Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 161 (2017)

Pseudorandom Pseudo-Distributions with Near-Optimal Error for Read-Once Branching Programs

Mark Braverman* Gil Cohen† Sumegha Garg‡

Abstract

Nisan [Nis92] constructed a pseudorandom generator for length n, width n read-once branching programs (ROBPs) with error ε and seed length $O(\log^2 n + \log n \cdot \log(1/\varepsilon))$. A major goal in complexity theory is to reduce the seed length, hopefully, to the optimal $O(\log n + \log(1/\varepsilon))$, or to construct improved hitting sets, as these would yield stronger derandomization of **BPL** and **RL**, respectively. In contrast to a successful line of work in restricted settings, no progress has been made for general, unrestricted, ROBPs. Indeed, Nisan's construction is the best pseudorandom generator and, prior to this work, also the best hitting set for unrestricted ROBPs.

In this work, we make the first improvement for the general case by constructing a hitting set with seed length $\widetilde{O}(\log^2 n + \log(1/\varepsilon))$. That is, we decouple ε and n, and obtain near-optimal dependence on the former. The regime of parameters in which our construction strictly improves upon prior works, namely, $\log(1/\varepsilon) \gg \log n$, is also motivated by the work of Saks and Zhou [SZ99] who use pseudorandom generators with error ε, for length n, width w read-once branching programs, such that $w, 1/\varepsilon = 2^{(\log n)^2}$ in their proof for $\mathbf{BPL} \subseteq \mathbf{L}^{3/2}$.

In fact, we introduce and construct a new type of primitive we call *pseudorandom pseudo-distributions*. Informally, this is a generalization of pseudorandom generators in which one may assign negative and unbounded weights to paths as opposed to working with probability distributions. We show that such a primitive yields hitting sets and, for derandomization purposes, can be used to derandomize two-sided error algorithms.

*Department of Computer Science, Princeton University, Princeton, USA. Email: mbraverm@cs.princeton.edu. Research supported in part by NSF Awards, DMS-1128155, CCF-1525342, and CCF-1149888, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry. Part of this work was done while MB was a Fellow at the Institute for Advanced Study.

†Department of Computer Science, Princeton University, Princeton, USA. Email: coheng@gmail.com.

‡Department of Computer Science, Princeton University, Princeton, USA. Email: sumeghag@cs.princeton.edu.

ISSN 1433-8092

Contents

1 Introduction

Understanding the role that randomness plays in computation is of central importance in complexity theory. While randomness is provably necessary in many computational settings such as cryptography, distributed computing, and interactive proofs, by now it is widely believed that randomness adds no computational power to time-bounded nor to space-bounded algorithms. Surprisingly, proving such a statement for time-bounded algorithms implies circuit lower bounds which seem to be out of reach of current proof techniques [NW94, IKW02, KI04].

On the other hand, there is no known barrier for proving such a statement in the space-bounded setting. Indeed, while we cannot even rule out a scenario in which randomness "buys" exponential time, the space-bounded setting is much better understood. Savitch's theorem [Sav70] already implies that any one-sided error randomized algorithm can be simulated deterministically with only a quadratic overhead in space. $\mathbf{BPL} \subseteq \mathbf{L}^2$ can be proved easily through a variant of Savitch's theorem and also follows from [BCP83]. Nisan [Nis92, Nis94] proved that $\mathbf{BPL} \subseteq \mathbf{DTISP}(\mathrm{poly}(n), \log^2(n))$ using pseudorandom generators. The state of the art result was obtained by Saks and Zhou [SZ99] that build on Nisan's work to deterministically simulate two-sided error space s randomized algorithms in space $O(s^{3/2})$, thus, establishing that $\mathbf{BPL} \subseteq \mathbf{L}^{3/2}$.

There has been much work on the study of derandomizing space-bounded computation (see [NSW92, ATSWZ00, RR99, Tri08, DSTS17, MRSV17] and references therein). Unfortunately, the progress in derandomizing general space-bounded computation halted at once with the work of Saks and Zhou [SZ99]. Research began to focus on natural restricted settings and several exciting results were obtained, perhaps most notable is Reingold's celebrated result $\mathbf{SL} = \mathbf{L}$ [Rei08].

1.1 Pseudorandom distributions for ROBPs

Space-bounded algorithms are typically studied by considering their non-uniform counterparts. A length n, width w *read-once branching program* (ROBP) is a directed graph whose nodes, called states, are partitioned to n layers, each consists of at most w states, as well as an additional "start" state. The last layer consists of 2 states called "accept" and "reject". From every state but for the latter two, there are two outgoing edges, labeled by 0 and 1, to the following layer. On input $x \in \{0,1\}^n$, the computation proceeds by following the edges according to the labels given by the bits of x starting from the start state. The string x is accepted by the program if the computation ends in the accept state.

A well-known fact (see, e.g., [AB09], Chapter 14.4.4) is that any space s randomized algorithm in the Turing model can be simulated by a length n, width w ROBP with $n, w = 2^{O(s)}$. Thus, one approach to derandomize two-sided error space-bounded algorithms is to construct, in bounded space, a distribution of small support that "looks random" to any such ROBP. We say that a distribution $\mathcal{D}$ on n-bit strings is (n, w, ε)*-pseudorandom* if for every length n, width w ROBP, a path (string) that is sampled from $\mathcal{D}$ has, up to an additive error ε, the same probability to end in the accept state as a truly random path. A truly random path corresponds to a path picked uniformly at random from the 2^n possible paths. An (n, w, ε)-*pseudorandom generator* (PRG) is an algorithm ($\{0,1\}^l \to \{0,1\}^n$) that takes in l bits and outputs n bits such that the output distribution (uniform distribution over the range) is (n, w, ε)-pseudorandom. The seed length of a PRG is the number of random bits it requires to generate the distribution, that is, l. Informally, we call the PRG explicit, if each output bit can be computed efficiently given the input and the index, that is, in $O(\log n)$-space.

Derandomizing using an explicit pseudorandom distribution is straightforward. By iterating over all the paths in the support of the distribution and calculating the fraction of those paths that end in the accept state, one obtains an ε-approximation to the probability of reaching the accept state while taking a truly random path in the program. The support size being small (or, equivalently, the seed being short) allows one to perform such an iteration in bounded space.

One can prove the existence of an (n, w, ε)-PRG with seed length $O(\log(nw/\varepsilon))$. The proof is non-constructive and hence, the PRG isn't efficient. In his seminal paper, Nisan [Nis92] gave an explicit construction of a PRG with seed length $O(\log n \cdot \log(nw/\varepsilon))$. Setting $n, w = 2^{\Theta(s)}$ and ε to a small constant, the seed length is $O(s^2)$ which yields derandomization with quadratic overhead in space. Saks and Zhou [SZ99] applied Nisan's generator in a far more sophisticated way than the naïve derandomization so to obtain their result (see Section 1.4).

While pseudorandom distributions are suitable for derandomizing two-sided error randomized algorithms, hitting sets are suitable for one-sided error. An (n, w, ε)-*hitting set* is a set of n-bit strings such that for every length n, width w ROBP, whenever a truly random path ends in the accept state with probability at least ε, then there exists a path in the set that ends at the accept state. Hitting sets can be used to derandomize **RL** (and **coRL**). The best known hitting set for width $w > 3$ is in fact Nisan's PRG. In particular, the same seed length is required and $\mathbf{RL} \subseteq \mathbf{L}^{3/2}$ is the strongest known inclusion. Even for the deceptively simple looking problem of constructing hitting sets for width $w = 3$ ROBPs, no progress was made for nearly two decades, until the works of [ŠŽ11, GMR+12]. In particular, Gopalan *et al.* [GMR+12] construct near-optimal hitting sets in that setting.

There has been much success in constructing PRGs for restricted types of ROBPs (see, e.g., [INW94, NZ96, RTV06, BPW11, Ste12, BPW12, KNP11, KNP11, De11, IMZ12, GMR+12, GMRZ13, RSV13, SVW14, GV17] and references therein) such as *permutation* and, more generally, *regular* ROBPs [BRRY14, BV10]. These are programs in which every state but for start, accept and reject, has in-degree 2.

1.2 Pseudorandom pseudo-distributions for ROBPs

In this work, we obtain the first improved constructions of hitting sets for unrestricted ROBPs (for any width w) by constructing hitting sets with near-optimal dependence on ε (precisely, the seed length is $O\left(\log(w/\varepsilon)\log\log(w/\varepsilon) + \log^2(n) \cdot \log\left(\frac{\log(1/\varepsilon)}{\log n}\right) + \log n \cdot \log w\right)$). In fact, we introduce and construct a new type of primitive we call a *pseudorandom pseudo-*

distribution [1] that, informally speaking, lies between hitting sets and pseudorandom distributions. We find this notion to be of independent interest.

Definition 1.1 (Pseudorandom pseudo-distributions)**.** *Let* $\rho_1, \dots, \rho_{2^s} \in \mathbb{R}$ *and* $p_1, \dots, p_{2^s} \in \{0,1\}^n$*. The sequence* $\widetilde{\mathcal{D}} = ((\rho_1, p_1), \dots, (\rho_{2^s}, p_{2^s}))$ *is an* (n, w, ε)-pseudorandom pseudo-distribution *if for every length n, width w ROBP, the sum of all* ρ_i*'s for which the respective paths* p_i *end in the accept state is an* ε*-approximation to the probability of ending at the accept state by taking a truly random path in the program.*

We stress that Definition 1.1 allows the ρ_i's to take both positive and negative values. These values are not necessarily bounded by 1 in absolute value, nor by any constant for that matter, and they do not necessarily sum up to 1. Indeed, in our construction, it is possible that $|\rho_i| = \mathrm{poly}(nw/\varepsilon)$. Nevertheless, the definition requires that the numbers cancel out nicely so that summing the ρ_i's of the respective paths that arrive to the accept state yields an ε-approximation for the probability of arriving to the accept state by taking a truly random path (and, in particular, the sum is a number in $[-\varepsilon, 1+\varepsilon]$). An (n, w, ε)-*pseudorandom pseudo-generator* (PRPG) is an algorithm ($\{0,1\}^l \to \mathbb{R} \times \{0,1\}^n$) that takes in l bits and outputs a real number and an n-bit string such that the output distribution (sequence achieved by iterating over all l-bit inputs) is a (n, w, ε)-pseudorandom pseudo-distribution. Informally, we call the PRPG explicit, if each output bit (and the real number) can be computed efficiently given the input and the index, that is, in $O(\log n)$-space[2].

Pseudorandom pseudo-distributions yield hitting sets. Observe that, if one simply ignores the ρ_i's, and considers the set of paths $\{p_1, \dots, p_{2^s}\}$ in an (n, w, ε)-pseudorandom pseudo-distribution, one obtains an (n, w, ε')-hitting set for any $\varepsilon' > \varepsilon$. Indeed, consider a program in which the probability to reach the accept state is at least ε'. Then, the sum of ρ_i's which correspond to paths p_i ending in the accept state is at least $\varepsilon' - \varepsilon > 0$. Surely then, at least one path p_i ends in the accept state.

Pseudo-distributions are as good as distributions for derandomizing BPL. By the above, a pseudorandom pseudo-distribution suffices to derandomize one-sided error randomized algorithms. In fact, more is true. While $\widetilde{\mathcal{D}}$ is not a distribution per se, it is as good as such for the purpose of derandomizing *two-sided* error randomized algorithms, at least when using the naïve derandomization method described above. Indeed, the straightforward derandomization using a pseudorandom (proper) distribution, which sums the probability mass of the relevant paths, works just as well for pseudo-distributions as one can sum up the ρ_i's which, in some sense, generalize the probability mass. Of course, the space requirement now depends on $\sum_i |\rho_i|$ (also assuming there's an explicit PRPG corresponding to the pseudo-distribution).

[1] The term "pseudo-distribution" is used in different contexts to mean different things, all under the general idea that the object at hand shares some desired properties with a "proper" distribution. The closest research field in which the term pseudo-distributions is used (with a different meaning than ours) is Sum of Squares. However, we do not believe this will cause any confusion.

[2] Our construction is $\tilde{O}(\log^2(n) + \log(n)\log(w) + \log(1/\varepsilon))$-computable.

1.3 Main result

The main contribution of this work is an explicit construction (computable in $\widetilde{O}(\log^2 n + \log n \cdot \log w + \log(1/\varepsilon))$ space) of a pseudorandom pseudo-distribution with near-optimal dependence on ε. This, in particular, yields the first improved construction of hitting sets for unrestricted ROBPs.

Theorem 1.2 (Main result)**.** *For all integers $n, w \geq 1$ and $0 < \varepsilon < 1/n$, there exists an explicit (n, w, ε)-pseudorandom pseudo-distribution with seed length*

$$\widetilde{O}\left(\log(n)\log(nw) + \log(1/\varepsilon)\right).$$

In particular, for $w = n$ the seed length is $\widetilde{O}(\log^2 n + \log(1/\varepsilon))$.

See Theorem 4.3 for the full statement and a discussion on the explicitness of our construction. Consider, for simplicity, the setting where $w = n$. Further, for ease of discussion, ignore double-logarithmic factors. Recall that Nisan's generator has seed length $O(\log n \cdot \log(n/\varepsilon))$ whereas the optimal seed length is $O(\log(n/\varepsilon))$. That is, the problem is all about "shaving off" the redundant $\log n$ factor. In Theorem 1.2, we are able to shave off this factor from the $\log(1/\varepsilon)$ term and obtain near-optimal dependence on ε in the setting of pseudorandom pseudo-distributions (and, thus, for hitting sets). Whereas the result doesn't give any better derandomization of **BPL** or **RL**, it strictly improves upon prior works when $\log(1/\varepsilon) = \omega(\log n)$, a regime of parameters that is well-motivated by the work of Saks and Zhou [SZ99] as discussed in Section 1.4.

At a very high level, the underlying idea behind our construction is to work with a rough approximation for the probability of acceptance together with a sequence of finer and finer correction terms, which add up to yield the desired error guarantee. Generating and maintaining these correction terms require the flexibility of working with negative, unbounded, weights. In Section 2, we give a detailed overview of the proof of Theorem 1.2 in which we emphasize the main ideas and new techniques. We hope that our techniques can find further applications for constructing hitting sets and pseudorandom generators for other computational models.

1.4 Towards $\mathbf{BPL} \subseteq \mathbf{L}^{4/3}$

Recall that the seed length of Nisan's PRG is $O(\log n \cdot \log(nw/\varepsilon))$. In particular, even for constant width w and constant error guarantee ε, the seed length is $O(\log^2 n)$ which is the best known result even in this setting, and is perhaps the most identified aspect of Nisan's PRG. Nevertheless, in their seminal paper, Saks and Zhou [SZ99] showed how to apply Nisan's generator in a sophisticated way so as to prove $\mathbf{BPL} \subseteq \mathbf{L}^{3/2}$. In this section, we give a very high-level sketch of their idea and lay down a research program towards improving the exponent to 4/3. Theorem 1.2 accomplishes one step in our program. We stress that the description we give here is very sketchy and the reader is referred to [SZ99] for a formal treatment.

1.4.1 A sketch of Saks-Zhou's argument

It can be easily seen, also discussed in Section 2, that derandomizing space $O(s)$ randomized algorithms is equivalent to approximating the matrix M^{2^s} for a given $2^s \times 2^s$ stochastic matrix M. More generally, a $(2^r, 2^s, \varepsilon)$-PRG can be used to approximate M^{2^r} for a given $2^s \times 2^s$ stochastic matrix M [SZ99].

Write $s = r_1 r_2$ for r_1, r_2 integers to be chosen later on. A first attempt at approximating M^{2^s} is to start by computing N_1–an ε-approximation of $M^{2^{r_1}}$. Then, computing N_2 which is an approximation for $N_1^{2^{r_1}} \approx M^{2^{2r_1}}$ and so on for r_2 steps. Consider a $(2^{r_1}, 2^s, \varepsilon)$-PRG. It can be shown that most seeds to the PRG yield an ε-approximation for $M^{2^{r_1}}$. One can then find a "good" seed by iterating over all seeds and test each against the given matrix M. A good seed can then be stored in memory. What Saks and Zhou showed is that by making certain random perturbations to the approximating matrix N_1, one can break the correlation the matrix has with the seed that was used to compute it. Thus, the *same* seed can be used throughout all r_2 recursive levels.

