Electronic Colloquium on Computational Complexity, Report No. 167 (2017)

More on bounded independence plus noise: Pseudorandom generators for read-once polynomials

Chin Ho Lee Emanuele Viola

November 2, 2017

Abstract

We construct pseudorandom generators with improved seed length for several classes of tests. First we consider the class of read-once polynomials over GF(2) in m variables. For error ε we obtain seed length $\tilde{O}(\log(m/\varepsilon))\log(1/\varepsilon)$, where $\tilde{O}$ hides lower-order terms. This is optimal up to the factor $\tilde{O}(\log(1/\varepsilon))$. The previous best seed length was polylogarithmic in m and $1/\varepsilon$.

Second we consider product tests $f\colon \{0,1\}^m \to \mathbb{C}_{\leq 1}$. These tests are the product of k functions $f_i\colon \{0,1\}^n \to \mathbb{C}_{\leq 1}$, where the inputs of the f_i are disjoint subsets of the m variables and $\mathbb{C}_{\leq 1}$ is the complex unit disk. Here we obtain seed length $n \cdot \operatorname{poly}\log(m/\varepsilon)$. This implies better generators for other classes of tests. If moreover the f_i have outputs independent of n and k (e.g., $\{-1,1\}$) then we obtain seed length $\tilde{O}(n + \log(k/\varepsilon))\log(1/\varepsilon)$. This is again optimal up to the factor $\tilde{O}(\log 1/\varepsilon)$, while the previous best seed length was $\geq \sqrt{k}$.

A main component of our proofs is showing that these classes of tests are fooled by almost d-wise independent distributions perturbed with noise.

ISSN 1433-8092

1 Introduction, results, and discussion

The construction of unconditional pseudorandom generators that fool restricted classes of tests is a fundamental research direction that has found disparate applications. In this work we obtain new generators for several classes. We start with the simplest.

Fooling read-once polynomials. Pseudorandom generators for *polynomials* have been studied since at least the 1993 work by Luby, Veličković, and Wigderson [LVW93], who gave a generator for GF(2) polynomials of size s with error ε and seed length $2^{O(\sqrt{s/\varepsilon})}$. See [Vio07] for an alternative proof. This seed length remains the best available after 25 years. For low-degree polynomials, better generators are known [BV10, Lov09, Vio09]. In this work we consider *read-once* polynomials, which are a sum of monomials on disjoint variables. For this class, a generator with seed length polylogarithmic in m and $1/\varepsilon$ is given in [GLS12] and it applies more generally to read-once ACC^0. We obtain a seed length which is optimal up to a factor of $\tilde{O}(\log 1/\varepsilon)$, where $\tilde{O}$ hides factors $\log\log(m/\varepsilon)$. In particular, when ε is not too small, our generator has seed length optimal up to $\operatorname{poly}\log\log m$.

Theorem 1. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ that fools any read-once GF(2) polynomial with error ε and seed length $\tilde{O}(\log(m/\varepsilon))\log(1/\varepsilon)$.*

The same result holds for polynomials modulo m for any fixed m; and in fact we obtain it as an easy corollary of a more general generator.

Fooling products. We consider tests on m bits that can be written as the product of k bounded functions on disjoint inputs of n bits. Such tests generalize the well-studied *combinatorial rectangles* [AKS87, Nis92, NZ96, INW94, EGL+98, ASWZ96, Lu02, Vio14, GMR+12, GY14] as well as other classes of tests, see [GKM15]. They were introduced in the latter paper by Gopalan, Kane, and Meka who call them *Fourier shapes*. However, in their definition the partition of the m-bit input into the k n-bit inputs to the functions is fixed and known to the generator. Following a recent push for breaking the mold of "fixed-order" tests, we consider such tests under arbitrary order. We call them *product tests* and define them formally next.

Definition 2 (Product tests)**.** *A function $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ is a* product test *with k functions of input length n if there exist k disjoint subsets $I_1, I_2, \ldots, I_k \subseteq \{1,2,\ldots,m\}$ of size $\le n$ such that $f(x) = \prod_{i\le k} f_i(x_{I_i})$ for some functions f_i with range in $\mathbb{C}_{\le 1}$. Here $\mathbb{C}_{\le 1}$ is the complex unit disk $\{z \in \mathbb{C} : |z| \le 1\}$, and x_{I_i} are the $|I_i|$ bits of x indexed by I_i.*

Handling arbitrary order is significantly more challenging because the classical space-bounded generators such as Nisan's [Nis92] only work in fixed order. Our previous work with Haramaty [HLV17] gave the first generators, but in it the dependency on k is poor: the seed length is always $\ge \sqrt{k}$. In this work we improve the dependency on k exponentially, though the results in [HLV17] are still unsurpassed when k is very small, e.g. $k = O(1)$. We actually obtain two incomparable generators.

Theorem 3. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ that fools any product test with k functions of input length n with error ε and seed length $\tilde{O}(n+\log k)\log(1/\varepsilon)\log(k/\varepsilon) = n\log^{O(1)}(m/\varepsilon)$.*

By the reductions in [GKM15], we also obtain generators that fool variants of product tests where the outputs of the f_i are not simply multiplied but combined in other ways. These variants include generalized halfspaces [GOWZ10] and combinatorial shapes [GMRZ13, De15], extended to arbitrary order. For those we obtain seed length $n^2\log^{O(1)}(m/\varepsilon)$, whereas the previous best was $\geq n\sqrt{k}$ [HLV17]. As this application amounts to plugging the above theorem in previous reductions, we don't discuss it further in this paper and instead refer the reader to Section 6 in [HLV17].

We then give another generator whose seed length is optimal up to a factor $\tilde{O}(\log 1/\varepsilon)$, just like Theorem 1. However, for this we need the functions f_i in the definition of product tests to take fixed values that are independent of n and k. This condition is satisfied by Boolean and most natural functions. For simplicity one can think of f_i having outputs $\{-1,1\}$.

Definition 4 (Nice product tests)**.** *A product test as in Definition 2 is* nice *if the functions f_i output values that are independent of n and k.*

Formally, one should talk about a nice *family* of product tests; but for simplicity we'll just say "nice product test."

Theorem 5. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ that fools any nice product test with k functions of input length n with error ε and seed length $\tilde{O}(n+\log(k/\varepsilon))\log(1/\varepsilon)$.*

This is the result from which the generator for polynomials in Theorem 1 follows easily.

Bounded independence plus noise. The framework in which we develop these generators was first laid out by Ajtai and Wigderson in their pioneering work [AW89] constructing generators for AC^0 with polynomial seed length. The framework seems to have been forgotten for a while, possibly due to the spectacular successes by Nisan who gave better and arguably simpler generators [Nis91, Nis92]. It has been recently revived in a series of papers starting with the impressive work by Gopalan, Meka, Reingold, Trevisan, and Vadhan [GMR$^+$12], who use it to obtain a generator for read-once CNF on m bits with error ε and seed length $\tilde{O}(\log(m/\varepsilon))$. This significantly improves on the previously available seed length of $O(\log m)\log(1/\varepsilon)$ when ε is small.

The Ajtai–Wigderson framework goes by showing that the test is fooled by a distribution with limited independence [NN93], *if we perturb it with noise.* (Previous papers use the equivalent language of *restrictions*, we instead follow [HLV17].) Then the high-level idea is to recurse on the noise. This has to be coupled with a separate, sometimes technical argument showing that each recursion simplifies the test, which we address later. [HLV17] show that this framework can be applied to product tests, but again their dependence on the number k of functions is poor. A main technical contribution of this work is obtaining

exponentially better dependency on k using different techniques from [HLV17]. We now state this result after defining almost bounded independence.

Definition 6 ((ε, d)-closeness)**.** *The random variables $X_1, \ldots, X_k$ are (ε, d)-close to $Y_1, \ldots, Y_k$ if for every $i_1, \ldots, i_d \in \{1, 2, \ldots, k\}$ the d-tuples $(X_{i_1}, \ldots, X_{i_d})$ and $(Y_{i_1}, \ldots, Y_{i_d})$ have statistical distance $\leq \varepsilon$.*

Theorem 7. *There is a universal constant C such that the following holds:*

Let $f\colon \{0,1\}^m \to \mathbb{C}_{\leq 1}$ be a product test with k functions of input length $\leq n$. Let D and T be two independent distributions over $\{0,1\}^m$ that are $(\delta, Cn\log(1/\varepsilon))$-close to uniform. Then

$$\left| \mathop{\mathbb{E}}_{D,T,U} \Big[f(D + T \wedge U) \Big] - \mathop{\mathbb{E}}_{U}[f(U)] \right| \leq \varepsilon,$$

where U is the uniform distribution, for the following choices of δ:

(1) $\delta = 2^{-C(\log k + n)\log(1/\varepsilon)}$;

(2) if f is nice then $\delta = 2^{-C(n + \log\log(1/\varepsilon))\log(1/\varepsilon)}$.

Above, $D + T \wedge U$ is the bit-wise Xor of D and $T \wedge U$, where $\wedge$ in turn is bit-wise And. We think of $T \wedge U$ as a noise vector: if a bit in T is 1 then we get a uniform bit. For the application it is important that T is selected pseudorandomly, though the result is interesting even if T is uniform in $\{0,1\}^m$.

An interesting feature of Theorem 7 is that for nice products the parameter δ can be independent of k. We complement this feature with a negative result showing that for general products a dependence on k is necessary. Thus, the distinction between products and nice products is not an artifact of our proof but is inherent.

Claim 8. *For every sufficiently large k, there exists a distribution D over $\{0,1\}^k$ that is $(k^{-\Omega(1)}, k^{\Omega(1)})$-close to uniform for every integer d, and a product test $f\colon \{0,1\}^k \to \mathbb{C}_{\leq 1}$ with k functions of input length 1 such that*

$$\Big| \mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)] \Big| \geq 1/10,$$

where T and U are the uniform distribution over $\{0,1\}^k$.

This claim also shows that for $n = 1$ and $1/\varepsilon = O(1)$ one needs $\delta \leq k^{-\Omega(1)}$, and even for distributions which are $(\delta, k^{\Omega(1)})$-close to uniform, instead of just $(\delta, O(1))$-close.

1.1 Techniques

We start by explaining how we obtain Theorem 7. Following the literature, at a high level we do a case analysis based on the *total variance* of the product test f we want to fool. This variance is defined as the sum of the variances $\mathrm{Var}(f_i)$ of the functions f_i in the definition of product test. The variance of a function g is $\mathbb{E}[|g(x)|^2] - |\mathbb{E}[g(x)]|^2$ where x is uniform. For the application to polynomials, the functions f_i are simply And; however, we do not know how to simplify the proofs in this case.

The high variance case. This case actually does not appear in the read-once CNF generator in [GMR+12]. This is because one can always *truncate* the CNF to have at most $2^w \log(1/\varepsilon)$ number of clauses of width w (this number of clauses suffices to determine the expected value of the CNF up to an additive error of ε) and such a CNF has low total variance (for this one argues that noise helps in reducing the variance a little.) To handle an arbitrary read-once CNF, [GMR+12] partition the clauses according to their width, and handle each width separately.

But one cannot truncate polynomials. To see why, consider for a simple example the linear polynomial $x_1 + x_2 + \ldots + x_m$ (corresponding to f_i that just compute parity). Here no strict subset of the monomials determines the value of the polynomial. Indeed, one can construct distributions which look random to $m-1$ monomials, but not to m.

The papers [GLS12, GKM15] essentially reduce the high variance case to the low variance. However their techniques either blow up the seed length polynomially [GLS12] or rely on space-bounded generators that only work in fixed order [GKM15].

We instead show that bounded independence plus noise fools even high-variance product tests. We now give some details of our approach. A standard fact is that the expectation of a function f_i can be bounded above by $(1-\mathrm{Var}(f_i))^{1/2} \leq e^{-\mathrm{Var}(f_i)/2}$, and so if the total variance $\sum_i \mathrm{Var}(f_i)$ is large then the expectation under the uniform distribution is small. Thus, it suffices to show that the expectation is also small under bounded independence plus noise. To show this, we argue that typically, the total variance remains high even considering the f_i as functions of the noise only. Specifically, we first show that on average over a uniform x and t, the variance of the functions $f_i'(y) := f_i(x + t \wedge y)$ is about as large as that of the f_i. This uses Fourier analysis. Then we use concentration inequalities for distributions with limited independence to derandomize this fact: we show that it also holds for typical x and t sampled from D and T.

The low variance case. Here our starting point is a compelling inequality in [GKM15] (cf. [GMR+12, GY14]) showing that bounded independence (even without noise) fools low-variance product tests. However, their result is only proved for *exact* bounded independence (any d bits are exactly uniform), whereas it is critical for our seed lengths to handle *almost* bounded independence (any d bits are close to uniform).

