Statistical Difference Beyond the Polarizing Regime

Itay Berman[1](✉), Akshay Degwekar[1], Ron D. Rothblum[2], and Prashant Nalini Vasudevan[3]

[1] MIT, Cambridge, USA
{itayberm,akshayd}@mit.edu
[2] Technion, Haifa, Israel
rothblum@cs.technion.ac.il
[3] UC Berkeley, Berkeley, USA
prashvas@berkeley.edu

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

This is an extended abstract. The full version of this paper, with formal statements of the theorems and proofs, may be found at: https://eccc.weizmann.ac.il/report/2019/038/.

I. Berman and A. Degwekar—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.

R. D. Rothblum—This research was supported in part by the Israeli Science Foundation (Grant No. 1262/18).

P. N. Vasudevan—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).

D. Hofheinz and A. Rosen (Eds.): TCC 2019, LNCS 11892, pp. 311–332, 2019.
https://doi.org/10.1007/978-3-030-36033-7_12

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.

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*

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

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

[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.

(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.

1.1 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 Sect. 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*. To start with, 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.

[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.

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* 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/poly)-separated *the set of all pairs of functions that are $(1/p)$-separated for some polynomial p. Similarly, we denote by* (1/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/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 seems inherent. 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/poly)-separated.

[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.

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)]):

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

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).

[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 $\mathrm{TD}(P,Q) > \mathrm{TD}(P',Q')$.

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 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. 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 Theorem 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* (α, β) *that are* $(1/\mathsf{poly})$*-separated, 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.

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]

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

[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 $|f^{-1}(y)|$.

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 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. In the process of 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 Sect. 2.3.3 for details).

[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 $(b_1^\sigma, \dots, b_\ell^\sigma)$ 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 $(b_1^\sigma, \dots, b_\ell^\sigma)$ and r^σ and outputs $(C_{b_1^\sigma}, \dots, C_{b_\ell^\sigma}, r^\sigma)$. 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 $\mathrm{SD}(D_0, D_1) \leq 2^{-k}$.

[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 $\mathrm{Fidelity}(P, Q) = \sum_y \sqrt{P_y \cdot Q_y}$.

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]

2 Techniques

We begin in Sect. 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 Sect. 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 Sect. 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 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

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

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]

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

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

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

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 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,

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

[15] In Sect. 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.

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 $\frac{1}{2}P_0 + \frac{1}{2}P_1$.)

Equation (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 $y_1, \ldots, y_t$ independently from $\frac{1}{2}P_0 + \frac{1}{2}P_1$.
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 conditioned 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.

The above proof strategy also yields a new proof for the strengthening of [Gol90] by Naor and Rothblum [NR06].[16] See Theorem 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

[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.

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 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

[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.

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 Sect. 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:

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

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

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

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,

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

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 $\left|\mathcal{I}_y\right|^k$. 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,

[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 $\mathrm{H}(X|Y) = \mathbb{E}_{y \sim Y}\left[\mathrm{H}(X|Y = y)\right]$.

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

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 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 $\text{TDP}^{\alpha,\beta}$ and $\text{JSP}^{\alpha,\beta}$ 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 $\text{JSP}^{\alpha,\beta}$ is SZK-complete.

2.3.1 Jensen-Shannon Divergence Problem Is SZK-Complete

We need to show that $\text{JSP}^{\alpha,\beta}$ 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 that:

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

We use this characterization to tie Jensen-Shannon Divergence Problem to another SZK-complete problem—the 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 $\text{H}(C_0) - \text{H}(C_1)$ 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 $\text{JSP}^{\alpha,\beta}$ 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].

$\text{JSP}^{\alpha,\beta}$ is in SZK: We reduce $\text{JSP}^{\alpha,\beta}$ to $\text{ED}^{(\alpha-\beta)/2}$. 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 $\text{H}(B') = 1 - \frac{\alpha+\beta}{2}$. The chain rule for entropy[19] implies that

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

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

$\text{JSP}^{\alpha,\beta}$ is SZK-hard: We reduce $\text{SDP}^{1-2^{-k}, 2^{-k}}$ to the problem $\text{JSP}^{\alpha,\beta}$, for some large enough k. This is sufficient since $\text{SDP}^{1-2^{-k}, 2^{-k}}$ 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

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

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

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]

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

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 $\mathrm{SD}(C_0, C_1) \ge 1 - 2^{-k}$, then $\mathrm{JS}(C_0, C_1) \ge \alpha$; and if $\mathrm{SD}(C_0, C_1) \le 2^{-k}$, then $\mathrm{JS}(C_0, C_1) \le \beta$. 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.

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

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

[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.

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. Equations (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 $\mathbb{E}_{y\sim Y}\left[g(\theta_y)\right]$, for some function $g : [-1, 1] \to [0, 1]$. For example:

- $\mathrm{SD}(C_0, C_1) = \mathbb{E}_{y\sim Y}\left[|\theta_y|\right]$ (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—is 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. 1 a plot comparing the different choices for the function g.