In terms of parameters, one must set $\varepsilon = 2^{-s}$, and $O(s)$ fresh random bits are required for the perturbations done in each of the r_2 recursive levels. Thus, if we denote the seed length of the $(2^{r_1}, 2^s, 2^{-s})$-PRG by d, the total number of random bits used to approximate M^{2^s} is $O(d + r_2 s)$. By using Nisan's $(2^{r_1}, 2^s, 2^{-s})$-PRG, which has seed length $d = O(r_1^2 + r_1 s) = O(r_1 s)$, one needs a seed of length $O((r_1 + r_2)s)$ for approximating M^{2^s}. By setting $r_1 = r_2 = \sqrt{s}$, one obtains a randomized algorithm with seed length $O(s^{3/2})$ for approximating M^{2^s}. This algorithm can then be derandomized by iterating over all seeds and taking the average of the results.

1.4.2 On the seed length dependence on ε and w

In Theorem 1.2, we gave a construction of a pseudorandom pseudo-distribution with seed length $\widetilde{O}(\log(n)\log(nw) + \log(1/\varepsilon))$. For ease of readability we ignore the double-logarithmic factors which anyhow do not make a significant difference in this discussion. We now show that if one decouples the width w, on top of ε, from n, to get a PRG of seed length $O(\log^2 n + \log(w/\varepsilon))$, one can apply[3] the Saks-Zhou scheme to obtain a stronger derandomization of **BPL**. By plugging r_1, s, the seed length of the generator is $O(r_1^2 + s)$ and so the total seed length required by the Saks-Zhou scheme is of the order of $r_1^2 + s + r_2 s = r_1^2 + s + s^2/r_1$. One can then set $r_1 = s^{2/3}$ to get seed length $s^{4/3}$ and deduce $\mathbf{BPL} \subseteq \mathbf{L}^{4/3}$. In fact, decoupling ε and w from n, even at some cost, would yield some improvement in derandomizing **BPL**. In particular, a PRG with seed length $O(\log^2 n + \log^c(w/\varepsilon))$ would yield $\mathbf{BPL} \subseteq \mathbf{L}^{\max(4/3,c)}$.

To summarize, even without improving upon the dependence of the seed length on n, one can obtain improved derandomization by decoupling both w, ε from n in the seed length of the PRG. In Theorem 1.2, we obtained the desired improvement for ε in the setting of pseudorandom pseudo-distributions, and we leave the task of doing the same for ε, w

[3] Assuming the PRG can be modified to have certain similar properties as Nisan's generator; such as in [Arm98].

in the setting of PRGs to future research. We stress that, unlike the naïve method of derandomization, the Saks-Zhou scheme does not work as it is with pseudo-distributions.

2 Proof Overview

Unfortunately, our construction is fairly involved and the analysis requires a significant amount of work. To guide the reader through the formal proof, in this section we give an informal overview of our construction and its analysis. This section is not required for the sequel and can be skipped, though we believe the informal manner in which it is written and the discussions it contains are of value.

We start this section by presenting the well-known reduction from the problem of constructing PRGs for ROBPs to the problem of sparsifying matrix product. Then, in Section 2.2, we rederive Nisan's result via samplers rather than using hash functions as was done originally [Nis92], expander graphs [INW94], or seeded extractors [RR99]. While not improving upon previous works, in this section we present the notion of a sampler [BR94], which plays a key role in our construction, and show how it can be used for constructing PRGs. In Section 2.3, we introduce and motivate the idea of working with differences, or delta, of samplers. This discussion, even being very informal, should be helpful in guiding the reader through the following sections. In Section 2.4 we introduce the notion of a matrix-bundle sequence (MBS) and its smallness; define multiplication rules for MBSs in Section 2.5 and Section 2.6, and proceed from there to describe our construction and its analysis.

2.1 The reduction to sparsifying matrix product

It is folklore that the problem of constructing PRGs for ROBPs can be reduced to the problem of sparsifying matrix product or, more precisely, the product of matrices when represented in a certain way. To describe this reduction, consider a length n, width w ROBP. The transition between a pair of consecutive layers $\mathrm{P}_t, \mathrm{P}_{t+1}$ in the program can be represented as the average of two $w \times w$ zero-one matrices $M_t = (M_t^0 + M_t^1)/2$, where $(M_t^0)_{i,j} = 1$ if and only if the edge labeled by 0 that is going out of state i in layer t ends in state j of layer $t+1$. M_t^1 is similarly defined with respect to edges labeled by 1. Note that for every t, the matrix M_t is stochastic. Thus, M_t represents a single step from layer t to $t+1$ when the bit is uniformly distributed over $\{0,1\}$. $M_{start,accept}$ represents the acceptance probability when traversing a truly random path in P. In these terms, the goal is then to approximate the matrix product $M = M_1 M_2 \cdots M_n$ in bounded space. More precisely, given indices $i, j \in [w]$ as inputs and access to any entry of the matrices, one would like to compute an ε-approximation to $M_{i,j}$.

Slightly deviating from previous works, the most suitable measure of approximation for our construction is obtained by using the infinity norm. The infinity norm of a $w \times w$ matrix A, is defined by $\|A\|_\infty = \max_{i \in [w]} \sum_{j=1}^{w} |A_{i,j}|$. We say that two matrices A, B are ε-close, or that A ε-approximates B, if $\|A - B\|_\infty \le \varepsilon$. As with any norm, $\|\cdot\|_\infty$ is sub-additive, namely, $\|A + B\|_\infty \le \|A\|_\infty + \|B\|_\infty$. We make use of two further properties of the infinity norm. First, $\|\cdot\|_\infty$ is sub-multiplicative, namely, $\|AB\|_\infty \le \|A\|_\infty \|B\|_\infty$. Second, $\|A\|_\infty = 1$ for any stochastic matrix A.

Now, clearly, one can expand

$$M = 2^{-n} \prod_{t=1}^{n} (M_t^0 + M_t^1) = \mathop{\mathbf{E}}_{r \sim \{0,1\}^n} \prod_{t=1}^{n} M_t^{r_t}.$$

Note that $M_t^{r_t}(i,j)$ is 1 if there is an edge from i^{th} vertex of layer P_t to j^{th} vertex of layer P_{t+1} labelled r_t. Therefore, the RHS can be thought of as taking all paths in the ROBP, namely, one for each choice of $r \in \{0,1\}^n$. A productive point of view for the construction of PRGs for ROBPs is that of sparsifying the above product such that M is ε-approximated by $\mathbf{E}_{r \sim H} \prod_{t=1}^{n} M_t^{r_t}$, ending up with a small set of paths $H \subseteq \{0,1\}^n$ to average on.

Firstly, we introduce some notation and give intuition for a recursive sparsification process. Let $\mathrm{A} = (A_1, \ldots, A_{2^s})$ be a sequence of $w \times w$ stochastic matrices. From here on, all matrices in this section are of order $w \times w$. The matrix that is *realized* by A is given by $\langle \mathrm{A} \rangle = \mathbf{E}_i [A_i]$. Similarly, let $\mathrm{B} = (B_1, \ldots, B_{2^s})$ and $\langle \mathrm{B} \rangle = \mathbf{E}_j [B_j]$. Assume that $\langle \mathrm{A} \rangle$ ε_A-approximates some matrix of interest $\widetilde{A}$ and $\langle \mathrm{B} \rangle$ ε_B-approximates $\widetilde{B}$ (such that $\|\widetilde{A}\|_\infty, \|\widetilde{B}\|_\infty \le 1$). We think of s as the complexity of the "representation" and would like to keep it small. If one wishes to approximate the product $\widetilde{A}\widetilde{B}$, the natural approach would be to consider the product of approximations $\langle \mathrm{A} \rangle \langle \mathrm{B} \rangle = \mathbf{E}_{i,j \sim [2^s]} A_i B_j$. Indeed, using the properties of $\|\cdot\|_\infty$, we have that

$$\begin{aligned}
\|\widetilde{A}\widetilde{B} - \langle \mathrm{A} \rangle \langle \mathrm{B} \rangle\|_\infty &= \|\widetilde{A}\widetilde{B} - \langle \mathrm{A} \rangle \widetilde{B} + \langle \mathrm{A} \rangle \widetilde{B} - \langle \mathrm{A} \rangle \langle \mathrm{B} \rangle\|_\infty \\
&\le \|\widetilde{A}\widetilde{B} - \langle \mathrm{A} \rangle \widetilde{B}\|_\infty + \|\langle \mathrm{A} \rangle \widetilde{B} - \langle \mathrm{A} \rangle \langle \mathrm{B} \rangle\|_\infty \\
&\le \|\widetilde{A} - \langle \mathrm{A} \rangle\|_\infty \|\widetilde{B}\|_\infty + \|\langle \mathrm{A} \rangle\|_\infty \|\widetilde{B} - \langle \mathrm{B} \rangle\|_\infty \\
&\le \|\widetilde{A} - \langle \mathrm{A} \rangle\|_\infty \cdot 1 + 1 \cdot \|\widetilde{B} - \langle \mathrm{B} \rangle\|_\infty \\
&\le \varepsilon_A + \varepsilon_B. \qquad (2.1)
\end{aligned}$$

Thus, taking the product of the approximations $\langle A \rangle, \langle B \rangle$ yields a very good approximation guarantee. However, taking this product is costly in terms of representation as it doubles the complexity of the representation from s to $2s$, that is, the expectation is over 2^{2s} terms. To reduce the number of terms, we want to sparsify the product of the two matrix representations.

This approach was taken by many previous works, either implicitly or explicitly using hash functions [Nis92, Nis94], expander graphs [INW94, RV05], and seeded extractors [NZ96, RR99, BRRY14, Arm98]. We are going to describe such derandomization based on *samplers*. Besides being a natural perspective, we work with samplers because, for our improved construction, we require flexibility that we only know how to obtain using samplers (see Section 2.5.1 for details). Interestingly enough, though, the constructions of the samplers we make use of are based on expander graphs and seeded extractors. In the next section we rederive Nisan's result [Nis92] via samplers. We do so mainly for preparing the ground for our improved construction that follows.

2.2 Deriving Nisan's result via samplers

In this section, we show, on a high level, that a product of n $w \times w$ stochastic matrices can be approximated to an error of ε by a set of $2^{O(\log^2(n)+\log(n)\log(w/\varepsilon))}$ many products (with each matrix being 0-1 stochastic) using the method of sparsification. This also gives Nisan's PRG through observations made in the last section. Informally, a *sampler* is a randomized algorithm that, with high probability over its randomness, yields a good approximation for the expectation of any bounded function by querying the latter on a small number of points. A sampler has two parameters: the query complexity that determines how many queries are required by the sampler, and its randomness complexity, which is the number of truly random bits required for the sampling. An *averaging sampler* is a special type of sampler where the randomness is only used to select the points on which to query the function, independently of the function being considered. Only then the function is queried, and the output is the average of the corresponding values.

In the following definition, and throughout the paper, we use the graph-theoretic perspective of averaging samplers and use the term sampler instead of an averaging sampler. More on samplers can be found in the excellent survey by Goldreich [Gol11] and in Vadhan's excellent monograph [Vad11].

Definition 2.1 (Samplers [BR94])**.** *A left-regular bipartite graph $G = (L, R, E)$ is an (ε, δ)-sampler if for every function $f : R \to [0, 1]$, for all but δ-fraction of vertices $v \in L$ it holds that*

$$\left| \mathop{\mathbf{E}}_{i\sim\Gamma(v)} [f(i)] - \mathop{\mathbf{E}}_{i\sim R} [f(i)] \right| \le \varepsilon.$$

Here $\Gamma(v)$ is the set of neighbors of v in G. The left-degree of G is called the degree *of the sampler.*

Observe that given a graph G as in Definition 2.1, the randomized algorithm that performs the sampling process simply uses its randomness to select a vertex $v \in L$ uniformly at random, and then outputs the average $\mathbf{E}_{i\sim\Gamma(v)} f(i)$.

Now that samplers have been defined, we show how they can be used to sparsify matrix product or, more precisely, the product of the representations of the respective matrices. Let $\mathrm{A} = (A_1, \ldots, A_{2^s})$, $\mathrm{B} = (B_1, \ldots, B_{2^s})$ be as before. Recall, $\forall i$ A_i, B_i are $w \times w$ stochastic matrices. Given a left-regular bipartite graph $G = ([2^s], [2^s], E)$ with degree 2^d, define the sequence

$$\mathrm{A} \circ_G \mathrm{B} = \mathrm{C} = (C_{i,j})_{i\in[2^s], j\in[2^d]}$$

as follows: for $i \in [2^s]$ and $j \in [2^d]$, $C_{i,j} = A_i B_{\Gamma(i,j)}$, where $\Gamma(i, j)$ denotes the j'th neighbor of vertex i in G. Note that $C_{i,j}$ are all stochastic. In particular, they are 0-1 if A_i and B_j are 0-1 (that is, all the entries are either 0 or 1). We now prove

Lemma 2.2. *Let* A *and* B *be sequence of $w\times w$ stochastic matrices as defined above. Let G be as above and $0 < \varepsilon, \delta < 1$. If G is an (ε, δ)-sampler then $\|\langle \mathrm{A} \circ_G \mathrm{B}\rangle - \langle \mathrm{A}\rangle\langle \mathrm{B}\rangle\|_\infty \le w^2(\varepsilon + \delta)$.*

Proof. Note that

$$\langle \mathrm{C} \rangle = \mathop{\mathbf{E}}_{i \sim [2^s], j \sim [2^d]} [\mathrm{C}_{i,j}] = \mathop{\mathbf{E}}_{i \sim [2^s]} \left[A_i \mathop{\mathbf{E}}_{j \sim \Gamma(i)} B_j \right].$$

Therefore, for every fixed $\alpha, \beta \in [w]$,

$$\langle \mathrm{C} \rangle_{\alpha,\beta} = \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i \sim [2^s]} \left[(A_i)_{\alpha,\gamma} \mathop{\mathbf{E}}_{j \sim \Gamma(i)} (B_j)_{\gamma,\beta} \right].$$

For a fixed $\gamma \in [w]$, consider the function $f_{\gamma,\beta} \colon [2^s] \to [0,1]$ that is given by $f_{\gamma,\beta}(j) = (B_j)_{\gamma,\beta}$. Note that the range of $f_{\gamma,\beta}$ is indeed $[0,1]$ as B_j are all stochastic matrices. Define

$$\varepsilon_{\gamma,\beta}(i) = \mathop{\mathbf{E}}_{j \sim \Gamma(i)} [f_{\gamma,\beta}(j)] - \langle \mathrm{B} \rangle_{\gamma,\beta}.$$

Informally, as $\langle \mathrm{B} \rangle_{\gamma,\beta} = \mathbf{E}_{j \sim [2^s]} [f_{\gamma,\beta}(j)]$, the quantity $\varepsilon_{\gamma,\beta}(i)$ measures the quality of the approximation for the function $f_{\gamma,\beta}$ from the point of view of vertex i, that is, when the points are sampled using the neighborhood of i. We have that

$$\begin{aligned}
\langle \mathrm{C} \rangle_{\alpha,\beta} &= \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i \sim [2^s]} [(A_i)_{\alpha,\gamma} (\langle \mathrm{B} \rangle_{\gamma,\beta} + \varepsilon_{\gamma,\beta}(i))] \\
&= \sum_{\gamma=1}^{w} \langle \mathrm{A} \rangle_{\alpha,\gamma} \langle \mathrm{B} \rangle_{\gamma,\beta} + \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i \sim [2^s]} [(A_i)_{\alpha,\gamma} \varepsilon_{\gamma,\beta}(i)] \\
&= (\langle \mathrm{A} \rangle \langle \mathrm{B} \rangle)_{\alpha,\beta} + \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i \sim [2^s]} [(A_i)_{\alpha,\gamma} \varepsilon_{\gamma,\beta}(i)].
\end{aligned}$$

As A_i are all stochastic, for every $i \in [2^s]$ we have that

$$\begin{aligned}
|\langle \mathrm{C} \rangle_{\alpha,\beta} - (\langle \mathrm{A} \rangle \langle \mathrm{B} \rangle)_{\alpha,\beta}| &\le \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i \sim [2^s]} [|(A_i)_{\alpha,\gamma}| \, |\varepsilon_{\gamma,\beta}(i)|] \\
&\le \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i \sim [2^s]} |\varepsilon_{\gamma,\beta}(i)|.
\end{aligned}$$

As G is an (ε, δ)-sampler, for all γ, for all but δ-fraction of $i \in [2^s]$, it holds that $|\varepsilon_{\gamma,\beta} s(i)| \le \varepsilon$ and so

$$|\langle \mathrm{C} \rangle_{\alpha,\beta} - (\langle \mathrm{A} \rangle \langle \mathrm{B} \rangle)_{\alpha,\beta}| \le w(\varepsilon + \delta).$$

The lemma follows as the above bound holds for every α, β. ☐

Equation (2.1) and Lemma 2.2 readily imply that if $\langle \mathrm{A} \rangle$ is an ε_A-approximation for some matrix $\widetilde{A}$ of interest and $\langle \mathrm{B} \rangle$ ε_B-approximates $\widetilde{B}$ then

$$\begin{aligned}
\|\langle \mathrm{A} \circ_G \mathrm{B} \rangle - \widetilde{A}\widetilde{B}\|_\infty &\le \|\langle \mathrm{A} \circ_G \mathrm{B} \rangle - \langle \mathrm{A} \rangle \langle \mathrm{B} \rangle\|_\infty + \|\langle \mathrm{A} \rangle \langle \mathrm{B} \rangle - \widetilde{A}\widetilde{B}\|_\infty \\
&\le \varepsilon_A + \varepsilon_B + w^2(\varepsilon + \delta). \qquad (2.2)
\end{aligned}$$

Thus, one pays an additional error of $w^2(\varepsilon+\delta)$ in the resulting approximation, compared to taking the actual product, when using the sparsified product parameterized by the (ε,δ)-sampler G. The advantage, however, is that now the expectation is over way less that 2^{2s} terms as indeed $\mathrm{A} \circ_G \mathrm{B}$ is a sequence of length 2^{s+d} (rather than 2^{2s}), where 2^d is the degree of the sampler.

