A Tight Parallel Repetition Theorem for Partially Simulatable Interactive Arguments via Smooth KL-Divergence*

Itay Berman**, Iftach Haitner*** †, and Eliad Tsfadia†

Abstract. Hardness amplification is a central problem in the study of interactive protocols. While "natural" parallel repetition transformation is known to reduce the soundness error of some special cases of interactive arguments: three-message protocols (Bellare, Impagliazzo, and Naor [FOCS '97]) and public-coin protocols (Håstad, Pass, Wikström, and Pietrzak [TCC '10], Chung and Liu [TCC '10] and Chung and Pass [TCC '15]), it fails to do so in the general case (the above Bellare et al.; also Pietrzak and Wikström [TCC '07]).

The only known round-preserving approach that applies to all interactive arguments is Haitner's *random-terminating* transformation [SICOMP '13], who showed that the parallel repetition of the transformed protocol reduces the soundness error at a *weak* exponential rate: if the original m-round protocol has soundness error $1-\varepsilon$, then the n-parallel repetition of its random-terminating variant has soundness error $(1-\varepsilon)^{\varepsilon n/m^4}$ (omitting constant factors). Håstad et al. have generalized this result to *partially simulatable interactive arguments*, showing that the n-fold repetition of an m-round δ-simulatable argument of soundness error $1-\varepsilon$ has soundness error $(1-\varepsilon)^{\varepsilon\delta^2 n/m^2}$. When applied to random-terminating arguments, the Håstad et al. bound matches that of Haitner.

In this work we prove that parallel repetition of random-terminating arguments reduces the soundness error at a much stronger exponential rate: the soundness error of the n parallel repetition is $(1-\varepsilon)^{n/m}$, only an m factor from the optimal rate of $(1-\varepsilon)^n$ achievable in public-coin and three-message arguments. The result generalizes to δ-simulatable arguments, for which we prove a bound of $(1-\varepsilon)^{\delta n/m}$. This is achieved by presenting a tight bound on a relaxed variant of the KL-divergence between the distribution induced by our reduction and its ideal variant, a result whose scope extends beyond parallel repetition proofs. We prove the tightness of the above bound for random-terminating arguments, by presenting a matching protocol.

* Due to space limitations, the reader is referred to the full version [2].

** MIT. E-mail:itayberm@mit.edu. Research supported in part by NSF Grants CNS-1413920 and CNS-1350619, and by the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Research Office under contracts W911NF-15-C-0226 and W911NF-15-C-0236.

*** Director of the Check Point Institute for Information Security.

† School of Computer Science, Tel Aviv University. E-mail:{iftachh@cs.tau.ac.il, eliadtsf@tau.ac.il}. Research supported by ERC starting grant 638121 and Israel Science Foundation grant 666/19.

Keywords: parallel repetition; interactive argument; partially simulatable; smooth KL-divergence

1 Introduction

Hardness amplification is a central question in the study of computation: can a somewhat secure primitive be made fully secure, and, if so, can this be accomplished without loss (i.e., while preserving certain desirable properties the original primitive may have). In this paper we focus on better understanding the above question with respect to interactive arguments (also known as, computationally sound proofs). In an interactive argument, a prover tries to convince a verifier in the validity of a statement. The basic properties of such proofs are *completeness* and *soundness*. Completeness means that the prover, typically using some extra information, convinces the verifier to accept valid statements with high probability. Soundness means that a cheating *polynomial-time* prover cannot convince the verifier to accept invalid statements, except with small probability. Interactive arguments should be compared with the related notion of *interactive proofs*, whose soundness should hold against *unbounded* provers. Interactive argument are important for being "sufficiently secure" proof system that sometimes achieve properties (e.g., compactness) that are beyond the reach of interactive proofs. Furthermore, the security of many cryptographic protocols (e.g., binding of a computationally binding commitment) can be cast as the soundness of a related interactive argument, but (being computational) cannot be cast as the soundness of a related interactive proof.

The question of hardness amplification with respect to interactive arguments is whether an argument with *non-negligible* soundness error, i.e., a cheating prover can convince the verifier to accept false statements with some non-negligible probability, can be transformed into a new argument, with similar properties, of negligible soundness error (i.e., the verifier almost never accepts false statements). The most common paradigm to obtain such an amplification is via *repetition*: repeat the protocol multiple times with independent randomness, and the verifier accepts only if the verifiers of the original protocol accept in *all* executions. Such repetitions can be done in two different ways, sequentially (known as *sequential repetition*), where the $(i+1)$ execution of the protocol starts only after the i^{th} execution has finished, or in parallel (known as *parallel repetition*), where the executions are all simultaneous. Sequential repetition is known to reduce the soundness error in most computational models (cf., Damgärd and Pfitzmann [9]), but has the undesired effect of increasing the round complexity of the protocol. Parallel repetition, on the other hand, does preserve the round complexity, and reduces the soundness error for (single-prover) interactive proofs (Goldreich [16]) and two-prover interactive proofs (Raz [25], Holenstein [19], Rao [24]). Parallel repetition was also shown to reduce the soundness error in three-message arguments ([1]) and public-coin arguments (Håstad, Pass, Wikström, and Pietrzak [18], Chung and Lu [5], Chung and Pass [8]). Unfortunately, as shown by Bellare et al. [1], and by Pietrzak and Wikström [23], parallel repeti-

tion *might not* reduce the soundness error of any interactive argument: assuming common cryptographic assumptions, [23] presented an 8-message interactive proof with constant soundness error, whose parallel repetition, for *any* polynomial number of repetitions, still has a constant soundness error.

Faced with the above barrier, Haitner [17] presented a simple method for transforming any interactive argument π into a slightly modified protocol $\widetilde{\pi}$, such that the parallel repetition of $\widetilde{\pi}$ does reduce the soundness error. Given any m-round interactive protocol $\pi = (\mathrm{P}, \mathrm{V})$, let $\widetilde{\mathrm{V}}$ be the following *random-terminating variant* of V: in each round, $\widetilde{\mathrm{V}}$ flips a coin that takes one with probability $1/m$ and zero otherwise. If the coin outcome is one, $\widetilde{\mathrm{V}}$ accepts and aborts the execution. Otherwise, $\widetilde{\mathrm{V}}$ acts as V would, and continues to the next round. At the end of the prescribed execution, if reached, $\widetilde{\mathrm{V}}$ accepts if and only if V would. Observe that if the original protocol π has soundness error $1-\varepsilon$, then the new protocol $\widetilde{\pi} = (\mathrm{P}, \widetilde{\mathrm{V}})$ has soundness error $1-\varepsilon/4$ (i.e., only slightly closer to one). Haitner [17] proved that the parallel repetition of $\widetilde{\pi}$ does reduce the soundness error (for any protocol π). Håstad, Pass, Wikström, and Pietrzak [18] have generalized the above to *partially-simulatable interactive arguments*, a family of interactive arguments that contains the random-terminating variant protocols as a special case. An interactive argument $\pi = (\mathrm{P}, \mathrm{V})$ is δ-simulatable if given any partial view v of an efficient prover P^* interacting with V, the verifier's future messages in $(\mathrm{P}^*, \mathrm{V})$ can be simulated with probability δ. This means that one can efficiently sample a random continuation of the execution conditioned on an event of density δ over V's coins consistent with v. It is easy to see that the random-terminating variant of any protocol is $1/m$ simulatable. Unfortunately, the soundness bound proved by Haitner [17], Håstad et al. [18] lags way behind what one might have hoped for, making parallel repetition impractical in many typical settings. Assuming a δ-simulatable argument π has soundness error is $1-\varepsilon$, then π^n, the n-parallel repetition of π, was shown to have soundness error $(1-\varepsilon)^{\varepsilon\delta^2 n/m^2}$ (equals $(1-\varepsilon)^{\varepsilon n/m^4}$ if π is a random-terminating variant), to be compared with the $(1-\varepsilon)^n$ bound achieved by parallel repetition of interactive proofs, and by three-message and public-coin interactive arguments.[1] Apart from the intellectual challenge, improving the above bound is important since repeating the random-termination variant in parallel is the *only* known unconditional round-preserving amplification method for arbitrary interactive arguments.

1.1 Proving Parallel Repetition

Let $\pi = (\mathrm{P}, \mathrm{V})$ be an interactive argument with assumed soundness error $1-\varepsilon$, i.e., a polynomial time prover cannot make the verifier accept a false statement with probability larger than $1-\varepsilon$. Proving amplification theorems for such proof

[1] As in all known amplifications of computational hardness, and proven to be an inherent limitation (at least to some extent) in Dodis et al. [11], the improvement in the soundness error does not go below negligible. We ignore this subtly in the introduction. We also ignore constant factors in the exponent.

systems is done via reduction: assuming the existence of a cheating prover P^{n*} making all the n verifiers in n-fold protocol $\pi^n = (\mathrm{P}^n, \mathrm{V}^n)$ accept a false statement "too well" (e.g., more than $(1-\varepsilon)^n$), this prover is used to construct a cheating prover P^* making V accept this false statement with probability larger than $1-\varepsilon$, yielding a contradiction. Typically, the cheating prover P^* emulates an execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$ while *embedding* the (real) verifier V as one of the n verifiers (i.e., by embedding its messages). Analyzing the success probability of this P^* is directly reduced to bounding the "distance" (typically statistical distance or KL-divergence) between the following Winning and Attacking distributions: the Winning distribution is the n verifiers' messages distribution in a winning (all verifiers accept) execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$. The Attacking distribution is the n verifiers' messages distribution in the emulated execution done by P^* (when interacting with V).

If the verifier is public-coin, or if the prover is unrestricted (as in single-prover interactive proofs), an optimal strategy for P^* is sampling the emulated verifiers messages uniformly at random conditioned on all verifiers accept, and the messages so far. Håstad et al. [18] have bounded the statistical distance between the induced Winning and Attacking distributions in such a case, while Chung and Pass [8] gave a tight bound for the KL-divergence between these distributions, yielding an optimal result for public-coin arguments.

For non public-coin protocols, however, a computationally bounded prover cannot always perform the above sampling task (indeed, this inability underneath the counter examples for parallel repetition of such arguments). However, if the argument is random terminating, the cheating prover can sample the following "skewed" variant of the desired distribution: it samples as described above, but conditioned that the real verifier *aborts at the end of the current round*, making the simulation of its future messages trivial. More generally, for partially-simulatable arguments, the cheating prover samples the future messages of the real verifier using the built-in mechanism for sampling a skewed sample of its coins. Analyzing the prover success probability for such an attack, and thus upper-bounding the soundness error of the parallel repetition of such arguments, reduces to understanding the (many-round) skewed distributions induced by the above attack. This will be discussed in the next section.

1.2 Skewed Distributions

The Attacking distribution induced by the security proof of parallel repetition of partially-simulatable arguments discussed in Section 1.1, gives rise to the following notion of (many-round) skewed distributions. Let $P = P_X$ be a distribution over an $m \times n$ size matrices, letting P_{X_i} and P_{X^j} denoting the induced distribution over the i^{th} row and j^{th} column of X, respectively. For an event W, let $\widetilde{P} = P|W$. The following distribution $Q_{X,J}$ is a skewed variant of $\widetilde{P}$ induced by an event family $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ over P: let $Q_J = U_{[n]}$, and let

$$Q_{X|J} = \prod_{i=1}^{m} P_{X_{i,J}|X_{<i,J}} \widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J},E_{i,J}} \tag{1}$$

for $X_{<i} = (X_1, \ldots, X_{i-1})$, $X_{<i,j} = (X_{<i})^j = (X_{1,j}, \ldots, X_{i-1,j})$ and $X_{i,-j} = X_{i,[n]\setminus\{j\}}$. That is, Q induced by first sampling $J \in [n]$ uniformly at random, and then sampling the following skewed variant of $\widetilde{P}$: At round i

1. Sample $X_{i,J}$ according to $P_{X_{i,J}|X_{<i,J}}$ (rather than $P_{X_{i,J}|X_{<i},W}$ as in $\widetilde{P}$),
2. Sample $X_{i,-J}$ according $\widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J},E_{i,J}}$ (rather than $\widetilde{P}_{X_{i,J}|X_{<i},X_{i,J}}$).

At a first glance, the distribution Q looks somewhat arbitrary. Nevertheless, as we explain below, it naturally arises in the analysis of parallel repetition theorem of partially-simulatable interactive arguments, and thus of random-terminating variants. Somewhat similar skewed distributions also come up when proving parallel repetition of two-prover proofs, though there we only care for single round distributions, i.e., $m = 1$.

