ for k times to get circuits D_0 and D_1.

The fact that D_0 and D_1 are polarized is proved in [SV03]. Here we only focus on showing the dependence of k on the input and output lengths.

Based on the above description, the input lengths of D_0 and D_1 are $m' = \left(\left(m \cdot \ell \cdot \frac{\lambda^\ell}{2\alpha^{2\ell}} + \frac{\lambda^\ell}{2\alpha^{2\ell}}\right) \cdot k\right)$ and their output length are $n' = \left(\ell \cdot \frac{\lambda^\ell}{2\alpha^{2\ell}} \cdot k \cdot n\right)$, where m (resp., n) is the input (resp., output) length of the input circuits C_0 and C_1. Note that if the input circuits are length-preserving, then the input of the resulting circuits is longer than their output.

We follow the proof of [CCKV08, Lemma 38] and note that the above setting guarantees the following. First, note that $\ell = O\left(\frac{\ln k}{\ln \lambda}\right)$. Moreover, since $\lambda \in (1, 2]$, we have that $\ln(\lambda) = \ln(1 + (\lambda - 1)) \geq (\lambda - 1)/2 \geq \Omega\left(\frac{\alpha^2 - \beta}{\beta}\right)$, where we used that $\ln(1+x) \geq x/2$ for all $x \in [0, 1]$. So, $\ell = O\left(\frac{\beta \ln k}{\alpha^2 - \beta}\right)$. Also note that $r \leq 1/2 \cdot (2/\alpha^2)^\ell = \exp\left(O\left(\frac{\beta \ln k \ln(2/\alpha^2)}{\alpha^2 - \beta}\right)\right)$.

Our starting point is an instance of $\mathrm{SDP}^{\alpha,\beta}$. We would first like to use the polarization lemma to reduce it to an instance of $\mathrm{SDP}^{3/4,1/4}$. Set $k = 2$, and so $\ell \leq O(\log(n))$, where we used that $\alpha^2 - \beta \geq \Omega(1/\log(n))$, for n being the output length of the given circuits. Further, it holds that $\beta \ln(2/\alpha^2) \leq \beta \ln(2/\beta) \leq 1$ for all $\beta \in (0, 1)$, and hence, $r = \mathsf{poly}(n)$. The resulting polarization procedure is polynomial in the description of the input circuits, and thus reduces $\mathrm{SDP}^{\alpha,\beta}$ to $\mathrm{SDP}^{3/4,1/4}$. This shows that $\mathrm{SDP}^{\alpha,\beta} \in \mathsf{SZK}$. To show that $\mathrm{SDP}^{\alpha,\beta}$ is SZK-hard, we further reduce $\mathrm{SDP}^{3/4,1/4}$ to $\mathrm{SDP}^{1-2^{-k},2^{-k}}$, for $k(n) = n^{1/2-\varepsilon}$.

We are given a pair of circuits that are an instance of $\mathrm{SDP}^{3/4,1/4}$. First, pad the input and output to get a length-preserving circuits whose input and output lengths are some integer m. We apply the polarization lemma twice, each time appropriately setting the parameter k. In the second time we apply it with k to be determined by the analysis below. This k also sets the parameters for the first application of the polarization lemma as follows.

Let $\ell = \log_2 4k$—this is the ℓ set by the second application of the polarization lemma (we are in a regime that $\alpha^2 > 2\beta$, so $\lambda = 2$). We first polarize the circuits so that $\alpha \geq (1 - 1/\ell)$. To do this, set $k' = \log \ell$ and $\ell' = \log_2 k'$. The input length of the resulting polarized circuits is $m' = \left(\left(m \cdot \ell' \cdot \frac{2^{\ell'}}{2(3/4)^{2\ell'}} + \frac{2^{\ell'}}{2(3/4)^{2\ell'}}\right) \cdot k'\right) = O(m \cdot \mathsf{polylog}(\log(k)))$. Now, apply the polarization lemma one more time with k and $\ell = \log_2 4k \geq 2$ (that we already set) to get our (almost) final circuits D_0 and D_1. The input length of D_0 or D_1 is given by $n = \left(\left(m' \cdot \ell \cdot \frac{2^\ell}{2(1-1/\ell)^{2\ell}} + \frac{2^\ell}{2(1-1/\ell)^{2\ell}}\right) \cdot k\right) = O(mk^2\mathsf{polylog}(k))$, where we used that $(1 - 1/\ell)^\ell = \Omega(1)$. The statistical distance between D_0 and D_1 is now either at least $1 - 2^{-k}$ or at most 2^{-k}. Also note that the output length of the circuits is shorter than their input length n.

We would like that (D_0, D_1) be an instance of $\mathrm{SDP}^{1-2^{-n^{1/2-\varepsilon}},2^{-n^{1/2-\varepsilon}}}$. It hence suffices that $2^{-k} \leq 2^{-n^{1/2-\varepsilon}}$, namely $k \geq n^{1/2-\varepsilon}$. Let $c = \frac{2\varepsilon}{1-2\varepsilon}$ and assume that $k \geq m^{1/c}$ (we will shortly set k to satisfy this assumption). This setting guarantee that $(mk^2)^{1/2-\varepsilon} \leq k^{1-\varepsilon}$. Thus, it holds that $n^{1/2-\varepsilon} \leq O(k^{1-\varepsilon}\mathsf{polylog}(k)) \leq k$, where the last inequality holds for all $k > k(\varepsilon)$, for $k(\varepsilon)$ being a constant depends on ε and the hidden constants in the O notation of n. We are now finally able to set $k = \max(m^{1/c}, k(\varepsilon))$.

Lastly, we also want to produce length-preserving circuits. Since the input length of the circuits is longer than their output, and we set k according to the input length, we can simply pad the output so it equals the input length n. The resulting circuits are thus instance of $\mathrm{SDP}^{1-2^{-n^{1/2-\varepsilon}},2^{-n^{1/2-\varepsilon}}}$, as required. □

Theorem 3.16 ([GV99, GSV98])**.** *For every efficiently computable $p = p(n) \in \mathsf{poly}(n)$, the problem $\mathrm{EDP}^{1/p}$ is SZK-complete.*

4 Complete Problems for SZK

In this section we prove Theorems 1.5 and 1.6. That is, we show that the Triangular Discrimination Problem (TDP) and Jensen-Shannon Divergence Problem (JSP) are SZK-complete for any noticeable gap between the YES and NO promises.

The outline of this section is as follows. We begin, in Section 4.1, with proving that JSP is SZK-complete (Theorem 1.6). This proof is closely related to known results in SZK, and in particular closely follow reductions related to the Entropy Difference Problem (EDP). In Section 4.2 we prove that TDP is SZK-complete (Theorem 1.5). The latter proof is actually via a reduction to the SZK completeness of JSP.

4.1 JSP is Complete for SZK

In this section we show that the promise problem $\mathrm{JSP}^{\alpha,\beta}$ (see Definition 1.3) is complete for SZK. To do so, we need to show that $\mathrm{JSP}^{\alpha,\beta}$ is both in SZK and hard for SZK. The proofs of both parts rely on the following characterization of the Jensen-Shannon divergence.

Proposition 4.1. *Let P and Q be two distributions over a universe $\mathcal{Y}$. Let (B, Y) be the jointly distributed random variables defined as follows: $B \sim \{0,1\}$ and if $B = 1$, then $Y \sim P$ (that is, Y is a random variable drawn according to P), and if $B = 0$, then $Y \sim Q$.*

Then, $\mathrm{JS}(P, Q) = 1 - \mathrm{H}(B|Y)$.

Proof. Assume without loss of generality that $\mathcal{Y}$ is the union of the supports of P and Q (otherwise, exclude those elements from $\mathcal{Y}$ and the following calculations remain intact). For $y \in \mathcal{Y}$, recall that we defined the *imbalance* (wrt P and Q) as $\theta_y = \Pr[B = 1|Y = y] - \Pr[B = 0|Y = y] = \frac{P_y - Q_y}{P_y + Q_y}$ (see Definition 3.2). Observe that $(B|Y = y)$ is a Bernoulli random variable with parameter $\frac{1+\theta_y}{2} = \frac{P_y}{P_y+Q_y}$.

We proceed to a straightforward but somewhat tedious calculation that establishes the propo-

sition (in the following y is always summed over $\mathcal{Y}$).

$$\begin{aligned}
\mathrm{JS}(P,Q) &= \frac{1}{2}\,\mathrm{KL}\left(P\middle\|\frac{P+Q}{2}\right)+\frac{1}{2}\,\mathrm{KL}\left(Q\middle\|\frac{P+Q}{2}\right)\\
&= \frac{1}{2}\left(\sum_y P_y\log\frac{2P_y}{P_y+Q_y}+\sum_y Q_y\log\frac{2Q_y}{P_y+Q_y}\right)\\
&= 1+\frac{1}{2}\left(\sum_y P_y\log\frac{P_y}{P_y+Q_y}+\sum_y Q_y\log\frac{Q_y}{P_y+Q_y}\right)\\
&= 1+\left(\sum_y \frac{P_y+Q_y}{2}\cdot\left(\frac{P_y}{P_y+Q_y}\log\frac{P_y}{P_y+Q_y}+\frac{Q_y}{P_y+Q_y}\log\frac{Q_y}{P_y+Q_y}\right)\right)\\
&= 1-\left(\sum_y \frac{P_y+Q_y}{2}\cdot h\left(\frac{1+\theta_y}{2}\right)\right)\\
&= 1-\mathbb{E}_{y\sim Y}\left[h\left(\frac{1+\theta_y}{2}\right)\right]\\
&= 1-\mathrm{H}(B|Y),
\end{aligned}$$

where we recall that h is the binary entropy function (see Definition 3.5). □

The above characterization naturally relates the Jensen-Shannon Divergence Problem to the Entropy Difference Problem (EDP), a promise problem already known to be complete for SZK (see Theorem 3.16). In particular, the proofs of the next two lemmas closely follow techniques from the reduction of Statistical Closeness (i.e., the reversal problem of SDP) to EDP [Vad99, Section 4.4].

Lemma 4.2 (JSP is in SZK)**.** *For every* $(1/\mathsf{poly})$*-separated pair of functions* (α,β) *(according to Definition 1.4), the promise problem* $\mathrm{JSP}^{\alpha,\beta}$ *is in* SZK*.*

Proof. The proof reduces $\mathrm{JSP}^{\alpha,\beta}$ to EDP^{g}, where $g(n)=(\alpha(n-1)-\beta(n-1))/2$ for every $n\geq 2$, and $g(1)=g(2)$.[24] Since (α,β) are polynomially separated, Theorem 3.16 completes the proof.

Given a pair of circuits (C_0,C_1) whose output length is n, let (B,Y) be the jointly distributed random variables from Proposition 4.1 with respect to the distributions C_0 and C_1 . The reduction outputs the pair of circuits (D_0,D_1) such that D_1 outputs a sample from (B,Y) and D_0 outputs a sample from (B',Y), where B' is an independent random bit with $\mathrm{H}(B')=1-\frac{\alpha(n)+\beta(n)}{2}$.[25] Note that the output length of D_0 or D_1 is $n+1\geq 2$. It holds that

$$\mathrm{H}(D_0)-\mathrm{H}(D_1)=\mathrm{H}(B',Y)-\mathrm{H}(B,Y)=\mathrm{H}(B')-\mathrm{H}(B|Y)=1-\frac{\alpha(n)+\beta(n)}{2}-\mathrm{H}(B|Y).$$

If $\mathrm{JS}(C_0,C_1)\geq\alpha(n)$, then by Proposition 4.1, $\mathrm{H}(B|Y)\leq 1-\alpha(n)$. It holds that

$$\mathrm{H}(D_0)-\mathrm{H}(D_1)\geq 1-\frac{\alpha(n)+\beta(n)}{2}-(1-\alpha(n))=\frac{\alpha(n)-\beta(n)}{2}=g(n+1).$$

[24] Seeting $g(1)=g(2)$ is done for technical reasons so g would be defined for all $n\in\mathbb{N}$. As we will soon see, the reduction always outputs circuits whose output length is at least 2.

[25] To sample such B' we require that $p=h^{-1}((\alpha+\beta)/2)$ can be described using polynomially many bits. While this might not always be true, we can efficiently compute $p'\approx p$ such that the difference between their binary entropies are negligible. For simplicity, we ignore this issue in this proof.

If $\mathrm{JS}(C_0, C_1) \leq \beta(n)$, then by Proposition 4.1, $\mathrm{H}(B|Y) \geq 1 - \beta(n)$. It holds that

$$\mathrm{H}(D_0) - \mathrm{H}(D_1) \leq 1 - \frac{\alpha(n) + \beta(n)}{2} - (1 - \beta(n)) = -\frac{\alpha(n) - \beta(n)}{2} = -g(n+1).$$

Finally, since the output length of D_0 or D_1 is $n+1$, the mapping $(C_0, C_1) \mapsto (D_0, D_1)$ is a polynomial-time reduction from $\mathrm{JSP}^{\alpha,\beta}$ to EDP^g. □

Lemma 4.3 (JSP is Hard for SZK)**.** *Let $\alpha, \beta \colon \mathbb{N} \to [0,1]$ be efficiently- computable functions such that there exists a constant $\varepsilon \in (0, 1/2)$ such that $2^{-n^{1/2-\varepsilon}} \leq \beta(n)$ and $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$ for every $n \in \mathbb{N}$. Then, the promise problem $\mathrm{JSP}^{\alpha,\beta}$ is hard for SZK.*

Proof. We reduce $\mathrm{SDP}^{1-2^{-n^{1/2-\varepsilon/2}}, 2^{-n^{1/2-\varepsilon/2}}}$ with length-preserving circuits to $\mathrm{JSP}^{\alpha,\beta}$. This suffices since the former problem is SZK-hard (Theorem 3.14).

Let (C_0, C_1) be a pair of circuits whose output length is n, and let (B, Y) be the jointly distributed random variables defined in Proposition 4.1 w.r.t. the distributions C_0 and C_1. Assume for now that $n \geq n(\varepsilon)$, where $n(\varepsilon)$ is some constant dependent on ε to be determined by the analysis later. Vadhan [Vad99, Claim 4.4.2] showed the following relation between the statistical difference of C_0 and C_1 to $\mathrm{H}(B|Y)$. Specifically, if $\mathrm{SD}(C_0, C_1) = \delta$, then

$$1 - \delta \leq \mathrm{H}(B|Y) \leq h\left(\frac{1-\delta}{2}\right). \tag{4.1}$$

If $\mathrm{SD}(C_0, C_1) \geq 1 - 2^{-n^{1/2-\varepsilon/2}}$, then Proposition 4.1 yields that

$$\begin{aligned}
\mathrm{JS}(C_0, C_1) = 1 - \mathrm{H}(B|Y) &\geq 1 - h\left(\frac{1 - (1 - 2^{-n^{1/2-\varepsilon/2}})}{2}\right) \\
&= 1 - h\left(2^{-n^{1/2-\varepsilon/2}-1}\right) \\
&\geq 1 - 2 \cdot 2^{-(n^{1/2-\varepsilon/2}+1)/2} \geq \alpha(n),
\end{aligned}$$

where we used that $h\left(\frac{1-\delta}{2}\right)$ is decreasing in $0 \leq \delta \leq 1$, that $h(p) \leq 2\sqrt{p}$ for all $p \in [0,1]$ and we set $n(\varepsilon)$ so that the last inequality holds. Specifically, recall that $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$ so there exists a constant $n(\varepsilon)$ such that $1 - 2 \cdot 2^{-(n^{1/2-\varepsilon/2}+1)/2} \geq 1 - 2^{-n^{1/2-\varepsilon}}$ for all $n \geq n(\varepsilon)$.