In this paper we extend the inequality in [GKM15] to work for almost bounded independence. The proof of the inequality in [GKM15] is somewhat technical, and our extension introduces several complications. For example, the expectations of the f_i under the pseudorandom distribution D and the uniform distribution U are not guaranteed to be equal, and this requires additional arguments. However our proof follows the argument in [GKM15], which we also present in a slightly different way, possibly of interest to some readers. Finally we mention that Claim 8 shows that our error term is close to tight in certain regimes, cf. Section A.

More on truncation. As mentioned earlier, [GMR+12] truncate read-once CNF, but we cannot do that for polynomials. However, we show that something almost as good can still be done, and this idea is critical to obtain our seed lengths. We show that the statistical

closeness parameter in D and T can be selected *as if the polynomial was truncated*: it is independent from the number k of functions. This is reflected in Item (2) in Theorem 7, where δ is independent from k. The proof goes by showing that if the number k of functions is much larger than 2^n then noise will be enough to fool the test, regardless of anything else. This proof critically uses noise: without noise a dependence on k is necessary. Also, for the proof to work the functions must have expectation at most $1 - 2^{-n}$. As mentioned earlier, we further prove that this last requirement is necessary (Claim 8): we construct functions whose expectation is about $1 - 1/k$ but whose product is not fooled by almost bounded independence plus noise, if the statistical closeness parameter is larger than $1/k^c$ for a suitable constant c.

Getting the pseudorandom generator. To construct the pseudorandom generator given in Theorem 5, we recurse on the noise following the Ajtai–Wigderson framework. As hinted earlier, we must also argue that the test gets simplified at each iteration. Our measure of simplification is the input length of the f_i, which corresponds to the *degree* of the polynomial. At stage i of the recursion, we have input length $n/2^i$. We show that this input length is halved using $O(2^i)$ iterations of Theorem 7. Each iteration costs $\tilde{O}(n/2^i \cdot \log(1/\varepsilon))$ in seed, for a total seed length of $\tilde{O}(n)\log(1/\varepsilon)$ for each stage. We perform $O(\log n)$ stages until the test is a constant. For read-once polynomials one should think of $n = \log m$, in which case the number of stages is $\log\log m$.

We remark that [GMR+12] instead use the *number of clauses* in the CNF as progress measure. They show that this number drops polynomially at each iteration. However this only holds for truncated CNF, and is false for example for the linear polynomial $x_1 + x_2 + \ldots + x_m$.

It is an interesting question whether in (2) in Theorem 7 one can set δ as large as $2^{-O(n+\log(1/\varepsilon))}$ instead of about $2^{-n\log(1/\varepsilon)}$. However we do not know how to get better seed length even if this improvement could be obtained, because we pay $\Omega(\log m)\log(1/\varepsilon)$ in seed length even if $n = 2$.

2 Bounded independence plus noise fools products

In this section we prove Theorem 7, except for the next lemma handling the low-variance case, which is proved in Section B.

Lemma 9. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\le 1}$ with $\min_{z\in\mathrm{Supp}(X_i)} \Pr[X_i = z] \ge 2^{-n}$ for each $i \in \{1,\ldots,k\}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon, 16d)$-close to $X_1, \ldots, X_k$. Then*

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} Y_i\right] - \mathbb{E}\left[\prod_{i=1}^{k} X_i\right]\right| \le 2^{O(d)}\left(\frac{\sum_{i=1}^{k}\mathrm{Var}[X_i]}{d}\right)^{d/2} + (k2^n)^{O(d)}\varepsilon.$$

We start with a claim that shows that for typical x and t, the variance of the function $g(y) := f(x + t \wedge y)$ is close to the variance of f.

Claim 10. *Let T be the distribution over $\{0,1\}^n$ where the T_j's are independent and $\mathbb{E}[T_j] = \eta$ for each j. Let $f\colon \{0,1\}^n \to \mathbb{C}$ be any function. Then*

$$\mathop{\mathbb{E}}_{U,T}[\mathop{\mathrm{Var}}_{U'}[f(U + T \wedge U')]] \geq \eta \,\mathrm{Var}[f].$$

Proof of Claim 10. By the definition of variance and linearity of expectation, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathrm{Var}}_{U'}[f(U + T \wedge U')]\Big] &= \mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}[|f(U + T \wedge U')|^2] - \Big|\mathop{\mathbb{E}}_{U'}[f(U + T \wedge U')]\Big|^2\Big] \\ &= \mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}[|f(U + T \wedge U')|^2]\Big] - \mathop{\mathbb{E}}_{U,T}\Big[\Big|\mathop{\mathbb{E}}_{U'}[f(U + T \wedge U')]\Big|^2\Big].\end{aligned}$$

The first term is equal to

$$\mathop{\mathbb{E}}_{U}[|f(U)|^2] = \sum_{\alpha,\alpha'} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \mathop{\mathbb{E}}_{U}[\chi_{\alpha-\alpha'}(U)] = \sum_{\alpha} |\hat{f}_\alpha|^2.$$

The second term is equal to

$$\begin{aligned}&\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\Big[\sum_{\alpha} \hat{f}_\alpha \chi_\alpha(U + T \wedge U')\Big] \overline{\mathop{\mathbb{E}}_{U''}\Big[\sum_{\alpha'} \hat{f}_{\alpha'} \chi_{\alpha'}(U + T \wedge U'')\Big]}\Big] \\ &= \mathop{\mathbb{E}}_{U,T}\Big[\sum_{\alpha,\alpha'} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \mathop{\mathbb{E}}_{U'}[\chi_\alpha(U + T \wedge U')] \mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(U + T \wedge U'')]\Big] \\ &= \sum_{\alpha,\alpha'} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \mathop{\mathbb{E}}_{U}[\chi_{\alpha+\alpha'}(U)] \mathop{\mathbb{E}}_{T}\Big[\mathop{\mathbb{E}}_{U'}[\chi_\alpha(T \wedge U')] \mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(T \wedge U'')]\Big] \\ &= \sum_{\alpha} |\hat{f}_\alpha|^2 \mathop{\mathbb{E}}_{T,U',U''}[\chi_\alpha(T \wedge (U' + U''))] \\ &= \sum_{\alpha} |\hat{f}_\alpha|^2 (1-\eta)^{|\alpha|}.\end{aligned}$$

Therefore,

$$\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathrm{Var}}_{U'}[f(U + T \wedge U')]\Big] = \sum_{\alpha} |\hat{f}_\alpha|^2 \Big(1 - (1-\eta)^{|\alpha|}\Big) \geq \eta \sum_{\alpha \neq 0} |\hat{f}_\alpha|^2 = \eta \,\mathrm{Var}[f],$$

where the inequality is because $1 - (1-\eta)^{|\alpha|} \geq 1 - (1-\eta) \geq \eta$ for any $\alpha \neq 0$. □

With the above two results, we can prove Item (1) of Theorem 7.

Proof of Item (1) in Theorem 7. Let σ denote $(\sum_{i \leq k} \mathrm{Var}[f_i])^{1/2}$. We will consider two cases: $\sigma^2 \leq C' \log(1/\varepsilon)$ and $\sigma^2 > C' \log(1/\varepsilon)$, where C' is a sufficiently large constant depending on C.

If $\sigma^2 \leq C' \log(1/\varepsilon)$, we use Lemma 9. Specifically, since $\Pr[f_i(U) = z] \geq 2^{-n}$ for any $z \in \mathrm{Supp}(f_i)$, it follows from Lemma 9 with $d = O(\log(1/\varepsilon))$ that

$$\left|\mathbb{E}\Big[\prod_{i=1}^{k} f_i(D)\Big] - \mathbb{E}\Big[\prod_{i=1}^{k} f_i(U)\Big]\right| \leq 2^{-\Omega(\log(1/\varepsilon))} + (k2^n)^{O(\log(1/\varepsilon)) - C\log(1/\varepsilon)} \leq \varepsilon,$$

and the desired bound holds for all fixing of T and U.

If $\sigma^2 \geq C' \log(1/\varepsilon)$, then we have

$$\Big|\prod_{i\leq k} \mathop{\mathbb{E}}_{U}[f_i(U)]\Big| = \prod_{i\leq k}(1-\mathrm{Var}[f_i])^{1/2} \leq e^{-\frac{1}{2}\sigma^2} \leq \varepsilon/2.$$

Thus, it suffices to show that

$$\Big|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i=1}^{k} f_i(D+T\wedge U)\Big]\Big| \leq \varepsilon/2.$$

For each $t, x \in \{0,1\}^m$, and each $i \in \{1,2,\ldots,k\}$, let $\sigma^2_{t,x,i}$ denote $\mathrm{Var}_{U'}[f_i(x+t\wedge U')]$. Let T' be the uniform distribution over $\{0,1\}^m$. By Claim 10 with $\eta = 1/2$, we have $\mathbb{E}_{T',U}[\sigma^2_{T',U,i}] \geq \mathrm{Var}[f_i]/2$. So by linearity of expectation we have

$$\mathop{\mathbb{E}}_{T',U}\Big[\sum_{i\leq k}\sigma^2_{T',U,i}\Big] \geq \sigma^2/2 \geq C'\log(1/\varepsilon)/2.$$

Since T and D are both $(\delta, Cn\log(1/\varepsilon))$-close to uniform, the random variables $\sigma^2_{T,D,1},\ldots,\sigma^2_{T,D,k}$ are $(2\delta, C\log(1/\varepsilon))$-close to $\sigma^2_{T',U,1},\ldots,\sigma^2_{T',U,k}$. Let $\mu = \mathbb{E}_{T',U}[\sum_{i\leq k}\sigma^2_{T',U,i}] \geq C'\log(1/\varepsilon)/2$. By Lemma 36,

$$\Pr_{T',U}\left[\sum_{i\leq k}\sigma^2_{T',U,i} \leq \mu/2\right] \leq 2^{-\Omega(\log(1/\varepsilon))} + k^{O(\log(1/\varepsilon))}\delta \leq \varepsilon/4.$$

Hence, except with probability $\varepsilon/4$ over $t \in T$ and $x \in D$, we have

$$\sum_{i\leq k}\sigma^2_{t,x,i} = \sum_{i\leq k}\mathop{\mathrm{Var}}_{U'}[f_i(x+t\wedge U')] \geq C'\log(1/\varepsilon)/4.$$

Therefore, for every such t and x, we have

$$\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\Big| \leq \prod_{i\leq k}\Big|\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\Big| = \prod_{i\leq k}(1-\sigma^2_{t,x,i})^{1/2} \leq e^{-\frac{1}{2}\sum_{i\leq k}\sigma^2_{t,x,i}} \leq \varepsilon/4.$$

In addition, we always have $|f| \leq 1$. Hence,

$$\left|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i\leq k} f_i(D+T\wedge U)\Big]\right| \leq \mathop{\mathbb{E}}_{D,T}\left[\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(D+T\wedge U)]\Big|\right] \leq \varepsilon/2. \qquad \square$$

To prove Item (2) in Theorem 7 we need the following claim showing that noise alone fools nice products when k is suitably larger than 2^n.

Claim 11 (Noise fools nice products with large k)**.** *Let $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ be a product test with $k \ge 2 \cdot 2^{(c+1)n}\log(4/\varepsilon)$ functions $f_1,\dots,f_k$ of input length $\le n$ and $|\mathbb{E}[f_i]| \le 1 - 2^{-cn}$ for every $i \in \{1,2,\dots,k\}$ for some constant c. Let C be a sufficiently large constant. Let T be a distribution over $\{0,1\}^m$ that is $(\delta, Cn\log(1/\varepsilon))$-close to uniform. Then*

$$\Big|\mathop{\mathbb{E}}_{T,U}[f(T\wedge U)] - \mathbb{E}[f(U)]\Big| \le \varepsilon/2 + 2^{O(n+\log\log(1/\varepsilon))\log(1/\varepsilon)}\delta.$$

Proof. We will bound above both expectations in absolute value. Let $k' = 2\cdot 2^{(c+1)n}\log(4/\varepsilon)$. Write $f = \prod_{i=1}^k f_i$, where $f_i\colon \{0,1\}^{I_i} \to \mathbb{C}_{\le 1}$. We have

$$|\mathbb{E}[f(U)]| = \prod_{i=1}^k |\mathbb{E}[f_i(U)]| \le (1-2^{-cn})^k \le e^{-k2^{-cn}} \le \varepsilon/8. \tag{1}$$

We also have

$$|\mathbb{E}[f(T\wedge U)]| \le \mathop{\mathbb{E}}_T\Big[\prod_{i=1}^k \Big|\mathop{\mathbb{E}}_U[f_i(T\wedge U)]\Big|\Big] \le \mathop{\mathbb{E}}_T\Big[\prod_{i=1}^{k'} \Big|\mathop{\mathbb{E}}_U[f_i(T\wedge U)]\Big|\Big].$$