The distributions $\widetilde{P}$ and Q relate to the Winning and Attacking distributions described in Section 1.1 in the following way: let $\pi = (\mathrm{P}, \mathrm{V})$ be an m-round δ-simulatable argument, and let P^{n*} be an efficient (for simplicity) deterministic cheating prover for π^n. Let P to be the distribution of the n verifiers messages in a random execution of π^n, and let W be the event that P^{n*} wins in $(\mathrm{P}^{n*}, \mathrm{V}^n)$. By definition, $\widetilde{P} = P|W$ is just the Winning distribution. Assume for sake of simplicity that V is a random-termination variant (halts at the end of each round with probability $1/m$), let $E_{i,j}$ be the set of coins in which the j^{th} verifier halts at the end of the i^{th} round of $(\mathrm{P}^n, \mathrm{V}^n)$, and let $Q = Q(P, W, \{E_{i,j}\})$ be according to Equation (1). Then, ignoring some efficiency concerns, Q is just the Attacking distribution. Consequently, a bound on the soundness error of π^n can be proved via the following result:

Lemma 1 (informal). *Let π be a partially simulatable argument of soundness error $(1-\varepsilon)$. Assume that for every efficient cheating prover for π^n and every event T, it holds that*

$$\Pr_{Q_X}[T] \leq \Pr_{\widetilde{P}_X}[T] + \gamma$$

where W, $\widetilde{P}$ and Q are as defined above with respect to this adversary, and that Q is efficiently samplable. Then π^n has soundness error $(1-\varepsilon)^{\log(1/P[W])/\gamma}$.

It follows that proving a parallel repetition theorem for partially simulatable arguments, reduces to proving that low probability events in $\widetilde{P}_X$ have low probability in Q_X (for the sake of the introduction, we ignore the less fundamental samplability condition assumed for Q). One can try to prove the latter, as implicitly done in [17, 18], by bounding the *statistical distance* between $\widetilde{P}$ and Q (recall that $\mathrm{SD}(P, Q) = \max_E(\Pr_P[E] - \Pr_Q[E])$). This approach, however, seems doomed to give non-tight bounds for several reasons: first, statistical distance is not geared to bound non-product distributions (i.e., iterative processes) as the one defined by Q, and one is forced to use a wasteful hybrid argument in order to bound the statistical distance of such distributions. A second reason is that statistical distance bounds the difference in probability between the two distributions for *any* event, where we only care that this difference is small

for low (alternatively, high) probability events. In many settings, achieving this (unneeded) stronger guarantee inherently yields a weaker bound.

What seems to be a more promising approach is bounding the *KL-divergence* between $\widetilde{P}$ and Q (recall that $D(P||Q) = \mathrm{E}_{x\sim P} \log \frac{P(x)}{Q(x)}$). Having a chain rule, KL-divergence is typically an excellent choice for non-product distributions. In particular, bounding it only requires understanding the non-product nature (i.e., the dependency between the different entries) of the left-hand-side distribution. This makes KL-divergence a very useful measure in settings where the iterative nature of the right-hand-side distribution is much more complicated. Furthermore, a small KL-divergence guarantees that low probability events in $\widetilde{P}$ happen with almost the same probability in Q, but it only guarantees a weaker guarantee for other events (so it has the potential to yield a tighter result). Chung and Pass [8] took advantage of this observation for proving their tight bound on parallel repetition of public-coin argument by bounding the KL-divergence between their variants of $\widetilde{P}$ and Q. Unfortunately, for partially simulatable (and for random terminating) arguments, the KL-divergence between these distributions might be infinite.

Faced with the above difficulty, we propose a relaxed variant of KL-divergence that we name *smooth KL-divergence*. On the one hand, this measure has the properties of KL-divergence that make it suitable for our settings. However, on the other hand, it is less fragile (i.e., oblivious to events of small probability), allowing us to tightly bound its value for the distributions under consideration.

1.3 Smooth KL-divergence

The KL-divergence between distributions P and Q is a very sensitive distance measure: an event x with $P(x) \gg Q(x)$ might make $D(P||Q)$ huge even if $P(x)$ is tiny (e.g., $P(x) > 0 = Q(x)$ implies $D(P||Q) = \infty$). While events of tiny probability are important in some settings, they have no impact in ours. So we seek a less sensitive measure that enjoys the major properties of KL-divergence, most notably having chain-rule and mapping low probability events to low probability events. A natural attempt would be to define it as $\inf_{P',Q'}\{D(P'||Q')\}$, where the infimum is over all pairs of distributions such that both $\mathrm{SD}(P,P')$ and $\mathrm{SD}(Q,Q')$ are small. This relaxation, however, requires an upper bound on the probability of events with respect to Q, which in our case is the complicated skewed distribution Q. Unfortunately, bounding the probability of events with respect to the distribution Q is exactly the issue in hand.

Instead, we take advantage of the asymmetric nature of the KL-divergence to propose a relaxation that only requires upper-bounding events with respect to P, which in our case is the much simpler $\widetilde{P}$ distribution. Assume P and Q are over a domain $\mathcal{U}$. The *α-smooth KL-divergence of P and Q* is defined by

$$D^{\alpha}(P||Q) = \inf_{(F_P,F_Q)\in\mathcal{F}}\{D(F_P(P)||F_Q(Q))\}$$

for $\mathcal{F}$ being the set of randomized function pairs, such that for any $(F_P,F_Q)\in\mathcal{F}$: (1) $\Pr_{x\sim P}[F_P(x)\neq x]\leq\alpha$, and (2) $\forall x\in\mathcal{U}$ and $C\in\{P,Q\}$: $F_C(x)\in\{x\}\cup\overline{\mathcal{U}}$.

Note that for any pair $(F_P, F_Q) \in \mathcal{F}$ and any event B over $\mathcal{U}$, it holds that $\Pr_Q[B] \geq \Pr_{F_Q(Q)}[B]$, and $\Pr_{F_P(P)}[B] \geq \Pr_P[B] - \alpha$. Thus, if $\Pr_P[B]$ is low, a bound on $D(F_P(P)||F_Q(Q))$ implies that $\Pr_Q[B]$ is also low. Namely, low probability events in P happen with low probability also in Q.

Bounding smooth KL-divergence. Like the (standard) notion of KL-divergence, the power of smooth KL-divergence is best manifested when applied to non-product distributions. Let P and Q be two distributions for which we would like to prove that small events in $P_{X=(X_1,\ldots,X_m)}$ are small in $Q_{X=(X_1,\ldots,X_m)}$ (as a running example, let P and Q be the distributions $\widetilde{P}_X$ and $Q_{X,J}$ from the previous section, respectively). By chain rule of KL-divergence, it suffices to show that for some events $B_1, \ldots, B_m$ over Q (e.g., B_i is the event that $J|X_{<i}$ has high min entropy) it holds that

$$\sum_{i=1}^{m} D(P_{X_i}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}}) \quad \left(\text{i.e.,} \sum_{i=1}^{m} \mathrm{E}_{x \leftarrow P_{X_{<i}}}\left[D\big(P_{X_i|X_{<i}=x}||Q_{X_i|X_{<i}=x,B_{\leq i}}\big)\right]\right) \tag{2}$$

is small, and $Q[B_{\leq m}]$ is large. Bounding Equation (2) only requires understanding P and simplified variants of Q (in which all but the i^{th} entry is sampled according to P). Unfortunately, bounding $Q[B_{\leq m}]$ might be hard since it requires a good understanding of the distribution Q itself. We would have liked to relate the desired bound to $P[B_{\leq m}]$, but the events $\{B_i\}$ might not even be defined over P (in the above example, P has no J part). However, smooth KL-divergence gives us the means to do almost that.

Lemma 2 (Bounding smooth KL-divergence, informal). *Let P, Q and $\{B_i\}$ be as above. Associate the events $\{\widetilde{B}_i\}$ with P, each $\widetilde{B}_i$ (independently) occur with probability $Q[B_i \mid B_{<i}, X_{<i}]$. Then*

$$D^{1-P[\widetilde{B}_{\leq m}]}(P_X||Q_X) \leq \sum_{i=1}^{m} D\Big(P_{X_i}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}|\widetilde{B}_{\leq i}}\Big).$$

Namely, $\{\widetilde{B}_i\}$ mimics the events $\{B_i\}$, defined over Q, in (an extension of) P. It follows that bounding the smooth KL-divergence of P_X and Q_X (and thus guarantee that small events in P_X are small in Q_X), is reduced to understanding P and *simplified* variants of Q.

1.4 Main Results

We prove the following results (in addition to Lemmas 1 and 2). The first result, which is the main technical contribution of this paper, is the following bound on the smooth KL-divergence between a distribution and its many-round skewed variant.

Theorem 1 (Smooth KL-divergence for skewed distributions, informal). *Let $P = P_X$ be a distribution over an $m \times n$ matrices with independent columns, and let W and $\mathcal{E} = \{E_{i,j}\}$ be events over P. Let $\widetilde{P} = P|W$ and let $Q = Q(P, W, \mathcal{E})$ be the skewed variant of $\widetilde{P}$ defined in Equation (1). Assume $\forall (i,j) \in [m] \times [n]$: (1) $E_{i,j}$ is determined by X^j and (2) There exists $\delta_{i,j} \in (0,1]$ such that $P[E_{i,j}|X_{\leq i,j}] = \delta_{i,j}$ for any fixing of $X_{\leq i,j}$. Then (ignoring constant factors, and under some restrictions on n and $P[W]$)*

$$D^{\varepsilon m + 1/\delta n}(\widetilde{P}_X || Q_X) \leq \varepsilon m + m/\delta n$$

for $\delta = \min_{i,j}\{\delta_{i,j}\}$ and $\varepsilon = \log(\frac{1}{P[W]})/\delta n$. In a special case where $E_{i,j}$ is determined by $X_{\leq i+1,j}$, it holds that

$$D^{\varepsilon + 1/\delta n}(\widetilde{P}_X || Q_X) \leq \varepsilon + m/\delta n.$$

Combining Lemma 1 and Theorem 1 yields the following bound on parallel repetition of partially simulatable arguments. We give separate bounds for partially simulatable argument and for *partially prefix-simulatable arguments*: a δ-simulatable argument is δ-prefix-simulatable if for any i-round view, the event E guaranteed by the simulatable property for this view is determined by the coins used in the first $i+1$ rounds. It is clear that the random-termination variant of an m-round argument is $1/m$-prefix-simulatable.

Theorem 2 (Parallel repetition for partially simulatable arguments, informal). *Let π be an m-round δ-simulatable interactive argument with soundness error $1-\varepsilon$, and let $n \in \mathbb{N}$. Then π^n has soundness error $(1-\varepsilon)^{\delta n/m}$. Furthermore, if π is δ-prefix-simulatable, then π^n has soundness error $(1-\varepsilon)^{\delta n}$.*[2]

A subtlety that arises when proving Theorem 2 is that a direct composition of Lemma 1 and Theorem 1 only yields the desired result when the number of repetitions n is "sufficiently" large compared to the number of rounds m (roughly, this is because we need the additive term $m/\delta n$ in Theorem 1 to be smaller than ε). We bridge this gap by presenting a sort of upward-self reduction from a few repetitions to many repetitions. The idea underlying this reduction is rather general and applies to other proofs of this type, and in particular to those of [18, 17, 6].[3]

We complete the picture by showing that an δ factor in the exponent in Theorem 2 is unavoidable.

[2] Throughout, we assume that the protocol transcript contains the verifier's Accept/Reject decision (which is without loss of generality for random-terminating variants). We deffer the more general case for the next version.

[3] Upward-self reductions trivially exist for interactive proof: assume the existence of a cheating prover P^{n*} breaking the α soundness error of π^n, then $(\mathrm{P}^{n*})^{\ell}$, i.e., the prover using P^{n*}in parallel for ℓ times, violates the assumed α^{ℓ} soundness error of $\pi^{n\ell}$. However, when considering interactive arguments, for which we cannot guarantee a soundness error below negligible (see Footnote 1), this approach breaks down when α^{ℓ} is negligible.

Theorem 3 (lower bound, informal). *Under suitable cryptographic assumptions, for any $n, m \in \mathbb{N}$ and $\varepsilon \in [0,1]$, there exists an m-round δ-prefix-simulatable interactive argument π with soundness error $1-\varepsilon$, such that π^n has soundness error at least $(1-\varepsilon)^{\delta n}$. Furthermore, protocol π is a random-terminating variant of an interactive argument.*

It follows that our bound for partially prefix-simulatable arguments and random-termination variants, given in Theorem 2, is tight.