On the other hand, if $\mathrm{SD}(C_0, C_1) \leq 2^{-n^{1/2-\varepsilon/2}}$, it holds that

$$\mathrm{JS}(C_0, C_1) = 1 - \mathrm{H}(B|Y) \leq 1 - (1 - 2^{-n^{1/2-\varepsilon/2}}) = 2^{-n^{1/2-\varepsilon/2}} \leq \beta(n),$$

where the last inequality holds for $n \geq n(\varepsilon)$ (recall that $\beta(n) \geq 2^{-n^{1/2-\varepsilon/2}}$).

Hence, the identity mapping $(C_0, C_1) \mapsto (C_0, C_1)$ is a reduction from $\mathrm{SDP}^{1-2^{-n^{1/2-\varepsilon/2}}, 2^{-n^{1/2-\varepsilon/2}}}$ to $\mathrm{JSP}^{\alpha,\beta}$, as long as the output length of the given circuits is larger than $n(\varepsilon)$. For circuits of shorter output, we use that the circuits are length-preserving, so their input is also shorter than $n(\varepsilon)$. For such input circuits the reduction can go over all inputs (at most $2^{n(\varepsilon)}$ strings, which is constant) and compute exactly the statistical distance. If that statistical distance is larger than $1 - 2^{-n^{1/2-\varepsilon}}$, the reduction outputs arbitrary disjoint circuits; and if that statistical distance is smaller than $2^{-n^{1/2-\varepsilon}}$, the reduction outputs arbitrary identical circuits. □

Lemmas 4.2 and 4.3 imply that $\mathrm{JSP}^{\alpha,\beta}$ is SZK-complete, for the desired set of (α, β), thereby proving Theorem 1.6.

4.2 TDP is Complete for SZK

In this section we show that the promise problem $\mathrm{TDP}^{\alpha,\beta}$ (see Definition 1.2) is SZK complete. To do so, we need to show that $\mathrm{TDP}^{\alpha,\beta}$ is both in SZK and hard for SZK. We start by proving the latter.

Lemma 4.4 (TDP is Hard for SZK)**.** *Let $\alpha, \beta \colon \mathbb{N} \to [0,1]$ be efficiently- computable functions such that there exists a constant $\varepsilon \in (0, 1/2)$ such that $2^{-n^{1/2-\varepsilon}} \leq \beta(n)$ and $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$ for every $n \in \mathbb{N}$. Then, the promise problem $\mathrm{TDP}^{\alpha,\beta}$ is hard for SZK.*

We prove Lemma 4.4 by using the fact that the triangular discrimination is polynomially related to the statistical distance.

Proof. We reduce $\mathrm{SDP}^{1-2^{-n^{1/2-\varepsilon/2}},2^{-n^{1/2-\varepsilon/2}}}$ with length-preserving circuits to $\mathrm{TDP}^{\alpha,\beta}$. This suffices since the former problem is SZK-hard (Theorem 3.14).

We will use the fact that triangular discrimination is sandwiched between the statistical difference squared and the statistical difference. Specifically, recall Eq. (1.1): for every distributions P and Q, it holds that

$$\mathrm{SD}(P,Q)^2 \leq \mathrm{TD}(P,Q) \leq \mathrm{SD}(P,Q).$$

Let (C_0, C_1) be a pair of circuits whose output length is n. Assume for now that $n \geq n(\varepsilon)$, where $n(\varepsilon)$ is some constant dependent on ε to be determined by the analysis later. If $\mathrm{SD}(C_0, C_1) \geq 1 - 2^{-n^{1/2-\varepsilon/2}}$, then

$$\mathrm{TD}(C_0, C_1) \geq (1 - 2^{-n^{1/2-\varepsilon/2}})^2 \geq 1 - 2^{-n^{1/2-\varepsilon/2}+1} \geq \alpha(n),$$

where we set $n(\varepsilon)$ so that the last inequality holds. Specifically, recall that $\alpha(n) \leq 1 - 2^{-n^{1/2-\varepsilon}}$ so there exists a constant $n(\varepsilon)$ such that $1 - 2^{-n^{1/2-\varepsilon/2}+1} \geq 1 - 2^{-n^{1/2-\varepsilon}}$ for all $n \geq n(\varepsilon)$.

On the other hand, if $\mathrm{SD}(C_0, C_1) \leq 2^{-n^{1/2-\varepsilon/2}}$ it holds that

$$\mathrm{TD}(C_0, C_1) \leq 2^{-n^{1/2-\varepsilon/2}} \leq \beta(n),$$

where the last inequality holds for every n by the assumption on β.

Hence, the identity mapping $(C_0, C_1) \mapsto (C_0, C_1)$ is a reduction from $\mathrm{SDP}^{1-2^{-n^{1/2-\varepsilon/2}},2^{-n^{1/2-\varepsilon/2}}}$ to $\mathrm{JSP}^{\alpha,\beta}$, as long as the output length of the given circuits is larger than $n(\varepsilon)$. For circuits of shorter output, we use the same procedure described at the end of the proof of Lemma 4.3. □

It is left to show that $\mathrm{TDP}^{\alpha,\beta}$ is in SZK. Given that the triangular discrimination is polynomially related to statistical difference, a natural approach to achieve such goal is to *polarize* the triangular discrimination of the given distributions. Namely, design an efficient procedure that takes as input a pair of circuits (C_0, C_1) and outputs a pair of circuits (D_0, D_1) such that if $\mathrm{TD}(C_0, C_1) \geq \alpha$ then $\mathrm{TD}(D_0, D_1) \geq 1 - 2^{-k}$, and if $\mathrm{TD}(C_0, C_1) \leq \beta$ then $\mathrm{TD}(D_0, D_1) \leq 2^{-k}$. Using Eq. (1.1), we would now be able to reduce $\mathrm{TDP}^{\alpha,\beta}$ to $\mathrm{SDP}^{2/3,1/3}$. Indeed, Sahai and Vadhan [SV03] used such a polarization lemma for statistical difference to show that $\mathrm{SDP}^{2/3,1/3}$ is in SZK.

We can adapt the polarization lemma of [SV03] to polarize triangular discrimination as well, because triangular discrimination behaves sufficiently like statistical distance under the repetition and xor operations. Analogous to the statistical distance polarization, where α^2 and β can be $(1/\mathsf{log})$-separated, this approach allows us to show that $\mathrm{TDP}^{\alpha,\beta} \in \mathsf{SZK}$ for all $(1/\mathsf{log})$-separated α and β.

To show the stronger claim that $\mathrm{TDP}^{\alpha,\beta}$ is in SZK for $(1/\mathsf{poly})$-separated α and β we take a different approach—we reduce $\mathrm{TDP}^{\alpha,\beta}$ to $\mathrm{JSP}^{\alpha'\beta'}$ for some $(1/\mathsf{poly})$-separated α' and β'. Since we already showed that $\mathrm{JSP}^{\alpha'\beta'}$ is in SZK (see Lemma 4.2), this shows that $\mathrm{TDP}^{\alpha,\beta}$ is also in SZK.

The reduction from $\mathrm{TDP}^{\alpha,\beta}$ to $\mathrm{JSP}^{\alpha'\beta'}$ is given in Section 4.2.1. Since we find the (direct) polarization lemma for triangular discrimination and its analogy to [SV03]'s polarization lemma for statistical difference interesting, we prove it in Section 4.2.2.

4.2.1 From TDP to JSP

In this section we prove that TDP with any noticeable gap is in SZK. This proof is via a Karp reduction to JSP with a noticeable gap, which we have already shown to be in SZK (see Lemma 4.2). Since SZK is closed under Karp reductions, this implies that TDP with any noticeable gap is in SZK.

Lemma 4.5. *Let (α,β) be $(1/\mathsf{poly})$-separated (according to Definition 1.4). Then there exist $(1/\mathsf{poly})$-separated (α',β') such that $\mathrm{TDP}^{\alpha,\beta}$ is polynomially (Karp) reducible to $\mathrm{JSP}^{\alpha',\beta'}$.*

Corollary 4.6. *For every $(1/\mathsf{poly})$-separated (α,β), the promise problem $\mathrm{TDP}^{\alpha,\beta}$ is in SZK.*

Lemma 4.4 and Corollary 4.6 together imply that $\mathrm{TDP}^{\alpha,\beta}$ is SZK-complete, for the desired set of (α,β), thereby proving Theorem 1.5.

The main technical tool we use to prove Lemma 4.5 is to consider the *convex combination* of a pair of circuits. Given a pair of circuits (C_0,C_1), consider the circuits (D_0,D_1), where $D_b = \lambda C_b + (1-\lambda)\frac{C_0+C_1}{2}$. Unsurprisingly, such an operation reduces the difference between the circuits. To analyze its exact effect on the triangular discrimination, it will be convenient to characterize the triangular discrimination in terms of the random variables (B,Y) and the imbalance θ_y (see Definition 3.2).

Proposition 4.7. *Let P and Q be two distributions over a universe $\mathcal{Y}$. Let (B,Y) be the jointly distributed random variables defined as follows: $B \sim \{0,1\}$ and if $B=1$, then $Y \sim P$ (that is, Y is a random variable drawn according to P), and if $B=0$, then $Y \sim Q$. Finally, for $y \in \mathrm{Supp}(Y)$, recall that $\theta_y = \Pr[B=1|Y=y] - \Pr[B=0|Y=y] = \frac{P_y-Q_y}{P_y+Q_y}$.*

Then, $\mathrm{TD}(P,Q) = \mathbb{E}_{y\sim Y}[\theta_y^2]$.

Proof. In the following y is summed over $\mathrm{Supp}(Y)$.

$$\mathrm{TD}(P,Q) = \frac{1}{2}\sum_y \frac{(P_y-Q_y)^2}{P_y+Q_y} = \sum_y \frac{P_y+Q_y}{2}\cdot\frac{(P_y-Q_y)^2}{(P_y+Q_y)^2} = \mathbb{E}_{y\sim Y}[\theta_y^2].$$

□

It is easy to see the effect of the convex combination operation on θ_y.

Proposition 4.8. *Let P and Q be two distributions over a universe $\mathcal{Y}$ and let $0 \le \lambda \le 1$. Define the distributions $P' = \lambda\cdot P + (1-\lambda)\cdot\frac{P+Q}{2}$ and $Q' = \lambda\cdot Q + (1-\lambda)\cdot\frac{P+Q}{2}$.*

Then, for every $y \in \mathrm{Supp}(P)\cup\mathrm{Supp}(Q)$, equivalently $\mathrm{Supp}(P')\cup\mathrm{Supp}(Q')$, it holds that $\theta_y^{P',Q'} = \lambda\cdot\theta_y^{P,Q}$.

Proof.

$$\theta_y^{P',Q'} = \frac{P'_y-Q'_y}{P'_y+Q'_y} = \frac{\lambda\cdot(P_y-Q_y)}{\lambda\cdot(P_y+Q_y)+(1-\lambda)\cdot(P_y+Q_y)} = \frac{\lambda\cdot(P_y-Q_y)}{P_y+Q_y} = \lambda\cdot\theta_y^{P,Q}.$$

□

Propositions 4.7 and 4.8 immediately yield that $\mathrm{TD}(D_0, D_1) = \lambda^2 \cdot \mathrm{TD}(C_0, C_1)$. So, as long as λ is not too small, a noticeable gap in the triangular discrimination is preserved. We can now finally prove Lemma 4.5. The main insight in the proof is that for small θ_y's the Jensen-Shannon divergence behaves like θ_y^2. The first step in the proof is to reduce the magnitude of the θ_y's by taking via a convex combination with some small parameter λ. Since, by Proposition 4.7, the triangular discrimination is *exactly* characterized by θ_y^2, we can now relate the two measures. One difficulty arises when performing the convex combination—the gap in triangular discrimination between the Yes and the No cases shrinks as well. We show that with a careful choice of λ, the Jensen-Shannon divergence is "closer" to θ_y^2 than the degree to which the gap decreases, thus ensuring that we preserve a noticeable gap.

Proof of Lemma 4.5. The proof relies on the Taylor series of the function $g(\theta) = 1 - h((1+\theta)/2)$ around 0:

$$g(\theta) = \frac{\theta^2}{2\ln 2} + \frac{1}{2\ln 2}\sum_{n=2}^{\infty} \frac{\theta^{2n}}{n(2n-1)}.$$

(This series is obtained from the Taylor series of the binary entropy function h around $1/2$.) The above series yields that for all $0 \le \lambda \le 1$ and $-1 \le \theta \le 1$,

$$\frac{\lambda^2\theta^2}{2\ln 2} \le g(\lambda\theta) \le \frac{\lambda^2\theta^2}{2\ln 2} + \frac{\lambda^4}{2\ln 2}. \tag{4.2}$$

To see that the right-hand side inequality holds, note that

$$\frac{1}{2\ln 2}\sum_{n=2}^{\infty} \frac{(\lambda\theta)^{2n}}{n(2n-1)} \le \frac{\lambda^4}{2\ln 2}\sum_{n=2}^{\infty} \frac{\theta^{2n}}{n(2n-1)} \le \frac{\lambda^4}{2\ln 2}\sum_{n=2}^{\infty} \frac{1}{n(2n-1)} \le \frac{\lambda^4}{2\ln 2}\sum_{n=2}^{\infty} \frac{1}{n^2} = \frac{\lambda^4}{2\ln 2}\left(\frac{\pi^2}{6} - 1\right).$$

Let λ be the largest number such that $1/\lambda$ is a power of 2 and that $\lambda^2 \le (\alpha-\beta)/2$. This implies that $1/\lambda \le 2\sqrt{2/(\alpha-\beta)}$, and thus $\lambda^2 \ge (\alpha-\beta)/8$. Let (C_0, C_1) be a pair of circuits and consider the pair of circuits (D_0, D_1), where $D_b = \lambda \cdot C_b + (1-\lambda) \cdot \frac{C_0+C_1}{2}$. We wish to analyze $\mathrm{JS}(D_0, D_1)$.