We will show that $\mathbb{E}_T[\prod_{i=1}^{k'}|\mathbb{E}_U[f_i(T\wedge U)]|]$ is at most $\varepsilon/2 + 2^{O(n+\log\log(1/\varepsilon))\log(1/\varepsilon)}\delta$. Let T' be the uniform distribution over $\{0,1\}^m$. Then

$$\mathbb{E}\Big[\sum_{i=1}^{k'} \mathbb{1}(T'_{I_i} = 1^{|I_i|})\Big] = \sum_{i=1}^{k'} \Pr[T'_{I_i} = 1^{|I_i|}] \ge k'2^{-n} \ge 2\cdot 2^{cn}\log(4/\varepsilon).$$

Since T is $(\delta, Cn\log(1/\varepsilon))$-close to uniform, the T_{I_i} are $(\delta, C\log(1/\varepsilon))$-close to uniform. By Lemma 36,

$$\begin{aligned}\Pr_T\Big[\sum_{i=1}^{k'} \mathbb{1}(T_{I_i} = 1^{|I_i|}) \le 2^{cn}\log(4/\varepsilon)\Big] &\le \varepsilon/8 + k'^{O(\log(1/\varepsilon))\delta} \\ &\le \varepsilon/8 + 2^{O(n+\log\log(1/\varepsilon))\log(1/\varepsilon)}\delta. \end{aligned}\tag{2}$$

Note that if $T_{I_i} = 1^{|I_i|}$, then $|\mathbb{E}_U[f_i(T\wedge U)]| = |\mathbb{E}[f]| \le 1 - 2^{-cn}$, and we always have $|f| \le 1$. Thus, conditioned on $\sum_{i=1}^{k'} \mathbb{1}(T_{I_i} = 1^{|I_i|}) \ge 2^{cn}\log(4/\varepsilon)$, we have

$$\prod_{i=1}^{k'} |\mathbb{E}[f_i(T\wedge U)]| \le (1-2^{-cn})^{2^{cn}\log(4/\varepsilon)} \le \varepsilon/4. \tag{3}$$

The error bound follows from summing the R.H.S. of (1), (2) and (3). □

Now we can prove Item (2) in Theorem 7.

Proof of Item (2) in Theorem 7. Suppose $|\mathbb{E}[f_i]| \le 1 - 2^{-cn}$ for some constant c depending on C. If $k \ge 2\cdot 2^{(c+1)n}\log(4/\varepsilon)$, then the theorem follows by Claim 11. Otherwise, $k \le 2\cdot 2^{(c+1)n}\log(4/\varepsilon) \le 2\cdot 2^{(c+1)n+\log\log(4/\varepsilon)}$ and we can apply Item (1) with $\delta = 2^{-C(n+\log\log(1/\varepsilon))\log(1/\varepsilon)} \le (k2^n)^{-C'\log(1/\varepsilon)}$ to get an error bound of ε. □

3 Pseudorandom generators

In this section we construct our generators. First we state a lemma that will be useful in our constructions, then we prove our generator theorems, and finally we prove the lemma.

Lemma 12. *There is a universal constant C such that the following holds:*

If there is an explicit generator $G'\colon \{0,1\}^{\ell'} \to \{0,1\}^m$ that fools product tests with k functions of input length r with error ε' and seed length ℓ', then there is an explicit generator $G\colon \{0,1\}^{\ell} \to \{0,1\}^m$ that fools product tests with k functions of input length n with error $\varepsilon'+t\varepsilon$ and seed length $\ell = \ell'+t\cdot(n\log(1/\varepsilon)+\log(1/\delta)+\log\log m)$, where $t = O(\log(k/\varepsilon)/(r+1)+\log(n/(r+1)))$, for the following choices of δ:

(1) $\delta = 2^{-C(\log k+n)\log(1/\varepsilon)}$;

(2) if the product tests are nice then $\delta = 2^{-C(n+\log\log(1/\varepsilon))\log(1/\varepsilon)}$.

Proof of Theorem 3. We will apply Lemma 12 with $r = 0$ with error $\varepsilon/O(\log(k/\varepsilon)+\log n)$. Note that a product test of input length 0 is a constant function, which can always be fooled with zero error. So we have a generator that fools product tests with k functions of input length n, with error ε and seed length $t\cdot O\left(n\log\left((\log(k/\varepsilon)+\log n)/\varepsilon\right)+\log(1/\delta)+\log\log m\right) = \tilde{O}(\log(k/\varepsilon))(n+\log k)\log(1/\varepsilon)$. □

Proof of Theorem 5. Let f be a nice product test with k functions of input length n. Note that by applying Lemma 12 with $r = n/2$ and error $\varepsilon/O(\log n\cdot(\log(k/\varepsilon)/n+1))$, we can halve its input length by incurring an error of $\varepsilon/O(\log n)$ and using a seed of length $t\cdot O\left(n\log\left((\log n\cdot(\log(k/\varepsilon)/n+1))/\varepsilon\right)+\log(1/\delta)+\log\log m\right) = \tilde{O}(\log(k/\varepsilon)+n)\log(1/\varepsilon)$. Now we repeat the argument for $s = O(\log n)$ steps until the input length is zero, which is a constant function and can be fooled with zero error. So we have a generator that fools nice product tests with k functions of input length n, with error ε and seed length $s\cdot\tilde{O}(\log(k/\varepsilon)+n)\log(1/\varepsilon) = \tilde{O}(\log(k/\varepsilon)+n)\log(1/\varepsilon)$. □

Proof of Theorem 1. Let c be a sufficiently large constant. Let D be a $(\varepsilon/m)^c$-biased distribution over $\{0,1\}^m$ [NN93]. Let G be the output distribution of the generator in Theorem 5 that fools product tests with m functions and input length $c\log(m/\varepsilon)$ with $\varepsilon/2$. The generator outputs $D+G$ and by [NN93] and Theorem 5 takes a seed of length $O(\log(m/\varepsilon))+\tilde{O}(\log(m/\varepsilon)+c\log(m/\varepsilon))\log(1/\varepsilon) = \tilde{O}(\log(m/\varepsilon))\log(1/\varepsilon)$.

Let $p\colon \{0,1\}^m \to \{-1,1\}$ be any read-once GF(2) polynomial. Consider the polynomial p' obtained from p by removing all the monomials with degree greater than $c\log(m/\varepsilon)$ in p. Note that for any $(\varepsilon/m)^c$-biased distribution X, the probability that any $c\log(m/\varepsilon)$ bits are 1 is at most $\varepsilon/4m$, and so we have $\Pr[p(X)\neq p'(X)] \le \varepsilon/4$. It follows that $|\mathbb{E}[p(D+G)]-\mathbb{E}[p(U)]| \le |\mathbb{E}[p'(D+G)]-\mathbb{E}[p'(U)]|+\varepsilon/2 \le \varepsilon$, where the last inequality holds for any fixed D because of Theorem 5. □

We now prove Lemma 12. First we state a claim that will be used in the proof.

Claim 13. *Let $T^{(1)},\ldots,T^{(t)}$ be t independent and identical distributions over $\{0,1\}^n$ that are ε-close to uniform. Then* $\Pr[wt(\wedge_{i=1}^t T^{(i)}) > r] \le \binom{n}{r+1}(2^{-(r+1)}+\varepsilon)^t$.

Proof. Since $T^{(1)}, \ldots, T^{(t)}$ are independent, and each $T^{(i)}$ is ε-close to uniform,

$$\Pr[wt(\wedge_{i=1}^{t} T^{(i)}) > r] \leq \sum_{S:|S|=r+1} \Pr\left[\wedge_{i=1}^{t} \wedge_{j\in S}(T_j^{(i)} = 1)\right] = \sum_{S:|S|=r+1} \prod_{i=1}^{t} \Pr[\wedge_{j\in S}(T_j^{(i)} = 1)]$$
$$\leq \sum_{S:|S|=r+1} (2^{-(r+1)} + \varepsilon)^t = \binom{n}{r+1}(2^{-(r+1)} + \varepsilon)^t. \quad \square$$

Proof of Lemma 12. For $S \subseteq \{1, 2, \ldots, m\}$, define the function $\mathrm{PAD}_S(x)\colon \{0,1\}^{|S|} \to \{0,1\}^m$ outputs m bits of which the positions in S are the first $|S|$ bits of $x0^{|S|}$ and the rest are 0.

Define the distribution $H^{(i)}$ recursively for $t = O(\log(k/\varepsilon)/(r+1) + \log(n/(r+1)))$ steps: At the i-th step, $H^{(i)}$ samples two independent distributions $D^{(i)}, T^{(i)}$ over $\{0,1\}^m$ that are $(\delta, Cn\log(1/\varepsilon))$-close to uniform. Then output $D^{(i)} + T^{(i)} \wedge \mathrm{PAD}_{T^{(i)}}(H^{(i+1)})$. We define $H^{(t+1)}$ to be $G(U_\ell)$. The generator G will output $H^{(1)}$.

By [NN93, Lemma 4.2], sampling $D^{(i)}$ and $T^{(i)}$ takes a seed of length $O(n\log(1/\varepsilon) + \log(1/\delta) + \log\log m)$. The total seed length of G is therefore $\ell = \ell' + t \cdot O(n\log(1/\varepsilon) + \log(1/\delta) + \log\log m)$.

We now analyze the error of G. For $i \in \{1, 2, \ldots, t\}$, consider a variant of $H^{(1)}$ which is the same as $H^{(1)}$ but at the i-th step replace $\mathrm{PAD}_{T^{(i)}}(H^{(i+1)})$ with $\mathrm{PAD}_{T^{(i)}}(U_m)$. Call this $H_U^{(i)}$. Let $H_U^{(0)} = U_m$.

For every $i \in \{1, \ldots, t\}$, for every fixed $D^{(1)}, \ldots, D^{(i-1)}$ and $T^{(1)}, \ldots, T^{(i-1)}$, the function f restricted to $\wedge_{j<i} T^{(j)}$ remains a product test with k functions of input length n, and remains nice if f is nice. Call the restricted function g. Then, by Theorem 7, we have

$$|\mathbb{E}[f(H_U^{(i-1)})] - \mathbb{E}[f(H_U^{(i)})]| = |\mathbb{E}[g(U)] - \mathbb{E}[g(D^{(i)} + T^{(i)} \wedge U_m)]| \leq \varepsilon.$$

Hence, $|\mathbb{E}[f(U_m)] - \mathbb{E}[f(H_U^{(t)}]| \leq t\varepsilon$.

We now show that $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \leq \varepsilon' + 2\varepsilon$. Write $f = \prod_{i\leq k} f_i$, where each f_i is defined on $\{0,1\}^{I_i}$ with $|I_i| \leq n$. We claim that

$$\Pr\left[wt(\wedge_{i=1}^{t} T_{I_j}^{(i)}) > r \text{ for some } j \in \{1, \ldots, k\}\right] \leq \varepsilon.$$

It suffices to analyze $\Pr[wt(\wedge_{i=1}^{t} T_{I_j}^{(i)}) > r]$ for each j and take a union bound over $j \leq k$.

Since $|I_j| \leq n$, $T_{I_j}^{(i)}$ is 2^{-Cn}-close to uniform. By Claim 13 and a union bound over $j \leq k$, the probability that some f_i has input length $> r$ is at most

$$k\binom{n}{r+1}(2^{-(r+1)} + 2^{-Cn})^t \leq k \cdot \left(\frac{ne}{r+1}\right)^{r+1} \cdot O(2^{-r})^{\Omega(\log(k/\varepsilon)/(r+1)+\log(n/(r+1)))} \leq \varepsilon.$$

Hence, for every $D^{(1)}, \ldots, D^{(t)}$, with probability $1 - \varepsilon$ over the choice of $T^{(1)}, \ldots, T^{(t)}$, the function f restricted to $\wedge_{i=1}^{t} T^{(i)}$ becomes a product with k functions of input length r, and remains nice if f is nice. Conditioned on this, we have by the definition of G' that $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \leq \varepsilon'$. Otherwise, as $|f|$ is bounded by 1, the absolute difference is always at most 2. Hence, $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \leq \varepsilon' + 2\varepsilon$, and so the total error is at most $\varepsilon' + (t+2)\varepsilon$. $\square$

Acknowledgments. We thank Daniel Kane for answering some questions about [GKM15].

References

[AKS87] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *19th ACM Symp. on the Theory of Computing (STOC)*, pages 132–140, 1987.

[ASWZ96] Roy Armoni, Michael E. Saks, Avi Wigderson, and Shiyu Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *37th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 412–421, 1996.

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant-depth circuits. *Advances in Computing Research - Randomness and Computation*, 5:199–223, 1989.

[BV10] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. on Computing*, 39(6):2464–2486, 2010.

[De15] Anindya De. Beyond the central limit theorem: Asymptotic expansions and pseudorandomness for combinatorial sums. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 883–902, 2015.

[EGL+98] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Efficient approximation of product distributions. *Random Struct. Algorithms*, 13(1):1–16, 1998.

[GKM15] Parikshit Gopalan, Daniek Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 903–922, 2015.