It is now a question of how the degree of a sampler relates to the parameters ε, δ. It turns out that, based on expander graphs and, in particular, Ramanujan graphs, one can construct an (ε,δ)-sampler with degree $O(\varepsilon^{-2}\delta^{-1})$ [GW97]. As ε,δ play the same role in the bound that was derived in Lemma 2.2 and the degree has roughly the same dependence on ε,δ (none of which is the case in our improved construction as discussed in Section 2.5) we set $\varepsilon=\delta$ and consider a sampler of degree $O(\delta^{-3})$ to illustrate Nisan's construction.

2.2.1 Going from 2 to n matrices

To approximate the product of n stochastic matrices $(M_1, M_2, ..., M_n)$, one can apply recursion. Let $M_{[i,j]}$ represent the product of matrices $M_i, M_{i+1}, ...M_j$. Let A be an approximation of $M_{[1,2^r]}$ and B be an approximation of $M_{[2^r+1,2^{2r}]}$, then $\langle \mathrm{A} \circ_G \mathrm{B}\rangle$ gives an approximation of $M_{[1,2^{2r}]}$. We iterate the process for $\log(n)$ levels to get an approximation of product of n matrices. If we denote the approximation guarantee for multiplying 2^r matrices by $\varepsilon(r)$ then Equation (2.2) yields the recursive relation $\varepsilon(r) = 2\varepsilon(r-1) + 2w^2\delta$, and so $\varepsilon(r) = O(2^r w^2 \delta)$. Further, if one denotes by $s(r)$ the complexity of the representation at level r, that is the expectation is over $2^{s(r)}$ terms, one has $s(r) = s(r-1) + O(\log(1/\delta))$, yielding $s(r) = O(r\log(1/\delta))$. If ε' is the approximation guarantee one is aiming for, one must set $\delta = O(2^{-r}\varepsilon'/w^2)$ which yields complexity $s(r) = O(r^2 + r\log(w/\varepsilon'))$. Plugging $r = \log n$, the depth of the recursion, we rederive Nisan's result, namely, the seed length of the respective PRG, is $O(\log n \cdot \log(nw/\varepsilon))$. The explicitness of the PRG (computability in at least $O(\log n \cdot \log(nw/\varepsilon))$ space) follows from the log-space computability (log in the size of the bipartite graph of the sampler) of the neighbourhood function of the sampler.

We remark that by using the samplers that are constructed via expander graphs, the construction above is in fact exactly the one introduced in [INW94], though the analysis is conceptually different. Building on the notations and ideas presented so far, in the following section we significantly deviate from existing ideas and start to describe our improved construction.

Before proceeding further, we observe that from the way in which one sparsifies matrix product, it is possible to obtain a description of the pseudorandom distribution or, equivalently, the PRG. Thus, throughout the paper we only consider sparsifying matrix products and do not explicitly define the induced pseudorandom pseudo-distribution for that matter. We find this point of view far more suitable for our construction and its analysis. We point out that the construction stated space complexity follows from the space complexity guaranteed by the samplers that we use. We elaborate more on the corresponding PRPG in Section 9.2.

2.3 Delta of samplers–a preliminary discussion

By inspecting the construction from the previous section, one can see that the reason the seed length ended up being $O(\log n \cdot \log(nw/\varepsilon))$ is that we had to set δ so low so as to guarantee that the accumulation of errors from all n products will not exceed ε. The main conceptual novelty of our construction is in working with differences, or delta, of samplers. We motivate this reasoning in the following informal discussion.

Assume, as before, that $\mathrm{A} = (A_1, \ldots, A_{2^s})$ and $\mathrm{B} = (B_1, \ldots, B_{2^s})$ are sequences such that $\langle \mathrm{A} \rangle, \langle \mathrm{B} \rangle$ are ε-approximations for some matrices of interest $\widetilde{A}, \widetilde{B}$, respectively. For an integer d, let $G_d = ([2^s], [2^s], E_d)$ be an (ε, δ)-sampler set with $\varepsilon = \delta = 2^{-d}$. Recall that the degree of G_d is $2^{O(d)}$. In the previous section, we used an expensive choice of $d = O(\log(wn/\varepsilon)) \triangleq k$. Instead, let's try to "break down" the matrix that is realized by this expensive product by suggestively writing $\langle \mathrm{A} \circ_{G_k} \mathrm{B} \rangle$ as

$$\begin{aligned}\langle \mathrm{A} \circ_{G_k} \mathrm{B} \rangle = &\langle \mathrm{A} \circ_{G_g} \mathrm{B} \rangle + \\ &\langle \mathrm{A} \circ_{G_{2g}} \mathrm{B} \rangle - \langle \mathrm{A} \circ_{G_g} \mathrm{B} \rangle + \\ &\langle \mathrm{A} \circ_{G_{4g}} \mathrm{B} \rangle - \langle \mathrm{A} \circ_{G_{2g}} \mathrm{B} \rangle + \\ &\quad\vdots \\ &\langle \mathrm{A} \circ_{G_k} \mathrm{B} \rangle - \langle \mathrm{A} \circ_{G_{k/2}} \mathrm{B} \rangle,\end{aligned} \tag{2.3}$$

where $g \ll k$ is some parameter such that k/g is conveniently a power of two. Consider now a summand in this telescopic sum, say, $\langle \mathrm{A} \circ_{G_{2g}} \mathrm{B} \rangle - \langle \mathrm{A} \circ_{G_g} \mathrm{B} \rangle$. We are going to define a new multiplication rule between matrix representations (that doesn't approximate the product), which for now we denote by $\circ_{G_{2g} - G_g}$ [4], that has the following three properties:

Property 1 (Linearity). First, our product is linear with respect to the samplers by which it is parameterized, namely,

$$\langle \mathrm{A} \circ_{G_{2g} - G_g} \mathrm{B} \rangle = \langle \mathrm{A} \circ_{G_{2g}} \mathrm{B} \rangle - \langle \mathrm{A} \circ_{G_g} \mathrm{B} \rangle.$$

That is, the matrix that is realized by the new product gives the desired difference.

Property 2 (Smallness is stored). The resulted object, $\mathrm{A} \circ_{G_{2g} - G_g} \mathrm{B}$, has "smallness" g and, more generally, for integers $D > d$, $\mathrm{A} \circ_{G_D - G_d} \mathrm{B}$ has smallness d such that: if one considers the product

$$(\mathrm{A} \circ_{G_D - G_d} \mathrm{B}) \circ_{G_{D'} - G_{d'}} \mathrm{C}$$

for some matrix representation C, the smallness of the product is $d + d'$. That is, smallness is being stored in the matrix representation and then added back when taking future products. In fact, the product will also inherit the smallness of the right operand. That is,

$$(\mathrm{A} \circ_{G_D - G_d} \mathrm{B}) \circ_{G_{D'} - G_{d'}} \left(\mathrm{C} \circ_{G_{D''} - G_{d''}} \mathrm{D}\right)$$

[4] Use of delta of samplers in the approximation is the reason we get a pseudorandom pseudo-distribution instead of a PRG.

has smallness $d + d' + d''$ [5], and so forth.

Property 3 (Smallness implies small norm). If A has smallness s then $\|\mathrm{A}\|_\infty \le 2^{-\Omega(s)}$.

Using the new product rule, an instructive way of thinking of Equation (2.3) is by rewriting it as

$$\begin{aligned}\langle \mathrm{A} \circ_{G_k} \mathrm{B}\rangle = &\langle \mathrm{A} \circ_{G_g} \mathrm{B}\rangle + \\ &\langle \mathrm{A} \circ_{G_{2g}-G_g} \mathrm{B}\rangle + \\ &\langle \mathrm{A} \circ_{G_{4g}-G_{2g}} \mathrm{B}\rangle + \\ &\qquad\vdots \\ &\langle \mathrm{A} \circ_{G_k - G_{k/2}} \mathrm{B}\rangle, \end{aligned} \tag{2.4}$$

and thinking of $\mathrm{A} \circ_{G_g} \mathrm{B}$ as a "rough approximation" of the product we care about (rough since $g \ll k$), which have 0 smallness. The object $\mathrm{A} \circ_{G_{2g}-G_g} \mathrm{B}$ is the first "correction term" having smallness g, $\mathrm{A} \circ_{G_{4g}-G_{2g}} \mathrm{B}$ the second correction term having smallness $2g$, and so forth.

In general, for $D > d$, the representation $\mathrm{A} \circ_{G_D - G_d} \mathrm{B}$ is going to "cost" D addition to the number of terms in the "expectation" and have smallness d. Setting $D = 2d$, as we did in Equation (2.3), guarantees that in some intuitive sense, up to a constant factor, what is being paid for is invested. Thus, as you increase the number of terms in the sparsification of the product, the terms become smaller and this investment by using a sampler does not go to waste as it is somehow stored as smallness in the object. And, matrices with large smallness can be discarded without much affect on the total error. Thus, intuitively, we get a $\widetilde{O}(\log(1/\varepsilon))$ dependence as any more investment makes the terms small enough to be discarded in the process of recursive sparsification.

2.4 How to "store" smallness

In the construction presented in Section 2.2, a matrix was represented by a "one dimensional" sequence $\mathrm{A} = (A_1, \ldots, A_{2^s})$ of $w \times w$ stochastic matrices, and the matrix that was represented, or realized, by this representation was defined by $\langle \mathrm{A}\rangle = \mathbf{E}_i[A_i]$. In order to "store" smallness, we first need to devise a more subtle representation of matrices. This will require a fair amount of preparation, and such representation is given in Section 2.7. To begin, in this section we define the notions of matrix bundles, matrix bundle sequences, and smallness.

Definition 2.3 (Matrix bundles)**.** *For an integer $\ell \ge 0$, an* ℓ-matrix bundle A *is a sequence*

$$\mathrm{A} = ((\alpha_1, A_1), \ldots, (\alpha_{2^\ell}, A_{2^\ell})),$$

where the α_i's are real numbers (that are not necessarily bounded, and can take both positive and negative values) and the A_i's are $w \times w$ stochastic matrices[6]*. The matrix that is* realized

[5] We don't achieve exactly the sum but with the right parameters, the smallness is effectively this.

[6] For the purpose of derandomizing ROBPs or randomized log-space, all A_i's are 0-1 matrices.

by A *is defined by* $\langle \mathrm{A} \rangle = \sum_{i=1}^{2^\ell} \alpha_i A_i$. *We extend any matrix norm* $\|\cdot\|$ *to matrix bundles by letting* $\|\mathrm{A}\| = \|\langle \mathrm{A} \rangle\|$. *We refer to the numbers* $\alpha_1, \dots, \alpha_{2^\ell}$ *as the* coefficients *of* A.

Definition 2.4 (Matrix bundle sequences (MBSs))**.** *Let* $d_{\mathsf{out}}, d_{\mathsf{in}} \geq 0$ *be integers. A* $(d_{\mathsf{out}}, d_{\mathsf{in}})$-matrix bundle sequence (MBS) $\mathcal{A}$ *is a sequence of* $2^{d_{\mathsf{out}}}$ *number of* d_{in}*-matrix bundles* $\mathcal{A} = (\mathrm{A}_1, \dots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$. *The matrix that is* realized by $\mathcal{A}$ *is defined by* $\langle \mathcal{A} \rangle = \mathbf{E}_{i \sim [2^{d_{\mathsf{out}}}]} \langle \mathrm{A}_i \rangle$. *We extend any matrix norm* $\|\cdot\|$ *to MBSs by letting* $\|\mathcal{A}\| = \|\langle \mathcal{A} \rangle\|$. *We refer to the union of the coefficients of* $\mathrm{A}_1, \dots, \mathrm{A}_{2^{d_{\mathsf{out}}}}$ *as the* coefficients of $\mathcal{A}$.

A matrix bundle sequence is not going to be the final representation of a matrix in our construction but rather it will be used to represent a "piece" of the matrix with some smallness, alluded to in the above discussion as a correction term or the first rough approximation term. Before presenting the final representation, we need to understand MBSs better. We start by giving the formal definition of "smallness", which we already informally discussed above. In the following section, we define multiplication rules for MBSs and show their interplay with smallness.

Definition 2.5 (Smallness)**.** *Let* $\mathcal{A} = (\mathrm{A}_1, \dots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$ *be a* $(d_{\mathsf{out}}, d_{\mathsf{in}})$*-MBS. The smallness of* $\mathcal{A}$*, denoted by* $\sigma(\mathcal{A})$*, is defined by*

$$\sigma(\mathcal{A}) = -\log_2 \mathop{\mathbf{E}}_{i \sim [2^{d_{\mathsf{out}}}]} \|\mathrm{A}_i\|_\infty^2.$$

It is straightforward to show that if $\sigma(\mathcal{A}) \leq s$ then $\|\mathcal{A}\|_\infty \leq 2^{s/2}$ (see Claim 5.6). Thus, if an MBS has a sufficiently large smallness, it can be discarded with low cost in error.

2.5 Multiplication rules for MBSs

In Section 2.2, we defined the multiplication rule $\circ_G$ between "one-dimensional" sequence of matrices. We now turn to define a multiplication rule for MBSs. In fact, we are going to introduce two types of multiplication rules which we refer to as outer-multiplication and inner-multiplication (for the actual construction, we need to consider four multiplication rules as we need to worry about the order in which we multiply matrices. However, in this informal proof overview, we allow ourselves to be somewhat informal regarding this point). We define these multiplication rules to ensure that the smallness is stored while keeping the number of the matrices in the representation in check. The outer-multiplication is an extension of the multiplication rule used in Section 2.2 whereas the inner-multiplication is carefully engineered to work with smallness. In the next section, we describe how the multiplication rule is defined when parameterized by a delta of samplers.

For the description of both multiplication rules, let $\mathcal{A}$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}))$-MBS and $\mathcal{B}$ a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}))$-MBS. Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be a left-regular bipartite graph with left-degree 2^d. Note that G may be unbalanced. Indeed, the flexibility of working with samplers for which $d_{\mathsf{out}}(\mathcal{A}) \gg d_{\mathsf{out}}(\mathcal{B})$ is pivotal for our construction.