Let (B, Y) be the random variables defined in Proposition 4.7 w.r.t. distributions C_0 and C_1, and let (B', Y') be defined similarly for D_0 and D_1. Note that Y is distributed the same as Y'. Indeed, Y' is sampled according to the distribution

$$\frac{D_0 + D_1}{2} = \frac{\lambda \cdot (C_0 + C_1) + (1-\lambda) \cdot (C_0 + C_1)}{2} = \frac{C_0 + C_1}{2},$$

and the latter is the distribution by which Y is sampled.

Assume that $\mathrm{TD}(C_0, C_1) \ge \alpha$. Then, Propositions 4.1 and 4.7 and Eq. (4.2) yield that

$$\mathrm{JS}(D_0, D_1) = \mathbb{E}_{y\sim Y'}\left[g(\lambda \cdot \theta_y)\right] \ge \frac{\lambda^2 \cdot \mathbb{E}_{y\sim Y}[\theta_y^2]}{2\ln 2} = \frac{\lambda^2 \cdot \mathrm{TD}(C_0, C_1)}{2\ln 2} \ge \frac{\lambda^2 \cdot \alpha}{2\ln 2}.$$

On the other hand, if $\mathrm{TD}(C_0, C_1) \le \beta$, then

$$\begin{aligned}\mathrm{JS}(D_0, D_1) = \mathbb{E}_{y\sim Y'}\left[g(\lambda \cdot \theta_y)\right] \le \frac{\lambda^2 \cdot \mathbb{E}_{y\sim Y}[\theta_y^2] + \lambda^4}{2\ln 2} &= \frac{\lambda^2 \cdot \mathrm{TD}(C_0, C_1) + \lambda^4}{2\ln 2} \\ &\le \frac{\lambda^2 \cdot (\beta + (\alpha-\beta)/2)}{2\ln 2} = \frac{\lambda^2}{2\ln 2} \cdot \left(\frac{\alpha+\beta}{2}\right).\end{aligned}$$

Set $\alpha' = \frac{\lambda^2\alpha}{2\ln 2}$ and $\beta' = \frac{\lambda^2}{2\ln 2} \cdot \left(\frac{\alpha+\beta}{2}\right)$. The mapping $(C_0, C_1) \mapsto (D_0, D_1)$ establishes the reduction from $\mathrm{TDP}^{\alpha,\beta}$ to $\mathrm{JSP}^{\alpha',\beta'}$. Note that the output lengths of the circuits is preserved, so the above calculation indeed guarantees the desired gap in the Jensen-Shannon divergence, as a function of the output length of D_0 and D_1. Since $1/\lambda$ is polynomial in $1/(\alpha-\beta)$, the reduction runs in polynomial time. Finally, it holds that

$$\alpha' - \beta' = \frac{\lambda^2}{2\ln 2} \cdot \left(\frac{\alpha-\beta}{2}\right) \geq \frac{(\alpha-\beta)^2}{32},$$

and since (α, β) are $(1/\mathsf{poly})$-separated, then so are (α', β'). □

4.2.2 A Polarization Lemma for Triangular Discrimination

In this section we give a procedure that polarizes the triangular discrimination of two input circuits. The procedure is practically identical to that of Sahai and Vadhan's polarization lemma [SV03, Lemma 3.3]. While [SV03, Lemma 3.3] requires that $\alpha^2 > \beta$, the polarization lemma for triangular discrimination only needs that $\alpha > \beta$. As already stated, the polarization that we show here is inferior to the (indirect) polarization obtained in Section 4.2.1, as it only supports an inverse *logarithmic* gap. Nevertheless, we include it since we find the approach appealing.

Lemma 4.9 (Polarization Lemma for triangular discrimination)**.** *There is an algorithm that takes as input the tuple $(X_0, X_1, \alpha, \beta, k)$, where X_0 and X_1 are circuits and $\alpha > \beta$, and outputs a pair of circuits (Y_0, Y_1) such that:*

$$\begin{aligned} \mathrm{TD}(X_0, X_1) \geq \alpha \quad &\implies \quad \mathrm{TD}(Y_0, Y_1) \geq 1 - 2^{-k} \\ \mathrm{TD}(X_0, X_1) \leq \beta \quad &\implies \quad \mathrm{TD}(Y_0, Y_1) \leq 2^{-k}. \end{aligned}$$

The running time of the algorithm is polynomial in the description of X_0 and X_1 as well as in k and $\exp\left(\frac{\beta\log(1/\alpha)}{\alpha-\beta}\right)$.

The proof of [SV03, Lemma 3.3] is done by considering two operations on a pair of circuits—the repetition (i.e., direct product) operation and the XOR operation. We analyze the effect of these operations on the triangular discrimination of the distributions.

Lemma 4.10 (Direct Product Lemma for triangular discrimination)**.** *Let P, Q be distributions such that $\mathrm{TD}(P, Q) = \delta$. Then for all $k \in \mathbb{N}$,*

$$1 - \exp(-\delta k/2) \leq \mathrm{TD}(P^{\otimes k}, Q^{\otimes k}) \leq 2k\delta.$$

This lemma is where the main difference between polarizing triangular discrimination and statistical difference lies. Specifically, the lower bound in the analogous lemma for statistical difference ([SV03, Lemma 3.4]) depends on δ^2, rather than δ here. This dependence is exactly why α^2 must be larger than β for statistical difference, but not for triangular discrimination.

Proof. The proof is done by considering the *Squared Hellinger distance*:

$$\mathrm{H}^2(P, Q) \triangleq \frac{1}{2}\sum_i \left(\sqrt{P_i} - \sqrt{Q_i}\right)^2 = 1 - \mathbb{E}_{X\sim Q}\left[\sqrt{\frac{P(X)}{Q(X)}}\right].$$

The squared Hellinger distance is useful in our context because of two properties: the triangular discrimination is sandwiched by constant factors of the squared Hellinger distance, and the squared

Hellinger distance tensorizes under product distributions. For the first property, Le Cam [Cam86, P. 48] showed that

$$\mathrm{H}^2(P,Q) \le \mathrm{TD}(P,Q) \le 2\,\mathrm{H}^2(P,Q). \tag{4.3}$$

For the second property, that $P^{\otimes k}, Q^{\otimes k}$ are product distributions yield that

$$\begin{aligned}
\mathrm{H}^2(P^{\otimes k}, Q^{\otimes k}) &= 1 - \mathbb{E}_{X^k \sim Q^{\otimes k}}\left[\sqrt{\frac{P^{\otimes k}(X^k)}{Q^{\otimes k}(X^k)}}\right] \\
&= 1 - \mathbb{E}_{X^k \sim Q^{\otimes k}}\left[\sqrt{\prod_{i=1}^{k}\frac{P(X_i)}{Q(X_i)}}\right] \\
&= 1 - \prod_{i=1}^{k}\mathbb{E}_{X_i \sim Q}\left[\sqrt{\frac{P(X_i)}{Q(X_i)}}\right] \\
&= 1 - (1 - \mathrm{H}^2(P,Q))^k.
\end{aligned} \tag{4.4}$$

Equipped with the above properties, we are now ready to prove the lemma. For the upper bound, it holds that

$$\begin{aligned}
\mathrm{TD}(P^{\otimes k}, Q^{\otimes k}) &\le 2\,\mathrm{H}^2(P^{\otimes k}, Q^{\otimes k}) \\
&= 2(1 - (1 - \mathrm{H}^2(P,Q))^k) \\
&\le 2(1 - (1 - k\,\mathrm{H}^2(P,Q))) \\
&= 2k\,\mathrm{H}^2(P,Q) \\
&\le 2k\,\mathrm{TD}(P,Q),
\end{aligned}$$

where the second inequality follows since $(1-x)^k \ge 1 - kx$ for all x and integer k.

For the lower bound, it holds that

$$\begin{aligned}
\mathrm{TD}(P^{\otimes k}, Q^{\otimes k}) &\ge \mathrm{H}^2(P^{\otimes k}, Q^{\otimes k}) \\
&= (1 - (1 - \mathrm{H}^2(P,Q))^k) \\
&\ge (1 - (1 - \delta/2)^k) \\
&\ge 1 - \exp(-\delta k/2),
\end{aligned}$$

where the last inequality follows since $1 - x \le e^{-x}$ for all x. □

Next we analyze the XOR operation for triangular discrimination, and see that it is idendical to the effect of this operation on statistical difference ([SV03, Lemma 3.5]).

Lemma 4.11 (XOR Lemma for triangular discrimination)**.** *There is a polynomial algorithm that takes as input $(X_0, X_1, 1^k)$, where X_0 and X_1 are circuits, and outputs a pair of circuits (Y_0, Y_1) such that $\mathrm{TD}(Y_0, Y_1) = \mathrm{TD}(X_0, X_1)^k$. Specifically, Y_0 and Y_1 are defined as follows:*

Y_0: *Sample $(b_1, \ldots, b_k) \sim \{0,1\}^k$ uniformly at random condition that $b_1 \oplus \cdots \oplus b_k = 0$ and output a sample from $X_{b_1} \cdot X_{b_2} \cdot \cdots \cdot X_{b_k}$.*

Y_1: *Sample $(b_1, \ldots, b_k) \sim \{0,1\}^k$ uniformly at random condition that $b_1 \oplus \cdots \oplus b_k = 1$ and output a sample from $X_{b_1} \cdot X_{b_2} \cdot \cdots \cdot X_{b_k}$.*

The proof of Lemma 4.11 follows from the next proposition and a straightforward induction.

Proposition 4.12. *Let P, P', Q, Q' be any distributions over $\mathcal{Y}$, and let $R = \frac{1}{2}(PP' + QQ')$ and $R' = \frac{1}{2}(PQ' + QP')$ be distributions over $\mathcal{Y} \times \mathcal{Y}$.*

Then $\mathrm{TD}(R, R') = \mathrm{TD}(P, Q) \cdot \mathrm{TD}(P', Q')$.

Proof. It is easy to verify that for every $i, j \in \mathcal{X}$, it holds that

$$R_{ij} = 0 \wedge R'_{ij} = 0 \iff (P_i = 0 \wedge Q_i = 0) \vee (P'_j = 0 \wedge Q'_j = 0). \tag{4.5}$$

Hence, the set $\mathcal{A} = \{(i,j) \colon R_{ij} > 0 \vee R'_{ij} > 0\}$ is exactly the product of $\mathcal{B} = \{i \colon P_i > 0 \vee Q_i > 0\}$ and $\mathcal{C} = \{j \colon P'_j > 0 \vee Q'_j > 0\}$.

Compute

$$\begin{aligned}
\mathrm{TD}(R, R') &= \frac{1}{2} \sum_{(i,j) \in \mathcal{A}} \frac{\left(\frac{1}{2}(P_i P'_j + Q_i Q'_j) - \frac{1}{2}(P_i Q'_j + Q_i P'_j)\right)^2}{\frac{1}{2}(P_i P'_j + Q_i Q'_j) + \frac{1}{2}(P_i Q'_j + Q_i P'_j)} \\
&= \frac{1}{4} \sum_{(i,j) \in \mathcal{A}} \frac{(P_i(P'_j - Q'_j) - Q_i(P'_j - Q'_j))^2}{P_i(P'_j + Q'_j) + Q_i(P'_j + Q'_j)} \\
&= \frac{1}{4} \sum_{i \in \mathcal{B}, j \in \mathcal{C}} \frac{((P_i - Q_i)(P'_j - Q'_j))^2}{(P_i + Q_i)(P'_j + Q'_j)} \\
&= \left(\frac{1}{2} \sum_{i \in \mathcal{B}} \frac{(P_i - Q_i)^2}{P_i + Q_i}\right) \cdot \left(\frac{1}{2} \sum_{j \in \mathcal{C}} \frac{(P'_j - Q'_j)^2}{P'_j + Q'_j}\right) \\
&= \mathrm{TD}(P, Q) \cdot \mathrm{TD}(P', Q').
\end{aligned}$$

□

Using the above analysis we can now prove Lemma 4.9. This proof follows similar lines to that of [SV03, Lemma 3.5].

Proof of Lemma 4.9. Let $\lambda = \min(\alpha/\beta, 2) > 1$,[26] and let $\ell = \lceil \log_\lambda 8k \rceil$. Apply the XOR Lemma (Lemma 4.11) to the input $(X_0, X_1, 1^\ell)$ to produce (X'_0, X'_1) such that

$$\begin{aligned}
\mathrm{TD}(X_0, X_1) \geq \alpha \quad &\implies \quad \mathrm{TD}(X'_0, X'_1) \geq \alpha^\ell \\
\mathrm{TD}(X_0, X_1) \leq \beta \quad &\implies \quad \mathrm{TD}(X'_0, X'_1) \leq \beta^\ell.
\end{aligned}$$

Let $m = \lambda^\ell/(4\alpha^\ell) \leq 1/(4\beta^\ell)$, let $X''_0 = (X'_0)^{\otimes m}$ and $X''_1 = (X'_1)^{\otimes m}$. The Direct Product Lemma (Lemma 4.10) now yields that

$$\begin{aligned}
\mathrm{TD}(X_0, X_1) \geq \alpha \quad &\implies \quad \mathrm{TD}(X''_0, X''_1) \geq 1 - \exp(-\alpha^\ell m/2) \geq 1 - e^{-k} \\
\mathrm{TD}(X_0, X_1) \leq \beta \quad &\implies \quad \mathrm{TD}(X''_0, X''_1) \leq 2m\beta^\ell \leq 1/2.
\end{aligned}$$

Finally, apply the XOR Lemma (Lemma 4.11) again to the input $(X''_0, X''_1, 1^k)$ to produce Y_0, Y_1 such that

$$\begin{aligned}
\mathrm{TD}(X_0, X_1) \geq \alpha \quad &\implies \quad \mathrm{TD}(Y_0, Y_1) \geq (1 - e^{-k})^k \geq 1 - ke^{-k} \geq 1 - 2^{-k} \\
\mathrm{TD}(X_0, X_1) \leq \beta \quad &\implies \quad \mathrm{TD}(Y_0, Y_1) \leq 1/2^k.
\end{aligned}$$

[26] This is the only place this proof diverges from that of [SV03, Lemma 3.5]. The latter sets $\lambda = \alpha^2/\beta$.

The last derivation holds for sufficiently large k, which we can obtain by increasing it at the start.

As for the running time, the analysis is similar to the one done in the proof sketch of Theorem 3.14 (which in turn follows [CCKV08, Lemma 38]), and we refer the reader there. □

5 One Way Functions from SDP with Any Noticeable Gap

In this section we construct a one-way function assuming the average case hardness of the Statistical Difference Problem (SDP), with any inverse polynomial gap. Actually, we will only construct a *distributional* one-way function but by a result of Impagliazzo and Luby [IL89], this yields a full-fledged one-way function. We first recall the (standard) definitions of a one-way function and a distributional one-way function.