[GLS12] Dmitry Gavinsky, Shachar Lovett, and Srikanth Srinivasan. Pseudorandom generators for read-once acc^0. In *IEEE Conf. on Computational Complexity (CCC)*, pages 287–297, 2012.

[GMR+12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2012.

[GMRZ13] Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM J. on Computing*, 42(3):1051–1076, 2013.

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *25th IEEE Conf. on Computational Complexity (CCC)*, pages 223–234. IEEE, 2010.

[GY14] Parikshit Gopalan and Amir Yehudayoff. Inequalities and tail bounds for elementary symmetric polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:19, 2014.

[HLV17] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. Bounded independence plus noise fools products. In *Conf. on Computational Complexity (CCC)*, 2017.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *26th ACM Symp. on the Theory of Computing (STOC)*, pages 356–364, 1994.

[Lov09] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.

[Lu02] Chi-Jen Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–433, 2002.

[LVW93] Michael Luby, Boban Veličković, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica. An Journal on Combinatorics and the Theory of Computing*, 11(1):63–70, 1991.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. of Computer and System Sciences*, 52(1):43–52, February 1996.

[Ste04] J. Michael Steele. *The Cauchy-Schwarz master class*. MAA Problem Books Series. Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004.

[Vio07] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. on Computing*, 36(5):1387–1403, 2007.

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.

[Vio14] Emanuele Viola. Randomness buys depth for approximate counting. *Computational Complexity*, 23(3):479–508, 2014.

A Proof of Claim 8

We more generally exhibit a distribution D that is $(d^2/10k, d)$-close to uniform. One can obtain Claim 8 by setting $d = k^{1/3}$. To simplify notation we will switch from $\{0,1\}$ to $\{-1,1\}$, and replace k with $2k$.

We define D to be the uniform distribution over strings in $\{-1,1\}^{2k}$ with equal number of -1's and 1's.

Claim 14. *D is $(10d^2/k, d)$-close to uniform for every integer d.*

Proof. We can assume $d^2 \leq k/10$, for otherwise the conclusion is trivial. Let $I \subseteq [k]$ be a subset of size d. For every $x \in \{-1,1\}^d$, we have

$$\Pr[D_I = x] = \frac{\binom{2k-d}{k-wt(x)}}{\binom{2k}{k}},$$

where $wt(x)$ is the number of -1's in x. We bound below the R.H.S. by

$$\begin{aligned}
\frac{\binom{2k-d}{k-d}}{\binom{2k}{k}} &= \frac{k(k-1)\cdots(k-d+1)}{2k(2k-1)\cdots(2k-d+1)} \\
&\geq \left(\frac{k-d+1}{2k}\right)^d \\
&= 2^{-d}\left(1-\frac{d-1}{k}\right)^d \\
&\geq 2^{-d}\left(1-\frac{d(d-1)}{k}\right) \\
&\geq 2^{-d}\cdot(1-d^2/k),
\end{aligned}$$

and bound it above by

$$\begin{aligned}
\frac{\binom{2k-d}{k-d/2}}{\binom{2k}{k}} &= \frac{(k(k-1)\cdots(k-d/2+1))^2}{2k(2k-1)\cdots(2k-d+1)} \\
&\leq \left(\frac{k}{2k-d+1}\right)^d \\
&= 2^{-d}\left(1+\frac{d-1}{2k-d+1}\right)^d \\
&\leq 2^{-d}\left(1+\sum_{i=1}^{d}\left(\frac{d(d-1)}{2k-d+1}\right)^i\right) \\
&\leq 2^{-d}\left(1+2\cdot\frac{d(d-1)}{2k-d+1}\right) \\
&\leq 2^{-d}\cdot(1+2d^2/k).
\end{aligned}$$

The third inequality is because the geometric sum has ratio $\leq 1/2$ as $d^2 \leq k/10$, and so is bounded by twice the first term. Hence, we have $|\Pr[D_I = x] - 2^{-d}| \leq 2^{-d}\cdot 2d^2/k$ for every $x \in \{-1,1\}^d$. The claim then follows from summing the inequality over every $x \in \{-1,1\}^d$. □

We now define our product test f. For each $j \in \{1, \ldots, 2k\}$, define $f_j \colon \{-1,1\}^{2k} \to \mathbb{C}_{\leq 1}$ to be $f_j(x) = \omega^{x_j}$, where $\omega := e^{-i/\sqrt{2k}}$. Let $f = \prod_{j \leq 2k} f_j$. We now show that for every large enough k we have

$$\Big| \mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)] \Big| \geq 1/10.$$

We now bound above and below the expectation of f under both distributions. We will use the fact that $1 - \theta^2/2 \leq \cos\theta \leq 1 - 2\theta^2/5$ for $\theta \in [-1, 1]$. First, we have

$$\mathbb{E}[f(U)] = \prod_{j \leq 2k} \mathop{\mathbb{E}}_{x \sim \{-1,1\}}[\omega^x] = \prod_{j \leq 2k} (\omega + \omega^{-1})/2 = \Big(\cos(1/\sqrt{2k}) \Big)^{2k} \leq (1 - 1/5k)^{2k}.$$

Next for every $j \in \{1, 2, \ldots, 2k\}$, we have

$$\mathop{\mathbb{E}}_{T,U}[f_j(x + T \wedge U)] = \frac{3}{4}\omega^{x_j} + \frac{1}{4}\omega^{-x_j}.$$

Define $\beta \colon \{-1,1\} \to \mathbb{C}_{\leq 1}$ to be $\beta(x) := \frac{3}{4}\omega^x + \frac{1}{4}\omega^{-x}$. Since D has the same number of -1's and 1's,

$$\begin{aligned}
\mathop{\mathbb{E}}_{D}\Big[\prod_{j \leq 2k} \beta_j(D) \Big] &= \beta(1)^k \beta(-1)^k \\
&= (10/16 + 3/16 \cdot (\omega^2 + \omega^{-2}))^k \\
&= (5/8 + 3/8 \cdot \cos(2/\sqrt{2k}))^k \\
&\geq (5/8 + 3/8 \cdot (1 - 1/k))^k \\
&= (1 - 3/8k)^k,
\end{aligned}$$

Therefore $|\mathbb{E}[f(D+T\wedge U] - \mathbb{E}[f(U)]| \geq (1-3/8k)^k - (1-1/5k)^{2k} \geq 1/10$, for every sufficiently large k, concluding the proof.

The f_i in this proof have variance $\Theta(1/k)$. So this counterexample gives a product test with total variance $O(1)$, and is relevant also to Lemma 9. Specifically it shows that for $n = 1$ and say $d = O(1)$, the error term $(k2^n)^{O(d)}\varepsilon$ in Lemma 9 cannot be replaced with $k^c\varepsilon$ for a certain constant c. Moreover, it cannot be replaced even if any $k^{\Omega(1)}$ of the Y_i are close to the X_i (as opposed to just $O(1)$).

B On almost k-wise independent variables with small total variance

In this section we will prove Lemma 9.

Our proof follows closely to the one in [GKM15], which proves the lemma for $\varepsilon = 0$, that is, when the X_i's are d-wise independent. We first give an overview of their proof.

For independent random variables $Z_1, \ldots, Z_k$, we will use $\sigma(Z)$ to denote the standard deviation of $\sum_{i \leq k} Z_i$, that is, $\sigma(Z) := (\sum_{i=1}^k \mathrm{Var}[Z_i])^{1/2}$.

As a first step, let us assume each $\mathbb{E}[X_i]$ is nonzero and normalize the variables X_i by writing

$$\prod_i X_i = \prod_i (\mathbb{E}[X_i] + (X_i - \mathbb{E}[X_i])) = \prod_i \mathbb{E}[X_i] \cdot \prod_i \left(1 + \frac{X_i - \mathbb{E}[X_i]}{\mathbb{E}[X_i]}\right).$$

Let Z_i denote $(X_i - \mathbb{E}[X_i])/\mathbb{E}[X_i]$. If $|Z_i|$ is small, then intuitively a low-order Taylor's expansion of $\prod_i (1 + Z_i)$ should approximate the original function well. To write down its Taylor's expansion, a convenient way is to rewrite $\prod_i (1 + Z_i)$ as $e^{\sum_i \log(1+Z_i)}$. It suffices to bound above its error term in expectation. This is equivalent to bounding the d-th moment of $\sum_i \log(1 + Z_i)$. A standard calculation gives a bound in terms of norm and variance of the $\log(1 + Z_i)$'s. Since $|Z_i|$ is small, $\log(1 + Z_i)$ behaves similarly as Z_i. So we can relate the error term in terms of $|Z_i|$ and $\sigma(Z)^2 := \sum_i \mathrm{Var}[Z_i]$. In particular if $|Z_i| \le B$ for all i then we would get an error bound of the form $2^{O(d)}(\sqrt{\sigma(Z)^2/d} + B)^{O(d)}$. For now let's think of $\mathbb{E}[X_i]$ is bounded away from 0 so that $\mathrm{Var}[Z_i] = \Theta(\mathrm{Var}[X_i])$.

Now we handle the case where $|Z_i|$ is large. Note that this implies either (1) $|X_i - \mathbb{E}[X_i]|$ is large, or (2) $\mathbb{E}[X_i]$ is small.

We will handle the two conditions separately by a reduction to the case where the $|Z_i|$'s are small.

The recurring idea throughout is that we can always tolerate $O(d)$ bad variables that violates the conditions, provided with high probability there can be at most $O(d)$ bad variables. This is because by affording an extra $O(d)$ amount of independence in the beginning, we can condition on the values of these variables and work with remaining variables.

As a simple illustration of this idea, throughout the proof we can assume $\mathrm{Var}[X_i] \le \sum_i \mathrm{Var}[X_i]/d =: \sigma(X)^2/d$, as there can be at most d bad variables that violates this inequality, and so we can start with $2d$-wise independence, then condition on values of the bad X_i's and the rest of the X_i would satisfy the inequality.

We first assume the $|\mathbb{E}[X_i]|$'s are large and handle (1), we will round the X_i to $\mathbb{E}[X_i]$ whenever $|X_i - \mathbb{E}[X_i]| \le B$. Note that by Chebyshev's inequality an X_i gets rounded with probability $\mathrm{Var}[X_i]/B^2$. It follows that the probability that there are more than d such X_i's is bounded by $(\sigma(X)/Bd)^d$. This suggests taking B to be $(\sigma(X)/d)^\alpha$ for some constant $\alpha \in (0, 1)$ to balance the error terms.

It remains to handle condition (2), for Z_i to be bounded by $B = (\sigma(X)^2/d)^{\Omega(1)}$, as explained above it suffices to show that all but $O(d)$ of the X_i's satisfy $|\mathbb{E}[X_i]| \ge (\sigma(X)^2/d)^{O(1)}$. If $|\mathbb{E}[X_i]| \ge (\sigma(X)/d)^{\Omega(1)}$ for $\Omega(d)$ of the X_i's, then by a similar argument as above one can show that with high probability at least half of them is bounded by $(\sigma(X)^2/d)^{\Omega(1)}$. Hence, $\mathbb{E}[\prod_i X_i]$ is at most $(\sigma(X)^2/d)^{\Omega(d)}$ when the X_i's are d-wise independent. This finishes the proof.

Note that in the case of $\varepsilon > 0$, each X_i is only ε-close to the corresponding Y_i and they are not exactly identical. As a result, throughout the proof we will often have to introduce hybrid terms to move from functions of X_i to functions of Y_i, and vice versa. Each of these steps introduces an error of at most $k^{O(d)}\varepsilon$.

Also, there is some loss in ε whenever we condition on the values of any subset of the

Y_i's, see Claim 22 for a formal claim. This induces the extra condition that each X_i must put a certain mass on each outcome.

B.1 Preliminaries

In this section, we prove several claims that will be used throughout the proof of Lemma 9.