The outer-multiplication denoted by $\mathcal{A} \circ_G \mathcal{B}$, is the $(d_{\mathsf{out}}(\mathcal{A}) + d, d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}))$-MBS $\mathcal{C} = (\mathrm{C}_{i,j})_{i \in [2^{d_{\mathsf{out}}(\mathcal{A})}], j \in [2^d]}$ that is defined as follows. For every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$ and $j \in [2^d]$, $\mathrm{C}_{i,j}$ is the $(d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}))$-matrix bundle that is obtained by taking all products of matrices, and of the respective coefficients, from the matrix bundles A_i and $\mathrm{B}_{\Gamma(i,j)}$ (the formal definition is given in Definition 6.1). Note that for every i, j,

$$\langle \mathrm{C}_{i,j} \rangle = \langle \mathrm{A}_i \rangle \langle \mathrm{B}_{\Gamma(i,j)} \rangle.$$

The inner-multiplication denoted by $\mathcal{A} \bullet_G \mathcal{B}$ is a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + d)$-MBS $\mathcal{C} = (\mathrm{C}_i)_{i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]}$, where C_i is the $(d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + d)$-matrix bundle that is obtained by taking the product of all matrices in the matrix bundle A_i with all the matrices in all of the matrix bundles in $\{\mathrm{B}_j \mid j \in \Gamma(i)\}$, where the respective coefficients are multiplied accordingly and then divided by 2^d to yield

$$\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \langle \mathrm{B}_j \rangle.$$

The formal definition is given in Definition 6.8. Note that when applying the outer-multiplication, we pay the degree of the sampler in d_{out}, whereas the inner-multiplication increases d_{in} by the degree. The fact that we need to normalize by 2^{-d} is one reason we need the flexibility of maintaining arbitrary coefficients in the definition of matrix bundles.

By adapting the proof of Lemma 2.2, we can prove that both the inner and outer multiplication rules, when parameterized by a good sampler, approximate the product.

Lemma 2.6. ***[Idealized]*** *Let $0 < \varepsilon, \delta < 1$. Let $\mathcal{A}$ and $\mathcal{B}$ be MBSs as defined above. If G is an (ε, δ)-sampler then*

$$\|\langle \mathcal{A} \circ_G \mathcal{B} \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle\|_\infty \le w^2(\varepsilon + \delta),$$
$$\|\langle \mathcal{A} \bullet_G \mathcal{B} \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle\|_\infty \le w^2(\varepsilon + \delta).$$

For a formal statement and its proof see Lemma 6.13. The key property of the multiplication rule $\bullet_G$ is that it preserves the smallness of both MBSs it operates on, when parameterized with a good enough sampler G. The following lemma is an idealized version of an assertion we can actually make (see Lemma 6.14 for the formal statement).

Lemma 2.7. ***[Idealized]*** *Let $\mathcal{A}$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}))$-MBS and let $\mathcal{B}$ be a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}))$-MBS. Let $0 < \varepsilon, \delta < 1$. Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be an (ε, δ)-sampler with*

$$\begin{aligned} \varepsilon &\le 2^{-\sigma(\mathcal{B})}, \\ \delta &\le 2^{-\sigma(\mathcal{A}) - \sigma(\mathcal{B})}. \end{aligned} \tag{2.5}$$

Then, $\sigma(\mathcal{A} \bullet_G \mathcal{B}) \ge \sigma(\mathcal{A}) + \sigma(\mathcal{B})$.

We prove a weaker variant of Lemma 2.7 in Section 2.5.2. Before, in Section 2.5.1, we give some remarks on the asymmetry between the roles that ε and δ play in the lemma, and discuss unbalanced samplers.

2.5.1 Unbalanced samplers and the asymmetry between ε and δ

Lemma 2.7 states that the smallnesses of $\mathcal{A}, \mathcal{B}$ are completely preserved, or "stored" in $\mathcal{A} \bullet_G \mathcal{B}$, as long as the sampler G has good enough parameters. Note the asymmetry between ε and δ. Indeed, while δ is required to be taken exponentially small in the *sum* $\sigma(\mathcal{A}) + \sigma(\mathcal{B})$, ε only needs to be exponentially small in $\sigma(\mathcal{B})$. This may allow for a significant saving in cases where $\sigma(\mathcal{A}) \gg \sigma(\mathcal{B})$. However, the sampler used above has degree $\mathrm{poly}(1/\varepsilon, 1/\delta)$ and thus cannot exploit this saving. In fact, if one considers only balanced samplers, namely, samplers $G = (L, R, E)$ with $|L| = |R|$, then a polynomial dependence of the degree on $1/\varepsilon$ and $1/\delta$ is necessary. We are therefore led to consider unbalanced samplers.

As it turns out, unbalanced samplers are equivalent to seeded extractors [Zuc97], and the state of the art construction of unbalanced samplers is obtained by seeded extractors. In particular, for all integers ℓ, r and $0 < \delta < 1$ such that $\ell \geq r/\delta^2$ there exists an explicit (ε, δ)-sampler $G = ([\ell], [r], E)$ with degree $\mathrm{poly}(1/\varepsilon, \log(1/\delta))$ (see Theorem 3.10). That is, if the ratio between the sides of the sampler is large enough, the degree of the sampler has an exponentially better dependence on δ than what can be obtained by using balanced samplers. Thus, roughly speaking, by working with unbalanced samplers, Lemma 2.7 tells us that we gain the sum of smallnesses $\sigma(\mathcal{A}) + \sigma(\mathcal{B})$ by paying roughly $\min(\sigma(\mathcal{A}), \sigma(\mathcal{B}))$ in the degree.

2.5.2 Proof of a weaker version of Lemma 2.7

Next, we give a proof for a weaker version of Lemma 2.7. We give the proof for a relaxed setting in which the matrix bundles $\mathrm{A}_i, \mathrm{B}_j$ that compose $\mathcal{A}, \mathcal{B}$ are of bounded norm, in particular $\|\mathrm{A}_i\|_\infty$ and $\|\mathrm{B}_j\|_\infty$ are all bounded by 1. Moreover, we will not prove a bound as strong as stated above for the smallness of $\sigma(\mathcal{A} \bullet_G \mathcal{B})$. Instead, we prove that $\sigma(\mathcal{A} \bullet_G \mathcal{B}) \geq \sigma(\mathcal{A}) + \sigma(\mathcal{B}) - 2$. In fact, even in the formal proof we cannot give a bound of $\sigma(\mathcal{A}) + \sigma(\mathcal{B})$ though it will be crucial to give a bound of the form $\sigma(\mathcal{A}) + \sigma(\mathcal{B}) - \tau$ for some suitable slowly growing function $\tau = o(1)$.

Proof of Lemma 2.7. Write $\mathcal{C} = \mathcal{A} \bullet_G \mathcal{B} = (\mathrm{C}_i)_{i=1}^{2^{d_{\mathsf{out}}(\mathcal{A})}}$. For $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, define

$$\varepsilon(i) = \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \|\mathrm{B}_j\|_\infty^2 - 2^{-\sigma(\mathcal{B})}.$$

As G is an (ε, δ)-sampler, and since we assume that for all $j \in [2^{d_{\mathsf{out}}(\mathcal{B})}]$, $\|\mathrm{B}_j\|_\infty \leq 1$, there exists a set $S \subseteq [2^{d_{\mathsf{out}}(\mathcal{A})}]$ of size $|S| \geq (1 - \delta) 2^{d_{\mathsf{out}}(\mathcal{A})}$ such that for every $i \in S$, $|\varepsilon(i)| \leq \varepsilon$. Recall that for every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$,

$$\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \langle \mathrm{B}_j \rangle.$$

By Jensen's inequality and since $\|\cdot\|_\infty$ is sub-multiplicative (and sub-additive),

$$\begin{aligned}2^{-\sigma(\mathcal{C})} &= \mathop{\mathbf{E}}_{i} \|\mathrm{C}_i\|_\infty^2 \\ &= \mathop{\mathbf{E}}_{i} \|\langle \mathrm{A}_i\rangle \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \langle \mathrm{B}_j\rangle\|_\infty^2 \\ &\le \mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \|\mathrm{B}_j\|_\infty^2\right].\end{aligned}$$

Thus,

$$\begin{aligned}2^{-\sigma(\mathcal{C})} &\le \mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 (2^{-\sigma(\mathcal{B})} + \varepsilon(i))\right] \\ &= 2^{-\sigma(\mathcal{A})-\sigma(\mathcal{B})} + \mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i)\right]. \end{aligned} \tag{2.6}$$

As we assume $\|\mathrm{A}_i\|_\infty^2 \le 1$ and since $|\varepsilon(i)| \le 1$ for all $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$,

$$\begin{aligned}\mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i)\right] &\le \mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i) \;\middle|\; i \in S\right] + \mathbf{Pr}[i \notin S] \\ &\le \varepsilon \cdot \mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \;\middle|\; i \in S\right] + \delta.\end{aligned}$$

Since we might as well assume $\delta \le 1/2$, we have that $\mathbf{Pr}[i \in S] \ge 1 - \delta \ge 1/2$, and so

$$\mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \;\middle|\; i \in S\right] \le \frac{\mathbf{E}_i \left[\|\mathrm{A}_i\|_\infty^2\right]}{\mathbf{Pr}[i \in S]} \le 2^{-\sigma(\mathcal{A})+1}.$$

Hence, $\mathbf{E}_i \left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i)\right] \le 2\varepsilon \cdot 2^{-\sigma(\mathcal{A})} + \delta$. Plugging this to Equation (2.6), we get

$$2^{-\sigma(\mathcal{C})} \le 2^{-\sigma(\mathcal{A})-\sigma(\mathcal{B})} + 2\varepsilon \cdot 2^{-\sigma(\mathcal{A})} + \delta.$$

Substituting for ε, δ, we conclude that $\sigma(\mathcal{C}) \ge \sigma(\mathcal{A}) + \sigma(\mathcal{B}) - 2$, as desired. □

2.6 Multiplication parameterized by a delta of samplers

Now that MBSs and the two multiplication rules are in place, we are ready to define a multiplication rule that is parameterized by a delta of samplers. Assume, as in the previous section, that $\mathcal{A}$ is a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}))$-MBS and $\mathcal{B}$ is a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}))$-MBS. Let $D > d$ be integers. Let $G_D = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_D)$ be a left-regular bipartite graph with left-degree 2^D and $G_d = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_d)$ a left-regular bipartite graph with left-degree 2^d.

Write $\mathcal{A} \bullet_{G_D} \mathcal{B} = \mathcal{C}^+ = (\mathrm{C}_i^+)_{i\in[2^{d_{\mathsf{out}}(\mathcal{A})}]}$ and $\mathcal{A} \bullet_{G_d} \mathcal{B} = \mathcal{C}^- = (\mathrm{C}_i^-)_{i\in[2^{d_{\mathsf{out}}(\mathcal{A})}]}$. We define $\mathcal{A} \bullet_{G_D - G_d} \mathcal{B}$ to be the sequence $(\mathrm{C}_i)_{i\in[2^{d_{\mathsf{out}}(\mathcal{A})}]}$ where C_i is the concatenation of the matrix bundle C_i^+ with $-\mathrm{C}_i^-$, where by the leading minus sign, we mean that one negates all coefficients in C_i^-. The formal definition is given in Definition 6.16. It is easy to see that

$$\langle \mathcal{A} \bullet_{G_D - G_d} \mathcal{B}\rangle = \langle \mathcal{A} \bullet_{G_D} \mathcal{B}\rangle - \langle \mathcal{A} \bullet_{G_d} \mathcal{B}\rangle,$$

a property that we refer to as the linearity of $\bullet$. Further, note that $2^{d_{\mathsf{in}}(\mathcal{C})} = 2^{d_{\mathsf{in}}(\mathcal{A})+d_{\mathsf{in}}(\mathcal{B})}(2^D + 2^d)$. Thus, as $D \geq d$, we have $d_{\mathsf{in}}(\mathcal{C}) \leq d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + D + 1$. We remark that the relaxation of using negative numbers in the definition of pseudo-distributions is required so as to allow taking delta of samplers.

The smallness of $\mathcal{A} \bullet_{G_D - G_d} \mathcal{B}$ is analyzed in the following lemma, which, again, is an idealized version of an assertion we can actually make (see Lemma 6.18).

Lemma 2.8. ***[Idealized]*** *Let $\mathcal{A}, \mathcal{B}$ be MBSs as above. Let $G_1 = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_1)$ be an $(\varepsilon_1, \delta_1)$-sampler and $G_2 = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_2)$ an $(\varepsilon_2, \delta_2)$-sampler. Assume that $0 < \varepsilon_1 \leq \varepsilon_2 < 1$ and $0 < \delta_1 \leq \delta_2 < 1$. Then,*

$$\sigma(\mathcal{A} \bullet_{G_1 - G_2} \mathcal{B}) \geq \min\left(\log(1/\varepsilon_2) + \sigma(\mathcal{A}),\, \log(1/\delta_2)\right).$$

Lemma 2.8 states that the smallness of the product grows with the parameters of the weaker $(\varepsilon_2, \delta_2)$-sampler. As in Lemma 2.7, the parameter ε_2, which is exponentially more "expensive" than δ_2 in terms of degree (at least for unbalanced samplers) is being added to $\sigma(\mathcal{A})$ and so can be set much larger than δ_2. Unlike Lemma 2.7, $\sigma(\mathcal{A} \bullet_{G_1 - G_2} \mathcal{B})$ can grow beyond $\sigma(\mathcal{A}) + \sigma(\mathcal{B})$ if one takes a pair of good enough samplers. That is, the smallness of the product is not bounded by the sum of smallnesses of the operands.

2.7 Matrix representations

We are finally ready to give a high level description of how a matrix is being represented by our construction and how to multiply two such matrix representations.

Definition 2.9. *Let $1 \leq k$ be an integer. A k-*matrix representation *is a sequence $\mathbf{A} = (\mathcal{A}_0, \ldots, \mathcal{A}_k)$ where $\mathcal{A}_i$ is an MBS with $\sigma(\mathcal{A}_i) \geq i$. The matrix that is* realized *by $\mathbf{A}$ is defined by $\langle \mathbf{A} \rangle = \sum_{i=0}^{k} \langle \mathcal{A}_i \rangle$.*

Informally, one should think of 2^{-k} as the desired error guarantee. We think of $\mathcal{A}_0$ as a rough approximation of the matrix of interest $\widetilde{A}$. Let $1 \leq g \ll k$ be an integer such that that the approximation is 2^{-g} rather than the desired 2^{-k}, that is, $\|\langle \mathcal{A}_0 \rangle - \widetilde{A}\|_\infty \leq 2^{-g}$. The remaining MBSs are the finer and finer correction terms. Adding them improves the approximation up to the point that $\|\langle \mathbf{A} \rangle - \widetilde{A}\|_\infty \leq 2^{-k}$. For the formal construction, we will need to weight the different MBSs and these weights, which we ignore in this high-level description, is why we allow the ρ_i's in a pseudo-distribution to be unbounded (see Section 7).

We would like to define a multiplication rule between matrix representations that approximates the respective matrices. Assume that $\mathbf{A} = (\mathcal{A}_0, \ldots, \mathcal{A}_k)$ and $\mathbf{B} = (\mathcal{B}_0, \ldots, \mathcal{B}_k)$ are two matrix representations. We are going to define a multiplication rule $\cdot$ for matrix representations such that the matrix that is realized by the product $\mathbf{A} \cdot \mathbf{B}$ is an $2^{-\Omega(k)}$-approximation for $\langle \mathbf{A} \rangle \langle \mathbf{B} \rangle$. To describe our product, we start by writing

$$\langle \mathbf{A} \rangle \langle \mathbf{B} \rangle = \left(\sum_{i=0}^{k} \langle \mathcal{A}_i \rangle\right)\left(\sum_{j=0}^{k} \langle \mathcal{B}_j \rangle\right) = \sum_{i,j=0}^{k} \langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle.$$

Consider an expensive sampler, say an (ε, δ)-sampler G_k with $\varepsilon = \delta = 2^{-k}$. By Idealized Lemma 2.7, for every i, j, we can $2^{-\Omega(k)}$-approximate $\langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle$ by $\langle \mathcal{A}_i \bullet_{G_k} \mathcal{B}_j \rangle$. Doing so, and adding the errors from all $O(k^2)$ pairs (i, j), we get a total error of $2^{-\Omega(k)} k^2 = 2^{-\Omega(k)}$. However, we do not want to pay for an expensive sampler in "one shot". Instead, for every pair of $i, j \in \{0, 1, \ldots, k\}$, consider the sequence of MBSs

$$\begin{aligned} &\mathcal{A}_i \bullet_{G_d} \mathcal{B}_j \\ &\mathcal{A}_i \bullet_{G_{2d} - G_d} \mathcal{B}_j \\ &\mathcal{A}_i \bullet_{G_{4d} - G_{2d}} \mathcal{B}_j \\ &\quad \vdots \\ &\mathcal{A}_i \bullet_{G_k - G_{k/2}} \mathcal{B}_j, \end{aligned} \tag{2.7}$$

where the choice of d, and whether to use a balanced or an unbalanced sampler depends on i, j and will be discussed later. Moreover, for some pairs i, j, we will use the outer-multiplication rule in some of the MBSs in the list. By Idealized Lemma 2.7, $\sigma(\mathcal{A}_i \bullet_{G_d} \mathcal{B}_j) \geq i + j$ when d is taken sufficiently large. Further, by Idealized Lemma 2.8, $\sigma(\mathcal{A}_i \bullet_{G_{2d} - G_d} \mathcal{B}_j) \geq i + j + d$ (the parameters are chosen such that the smallness is effectively this) and generally $\sigma(\mathcal{A}_i \bullet_{G_{2^{r+1}d} - G_{2^r d}} \mathcal{B}_j) \geq i + j + 2^r d$. That is, each MBS in the list has a certain smallness we know how to bound from below.