1.4.1 Proving Theorem 1 We highlight some details about the proof of Theorem 1. Using Lemma 2, we prove the theorem by showing that the following holds for a carefully chosen events $\{B_i\}$ over $Q_{X,J}$:

- $\sum_{i=1}^{m} D\Big(\widetilde{P}_{X_i} || Q_{X_i|B_{\leq i}} \mid \widetilde{P}_{X_{<i}|\widetilde{B}_{\leq i}}\Big)$ is small, and
- $\widetilde{P}[\widetilde{B}_{\leq m}]$ is large,

where $\{\widetilde{B}_i\}$ are events over (extension of) $\widetilde{P}$, with $\widetilde{B}_i$ taking the value 1 with probability $Q[B_i \mid B_{<i}, X_{<i}]$. We chose the events $\{B_i\}$ so that we have the following guarantees on $Q_{X_i,J|B_{\leq i},X_{<i}}$:

1. $J|X_{<i}$ has high entropy (like it has without any conditioning), and
2. $P[W \mid X_{<i}, X_{i,J}, E_{i,J}] \geq P[W|X_{<i}]/2$.

Very roughly, these guarantees make the task of bounding the required KL-divergence much simpler since they guarantee that the skewing induced by Q does not divert it too much (compared to $\widetilde{P}$). The remaining challenge is therefore lower-bounding $\widetilde{P}[\widetilde{B}_{\leq m}]$. We bound the latter distribution by associating a martingale sequence with the distribution Winning. In order to bound this sequence, we prove a new concentration bound for "slowly evolving" martingale sequences, Lemma 3, that we believe to be of independent interest.

1.5 Related Work

1.5.1 Interactive Arguments

Positive results. Bellare et al. [1] proved that the parallel repetition of three-message interactive arguments reduces the soundness error at an exponential, but not optimal, rate. Canetti et al. [4] later showed that parallel repetition does achieve an optimal exponential decay in the soundness error for such arguments. Pass and Venkitasubramaniam [22] have proved the same for constant-round public-coin arguments. For public-coin arguments of any (polynomial) round complexity, Håstad et al. [18] were the first to show that parallel repetition reduces the soundness error exponentially, but not at an optimal rate. The first optimal analysis of parallel repetition in public-coin arguments was that of Chung and Liu [6], who showed that the soundness error of the k repetitions improves to $(1-\varepsilon)^k$. Chung and Pass [8] proved the same bound using KL-divergence. For

non-public coin argument (of any round complexity), Haitner [17] introduced the random-terminating variant of a protocol, and proved that the parallel repetition of these variants improves the soundness error at a weak exponential rate. Håstad et al. [18] proved the same, with essentially the same parameters, for partially-simulatable arguments, that contain random-terminating protocols as a special case. All the above results extend to "threshold verifiers" where the parallel repetition is considered accepting if the number of accepting verifiers is above a certain threshold. Our result rather easily extends to such verifiers, but we defer the tedious details to the next version. Chung and Pass [7] proved that full independence of the parallel executions is not necessary to improve the soundness of public-coin arguments, and that the verifier can save randomness by carefully correlating the different executions. It is unknown whether similar savings in randomness can be achieved for random-terminating arguments. Finally, the only known round-preserving alternative to the random-terminating transformation is the elegant approach of Chung and Liu [6], who showed that a fully-homomorphic encryption (FHE) can be used to compile any interactive argument to a one (with the same soundness error) for which parallel repetition improves the soundness error at ideal rate, i.e., $(1-\varepsilon)^n$. However, in addition to being conditional (and currently it is only known how to construct FHE assuming hardness of learning with errors [3]), the compiled protocol might lack some of the guarantees of the original protocol (e.g., fairness). Furthermore, the reduction is *non* black box (the parties homomorphically evaluate *each* of the protocol's gates), making the resulting protocol highly impractical, and preventing the use of this approach when only black-box access is available (e.g., the weak protocol is given as a DLL or implemented in hardware).

Negative results. Bellare et al. [1] presented for any $n \in \mathbb{N}$, a four-message interactive argument of soundness error $1/2$, whose n-parallel repetition soundness remains $1/2$. Pietrzak and Wikström [23] ruled out the possibility that enough repetitions will eventually improve the soundness of an interactive argument. They presented a *single* 8-message argument for which the above phenomenon holds for all polynomial n simultaneously. Both results hold under common cryptographic assumptions.

1.5.2 Two-Prover Interactive Proofs The techniques used in analyzing parallel-repetition of interactive arguments are closely related to those for analyzing parallel repetition of two-prover one-round games. Briefly, in such a game, two unbounded *isolated* provers try to convince a verifier in the validity of a statement. Given a game of soundness error $(1-\varepsilon)$, one might expect the soundness error of its n parallel repetition to be $(1-\varepsilon)^n$, but as in the case of interactive arguments, this turned out to be false [13, 14, 15]. Nonetheless, Raz [25] showed that parallel repetition does achieve an exponential decay for any two-prover one-round game, and in particular reduces the soundness error to $(1-\varepsilon)^{\varepsilon^{O(1)}n/s}$, where s is the provers' answer length. These parameters were later improved by Holenstein [19], and improved further for certain types of games by Rao

[24], Dinur and Steurer [10], Moshkovitz [20]. The core challenge in the analysis of parallel repetition of interactive arguments and of multi-prover one-round games is very similar: how to simulate a random accepting execution of the proof/game given the verifier messages. In interactive arguments, this is difficult since the prover lacks computational power. In multi-prover one-round games, the issue is that the different provers cannot communicate.

Open Questions

While our bound for the parallel repetition of partially prefix-simulatable arguments is tight, this question for (non prefix) partially simulatable arguments is still open (there is a $1/m$ gap in the exponent). A more important challenge is to develop a better (unconditional) round-preserving amplification technique for arbitrary interactive arguments (which cannot be via random termination), or alternatively to prove that such an amplification does not exist.

Paper Organization

Basic notations, definitions and tools used throughout the paper are stated in Section 2. The definition of smooth KL-divergence and some properties of this measure are given in Section 3. The definition of many-round skewed distributions and our main bound for such distributions are given in Section 4. A proof sketch of the aforementioned bound is given in Section 6, and is used in Section 5 for proving our bound on the parallel repetition of partially simulatable arguments. Due to space limitations, the full proof of our main bound, the matching lower bound (Theorem 3) and other missing proofs are only given in the full version of this paper [2].

2 Preliminaries

2.1 Notation

We use calligraphic letters to denote sets, uppercase for random variables, and lowercase for values and functions. All logarithms considered here are natural logarithms (i.e., in base e). For $n \in \mathbb{N}$, let $[n] := \{1, \ldots, n\}$. Given a vector $v \in \Sigma^m$, we let v_i be its i^{th} entry, and let $v_{<i} = v_{1,\ldots,i-1}$ and $v_{\leq i} = v_{1,\ldots,i}$. For $v \in \{0,1\}^n$, let $1_v = \{i \in [n] \colon v_i = 1\}$. For $m \times n$ matrix x, let x_i and x^j denote their i^{th} row and j^{th} column respectively, and defined $x_{<i}$, $x_{\leq i}$, $x^{<j}$ and $x^{\leq j}$ respectively. Given a Boolean statement S (e.g., $X \geq 5$), let $\mathbb{1}_S$ be the indicator function that outputs 1 if S is a true statement and 0 otherwise.

Let poly denote the set of all polynomials, PPT denote for probabilistic polynomial time, and PPTM denote a PPT algorithm (Turing machine). A function $\nu \colon \mathbb{N} \to [0,1]$ is *negligible*, denoted $\nu(n) = \text{neg}(n)$, if $\nu(n) < 1/p(n)$ for every $p \in \text{poly}$ and large enough n. Function ν is *noticeable*, denoted $\nu(n) \geq 1/\text{poly}(n)$, if exists $p \in \text{poly}$ such that $\nu(n) \geq 1/p(n)$ for all n.

2.2 Distributions and Random Variables

A discrete random variable X over $\mathcal{X}$ is sometimes defined by its probability mass function (pmf) P_X (P is an arbitrary symbol). A conditional probability distribution is a function $P_{Y|X}(\cdot|\cdot)$ such that for any $x \in \mathcal{X}$, $P_{Y|X}(\cdot|x)$ is a pmf over $\mathcal{Y}$. The joint pmf P_{XY} can be written the product $P_X P_{Y|X}$, where $(P_X P_{Y|X})(x,y) = P_X(x)P_{Y|X}(y|x) = P_{XY}(xy)$. The marginal pmf P_Y can be written as the composition $P_{Y|X} \circ P_X$, where $(P_{Y|X} \circ P_X)(y) = \sum_{x\in\mathcal{X}} P_{Y|X}(y|x)P_X(x) = P_Y(y)$. We sometimes write $P_{\cdot,Y}$ to denote a pmf $P_{X,Y}$ for which we do not care about the random variable X. We denote by $P_X[W]$ the probability that an event W over P_X occurs, and given a set $\mathcal{S} \subseteq \mathcal{X}$ we define $P_X(\mathcal{S}) = P_X[X \in \mathcal{S}]$. Distribution P'_{XY} is an extension of P_X if $P'_X \equiv P_X$. Random variables and events defined over P_X are defined over the extension P'_{XY} by ignoring the value of Y. We sometimes abuse notation and say that P_{XY} is an extension of P_X.

The support of a distribution P over a finite set $\mathcal{X}$, denoted $\mathrm{Supp}(P)$, is defined as $\{x \in \mathcal{X} : P(x) > 0\}$. The *statistical distance* of two distributions P and Q over a finite set $\mathcal{X}$, denoted as $\mathrm{SD}(P,Q)$, is defined as $\max_{\mathcal{S}\subseteq\mathcal{X}} |P(\mathcal{S}) - Q(\mathcal{S})| = \frac{1}{2}\sum_{x\in\mathcal{S}} |P(x) - Q(x)|$. Given a set $\mathcal{S}$, let $U_{\mathcal{S}}$ denote the uniform distribution over the elements of $\mathcal{S}$. We sometimes write $x \sim \mathcal{S}$ or $x \leftarrow \mathcal{S}$, meaning that x is uniformly drawn from $\mathcal{S}$. For $p \in [0,1]$, let $\mathrm{Bern}(p)$ be the Bernoulli distribution over $\{0,1\}$, taking the value 1 with probability p.

2.3 KL-Divergence

Definition 1. *The* KL-divergence *(also known as, Kullback-Leibler divergence and relative entropy) between two distributions P, Q on a discrete alphabet $\mathcal{X}$ is*

$$D(P||Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)} = \mathrm{E}_{x\sim P}\log\frac{P(x)}{Q(x)},$$

where $0\cdot\log\frac{0}{0} = 0$ and if $\exists x \in \mathcal{X}$ such that $P(x) > 0 = Q(x)$ then $D(P||Q) = \infty$.

Definition 2. *Let P_{XY} and Q_{XY} be two probability distributions over $\mathcal{X}\times\mathcal{Y}$. The* conditional divergence *between $P_{Y|X}$ and $Q_{Y|X}$ is*

$$D(P_{Y|X}||Q_{Y|X}|P_X) = \mathrm{E}_{x\sim P_X}[D(P_{Y|X=x}||Q_{Y|X=x})] = \sum_{x\in\mathcal{X}} P_X(x)D(P_{Y|X=x}||Q_{Y|X=x}).$$

Fact 4 (Properties of divergence) *P_{XY} and Q_{XY} be two probability distributions over $\mathcal{X}\times\mathcal{Y}$. It holds that:*

1. *(Information inequality) $D(P_X||Q_X) \geq 0$, with equality holds iff $P_X = Q_X$.*
2. *(Monotonicity) $D(P_{XY}||Q_{XY}) \geq D(P_Y||Q_Y)$.*
3. *(Chain rule) $D(P_{X_1\cdots X_n}||Q_{X_1\cdots X_n}) = \sum_{i=1}^n D(P_{X_i|X_{<i}}||Q_{X_i|X_{<i}}|P_{X_{<i}})$ If $Q_{X_1\cdots X_n} = \prod_{i=1}^n Q_{X_i}$ then*

$$D(P_{X_1\cdots X_n}||Q_{X_1\cdots X_n}) = D(P_{X_1\cdots X_n}||P_{X_1}P_{X_2}\cdots P_{X_n}) + \sum_{i=1}^n D(P_{X_i}||Q_{X_i}).$$

4. *(Conditioning increases divergence) If* $Q_Y = Q_{Y|X} \circ P_X$ *(and* $P_Y = P_{Y|X} \circ P_X$*), then* $D(P_Y||Q_Y) \leq D(P_{Y|X}||Q_{Y|X}|P_X)$.
5. *(Data-processing) If* $Q_Y = P_{Y|X} \circ Q_X$ *(and* $P_Y = P_{Y|X} \circ P_X$*), it holds that* $D(P_Y||Q_Y) \leq D(P_X||Q_X)$.