Lemma 15. *For any $z \in \mathbb{C}$ with $|z| \le 1/2$, $|\log(1+z)| \le 2|z|$, where we take the principle branch of the logarithm.*

Proof. From the Taylor series expansion of the complex-valued log function we have

$$|\log(1+z)| = \left|\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n!} z^n\right| \le \sum_{n=1}^{\infty} |z|^n \le |z| \sum_{n=0}^{\infty} (1/2)^n = 2|z|. \qquad \square$$

Lemma 16. *Let $Z \in \mathbb{C}$ be a random variable with $|Z| \le 1/2$, $\mathbb{E}[Z] = 0$ and $W = \log(1+Z)$ the principle branch of the logarithm function (phase between $(-\pi, \pi)$). We have $\mathrm{Var}[W] \le 4\,\mathrm{Var}[Z]$.*

Proof. By the definition of Variance, Lemma 15, and that $\mathbb{E}[Z] = 0$,

$$\begin{aligned} \mathrm{Var}[W] &= \mathbb{E}[|W|^2] - |\mathbb{E}[W]|^2 \\ &\le \mathbb{E}[|W|^2] \\ &\le 4\,\mathbb{E}[|Z|^2] \\ &= 4\,\mathrm{Var}[Z]. \end{aligned} \qquad \square$$

Lemma 17 (Taylor's approximation)**.** *For $w \in \mathbb{C}$ and $d > 0$,*

$$\left|e^w - \sum_{j=0}^{d-1} w^j/j!\right| \le O(1)\frac{|w|^d}{d!} \cdot \max\{1, e^{\Re(w)}\}.$$

Lemma 18. *For any random variable $W \in \mathbb{C}$, $|e^{\mathbb{E}[W]}| \le \mathbb{E}[|e^W|]$.*

Proof. By Jensen's inequality, we have

$$|e^{\mathbb{E}[W]}| = |e^{\mathbb{E}[\Re(W)]}| \le |\mathbb{E}[e^{\Re(W)}]| = \mathbb{E}[|e^W|]. \qquad \square$$

Claim 19. *$|e^{z_1} - e^{z_2}| \le |e^{z_2}| \cdot O(|z_1 - z_2|)$ if $|z_1 - z_2| \le 1$,*

Proof. By Lemma 17 with $d = 1$,

$$|e^{z_1 - z_2} - 1| \le O(1) \cdot |z_1 - z_2| \cdot \max\{1, e^{\Re(z_1 - z_2)}\} = O(|z_1 - z_2|),$$

because $\Re(z_1 - z_2) \le |z_1 - z_2| \le 1$. Therefore,

$$\begin{aligned} |e^{z_1} - e^{z_2}| &= |e^{z_2}(e^{z_1 - z_2} - 1)| \\ &= |e^{z_2}||e^{z_1 - z_2} - 1| \\ &\le |e^{z_2}| \cdot O(|z_1 - z_2|). \end{aligned} \qquad \square$$

Claim 20. *Let $X, Y \in \Omega$ be two discrete random variables such that $\mathrm{sd}(X,Y) \le \varepsilon$. Let $f\colon \Omega \to \mathbb{C}$ be any function. We have $|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \le 2\max_z |f(z)| \cdot \mathrm{sd}(X,Y)$.*

Proof. Let p and q be the probability function of X and Y. Using the fact that $\mathrm{sd}(X,Y) = \frac{1}{2}\sum_z |p(z) - q(z)|$, we have

$$\begin{aligned}
\Big|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]\Big| &= \Big|\sum_z p(z)f(z) - \sum_z q(z)f(z)\Big| \\
&\le \sum_z |f(z)||p(z) - q(z)| \\
&\le \max_z |f(z)| \cdot \sum_z |p(z) - q(z)| \\
&= 2\max_z |f(z)| \cdot \mathrm{sd}(X,Y).
\end{aligned}$$

□

Claim 21 (Maclaurin's inequality (cf. [Ste04]))**.** *Let $z_1, \ldots, z_k$ be k non-negative numbers. For any $i \in \{0, \ldots, k\}$, we have*

$$S_i(z_1, \ldots, z_k) := \sum_{S:|S|=i} \prod_{j \in S} z_j \le (e/i)^i \Big(\sum_{j=1}^k z_j\Big)^i.$$

B.2 Proof of Lemma 9

We now prove Lemma 9. For independent random variables $Z_1, \ldots, Z_k$, we will use $\sigma(Z)$ to denote the standard deviation of $\sum_{i \le k} Z_i$, that is, $\sigma(Z) := (\sum_{i=1}^k \mathrm{Var}[Z_i])^{1/2}$. We will also denote $\sigma(Z)^2/d$ by v for notational simplicity.

B.2.1 Assuming the variances are not too small

As hinted in the overview above, throughout the proof we will without loss of generality assume $\mathrm{Var}[X_i] \le \sigma(X)^2/d$ for every $i \in \{1, \ldots, k\}$. This assumption will be used in the proof of Lemma 28 to give a uniform bound on how close the rounded X_i's and X_i's are in expectation.

We first prove a claim that shows the Y_i's remains close to the X_i even if we condition on the values of a few of the Y_i's. This claim will be used multiple times throughout the proof. Note that this claim is immediate for exact independence ($\varepsilon = 0$) but less for almost independence. We shall use the assumption that the X_i take any value with probability at least 2^{-n}.

Claim 22. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\le 1}$ with $\min_{z \in \mathrm{Supp}(X_i)} \Pr[X_i = z] \ge 2^{-n}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\le 1}$ that are (ε, d)-close to $X_1, X_2, \ldots, X_k$. Let $S \subseteq \{1, \ldots, k\}$ be a subset of size t. Then conditioned on any values of the Y_i for $i \in S$, the Y_i for $i \notin S$ are $(3 \cdot 2^{2tn}\varepsilon, d - t)$-close to the X_i for $i \notin S$.*

Proof. Let $T \subseteq [k] - S$ be a subset of size at most $d-t$. We have for any value z_ℓ for $\ell \in S$,

$$\sum_{z_j : j \in T} \left| \Pr\Big[\bigwedge_{j \in T} Y_j = z_j \mid \bigwedge_{\ell \in S} Y_\ell = z_\ell \Big] - \Pr\Big[\bigwedge_{j \in T} X_j = z_j \Big] \right|$$
$$= \sum_{z_j : j \in T} \left| \frac{\Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j \Big]}{p_Y} - \frac{\Pr\Big[\bigwedge_{j \in S \cup T} X_j = z_j \Big]}{p_X} \right|,$$

where $p_X := \Pr[\wedge_{\ell \in S} X_\ell = z_\ell]$ and $p_Y := \Pr[\wedge_{\ell \in S} Y_\ell = z_\ell]$. Hence, we can rewrite above as

$$\sum_{z_j : j \in T} \left| \Big(\frac{1}{p_Y} - \frac{1}{p_X} \Big) \Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j \Big] + \frac{1}{p_X} \Big(\Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j \Big] - \Pr\Big[\bigwedge_{j \in S \cup T} X_j = z_j \Big] \Big) \right|$$
$$\le \Big| \frac{1}{p_Y} - \frac{1}{p_X} \Big| \sum_{z_j : j \in T} \Pr\Big[\bigwedge_{j \in S \cup T} Y_j = z_j \Big] + \frac{\varepsilon}{p_X}$$
$$\le |1/p_Y - 1/p_X| + \varepsilon/p_X$$
$$\le (1/p_X p_Y + 1/p_X)\varepsilon.$$

The first and last inequalities are because the X_i's are (ε, d)-close to the Y_i's. As the X_i's are independent, by our assumption we have $p_X = \prod_{\ell \in S} \Pr[X_\ell = z_\ell] \ge 2^{-tn}$, and so $p_Y \ge 2^{-tn} - \varepsilon \ge 2^{-tn}/2$. (Otherwise the conclusion is trivial.) Therefore, $(1/p_X p_Y + 1/p_X)\varepsilon \le 3 \cdot 2^{2tn}\varepsilon$, and the proof follows. □

Claim 23. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\le 1}$ with $\min_{z \in \mathrm{Supp}(X_i)} \Pr[X_i = z] \ge 2^{-n}$ for each $i \in \{1, \ldots, k\}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\le 1}$. If Lemma 9 holds when the Y_i's are (Cd, ε)-close to the X_i's assuming $\mathrm{Var}[X_i] \le \sigma(X)^2/d$ for every $i \in [k]$, then Lemma 9 holds when the Y_i's are $((C+1)d, \varepsilon)$-close the X_i's without the assumption.*

Proof. Note that there can be at most d different such indices. Let J be the set of these indices. We have

$$\prod_i X_i - \prod_i Y_i = \prod_{j \in J} X_j \prod_{i \notin J} X_i - \prod_{j \in J} Y_j \prod_{i \notin J} Y_i$$
$$= \Big(\prod_{j \in J} X_j - \prod_{j \in J} Y_j \Big) \prod_{i \notin J} X_j + \prod_{j \in J} Y_j \Big(\prod_{i \notin J} X_j - \prod_{i \notin J} Y_j \Big).$$

We first bound the expectation of the first term. Since the X_i's are independent,

$$\left| \mathop{\mathbb{E}}_{X,Y} \Big[\Big(\prod_{j \in J} X_j - \prod_{j \in J} Y_j \Big) \prod_{i \notin J} X_j \Big] \right| = \left| \mathbb{E}\Big[\prod_{j \in J} X_j \Big] - \mathbb{E}\Big[\prod_{j \in J} Y_j \Big] \right| \cdot \left| \mathbb{E}\Big[\prod_{i \notin J} X_j \Big] \right|$$
$$\le \left| \mathbb{E}\Big[\prod_{j \in J} X_j \Big] - \mathbb{E}\Big[\prod_{j \in J} Y_j \Big] \right|$$
$$\le \varepsilon.$$

For the second term, note that conditioning on the values of the Y_j for which $j \in J$, by Claim 22, the remaining variables are $(2^{O(dn)}\varepsilon, Cd)$-close to the corresponding X_j's. So we can apply the above Lemma 9 with our assumption and the claim follows. □

B.2.2 Assuming the variables are close to their expectations and the expectations are large

Lemma 24. *Let $X_1, X_2, \ldots, X_k$ be k independent discrete random variables over $\mathbb{C}_{\leq 1}$. Let $Y_1, Y_2, \ldots, Y_k$ be k discrete random variables over $\mathbb{C}_{\leq 1}$ that are (ε, d)-close to $X_1, \ldots, X_k$. Assume for each X_i and Y_i, there exist Z_i and Z_i' such that*

$$X_i = \mathbb{E}[X_i](1 + Z_i) \quad \text{and} \quad Y_i = \mathbb{E}[X_i](1 + Z_i'),$$

where $|Z_i| \leq B \leq 1/2$ and $|Z_i'| \leq B \leq 1/2$. Then

$$\left| \mathbb{E}\Big[\prod_{i=1}^{k} X_i\Big] - \mathbb{E}\Big[\prod_{i=1}^{k} Y_i\Big] \right| \leq 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d} \right)^d + (Bk)^{O(d)}\varepsilon.$$

Remark 25. *Note that we define Y_i above in terms of $\mathbb{E}[X_i]$ but not $\mathbb{E}[Y_i]$. The Z_i's are independent, but the Z_i's may not be. Also, later we will take B to be $v^{1/3}$.*

Proof. Define $W_i, \hat{W}_i$ such that

$$W_i = \log(1 + Z_i) \quad \text{and} \quad \hat{W}_i = W_i - \mathbb{E}[W_i].$$

Likewise, define $W_i', \hat{W}_i'$ such that

$$W_i' = \log(1 + Z_i') \quad \text{and} \quad \hat{W}_i' = W_i' - \mathbb{E}[W_i'].$$

Let $\hat{W} = \sum_i \hat{W}_i$ and $\hat{W}' = \sum_i \hat{W}_i'$. Note that $X_i = \mathbb{E}[X_i]e^{\hat{W}_i + \mathbb{E}[W_i]}$ and $Y_i = \mathbb{E}[Y_i]e^{\hat{W}_i' + \mathbb{E}[W_i']}$. We have

$$\prod_{i=1}^{k} X_i = \Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i]}\Big) e^{\hat{W}} \quad \text{and} \quad \prod_{i=1}^{k} Y_i = \Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}'}.$$

Hence the difference is

$$\begin{aligned}
\prod_{i=1}^{k} X_i - \prod_{i=1}^{k} Y_i &= \Big(\prod_{i=1}^{k} \mathbb{E}[X_i]\Big)\Big(\prod_{i=1}^{k} e^{\mathbb{E}[W_i]} \cdot e^{\hat{W}} - \prod_{i=1}^{k} e^{\mathbb{E}[W_i']} \cdot e^{\hat{W}'}\Big) \\
&= \Big(\prod_{i=1}^{k} \mathbb{E}[X_i]\Big)\left(\Big(\prod_{i=1}^{k} e^{\mathbb{E}[W_i]} - \prod_{i=1}^{k} e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}} + \prod_{i=1}^{k} e^{\mathbb{E}[W_i']} \cdot \Big(e^{\hat{W}} - e^{\hat{W}'}\Big)\right).
\end{aligned}$$

The lemma follows from the two claims below:

Claim 26. *For every outcome of $\hat{W}$, $\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i]\Big)\Big(\prod_{i=1}^{k} e^{\mathbb{E}[W_i]} - \prod_{i=1}^{k} e^{\mathbb{E}[W_i']}\Big) e^{\hat{W}}\right| \leq O(k\varepsilon)$.*

Claim 27. $\left|\left(\prod_{i=1}^{k}\mathbb{E}[X_i]e^{\mathbb{E}[W_i']}\right)\left(\mathbb{E}[e^{\hat{W}}]-\mathbb{E}[e^{\hat{W}'}]\right)\right| \le 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d}+Bd}{d}\right)^d + (Bk)^{O(d)}\varepsilon.$