Definition 5.1 (One-Way Function)**.** *A polynomial-time computable function $f\colon \{0,1\}^* \to \{0,1\}^*$ is* one-way *if for every probabilistic polynomial-time algorithm* A*,*

$$\Pr\left[\mathsf{A}(1^n, Y) \in f^{-1}(Y)\right] = \mathsf{negl}(n),$$

where $Y = f(X)$ for $X \sim \{0,1\}^n$ (and the probability is also over the coin tosses of A).

Definition 5.2 (Distributionally One-Way Function)**.** *A polynomial-time computable function $f\colon \{0,1\}^* \to \{0,1\}^*$ is* distributionally one-way *if there exists a polynomial p such that for every probabilistic polynomial-time algorithm* A *and all large enough n*

$$\mathrm{SD}\left((X,Y),(\mathsf{A}(1^n,Y),Y)\right) \geq \frac{1}{p(n)},$$

where $X \sim \{0,1\}^n$ and $Y = f(X)$.

Any one-way function is also a distributionally one-way function. While the other direction is not always true, [IL89] showed that the existence of both primitives is equivalent.

Lemma 5.3 ([IL89, Lemma 1])**.** *If there exists a distributionally one-way function then there exists a one-way function.*

Hence, to show the existence of one-way functions, it suffices to show that distributionally one-way functions exist. As noted above we will do so based on the *average-case hardness* of SDP.

Definition 5.4 (Average-case Hardness)**.** *We say that a promise problem $\Pi = (\mathrm{YES}, \mathrm{NO})$ is* average-case hard *if there exists a probabilistic polynomial-time algorithm* S *such that $\mathsf{S}(1^n)$ outputs samples from $\mathrm{YES} \cup \mathrm{NO}$, and for every probabilistic polynomial-time distinguisher* D*,*

$$\Pr_{x \sim \mathsf{S}(1^n)}\left[\mathsf{D}(1^n, x) = \Pi(x)\right] \leq \frac{1}{2} + \mathsf{negl}(n),$$

where $\Pi(x) = 1$ if $x \in \mathrm{YES}$ and $\Pi(x) = 0$ if $x \in \mathrm{NO}$. The above probability is taken also over the randomness of D*. We call* S *a* hard-instance sampler *for* Π*.*

The main result of this section is that the average-case hardness of SDP with any noticeable gap implies the existence of distributionally one-way functions.

Theorem 5.5. *Let (α, β) be $(1/\mathsf{poly})$-separated functions (according to Definition 1.4). Assume that $\mathrm{SDP}^{\alpha,\beta}$ is average-case hard with hard-instance sampler S. Then, the function f defined as*

$$f(1^n, r, b, x) = (1^n, C_0, C_1, y),$$

where[27] *$(C_0, C_1) = \mathsf{S}(1^n; r)$ and $y = C_b(x)$, is distributionally one-way.*[28]

It may be useful for the reader to think of f as a collection of functions indexed by the pair of circuits (C_0, C_1) that map (b, x) to $C_b(x)$. We refrain from describing the function as a collection as the definition of collections of *distributionally* one-way functions is slightly more cumbersome.

Theorem 5.5 immediately proves Theorem 1.8. Furthermore, our proof implicitly implies two additional results. First, it proves that the function f is distributional k-multi-collision resistant hash function, for $k = O\left(\frac{\log n}{(\alpha-\beta)^2}\right)$; that is, for a random output (c_0, c_1, y) of f, it is difficult to find k random preimages of (c_0, c_1, y). Second, the proof of Theorem 5.5 also implicitly shows the following strengthening of Goldreich's [Gol90] result: the existence of efficiently sampleable distributions whose statistical distance is α but no efficient algorithm can distinguish between them with advantage more than β, for any $(1/\mathsf{poly})$-separated (α, β), implies the existence of one-way functions.

The rest of this section is dedicated to proving Theorem 5.5.

5.1 Proving Theorem 5.5

The proof of Theorem 5.5 is via a reduction. We show that given an adversary that distributionally inverts f, we can construct a distinguisher that breaks the average case hardness of $\mathrm{SDP}^{\alpha,\beta}$.

We begin with defining the following jointly distributed random variables with respect to the security parameter n: Let $R \sim \{0,1\}^{\rho(n)}$, for $\rho(n)$ being a bound on the number of random bits that $\mathsf{S}(1^n)$ uses, let $(C_0, C_1) \sim \mathsf{S}(1^n; R)$, let $B \sim \{0,1\}$, let $X \sim \{0,1\}^{m(n)}$, for $m(n)$ being a bound on the input length of C_0 and C_1, and let $Y = C_B(X)$. Finally, let $W = (1^n, R, B, X)$ and $Z = (1^n, C_0, C_1, Y)$. Note that C_0 and C_1 are random variables taking values of circuits (i.e., a description of the circuit itself). This is in contrast to other parts of this paper in which we (abuse notation and) use C to denote also the output distribution of the circuit C. To avoid confusion, in this section we denote by P_0 and P_1 the output distributions of the circuits C_0 and C_1, respectively.

Assume toward a contradiction that f is not distributionally one-way and let $p(n) \in \mathsf{poly}(n)$ be some polynomial to be determined by the analysis. Then, there exists a probabilistic polynomial-time inverter A such that

$$\mathrm{SD}\left((W, Z), (\mathsf{A}(1^n, Z), Z)\right) < \frac{1}{p(n)}, \tag{5.1}$$

for every n in an infinite set $\mathcal{I} \subseteq \mathbb{N}$.

[27] Recall that $\mathsf{S}(1^n; r)$ stands for the output of $\mathsf{S}(1^n)$ when its random coins are set to be r.

[28] Definition 5.2 only considers functions whose domain is $\{0,1\}^*$, i.e., functions defined for every input length. Although this function is not defined for every input length (and has 1^n as an input), using the fact that it is defined on $\{0,1\}^{q(n)}$ for some $q(n) \in \mathsf{poly}(n)$ and standard padding techniques, such restricted distributionally one-way function imply the existence of standard distributionally one-way function, per Definition 5.2. In the rest of this section we ignore this issue and assume that distributionally one-way functions can be defined for inputs in $\{0,1\}^{q(n)}$, for some $q(n) \in \mathsf{poly}(n)$.

Using the inverter A we construct a distinguisher D such that for large enough $n \in \mathcal{I}$, it holds that[29]

$$\Pr\Big[\mathsf{D}(1^n,(C_0,C_1)) = \mathbb{1}\{\mathrm{SD}(P_0,P_1) \geq \alpha(n)\}\Big] \geq \frac{1}{2} + \frac{1}{q(n)}, \tag{5.2}$$

for some $q(n) \in \mathsf{poly}(n)$ to be determined by the analysis.[30] The existence of such a D contradicts the average-case hardness of $\mathrm{SDP}^{\alpha,\beta}$ and so it remains to establish Eq. (5.2).

Fix some large enough $n \in \mathcal{I}$. When it is clear from the context, we will sometimes omit n from the notations. Our first step is to show that for a large fraction of the circuit pairs sampled by S, the inverter A inverts the function well. Let R', B', X' be random variables induced by the output of $\mathsf{A}(1^n, Z)$. By Eq. (5.1) (and using the data-processing inequality for statistical distance) it holds that

$$\mathrm{SD}\big((B,Y,C_0,C_1),(B',Y,C_0,C_1)\big) < \frac{1}{p}.$$

Let

$$\mathcal{U} = \Big\{(c_0,c_1)\colon \mathrm{SD}\Big(\big((B,Y)|(C_0,C_1)=(c_0,c_1)\big),\big((B',Y)|(C_0,C_1)=(c_0,c_1)\big)\Big) < 1/\sqrt{p}\Big\}. \tag{5.3}$$

Namely, $\mathcal{U}$ is the set containing the pair of circuits for which the inverter's output B', when given a random Y, is statistically close to B. It follows that

$$\frac{1}{p} > \mathrm{SD}\big((B,Y,C_0,C_1),(B',Y,C_0,C_1)\big) \geq \Pr[(C_0,C_1)\notin\mathcal{U}]\cdot\frac{1}{\sqrt{p}}, \tag{5.4}$$

and thus $\Pr[(C_0,C_1)\in\mathcal{U}] > 1-\frac{1}{\sqrt{p}}$.

We design a distinguisher D that works particularly well when it is given a circuit pair belonging to $\mathcal{U}$. Specifically, we describe D for which, for all $(c_0,c_1)\in\mathcal{U}$, it holds that

$$\Pr\big[\mathsf{D}(c_0,c_1) = \mathbb{1}\{\mathrm{SD}(P_0,P_1)\geq\alpha(n)\}\big] \geq 1-\frac{4}{n}, \tag{5.5}$$

where the probability is over the randomness of D. Such a distinguisher yields Eq. (5.2) as follows:

$$\begin{aligned}\Pr\big[\mathsf{D}(C_0,C_1) = \mathbb{1}\{\mathrm{SD}(P_0,P_1)\geq\alpha(n)\}\big] &\geq \Pr[(C_0,C_1)\in\mathcal{U}]\cdot\left(1-\frac{4}{n}\right)\\ &\geq \left(1-\frac{1}{\sqrt{p(n)}}\right)\cdot\left(1-\frac{4}{n}\right)\\ &\geq \frac{1}{2}+\frac{1}{q(n)},\end{aligned}$$

for large enough $q(n)\in\mathsf{poly}(n)$.

In the rest of the proof we establish Eq. (5.5). For $(c_0,c_1,y)\in\mathrm{Supp}(C_0,C_1,Y)$, let

$$\theta_{c_0,c_1}(y) \triangleq \Pr\big[B=1|C_0=c_0,C_1=c_1,Y=y\big] - \Pr\big[B=0|C_0=c_0,C_1=c_1,Y=y\big].$$

[29] Recall that for a boolean statement S (e.g., $X\geq 5$), $\mathbb{1}\{S\}$ stands for the indicator function that outputs 1 if S is a true statement and 0 otherwise.

[30] The probability is over the choices of (C_0,C_1) and the randomness of D. The statistical difference in the probability is between the output distributions of C_0 and C_1 after those circuits were drawn and fixed.

Namely, $\theta_{c_0,c_1}(y)$ measures the difference between the likelihoods that each circuit outputs y. A perfect inverter for such f, when given an output y, would return 1 with probability $\frac{1+\theta_{c_0,c_1}(y)}{2}$, and 0 with probability $\frac{1-\theta_{c_0,c_1}(y)}{2}$.

The quantity $\theta_{c_0,c_1}(y)$ plays a crucial role in the proof. First, we show that it can be used to characterize the statistical distance between the output distributions of c_0 and c_1. Second, we show that it can be well-approximated using the inverter A, specifically using random samples of B'. Thus, the distinguisher D uses A to approximate $\mathrm{SD}(P_0, P_1)$ and answers accordingly. We proceed to the actual proof.

For $(c_0, c_1) \in \mathrm{Supp}(C_0, C_1)$, let Y_{c_0,c_1} denote the random variable sampled according to $(Y|(C_0, C_1) = (c_0, c_1))$ (i.e., Y_{c_0,c_1} drawn from the distribution $\frac{P_0+P_1}{2}$). For $y \in \mathrm{Supp}(Y_{c_0,c_1})$ let $B_{c_0,c_1,y}$ denote the random variable sampled according to $(B|(C_0 = c_0, C_1 = c_1, Y = y))$, and similarly for $B'_{c_0,c_1,y}$.

Using the above notations, Proposition 3.3 states that for every $(c_0, c_1) \in \mathrm{Supp}(C_0, C_1)$, it holds that

$$\mathrm{SD}(P_0, P_1) = \mathop{\mathbb{E}}_{y \sim Y_{c_0,c_1}} [|\theta_{c_0,c_1}(y)|]. \tag{5.6}$$

Namely, the statistical distance between the output distributions of c_0 and c_1 is the expected value over $y \sim Y_{c_0,c_1}$ of $|\theta_{c_0,c_1}(y)|$. So, the distinguisher's task is to approximate the expected value of $|\theta_{c_0,c_1}(y)|$, for $(c_0, c_1) \in \mathcal{U}$. We design such a distinguisher in two steps. First, we show how to estimate $\theta_{c_0,c_1}(y)$ for a random y. Then, we use this estimator and Eq. (5.6) to calculate an approximation for the statistical distance (up to some inverse polynomial additive error). In the following we let $\varepsilon = \frac{\alpha-\beta}{4}$, $k = \left\lceil \frac{\log(n)}{\varepsilon^2} \right\rceil$ and $\ell = \left\lceil \frac{2+\log(k \cdot n)}{2(\varepsilon/2)^2} \right\rceil$. Note that since $\alpha > \beta$ are noticeably separated, it holds that $k, \ell \in \mathsf{poly}(n)$.

5.1.1 Estimating $\theta_{c_0,c_1}(y)$

Consider the following algorithm Est whose goal is to estimate $\theta_{c_0,c_1}(y)$. The algorithm gets access to an oracle O that takes as input $(c_0, c_1, y) \in \mathrm{Supp}(C_0, C_1, Y)$ and outputs a bit b'.

Estimator $\mathsf{Est}^{\mathsf{O}}(c_0, c_1, y)$:

1. For every $i \in [\ell]$, run $\mathsf{O}(c_0, c_1, y)$ to get $(b'_1, \ldots, b'_\ell)$.
2. Let $m = |\{b'_i \colon b'_i = 1\}|$ be number of ones in $(b'_1, \ldots, b'_\ell)$.
3. Set $\widehat{P}_B(1) = m/\ell$ and $\widehat{P}_B(0) = (\ell - m)/\ell$.
4. Return $\widehat{P}_B(1) - \widehat{P}_B(0)$.

The next claim shows that if the oracle $\mathsf{O}(c_0, c_1, y)$ perfectly samples from $B_{c_0,c_1,y}$, then indeed the output of $\mathsf{Est}^{\mathsf{O}}(c_0, c_1, y)$ is a good estimator for $\theta_{c_0,c_1}(y)$.

Claim 5.1. *Let $(c_0, c_1, y) \in \mathrm{Supp}(C_0, C_1, Y)$ and assume that $\mathsf{O}(c_0, c_1, y) \sim B_{c_0,c_1,y}$. Then, it holds that*

$$\Pr\left[\left|\theta_{C_0,C_1}(y) - \mathsf{Est}^{\mathsf{O}}(c_0, c_1, y)\right| > \varepsilon\right] \leq \frac{1}{kn},$$

where the above probability is over the randomness of O.