Fact 5 *Let* X *be random variable drawn from* P *and let* W *be an event defined over* P*. Then* $D\big(P_{X|W}||P_X\big) \leq \log \frac{1}{P[W]}$.

Definition 3. *For* $p, q \in [0,1]$ *let* $D(p||q) := D(\mathrm{Bern}(p)||\,\mathrm{Bern}(q))$.

Fact 6 ([21, Implicit in Corollary 3.2 to 3.4]) *For any* $p \in [0,1]$*:*

1. $D((1-\delta)p||p) \geq \delta^2 p/2$ *for any* $\delta \in [0,1]$.
2. $D((1+\delta)p||p) \geq \min\{\delta, \delta^2\}p/4$ *for any* $\delta \in [0, \frac{1}{p} - 1]$.

The proof of the following proposition, which relies on Donsker and Varadhan [12]'s inequality, is given in the full version.

Proposition 1. *Let* X *be a random variable drawn form either* P *or* Q*. Assume that* $\Pr_P[|X| \leq 1] = 1$ *(i.e., if* X *is drawn from* P *then* $|X| \leq 1$ *almost surely) and that there exist* $\varepsilon, \sigma^2, K_1, K_2 > 0$ *such that* $\Pr_Q[|X| \leq 1] \geq 1 - \varepsilon$ *and*

$$\Pr_Q[|X| \geq t] \leq K_2 \cdot \exp\left(-\frac{t^2}{K_1 \sigma^2}\right) \quad \text{for all } 0 \leq t \leq 1.$$

Then, $\exists K_3 = K_3(K_1, K_2, \varepsilon) > 0$ *such that* $\mathrm{E}_P[X^2] \leq K_3 \cdot \sigma^2 \cdot (D(P||Q) + 1)$.

2.4 Concentration Bounds

The following concentration bound is proven in the full version.

Fact 7 *Let* $L_1, \ldots, L_n$ *be independent random variables over* $\mathbb{R}$ *with* $|L_i| \leq \ell$ *for all* $i \in [n]$ *and let* $Z_i = (L_i/p_i) \cdot \mathrm{Bern}(p_i)$ *with* $p_i > 0$ *for all* $i \in [n]$*. Let* $L = \sum_{i=1}^n L_i$*, let* $Z = \sum_{i=1}^n Z_i$*, let* $\mu = \mathrm{E}[L]$ *and let* $p = \min_{i \in [n]}\{p_i\}$*. Finally, let* $\Gamma = Z/\mu - 1$*. Then for any* $\gamma \in [0,1]$ *it holds that*

$$\Pr[|\Gamma| \geq \gamma] \leq 4 \exp\left(-\frac{p\mu^2\gamma^2}{5\ell^2 n}\right)$$

2.4.1 Martingales

Definition 4. *A sequence of random variables* $Y_0, Y_1, \ldots, Y_n$ *is called a* ***martingale sequence with respect to*** *a sequence* $X_0, X_1, \ldots, X_n$*, if* $\forall i \in [n]$*: (1)* Y_i *is a deterministic function of* $X_0, \ldots, X_i$*, and (2)* $\mathrm{E}[Y_i \mid X_0, \ldots, X_{i-1}] = Y_{i-1}$.

The following lemma (proven in the full version) is a new concentration bound on "slowly evolving" martingales.

Lemma 3 (A bound on slowly evolving martingales). *Let $Y_0 = 1, Y_1, \ldots, Y_n$ be a martingale w.r.t $X_0, X_1, \ldots, X_n$ and assume that $Y_i \geq 0$ for all $i \in [n]$. Then for every $\lambda \in (0, \frac{1}{4}]$ it holds that*

$$\Pr[\exists i \in [n] \text{ s.t. } |Y_i - 1| \geq \lambda] \leq \frac{23 \cdot \mathrm{E}\left[\sum_{i=1}^{n} \min\{|R_i|, R_i^2\}\right]}{\lambda^2}$$

for $R_i = \frac{Y_i}{Y_{i-1}} - 1$, letting $R_i = 0$ in case $Y_{i-1} = Y_i = 0$.

That is, if Y_i is unlikely to be far from Y_{i-1} in a multiplicative manner, then the sequence is unlikely to get far from 1.

2.5 Interactive Arguments

Definition 5 (Interactive arguments). *A* PPT *protocol* (P, V) *is an* interactive argument *for a language* $\mathrm{L} \in \mathrm{NP}$ *with* completeness α *and* soundness error β*, if the following holds:*

- $\Pr[(\mathrm{P}(w), \mathrm{V})(x) = 1] \geq \alpha(|x|)$ *for any* $(x, w) \in R_{\mathrm{L}}$.
- $\Pr[(\mathrm{P}^*, \mathrm{V})(x) = 1] \leq \max\{\beta(|x|), \mathrm{neg}(|x|)\}$ *for any* PPT P^* *and large enough* $x \notin \mathrm{L}$.

We refer to party P *as the* prover*, and to* V *as the* verifier.

Soundness against *non-uniform* provers is analogously defined, and all the results in this paper readily extend to this model.

Since in our analysis we only care about soundness amplification, in the following we fix L to be the empty language, and assume the input to the protocol is just a string of ones, which we refer to as the *security parameter*, a parameter we omit when cleared from the context.

2.5.1 Random-Terminating Variant

Definition 6 (Random-terminating variant, [17]). *Let* V *be a m-round randomized interactive algorithm. The* random-terminating variant of V*, denoted* $\widetilde{\mathrm{V}}$*, is defined as follows: algorithm* V *acts exactly as* V *does, but adds the following step at the end of each communication round: it tosses an $(1 - 1/m, 1/m)$ biased coin (i.e., 1 is tossed with probability $1/m$), if the outcome is one then it outputs 1 (i.e., accept) and halts. Otherwise, it continues as* V *would.*

For a protocol $\pi = (\mathrm{P}, \mathrm{V})$, the protocol $\widetilde{\pi} = (\mathrm{P}, \widetilde{\mathrm{V}})$ is referred to as the random-terminating variant of π.

2.5.2 Partially Simulatable Interactive Arguments

Definition 7 (Partially simulatable protocols, [18]). *A randomized interactive algorithm* V *is* δ-simulatable*, if there exists an oracle-aided* S *(simulator) such that the following holds: for every strategy* P^* *and a partial view* v *of* P^* *in an interaction of* $(\mathrm{P}^*, \mathrm{V})(1^\kappa)$*, the output of* $\mathsf{S}^{\mathrm{P}^*}(1^\kappa, v)$ *is* P^**'s view in a random continuation of* $(\mathrm{P}^*, \mathrm{V})(1^\kappa)$ *conditioned on* v *and* Δ*, for* Δ *being a* δ*-dense subset of the coins of* V *that are consistent with* v*. The running time of* $\mathsf{S}^{\mathrm{P}^*}(1^\kappa, v)$ *is polynomial in* κ *and the running time of* $\mathrm{P}^*(1^\kappa)$*.*

Algorithm V *is* δ-prefix-simulatable *if membership in the guaranteed event* Δ *is determined by the coins* V *uses in the first* $\mathrm{round}(v)+1$ *rounds.*[4]

An interactive argument (P, V) *is* δ-simulatable/ δ-prefix-simulatable, if V is.

It is clear that random termination variant of an m-round interactive argument is $1/m$-prefix-simulatable.

Remark 1. One can relax the above definition and allow a different (non-black) simulator per P^*, and then only require it to exists for poly-time P^*. While our proof readily extends to this relaxation, we prefer to use the above definition for presentation clarity.

2.5.3 Parallel Repetition

Definition 8 (Parallel repetition). *Let* (P, V) *be an interactive protocol, and let* $n \in \mathbb{N}$*. We define the* n-parallel-repetition *of* (P, V) *to be the protocol* $(\mathrm{P}^n, \mathrm{V}^n)$ *in which* P^n *and* V^n *execute* n *copies of* (P, V) *in parallel, and at the end of the execution,* V^n *accepts if all copies accept.*

Black-box soundness reduction. As in most such proofs, our proof for the parallel repetition of partially-simulatable arguments has the following black-box form.

Definition 9 (Black-box reduction for parallel repetition). *Let* $\pi = (\mathrm{P}, \mathrm{V})$ *be an interactive argument. An oracle-aided algorithm* R *is a* black-box reduction for the g-soundness of the parallel repetition of π*, if the following holds for any poly-bounded* n*: let* $\kappa \in \mathbb{N}$ *and* P^{n*} *be deterministic cheating prover breaking the soundness of* $\pi^{n=n(\kappa)}(1^\kappa)$ *with probability* $\varepsilon' \geq g(n, \varepsilon = \varepsilon(\kappa))$*. Then*

Sucesss probability. $\mathsf{R} = \mathsf{R}^{\mathrm{P}^{n*}}(1^\kappa, 1^n)$ *breaks the soundness of* π *with probability at least* $1 - \varepsilon/3$*.*

Running time. *Except with probability* $\varepsilon/3$*, the running time of* R *is polynomial in* κ*, the running time of* $\mathrm{P}^{n*}(1^\kappa)$ *and* $1/\varepsilon'$*.*

We use the following fact (proven in the full version).

Proposition 2. *Assume there exists a black-box reduction for the* g*-soundness of the parallel repetition of any* δ*-simulatable [resp.,* δ*-prefix-simulatable] interactive argument, then for any poly-bounded* n*, the soundness error of the* n*-fold repetition of any such argument is bounded by* $g(n, \varepsilon)$*.*

[4] $\Delta = \Delta_1 \times \Delta_2$, for Δ_1 being a (δ-dense) subset of the possible values for first $\mathrm{round}(v)+1$ round coins, and Δ_2 is the set of all possible values for the coins used in rounds $\mathrm{round}(v)+2, \ldots, m$, for m being the round complexity of V.

3 Smooth KL-Divergence

In this section we formally define the notion of smooth KL-divergence, state some basic properties of this measure in Section 3.1, and develop a tool to help bounding it in Section 3.2.

Definition 10 (α-*smooth* divergence). *Let P and Q be two distributions over a universe $\mathcal{U}$ and let $\alpha \in [0,1]$. The* α-smooth divergence of P and Q, *denoted $D^{\alpha}(P||Q)$, is defined as $\inf_{(F_P,F_Q)\in\mathcal{F}}\{D(F_P(P)||F_Q(Q))\}$, for $\mathcal{F}$ being the set of randomized functions pairs such that for every $(F_P,F_Q)\in\mathcal{F}$:*

1. $\Pr_{x\sim P}[F_P(x)\neq x]\leq\alpha$, *where the probability is also over the coins of F_P.*
2. $\forall x\in\mathcal{U}$: $\mathrm{Supp}(F_P(x))\cap\mathcal{U}\subseteq\{x\}$ *and* $\mathrm{Supp}(F_Q(x))\cap\mathcal{U}\subseteq\{x\}$.

See the full version for comparison to the H-technique.

3.1 Basic Properties

The following proposition (proven in the full version) states that small smooth KL-divergence guarantees that small events with respect to the left-hand-side distribution are also small with respect to the right-hand-side distribution.

Proposition 3. *Let P and Q be two distributions over $\mathcal{U}$ with $D^{\alpha}(P||Q)<\beta$. Then for every event E over $\mathcal{U}$, it holds that $Q[E]<2\cdot\max\{\alpha+P[E],4\beta\}$.*

Like any useful distribution measure, smooth KL-divergence posses a data-processing property. The following proposition is proven in the full version.

Proposition 4 (Data processing of smooth KL-divergence). *Let P and Q be two distributions over a universe $\mathcal{U}$, let $\alpha\in[0,1]$ and let H be a randomized function over $\mathcal{U}$. Then $D^{\alpha}(H(P)||H(Q))\leq D^{\alpha}(P||Q)$.*

3.2 Bounding Smooth KL-Divergence

The following lemma allow us to bound the smooth KL-divergence between P and Q, while only analyzing simpler variants of Q.