□

Proof of Claim 26. We have

$$\left|\Big(\prod_{i=1}^{k}\mathbb{E}[X_i]\Big)\Big(\prod_{i=1}^{k}e^{\mathbb{E}[W_i]}-\prod_{i=1}^{k}e^{\mathbb{E}[W_i']}\Big)e^{\hat{W}}\right| = \left|\prod_{i=1}^{k}\mathbb{E}[X_i]\right|\cdot\left|\prod_{i=1}^{k}e^{\mathbb{E}[W_i]}-\prod_{i=1}^{k}e^{\mathbb{E}[W_i']}\right|\cdot\left|e^{\hat{W}}\right|$$

Since $|\sum_i \mathbb{E}[W_i]-\sum_i\mathbb{E}[W_i']| \le \sum_i|\mathbb{E}[W_i]-\mathbb{E}[W_i']| \le k\varepsilon$, by Claim 19,

$$\begin{aligned}\left|\prod_{i=1}^{k}e^{\mathbb{E}[W_i]}-\prod_{i=1}^{k}e^{\mathbb{E}[W_i']}\right| &= \left|e^{\sum_i\mathbb{E}[W_i]}-e^{\sum_i\mathbb{E}[W_i']}\right| \\ &\le \left|e^{\sum_i\mathbb{E}[W_i]}\right|\cdot O(k\varepsilon) \\ &= \left|\prod_{i=1}^{k}e^{\mathbb{E}[W_i]}\right|\cdot O(k\varepsilon),\end{aligned}$$

Therefore,

$$\begin{aligned}\left|\prod_{i=1}^{k}\mathbb{E}[X_i]\right|\cdot\left|\prod_{i=1}^{k}e^{\mathbb{E}[W_i]}-\prod_{i=1}^{k}e^{\mathbb{E}[W_i']}\right|\cdot\left|e^{\hat{W}}\right| &\le \left|\prod_{i=1}^{k}\mathbb{E}[X_i]\right|\cdot\left|\prod_{i=1}^{k}e^{\mathbb{E}[W_i]}\right|\cdot O(k\varepsilon)\cdot\left|e^{\hat{W}}\right| \\ &= \left|\Big(\prod_{i=1}^{k}\mathbb{E}[X_i]e^{\mathbb{E}[W_i]}\Big)e^{\hat{W}}\right|\cdot O(k\varepsilon) \\ &= \left|\prod_{i=1}^{k}X_i\right|\cdot O(k\varepsilon) \\ &\le O(k\varepsilon).\end{aligned}$$

□

Proof of Claim 27. We first rewrite $e^{\hat{W}}-e^{\hat{W}'}$ as a sum of 3 terms:

$$e^{\hat{W}}-e^{\hat{W}'} = \Big(e^{\hat{W}}-\sum_{j=0}^{d-1}\hat{W}^j/j!\Big)+\Big(\sum_{j=0}^{d-1}(\hat{W}^j-\hat{W}'^j)/j!\Big)+\Big(\sum_{j=0}^{d-1}\hat{W}'^j/j!-e^{\hat{W}'}\Big).$$

It suffices to bound above the expectation of each term multiplied by $\gamma := \prod_{i=1}^{k}\mathbb{E}[X_i]e^{\mathbb{E}[W_i']}$. We bound the first and last terms using Taylor's approximation (Lemma 17), and the second

term using (ε, d)-closeness of the variables. We will show the following:

$$\mathbb{E}\left[\left|\gamma \cdot \left(e^{\hat{W}'} - \sum_{j=0}^{d-1} \hat{W'}^j/j!\right)\right|\right] \leq 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d + (kB)^{O(d)}\varepsilon \tag{4}$$

$$\mathbb{E}\left[\left|\gamma \cdot \left(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\right)\right|\right] \leq 2^{O(d)} \left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d \tag{5}$$

$$\left|\gamma \cdot \mathbb{E}\left[\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W'}^j)/j!\right]\right| \leq k^d\varepsilon. \tag{6}$$

For (4), by Lemma 17 we have

$$\left|\gamma \cdot \left(e^{\hat{W}'} - \sum_{j=0}^{d-1} \hat{W'}^j/j!\right)\right| \leq |\gamma| \cdot O(1)\frac{|\hat{W'}|^d}{d!} \cdot \max\{1, e^{\Re(\hat{W'})}\}.$$

We now bound above $|\gamma \cdot \max\{1, e^{\Re(\hat{W'})}\}|$ by 1. We have

$$\begin{aligned}
|\gamma| &= \left|\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\right| \\
&= \left|\prod_{i=1}^{k} \mathbb{E}[X_i]\right| \cdot \left|e^{\mathbb{E}[\sum_i W_i']}\right| \\
&\leq \left|\prod_{i=1}^{k} \mathbb{E}[X_i]\right| \cdot \mathbb{E}[|e^{\sum_i W_i'}|] && \text{(Jensen's inequality, see Lemma 18)} \\
&= \mathbb{E}\left[\left|\prod_{i=1}^{k} \mathbb{E}[X_i] \cdot e^{\sum_i W_i'}\right|\right] \\
&= \mathbb{E}\left[\left|\prod_{i=1}^{k} Y_i\right|\right] \\
&\leq 1.
\end{aligned}$$

Moreover,

$$|\gamma \cdot e^{\Re(\hat{W'})}| = \left|\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\right| \cdot e^{\Re(\hat{W'})} = \left|\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']} e^{\hat{W'}}\right| = \left|\prod_{i=1}^{k} Y_i\right| \leq 1.$$

Hence, it suffices to bound above $\mathbb{E}[|\hat{W'}|^d]$. Note that the $\hat{W}_i'$'s are (ε, d)-close to the $\hat{W}_i$'s. So we bound above $|\hat{W}_i|$ and $\mathrm{Var}[\hat{W}_i]$ and then apply Lemma 35. First, since $|Z_i| \leq B$, we

have $|W_i| \le 2B$ because of Lemma 15, and so $|\hat{W}_i| \le |W_i| + |\mathbb{E}[W_i]| \le 4B$. Next, we have $\operatorname{Var}[\hat{W}_i] \le 4 \operatorname{Var}[Z_i]$ because of Lemma 16, and so $\sigma(\hat{W}) \le 2\sigma(Z)$. Therefore, by Lemma 35,

$$\begin{aligned}
\mathbb{E}\left[\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(e^{\hat{W}'} - \sum_{j=0}^{d-1} \hat{W}'^{j}/j!\Big)\right|\right] &\le O(1)\frac{\mathbb{E}[|\hat{W}'|^d]}{d!} \\
&\le 2^{O(d)}\left(\frac{\sigma(\hat{W})\sqrt{d} + 4Bd}{d}\right)^d + (kB)^{O(d)}\varepsilon \\
&\le 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d + (kB)^{O(d)}\varepsilon.
\end{aligned}$$

We prove Inequality (5) similarly. Note that

$$\begin{aligned}
\frac{|e^{\sum_i \mathbb{E}[W_i']}|}{|e^{\sum_i \mathbb{E}[W_i]}|} &= |e^{\sum_i \mathbb{E}[W_i'] - \sum_i \mathbb{E}[W_i]}| \\
&\le e^{|\sum_i \mathbb{E}[W_i'] - \sum_i \mathbb{E}[W_i]|} \\
&\le e^{\sum_i |\mathbb{E}[W_i'] - \mathbb{E}[W_i]|} \\
&\le e^{k\varepsilon} \\
&\le O(1),
\end{aligned}$$

because $\varepsilon < 1/k$, otherwise the conclusion is trivial. Hence,

$$\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right| \le \left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i]}\Big)\Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right| \cdot O(1).$$

Therefore, it follows by Inequality (1) by considering $\varepsilon = 0$ that

$$\mathbb{E}\left[\left|\Big(\prod_{i=1}^{k} \mathbb{E}[X_i] e^{\mathbb{E}[W_i']}\Big)\Big(e^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j/j!\Big)\right|\right] \le 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d} + Bd}{d}\right)^d.$$

Finally we prove Inequality (6). By linearity of expectation,

$$\mathbb{E}\Big[\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W}'^j)/j!\Big] = \sum_{j=0}^{d-1} (\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W}'^j])/j!\,.$$

Note that $\hat{W}^j = (\sum_i \hat{W}_i)^j$ can be written as a sum of k^j terms where each term is a product of at most $j \le d$ different W_i's. Moreover, we have $|W_i| \le 2B \le 1$ for each i because of

Lemma 15. So we have $|\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W'}^j]| \le k^j\varepsilon$. Hence,

$$\begin{aligned}\left|\mathbb{E}\Big[\sum_{j=0}^{d-1}(\hat{W}^j - \hat{W'}^j)/j!\Big]\right| &\le \sum_{j=0}^{d-1}|\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W'}^j]| \\ &\le \sum_{j=0}^{d-1} k^j\varepsilon \\ &\le k^d\varepsilon.\end{aligned}$$

Recall that $|\gamma| \le 1$, this concludes (6). □

B.2.3 Assuming the expectations are large

We now prove the main lemma assuming the expectation of the X_i are far from zero.

Lemma 28. *Let $X_1, X_2, \ldots, X_k$ be k independent random variables over $\mathbb{C}_{\le 1}$, with $\min_{z\in\mathrm{Supp}(X_i)} \Pr[X_i = z] \ge 2^{-n}$. Let $Y_1, Y_2, \ldots, Y_k$ be k random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon, 9d)$-close to $X_1, \ldots, X_k$. Assume $|\mathbb{E}[X_i]| \ge (\sigma(X)^2/d)^{1/6}$ for each i. We have*

$$\left|\mathbb{E}\Big[\prod_{i=1}^{k} X_i\Big] - \mathbb{E}\Big[\prod_{i=1}^{k} Y_i\Big]\right| \le 2^{O(d)}\left(\frac{\sigma(X)^2}{d}\right)^d + (k2^n)^{O(d)}\varepsilon.$$

Proof of Lemma 28. We will assume $\sigma(X)^2/d$ is less than a sufficiently small constant and $\varepsilon \le (k2^n)^{-Cd}$ for a sufficiently large C; otherwise the R.H.S. of the inequality is greater than 2 and there is nothing to prove.

For each $i \in \{1, 2, \ldots, k\}$, we define a new function $\mathrm{rd}_i \colon \mathbb{C}_{\le 1} \to \mathbb{C}_{\le 1}$ that will be used to round the variables X_i and Y_i. We define rd_i as

$$\mathrm{rd}_i(z) := \begin{cases} z & \text{if } |z - \mathbb{E}[X_i]| \le (\sigma(X)^2/d)^{1/3} \\ \mathbb{E}[X_i] & \text{otherwise.}\end{cases}$$

Let $\tilde{X}_i = \mathrm{rd}_i(X_i)$ and $\tilde{Y}_i = \mathrm{rd}_i(Y_i)$. We will write both $\prod_i X_i$ and $\prod_i Y_i$ as

$$\prod_{i=1}^{k} X_i = \prod_{i=1}^{k}(X_i - \tilde{X}_i + \tilde{X}_i) = \sum_{S\subseteq\{1,2,\ldots,k\}}\prod_{i\in S}(X_i - \tilde{X}_i)\prod_{i\notin S}\tilde{X}_i,$$

and

$$\prod_{i=1}^{k} Y_i = \prod_{i=1}^{k}(Y_i - \tilde{Y}_i + \tilde{Y}_i) = \sum_{S\subseteq\{1,2,\ldots,k\}}\prod_{i\in S}(Y_i - \tilde{Y}_i)\prod_{i\notin S}\tilde{Y}_i.$$

Let $m = 3d$. Define

$$P_m(z_1, z_2, \ldots, z_k) = \sum_{|S|<m}\prod_{i\in S}(z_i - \mathrm{rd}_i(z_i))\prod_{i\notin S}\mathrm{rd}_i(z_i).$$

We will show that P_m is a good approximation of the product in expectation over both X_i's and Y_i's and then show that the expectations of P_m under X_i's and Y_i's are close.

We will use the following inequalities repeatedly.

Claim 29. $\Pr[\tilde{X}_i \neq X_i] \leq \mathrm{Var}[X_i]v^{-2/3} \leq v^{1/3}$. *In particular,* $\sum_i \Pr[\tilde{X}_i \neq X_i] \leq (d\sigma)^{2/3}$.

Proof. The first inequality follows from Chebyshev's inequality and second follows from the assumption $\mathrm{Var}[X_i] \leq v$. The last sentence is implied by the first inequality. □

Claim 30. $\left|\mathbb{E}\left[\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\right]\right| \leq 2^{O(d)}v^d + k^{O(d)}\varepsilon$.