Proof. We use Fact 3.9 to show that $\mathsf{Est}^{\mathsf{O}}(c_0, c_1, y)$ indeed approximates $\theta_{c_0,c_1}(y)$. Let P_B denote the distribution of $B_{c_0,c_1,y}$. By the assumption on $\mathsf{O}(c_0, c_1, y)$ and the definition of Est it follows

that $\widehat{P}_B$ (defined in line 3 of Est) is the empirical distribution of P_B, computed using ℓ samples. Since the domain size of B is 2, the setting of ℓ and Fact 3.9 yield that

$$\begin{aligned}\Pr\Big[\Big|\theta_{c_0,c_1}(y) - \mathsf{Est}^{\mathsf{O}}(c_0,c_1,y)\Big| > \varepsilon\Big] &= \Pr\Big[\Big|(P_B(1)-P_B(0)) - (\widehat{P}_B(1)-\widehat{P}_B(0))\Big| > \varepsilon\Big] \\ &\le \Pr\Big[\mathrm{SD}(P_B,\widehat{P}_B) \ge \varepsilon/2\Big] \\ &\le \frac{1}{kn}.\end{aligned} \tag{5.7}$$

□

Unfortunately, even with the inverter A we cannot implement an oracle O that perfectly samples $B_{c_0,c_1,y}$. However, we can show that for $(c_0,c_1) \in \mathcal{U}$, the inverter approximates such an oracle for a *random* y. In the following we let $\widetilde{\mathsf{A}}$ be the projection variant of A that outputs A's second output (i.e., the bit b').

Claim 5.2. *Let $(c_0,c_1) \in \mathcal{U}$ (see Eq. (5.3)). Then*

$$\Pr_{y\sim Y_{c_0,c_1}}\Big[\Big|\theta_{c_0,c_1}(y) - \mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0,c_1,y)\Big| > \varepsilon\Big] \le \frac{1}{kn} + \frac{\ell+1}{\sqrt[4]{p}},$$

where the above probability is also over the randomness of $\widetilde{\mathsf{A}}$.

Proof. Consider the set

$$\mathcal{V} = \Big\{y\colon \mathrm{SD}\big(B_{c_0,c_1,y}, B'_{c_0,c_1,y}\big) < \frac{1}{\sqrt[4]{p}}\Big\}.$$

Similar calculations as those made in Eq. (5.4), and also using that $(c_0,c_1) \in \mathcal{U}$ (and thus $\mathrm{SD}\Big((B_{c_0,c_1,Y}, Y_{c_0,c_1}), (B'_{c_0,c_1,Y}, Y_{c_0,c_1})\Big) \le 1/\sqrt{p}$), show that $\Pr_{y\sim Y_{c_0,c_1}}[y \notin \mathcal{V}] \le \frac{1}{\sqrt[4]{p}}$. Let $y \in \mathcal{V}$. It follows that $\mathrm{SD}(\widetilde{\mathsf{A}}(c_0,c_1,y), \mathsf{O}(c_0,c_1,y)) \le 1/\sqrt[4]{p}$, where the oracle O is from Claim 5.1. Since Est makes ℓ oracle calls, a standard argument using the data-processing inequality for statistical distance and Claim 5.1 yield that for every $y \in \mathcal{V}$

$$\begin{aligned}\Pr\Big[\Big|\theta_{c_0,c_1}(y) - \mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0,c_1,y)\Big| > \varepsilon\Big] &\le \Pr\Big[\Big|\theta_{c_0,c_1}(y) - \mathsf{Est}^{\mathsf{O}}(c_0,c_1,y)\Big| > \varepsilon\Big] \\ &\quad + \ell \cdot \mathrm{SD}(\widetilde{\mathsf{A}}(c_0,c_1,y), \mathsf{O}(c_0,c_1,y)) \\ &\le \frac{1}{kn} + \frac{\ell}{\sqrt[4]{p}}.\end{aligned}$$

All in all,

$$\begin{aligned}&\Pr_{y\sim Y_{c_0,c_1}}\Big[\Big|\theta_{c_0,c_1}(y) - \mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0,c_1,y)\Big| > \varepsilon\Big] \\ &\le \Pr_{y\sim Y_{c_0,c_1}}\Big[\Big|\theta_{c_0,c_1}(y) - \mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0,c_1,y)\Big| > \varepsilon \Big| y \in \mathcal{V}\Big] + \Pr_{y\sim Y_{c_0,c_1}}[y \notin \mathcal{V}] \\ &\le \frac{1}{kn} + \frac{\ell}{\sqrt[4]{p}} + \frac{1}{\sqrt[4]{p}},\end{aligned}$$

as required.

□

5.1.2 Approximating the Statistical Distance

Using the $\theta_{c_0,c_1}(y)$ estimator Est, we now describe the distinguisher D (recall that $\widetilde{\mathsf{A}}$ is the projection variant of A that outputs A's second output (i.e., the bit b')).

Distinguisher $\mathsf{D}^{\mathsf{A}}(c_0, c_1)$:

1. For every $i \in [k]$, draw $b_i \sim B$, $x_i \sim X$, and set $y_i = c_{b_i}(x_i)$.
2. Compute $\widehat{\Delta} = \frac{1}{k}\sum_{i=1}^{k} |\mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0, c_1, y_i)|$.
3. Output 1 if $\widehat{\Delta} > \frac{\alpha+\beta}{2}$ and 0 otherwise.

We show that for $(c_0, c_1) \in \mathcal{U}$, the distinguisher approximates the relevant statistical distance well.

Claim 5.3. *Let $(c_0, c_1) \in \mathcal{U}$ and let $\widehat{\Delta}(c_0, c_1)$ be the value set to $\widehat{\Delta}$ in line 2 in a random execution of D^{A} on input (c_0, c_1).*

It holds that

$$\Pr\left[\left|\widehat{\Delta}(c_0, c_1) - \mathrm{SD}(P_0, P_1)\right| \geq \frac{\alpha-\beta}{2}\right] \leq \frac{4}{n},$$

where the probability is over the randomness of D and A.

The proof of Claim 5.3 goes as follows. First, using a union bound, we argue that with high probability for every y_i sampled by D, it holds that $\mathsf{Est}(c_0, c_1, y_i)$ is close to $\theta_{c_0,c_1}(y_i)$. Then, we use Eq. (5.6) and the Chernoff-Hoeffding bound to argue that the average of $|\mathsf{Est}(c_0, c_1, y_i)|$'s is a good approximation for the statistical distance.

Before formally proving Claim 5.3, let us use it to derive Theorem 5.5.

Proof of Theorem 5.5. Recall that in order to prove Theorem 5.5 it suffices to establish Eq. (5.5). Let $(c_0, c_1) \in \mathcal{U}$. Note that if $\left|\widehat{\Delta}(c_0, c_1) - \mathrm{SD}(P_0, P_1)\right| < \frac{\alpha-\beta}{2}$, then the distinguisher D always outputs the correct answer. Indeed, if $\mathrm{SD}(P_0, P_1) \geq \alpha$, then $\widehat{\Delta}(c_0, c_1) > \frac{\alpha+\beta}{2}$ and D outputs 1; and if $\mathrm{SD}(P_0, P_1) \leq \beta$, then $\widehat{\Delta}(c_0, c_1) < \frac{\alpha+\beta}{2}$ and D outputs 0. Hence, Claim 5.3 immediately establishes Eq. (5.5).

Lastly, since $k, \ell \in \mathsf{poly}(n)$, the pair of function α and β are efficiently computable, and A runs in polynomial time, then so does D^{A}. This completes the proof of Theorem 5.5. □

It it left to prove Claim 5.3.

Proof of Claim 5.3. For $i \in [k]$ let Y_i be the values set to y_i in a random execution of $\mathsf{D}^{\mathsf{A}}(c_0, c_1)$. Let $V_i = |\theta_{c_0,c_1}(Y_i)|$ and $\widehat{V}_i = \left|\mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0, c_1, Y_i)\right|$. The definition of $\widehat{\Delta}$ yields that

$$\begin{aligned}
\Pr\left[\left|\widehat{\Delta}(c_0, c_1) - \mathrm{SD}(P_0, P_1)\right| \geq \frac{\alpha-\beta}{2}\right] &= \Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k}\widehat{V}_i - \mathrm{SD}(P_0, P_1)\right| \geq \frac{\alpha-\beta}{2}\right] \qquad (5.8)\\
&\leq \Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} V_i - \mathrm{SD}(P_0, P_1)\right| \geq \frac{\alpha-\beta}{4}\right]\\
&\quad + \Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} V_i - \frac{1}{k}\sum_{i=1}^{k}\widehat{V}_i\right| \geq \frac{\alpha-\beta}{4}\right].
\end{aligned}$$

We bound each summand in the right-hand side of Eq. (5.8) separately.

To bound the first summand, note that by definition Y_i is always drawn from Y_{c_0,c_1}. Hence, Proposition 3.3 yields that $\mathbb{E}[V_i] = \mathrm{SD}(P_0, P_1)$ for every $i \in [k]$. Fact 3.8 now yields that

$$\begin{aligned}\Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} V_i - \mathrm{SD}(P_0,P_1)\right| \geq \frac{\alpha-\beta}{4}\right] &= \Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} V_i - \mathrm{SD}(P_0,P_1)\right| \geq \varepsilon\right] \\ &\leq 2e^{-2k\varepsilon^2} \\ &\leq \frac{2}{n},\end{aligned} \tag{5.9}$$

where the last inequality follows from the definition of k.

To bound the second summand in the right-hand side of Eq. (5.8), we apply Claim 5.2.

$$\begin{aligned}\Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} V_i - \frac{1}{k}\sum_{i=1}^{k} \widehat{V}_i\right| \geq \frac{\alpha-\beta}{4}\right] &= \Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} V_i - \frac{1}{k}\sum_{i=1}^{k} \widehat{V}_i\right| \geq \varepsilon\right] \\ &\leq \Pr\left[\exists i \in [k] \colon \left|V_i - \widehat{V}_i\right| \geq \varepsilon\right] \\ &\leq \sum_{i=1}^{k} \Pr\left[\left|V_i - \widehat{V}_i\right| \geq \varepsilon\right] \\ &\leq \sum_{i=1}^{k} \Pr\left[\left|\theta_{c_0,c_1}(Y_i) - \mathsf{Est}^{\widetilde{\mathsf{A}}}(c_0,c_1,Y_i)\right| \geq \varepsilon\right] \\ &\leq k \cdot \left(\frac{1}{kn} + \frac{\ell+1}{\sqrt[4]{p}}\right) \\ &\leq \frac{2}{n},\end{aligned} \tag{5.10}$$

where the first inequality follows since $|x-y| \geq ||x|-|y||$ for all x and y, the penultimate inequality follows from Claim 5.2, and the last inequality follows by setting $p = (k(\ell+1)n)^4 \in \mathsf{poly}(n)$.

Plugging Eqs. (5.9) and (5.10) into Eq. (5.8) completes the proof of the claim. □

6 Estimating Statistical Distance in AM ∩ coAM

In this section, we present a constant-round public-coin interactive protocol in which, given circuits C_0 and C_1 and a parameter $\Delta \in [0,1]$, a computationally unbounded prover can prove to a computationally bounded verifier that $\mathrm{SD}(C_0, C_1) \approx \Delta$, up to any arbitrary inverse-polynomial precision (and thereby proving Theorem 1.9). This immediately gives an AM as well as a coAM protocol for the Statistical Difference Problem problem $\mathrm{SDP}^{\alpha,\beta}$ for any noticeably separated α and β (which, as discussed in Section 1.1.3, were already known).

Fix a pair of circuits $C_0, C_1 : \{0,1\}^m \to \{0,1\}^n$, and define the circuit $C : \{0,1\}^{m+1} \to \{0,1\}^n$ that on input (b,x) outputs $C_b(x)$.[31] For any y, define the "pre-image set" $\mathcal{I}_y = \{(b,x) \mid C_b(x) = y\}$. We shall once more use the *imbalance* θ_y, defined in Definition 3.2. In terms that will be more

[31] The output distribution of C is exactly $\frac{C_0+C_1}{2}$ – the distribution from which the random variable Y from previous sections was drawn. In this section is will be convenient to explicitly describe this distribution as an output distribution of a circuit.

convenient for our application, for any y in the support of C, we note that θ_y may be written as follows:

$$\theta_y = \Pr_{(b,r)\sim\mathcal{I}_y}[b=1] - \Pr_{(b,r)\sim\mathcal{I}_y}[b=0] = \mathop{\mathbb{E}}_{(b,r)\sim\mathcal{I}_y}[b] - \left(1 - \mathop{\mathbb{E}}_{(b,r)\sim\mathcal{I}_y}[b]\right) = 2\cdot\mathop{\mathbb{E}}_{(b,r)\sim\mathcal{I}_y}[b] - 1.$$

Our protocol for statistical distance is based on the relation between statistical distance and statistics of the above quantity θ_y, which we proved in Proposition 3.3. Specifically, for any pair of circuits C_0 and C_1,

$$\mathrm{SD}(C_0, C_1) = \mathop{\mathbb{E}}_{y\sim C}[|\theta_y|] \tag{6.1}$$

In our protocol, the verifier will estimate this expectation by picking several y's at random from the distribution of C, and asking the prover for the values of the corresponding θ_y's. The prover is then asked to prove that the values of the θ_y's that it provided are (approximately) correct. In order to do this, we make use of the following formulation of "typical sets" that will provide us with a measure that distinguishes between values close to θ_y and those far from it.

For any y in the co-domain of C, any $\theta \in [-1,1]$ and $\delta \in [0,1]$, and any $k \in \mathbb{N}$, define the set $\mathcal{T}_y^{\theta,\delta,k}$ as:

$$\mathcal{T}_y^{\theta,\delta,k} = \left\{(b_1, x_1, \ldots, b_k, x_k) \mid C_{b_i}(x_i) = y \text{ for all } i, \text{ and } \left(2\cdot\frac{\sum_i b_i}{k} - 1\right) \in [\theta-\delta, \theta+\delta]\right\}$$

We refer to the set $\mathcal{T}_y^{\theta_y,\delta,k}$ (where θ is set to be θ_y) as the typical set corresponding to y, and sets $\mathcal{T}_y^{\theta,\delta,k}$ for values of θ close to θ_y as "nearly typical sets". We claim that, after sufficient repetition (that is, for large enough k), any nearly typical set contains a large fraction of the pre-images of y.

Proposition 6.1. *Consider any $y \in \mathrm{Supp}(C)$ and $\theta \in [-1,1]$, and let $\varepsilon = |\theta - \theta_y|$. For any $\delta \geq \varepsilon$ and $k \in \mathbb{N}$, the following holds:*

$$\frac{\left|\mathcal{T}_y^{\theta,\delta,k}\right|}{|\mathcal{I}_y|^k} \geq 1 - 2e^{-k(\delta-\varepsilon)^2/2}$$

This proposition follows from concentration of measure. We relegate the formal proofs of this and later propositions to the end of this section. Next, we claim that if θ is far from θ_y, then the set $\mathcal{T}_y^{\theta,\delta,k}$ contains very few pre-images of y under C.