Lemma 4 (Bounding smooth KL-Divergence, restatement of Lemma 2).
Let P and Q be distributions with P_X and Q_X being over universe $\mathcal{U}^m$, and let $A_1,\ldots,A_m$ and $B_1,\ldots,B_m$ be two sets of events over P and Q respectively. Let $P_{\cdot,XY}$ be an extension of $P=P_{\cdot,X}$ defined by $P_{Y|\cdot,X}=\prod_i P_{Y_i|X}$ for $P_{Y_i|X}=\mathrm{Bern}(P[A_i\mid X,A_{<i}]\cdot Q[B_i\mid X_{<i},B_{<i}])$, letting $P_{Y_i|X}=0$ if $P[A_{<i}\mid X]=0$ or $Q[B_{<i}\mid X_{<i}]=0$, and let $C_i=\{Y_i=1\}$. Then[5]

$$D^{1-P[C_{\leq m}]}(P_X||Q_X)\leq\sum_{i=1}^{m}D(P_{X_i|A_{\leq i}}||Q_{X_i|B_{\leq i}}\mid P_{X_{<i}|C_{\leq i}}).$$

[5] Note that Lemma 2 is a special case of Lemma 4 that holds when choosing $A_1,\ldots,A_m$ with $P[A_{\leq m}]=1$.

Proof. Let $Q_{\cdot,XY}$ be an extension of $Q = Q_{\cdot,X}$ defined by $Q_{Y|\cdot,X} = \prod_i Q_{Y_i|X}$ for $Q_{Y_i|X} = \mathrm{Bern}(P[A_i \mid X_{<i}, A_{<i}] \cdot Q[B_i \mid X, B_{<i}])$, letting $Q_{Y_i|X} = 0$ if $P[A_{<i} \mid X_{<i}] = 0$ or $Q[B_{<i} \mid X] = 0$. Our goal is to show that

$$D^{1-P[C_{\le m}]}(P_{Y_1,X_1,\ldots,Y_m,X_m} || Q_{Y_1,X_1,\ldots,Y_m,X_m}) \le \sum_{i=1}^{m} D(P_{X_i|A_{\le i}} || Q_{X_i|B_{\le i}} \mid P_{X_{<i}|C_{\le i}}) \tag{3}$$

The proof then follows by data processing of smooth KL-divergence (Proposition 4). By definition, for any $i \in [m]$:

$$P_{X_{<i}|Y_{\le i}=1^i} \equiv P_{X_{<i}|C_{\le i}} \tag{4}$$

and for any fixing of $x_{<i} \in \mathrm{Supp}(P_{X_{<i}|Y_{\le i}=1^i})$:

$$P_{X_i|Y_{\le i}=1^i, X_{<i}=x_{<i}} \equiv P_{X_i|X_{<i},A_{\le i}} \tag{5}$$

$$Q_{X_i|Y_{\le i}=1^i, X_{<i}=x_{<i}} \equiv Q_{X_i|X_{<i},B_{\le i}} \tag{6}$$

and for any fixing of $x_{<i} \in \mathrm{Supp}(P_{X_{<i}|Y_{<i}=1^{i-1}})$:

$$\begin{aligned} &P_{Y_i|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}(1) \\ &\equiv \mathrm{E}_{x \leftarrow P_{X|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}}[P[A_i \mid X = x, A_{<i}] \cdot Q[B_i \mid X_{<i} = x_{<i}, B_{<i}]] \\ &\equiv P[A_i \mid X_{<i} = x_{<i}, A_{<i}] \cdot Q[B_i \mid X_{<i} = x_{<i}, B_{<i}] \\ &\equiv \mathrm{E}_{x \leftarrow Q_{X|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}}[P[A_i \mid X_{<i} = x_{<i}, A_{<i}] \cdot Q[B_i \mid X = x, B_{<i}]] \\ &\equiv Q_{Y_i|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}(1). \end{aligned} \tag{7}$$

By Equations (4) to (6):

$$\mathrm{E}_{P_{X_{<i}|Y_{\le i}=1^i}}\left[D\left(P_{X_i|X_{<i},Y_{\le i}=1^i} || Q_{X_i|X_{<i},Y_{\le i}=1^i}\right)\right] = \mathrm{E}_{P_{X_{<i}|C_{\le i}}}\left[D\left(P_{X_i|X_{<i},A_{\le i}} || Q_{X_i|X_{<i},B_{\le i}}\right)\right] \tag{8}$$

and by Equation (7), for any fixing of $x \in \mathrm{Supp}(P_{X_{<i}|Y_{<i}=1^{i-1}})$:

$$D\left(P_{Y_i|X_{<i}=x,Y_{<i}=1^{i-1}} || Q_{Y_i|X_{<i}=x,Y_{<i}=1^{i-1}}\right) = 0 \tag{9}$$

We use Equations (8) and (9) for proving Equation (3), by applying on both distributions a function that "cuts" all values after the first appearance of $Y_i = 0$. Let $f_{\mathsf{cut}}(y_1, x_1, \ldots y_m, x_m) = (y_1, x_1, \ldots y_m, x_m)$ if $y = (y_1, \ldots, y_m) = 1^m$, and $f_{\mathsf{cut}}(y_1, x_1, \ldots y_m, x_m) = (y_1, x_1, \ldots y_{i-1}, x_{i-1}, y_i, \perp^{2n-2i+1})$ otherwise, where i is the minimal index with $y_i = 0$, and $\perp$ is an arbitrary symbol $\notin \mathcal{U}$. By definition,

$$\Pr_{s \sim P_{Y_1,X_1,\ldots,Y_m,X_m}}[f_{\mathsf{cut}}(s) \neq s] = P[Y \neq 1^m] = 1 - P[C_{\le m}],$$

and by Equations (8) and (9) along with data-processing of standard KL-divergence (Fact 4(3)),

$$D(f_{\mathsf{cut}}(P_{Y_1,X_1,\ldots,Y_m,X_m}) || f_{\mathsf{cut}}(Q_{Y_1,X_1,\ldots,Y_m,X_m})) \le \sum_{i=1}^{m} D(P_{X_i|A_{\le i}} || Q_{X_i|B_{\le i}} \mid P_{X_{<i}|C_{\le i}}).$$

That is, f_{cut} is the function realizing the stated bound on the smooth KL-divergence of P_X and Q_X.

4 Skewed Distributions

In this section we formally define the notion of many-round skewed distributions and state our main result for such distributions.

Definition 11 (The skewed distribution Q). *Let P be a distribution with P_X being a distribution over $m \times n$ matrices, and let W and $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ be events over P. We define the skewed distribution $Q_{X,J} = Q(P, W, \mathcal{E})$ of $\widetilde{P}_X = P|W$, by $Q_J = U_{[n]}$ and*

$$Q_{X|J} = \prod_{i=1}^{m} P_{X_{i,J}|X_{<i,J}} \widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J},E_{i,J}}$$

Definition 12 (dense and prefix events). *Let P_X be a distribution over $m \times n$ matrices, and let $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ be an event family over P_X such that $E_{i,j}$, for each i, j, is determined by X^j. The family $\mathcal{E}$ has* density δ *if $\forall(i,j) \in [m] \times [n]$ and for any fixing of $X_{\leq i,j}$, it holds that $P[E_{i,j}|X_{\leq i,j}] = \delta_{i,j} \geq \delta$. The family $\mathcal{E}$ is a* prefix family *if $\forall(i,j) \in [m] \times [n]$ the event $E_{i,j}$ is determined by $X_{\leq i+1,j}$.*

Bounding smooth KL-divergence of smooth distributions. The following theorem states our main result for skewed distributions. In Section 6 we give a proof sketch of Theorem 8, and in the full version we give the full details.

Theorem 8. *Let P be a distribution with P_X being a distribution over $m \times n$ matrices with independent columns, let W be an event over P and let $\mathcal{E} = \{E_{i,j}\}$ be a δ-dense event family over P_X. Let $\widetilde{P} = P|W$ and let $Q_{X,J} = Q(P, W, \mathcal{E})$ be the skewed variant of $\widetilde{P}$ defined in Definition 11. Let $Y_i = (Y_{i,1}, \ldots, Y_{i,n})$ for $Y_{i,j}$ being the indicator for $E_{i,j}$, and let $d = \sum_{i=1}^{m} D(\widetilde{P}_{X_iY_i}||P_{X_iY_i}|\widetilde{P}_{X_{<i}})$. Assuming $n \geq c \cdot m/\delta$ and $d \leq \delta n/c$, for a universal constant $c > 0$, then*

$$D^{\frac{c}{\delta n}(d+1)}(\widetilde{P}||Q) \leq \frac{c}{\delta n}(d+m).$$

We now prove that Theorem 1 is an immediate corollary of Theorem 8.

Corollary 1 (Restatement of Theorem 1). *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta$ and c be as in Theorem 8, and let $\varepsilon = \log(\frac{1}{P[W]})/\delta n$. Then the following hold assuming $n \geq c \cdot m/\delta$:*

- *if $P[W] \geq \exp(-\delta n/cm)$, then $D^{c\cdot(\varepsilon m+1/\delta n)}(\widetilde{P}||Q) \leq c \cdot (\varepsilon m + m/\delta n)$, and*
- *if $P[W] \geq \exp(-\delta n/2c)$ and $\mathcal{E}$ is a prefix family, then $D^{2c\cdot(\varepsilon+1/\delta n)}(\widetilde{P}||Q) \leq 2c \cdot (\varepsilon + m/\delta n)$.*

Proof. Let $\{Y_{i,j}\}$ be as in Theorem 8. Note that for each $i \in [m]$:

$$D(\widetilde{P}_{X_iY_i}||P_{X_iY_i} \mid \widetilde{P}_{X_{<i}}) \leq D(\widetilde{P}_{X_{\geq i}}||P_{X_{\geq i}} \mid \widetilde{P}_{X_{<j}}) \leq D(\widetilde{P}_X||P_X) \leq \log \frac{1}{P[W]}.$$

The first inequality holds by data-processing of KL-divergence (Fact 4(5)). The second inequality holds by chain-rule of KL-divergence (Fact 4(3)). The last inequality holds by Fact 5. Assuming $P[W] \geq \exp(-\delta n/cm)$, it holds that

$$d \leq m \cdot \log \frac{1}{P[W]} \leq \delta n/c,$$

concluding the proof of the first part.

Assuming $P[W] \geq \exp(-\delta n/c)$ and $\mathcal{E}$ is a prefix family (i.e., $E_{i,j}$ is a function of $X_{\leq i+1}$), then

$$\begin{aligned}
d &\leq \sum_{i=1}^{m-1} D(\widetilde{P}_{X_i X_{i+1}} || P_{X_i X_{i+1}} \mid \widetilde{P}_{X_{<i}}) + D(\widetilde{P}_{X_m} || P_{X_m} \mid \widetilde{P}_{X_{<m}}) \\
&= \sum_{i \in [m-1] \cap \mathbb{N}_{even}} D(\widetilde{P}_{X_i X_{i+1}} || P_{X_i X_{i+1}} \mid \widetilde{P}_{X_{<i}}) + \sum_{i \in [m-1] \cap \mathbb{N}_{odd}} D(\widetilde{P}_{X_i X_{i+1}} || P_{X_i X_{i+1}} \mid \widetilde{P}_{X_{<i}}) \\
&+ D(\widetilde{P}_{X_m} || P_{X_m} \mid \widetilde{P}_{X_{<m}}) \leq 2 \cdot D(\widetilde{P}_X || P_X) \\
&\leq 2 \cdot \log \frac{1}{P[W]} \leq \delta n/c,
\end{aligned}$$

concluding the proof of the second part. The first inequality holds by data-processing of KL-divergence, and the second one holds by chain-rule and data-processing of KL-divergence.

In order to show that the attacking distribution Q can be carried out efficiently, it suffice to show that with high probability over $(x, j) \sim Q_{X,J}$, we have for all $i \in [m]$ that $P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}]$ is not much smaller than $P[W]$. The following lemma (proven in the full version) states that the above holds under $\widetilde{P}_X$. Namely, when sampling $x \sim \widetilde{P}_X$ (instead of $x \sim Q_X$) and then $j \sim Q_{J|X=x}$, then $P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}]$ is indeed not too low.

Lemma 5. *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta, d$ be as in Theorem 8, let $t > 0$ and let*

$$p_t := \Pr_{x \sim \widetilde{P}_X \,;\, j \sim Q_{J|X=x}} [\exists i \in [m] : P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] < P[W]/t]$$

Assuming $n \geq c \cdot m/\delta$ and $d \leq \delta n/c$, for a universal constant $c > 0$, then

$$p_t \leq 2m/t + c(d+1)/(\delta n).$$

As an immediate corollary, we get the following result.