Proof. Consider the product $\prod_{i\in S}(Y_i - \tilde{Y}_i)$. Let N' be the number of $i \in \{0, 1, 2, \ldots, k\}$ such that $\tilde{Y}_i \neq Y_i$. If $N' < m$ then any set S of size at least m must contain an i such that $\tilde{Y}_i = Y_i$. In this case the product is 0 and thus

$$\prod_i Y_i - P_m(Y_1, \ldots, Y_k) = \sum_{|S|\geq m} \prod_{i\in S}(Y_i - \tilde{Y}_i) \prod_{i\notin S} \tilde{Y}_i = 0.$$

So,

$$\begin{aligned}
\left|\mathbb{E}\left[\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\right]\right| &= \left|\mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left(\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\right)\right]\right| \\
&\leq \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left(\left|\prod_i Y_i\right| + |P_m(Y_1, \ldots, Y_k)|\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left|\prod_i Y_i\right|\right] + \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot |P_m(Y_1, \ldots, Y_k)|\right].
\end{aligned}$$

If $N' \geq m$ then there can be at most $\sum_{\ell=0}^{m-1}\binom{N'}{\ell} \leq \sum_{\ell=0}^{m-1}\binom{N'}{m}\binom{m}{\ell} \leq 2^m\binom{N'}{m}$ subsets in the sum in P_m for which the product is nonzero, and each such product can be at most 2^m because $|S| < m$. Thus,

$$\begin{aligned}
\mathbb{1}(N' \geq m) \cdot \left|P_m(Y_1, \ldots, Y_k)\right| &\leq \mathbb{1}(N' \geq m) \cdot 2^m \sum_{\ell=0}^{m-1}\binom{N'}{\ell} \\
&\leq \mathbb{1}(N' \geq m) \cdot 2^m \cdot 2^m\binom{N'}{m} \\
&\leq 2^{2m}\binom{N'}{m}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left(\left|\prod_i Y_i\right| + |P_m(Y_1, \ldots, Y_k)|\right)\right] &\leq \mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \left|\prod_i Y_i\right|\right] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right] \\
&\leq \mathbb{E}[\mathbb{1}(N' \geq m)] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right].
\end{aligned}$$

We will show the following

Claim 31. $\Pr[N' \geq m] \leq \mathbb{E}\left[\binom{N'}{m}\right] \leq v^d + k^{O(d)}\varepsilon$.

Assuming the claim it follows that

$$\begin{aligned}\left|\mathbb{E}\left[\mathbb{1}(N' \geq m) \cdot \Big(\prod_i Y_i - P_m(Y_1, \ldots, Y_k)\Big)\right]\right| &\leq \mathbb{E}[\mathbb{1}(N' \geq m)] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right] \\ &\leq (1 + 2^{6d})((2v)^d + k^{O(d)}\varepsilon) \qquad (m = 6d) \\ &\leq 2^{O(d)}v^d + k^{O(d)}\varepsilon. \qquad \square\end{aligned}$$

We now prove Claim 31.

Proof of Claim 31. The first inequality is clear.

$$\begin{aligned}\mathbb{E}\left[\binom{N'}{m}\right] &\leq \sum_{|S|=m} \Pr[\wedge_{i\in S} Y_i \neq \tilde{Y}_i] \\ &\leq \sum_{|S|=m} \Big(\prod_{i\in S} \Pr[X_i \neq \tilde{X}_i] + \varepsilon\Big) && \text{(each } Y_i \text{ is } \varepsilon\text{-close to } X_i\text{)} \\ &\leq \sum_{|S|=m} \prod_{i\in S} \Pr[X_i \neq \tilde{X}_i] + k^m\varepsilon \\ &\leq \left(\frac{e\sum_{i=1}^k \Pr[X_i \neq \tilde{X}_i]}{m}\right)^m + k^m\varepsilon && \text{(Maclaurin's inequality)} \\ &\leq \left(\frac{e(d\cdot\sigma(X))^{2/3}}{3d}\right)^{3d} + k^m\varepsilon && \text{(Claim 29)} \\ &\leq v^d + k^{O(d)}\varepsilon. \qquad \square\end{aligned}$$

Now, we show that $P_m(Y_1, \ldots, Y_k)$ is close to $P_m(X_1, \ldots, X_k)$ in expectation.

Claim 32. $|\mathbb{E}[P_m(X_1, \ldots, X_k)] - \mathbb{E}[P_m(Y_1, \ldots, Y_k)]| \leq 2^{O(d)}v^d + O(k)^{3d}\varepsilon$.

Proof. The difference between $P_m(X_1, \ldots, X_k)$ and $P_m(Y_1, \ldots, Y_k)$ equals

$$P_m(X_1, \ldots, X_k) - P_m(Y_1, \ldots, Y_k) = \sum_{|S|<m} \Big(\prod_{i\in S}(X_i - \tilde{X}_i)\prod_{i\notin S}\tilde{X}_i - \prod_{i\in S}(Y_i - \tilde{Y}_i)\prod_{i\notin S}\tilde{Y}_i\Big).$$

We can rewrite the R.H.S. as

$$\sum_{|S|<m}\left(\Big(\prod_{i\in S}(X_i - \tilde{X}_i) - \prod_{i\in S}(Y_i - \tilde{Y}_i)\Big)\prod_{i\notin S}\tilde{X}_i + \prod_{i\in S}(Y_i - \tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i - \prod_{i\notin S}\tilde{Y}_i\Big)\right).$$

It suffices to show that

$$\left|\mathbb{E}\left[\sum_{|S|<m}\Big(\prod_{i\in S}(X_i-\tilde{X}_i)-\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big)\prod_{i\notin S}\tilde{X}_i\right]\right| \le k^{O(d)}\varepsilon \tag{7}$$

$$\left|\mathbb{E}\left[\sum_{|S|<m}\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i-\prod_{i\notin S}\tilde{Y}_i\Big)\right]\right| \le 2^{O(d)}v^d+(k2^n)^{O(d)}\varepsilon. \tag{8}$$

We first prove Inequality (7). Because the X_i's are independent, the L.H.S. of the inequality equals

$$\begin{aligned}
&\left|\sum_{|S|<m}\left(\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\right)\mathbb{E}\Big[\prod_{i\notin S}\tilde{X}_i\Big]\right|\\
&\le\sum_{\ell=1}^{m-1}\sum_{|S|=\ell}\left|\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\right|\cdot\left|\mathbb{E}\Big[\prod_{i\notin S}\tilde{X}_i\Big]\right|\\
&\le\sum_{\ell=1}^{m-1}\sum_{|S|=\ell}\left|\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\right|\\
&\le\sum_{\ell=1}^{m-1}\sum_{|S|=\ell}2\cdot 2^{\ell}\varepsilon\\
&\le\sum_{\ell=1}^{m-1}k^{\ell}\cdot 2\cdot 2^{\ell}\varepsilon\\
&\le 2(2k)^m\varepsilon\\
&= k^{O(d)}\varepsilon.
\end{aligned}$$

To see the third inequality, note that $|z-\mathrm{rd}_i(z)|\le 2$, and so $|\prod_{i\in S}(z_i-\mathrm{rd}_i(z_i))|\le 2^{|S|}$. So we can apply Claim 20 to bound above the absolute difference by $2\cdot 2^{|S|}\varepsilon$.

Now we prove Inequality (8). As $|S|\le m=3d$ and Y_i's are $(\varepsilon,9d)$-close to X_i's, conditioned on the values of $\tilde{X}_i$ for which $i\in S$, by Claim 22, the remaining $\tilde{Y}_i$'s for which $i\notin S$ are still $(2^{O(m\cdot n)}\varepsilon,6d)$-close to the corresponding $\tilde{X}_i$'s. (Recall that we can assume $\varepsilon=(k2^n)^{-Cd}$ for a sufficiently large C.) We will apply Lemma 24 to them.

Define Z_i, Z_i' such that $\tilde{X}_i=\mathbb{E}[\tilde{X}_i](1+Z_i)$ and $\tilde{Y}_i=\mathbb{E}[\tilde{X}_i](1+Z_i')$. To apply Lemma 24, we need the following two claims to bound above $|Z_i|, |Z_i'|$ and $\sigma(Z)^2$. We defer their proofs to the end.

Claim 33. *Let $B=4v^{1/6}$. Then $|Z_i|\le B$ and $|Z_i'|\le B$.*

Claim 34. $\sigma(Z)^2\le 4\sigma(X)^2v^{-1/3}$.

Therefore, by Lemma 24 with $\varepsilon' = 2^{O(m\cdot n)}\varepsilon$ and $B = 4(\sigma(X)^2/d)^{1/6} \le 1/2$ (Recall that we can assume $\sigma(X)^2/d$ less than a sufficiently small constant),

$$\left|\mathbb{E}\left[\sum_{|S|<m}\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i-\prod_{i\notin S}\tilde{Y}_i\Big)\right]\right| \le \sum_{|S|<m}\mathbb{E}\left[\left|\prod_{i\in S}(Y_i-\tilde{Y}_i)\right|\right]\cdot M,$$

where

$$\begin{aligned}
M &\le 2^{O(d)}\left(\frac{\sigma(Z)\sqrt{d}+dB}{d}\right)^{6d} + (Bk)^{O(d)}\varepsilon' \\
&\le 2^{O(d)}\left(\frac{\sigma(X)(\sigma(X)/\sqrt{d})^{-1/3}}{\sqrt{d}}+B\right)^{6d} + (Bk2^n)^{O(d)}\varepsilon \\
&\le 2^{O(d)}\left(\frac{\sigma(X)(\sigma(X)/\sqrt{d})^{-1/3}}{\sqrt{d}}+4v^{1/6}\right)^{6d} + (k2^n)^{O(d)}\varepsilon \\
&= 2^{O(d)}\left(v^{1/3}+v^{1/6}\right)^{6d} + (k2^n)^{O(d)}\varepsilon \\
&= 2^{O(d)}v^d + (k2^n)^{O(d)}\varepsilon.
\end{aligned}$$

We now bound above $\mathbb{E}[|\prod_{i\in S}(Y_i-\tilde{Y}_i)|]$. Note that $|\prod_{i\in S}(z_i-\mathrm{rd}_i(z_i))| \le 2^{|S|}$. Hence by Claim 20,

$$\mathbb{E}\left[\left|\prod_{i\in S}(Y_i-\tilde{Y}_i)\right|\right] \le \mathbb{E}\left[\left|\prod_{i\in S}(X_i-\tilde{X}_i)\right|\right] + 2^{|S|}\varepsilon.$$

Let N be the number of $i \in \{0,1,\dots,k\}$ such that $\tilde{X}_i \ne X_i$. Note that

$$\begin{aligned}
\sum_{|S|<m}\mathbb{E}\left[\left|\prod_{i\in S}(X_i-\tilde{X}_i)\right|\right] &\le \sum_{\ell=0}^{m-1}\left(2^\ell\,\mathbb{E}\left[\binom{N}{\ell}\right]\right) \\
&\le 2^m\,\mathbb{E}[2^N] \\
&= 2^m\prod_{i=1}^{k}\Big(1+\Pr[X_i\ne\tilde{X}_i]\Big) \\
&\le 2^m e^{\sum_i \Pr[X_i\ne\tilde{X}_i']} \\
&\le 2^m e^{(d\,\sigma(X))^{2/3}} \\
&\le 2^{O(d)},
\end{aligned}$$

where the last inequality is because $\sigma(X)^2/d \le 1$ and so $\sigma(X)^{2/3} \le d^{1/3}$. Therefore,

$$\sum_{|S|<m}\mathbb{E}\left[\left|\prod_{i\in S}(Y_i-\tilde{Y}_i)\right|\right] \le 2^{O(d)} + \sum_{|S|<m}2^{|S|}\varepsilon \le 2^{O(d)} + (2k)^m\varepsilon \le 2^{O(d)},$$

where the last inequality is because $\varepsilon \le k^{-Cd}$ for a sufficiently large C. So altogether the bound is $2^{O(d)} \cdot M$ as desired. □

□

We now prove Claim 33 and 34. By Claim 29, $|\mathbb{E}[X_i] - \mathbb{E}[\tilde{X}_i]| \le (\sigma(X)^2/d)^{1/3}$. Also by assumption, $|\mathbb{E}[X_i]| \ge (\sigma(X)^2/d)^{1/6}$. So, we have $|\mathbb{E}[\tilde{X}_i]| \ge |\mathbb{E}[X_i]|/2 \ge (\sigma(X)^2/d)^{1/6}/2$.