Proposition 6.2. *Consider any $y \in \mathrm{Supp}(C)$ and $\theta \in [-1,1]$, and let $\varepsilon = |\theta - \theta_y|$. For any $\delta < \varepsilon$ and $k \in \mathbb{N}$, the following holds:*

$$\frac{\left|\mathcal{T}_y^{\theta,\delta,k}\right|}{|\mathcal{I}_y|^k} \leq e^{-k(\varepsilon-\delta)^2/2}$$

Thus, in order for the prover to prove that a value θ is approximately equal to θ_y for a given y, all it has to do is show that the set $\mathcal{T}_y^{\theta,\delta,k}$ is large (relative to $|I_y|^k$). In order to do this, we will be using the set lower-bound protocol of Goldwasser and Sipser [GS89] (with the tighter analysis of Aiello and Hastad [AH91]).

Lemma 6.3 ([AH91], Lemma 4.1)**.** *There is a constant-round public-coin interactive protocol that, for any set $\mathcal{S} \subseteq \{0,1\}^n$ such that membership in $\mathcal{S}$ can be computed in time t and for any $b \in \mathbb{N}$, proves the statement "$|\mathcal{S}| \geq 2^b$". More precisely, the protocol satisfies the following properties for every such set $\mathcal{S}$:*

- **Completeness:** *When interacting with the honest prover, the verifier accepts with probability at least* $1 - \frac{2^b}{|\mathcal{S}|}$
- **Soundness:** *When interacting with any (possibly cheating) prover, the verifier accepts with probability at most* $\frac{|\mathcal{S}|}{2^b}$
- **Efficiency:** *The verifier runs in time* $\mathsf{poly}(n, t)$.

Finally, note that the size of the set $\mathcal{T}_y^{\theta,\delta,k}$ has to be shown to be large *relative to* $|\mathcal{I}_y|^k$. Hence, in order to be able to use the above set lower-bound protocol for this purpose, the verifier needs to know the value of $|\mathcal{I}_y|^k$, which it may not be able to compute efficiently.[32] However, as a random variable (when y is drawn from C), the quantity $|\mathcal{I}_y|$ behaves predictably.

Let B, X and Y denote the random variables induced by drawing a bit b and a string $x \in \{0,1\}^m$ uniformly at random and y computed as $C(b, x)$. By the definition of conditional entropy, we can write the expectation of $\log|\mathcal{I}_y|$ as follows:

$$\mathop{\mathbb{E}}_{y \sim Y}[\log|\mathcal{I}_y|] = \mathrm{H}(B, X|Y).$$

So if we pick several y's independently, the mean of the $\log|\mathcal{I}_y|$'s will be concentrated around $\mathrm{H}(B, X|Y)$. We state this in a more convenient form as follows.

Proposition 6.4. *For any $t \in \mathbb{N}$, suppose $y_1, \ldots, y_t$ are independently sampled as $C(b_i, x_i)$ for b_i and x_i chosen uniformly at random. Let $\bar{\mathcal{I}} = \mathcal{I}_{y_1} \times \cdots \times \mathcal{I}_{y_t}$. Then, for any $\eta \geq 0$,*

$$\Pr\left[|\bar{\mathcal{I}}| \in \left[2^{t\cdot(\mathrm{H}(B,X|Y)-\eta)}, 2^{t\cdot(\mathrm{H}(B,X|Y)+\eta)}\right]\right] \geq 1 - 2e^{-2t\eta^2/(m+1)^2}$$

Thus, if the verifier can estimate $\mathrm{H}(B, X|Y)$ reliably, we can then implement the earlier lower-bound protocol collectively for all the y_i's sampled without knowing any of the individual $|\mathcal{I}_{y_i}|$'s, but confident that the product of all of them is approximately $2^{t\cdot\mathrm{H}(B,X|Y)}$.

We will use the prover to enable the verifier to perform this estimation, using the fact that the problem of approximating the entropy of the output of a given circuit is in NISZK [GSV99], and hence can has a constant-round interactive protocol. While $(B, X|Y)$ is not the output distribution of any circuit, computing this conditional entropy reduces to computing the entropy of just Y by the following calculation:

$$\mathrm{H}(B, X|Y) = \mathrm{H}(B, X, Y) - \mathrm{H}(Y) = \mathrm{H}(B, X) - \mathrm{H}(Y) = (m + 1) - \mathrm{H}(Y),$$

where the first equality follows the chain rule for Shannon entropy (Fact 3.7), and the second from the fact that Y is a deterministic function of (B, X). Noting that Y is the random variable corresponding to the output of the circuit C, the following lemma is now implied by the results of Goldreich, Sahai and Vadhan [GSV99].

Lemma 6.5. *There is a constant-round public-coin interactive protocol that takes as input a tuple of the form (C, h, γ), where C is a circuit, $h \in \mathbb{R}$ is an entropy estimate, and $\gamma > 0$ is a gap parameter, and has the following properties:*

[32] An alternative to deal with this would be to ask the prover to prove to the verifier that $\mathcal{I}_y$ is of a certain size using set lower-bound and upper-bound protocols (an example of the latter may be found in [For89]). However, it is unclear how to perform the upper-bound protocol with sufficiently small soundness error given the inability of the verifier to sample at will several random elements from $\mathcal{I}_y$.

- **Completeness:** *If* $\mathrm{H}(C) \in [h-\gamma, h+\gamma]$*, the verifier accepts with probability at least* 0.9 *when interacting with the honest prover.*
- **Soundness:** *If* $\mathrm{H}(C) \notin [h-3\gamma, h+3\gamma]$*, then for every (possibly cheating) prover, the verifier accepts with probability at most* 0.1*.*
- **Efficiency:** *The verifier runs in time* $\mathsf{poly}(|C|, 1/\gamma)$*.*

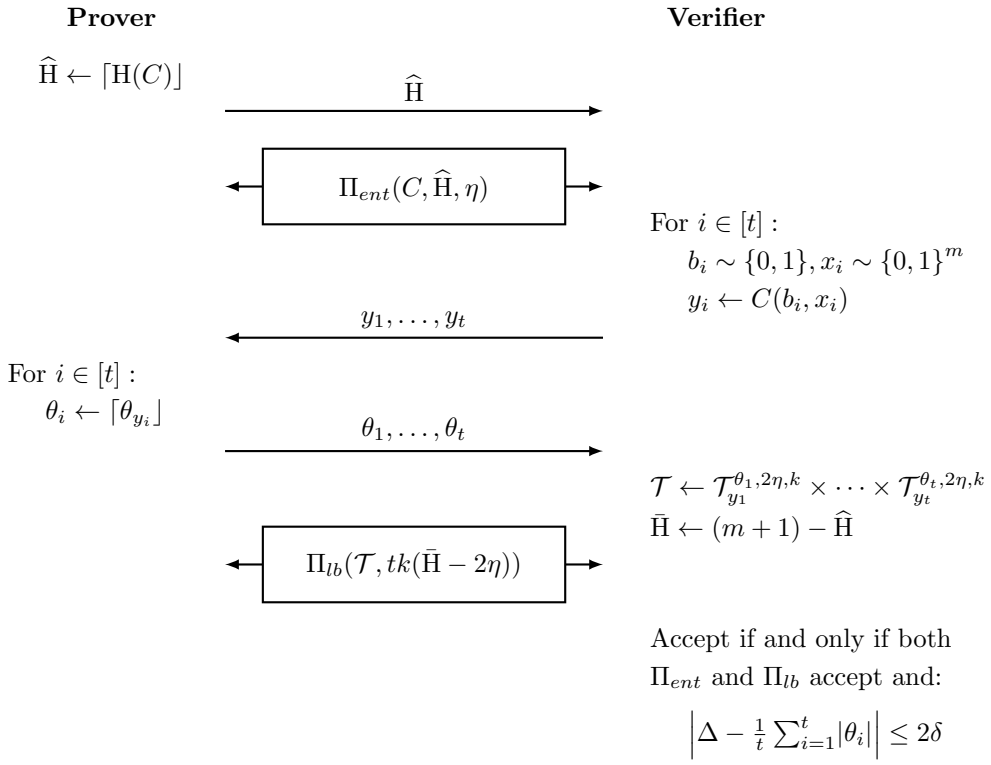
The entirety of our protocol is described formally as Protocol 6.1; Lemma 6.6 states its properties and immediately implies Theorem 1.9.

Input: Pair of circuits $C_0, C_1 \colon \{0,1\}^m \to \{0,1\}^n$, claimed distance Δ, and tolerance δ

Set $\eta = \delta^2/200$, $t = \lceil 8\ln(40)(m+1)^2/\eta^2 \rceil$, and $k = \lceil 4\ln(2t)/\eta^2 \rceil$. All numerical quantities in the protocol will be specified to within an additive error of η—to write down any $q \in \mathbb{R}$ like this uses $(\log\lceil |q| \rceil + O(\log(1/\eta))$ bits. We denote the result of rounding the quantity q in this manner by $\lceil q \rfloor$.

We will be using the following sub-protocols:

- The entropy approximation protocol from Lemma 6.5, denoted by Π_{ent}, which takes input of the form (C, k, γ).
- The set lower-bound protocol from Lemma 6.3, denoted by Π_{lb}, which takes input of the form $(\mathcal{S}, b)$.



Protocol 6.1: Estimating Statistical Distance

Lemma 6.6. *There is a constant-round public-coin interactive protocol that, given as input a pair of circuits (C_0, C_1), a claim $\Delta \in [0,1]$ for their statistical distance, and a tolerance $\delta \in [0,1]$, satisfies the following properties:*

- **Completeness:** *If $|\mathrm{SD}(C_0, C_1) - \Delta| \leq \delta$, the verifier accepts with probability at least $2/3$ when interacting with the honest prover.*
- **Soundness:** *If $|\mathrm{SD}(C_0, C_1) - \Delta| \geq 3\delta$, when interacting with any (possibly cheating) prover, the verifier accepts with probability at most $1/3$.*
- **Efficiency:** *The verifier runs in time $\mathsf{poly}(|C_0|, |C_1|, 1/\delta)$.*

That is, it proves that $\mathrm{SD}(C_0, C_1)$ is within an additive error δ of Δ, failing with probability at most $1/3$. Setting the tolerance factor δ in the above protocol to be a third of the one from Theorem 1.9 proves the latter. We now prove Lemma 6.6 by the approach outlined so far.

Proof of Lemma 6.6. We show that Protocol 6.1 satisfies the properties required by the lemma.

Completeness. Fix an input (C_0, C_1, δ), and suppose that $|\mathrm{SD}(C_0, C_1) - \Delta| \leq \delta$. We now show that the prover makes the verifier accept with high probability.

First, if the prover computes $\widehat{\mathrm{H}}$ correctly then $\left|\widehat{\mathrm{H}} - \mathrm{H}(C)\right| \leq \eta$ by the precision of our chosen representation of numbers. So by Lemma 6.5, the execution of the subprotocol Π_{ent} accepts except with probability 0.1.

Next, by Lemma 6.3 the subprotocol Π_{lb} accepts with probability at least $1 - 2^{(tk(\bar{\mathrm{H}} - 2\eta))}/|\mathcal{T}|$. So if $|\mathcal{T}|$ is more than $2^{tk(\bar{\mathrm{H}} - 3\eta/2)}$, then Π_{lb} accepts with probability more than $1 - 2^{-tk\eta/2} > 1 - 1/20$, where we use that $t \geq 1$ and $k\eta \geq 1/\eta \geq 200$. We claim that, if the θ_i's are reported correctly, this happens with high probability. Recall that $\bar{\mathrm{H}} = (m+1) - \widehat{\mathrm{H}}$ is supposed to be roughly $\mathrm{H}(B, X|Y)$.

Claim 6.1. *Suppose that for all $i \in [t]$, we have $\theta_i = \lceil \theta_{y_i} \rfloor$, and that $\widehat{\mathrm{H}} = \lceil \mathrm{H}(C) \rfloor$. Then,*

$$\Pr\left[|\mathcal{T}| < 2^{tk(\bar{\mathrm{H}} - 3\eta/2)}\right] \leq \frac{1}{20}.$$

Proof of Claim 6.1. This follows from Proposition 6.1 and Proposition 6.4. Recall that $\mathcal{T}$ is defined as $\mathcal{T}_{y_1}^{\theta_1, 2\eta, k} \times \cdots \times \mathcal{T}_{y_t}^{\theta_t, 2\eta, k}$, that for each i we assume that $|\theta_i - \theta_{y_i}| \leq \eta$, that $t \geq 1$, and that $k \geq 4\ln(2t)/\eta^2$. By Proposition 6.1, we have:

$$\begin{aligned} \frac{|\mathcal{T}|}{(|\mathcal{I}_{y_1}| \cdots |\mathcal{I}_{y_t}|)^k} \geq \left(1 - 2e^{-k(2\eta - \eta)^2/2}\right)^t &\geq 1 - 2te^{-k\eta^2/2} \\ &= 1 - e^{-k\eta^2/2 + \ln(2t)} \geq 1 - e^{-2\ln(2t)} \geq 3/4 \geq 2^{-tk\eta/4}. \end{aligned} \tag{6.2}$$

Let $\bar{\mathcal{I}} = \mathcal{I}_{y_1} \times \cdots \times \mathcal{I}_{y_t}$. Proposition 6.4 now lets us approximate $\left|\bar{\mathcal{I}}\right|^k$ as follows.

$$\begin{aligned} \Pr\left[\left|\bar{\mathcal{I}}\right|^k \geq 2^{tk(\bar{\mathrm{H}} - 5\eta/4)}\right] &\geq \Pr\left[\left|\bar{\mathcal{I}}\right|^k \geq 2^{tk(\mathrm{H}(B,X|Y) - \eta/4)}\right] \\ &= \Pr\left[\left|\bar{\mathcal{I}}\right| \geq 2^{t(\mathrm{H}(B,X|Y) - \eta/4)}\right] \\ &\geq 1 - 2e^{-t\eta^2/8(m+1)^2} \geq 1 - 1/20, \end{aligned} \tag{6.3}$$

where the first inequality follows from the precision of $\bar{\mathrm{H}}$ (since $\widehat{H} = \lceil H(C) \rfloor$, we have that $\left|\bar{H} - H(B, X|Y)\right|$ is at most η), the second from Proposition 6.4, and the last from the chosen setting of t. Eqs. (6.2) and (6.3) together imply the claim. □

Claim 6.1 along with the argument just preceding it now implies that the probability that the execution of Π_{lb} rejects is at most $(1/20 + 1/20) = 1/10$.

And third, we need to argue that the final check of the verifier passes. As the prover is honest, the mean $\sum_i |\theta_i|/t$ is within $\eta = \delta^2/200$ of $\sum_i |\theta_{y_i}|/t$. By Eq. (6.1), $\mathbb{E}_{y_i \sim C}[|\theta_{y_i}|] = \mathrm{SD}(C_0, C_1)$ for each i. And by our hypothesis, $|\Delta - \mathrm{SD}(C_0, C_1)| \leq \delta$. Using these facts and by the Chernoff-Hoeffding bound (Fact 3.8), noting that the y_i's sent by the verifier are indeed sampled from C, we have:

$$\begin{aligned}
\Pr\left[\left|\Delta - \frac{1}{t}\sum_i |\theta_i|\right| > 2\delta\right] &\leq \Pr\left[\left|\Delta - \frac{1}{t}\sum_i |\theta_{y_i}|\right| > \frac{3\delta}{2}\right] \\
&\leq \Pr\left[\left|\mathrm{SD}(C_0, C_1) - \frac{1}{t}\sum_i |\theta_{y_i}|\right| > \frac{\delta}{2}\right] \\
&\leq 0.1, \qquad (6.4)
\end{aligned}$$

where the last inequality follows from the choice of t.