Corollary 2. *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta$ be as in Theorem 8, let $\varepsilon = \log(\frac{1}{P[W]})/\delta n$, let $t > 0$ and let c and p_t as in Lemma 5. Assuming $n \geq c \cdot m/\delta$, it holds that*

- *if $P[W] \geq \exp(-\delta n/cm)$, then $p_t \leq 2m/t + c \cdot (\varepsilon m + 1/\delta n)$.*
- *if $P[W] \geq \exp(-\delta n/2c)$ and $\mathcal{E}$ is a prefix family, then $p_t \leq 2m/t + 2c \cdot (\varepsilon + 1/\delta n)$.*

5 The Parallel Repetition Theorem

In this section, we use Theorem 8 to prove Theorem 2, restated below.

Theorem 9 (Parallel repetition for partially simulatable arguments, restatement of Theorem 2). *Let π be an m-round δ-simulatable [resp., prefix δ-simulatable] interactive argument of soundness error $1-\varepsilon$. Then π^n has soundness error $(1-\varepsilon)^{cn\delta/m}$ [resp., $(1-\varepsilon)^{cn\delta}$], for a universal constant $c>0$.*

Since the random terminating variant of an m-round interactive argument is $1/m$-prefix-simulatable, the (tight) result for such protocols immediately follows. The proof of Theorem 9 follows from our bound on the smooth KL-divergence of skewed distributions, Theorem 8, and Lemma 6, stated and proven below.

Definition 13 (bounding function for many-round skewing). *A function f is a* bounding function for many-round skewing *if there exists a polynomial $p(\cdot,\cdot)$ such that the following holds for every $\delta\in(0,1]$ and every $m,n\in\mathbb{N}$ with $n>p(m,1/\delta)$: let P be a distribution with P_X being a column independent distribution over $m\times n$ matrices. Let W be an event and let $\mathcal{E}$ be a δ-dense [resp., prefix δ-dense] event family over P (see Definition 12). Let $\widetilde{P}=P|W$ and let $Q=Q(P,W,\mathcal{E})$ be according to Definition 11. Then the following holds for $\gamma=\log(1/P[W])/f(n,m,\delta)$:*

1. *$Q_X[T]\le 2\cdot\widetilde{P}_X[T]+\gamma$ for every event T,[6] and*
2. *$\forall t>0:\ \Pr_{x\sim\widetilde{P}_X\ ;\ j\sim Q_{J|X=x}}[(x,j)\in\mathsf{Bad}_t]\quad\le\quad p(m,1/\delta)/t+\gamma$, letting*

$$\mathsf{Bad}_t:=\{(x,j)\colon\exists i\in[m]:P[W\mid(X_{<i},X_{i,j})=(x_{<i},x_{i,j}),E_{i,j}]<P[W]/t\}.$$

Lemma 6 (Restatement of Lemma 1). *Let π be an m-round δ-simulatable [resp., prefix δ-simulatable] interactive argument of soundness error $1-\varepsilon$, let f be a bounding function for many-round skewing (according to Definition 13). Then π^n has soundness error $(1-\varepsilon)^{f(n,m,\delta)/80}$.*

That is, Lemma 6 tells us that the task of maximizing the decreasing rate of π^n directly reduces to the task of maximizing a bounding function for many-round skewing. A larger bounding function yields a smaller γ in Definition 13. This γ both defines an additive bound on the difference between a small event in $\widetilde{P}$ to a small event in Q, and bounds a specific event in $\widetilde{P}$ that captures the cases in which an attack can be performed efficiently.

We first prove Theorem 9 using Lemma 6.

[6] The constant 2 can be replaced with any other constant without changing (up to a constant factor) the decreasing rate which is promised by Lemma 6.

Proof of Theorem 9.

Proof. We prove for δ-simulatable arguments, the proof for δ-prefix-simulatable arguments follows accordingly. Let m, n, P, δ, $\mathcal{E}$, W, $\widetilde{P}$ and Q be as in Lemma 6, where $\mathcal{E}$ is δ-dense, and let $c = \max\{c', c''\}$ where c' is the constant from Corollary 1 and c'' is the constant from Corollary 2. By Corollary 1, if $n \geq c \cdot m/\delta$ and $P[W] \geq \exp(-\delta n/cm)$, then

$$D^{3cm\mu}(\widetilde{P}||Q) \leq 3cm\mu \tag{10}$$

for $\mu = \log(1/P[W])/\delta n$, where we assumed without loss of generality that $P[W] \leq 1/2$. Hence, assuming that $n \geq c \cdot m/\delta$ and $P[W] \geq \exp(-\delta n/cm)$, Proposition 3 and eq. (10) yields that for every event T:

$$Q[T] \leq 2 \cdot \widetilde{P}[T] + \gamma, \tag{11}$$

where $\gamma = \log(1/P[W])/f(n, m, \delta)$ for $f(n, m, \delta) = \delta n/(24cm)$. For an event W of smaller probability, it holds that $\gamma \geq 24$, and therefore Equation (11) trivially holds for such events. In addition, by Corollary 2, if $n \geq c \cdot m/\delta$ then

$$\Pr_{x\sim\widetilde{P}_X\ ;\ j\sim Q_{J|X=x}}[\exists i \in [m] : P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] < P[W]/t] \leq 2m/t + \gamma \tag{12}$$

(We assume $P[W] \geq \exp(-\delta n/cm)$, as otherwise Equation (12) trivially holds.) By Equations (11) and (12), f is a bounding function for many-round skewing with the polynomial $p(m, 1/\delta) = c \cdot m/\delta$. Therefore, Lemma 6 yields that the soundness error of π^n is bounded by $(1-\varepsilon)^{f(n,m,\delta)/80} = (1-\varepsilon)^{\delta n/(c'm)}$, for $c' = 1920c$.

5.1 Proving Lemma 6

Let f be a bounding function for many-round skewing with the polynomial $p(\cdot, \cdot) \in \text{poly}$. We first prove the case when the number of repetition n is at least $p(m, 1/\delta)$, and then show how to extend the proof for the general case.

Many repetitions case.

Proof (Proof of Lemma 6, many repetitions). Fix an m-round δ-simulatable interactive argument $\pi = (\mathrm{P}, \mathrm{V})$ of soundness error $1-\varepsilon$ (the proof of the δ-prefix-simulatable case follows the same lines), and let $n = n(\kappa) > p(m(\kappa), 1/\delta(\kappa))$. Note that without loss of generality $\varepsilon(\kappa) \geq 1/\text{poly}(\kappa)$.

Our proof is a black-box reduction according to Definition 9: we present an oracle-aided algorithm that given access to a deterministic cheating prover for π^n violating the claimed soundness of π^n, uses it to break the assumed soundness of π while not running for too long. The lemma then follows by Proposition 2.

Let S be the oracle-aided simulator guaranteed by the δ-simulatablily of V. For a cheating prover P^{n*} for π^n, let P^* be the cheating prover that when

interacting with V, emulates a random execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$, letting V plays one of the n verifiers at a random location. (Clearly, P^* only requires oracle access to P^{n*}.) Assume without loss of generality that in each round V flips $t = t(\kappa)$ coins. The oracle-aided algorithm P^* is defined as follows.

Algorithm 10 (P^*)
Input: 1^κ, $m = m(\kappa)$ *and* $n = n(\kappa)$.
Oracles: cheating prover P^{n*} *for* π^n.
Operation:

1. *Let* $j \leftarrow [n]$.
2. *For* $i = 1$ *to* m *do:*
   (a) *Let* a_i *be the* i^{th} *message sent by* V.
   (b) *Do the following ("rejection continuation"):*
      i. *Let* $x_{i,-j} \leftarrow (\{0,1\}^t)^{n-1}$
      ii. *Let* $v = \mathsf{S}^{\mathrm{P}^{n*}}(1^\kappa, (j, x_{\leq i,-j}, a_{\leq i}))$.
      iii. *If all* n *verifiers accept in* v, *break the inner loop.*
   (c) *Send to* V *the* i^{th} *message* P^{n*} *sends in* v.

Fix a cheating prover P^{n*}. We also fix $\kappa \in \mathbb{N}$, and omit it from the notation. Let $P = P_X$ denotes the coins V^n uses in a uniform execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$. (Hence P_X is uniformly distributed over $m \times n$ matrices.) Let W be the event over P that P^{n*} wins in $(\mathrm{P}^{n*}, \mathrm{V}^n)$ (i.e., all verifiers accept), and let $\widetilde{P}_X = P_X|W$. For an i rounds view $v = (j, \cdot)$ of P^{n*} in $(\mathrm{P}^{n*}, \mathrm{V})$, let Δ_v be the δ-dense subset of V's coins describing the output distribution of $\mathsf{S}^{\mathrm{P}^{n*}}(v)$. Let $\mathcal{T}_{i,j}$ be all possible i round views of P^{n*} in $(\mathrm{P}^{n*}, \mathrm{V})$ that are starting with j. Finally, let $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ be the event family over P defined by $E_{i,j} = \bigcup_{v\in\mathcal{T}_{i,j}} \Delta_v$, and let $Q_{X,J}$ be the e (skewed) distribution described in Definition 11 with respect to $P, W, \mathcal{E}$. By inspection, Q describes the distribution of $(j, x_{\leq m})$ in a random execution of $(\mathrm{P}^*, \mathrm{V}^n)$, where $x_{\leq m,j}$ denotes the coins of V, and $x_{\leq m,-j}$ denote the final value of this term in the execution. Assume

$$\Pr[(\mathrm{P}^{n*}, \mathrm{V}^n) = 1] = P[W] > (1-\varepsilon)^{f(n,m,\delta)/80}, \tag{13}$$

and let $\gamma = \log(1/P[W])/f(n, m, \delta)$. By Equation (13) it holds that

$$\gamma < -\log(1-\varepsilon)/80 \leq \varepsilon/80 \tag{14}$$

Since $\widetilde{P}[W] = 1$, we deduce by Property 13(1) of f on the event $\neg W$ that

$$\Pr[(\mathrm{P}^*, \mathrm{V}) = 1] \geq Q_X[W] > 1 - \gamma > 1 - \varepsilon/80 \tag{15}$$

So it is left to argue about the running time of P^*. By Property 13(2) of f on $t = 80 \cdot p(m, 1/\delta)/\varepsilon$ it holds that

$$\Pr_{x\sim\widetilde{P}_X\ ;\ j\sim Q_{J|X=x}}[(x, j) \in \mathsf{Bad}_t] \quad \leq \quad p(m, 1/\delta)/t + \gamma < \varepsilon/40 \tag{16}$$

Consider the extension $\widetilde{P}_{XJ}$ of $\widetilde{P}_X$, where $\widetilde{P}_{J|X} = Q_{J|X}$. Note that by Property 13(1), for any event T over (X, J) it holds that $Q_{XJ}[T] \leq 2 \cdot \widetilde{P}_{XJ}[T] + \gamma$. In particular, this holds for the event $(X, J) \in \mathsf{Bad}_t$. Therefore, we deduce from Equations (14) and (16) that

$$\Pr_{x \sim Q_X \,;\, j \sim Q_{J|X=x}}[(x, j) \in \mathsf{Bad}_t] \quad \leq \quad 2\varepsilon/40 + \gamma < \varepsilon/10 \tag{17}$$

By Equations (15) and (17) we obtain that

$$\Pr_{(x,j) \sim Q_{X,J}}[W \wedge ((x, j) \notin \mathsf{Bad}_t)] > 1 - \varepsilon/5 \tag{18}$$

Namely, with probability larger than $1 - \varepsilon/5$, the attacker P^* wins and its expected running time in each round is bounded by $O(t/P[W]) \leq \mathrm{poly}(\kappa)$. This contradicts the soundness guaranty of π.

Any number of repetitions. See the full version.

6 Bounding Smooth KL-Divergence of Skewed Distributions - Proof Sketch

In this section we give a rather detailed proof sketch (more accurately, an attempt proof sketch) for proving Theorem 8 in which we explain the difficulties that arise. The actual proof appears in the full version due to page limitation.