Consider the collection of all MBSs obtained by considering the MBSs in Equation (2.7) for all $i, j \in \{0, \ldots, k\}$. We denote this set of MBSs by $\mathcal{F}(\mathbf{A}, \mathbf{B})$ (see Definition 8.1). To obtain the matrix representation $\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = (\mathcal{C}_0, \ldots, \mathcal{C}_k)$, we collect MBSs from $\mathcal{F}(\mathbf{A}, \mathbf{B})$ with a common smallness s (or, more precisely, MBSs for which the best lower bound we have on their smallness is s) and "glue" them to form the MBS $\mathcal{C}_s$. We discard MBSs that have smallness larger than k. We glue MBSs by concatenating the two sequence of matrix bundles and factor the coefficients accordingly to yield an MBS with a slightly larger d_{out} (see Section 5.3).

2.8 Leveled matrix representations and setting of parameters

In this section, we give further information regarding the multiplication rule between matrix representation discussed in the previous section. In particular, we left out details about how to set d as a function of i, j, and whether the multiplication is parameterized by a balanced or an unbalanced sampler. The way we set things is as follows. Let $\mathbf{A} = (\mathcal{A}_0, \mathcal{A}_g, \ldots, \mathcal{A}_k)$ be a k-matrix representation, where k would be chosen later on. We maintain the invariant that there are no MBSs but for $\mathcal{A}_0$ with smallness less than g in $\mathbf{A}$. $g << k$ would be chosen later too. We partition the latter sequence to *levels*. The first MBS, $\mathcal{A}_0$ is in level 0. MBSs with smallness $[g, 2g)$ are in level 1; MBSs with smallness $[2g, 4g)$ are in level 2, and so forth. In fact, we are also required to maintain the invariant that all smallnesses are multiples of g. We do the same for a second k-matrix representation $\mathbf{B} = (\mathcal{B}_0, \mathcal{B}_g, \ldots, \mathcal{B}_k)$. For a formal treatment, see the definition of leveled matrix representation given in Section 7.

Consider now any $i, j > 0$. If $\mathcal{A}_i, \mathcal{B}_j$ belong to the same level (implying that $i/2 \leq j \leq 2i$) we use the inner-multiplication rule to multiply $\mathcal{A}_i, \mathcal{B}_j$ using balanced samplers. If i, j belong to different levels, we use unbalanced samplers instead. In all such cases we are going to set $d = O(\min(i,j))$. Handling $i = 0$ or $j = 0$ is done similarly, using unbalanced samplers, but using the outer-multiplication rule for the first MBS in Equation (2.7). In such cases, d is set to $O(g)$.

Every stochastic matrix in our construction corresponds to a path in our pseudo-distribution. As every MBS $\mathcal{A}_i$ in $\mathbf{A}$ consists of $2^{d_{\mathsf{in}}(\mathcal{A}_i)+d_{\mathsf{out}}(\mathcal{A}_i)}$ such matrices, the total number of paths is $\sum_{i=0}^{k} 2^{d_{\mathsf{in}}(\mathcal{A}_i)+d_{\mathsf{out}}(\mathcal{A}_i)}$. As $d_{\mathsf{in}}, d_{\mathsf{out}}$ are increasing functions of i, the seed length is dominated by $d_{\mathsf{in}}(\mathcal{A}_k) + d_{\mathsf{out}}(\mathcal{A}_k)$. We turn to analyze each of $d_{\mathsf{out}}, d_{\mathsf{in}}$.

Analyzing d_{out}. Our unbalanced samplers are all set with $\delta = 2^{-\Omega(k)}$ and so we are required to maintain the invariant that the d_{out} of MBSs increases in "jumps" of $\Omega(k)$ across levels. As the number of levels is logarithmic in k, this requires d_{out} to be as large as $k \log k$ for MBSs with smallness k. The fact that we set $d = g$ when using the outer-multiplication rule with $i = 0$ or $j = 0$ causes d_{out} to further increase by g in every recursive level. As described before, we multiply n matrices recursively, for example, we multiply the first $n/2$ matrices and the last $n/2$ matrices separately and then multiply the outputs to get the product of n matrices. As we have $\log n$ recursive levels, the bound that we get on the maximum d_{out} is $O(k \log k + g \log n)$.

Analyzing d_{in}. Using the interleaved use of balanced and unbalanced samplers, we are able to maintain the invariant $d_{\mathsf{in}}(\mathcal{A}_i) = O(i \log i)$ throughout the recursion, independently of the level of the recursion. In particular, d_{in} of all MBSs is bounded by $O(k \log k)$ and is thus dominated by d_{out}. To give some idea of why such bound is obtained, note that for every i, j, the first MBS in Equation (2.7) has smallness $i + j$ and for that we pay $\min(i,j)$ in d_{in}. For the remaining MBSs, paying $\min(i,j)$ in d_{in} credits one with a proportional smallness. Solving for the respective recursive relation gives the stated bound.

Setting k, g. So far, while we paid for choosing a large value of g in d_{out}, the role of g in the analysis was not explained. Without getting into the technical details, the finer-grained error analysis that we conduct, guarantees that at recursive-level t, the total error is bounded above by

$$\varepsilon(t) = w \cdot (k/g)^{kt/g} \cdot 2^{-k},$$

and so we set $g \approx \log n \cdot \log k$ to yield $\varepsilon(\log n) = w \cdot 2^{-\Omega(k)}$ and then $k = \Omega(\log(w/\varepsilon))$ to guarantee total error ε. For simplicity, set $w = n$. In such case, $k = O(\log(n/\varepsilon))$ and $g = O(\log n \cdot \log\log(n/\varepsilon))$. Plugging this to our bound on d_{out}, we get seed length of $O(k \log k + g \log n) = \widetilde{O}(\log^2 n + \log(1/\varepsilon))$. To obtain our result, which note is slightly stronger, we make a more careful setting of parameters.

3 Preliminaries

All logarithms in this paper are of base 2. For ease of readability, we avoid the use of floor and ceiling. This does not affect the stated results. For an integer $n \geq 1$ we use U_n to denote the uniform distribution over n-bit strings. Let b be a boolean expression. We define the indicator $\mathbf{1}_b$ to be 1 if b holds and 0 otherwise. For an integer $n \geq 1$ we let $[n] = \{1, 2, \ldots, n\}$. Let $A \subseteq B$ be finite sets. We denote by $\mu_B(A)$ the density of A within B, namely, $\mu_B(A) = |A|/|B|$. Typically, B will be clear from context, in which case we write $\mu(A)$.

Let $G = (L, R, E)$ be a bipartite graph. We say G is left-regular if all nodes in L have the same degree. If G is left-regular with left-degree d and edges labeled by $\{1, \ldots, d\}$, we define the neighborhood function $\Gamma_G \colon L \times [d] \to R$ to be such that the i'th neighbor of node $v \in L$ is given by $\Gamma_G(v, i)$. We denote the set of neighbors of v by $\Gamma_G(v)$. If G is clear from context we sometimes omit it from the subscript and simply write $\Gamma(v, i)$ and $\Gamma(v)$ for $\Gamma_G(v, i)$ and $\Gamma_G(v)$, respectively.

3.1 Read-once branching programs, hitting sets, and pseudorandom distributions

In this section we recall basic definitions related to read-once branching programs. Definition 3.1 below is slightly different from the informal definition that was used in the introduction, though the two definitions can be easily shown to be equivalent.

Definition 3.1. *Let $n, w \geq 1$ be integers. An (n, w)-read-once branching program (ROBP for short) P is a directed graph on the vertex set $V = \{s\} \cup \bigcup_{i=1}^{n} \mathrm{P}_i$, where the P_i's are disjoint sets of size w each. We refer to P_i as* layer i *of the program P. From every node but for those that belong to P_n there are two outgoing edges, labeled by 0 and 1. The pair of edges from s ends in P_1 and for every $1 \leq i < n$ and $v \in \mathrm{P}_i$, the pair of edges going out of v end in nodes that belong to P_{i+1}. There are no edges leaving P_n. The node s is called the start node of the program P.*

Given a string $p \in \{0, 1\}^{\ell}$, with $\ell \leq n$, we denote by $\mathrm{P}(p)$ the node that is reached by traversing the ROBP P according to the path p starting at the start node. The set of all (w, n)-ROBPs is denoted by $\mathcal{P}_{w,n}$.

Definition 3.2 (Hitting sets)**.** *A set $\{p_1, \ldots, p_{2^s}\} \subseteq \{0, 1\}^n$ is an (n, w, ε)-*hitting set *if for every $\mathrm{P} \in \mathcal{P}_{w,n}$ and node $v \in \mathrm{P}_n$ for which $\mathbf{Pr}[\mathrm{P}(U_n) = v] \geq \varepsilon$, there exists $j \in [2^s]$ such that $\mathrm{P}(p_j) = v$.*

It is sometimes convenient to address the function that generates the hitting set.

Definition 3.3 (Hitting set generators)**.** *A function $\mathsf{HSG} \colon \{0, 1\}^s \to \{0, 1\}^n$ is an (n, w, ε)-*hitting set generator *(HSG for short) if the image of HSG is an (n, w, ε)-hitting set. We refer to the input of HSG as the seed. Note that 2^s is an upper bound on the size of the hitting set.*

Definition 3.4 (Pseudorandom distributions)**.** *A distribution $\mathcal{D}$ over n-bit string is an* (n, w, ε)-pseudorandom distribution *if for every* $\mathrm{P} \in \mathcal{P}_{w,n}$ *and* $v \in \mathrm{P}_n$*,*

$$\left| \mathbf{Pr}[\mathrm{P}(U_n) = v] - \mathbf{Pr}[\mathrm{P}(\mathcal{D}) = v] \right| \le \varepsilon.$$

Clearly, the support of every (n, w, ε)-pseudorandom distribution is an (n, w, ε')-hitting set for any $\varepsilon' > \varepsilon$. As with hitting sets, it is sometimes convenient to address the function that generates the pseudorandom distribution.

Definition 3.5 (Pseudorandom generators)**.** *A function* $\mathsf{PRG}\colon \{0,1\}^s \to \{0,1\}^n$ *is an* (n, w, ε)-pseudorandom generator *(PRG for short) if the distribution* $\mathsf{PRG}(U_s)$ *is* (n, w, ε)*-pseudorandom. We refer to the input of* PRG *as the seed.*

3.2 Matrix norms

Throughout the paper, we make use of two matrix norms. Let A be a $w \times w$ real matrix. Recall that the *infinity norm* of A is defined by $\|A\|_\infty = \max_{i \in [w]} \sum_{j=1}^{w} |A_{i,j}|$. The *max norm* of A is given by $\|A\|_{\max} = \max_{i,j \in [w]} |A_{i,j}|$. We denote the set of $w \times w$ stochastic matrices by $\mathbf{S}_w$. We make use of the following well-known, easy to verify, facts:

Claim 3.6. *Let A, B be $w \times w$ real matrices. Then,*

- *The norm $\| \cdot \|_\infty$ is sub multiplicative, namely, $\|AB\|_\infty \le \|A\|_\infty \|B\|_\infty$.*
- *Both norms (by definition) are sub-additive, that is, $\|A + B\|_\infty \le \|A\|_\infty + \|B\|_\infty$ and $\|A + B\|_{\max} \le \|A\|_{\max} + \|B\|_{\max}$.*
- $\|A\|_{\max} \le \|A\|_\infty \le w\|A\|_{\max}$.
- *If $A \in \mathbf{S}_w$ then $\|A\|_\infty = 1$.*

3.3 Samplers

Definition 3.7 ([BR94])**.** *Let $0 < \varepsilon, \delta < 1$. A left-regular bipartite graph $G = (L, R, E)$ is an* (ε, δ)-sampler *if for every function $f\colon R \to [0,1]$, for all but δ-fraction of vertices $v \in L$ it holds that*

$$\left| \mathop{\mathbf{E}}_{i \sim \Gamma(v)} [f(i)] - \mathop{\mathbf{E}}_{i \sim R} [f(i)] \right| \le \varepsilon.$$

The left-degree of G is called the degree *of the sampler.*

In many cases, the range of the function f, whose expectation we want to approximate, is not bounded to $[0,1]$. We thus use the following easy claim.

Claim 3.8. *Let $m_1, m_2 \ge 0$ be real numbers. Let $G = (L, R, E)$ be an (ε, δ)-sampler and $f\colon R \to [-m_1, m_2]$. Then, for all but δ-fraction of vertices $v \in L$,*

$$\left| \mathop{\mathbf{E}}_{i \sim \Gamma(v)} [f(i)] - \mathop{\mathbf{E}}_{i \sim R} [f(i)] \right| \le \varepsilon(m_1 + m_2).$$

For our construction of pseudorandom pseudo-distributions, we make use of two constructions of samplers. The first has equal sides, namely, $|L| = |R|$ whereas the second sampler has better parameters, albeit, it requires $|L| \gg |R|$. We refer to the first one, informally, as a *balanced sampler* and to the second one as an *unbalanced sampler*. The constructions of these samplers rely on expander graphs and seeded extractors, respectively, and we refer the reader to the excellent survey by Goldreich [Gol11] for more information.

Theorem 3.9 ([GW97])**.** *For every integer n and all $\varepsilon, \delta > 0$, there exists an (ε, δ)-sampler $\mathsf{BSamp}(n, \varepsilon, \delta) = (L, R, E)$, with $|L| = |R| = n$, having degree $d = O(\delta^{-1}\varepsilon^{-2})$.*

Theorem 3.10 ([RVW01], Corollary 7.3 [7])**.** *There exists a universal constant $c \geq 1$ such that the following holds. For all $\varepsilon, \delta > 0$ such that $\log(1/\delta) > \log(1/\varepsilon)c^{\log^*(1/\delta)}$ and for all integers ℓ, r such that $\ell \geq r/\delta^2$ there exists an (ε, δ)-sampler $\mathsf{UBSamp}(\ell, r, \varepsilon, \delta) = ([\ell], [r], E)$ with degree $d = ((1/\varepsilon)\log(1/\delta))^c$.*

It can be shown that both samplers are log-space computable, namely, given $i \in L$ and $j \in [d]$, the j'th neighbor of vertex i can be computed in $O(\log |L|)$ space (and in time $\text{poly}\log |L|$). This assertion is well-known for the sampler that is given by Theorem 3.9, whose construction is based on expander graphs, as was used in [INW94]. The assertion with respect to Theorem 3.10 is only implicit in the literature. The assertion can be shown to hold because the samplers are obtained by composing expander graphs, hash functions and k-wise independent distributions in simple ways (simple to compute, not to analyze).

Working with the parameters of the sampler given by Theorem 3.10 is cumbersome. Thus, for the sake of readability, we make use of the following sampler which has parameters that are easier to work with. We stress that this sampler in *not* space-efficient. It is easy to verify that our result holds as is when using the space-efficient sampler that is given by Theorem 3.10. Indeed, the seed length of our construction only deteriorates by a factor of $2^{O(\log^*(nw/\varepsilon))}$ which is then hidden under the $\widetilde{O}$-notation. Further, the space complexity is linear in the seed length.