Proof of Claim 33. As $|\mathbb{E}[\tilde{X}_i]| \ge v^{1/6}/2$, we have

$$\begin{aligned}
|\tilde{Z}_i| &= \frac{|\tilde{X}_i - \mathbb{E}[\tilde{X}_i]|}{|\mathbb{E}[\tilde{X}_i]|} \\
&\le \frac{|\tilde{X}_i - \mathbb{E}[X_i]| + |\mathbb{E}[\tilde{X}_i] - \mathbb{E}[X_i]|}{|\mathbb{E}[\tilde{X}_i]|} \\
&\le 4v^{1/3}/v^{1/6} \\
&\le 4v^{1/6},
\end{aligned}$$

and the same argument holds for $|\tilde{Z}'_i|$ because $|\tilde{Y}_i - \mathbb{E}[X_i]| \le v^{1/3}$. □

Proof of Claim 34. Since $z^* = \mathbb{E}[Z]$ is the minimizer of $\mathbb{E}[|Z - z|^2]$, we have

$$\begin{aligned}
\operatorname{Var}[\tilde{X}_i] &= \mathbb{E}[|\tilde{X}_i - \mathbb{E}[\tilde{X}_i]|^2] \\
&\le \mathbb{E}[|\tilde{X}_i - \mathbb{E}[X_i]|^2] \\
&\le \mathbb{E}[|X_i - \mathbb{E}[X_i]|^2] && (\tilde{X}_i = \mathrm{rd}_i(X_i)) \\
&= \operatorname{Var}[X_i].
\end{aligned}$$

Therefore, $\operatorname{Var}[\tilde{Z}_i] = \operatorname{Var}[\tilde{X}_i]/|\mathbb{E}[\tilde{X}_i]|^2 \le 4 \operatorname{Var}[X_i] v^{-1/3}$ and thus $\sum_i \operatorname{Var}[\tilde{Z}_i] \le 4\sigma(X)^2 v^{-1/3}$. □

B.2.4 The general case

Proof of Lemma 9. We will again assume $\sigma(X)^2/d$ is less than a sufficiently small constant and $\varepsilon \le (k2^n)^{-Cd}$ for a sufficiently large constant C. We first assume $\operatorname{Var}[X_j] \le \sigma(X)^2/d$ for all j and prove the lemma when the Y_i's are $(\varepsilon, 15d)$-close to the X_i's. Later we will handle the general case.

Let m be the number of i such that $|\mathbb{E}[X_i]| \le v^{1/6}$.

If $m \le 6d$, let J be the set of indices for which $|\mathbb{E}[X_i]| \le v^{1/6}$. We can write

$$\prod_i X_i - \prod_i Y_i = \left(\prod_{j \in J} X_j - \prod_{j \in J} Y_j\right) \prod_{j \notin J} X_j + \prod_{j \in J} Y_j \left(\prod_{j \notin J} X_j - \prod_{j \notin J} Y_j\right).$$

It suffices to show that

$$\left|\mathbb{E}\left[\Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big)\prod_{j\notin J} X_j\right]\right| \le \varepsilon \tag{9}$$

$$\left|\mathbb{E}\left[\prod_{j\in J} Y_j\left(\prod_{j\notin J} X_j - \prod_{j\notin J} Y_j\right)\right]\right| \le 2^{O(d)}v^d + (k2^n)^{O(d)}\varepsilon. \tag{10}$$

We first show Inequality (9). Since the X_i's are independent, the L.H.S. of (9) is

$$\begin{aligned}\left|\left(\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right)\mathbb{E}\Big[\prod_{j\notin J} X_j\Big]\right| &\le \left|\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right| \\ &\le \varepsilon.\end{aligned}$$

To prove Inequality (10), note that conditioned on the values of the Y_i's for which $i \in J$, by Claim 22, the rest of the Y_i's are still $(2^{O(dn)}\varepsilon, 9d)$-close to the corresponding X_i's with $|\mathbb{E}[X_i]| \ge v^{1/6}$. (Recall that we can assume $\varepsilon = (k2^n)^{-Cd}$ for a sufficiently large C.) So the bound follows from Lemma 28.

If $m \ge 6d$, then note that

$$\left|\mathbb{E}\Big[\prod_{i=1}^{k} X_i\Big]\right| = \prod_{i=1}^{k}|\mathbb{E}[X_i]| \le v^{m/6} \le v^d.$$

So it suffices to show that

$$\left|\mathbb{E}\Big[\prod_{i=1}^{k} Y_i\Big]\right| \le 2^{O(d)}v^d + k^{O(d)}\varepsilon.$$

Consider the event E that at least $3d$ of the the Y_i for $i \in J$ have absolute value less than $2v^{1/6}$. Then we know that

$$\Big|\prod_{i=1}^{k} Y_i\Big| \le 2^{3d}\cdot v^{d/2}.$$

We will show that E happens except with probability at most $v^{2d} + k^{3d}\varepsilon$. Let $N \in \{0, 1, 2, \ldots, m\}$ be the number of $i \in J$ such that $|Y_i| \ge 2v^{1/6}$. Note that

$$\begin{aligned}\Pr[N \ge 3d] &\le \sum_{S\subseteq J:|S|=3d} \Pr\left[\bigwedge_{i\in S}\left(|Y_i| \ge 2v^{1/6}\right)\right] \\ &\le \sum_{S\subseteq J:|S|=3d}\prod_{i\in S}\Pr\left[|X_i| \ge 2v^{1/6}\right] + k^{3d}\varepsilon.\end{aligned}$$

By Chebyshev's inequality,

$$\Pr[|X_i| \ge 2v^{1/6}] \le \Pr[|X_i - \mathbb{E}[X_i]| \ge v^{1/6}] \le \mathrm{Var}[X_i]v^{-1/3}.$$

Hence, by Maclaurin's inequality,

$$\begin{aligned}\sum_{S\subseteq J:|S|=3d}\prod_{i\in S}\Pr\left[|X_i|\geq 2v^{1/6}\right] &\leq \left(\frac{e\sum_{i=1}^m \Pr[|X_i|\geq 2v^{1/6}]}{3d}\right)^{3d}\\ &\leq \left(\frac{e\sum_{i=1}^m \mathrm{Var}[X_i]v^{-1/3}}{3d}\right)^{3d}\\ &\leq \left(\frac{e\sigma(X)^2 v^{-1/3}}{3d}\right)^{3d}\\ &\leq v^{2d}.\end{aligned}$$

So,

$$\Pr[N\geq 3d]\leq v^{2d}+k^{3d}\varepsilon.$$

Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\prod_i Y_i\right]\right| &\leq 2^{3d}v^{d/2}+v^{2d}+k^{3d}\varepsilon\\ &\leq 2^{O(d)}v^{d/2}+k^{O(d)}\varepsilon.\end{aligned}$$

□

C Moment bounds for sum of almost d-wise independent variables

In this section we prove some moment bounds and tail bounds for sum of almost d-wise independent variables.

Lemma 35. *Let $Z_1, Z_2, \ldots, Z_k \in \mathbb{C}$ be independent random variables with $\mathbb{E}[Z_i]=0$, $|Z_i|<B$. Let d be an even positive integer. Let $W_1, W_2, \ldots, W_k \in \mathbb{C}$ be random variables that are (ε, d)-close to $Z_1, \ldots, Z_k$. Then,*

$$\mathbb{E}\left[\left|\sum_{i=1}^k W_i\right|^d\right]\leq 2^d\Big(\Big(\sum_i \mathrm{Var}[Z_i]\cdot d\Big)^{1/2}+dB\Big)^d+(2kB)^d\varepsilon.$$

Proof of Lemma 35. Note that for any random variable $W\in\mathbb{C}$ we have

$$\begin{aligned}\mathbb{E}\left[|W|^d\right] &= \mathbb{E}\left[\Big(|\Re(W)|^2+|\Im(W)|^2\Big)^{d/2}\right]\\ &\leq \mathbb{E}\left[\Big(2\max\{|\Re(W)|^2,|\Im(W)|^2\}\Big)^{d/2}\right]\\ &\leq 2^{d/2}\cdot\mathbb{E}\left[|\Re(W)|^d+|\Im(W)|^d\right],\end{aligned}$$

and $\mathrm{Var}[W]=\mathrm{Var}[\Re(W)]+\mathrm{Var}[\Im(W)]$. We will first prove the lemma when W is real-valued.

Since $W_1, \ldots, W_k$ are (ε, d)-close to $Z_1, \ldots, Z_k$, and d is even, we have

$$\begin{aligned}\mathbb{E}\left[\Big|\sum_{i=1}^{k} W_i\Big|^d\right] &= \mathbb{E}\left[\Big(\sum_i W_i\Big)^d\right] \\ &\leq \sum_{i_1,\ldots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right] + k^d B^d \varepsilon,\end{aligned}$$

because there are k^d products in the sum, each product is bounded by B^d and Claim 20. We now estimate the quantity $\sum_{i_1,\ldots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right]$. We have

$$\sum_{i_1,\ldots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right] = \sum_{m=1}^{d} \sum_{|S|=m} \sum_{\substack{i_1,\ldots,i_d \in S: \\ \{i_j\}_j = S}} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right].$$

The expectation is zero whenever Z_{i_j} appears only once for some $i_j \in S$. So each Z_{i_j} must appear at least twice. So the expectation is 0 whenever $m > d/2$. As each Z_i is bounded by B, each product is bounded by $B^{d-2m} \prod_{j\in S} \mathbb{E}[Z_j^2] = B^{d-2m} \prod_{j\in S} \mathrm{Var}[Z_j]$. For each $S \subseteq [k]$ of size m, there are at most m^d such terms. Let σ denote $(\sum_{i=1}^{k} \mathrm{Var}[Z_i])^{1/2}$. Then,

$$\begin{aligned}\sum_{i_1,\ldots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right] &\leq \sum_{m=1}^{d/2} B^{d-2m} m^d \sum_{|S|=m} \prod_{j\in S} \mathrm{Var}[Z_j] \\ &\leq \sum_{m=1}^{d/2} B^{d-2m} m^{d-m} e^m \sigma^{2m} && \text{(Maclaurin's inequality, see Claim 21)} \\ &\leq e^{d/2} \sum_{m=1}^{d/2} B^{d-2m} (d/2)^{d-m} \sigma^{2m} \\ &\leq e^{d/2} (d/2)^d B^d \sum_{m=0}^{d/2} \Big(\frac{\sigma^2}{(d/2)B^2}\Big)^m \\ &\leq e^{d/2} (d/2)^d B^d \cdot \left(d\Big(1 + \frac{\sigma^d}{(d/2)^{d/2} B^d}\Big)\right) && (\sum_{m=0}^{d-1} \alpha^m \leq d(\alpha^0 + \alpha^{d-1}), \forall \alpha > 0) \\ &\leq d e^{d/2} \Big((d/2)^d B^d + (d/2)^{d/2} \sigma^d\Big) \\ &\leq 2^{d/2} (dB + \sigma\sqrt{d})^d.\end{aligned}$$

Putting everything together, we have

$$\begin{aligned}\mathbb{E}\left[\Big|\sum_{i=1}^{k} W_i\Big|^d\right] &\leq 2^{d/2} \left(2^{d/2} (\sigma\sqrt{d} + dB)^d + (kB)^d \varepsilon\right) \\ &\leq 2^d (\sigma\sqrt{d} + dB)^d + (2kB)^d \varepsilon. \qquad \square\end{aligned}$$

Lemma 36. *Let $X_1, X_2, \ldots, X_k \in [0,1]$ be independent random variables. Let d be an even positive integer. Let $Y_1, Y_2, \ldots, Y_k \in [0,1]$ be random variables that are (ε, d)-close to $X_1, \ldots, X_k$. Let $Y = \sum_{i \le k} Y_i$ and $\mu = \mathbb{E}[\sum_i X_i]$. Then,*

$$\Pr[|Y - \mu| \ge \delta\mu] \le 2^d \left(\frac{\sqrt{\mu d} + d}{\delta\mu}\right)^d + \left(\frac{2k}{\delta\mu}\right)^d \varepsilon.$$

In particular, if $\mu \ge 25d$ and $\delta = 1/2$, we have $\Pr[|Y - \mu| \ge \mu/2] \le 2^{-\Omega(d)} + k^d\varepsilon$.

Proof. Let $X_i' = X_i - \mathbb{E}[X_i]$, $Y_i' = Y_i - \mathbb{E}[X_i]$ and $Y' = \sum_i Y_i'$. Note that $X_i' \in [-1, 1]$ and $\mathbb{E}[X_i'] = 0$. Since $X_i \in [0,1]$, we have

$$\mathbb{E}[X_i] \ge \mathbb{E}[X_i^2] \ge \mathrm{Var}[X_i] = \mathrm{Var}[X_i - \mathbb{E}[X_i]] = \mathrm{Var}[X_i'].$$

By Lemma 35 and Markov's inequality,

$$\begin{aligned}
\Pr[|Y - \mu| \ge \delta\mu] &= \Pr[|Y'|^d \ge (\delta\mu)^d] \\
&\le 2^d \left(\frac{(\sum_i \mathrm{Var}[X_i'] \cdot d)^{1/2} + d}{\delta\mu}\right)^d + \left(\frac{2k}{\delta\mu}\right)^d \varepsilon \\
&\le 2^d \left(\frac{\sqrt{\mu d} + d}{\delta\mu}\right)^d + \left(\frac{2k}{\delta\mu}\right)^d \varepsilon,
\end{aligned}$$

where in the last inequality we used $\mu \ge \sum_i \mathrm{Var}[X_i']$. □

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il