Fix a distribution P with P_X being a distribution over $\mathcal{U}^{m \times n}$ matrices with independent columns, event W over P and δ-dense event family $\mathcal{E} = \{E_{i,j}\}$ over P_X. Let $\widetilde{P} = P|W$ and let $Q_{X,J} = Q(P, W, \mathcal{E})$ be the skewed variant of $\widetilde{P}$ defined in Definition 11. Let $Y_i = (Y_{i,1}, \ldots, Y_{i,n})$ for $Y_{i,j}$ be the indicator for $E_{i,j}$, and let $d = \sum_{i=1}^{m} D(\widetilde{P}_{X_i Y_i} || P_{X_i Y_i} | \widetilde{P}_{X_{<i}})$.

In the following we present an attempt to bound the divergence between $\widetilde{P}$ and Q. That is, to show that

$$D(\widetilde{P}||Q) \leq O\left(\frac{1}{\delta n}\right) \cdot (d + m) \tag{19}$$

We try to do so by showing that for every $i \in [m]$ it holds that

$$D(\widetilde{P}_{X_i}||Q_{X_i}|\widetilde{P}_{X_{<i}}) \leq O\left(\frac{1}{\delta n}\right) \cdot (d_i + 1) \tag{20}$$

for $d_i = D(\widetilde{P}_{X_i Y_i}||P_{X_i Y_i}|\widetilde{P}_{X_{<i}})$, and applying chain-rule of KL-divergence for deducing Equation (19). By data-processing of KL-divergence (Fact 4(5)), it holds that

$$D(\widetilde{P}_{X_i}||Q_{X_i}|\widetilde{P}_{X_{<i}}) \leq D(\widetilde{P}_{X_i Y_i}||Q'_{X_i Y_i}|\widetilde{P}_{X_{<i}}), \tag{21}$$

where

$$Q'_{X_iY_i|X_{<i}} = \widetilde{P}_{X_iY_i|X_{<i},X_{i,J},Y_{i,J}=1} \circ Q_{J,X_{i,J}|X_{<i}} \equiv P_{X_{i,J}|X_{<i}} \widetilde{P}_{X_iY_i|X_{<i},X_{i,J},Y_{i,J}=1} \circ Q_{J|X_{<i}}$$

(note that $Q'_{X_i} \equiv Q_{X_i}$ and that $P_{X_{i,J}|X_{<i}} \equiv P_{X_{i,J}|X_{<i,J}}$ because the columns under P are independent). By definition of Q', for any fixing of $x_{\leq i}y_i \in \mathrm{Supp}(\widetilde{P}_{X_{\leq i}Y_i})$ it holds that

$$\begin{aligned}
&Q'_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i) \qquad (22)\\
&= \mathrm{E}_{j\sim Q_{J|X_{<i}=x_{<i}}}\left[P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}) \cdot \widetilde{P}_{X_iY_i|X_{<i}=x_{<i},X_{i,j}=x_{i,j},Y_{i,j}=1}(x_iy_i)\right]\\
&= \sum_{j=1}^{n} Q_{J|X_{<i}=x_{<i}}(j) \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}) \cdot \frac{\widetilde{P}_{X_iY_iX_{i,j}Y_{i,j}|X_{<i}=x_{<i}}(x_iy_ix_{i,j}1)}{\widetilde{P}_{X_{i,j},Y_{i,j}|X_{<i}=x_{<i}}(x_{i,j},1)}\\
&= \sum_{j\in 1_{y_i}} Q_{J|X_{<i}=x_{<i}}(j) \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}) \cdot \frac{\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}{\widetilde{P}_{X_{i,j},Y_{i,j}|X_{<i}=x_{<i}}(x_{i,j},1)}\\
&= \sum_{j\in 1_{y_i}} Q_{J|X_{<i}=x_{<i}}(j) \cdot \frac{\beta_{i,j}(x_{i,j}) \cdot \widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}{\tilde{\delta}_{i,j}},
\end{aligned}$$

for $\beta_{i,j}(x_{i,j}) = \beta_{i,j}(x_{i,j};x_{<i}) = \frac{P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})}{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}(x_{i,j})}$ and $\tilde{\delta}_{i,j} = \tilde{\delta}_{i,j}(x_{<i}) = \widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}(1)$ $(= \widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}])$, where recall that we denote $1_{y_i} = \{j \in [n] \colon y_{i,j} = 1\}$. We now use the following claim (proven in the full version) that calculates the probability to get j in the conditional distribution $Q_{J|X_{<i}=x_{<i}}$.

Claim. Let $\omega'_{i,j} = \omega'_{i,j}(x_{<i}) := \prod_{s=1}^{i-1} \frac{P[X_{s,j}=x_{s,j}|X_{<s}=x_{<s}]}{\widetilde{P}[X_{s,j}=x_{s,j}|X_{<s}=x_{<s}]}$ and let

$$\omega_{i,j} = \omega_{i,j}(x_{<i}) := \frac{n \cdot \omega'_{i,j}}{\sum_{t=1}^{n} \omega'_{i,t}} \cdot \prod_{s=1}^{i-1} \frac{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}, X_{s,j} = x_{s,j}]} \cdot \frac{\widetilde{P}[E_{s,j} \mid X_{\leq s} = x_{\leq s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]}.$$

Then it holds that

$$Q_{J|X_{<i}=x_{<i}}(j) = \frac{\omega_{i,j}}{\sum_{t=1}^{n} \omega_{i,t}}.$$

Note that $\omega_{i,j}$ is basically a relative "weight" for the column j, where a large $\omega_{i,j}$ with respect to the other $\omega_{i,t}$'s means that $Q_{J|X_{<i}=x_{<i}}(j)$ is higher. In an extreme case it is possible that $\omega_{i,j} = \infty$, meaning that $Q_{J|X_{<i}=x_{<i}}(j) = 1$. However, we assume for now that all $\omega_{i,j} < \infty$. Later in this proof attempt we even assume that all the terms are close to 1, meaning that $Q_{J|X_{<i}=x_{<i}}$ has high min entropy (assumptions that are eliminated in the full version). As a side note, observe that $\omega_{1,j} = 1$ for all $j \in [n]$ (meaning that Q_J, without any conditioning, is the uniform distribution over $[n]$). At this point, we just mention that we added (the same) multiplicative factor of $\frac{n}{\sum_{t=1}^{n} \omega'_{i,t}}$ to all $\{\omega_{i,j}\}_{j=1}^{n}$. On the one hand this does not change the relative weight, but on the other hand it

will help us to claim in the full version that these $\omega_{i,j}$'s are indeed close to 1. By Equations (21) and (22) and Claim 6, it holds that

$$\begin{aligned} D(\widetilde{P}_{X_i}||Q_{X_i}|\widetilde{P}_{X_{<i}}) &\leq D(\widetilde{P}_{X_iY_i}||Q'_{X_iY_i}|\widetilde{P}_{X_{<i}}) \qquad (23)\\ &= \mathrm{E}_{x_{<i}\sim X_{<i}}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\log\frac{\widetilde{P}_{X_iY_i|X_{<i}=X_{<i}}(x_iy_i)}{Q_{X_iY_i|X_{<i}=X_{<i}}(x_iy_i)}\right]\\ &= \mathrm{E}_{x_{<i}\sim X_{<i}}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\log\frac{\sum_{j=1}^n\omega_{i,j}}{\sum_{j\in 1_{y_i}}\frac{\omega_{i,j}\cdot\beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}}\right]\\ &= \mathrm{E}_{x_{<i}\sim X_{<i}}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}[-\log(1+\gamma_i(x_iy_i))], \end{aligned}$$

for

$$\gamma_i(x_iy_i)=\gamma_i(x_iy_i;x_{<i})=\left(\sum_{j\in 1_{y_i}}\frac{\omega_{i,j}\cdot\beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right)/\left(\sum_{j=1}^n\omega_{i,j}\right)-1 \qquad (24)$$

Naturally, we would like to approximate the logarithm in the above equation with a low-degree polynomial. However, we can only do if γ_i is far away from -1. In particular, if $\widetilde{P}[\gamma_i(X_iY_i;X_{<i})=-1]>0$ (which happens if the event W allows for none of the events $\{E_{i,j}\}_{i=1}^n$ to occur), the above expectation is unbounded. At that point, we only show how to bound Equation (23) under simplifying assumptions, while in the full version we present how to eliminate the assumptions via smooth KL-divergence. We now assume that for any $x_{<i}\in\mathrm{Supp}(\widetilde{P}_{X_{<i}})$ and any $j\in[n]$, the following holds:

Assumption 11

1. $|\gamma_i(x_iy_i)|\leq 1/2$ *for any* $x_iy_i\in\mathrm{Supp}(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}})$.
2. $\tilde{\delta}_{i,j}\geq 0.9\delta_{i,j}$ *(recall that* $\delta_{i,j}=P[E_{i,j}]=P[E_{i,j}\mid X_{\leq i}]$ *for any fixing of* $X_{\leq i}$*).*
3. $\omega_{i,j}\in 1\pm 0.1$.
4. $\mathrm{Supp}(P_{X_{i,j}|X_{<i}=x_{<i}})\subseteq\mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})$.
5. $\beta_{i,j}(x_{i,j})\leq 1.1$ *for any* $x_{i,j}\in\mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i}})$.

Note that Assumption 3 implies that $Q_{J|X_{<i}}$ has high min-entropy, and Assumptions 2 along with 5 imply that for all j:

$$\begin{aligned} &P[W\mid(X_{<i},X_{i,j})=(x_{<i},x_{i,j}),E_{i,j}]\\ &=\frac{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})}{P_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})}\cdot\frac{\widetilde{P}[E_{i,j}\mid X_{<i}=x_{<i}]}{P[E_{i,j}\mid X_{<i}=x_{<i}]}\cdot P[W\mid X_{<i}=x_{<i}]\\ &=\beta_{i,j}(x_{i,j})\cdot\left(\tilde{\delta}_{i,j}/\delta_{i,j}\right)\cdot P[W\mid X_{<i}=x_{<i}]\geq P[W\mid X_{<i}=x_{<i}]/2, \end{aligned}$$

which fits the explanation in Section 1.4.1 (note that in the second equality we used the fact that $P_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j}) = P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})$ by assumption). By Equation (23), note that in order to prove Equation (20), it is enough to show that for any $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{x_{<i}})$ it holds that

$$\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}[-\log(1+\gamma_i(x_i y_i))] \leq O\left(\frac{1}{\delta n}\right) \cdot \left(D(\widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}} || P_{X_i Y_i | X_{<i}=x_{<i}}) + 1\right) \tag{25}$$

In the following, fix $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{x_{<i}})$. We now focus on proving Equation (25). Using the inequality $-\log(1+x) \leq -x + x^2$ for $|x| \leq \frac{1}{2}$, we deduce from Assumption 1 that

$$\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}[-\log(1+\gamma_i(x_i y_i))] \leq \mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[-\gamma_i(x_i y_i) + \gamma_i(x_i y_i)^2\right] \tag{26}$$

Note that

$$\begin{aligned}
&\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[\sum_{j \in 1_{y_i}} \frac{\omega_{i,j} \cdot \beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right] \\
&= \sum_{j=1}^{n} \mathrm{E}_{x_{i,j} y_{i,j} \sim \widetilde{P}_{X_{i,j} Y_{i,j} | X_{<i}=x_{<i}}}\left[y_{i,j} \cdot \frac{\omega_{i,j} \cdot \beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right] \\
&= \sum_{j=1}^{n} \omega_{i,j} \cdot \mathrm{E}_{x_{i,j} \sim \widetilde{P}_{X_{i,j} | X_{<i}=x_{<i}, Y_{i,j}=1}}[\beta_{i,j}(x_{i,j})] \\
&= \sum_{j=1}^{n} \omega_{i,j} \cdot \mathrm{E}_{x_{i,j} \sim \widetilde{P}_{X_{i,j} | X_{<i}=x_{<i}, Y_{i,j}=1}}\left[\frac{P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})}{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}(x_{i,j})}\right] \\
&= \sum_{j=1}^{n} \omega_{i,j} \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})) = \sum_{j=1}^{n} \omega_{i,j}.
\end{aligned} \tag{27}$$

The second equality holds since $y_{i,j} \in \{0,1\}$ and since Assumption 2 implies that $\widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}(1) = \tilde{\delta}_{i,j} > 0$ for all $j \in [n]$, and the last equality holds by Assumption 4. Therefore, we deduce from Equation (27) that

$$\begin{aligned}
&\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}[\gamma_i(x_i y_i)] \\
&= \left(\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[\sum_{j \in 1_{y_i}} \frac{\omega_{i,j} \cdot \beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right]\right) / \left(\sum_{j=1}^{n} \omega_{i,j}\right) - 1 = 0.
\end{aligned} \tag{28}$$