Thus, we have shown that if this prover strategy is used, then the execution of Π_{ent} rejects with probability at most 0.1, that of Π_{lb} with probability at most 0.1, and the final check of the verifier fails with probability at most 0.1. By the union bound, the whole protocol rejects with probability less than 1/3, proving completeness.

Soundness. Now suppose that $|\mathrm{SD}(C_0, C_1) - \Delta| \geq 3\delta$. First, consider the case where the $\widehat{\mathrm{H}}$ sent by the prover is such that $\left|\widehat{\mathrm{H}} - \mathrm{H}(C)\right| \geq 3\eta$. In this case, by Lemma 6.5, the probability that $\Pi_{ent}(C, \widehat{\mathrm{H}}, \eta)$ accepts is less than 0.1.

Next, consider the case where $\left|\widehat{\mathrm{H}} - \mathrm{H}(C)\right| < 3\eta$, but the estimate $\sum_i |\theta_i|/t$ differs substantially from $\sum_i |\theta_{y_i}|/t$ (because the prover reported wrong values for θ_i). In this case, we would like to show that the execution of Π_{lb} rejects with high probability.

Specifically, suppose we have:

$$\frac{1}{t}\sum_i |\theta_i| - \frac{1}{t}\sum_i |\theta_{y_i}| \geq \frac{\delta}{2}$$

This implies the following:

$$\sum_i |\theta_i - \theta_{y_i}| \geq \sum_i |\theta_i| - \sum_i |\theta_{y_i}| \geq \frac{\delta t}{2} \qquad (6.5)$$

By Lemma 6.3, Π_{lb} accepts with probability at most $|\mathcal{T}|/2^{tk(\bar{\mathrm{H}} - 2\eta)}$. Recall that $\mathcal{T} = \mathcal{T}_{y_1}^{\theta_1, 2\eta, k} \times \cdots \times \mathcal{T}_{y_t}^{\theta_t, 2\eta, k}$. Let $\bar{\mathcal{I}} = \mathcal{I}_{y_1} \times \cdots \times \mathcal{I}_{y_t}$. We show that the condition (6.5) implies that $\mathcal{T}$ is much smaller than $\left|\bar{\mathcal{I}}\right|^k$; this is because (6.5) implies that a number of the θ_i's are far from the corresponding θ_{y_i}'s, and thus the sets formed using them are not typical.

Claim 6.2. *If $\sum_i |\theta_i - \theta_{y_i}| \geq \delta t/2$, then:*

$$\frac{|\mathcal{T}|}{\left|\bar{\mathcal{I}}\right|^k} \leq e^{-2tk\delta^2/25}$$

Proof of Claim 6.2. For each $i \in [t]$ such that $|\theta_i - \theta_{y_i}| > 2\eta$, Proposition 6.2 implies that:

$$\left|\mathcal{T}_{y_i}^{\theta_i,2\eta,k}\right| \leq e^{-k\left(|\theta_i-\theta_{y_i}|-2\eta\right)^2/2} \cdot |\mathcal{I}_{y_i}|^k.$$

Noting that $\mathcal{T}_{y_i}^{\theta_i,2\eta,k}$ is at most as large as $|\mathcal{I}_{y_i}|^k$, we multiply this bound by a correcting factor of $e^{k(2\eta)^2/2}$ to handle the case of $|\theta_i - \theta_{y_i}| \leq 2\eta$, so that the following holds for all $i \in [t]$,

$$\left|\mathcal{T}_{y_i}^{\theta_i,2\eta,k}\right| \leq e^{-k\left(|\theta_i-\theta_{y_i}|-2\eta\right)^2/2} \cdot e^{k(2\eta)^2/2} \cdot |\mathcal{I}_{y_i}|^k.$$

Taking the product of this over all i gives us:

$$\frac{|\mathcal{T}|}{|\bar{\mathcal{I}}|^k} \leq \exp\left(-k\sum_i(\theta_i - \theta_{y_i})^2/2 + 2k\eta\sum_i|\theta_i - \theta_{y_i}|\right)$$

By the Cauchy-Schwarz inequality and (6.5), we have $\sum_i(\theta_i - \theta_{y_i})^2 \geq \delta^2 t/4$. Further, we know that $\sum_i|\theta_i - \theta_{y_i}| \leq 2t$ (since $\theta_i, \theta_{y_i} \in [-1, 1]$). Together with these and the fact that $\eta = \delta^2/200$, the above expression gives us:

$$\frac{|\mathcal{T}|}{|\bar{\mathcal{I}}|^k} \leq \exp\left(-kt\delta^2/8 + 4kt\delta^2/200\right) \leq e^{-2kt\delta^2/25}.$$

□

Under the other part of our current hypothesis – that $\left|\widehat{\mathrm{H}} - \mathrm{H}(C)\right| < 3\eta$ (which implies that $\left|\bar{\mathrm{H}} - \mathrm{H}(B, X|Y)\right| < 3\eta$) – we show that $\bar{\mathcal{I}}$ is, most of the time, not very large.

Claim 6.3. *Suppose that* $\left|\bar{\mathrm{H}} - \mathrm{H}(B, X|Y)\right| < 3\eta$. *Then,*

$$\Pr\left[|\bar{\mathcal{I}}| \geq 2^{t(\bar{\mathrm{H}}+7\eta/2)}\right] \leq \frac{1}{20}$$

Proof of Claim 6.3. This follows from Proposition 6.4. We have:

$$\Pr\left[|\bar{\mathcal{I}}| \geq 2^{t(\bar{\mathrm{H}}+7\eta/2)}\right] \leq \Pr\left[|\bar{\mathcal{I}}| \geq 2^{t(\mathrm{H}(B,X|Y)+\eta/2)}\right] \leq 2e^{-t\eta^2/2(m+1)^2} \leq \frac{1}{20}$$

where the first inequality is from the accuracy of $\bar{\mathrm{H}}$, the second from Proposition 6.4, and the last from the value of t. □

Together, Claims 6.2 and 6.3 imply the following. If $\left|\widehat{\mathrm{H}} - \mathrm{H}(C)\right| < 3\eta$ then, except with probability 1/20 over the choice of the y_i's, unless the estimate $\sum_i|\theta_i|/t$ is within $\delta/2$ of $\sum_i|\theta_{y_i}|/t$, the subprotocol Π_{lb} accepts with probability at most:

$$\frac{|\mathcal{T}|}{2^{tk(\bar{\mathrm{H}}-2\eta)}} \leq \frac{e^{-2tk\delta^2/25} \cdot |\bar{\mathcal{I}}|^k}{2^{tk(\bar{\mathrm{H}}-2\eta)}} \leq \frac{e^{-2tk\delta^2/25} \cdot 2^{tk(\bar{\mathrm{H}}+7\eta/2)}}{2^{tk(\bar{\mathrm{H}}-2\eta)}} \leq \frac{1}{20}$$

where the first inequality follows from Claim 6.2, the second from our conditioning that the event from Claim 6.3 does not happen, and the last from the values of the quantities involved.

The case we are left with is where $\sum_i|\theta_i|/t$ is within $\delta/2$ of $\sum_i|\theta_{y_i}|/t$. The final check by the verifier is whether $\sum_i|\theta_i|/t$ is within 2δ of Δ. For this to happen, as $|\Delta - \mathrm{SD}(C_0, C_1)| \geq 3\delta$, the mean $\sum_i|\theta_{y_i}|/t$ has to be more than $\delta/2$ away from $\mathrm{SD}(C_0, C_1)$. For random y_i's drawn from C, as calculated in Eq. (6.4) in the proof of completeness, this happens with probability less than 0.1.

We summarize the argument for soundness as follows:

1. If the prover sends $\widehat{\mathrm{H}}$ that is 3η-far from $\mathrm{H}(C)$, the verifier rejects except with probability at most 0.1.

2. Otherwise, except with probability at most 1/20 over the choice of the y_i's, unless $\sum_i |\theta_i|/t$ is within $\delta/2$ of $\sum_i |\theta_{y_i}|/t$, the verifier rejects except with probability at most 1/20.

3. Also, except with probability at most 0.1 over the choice of the y_i's, if $\sum_i |\theta_i|/t$ is within $\delta/2$ of $\sum_i |\theta_{y_i}|/t$, the verifier rejects.

Thus, the total probability of the verifier accepting is at most $(0.1 + 1/20 + 1/20 + 0.1) < 1/3$, as required.

Efficiency. The running time of the verifier is the sum of those of the verifiers in the calls to Π_{ent} and Π_{lb}, and $\mathsf{poly}(t, 1/\delta)$ (for sampling and the final check). Membership in the set $\mathcal{T}$ can be verified using its definition in time $\mathsf{poly}(|C|, k, t)$. The entire running time may now be verified to be $\mathsf{poly}(|C_0|, |C_1|, 1/\delta)$, as required.

While the protocol as written is private-coin, note that, since Π_{ent} and Π_{lb} are both public-coin, the only instance where the verifier's coins are not sent over is when it sends the y_i's to the prover. This is remedied by having the verifier send the (b_i, x_i)'s to the prover instead, and noting that this does not affect the soundness of the protocol. □

6.1 Proofs of Intermediates

We complete this section by proving the intermediate propositions and lemmas used in the proof of Lemma 6.6.

Proof of Proposition 6.1. The proposition is proven using the additive Chernoff bound. Under the uniform distribution over $\mathcal{I}_y$, we have $\Pr_{(b,x)\sim\mathcal{I}_y}[b = 1] = (1 + \theta_y)/2$. So if we take many independent samples (b, x) from this distribution and looked at the empirical mean of the b's, it would be concentrated around $(1 + \theta_y)/2$. Then, if θ is close to θ_y, this empirical mean is, with good probability, close to θ as well.

We abuse notation slightly and write $b^k \sim \mathcal{I}_y^k$ to indicate the vector $(b_1, \ldots, b_k)$ obtained by sampling $(b_i, x_i)_{i\in[k]}$ uniformly and independently from $\mathcal{I}_y$ and dropping the x's. Observe that:

$$\begin{aligned}
\frac{\left|\mathcal{T}_y^{\theta,\delta,k}\right|}{|\mathcal{I}_y|^k} &= \Pr_{b^k\sim\mathcal{I}_y^k}\left[2\frac{\sum b_i}{k} - 1 \in [\theta - \delta, \theta + \delta]\right] \\
&\geq \Pr_{b^k\sim\mathcal{I}_y^k}\left[\frac{\sum b_i}{k} \in \left[\frac{1+\theta_y}{2} - \frac{\delta-\varepsilon}{2}, \frac{1+\theta_y}{2} + \frac{\delta-\varepsilon}{2}\right]\right] \\
&\geq 1 - 2\cdot e^{-k(\delta-\varepsilon)^2/2}
\end{aligned}$$

where the first inequality follows from the observation that an interval of size δ around θ contains an interval of size $(\delta - \varepsilon)$ around θ_y, and the second inequality follows from the Chernoff-Hoeffding Bound (Fact 3.8). □

Proof of Proposition 6.2. The proof of this proposition also uses the Chernoff-Hoeffding bound. The central idea here is that this claimed typical set is almost disjoint from the actual typical set, and most probability mass lies inside the actual typical set.

Consider the case where $\theta_y > \theta$; the argument for the other case is identical. We abuse notation slightly and write $b^k \sim I_y^k$ to indicate the vector $(b_1, \ldots, b_k)$ obtained by sampling $(b_i, x_i)_{i\in[k]}$ uniformly from $\mathcal{I}_y$ and dropping the r's. Let $\bar{b} = \sum b_i/k$. Observe that:

$$\begin{aligned}
\frac{\left|\mathcal{T}_y^{\theta,\delta,k}\right|}{|\mathcal{I}_y|^k} &= \Pr_{b^k\sim\mathcal{I}_y^k}\left[2\bar{b} - 1 \in [\theta-\delta, \theta+\delta]\right] \\
&\leq \Pr_{b^k\sim\mathcal{I}_y^k}\left[\bar{b} \leq \frac{1+\theta}{2} + \frac{\delta}{2}\right] \\
&= \Pr_{b^k\sim\mathcal{I}_y^k}\left[\bar{b} \leq \frac{1+\theta_y}{2} - \frac{\varepsilon-\delta}{2}\right] \\
&\leq e^{-k(\varepsilon-\delta)^2/2}
\end{aligned}$$

where the first equality follows from the definition of the distribution, the second equality follows from the definition of ε and the assumption that $\theta_y > \theta$, and the final inequality follows from the Chernoff-Hoeffding bound (Fact 3.8). □

Proof of Proposition 6.4. We prove this again using the Chernoff-Hoeffding bound on independent instances of the random variable $\log|\mathcal{I}_y|$, which is contained in $[0, m+1]$. This is done as follows.

$$\Pr\left[|\bar{\mathcal{I}}| \in \left[2^{t(\mathrm{H}(B,X|Y)-\eta)}, 2^{t(\mathrm{H}(B,X|Y)+\eta)}\right]\right] = \Pr\left[\frac{\log|\bar{\mathcal{I}}|}{t} \in [\mathrm{H}(B,X|Y)-\eta, \mathrm{H}(B,X|Y)+\eta]\right].$$

We now note that $\log|\bar{\mathcal{I}}|/t$ is simply the mean of several i.i.d. variables:

$$\frac{\log|\bar{\mathcal{I}}|}{t} = \frac{1}{t}\sum_{i=1}^{t}\log|\mathcal{I}_{y_i}|,$$

where for each i we know that $\log|\mathcal{I}_{y_i}|$ is contained in $[0, m+1]$ (as $\mathcal{I}_{y_i}$ is a set of (b,x) where b is a bit and x is of length m), and its expectation is $\mathrm{H}(B, X|Y)$. Applying the Chernoff-Hoeffding bound (Fact 3.8), after scaling the values withing the probability expression down by $(m+1)$, now gives us what we want:

$$\Pr\left[\frac{\log|\bar{\mathcal{I}}|}{t} \in [\mathrm{H}(B,X|Y)-\eta, \mathrm{H}(B,X|Y)+\eta]\right] \geq 1 - 2e^{-2t\eta^2/(m+1)^2}$$

□

Proof of Lemma 6.5. The protocol is a straightforward combination of the NISZK and coNISZK protocols for the Entropy Approximation and its complement from Goldreich et al [GSV99]. We make use of the following lemma from their work (with the inequalities made non-strict for ease of use).