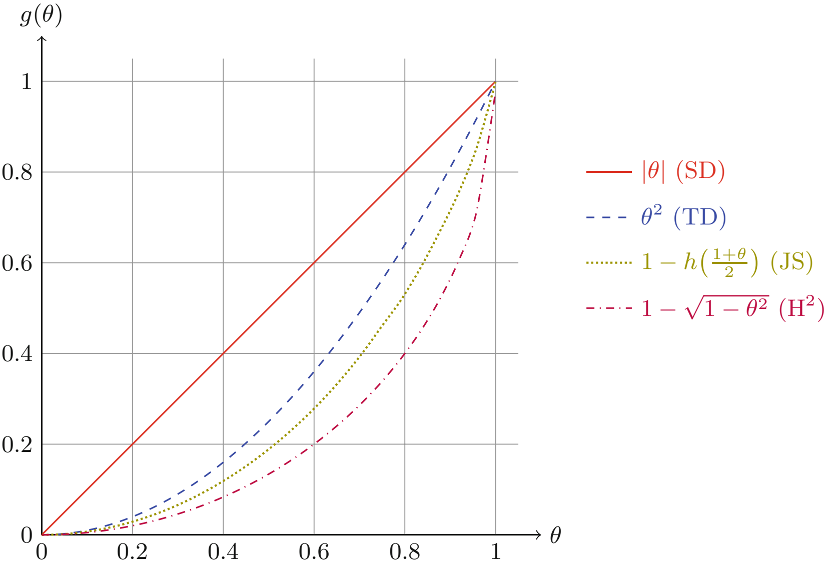

Fig. 1. Comparison between the difference choices of the function g that were discussed. Since all functions are symmetric around 0, we restrict to the domain $[0, 1]$. Recall that $g_1(\theta) = |\theta|$ corresponds to SD, $g_2(\theta) = \theta^2$ to TD, $g_3(\theta) = 1 - h\left(\frac{1+\theta}{2}\right)$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

Another popular f-divergence that we have not discussed thus far[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

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

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 $g(z) = 1 - \sqrt{1-z^2}$).

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

References

[AARV17] Applebaum, B., Arkis, B., Raykov, P., Vasudevan, P.N.: Conditional disclosure of secrets: amplification, closure, amortization, lower-bounds, and separations. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Part I. LNCS, vol. 10401, pp. 727–757. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63688-7_24

[AGGM06] Akavia, A., Goldreich, O., Goldwasser, S., Moshkovitz, D.: On basing one-way functions on NP-hardness. In: Kleinberg, J.M. (ed.) Symposium on Theory of Computing, pp. 701–710. ACM (2006)

[AH91] Aiello, W., Hastad, J.: Statistical zero-knowledge languages can be recognized in two rounds. J. Comput. Syst. Sci. **42**(3), 327–345 (1991)

[BB15] Bogdanov, A., Brzuska, C.: On basing size-verifiable one-way functions on NP-hardness. In: Dodis, Y., Nielsen, J.B. (eds.) TCC 2015. LNCS, vol. 9014, pp. 1–6. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46494-6_1

[BBF16] Brakerski, Z., Brzuska, C., Fleischhacker, N.: On statistically secure obfuscation with approximate correctness. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016, Part II. LNCS, vol. 9815, pp. 551–578. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-53008-5_19

[BBM11] Bhatnagar, N., Bogdanov, A., Mossel, E.: The computational complexity of estimating MCMC convergence time. In: Goldberg, L.A., Jansen, K., Ravi, R., Rolim, J.D.P. (eds.) APPROX/RANDOM -2011. LNCS, vol. 6845, pp. 424–435. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22935-0_36

[BCH+17] Bouland, A., Chen, L., Holden, D., Thaler, J., Vasudevan, P.N.: On the power of statistical zero knowledge. In: FOCS (2017)

[BDRV18] Berman, I., Degwekar, A., Rothblum, R.D., Vasudevan, P.N.: Multi-collision resistant hash functions and their applications. In: Nielsen, J.B., Rijmen, V. (eds.) EUROCRYPT 2018. LNCS, vol. 10821, pp. 133–161. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78375-8_5

[BDV17] Bitansky, N., Degwekar, A., Vaikuntanathan, V.: Structure vs. hardness through the obfuscation lens. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017. LNCS, vol. 10401, pp. 696–723. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63688-7_23

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

[BHZ87] Boppana, R.B., Håstad, J., Zachos, S.: Does co-NP have short interactive proofs? Inf. Process. Lett. **25**(2), 127–132 (1987)

[BKP18] Bitansky, N., Kalai, Y.T., Paneth, O.: Multi-collision resistance: a paradigm for keyless hash functions. In: STOC (2018)

[BL13] Bogdanov, A., Lee, C.H.: Limits of provable security for homomorphic encryption. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013. LNCS, vol. 8042, pp. 111–128. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40041-4_7

[Cam86] Le Cam, L.: Part I. Springer, New York (1986). https://doi.org/10.1007/978-1-4612-4946-7