Theorem 3.11 ([Zuc07])**.** *There exists a universal constant $c_{\mathsf{samp}} \geq 1$ such that the following holds. For all integers ℓ, r and all $\varepsilon, \delta > 0$ for which $\ell \geq r/\delta^2$, there exists an (ε, δ)-sampler $\mathsf{UBSamp}(\ell, r, \varepsilon, \delta) = ([\ell], [r], E)$ with degree $d = ((1/\varepsilon) \cdot \log(1/\delta))^{c_{\mathsf{samp}}}$.*

From here on, we suppress the size of the samplers n, ℓ, r and simply write $\mathsf{BSamp}(\varepsilon, \delta)$ for the sampler that is given by Theorem 3.9 and $\mathsf{UBSamp}(\varepsilon, \delta)$ for the sampler from Theorem 3.11.

4 Pseudorandom Pseudo-Distributions and Main Result

In this section we introduce the notion of a pseudorandom pseudo-distribution.

[7] We note that there are several versions of the cited paper. The conference and journal versions do not contain the results we need, and so we cite the version posted on ECCC.

Definition 4.1 (Pseudorandom pseudo-distributions)**.** *Let* $\rho_1, \ldots, \rho_{2^s} \in \mathbb{R}$ *and* $p_1, \ldots, p_{2^s} \in \{0,1\}^n$*. The sequence* $\widetilde{\mathcal{D}} = ((\rho_1, p_1), \ldots, (\rho_{2^s}, p_{2^s}))$ *is an* (n, w, ε)-pseudorandom pseudo-distribution *if for every* $\mathrm{P} \in \mathcal{P}_{w,n}$ *and* $v \in \mathrm{P}_n$*,*

$$\left| \mathbf{Pr}[\mathrm{P}(U_n) = v] - \sum_{i=1}^{2^s} \rho_i \mathbf{1}_{\mathrm{P}(p_i)=v} \right| \leq \varepsilon.$$

For a real number $b \geq 0$*, we say that* $\widetilde{\mathcal{D}}$ *is b-bounded if* $|\rho_i| \leq b$ *for all* $i \in [2^s]$*.*

We refer to s as the seed length of the pseudorandom pseudo-distribution.

We observe that pseudo-distributions readily yield hitting sets.

Claim 4.2. *Let* $((\rho_1, p_1), \ldots, (\rho_{2^s}, p_{2^s}))$ *be an* (n, w, ε)*-pseudorandom pseudo-distribution. Then, for every* $\varepsilon' > \varepsilon$*,* $p_1, \ldots, p_{2^s}$ *is an* (n, w, ε')*-hitting set.*

Proof. Let $\varepsilon' > \varepsilon$ be a real number. Let $\mathrm{P} \in \mathcal{P}_{w,n}$ and consider $v \in \mathrm{P}_n$ for which $\mathbf{Pr}[\mathrm{P}(U_n) = v] \geq \varepsilon'$. We have that

$$\sum_{i=1}^{2^s} \rho_i \mathbf{1}_{\mathrm{P}(p_i)=v} \geq \mathbf{Pr}[\mathrm{P}(U_n) = v] - \varepsilon \geq \varepsilon' - \varepsilon > 0$$

which readily implies the existence of $g \in [2^s]$ such that $\mathrm{P}(p_g) = v$. □

We are now ready to give a formal statement of our main result.

Theorem 4.3 (Main result)**.** *For every integers* $n, w \geq 1$ *and* $0 < \varepsilon < 1/n$*, there exists an* (n, w, ε)*-pseudorandom pseudo-distribution* $\widetilde{\mathcal{D}}$ *with seed length*

$$d = \widetilde{O}(\log(n)\log(nw) + \log(1/\varepsilon)).$$

Furthermore, $\widetilde{\mathcal{D}}$ *is* $\mathrm{poly}(w/\varepsilon)$*-bounded, and can be computed in space* $\widetilde{O}(d)$*.*

Here, by computability in space-s, we mean that the pseudorandom pseudo-distribution $\widetilde{\mathcal{D}}$ is generated by a pseudorandom pseudo-generator (PRPG) which is computable in space-s, that is given the seed $j \in [2^d]$ and the index $i \in [n]$, the real number ρ_j and i^{th} bit of the path p_j can be computed in $O(s)$ space.

Remark regarding explicitness. Note that in our proof of Theorem 4.3, we use the unbalanced sampler that is given by Theorem 3.11, whose parameters are easy to work with, though its space-complexity is high. By plugging-in, instead, the space-efficient sampler that is given by Theorem 3.10, one can easily show that the seed length and space complexity are as stated. Indeed, the seed length of our construction only deteriorates by a factor of $2^{O(\log^*(nw/\varepsilon))}$ when using the space-efficient sampler from Theorem 3.10. This small loss is anyhow hidden under the $\widetilde{O}$-notation. We choose to omit the cumbersome details as this complicates the already involved proof.

5 Matrix Bundle Sequences

In this section, we introduce the notion of a *matrix bundle sequence* (MBS for short). Informally speaking, an MBS is a "piece" of a matrix that we are interested in. To represent a matrix we make use of several MBSs. An MBS has a property we call *smallness* that, informally, captures how small the piece is. This is somewhat analogous to the digits of a number when represented in a decimal expansion, where the location of the digit are analogous to its smallness. We start by defining *matrix bundles*.

5.1 Matrix bundles

Definition 5.1. *Let $\ell \geq 0$, $w \geq 1$ be integers. An (ℓ, w)-*matrix bundle A *is an element of $(\mathbb{R} \times \mathbf{S}_w)^{2^\ell}$. Namely, $\mathrm{A} = ((\alpha_1, A_1), \ldots, (\alpha_{2^\ell}, A_{2^\ell}))$, where the α_i's are real numbers (that are unbounded and can be negative) and the A_i's are $w \times w$ stochastic matrices* [8]*. The matrix that is* realized by A *is defined by $\langle \mathrm{A} \rangle = \sum_{i=1}^{2^\ell} \alpha_i A_i$. We extend any matrix norm $\|\cdot\|$ to matrix bundles by letting $\|\mathrm{A}\| = \|\langle \mathrm{A} \rangle\|$. We refer to the numbers $\alpha_1, \ldots, \alpha_{2^\ell}$ as the* coefficients *of* A.

Next, we define the product of a scalar by a matrix bundle.

Definition 5.2. *For a real number β and an (ℓ, w)-matrix bundle $\mathrm{A} = ((\alpha_1, A_1), \ldots, (\alpha_{2^\ell}, A_{2^\ell}))$, we define $\beta \cdot \mathrm{A}$ to be the (ℓ, w)-matrix bundle $((\beta\alpha_1, A_1), \ldots, (\beta\alpha_{2^\ell}, A_{2^\ell}))$. We sometimes write $\beta\mathrm{A}$ instead of $\beta \cdot \mathrm{A}$. Note that $\langle \beta\mathrm{A} \rangle = \beta \langle \mathrm{A} \rangle$.*

5.2 Matrix bundle sequences

Definition 5.3. *Let $d_{\mathsf{out}}, d_{\mathsf{in}} \geq 0$ and $w \geq 1$ be integers. A $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-*matrix bundle sequence (MBS) $\mathcal{A}$ *is a sequence of $2^{d_{\mathsf{out}}}$ number of (d_{in}, w)-matrix bundles $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$. The matrix that is* realized by $\mathcal{A}$ *is defined by $\langle \mathcal{A} \rangle = \mathbf{E}_{i \sim [2^{d_{\mathsf{out}}}]} \langle \mathrm{A}_i \rangle$. We extend any matrix norm $\|\cdot\|$ to MBSs by letting $\|\mathcal{A}\| = \|\langle \mathcal{A} \rangle\|$. We refer to the union of the coefficients of $\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}}}$ as the* coefficients of $\mathcal{A}$.

Definition 5.4. *An MBS $\mathcal{A}$ is called* thin *if $d_{\mathsf{in}}(\mathcal{A}) = 0$ and all coefficients of $\mathcal{A}$ equal 1.*

Definition 5.5. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$ be a $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS. The smallness of $\mathcal{A}$, denoted by $\sigma(\mathcal{A})$, is defined by*

$$\sigma(\mathcal{A}) = -\log \mathop{\mathbf{E}}_{i \sim [2^{d_{\mathsf{out}}}]} \|\mathrm{A}_i\|_\infty^2,$$

where recall that all logarithms in this paper are to the base 2. The magnitude of $\mathcal{A}$, denoted by $\mu(\mathcal{A})$, is defined by

$$\mu(\mathcal{A}) = \log \max_{i \in [2^{d_{\mathsf{out}}}]} \|\mathrm{A}_i\|_\infty^2.$$

[8]For the purpose of derandomizing ROBPs, think of each A_i as a matrix corresponding to a single path (of some length $\leq n$) and is thus, a 0-1 matrix.

Claim 5.6. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$ be a $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS. Then, $\|\mathcal{A}\|_\infty \leq 2^{-\sigma(\mathcal{A})/2}$.*

Proof. By the sub-additivity of $\|\cdot\|_\infty$,

$$\|\mathcal{A}\|_\infty = \|\langle\mathcal{A}\rangle\|_\infty = \|\mathop{\mathbf{E}}_i \langle \mathrm{A}_i\rangle\|_\infty \leq \mathop{\mathbf{E}}_i \|\mathrm{A}_i\|_\infty.$$

By Jensen's inequality,

$$\left(\mathop{\mathbf{E}}_i \|\mathrm{A}_i\|_\infty\right)^2 \leq \mathop{\mathbf{E}}_i \|\mathrm{A}_i\|_\infty^2 = 2^{-\sigma(\mathcal{A})},$$

and so $\|\mathcal{A}\|_\infty^2 \leq 2^{-\sigma(\mathcal{A})}$, which completes the proof. □

Remarks regarding the monotonicity of $d_{\mathsf{in}}, d_{\mathsf{out}}$. Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$ be a $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS. For any $d'_{\mathsf{in}} \geq d_{\mathsf{in}}$, one can consider the $(d_{\mathsf{out}}, d'_{\mathsf{in}}, w)$-MBS $\mathcal{A}' = (\mathrm{A}'_1, \ldots, \mathrm{A}'_{2^{d_{\mathsf{out}}}})$ that is obtained by extending each of the (d_{in}, w)-matrix bundles A_i to a (d'_{in}, w)-matrix bundle A'_i by appending $2^{d'_{\mathsf{in}} - d_{\mathsf{in}}}$ zero coefficients and arbitrary stochastic matrices. Note that $\langle \mathrm{A}_i \rangle = \langle \mathrm{A}'_i \rangle$ and so this operation has no affect on the parameters of $\mathcal{A}$ other than d_{in}, and in particular, $\sigma(\mathcal{A}') = \sigma(\mathcal{A})$ and $\mu(\mathcal{A}') = \mu(\mathcal{A})$. Therefore, using this padding argument, one can think of every $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS as an $(d_{\mathsf{out}}, d'_{\mathsf{in}}, w)$-MBS with the same parameters for any $d'_{\mathsf{in}} \geq d_{\mathsf{in}}$.

Note that the same argument holds even if d_{in} is not an integer (this happens when we concatenate the matrix bundles of two MBSs with $(d_{\mathsf{in}})_1 \neq (d_{\mathsf{in}})_2$, resulting in $2^{d_{\mathsf{in}}} = 2^{(d_{\mathsf{in}})_1} + 2^{(d_{\mathsf{in}})_2}$, which indeed is not a power of 2). In particular, we implicitly always round d_{in} up to an integer by using this padding argument.

Similarly, one can consider $\mathcal{A}$ to be a $(d'_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS for any $d'_{\mathsf{out}} \geq d_{\mathsf{out}}$. This is because one can take the MBS $\mathcal{A}''$ with A_i duplicated $2^{d'_{\mathsf{out}} - d_{\mathsf{out}}}$ times to form a sequence of length $2^{d'_{\mathsf{out}}}$. Clearly, $d_{\mathsf{out}}(\mathcal{A}'') = d'_{\mathsf{out}}$ and $\langle\mathcal{A}\rangle = \langle\mathcal{A}''\rangle$. Note that this transformation has no effect on d_{in}, μ, σ.

Definition 5.7. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$ be a $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS. For a real number $\alpha \geq 0$, define $\alpha \cdot \mathcal{A}$, which we also write as $\alpha\mathcal{A}$, to be the $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS $(\alpha\mathrm{A}_1, \ldots, \alpha\mathrm{A}_{2^{d_{\mathsf{out}}}})$.*

Claim 5.8. *Let $\mathcal{A}$ be a $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBS and $\alpha > 0$ a real number. Then,*

- $\langle\alpha\mathcal{A}\rangle = \alpha\langle\mathcal{A}\rangle$;
- $\sigma(\alpha\mathcal{A}) = \sigma(\mathcal{A}) + 2\log(1/\alpha)$;
- $\mu(\alpha\mathcal{A}) = \mu(\mathcal{A}) - 2\log(1/\alpha)$.

Proof. The first item follows as

$$\langle\alpha\mathcal{A}\rangle = \mathop{\mathbf{E}}_i \langle\alpha\mathrm{A}_i\rangle = \alpha \mathop{\mathbf{E}}_i \langle\mathrm{A}_i\rangle = \alpha\langle\mathcal{A}\rangle.$$

As for the second item,

$$2^{-\sigma(\alpha\mathcal{A})} = \mathop{\mathbf{E}}_i \|\alpha\mathrm{A}_i\|_\infty^2 = \alpha^2 \mathop{\mathbf{E}}_i \|\mathrm{A}_i\|_\infty^2 = \alpha^2 2^{-\sigma(\mathcal{A})},$$

and so $\sigma(\alpha\mathcal{A}) = \sigma(\mathcal{A}) + 2\log(1/\alpha)$. As for the magnitude,

$$2^{\mu(\alpha\mathcal{A})} = \max_i \|\alpha \mathrm{A}_i\|_\infty^2 = \alpha^2 \max_i \|\mathrm{A}_i\|_\infty^2 = \alpha^2 2^{\mu(\mathcal{A})},$$

and so $\mu(\alpha\mathcal{A}) = \mu(\mathcal{A}) - 2\log(1/\alpha)$. □

5.3 Gluing MBSs

For our construction, we will need to "glue" MBSs, namely, stack the matrix bundles that compose two or more MBSs to one sequence. In this section, we formally define this operation and analyze the resulting "glued" MBS. In the following definition, we assume that the two MBSs to be glued have the same d_{out}. This is essentially without loss of generality as explained in the remark in Section 5.2.