Hence, in order to prove Equation (25), we deduce from Equations (26) and (28) that it is left to prove that

$$\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\gamma_i(x_iy_i)^2\right] \le O\left(\frac{1}{\delta n}\right)\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||P_{X_iY_i|X_{<i}=x_{<i}})+1\right) \tag{29}$$

In the following, rather than directly bounding the expected value of $\gamma_i(x_iy_i)^2$ under $\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}$, we show that under the product of the marginals of $\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}$ (namely, under the distribution $\prod_{j=1}^{n}\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$), the value of $\gamma_i(x_iy_i)$ is well concentrated around its mean (i.e., zero), and the proof will follow by Proposition 1. More formally, let Γ be the value of $\gamma_i(x_iy_i)$ when x_iy_i is drawn from either $\widetilde{P}=\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}$ or $\widetilde{P}^{\Pi}=\prod_{j=1}^{n}\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$. We prove that there exist two constants $K_1,K_2>0$ such that for any $\gamma\in[0,1]$:

$$\widetilde{P}^{\Pi}[|\Gamma|\ge\gamma]\le K_2\cdot\exp\left(-\frac{\gamma^2}{K_1\cdot\sigma^2}\right) \tag{30}$$

for $\sigma^2=1/\delta n$. Using Equation (30) and the fact that $|\Gamma|\le 1$ (Assumption 1), Proposition 1 yields that

$$\begin{aligned}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\gamma_i(x_iy_i)^2\right]=\mathrm{E}_{\widetilde{P}}\left[\Gamma^2\right] &\le \frac{K_3}{\delta n}\cdot\left(D(\widetilde{P}||\widetilde{P}^{\Pi})+1\right) \qquad (31)\\ &= \frac{K_3}{\delta n}\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||\prod_{j=1}^{n}\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}})+1\right)\\ &\le \frac{K_3}{\delta n}\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||P_{X_iY_i|X_{<i}=x_{<i}})+1\right).\end{aligned}$$

The last inequality holds by chain rule of KL-divergence when the right-hand side distribution is product (Fact 4(3), where recall that $P_{X_iY_i|X_{<i}=x_{<i}}=\prod_{j=1}^{n}P_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$). This concludes the proof of Equation (29). It is left to prove Equation (30). In the following, given x_iy_i which are drawn from either $\widetilde{P}^{\Pi}=\prod_{j=1}^{n}\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$ or $\widetilde{P}^{\Pi'}=\prod_{j=1}^{n}\widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}\cdot\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}$, we define the random variables L_j,Z_j,L and Z (in addition to Γ), where L_j is the value of $\omega_j\cdot\beta_j(x_{i,j})$, $L=\sum_{j=1}^{n}L_j$, $Z_j=\begin{cases}L_j/\tilde{\delta}_j & y_{i,j}=1\\ 0 & y_{i,j}=0\end{cases}$ and $Z=\sum_{j=1}^{n}Z_j$, letting $\omega_j=\omega_{i,j}$, $\beta_j(\cdot)=\beta_{i,j}(\cdot)$ and $\tilde{\delta}_j=\tilde{\delta}_{i,j}$. Note that by definition, $Z=(1+\Gamma)\mu$ for $\mu=\sum_{j=1}^{n}\omega_j$. Namely, Γ measures how far Z is from its expected value μ (follows by Equation (27) that calculates $\mathrm{E}_{\widetilde{P}}[Z]$, which also equals to $\mathrm{E}_{\widetilde{P}^{\Pi}}[Z]$ and $\mathrm{E}_{\widetilde{P}^{\Pi'}}[Z]$). Note that the distribution of Z and Γ when x_iy_i is drawn from $\widetilde{P}^{\Pi}$ is identical to the distribution of Z and Γ (respectively) when x_iy_i is drawn from $\widetilde{P}^{\Pi'}$. Therefore, in particular it holds that

$$\widetilde{P}^{\Pi}[|\Gamma|\ge\gamma]=\widetilde{P}^{\Pi'}[|\Gamma|\ge\gamma] \tag{32}$$

Under $\widetilde{P}^{\Pi'}$, the L_j's are independent random variables with $\mathrm{E}_{\widetilde{P}^{\Pi'}}[L_j] = \omega_j$ and $\mathrm{E}_{\widetilde{P}^{\Pi'}}[L] = \mu$ where $\mu = \sum_{j=1}^{n} \omega_j \geq n/2$ and $|L_j| \leq 2$ (by Assumptions 3 and 5). Moreover, for all $j \in [n]$, $Z_j = (L_j/\tilde{\delta}_j) \cdot \mathrm{Bern}(\tilde{\delta}_j)$ where $\tilde{\delta}_j \geq 0.9\delta_{i,j} \geq 0.9\delta$ (by Assumption 2). Hence, Fact 7 yields that

$$\widetilde{P}^{\Pi'}[|\Gamma| \geq \gamma] \leq 4\exp\left(-\frac{\delta n \gamma^2}{100}\right) \tag{33}$$

The proof of Equation (30) now follows by Equations (32) and (33), which ends the proof of Theorem 8 under the assumptions in 11.

6.1 Eliminating the Assumptions

The assumptions we made in 11 may seem unjustified at first glance. For instance, even for $j = 1$, there could be "bad" columns $j \in [n]$ with $\tilde{\delta}_{1,j} < 0.9\delta_{1,j}$. We claim, however, that the probability that a uniform J (chosen by Q) will hit such a "bad" column j is low. For showing that, let $\mathcal{B}_1 = \{j \in [n] \colon \tilde{\delta}_{1,j} < 0.9\delta_{1,j}\}$ be the set of "bad" columns $j \in [n]$ for $i = 1$. A simple calculation yields that

$$\begin{aligned} d_1 &= D(\widetilde{P}_{X_1Y_1}||P_{X_1Y_1}) \geq D(\widetilde{P}_{Y_1}||P_{Y_1}) \geq \sum_{j=1}^{n} D(\widetilde{P}_{Y_{1,j}}||P_{Y_{1,j}}) \\ &= \sum_{j=1}^{n} D(\tilde{\delta}_{1,j}||\delta_{1,j}) \geq \sum_{j\in\mathcal{B}_1} D(\tilde{\delta}_{1,j}||\delta_{1,j}) \geq \sum_{j\in\mathcal{B}_1} \delta_{1,j}/200 \geq |\mathcal{B}_1| \cdot \delta/200. \end{aligned}$$

The second inequality holds by chain-rule of KL-divergence when the right-hand side distribution is product (Fact 4(3))) and the penultimate inequality holds by Fact 6(1). This implies that $|\mathcal{B}_1| \leq 200d_1/\delta$, and hence, $Q_J[J \in \mathcal{B}_1] < 200d_1/(\delta n)$. Extending the above argument for a row $i > 1$ is a much harder task. As we saw in Claim 6, the conditional distribution $Q_{J|X_{<i}}$ is much more complicated, and it also seems not clear how to bound $|\mathcal{B}_i|$ (now a function of $X_{<i}$) as we did for $i = 1$, when $X_{<i}$ is drawn from Q. Yet, we show in the full version that when $X_{<i}$ is drawn from $\widetilde{P}$ (and not from Q), then we are able to understand $Q_{J|X_{<i}}$ and $\mathcal{B}_i(X_{<i})$ better and bound by $O(d/(\delta n))$ the probability of hitting a "bad" column for all $i \in [m]$. This is done by relating martingale sequences for each sequence $\{\omega_{i,j}\}_{i=1}^{m}$ under $\widetilde{P}$, and by showing (using Lemma 3) that with high probability, the sequences of most $j \in [n]$ remains around 1.

Following the above discussion, the high level plan of our proof is to define the "good" events $A_1, \ldots, A_n$ for $\widetilde{P}$ and $B_1, \ldots, B_n$ for Q such that for all $i \in [m]$, the conditional distributions $\widetilde{P}_{X_i|A_{\leq i}}$ and $Q_{X_i|B_{\leq i}}$ satisfy the assumptions in 11. Then, by only bounding the probability of "bad" events under $\widetilde{P}$, the proof of Theorem 8 will follow by Lemma 4. For details, see the full version.

Bibliography

[1] Bellare, M., Impagliazzo, R., Naor, M.: Does parallel repetition lower the error in computationally sound protocols? In: 38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997. pp. 374–383 (1997)

[2] Berman, I., Haitner, I., Tsfadia, E.: A tight parallel repetition theorem for partially simulatable interactive arguments via smooth kl-divergence. Cryptology ePrint Archive, Report 2019/393 (2019), https://eprint.iacr.org/2019/393

[3] Brakerski, Z., Vaikuntanathan, V.: Efficient fully homomorphic encryption from (standard) lwe. Journal of the ACM 43(2), 831–871 (2014)

[4] Canetti, R., Halevi, S., Steiner, M.: Hardness amplification of weakly verifiable puzzles. In: Theory of Cryptography, Second Theory of Cryptography Conference, TCC 2005. pp. 17–33 (2005)

[5] Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. Annals of combinatorics (2002)

[6] Chung, K., Liu, F.: Parallel repetition theorems for interactive arguments. In: Theory of Cryptography, Sixth Theory of Cryptography Conference, TCC 2010. pp. 19–36 (2010)

[7] Chung, K.M., Pass, R.: The randomness complexity of parallel repetition. In: Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS). pp. 658–667 (2011)

[8] Chung, K., Pass, R.: Tight parallel repetition theorems for public-coin arguments using kl-divergence. In: Theory of Cryptography, 11th Theory of Cryptography Conference, TCC 2015. pp. 229–246 (2015)

[9] Damgärd, I.B., Pfitzmann, B.: Sequential iteration arguments and an efficientzero-knowledge argument for NP. In: Annual International Colloquium on Automata, Languages and Programming (ICALP). pp. 772–783 (1998)

[10] Dinur, I., Steurer, D.: Analytical approach to parallel repetition. In: Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014. pp. 624–633 (2014)

[11] Dodis, Y., Jain, A., Moran, T., Wichs, D.: Counterexamples to hardness amplification beyond negligible. In: Theory of Cryptography, 8th Theory of Cryptography Conference, TCC 2012. pp. 476–493 (2012)

[12] Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain markov process expectations for large time. iv. Communications on Pure and Applied Mathematics 36(2), 183–212 (1983)

[13] Feige, U.: On the success probability of the two provers in one-round proof systems. In: Proceedings of the Sixth Annual Structure in Complexity Theory Conference, Chicago, Illinois, USA, June 30 - July 3, 1991. pp. 116–123 (1991)

[14] Feige, U., Verbitsky, O.: Error reduction by parallel repetition - A negative result. Combinatorica 22(4), 461–478 (2002)
[15] Fortnow, L., Rompel, J., Sipser, M.: Errata for on the power of multi-prover interactive protocols. In: Proceedings: Fifth Annual Structure in Complexity Theory Conference, Universitat Politècnica de Catalunya, Barcelona, Spain, July 8-11, 1990. pp. 318–319 (1990)
[16] Goldreich, O.: Modern Cryptography, Probabilistic Proofs and Pseudorandomness. Springer (1999)
[17] Haitner, I.: A parallel repetition theorem for any interactive argument. SIAM J. Comput. 42(6), 2487–2501 (2013), https://doi.org/10.1137/100810630
[18] Håstad, J., Pass, R., Wikström, D., Pietrzak, K.: An efficient parallel repetition theorem. In: Theory of Cryptography, Sixth Theory of Cryptography Conference, TCC 2010. pp. 1–18 (2010)
[19] Holenstein, T.: Parallel repetition: Simplification and the no-signaling case. Theory of Computing 5(1), 141–172 (2009)
[20] Moshkovitz, D.: Parallel repetition from fortification. In: 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014. pp. 414–423 (2014)
[21] Mulzer, W.: Chernoff bounds (2018), https://page.mi.fu-berlin.de/mulzer/notes/misc/chernoff.pdf
[22] Pass, R., Venkitasubramaniam, M.: A parallel repetition theorem for constant-round arthur-merlin proofs. TOCT 4(4), 10:1–10:22 (2012)
[23] Pietrzak, K., Wikström, D.: Parallel repetition of computationally sound protocols revisited. Journal of Cryptology 25(1), 116–135 (2012)
[24] Rao, A.: Parallel repetition in projection games and a concentration bound. SIAM J. Comput. 40(6), 1871–1891 (2011)
[25] Raz, R.: A parallel repetition theorem. SIAM J. Comput. 27(3), 763–803 (1998)