Lemma 6.7 ([GSV99, Lemma 3.2])**.** *There is a polynomial-time computable function that takes input $(1^s, C, h)$, where C is a circuit, $h \in \mathbb{R}^+$, and $s \in \mathbb{N}$, and produces a circuit C' that outputs ℓ bits such that:*

1. *If $\mathrm{H}(C) \geq h+1$, then $\mathrm{SD}(C', U_\ell) \leq 2^{-s}$, where U_ℓ is the uniform distribution on $\{0,1\}^\ell$; and*

2. *If* $\mathrm{H}(C) \leq h-1$, *then* $\frac{|\mathrm{Supp}(C')|}{2^\ell} \leq 2^{-s}$

We start with a constant-round private-coin protocol to approximate entropy that works as follows given input (C, k, γ):

1. Let $s = 3$, let $t = \lceil 1/\gamma \rceil$, and let C^t denote the circuit obtained by concatenating t copies of C.
2. The verifier invokes the function from Lemma 6.7 with the input $(1^s, C^t, tk+2)$ to get a circuit C_{upper} that outputs ℓ_{upper} bits.
3. The verifier picks a random bit b. If $b = 0$, it samples a random output of C_{upper} and sends it to the prover. Else it samples a uniform string from $\{0,1\}^{\ell_{upper}}$ and sends it to the prover.
4. The prover responds with a bit b'. If $b' \neq b$ the verifier rejects.
5. The verifier then invokes the function from Lemma 6.7 with the input $(1^s, C^t, tk-2)$ to get a circuit C_{lower} that takes n_{lower} bits as input and outputs ℓ_{lower} bits.
6. The verifier picks a uniformly random string y from $\{0,1\}^{\ell_{lower}}$ and sends it to the prover.
7. The prover responds with a string $x \in \{0,1\}^{n_{lower}}$.
8. If $C_{lower}(x) = y$, the verifier accepts. Else it rejects.

That the running time of the verifier above is $\mathsf{poly}(|C|, 1/\gamma)$ may be verified by inspection, noting the efficient computability of the function from Lemma 6.7.

To show completeness, suppose $\mathrm{H}(C) \in [k-\gamma, k+\gamma]$. For the sake of simplicity, suppose that $1/\gamma \in \mathbb{N}$ (so that $t = 1/\gamma$); the arguments for the more general case are identical. Then, $\mathrm{H}(C^t) \in [t(k-\gamma), t(k+\gamma)] = [tk-1, tk+1]$. By Lemma 6.7, this implies two things for the circuits constructed in our protocol:

$$\frac{|\mathrm{Supp}(C_{upper})|}{2^{\ell_{upper}}} \leq \frac{1}{8}$$
$$\mathrm{SD}(C_{lower}, U_{\ell_{lower}}) \leq \frac{1}{8}$$

Together with Proposition 3.1, these properties imply, respectively, that:

$$\mathrm{SD}(C_{upper}, U_{\ell_{upper}}) \geq \frac{7}{8} \tag{6.6}$$
$$\frac{|\mathrm{Supp}(C_{lower})|}{2^{\ell_{lower}}} \geq \frac{7}{8} \tag{6.7}$$

Eq. (6.6) and Proposition 3.4 imply that there is a prover strategy (specifically, to send the maximal likelihood bit) such that the probability that $b' \neq b$ in step 4 of our protocol is at most $1/2 - (7/8)/2 = 1/16$. And Eq. (6.7) implies that the probability that the prover is not able to produce a valid x in step 7 is at most $1/8$. Thus, the total probability that the verifier rejects is less than $3/16$.

To show soundness, first consider the case where $\mathrm{H}(C) > k + 3\gamma$. This implies that $\mathrm{H}(C^t)$ is at least $(tk+3)$. By the guarantees of Lemma 6.7, this means that $\mathrm{SD}(C_{upper}, U_{\ell_{upper}}) \leq 1/8$, which implies, together with Proposition 3.4, that the probability that the verifier does not reject in step 4 is at most $1/2 + (1/8)/2 = 9/16$.

On the other hand, if $\mathrm{H}(C) < k - 3\gamma$, then $\mathrm{H}(C^t)$ is at most $(tk - 3)$. In this case, Lemma 6.7 impies that $|\mathrm{Supp}(C_{lower})|/2^{\ell_{lower}} \leq 1/8$. Thus, the probability that the prover is able to produce a valid pre-image x in step 7 is at most $1/8$.

Overall, we have a protocol with completeness at least $1 - 3/16 = 13/16$ and soundness error at most $9/16$. By repetition in parallel, with an $O(1)$ blowup in the complexity of the verifier, both completeness and soundness error can be made smaller than 0.1, giving the desired parameters. And finally, it can be made public-coin while retaining a constant number rounds and an efficient verifier using the results of Goldwasser and Sipser [GS89]. □

Acknowledgments

We thank Andrej Bogdanov, Serge Fehr, Oded Goldreich, Siyao Guo, Guy Rothblum, Salil Vadhan and Serge Vaudenay for helpful discussions. We also thank Oded for useful comments on the writeup. The first author thanks Yury Polyanskiy for teaching an enlightening course on information theory.

References

[AARV17] Benny Applebaum, Barak Arkis, Pavel Raykov, and Prashant Nalini Vasudevan. Conditional disclosure of secrets: Amplification, closure, amortization, lower-bounds, and separations. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part I*, volume 10401 of *Lecture Notes in Computer Science*, pages 727–757. Springer, 2017.

[AGGM06] Adi Akavia, Oded Goldreich, Shafi Goldwasser, and Dana Moshkovitz. On basing one-way functions on np-hardness. In Jon M. Kleinberg, editor, *Symposium on Theory of Computing*, pages 701–710. ACM, 2006.

[AH91] William Aiello and Johan Hastad. Statistical Zero-knowledge Languages can be recognized in two rounds. *Journal of Computer and System Sciences*, 42(3):327–345, 1991.

[BB15] Andrej Bogdanov and Christina Brzuska. On basing size-verifiable one-way functions on np-hardness. In *TCC*, pages 1–6, 2015.

[BBF16] Zvika Brakerski, Christina Brzuska, and Nils Fleischhacker. On statistically secure obfuscation with approximate correctness. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part II*, pages 551–578, 2016.

[BBM11] Nayantara Bhatnagar, Andrej Bogdanov, and Elchanan Mossel. The computational complexity of estimating MCMC convergence time. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 14th International Workshop, APPROX 2011, and 15th International Workshop, RANDOM 2011, Princeton, NJ, USA, August 17-19, 2011. Proceedings*, pages 424–435, 2011.

[BCH+17] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the power of statistical zero knowledge. In *FOCS*, 2017.

[BDRV18] Itay Berman, Akshay Degwekar, Ron D. Rothblum, and Prashant Nalini Vasudevan. Multi-collision resistant hash functions and their applications. In *EUROCRYPT*, 2018.

[BDV17] Nir Bitansky, Akshay Degwekar, and Vinod Vaikuntanathan. Structure vs hardness through the obfuscation lens. *CRYPTO*, 2017.

[BG03] Michael Ben-Or and Danny Gutfreund. Trading help for interaction in statistical zero-knowledge proofs. *J. Cryptology*, 16(2):95–116, 2003.

[BHZ87] Ravi B. Boppana, Johan Håstad, and Stathis Zachos. Does co-np have short interactive proofs? *Inf. Process. Lett.*, 25(2):127–132, 1987.

[BKP18] Nir Bitansky, Yael Tauman Kalai, and Omer Paneth. Multi-collision resistance: a paradigm for keyless hash functions. In *STOC*, 2018.

[BL13] Andrej Bogdanov and Chin Ho Lee. Limits of provable security for homomorphic encryption. In *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, pages 111–128, 2013.

[Cam86] Lucien Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York, NY, 1986.

[CCKV08] André Chailloux, Dragos Florin Ciocan, Iordanis Kerenidis, and Salil P. Vadhan. Interactive and noninteractive zero knowledge are equivalent in the help model. In *Theory of Cryptography, Fifth Theory of Cryptography Conference, TCC 2008, New York, USA, March 19-21, 2008.*, pages 501–534, 2008.

[CGVZ18] Yi-Hsiu Chen, Mika Göös, Salil P. Vadhan, and Jiapeng Zhang. A tight lower bound for entropy flattening. In *CCC*, 2018.

[DNR04] Cynthia Dwork, Moni Naor, and Omer Reingold. Immunizing encryption schemes from decryption errors. In Christian Cachin and Jan Camenisch, editors, *EUROCRYPT*, volume 3027 of *Lecture Notes in Computer Science*, pages 342–360. Springer, 2004.

[FGM+89] Martin Fürer, Oded Goldreich, Yishay Mansour, Michael Sipser, and Stathis Zachos. On completeness and soundness in interactive proof systems. *Advances in Computing Research*, 5:429–442, 1989.

[For89] Lance Fortnow. The complexity of perfect zero-knowledge. *Advances in Computing Research*, 5:327–343, 1989.

[FV17] Serge Fehr and Serge Vaudenay. Personal Communication, 2017.

[Gol90] Oded Goldreich. A note on computational indistinguishability. *Inf. Process. Lett.*, 34(6):277–281, 1990.

[Gol17] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

[GS89] Shafi Goldwasser and Michael Sipser. Private coins versus public coins in interactive proof systems. *Advances in Computing Research*, 5:73–90, 1989.

[GSV98] Oded Goldreich, Amit Sahai, and Salil Vadhan. Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. In *STOC*, 1998.

[GSV99] Oded Goldreich, Amit Sahai, and Salil P. Vadhan. Can statistical zero knowledge be made non-interactive? or on the relationship of SZK and NISZK. In Michael J. Wiener, editor, *CRYPTO*, volume 1666 of *Lecture Notes in Computer Science*, pages 467–484. Springer, 1999.

[GV99] Oded Goldreich and Salil P. Vadhan. Comparing entropies in statistical zero knowledge with applications to the structure of SZK. In *CCC*, 1999.

[GV11] Oded Goldreich and Salil P. Vadhan. On the complexity of computational problems regarding distributions. In Oded Goldreich, editor, *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation - In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, volume 6650 of *Lecture Notes in Computer Science*, pages 390–405. Springer, 2011.

[GVW02] Oded Goldreich, Salil Vadhan, and Avi Wigderson. On interactive proofs with a laconic prover. *Computational Complexity*, 11(1-2):1–53, 2002.

[HR05] Thomas Holenstein and Renato Renner. One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption. In *CRYPTO*, pages 478–493, 2005.

[IL89] Russell Impagliazzo and Michael Luby. One-way functions are essential for complexity based cryptography. In *STOC*, pages 230–235, 1989.

[KNY17] Ilan Komargodski, Moni Naor, and Eylon Yogev. White-box vs. black-box complexity of search problems: Ramsey and graph property testing. In *FOCS*, 2017.

[KNY18] Ilan Komargodski, Moni Naor, and Eylon Yogev. Collision resistant hashing for paranoids: Dealing with multiple collisions. In *EUROCRYPT*, pages 162–194, 2018.

[KY18] Ilan Komargodski and Eylon Yogev. On distributional collision resistant hashing. In *CRYPTO*, 2018.

[LZ17] Shachar Lovett and Jiapeng Zhang. On the impossibility of entropy reversal, and its application to zero-knowledge proofs. In *TCC*, 2017.

[NR06] Moni Naor and Guy N. Rothblum. Learning to impersonate. In *ICML*, pages 649–656, 2006.

[Ost91] Rafail Ostrovsky. One-way functions, hard on average problems, and statistical zero-knowledge proofs. In *Structure in Complexity Theory Conference*, pages 133–138, 1991.

[OV08] Shien Jin Ong and Salil P. Vadhan. An equivalence between zero knowledge and commitments. In *TCC*, pages 482–500, 2008.

[OW93] Rafail Ostrovsky and Avi Wigderson. One-way fuctions are essential for non-trivial zero-knowledge. In *ISTCS*, pages 3–17, 1993.

[PW17] Yury Polyanskiy and Yihong Wu. Lecture notes on information theory. Available at: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf, 2017.

[SV03] Amit Sahai and Salil Vadhan. A complete problem for statistical zero knowledge. *Journal of the ACM (JACM)*, 50(2):196–249, 2003.

[Top00] Flemming Topsøe. Some inequalities for information divergence and related measures of discrimination. *IEEE Transactions on Information Theory*, 46(4):1602–1609, July 2000.

[Vad99] Salil Pravin Vadhan. *A study of statistical zero-knowledge proofs*. PhD thesis, Massachusetts Institute of Technology, 1999.

[Yeh16] Amir Yehudayoff. Pointer chasing via triangular discrimination. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:151, 2016.

A Triangular Discrimination Inequalities

Proposition A.1 ([Top00])**.** *For distributions P, Q it holds that:*

$$\mathrm{SD}(P,Q)^2 \leq \mathrm{TD}(P,Q) \leq \mathrm{SD}(P,Q).$$

Proof. Let $\mathcal{Y}$ be the union of the supports of P and Q. Observe that,

$$\mathrm{TD}(P,Q) = \frac{1}{2}\sum_{y\in\mathcal{Y}} \frac{(P_y - Q_y)^2}{P_y + Q_y} = \mathop{\mathbb{E}}_{y\sim(\frac{1}{2}P+\frac{1}{2}Q)}\left[\left(\frac{P_y - Q_y}{P_y + Q_y}\right)^2\right],$$

and

$$\mathrm{SD}(P,Q) = \frac{1}{2}\sum_{y\in\mathcal{Y}} |P_y - Q_y| = \mathop{\mathbb{E}}_{y\sim(\frac{1}{2}P+\frac{1}{2}Q)}\left[\left|\frac{P_y - Q_y}{P_y + Q_y}\right|\right].$$

It follows that,

$$\begin{aligned}\mathrm{TD}(P,Q) - \mathrm{SD}(P,Q)^2 &= \mathop{\mathbb{E}}_{y\sim(\frac{1}{2}P+\frac{1}{2}Q)}\left[\left(\frac{P_y - Q_y}{P_y + Q_y}\right)^2\right] - \mathop{\mathbb{E}}_{y\sim(\frac{1}{2}P+\frac{1}{2}Q)}\left[\left|\frac{P_y - Q_y}{P_y + Q_y}\right|\right]^2 \\ &= \mathrm{Var}\left(\left|\frac{P_y - Q_y}{P_y + Q_y}\right|\right) \geq 0.\end{aligned}$$

Hence, that $\mathrm{SD}(P,Q)^2 \leq \mathrm{TD}(P,Q)$ follows from the non-negativity of variance. That $\mathrm{TD}(P,Q) \leq \mathrm{SD}(P,Q)$ follows from the fact that $\left|\frac{P_y-Q_y}{P_y+Q_y}\right| \geq \left(\frac{P_y-Q_y}{P_y+Q_y}\right)^2$. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il