Definition 5.9. *Let* $\mathcal{A} = (\mathrm{A}_1, \dots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$, $\mathcal{B} = (\mathrm{B}_1, \dots, \mathrm{B}_{2^{d_{\mathsf{out}}}})$ *be a pair of* $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$*-MBSs. We define the* gluing *of* $\mathcal{A}$ *and* $\mathcal{B}$*, denoted by* $\mathsf{glue}(\mathcal{A},\mathcal{B})$ *to be the* $(d_{\mathsf{out}}+1, d_{\mathsf{in}}, w)$*-MBS* $\mathcal{C} = (\mathrm{C}_1, \dots, \mathrm{C}_{2^{d_{\mathsf{out}}+1}})$ *that is defined by*

$$\mathrm{C}_i = \begin{cases} \mathrm{A}_i, & i \in [1, 2^{d_{\mathsf{out}}}]; \\ \mathrm{B}_{i-2^{d_{\mathsf{out}}}}, & i \in [2^{d_{\mathsf{out}}}+1, 2^{d_{\mathsf{out}}+1}]. \end{cases}$$

Claim 5.10. *Let* $\mathcal{A} = (\mathrm{A}_1, \dots, \mathrm{A}_{2^{d_{\mathsf{out}}}})$, $\mathcal{B} = (\mathrm{B}_1, \dots, \mathrm{B}_{2^{d_{\mathsf{out}}}})$ *be a pair of* $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$*-MBSs. Then,*

$$\langle \mathsf{glue}(\mathcal{A},\mathcal{B}) \rangle = \frac{\langle\mathcal{A}\rangle + \langle\mathcal{B}\rangle}{2}.$$

Moreover,

$$\sigma(\mathsf{glue}(\mathcal{A},\mathcal{B})) \ge \min(\sigma(\mathcal{A}), \sigma(\mathcal{B})),$$
$$\mu(\mathsf{glue}(\mathcal{A},\mathcal{B})) \le \max(\mu(\mathcal{A}), \mu(\mathcal{B})).$$

Proof. We have that

$$\begin{aligned} \langle \mathsf{glue}(\mathcal{A},\mathcal{B}) \rangle &= \mathop{\mathbf{E}}_{i \sim [2^{d_{\mathsf{out}}+1}]} \langle \mathrm{C}_i \rangle \\ &= \frac{1}{2^{d_{\mathsf{out}}+1}} \left(\sum_{i=1}^{2^{d_{\mathsf{out}}}} \langle \mathrm{A}_i \rangle + \sum_{i=1}^{2^{d_{\mathsf{out}}}} \langle \mathrm{B}_i \rangle \right) \\ &= \frac{1}{2} \left(\mathop{\mathbf{E}}_{i \sim [2^{d_{\mathsf{out}}}]} \langle \mathrm{A}_i \rangle + \mathop{\mathbf{E}}_{i \sim [2^{d_{\mathsf{out}}}]} \langle \mathrm{B}_i \rangle \right) \\ &= \frac{\langle\mathcal{A}\rangle + \langle\mathcal{B}\rangle}{2}. \end{aligned}$$

As for the smallness of $\mathsf{glue}(\mathcal{A}, \mathcal{B})$,

$$\begin{aligned}
2^{-\sigma(\mathsf{glue}(\mathcal{A},\mathcal{B}))} &= \mathop{\mathbf{E}}_{i\sim[2^{d_{\mathsf{out}}+1}]} \|\mathrm{C}_i\|_\infty^2 \\
&= \frac{1}{2^{d_{\mathsf{out}}+1}} \left(\sum_{i=1}^{2^{d_{\mathsf{out}}}} \|\mathrm{A}_i\|_\infty^2 + \sum_{i=1}^{2^{d_{\mathsf{out}}}} \|\mathrm{B}_i\|_\infty^2 \right) \\
&= \frac{1}{2} \left(\frac{1}{2^{d_{\mathsf{out}}}} \sum_{i=1}^{2^{d_{\mathsf{out}}}} \|\mathrm{A}_i\|_\infty^2 + \frac{1}{2^{d_{\mathsf{out}}}} \sum_{i=1}^{2^{d_{\mathsf{out}}}} \|\mathrm{B}_i\|_\infty^2 \right) \\
&= \frac{1}{2} \left(2^{-\sigma(\mathcal{A})} + 2^{-\sigma(\mathcal{B})} \right) \\
&\le \max\left(2^{-\sigma(\mathcal{A})}, 2^{-\sigma(\mathcal{B})} \right),
\end{aligned}$$

which implies that $\sigma(\mathsf{glue}(\mathcal{A}, \mathcal{B})) \ge \min(\sigma(\mathcal{A}), \sigma(\mathcal{B}))$, as claimed. The proof regarding the magnitude of $\mathsf{glue}(\mathcal{A}, \mathcal{B})$ is straightforward, and so we omit it. ☐

Generally, we may need to "glue" more than two MBSs. Let $\mathcal{A}_1, \ldots, \mathcal{A}_r$ be r $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBSs. We extend Definition 5.9 in the natural way to define the gluing of $\mathcal{A}_1, \ldots, \mathcal{A}_r$ which we denote by $\mathsf{glue}(\mathcal{A}_1, \ldots, \mathcal{A}_r)$. The following claim can be proved similarly to the way we proved Claim 5.10 and we omit the details.

Claim 5.11. *Let $r \ge 1$ be an integer. Let $\mathcal{A}_1, \ldots, \mathcal{A}_r$ be $(d_{\mathsf{out}}, d_{\mathsf{in}}, w)$-MBSs. Let $\mathcal{B} = \mathsf{glue}(\mathcal{A}_1, \ldots, \mathcal{A}_r)$. Then, $\langle \mathcal{B} \rangle = \mathbf{E}_i \langle \mathcal{A}_i \rangle$. Moreover,*

$$\begin{aligned}
\sigma(\mathcal{B}) &\ge \min_i \sigma(\mathcal{A}_i), \\
\mu(\mathcal{B}) &\le \max_i \mu(\mathcal{A}_i), \\
d_{\mathsf{out}}(\mathcal{B}) &= d_{\mathsf{out}} + \log r, \\
d_{\mathsf{in}}(\mathcal{B}) &= d_{\mathsf{in}}.
\end{aligned}$$

6 Multiplication Rules for Matrix Bundle Sequences

In this section we define several multiplication rules for MBSs and analyze the products.

6.1 The multiplication rules $\overrightarrow{\circ}$, $\overleftarrow{\circ}$ parameterized by a sampler

Definition 6.1. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS, where*

$$\mathrm{A}_i = \left(((\alpha_i)_1, (A_i)_1), \ldots, ((\alpha_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}, (A_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}) \right).$$

Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ be a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS, where

$$\mathrm{B}_i = \left(((\beta_i)_1, (B_i)_1), \ldots, ((\beta_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}, (B_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}) \right).$$

Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be a left-regular bipartite graph with left-degree 2^d. We define the $(d_{\mathsf{out}}(\mathcal{A}) + d, d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}), w)$-MBS $\mathcal{C} = \mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B}$ as follows: $\mathcal{C} = (\mathrm{C}_{i,j})_{i \in [2^{d_{\mathsf{out}}(\mathcal{A})}], j \in [2^d]}$, where the $(d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}), w)$-matrix bundle $\mathrm{C}_{i,j}$ is defined by

$$(\mathrm{C}_{i,j})_{k,\ell} = ((\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell, (A_i)_k (B_{\Gamma_G(i,j)})_\ell),$$

with $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$, $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$.

Note that $\mathcal{C}$ is indeed an MBS as the product of the stochastic matrices $(A_i)_k$, $(B_{\Gamma_G(i,j)})_\ell$ is stochastic. Moreover, $\mathcal{C}$ has the dimensions that were claimed in the definition, namely, $\mathcal{C}$ is a $(d_{\mathsf{out}}(\mathcal{A}) + d, d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}), w)$-MBS.

Claim 6.2. *For every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, $j \in [2^d]$, $\langle \mathrm{C}_{i,j} \rangle = \langle \mathrm{A}_i \rangle \langle \mathrm{B}_{\Gamma_G(i,j)} \rangle$.*

Proof. We have that

$$\begin{aligned}
\langle \mathrm{C}_{i,j} \rangle &= \sum_{k,\ell} (\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell (A_i)_k (B_{\Gamma_G(i,j)})_\ell \\
&= \sum_k (\alpha_i)_k (A_i)_k \sum_\ell (\beta_{\Gamma_G(i,j)})_\ell (B_{\Gamma_G(i,j)})_\ell \\
&= \langle \mathrm{A}_i \rangle \langle \mathrm{B}_{\Gamma_G(i,j)} \rangle.
\end{aligned}$$

□

By Claim 6.2,

$$\langle \mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B} \rangle = \mathop{\mathbf{E}}_{i,j} \left[\langle \mathrm{A}_i \rangle \langle \mathrm{B}_{\Gamma_G(i,j)} \rangle \right] = \mathop{\mathbf{E}}_i \left[\langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \right]. \tag{6.1}$$

In particular, note that if K is the complete bipartite graph on $[2^{d_{\mathsf{out}}(\mathcal{A})}] \times [2^{d_{\mathsf{out}}(\mathcal{B})}]$ then

$$\langle \mathcal{A} \overset{\rightarrow}{\circ}_K \mathcal{B} \rangle = \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle. \tag{6.2}$$

Similarly to the definition of $\overset{\rightarrow}{\circ}$, we define $\overset{\leftarrow}{\circ}$ as follows. Informally, the difference between $\overset{\rightarrow}{\circ}$ and $\overset{\leftarrow}{\circ}$ is that while sparsifying the product of $\mathcal{A}$ and $\mathcal{B}$, whether we use $\mathcal{A}$ or $\mathcal{B}$ as the left-side of the bipartite left-regular graph. This choice depends on the parameters of the MBSs.

Definition 6.3. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS, where*

$$\mathrm{A}_i = \left(((\alpha_i)_1, (A_i)_1), \ldots, ((\alpha_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}, (A_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}) \right).$$

Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ be a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS, where

$$\mathrm{B}_i = \left(((\beta_i)_1, (B_i)_1), \ldots, ((\beta_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}, (B_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}) \right).$$

Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be a bipartite left-regular graph with left-degree 2^d. We define the $(d_{\mathsf{out}}(\mathcal{A}) + d, d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}), w)$-MBS $\mathcal{C} = \mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}$ as follows: $\mathcal{C} = (\mathrm{C}_{i,j})_{i \in [2^{d_{\mathsf{out}}(\mathcal{A})}], j \in [2^d]}$, where

$$(\mathrm{C}_{i,j})_{k,\ell} = ((\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell, (B_{\Gamma_G(i,j)})_\ell (A_i)_k),$$

with $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$, $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$.

Similarly to Claim 6.2, we have that

Claim 6.4. $\langle \mathrm{C}_{i,j} \rangle = \langle \mathrm{B}_{\Gamma_G(i,j)} \rangle \langle \mathrm{A}_i \rangle$.

Proof. We have that

$$\begin{aligned}
\langle \mathrm{C}_{i,j} \rangle &= \sum_{k,\ell} (\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell (B_{\Gamma_G(i,j)})_\ell (A_i)_k \\
&= \sum_{\ell} (\beta_{\Gamma_G(i,j)})_\ell (B_{\Gamma_G(i,j)})_\ell \sum_k (\alpha_i)_k (A_i)_k \\
&= \langle \mathrm{B}_{\Gamma_G(i,j)} \rangle \langle \mathrm{A}_i \rangle.
\end{aligned}$$

□

By Claim 6.4,

$$\langle \mathcal{A} \overleftarrow{\circ}_G \mathcal{B} \rangle = \mathop{\mathbf{E}}_{i,j} \left[\langle \mathrm{B}_{\Gamma_G(i,j)} \rangle \langle \mathrm{A}_i \rangle \right] = \mathop{\mathbf{E}}_{i} \left[\left(\mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \right) \langle \mathrm{A}_i \rangle \right]. \tag{6.3}$$

In particular, if K is the complete bipartite graph on $[2^{d_{\mathsf{out}}(\mathcal{A})}] \times [2^{d_{\mathsf{out}}(\mathcal{B})}]$ then

$$\langle \mathcal{A} \overleftarrow{\circ}_K \mathcal{B} \rangle = \langle \mathcal{B} \rangle \langle \mathcal{A} \rangle. \tag{6.4}$$

The following lemma relates the properties of the MBS $\mathcal{A} \overrightarrow{\circ}_G \mathcal{B}$ to those of $\mathcal{A}, \mathcal{B}$. Throughout the paper, we will only apply the product $\overrightarrow{\circ}$ with the right operand being a thin MBS, and so we restrict ourselves to that case.

Lemma 6.5. *Let* $\mathcal{A} = (\mathrm{A}_1, \dots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$*-MBS. Let* $\mathcal{B} = (\mathrm{B}_1, \dots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{B}), 0, w)$*-thin MBS. Let* $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ *be a left-regular bipartite graph with left-degree* 2^d*. Then,*

$$\begin{aligned}
\sigma(\mathcal{A} \overrightarrow{\circ}_G \mathcal{B}) &\geq \sigma(\mathcal{A}), \\
\mu(\mathcal{A} \overrightarrow{\circ}_G \mathcal{B}) &\leq \mu(\mathcal{A}), \\
d_{\mathsf{out}}(\mathcal{A} \overrightarrow{\circ}_G \mathcal{B}) &= d_{\mathsf{out}}(\mathcal{A}) + d, \\
d_{\mathsf{in}}(\mathcal{A} \overrightarrow{\circ}_G \mathcal{B}) &= d_{\mathsf{in}}(\mathcal{A}).
\end{aligned}$$

Proof. The assertions regarding $d_{\mathsf{in}}, d_{\mathsf{out}}$ follow by the definition of $\overrightarrow{\circ}$ and by $d_{\mathsf{in}}(\mathcal{B}) = 0$. Write $\mathcal{C} = \mathcal{A} \overrightarrow{\circ}_G \mathcal{B}$ and let $\Gamma \colon [2^{d_{\mathsf{out}}(\mathcal{A})}] \times [2^d] \to [2^{d_{\mathsf{out}}(\mathcal{B})}]$ be the neighborhood function of G. By Claim (6.2), $\langle \mathrm{C}_{i,j} \rangle = \langle \mathrm{A}_i \rangle \langle \mathrm{B}_{\Gamma(i,j)} \rangle$. As $\| \cdot \|_\infty$ is sub-multiplicative and since $\langle \mathrm{B}_{\Gamma(i,j)} \rangle$ is stochastic (due to $\mathcal{B}$'s thinness),

$$\begin{aligned}
\| \mathrm{C}_{i,j} \|_\infty &= \| \langle \mathrm{A}_i \rangle \langle \mathrm{B}_{\Gamma(i,j)} \rangle \|_\infty \\
&\leq \| \mathrm{A}_i \|_\infty \| \mathrm{B}_{\Gamma(i,j)} \|_\infty \\
&= \| \mathrm{A}_i \|_\infty.
\end{aligned}$$

This proves that $\mu(\mathcal{C}) \le \mu(\mathcal{A})$. As for the smallness,

$$2^{-\sigma(\mathcal{C})} = \mathop{\mathbf{E}}_{i,j} \|\mathrm{C}_{i,j}\|_\infty^2 \le \mathop{\mathbf{E}}_{i} \|\mathrm{A}_i\|_\infty^2 = 2^{-\sigma(\mathcal{A})}.$$

□

The proof of Lemma 6.5, which considers the product $\overset{\rightarrow}{\circ}$ can be adapted to prove the same result for $\overset{\leftarrow}{\circ}$. We summarize this in the following lemma.

Lemma 6.6. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS. Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ be a $(d_{\mathsf{out}}(\mathcal{B}), 0, w)$-thin MBS. Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be a left-regular bipartite graph with left-degree 2^d. Then,*

$$\begin{aligned}
\sigma(\mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}) &\ge \sigma(\mathcal{A}), \\
\mu(\mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}) &\le \mu(\mathcal{A}), \\
d_{\mathsf{out}}(\mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}) &= d_{\mathsf{out}}(\mathcal{A}) + d, \\
d_{\mathsf{in}}(\mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}) &= d_{\mathsf{in}}(\mathcal{A}).
\end{aligned}$$

We make use of the following claim regarding thinness under the products $\overset{\rightarrow}{\circ}, \overset{\leftarrow}{\circ}$.

Claim 6.7. *Let $\mathcal{A}, \mathcal{B}$ be a pair of $(d_{\mathsf{out}}, 0, w)$-MBSs, both thin. Let $G = ([2^{d_{\mathsf{out}}}], [2^{d_{\mathsf{out}}}], E)$ be a left-regular bipartite graph. Then, both $\mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B}$ and $\mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}$ are thin.*

Proof. By Definition 6.1, $d_{\mathsf{in}}(\mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B}) = d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B})$. As both $\mathcal{A}, \mathcal{B}$ are thin, $d_{\mathsf{in}}(\mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B}) = 0$. Moreover, by Definition 6.1, every coefficient of $\mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B}$ is a product of some coefficient of $\mathcal{A}$ with some coefficient of $\mathcal{B}$. As both $\mathcal{A}, \mathcal{B}$ are thin, their coefficients all equal 1 and so the coefficients of $\mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B}$ are all 1. The proof for $\mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B}$ is similar and we omit it. □

6.2 The multiplication rules $\overset{\rightarrow}{\bullet}, \overset{\leftarrow}{\bullet}$ parameterized by a sampler

Definition 6.8. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS, where*

$$\mathrm{A}_i = \left(\left((\alpha_i)_1, (A_i)_1 \right), \ldots, \left((\alpha_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}, (A_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}} \right) \right).$$

Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ be a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS, where

$$\mathrm{B}_i = \left(\left((\beta_i)_1, (B_i)_1 \right), \ldots, \left((\beta_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}, (B_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}} \right) \right).$$

Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be a left-regular bipartite graph with left-degree 2^d. We define the $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + d, w)$-MBS, $\mathcal{C} = \mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B}$ as follows. For $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$, $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$, and $j \in [2^d]$ define

$$(\mathrm{C}_i)_{j,k,\ell} = (2^{-d} (\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell, (A_i)_k (B_{\Gamma_G(i,j)})_\ell).$$

Note that $\mathcal{C}$ is an MBS as the product of the stochastic matrices $(A_i)_k$, $(B_{\Gamma(i,j)})_\ell$ is stochastic. Moreover, the dimensions of $\mathcal{C}$ is as asserted in the definition. That is, $\mathcal{C}$ is a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + d, w)$-MBS.