[CCKV08] Chailloux, A., Ciocan, D.F., Kerenidis, I., Vadhan, S.: Interactive and non-interactive zero knowledge are equivalent in the help model. In: Canetti, R. (ed.) TCC 2008. LNCS, vol. 4948, pp. 501–534. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78524-8_28

[CGVZ18] Chen, Y.-H., Göös, M., Vadhan, S.P., Zhang, J.: A tight lower bound for entropy flattening. In: CCC (2018)

[DNR04] Dwork, C., Naor, M., Reingold, O.: Immunizing encryption schemes from decryption errors. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 342–360. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24676-3_21

[FGM+89] Fürer, M., Goldreich, O., Mansour, Y., Sipser, M., Zachos, S.: On completeness and soundness in interactive proof systems. Adv. Comput. Res. **5**, 429–442 (1989)

[For89] Fortnow, L.: The complexity of perfect zero-knowledge. Adv. Comput. Res. **5**, 327–343 (1989)

[FV17] Fehr, S., Vaudenay, S.: Personal Communication (2017)

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

[Gol17] Goldreich, O.: Introduction to Property Testing. Cambridge University Press, Cambridge (2017)

[GS89] Goldwasser, S., Sipser, M.: Private coins versus public coins in interactive proof systems. Adv. Comput. Res. **5**, 73–90 (1989)

[GSV98] Goldreich, O., Sahai, A., Vadhan, S.: Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. In: STOC (1998)

[GSV99] Goldreich, O., Sahai, A., Vadhan, S.: Can statistical zero knowledge be made non-interactive? Or on the relationship of SZK and NISZK. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 467–484. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48405-1_30

[GV99] Goldreich, O., Vadhan, S.P.: Comparing entropies in statistical zero knowledge with applications to the structure of SZK. In: CCC (1999)

[GV11] Goldreich, O., Vadhan, S.: On the complexity of computational problems regarding distributions. In: Goldreich, O. (ed.) Studies in Complexity and Cryptography. Miscellanea on the Interplay Between Randomness and Computation. LNCS, vol. 6650, pp. 390–405. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22670-0_27

[GVW02] Goldreich, O., Vadhan, S., Wigderson, A.: On interactive proofs with a laconic prover. Comput. Complex. **11**(1–2), 1–53 (2002)

[HR05] Holenstein, T., Renner, R.: One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 478–493. Springer, Heidelberg (2005). https://doi.org/10.1007/11535218_29

[IL89] Impagliazzo, R., Luby, M.: One-way functions are essential for complexity based cryptography. In: STOC, pp. 230–235 (1989)

[KNY17] Komargodski, I., Naor, M., Yogev, E.: White-box vs. black-box complexity of search problems: Ramsey and graph property testing. In: FOCS (2017)

[KNY18] Komargodski, I., Naor, M., Yogev, E.: Collision resistant hashing for paranoids: dealing with multiple collisions. In: Nielsen, J.B., Rijmen, V. (eds.) EUROCRYPT 2018. LNCS, vol. 10821, pp. 162–194. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78375-8_6

[KY18] Komargodski, I., Yogev, E.: On distributional collision resistant hashing. In: Shacham, H., Boldyreva, A. (eds.) CRYPTO 2018. LNCS, vol. 10992, pp. 303–327. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96881-0_11

[LZ17] Lovett, S., Zhang, J.: On the impossibility of entropy reversal, and its application to zero-knowledge proofs. In: Kalai, Y., Reyzin, L. (eds.) TCC 2017. LNCS, vol. 10677, pp. 31–55. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70500-2_2

[NR06] Naor, M., Rothblum, G.N.: Learning to impersonate. In: ICML, pp. 649–656 (2006)

[Ost91] Ostrovsky, R.: One-way functions, hard on average problems, and statistical zero-knowledge proofs. In: Structure in Complexity Theory Conference, pp. 133–138 (1991)

[OV08] Ong, S.J., Vadhan, S.: An equivalence between zero knowledge and commitments. In: Canetti, R. (ed.) TCC 2008. LNCS, vol. 4948, pp. 482–500. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78524-8_27

[OW93] Ostrovsky, R., Wigderson, A.: One-way functions are essential for non-trivial zero-knowledge. In: ISTCS, pp. 3–17 (1993)

[PW17] Polyanskiy, Y., Wu, Y.: Lecture notes on information theory (2017). http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf

[SV03] Sahai, A., Vadhan, S.: A complete problem for statistical zero knowledge. J. ACM (JACM) **50**(2), 196–249 (2003)

[Top00] Topsøe, F.: Some inequalities for information divergence and related measures of discrimination. IEEE Trans. Inf. Theory **46**(4), 1602–1609 (2000)

[Vad99] Vadhan, S.P.: A study of statistical zero-knowledge proofs. Ph.D. thesis, Massachusetts Institute of Technology (1999)

[Yeh16] Yehudayoff, A.: Pointer chasing via triangular discrimination. Electron. Colloq. Comput. Complex. (ECCC) **23**, 151 (2016)