Claim 6.9. $\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \, \mathbf{E}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle$.

Proof.

$$\begin{aligned}
\langle \mathrm{C}_i \rangle &= \sum_{j,k,\ell} 2^{-d} (\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell (A_i)_k (B_{\Gamma_G(i,j)})_\ell \\
&= \sum_k (\alpha_i)_k (A_i)_k 2^{-d} \sum_{j \in [2^d]} \sum_\ell (\beta_{\Gamma_G(i,j)})_\ell (B_{\Gamma_G(i,j)})_\ell \\
&= \left(\sum_k (\alpha_i)_k (A_i)_k \right) \mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \\
&= \langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle .
\end{aligned}$$

□

Claim 6.9 readily implies that

$$\langle \mathcal{A} \overrightarrow{\bullet}_G \mathcal{B} \rangle = \mathop{\mathbf{E}}_i \left[\langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \right]. \tag{6.5}$$

Definition 6.10. *Let* $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$*-MBS, where*

$$\mathrm{A}_i = \left(((\alpha_i)_1, (A_i)_1), \ldots, ((\alpha_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}, (A_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}) \right).$$

Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$*-MBS, where*

$$\mathrm{B}_i = \left(((\beta_i)_1, (B_i)_1), \ldots, ((\beta_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}, (B_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}) \right).$$

Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ *be a left-regular bipartite graph with left-degree* 2^d*. We define the* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + d, w)$*-MBS* $\mathcal{C} = \mathcal{A} \overleftarrow{\bullet}_G \mathcal{B}$ *as follows. For* $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$*,* $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$*, and* $j \in [2^d]$ *define*

$$(\mathrm{C}_i)_{j,k,\ell} = (2^{-d} (\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell, (B_{\Gamma_G(i,j)})_\ell (A_i)_k).$$

Claim 6.11. $\langle \mathrm{C}_i \rangle = \left(\mathbf{E}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \right) \langle \mathrm{A}_i \rangle$.

Proof.

$$\begin{aligned}
\langle \mathrm{C}_i \rangle &= \sum_{j,k,\ell} 2^{-d} (\alpha_i)_k (\beta_{\Gamma_G(i,j)})_\ell (B_{\Gamma_G(i,j)})_\ell (A_i)_k \\
&= \left(2^{-d} \sum_{j \in [2^d]} \sum_\ell (\beta_{\Gamma_G(i,j)})_\ell (B_{\Gamma_G(i,j)})_\ell \right) \sum_k (\alpha_i)_k (A_i)_k \\
&= \left(\mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \right) \langle \mathrm{A}_i \rangle .
\end{aligned}$$

□

By Claim 6.11,

$$\langle \mathcal{A} \overset{\leftarrow}{\bullet}_G \mathcal{B} \rangle = \mathop{\mathbf{E}}_{i}\left[\left(\mathop{\mathbf{E}}_{j \sim \Gamma_G(i)} \langle \mathrm{B}_j \rangle \right) \langle \mathrm{A}_i \rangle \right]. \tag{6.6}$$

The following claim readily follows by Equations (6.1), (6.3), (6.5), and (6.6).

Claim 6.12. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS. Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ be a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS. Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be a left-regular bipartite graph. Then,*

$$\langle \mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B} \rangle = \langle \mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B} \rangle,$$
$$\langle \mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B} \rangle = \langle \mathcal{A} \overset{\leftarrow}{\bullet}_G \mathcal{B} \rangle.$$

Claim 6.12 together with Equation (6.2) and Equation (6.4) implies that

$$\langle \mathcal{A} \overset{\rightarrow}{\bullet}_K \mathcal{B} \rangle = \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle,$$
$$\langle \mathcal{A} \overset{\leftarrow}{\bullet}_K \mathcal{B} \rangle = \langle \mathcal{B} \rangle \langle \mathcal{A} \rangle, \tag{6.7}$$

where K is the complete bipartite graph on $[2^{d_{\mathsf{out}}(\mathcal{A})}] \times [2^{d_{\mathsf{out}}(\mathcal{B})}]$.

The following lemma shows that the matrix that is realized by the product $\mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B}$ approximates $\langle \mathcal{A} \rangle \langle \mathcal{B} \rangle$, where the approximation guarantee is determined by the parameters of the sampler G (and those of $\mathcal{A}, \mathcal{B}$).

Lemma 6.13. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS and $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS. Let $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be an (ε, δ)-sampler with $\delta \le 1/2$. Then,*

$$\| \langle \mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B} \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \|_{\max} \le 4w 2^{\frac{\mu(\mathcal{B})}{2}} \left(2^{\frac{\mu(\mathcal{A})}{2}} \delta + 2^{-\frac{\sigma(\mathcal{A})}{2}} \varepsilon \right). \tag{6.8}$$

Furthermore, the same bound holds also for

$$\| \langle \mathcal{A} \overset{\leftarrow}{\bullet}_G \mathcal{B} \rangle - \langle \mathcal{B} \rangle \langle \mathcal{A} \rangle \|_{\max},$$
$$\| \langle \mathcal{A} \overset{\rightarrow}{\circ}_G \mathcal{B} \rangle - \langle \mathcal{A} \rangle \langle \mathcal{B} \rangle \|_{\max},$$
$$\| \langle \mathcal{A} \overset{\leftarrow}{\circ}_G \mathcal{B} \rangle - \langle \mathcal{B} \rangle \langle \mathcal{A} \rangle \|_{\max}.$$

Proof. We prove Equation (6.8). A similar proof gives the same bound for $\| \langle \mathcal{A} \overset{\leftarrow}{\bullet}_G \mathcal{B} \rangle - \langle \mathcal{B} \rangle \langle \mathcal{A} \rangle \|_{\max}$. The bound for the third and fourth expressions then follows by Claim 6.12. Write $\mathcal{C} = \mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B} = (\mathrm{C}_i)_{i=1}^{2^{d_{\mathsf{out}}(\mathcal{A})}}$ and let $\Gamma \colon [2^{d_{\mathsf{out}}(\mathcal{A})}] \times [2^d] \to [2^{d_{\mathsf{out}}(\mathcal{B})}]$ be the neighborhood function of G, where 2^d is the degree of the sampler. By Claim 6.9, for every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, $\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \mathbf{E}_{j \sim \Gamma(i)} \langle \mathrm{B}_j \rangle$. Therefore, for every $\alpha, \beta \in [w]$,

$$\langle \mathrm{C}_i \rangle_{\alpha,\beta} = \sum_{\gamma=1}^{w} \langle \mathrm{A}_i \rangle_{\alpha,\gamma} \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \langle \mathrm{B}_j \rangle_{\gamma,\beta},$$

and so

$$\begin{aligned}\langle \mathcal{C}\rangle_{\alpha,\beta} &= \mathop{\mathbf{E}}_{i} \langle \mathrm{C}_i\rangle_{\alpha,\beta} \\ &= \sum_{\gamma=1}^{w} \mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma} \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \langle \mathrm{B}_j\rangle_{\gamma,\beta}\right]. \end{aligned} \tag{6.9}$$

For fixed $\alpha,\beta,\gamma \in [w]$, define

$$\varepsilon_{\gamma,\beta}(i) = \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \langle \mathrm{B}_j\rangle_{\gamma,\beta} - \langle \mathcal{B}\rangle_{\gamma,\beta}.$$

Note that $|\varepsilon_{\gamma,\beta}(i)| \le 2^{\mu(\mathcal{B})/2+1}$ for all $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$. Moreover, as $\langle \mathcal{B}\rangle_{\gamma,\beta} = \mathbf{E}_{j\sim[2^{d_{\mathsf{out}}(\mathcal{B})}]} \langle \mathrm{B}_j\rangle_{\gamma,\beta}$ and since $|\langle \mathrm{B}_j\rangle_{\gamma,\beta}| \le 2^{\mu(\mathcal{B})/2}$, Claim 3.8 implies that there exists a set $S \subseteq [2^{d_{\mathsf{out}}(\mathcal{A})}]$ with $|S| \ge (1-\delta)\cdot 2^{d_{\mathsf{out}}(\mathcal{A})}$ such that for all $i \in S$, $|\varepsilon_{\gamma,\beta}(i)| \le \varepsilon \cdot 2^{\mu(\mathcal{B})/2+1}$. Therefore,

$$\begin{aligned}\mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma} \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \langle \mathrm{B}_j\rangle_{\gamma,\beta}\right] &= \mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\left(\langle \mathcal{B}\rangle_{\gamma,\beta} + \varepsilon_{\gamma,\beta}(i)\right)\right] \\ &= \langle \mathcal{A}\rangle_{\alpha,\gamma}\langle \mathcal{B}\rangle_{\gamma,\beta} + \mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i)\right]. \end{aligned} \tag{6.10}$$

As $|\langle \mathrm{A}_i\rangle_{\alpha,\gamma}| \le 2^{\mu(\mathcal{A})/2}$ and $|\varepsilon_{\gamma,\beta}(i)| \le 2^{\mu(\mathcal{B})/2+1}$ for all $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, we have that

$$\begin{aligned}\mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i)\right] &\le \mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i) \mid i \in S\right] + 2^{\frac{\mu(\mathcal{A})+\mu(\mathcal{B})}{2}+1}\,\mathbf{Pr}[i \notin S] \\ &\le \mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i) \mid i \in S\right] + 2^{\frac{\mu(\mathcal{A})+\mu(\mathcal{B})}{2}+1}\delta. \end{aligned} \tag{6.11}$$

By Jensen's inequality, and using the fact that $(\langle \mathrm{A}_i\rangle_{\alpha,\gamma})^2 \ge 0$,

$$\begin{aligned}\left(\mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i) \mid i \in S\right]\right)^2 &\le \mathop{\mathbf{E}}_{i}\left[\left(\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i)\right)^2 \mid i \in S\right] \\ &\le \left(\varepsilon 2^{\mu(\mathcal{B})/2+1}\right)^2 \mathop{\mathbf{E}}_{i}\left[\left(\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\right)^2 \mid i \in S\right] \\ &\le \left(\varepsilon 2^{\mu(\mathcal{B})/2+1}\right)^2 \frac{\mathbf{E}_i\left[\left(\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\right)^2\right]}{\mathbf{Pr}[i \in S]} \\ &\le \left(\varepsilon 2^{\mu(\mathcal{B})/2+1}\right)^2 \frac{2^{-\sigma(\mathcal{A})}}{\mathbf{Pr}[i \in S]}. \end{aligned}$$

As $\delta \le 1/2$, we have $\mathbf{Pr}[i \in S] \ge 1-\delta \ge 1/2$ and so

$$\left|\mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma}\varepsilon_{\gamma,\beta}(i) \mid i \in S\right]\right| \le 2^{\frac{\mu(\mathcal{B})}{2}-\frac{\sigma(\mathcal{A})}{2}+2}\varepsilon.$$

Equations (6.10), (6.11) then implies

$$\left|\mathop{\mathbf{E}}_{i}\left[\langle \mathrm{A}_i\rangle_{\alpha,\gamma} \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \langle \mathrm{B}_j\rangle_{\gamma,\beta}\right] - \langle \mathcal{A}\rangle_{\alpha,\gamma}\langle \mathcal{B}\rangle_{\gamma,\beta}\right| \le 2^{\frac{\mu(\mathcal{A})+\mu(\mathcal{B})}{2}+1}\delta + 2^{\frac{\mu(\mathcal{B})}{2}-\frac{\sigma(\mathcal{A})}{2}+2}\varepsilon.$$

As the bound holds for all $\gamma \in [w]$, Equation (6.9) yields

$$|\langle \mathcal{C} \rangle_{\alpha,\beta} - (\langle \mathcal{A} \rangle \langle \mathcal{B} \rangle)_{\alpha,\beta}| \leq 4w2^{\frac{\mu(\mathcal{B})}{2}} \left(2^{\frac{\mu(\mathcal{A})}{2}} \delta + 2^{-\frac{\sigma(\mathcal{A})}{2}} \varepsilon \right).$$

The proof follows as the bound holds for every $\alpha, \beta \in [w]$. □

Next, we show that by taking a good enough sampler, the smallness of the product $\mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B}$ (and of the other products) approaches the sum $\sigma(\mathcal{A}) + \sigma(\mathcal{B})$ and that the magnitude of the product is bounded by $\mu(\mathcal{A}) + \mu(\mathcal{B})$.

Lemma 6.14. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS and $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS. Let $\tau \in (0, 1]$ and $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be an (ε, δ)-sampler with*

$$\begin{aligned} \varepsilon &\leq 2^{-\lambda(\mathcal{B}) - \mu(\mathcal{B}) - \log(1/\tau) - 3}, \\ \delta &\leq 2^{-\lambda(\mathcal{A}) - \lambda(\mathcal{B}) - \mu(\mathcal{A}) - \mu(\mathcal{B}) - \log(1/\tau) - 3}, \end{aligned} \tag{6.12}$$

for some $\lambda(\mathcal{A}), \lambda(\mathcal{B})$ such that $0 \leq \lambda(\mathcal{A}) \leq \sigma(\mathcal{A})$ and $0 \leq \lambda(\mathcal{B}) \leq \sigma(\mathcal{B})$. Then,

$$\begin{aligned} \sigma(\mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B}) &\geq \lambda(\mathcal{A}) + \lambda(\mathcal{B}) - \tau, \\ \mu(\mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B}) &\leq \mu(\mathcal{A}) + \mu(\mathcal{B}). \end{aligned}$$

Proof. Write $\mathcal{C} = \mathcal{A} \overset{\rightarrow}{\bullet}_G \mathcal{B} = (\mathrm{C}_i)_{i=1}^{2^{d_{\mathsf{out}}(\mathcal{A})}}$ and let $\Gamma \colon [2^{d_{\mathsf{out}}(\mathcal{A})}] \times [2^d] \to [2^{d_{\mathsf{out}}(\mathcal{B})}]$ be the neighborhood function of G, where 2^d is the degree of the sampler. For $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, define

$$\varepsilon(i) = \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \|\mathrm{B}_j\|_\infty^2 - 2^{-\sigma(\mathcal{B})}.$$

As G is an (ε, δ)-sampler, and since $0 \leq \|\mathrm{B}_j\|_\infty^2 \leq 2^{\mu(\mathcal{B})}$ for all $j \in [2^{d_{\mathsf{out}}(\mathcal{B})}]$, Claim 3.8 implies that there exists a set $S \subseteq [2^{d_{\mathsf{out}}(\mathcal{A})}]$ with $|S| \geq (1 - \delta) 2^{d_{\mathsf{out}}(\mathcal{A})}$ such that for every $i \in S$, $|\varepsilon(i)| \leq \varepsilon 2^{\mu(\mathcal{B})}$. By Claim 6.9, for every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$,

$$\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \langle \mathrm{B}_j \rangle.$$

By Jensen's inequality and since $\|\cdot\|_\infty$ is sub-multiplicative (and sub-additive),

$$\begin{aligned} 2^{-\sigma(\mathcal{C})} &= \mathop{\mathbf{E}}_{i} \|\mathrm{C}_i\|_\infty^2 \\ &= \mathop{\mathbf{E}}_{i} \|\langle \mathrm{A}_i \rangle \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \langle \mathrm{B}_j \rangle \|_\infty^2 \\ &\leq \mathop{\mathbf{E}}_{i} \left[\|\mathrm{A}_i\|_\infty^2 \mathop{\mathbf{E}}_{j \sim \Gamma(i)} \|\mathrm{B}_j\|_\infty^2 \right]. \end{aligned}$$

Thus,

$$2^{-\sigma(\mathcal{C})} \le \mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2 (2^{-\sigma(\mathcal{B})} + \varepsilon(i))\right]$$
$$= 2^{-\sigma(\mathcal{A})-\sigma(\mathcal{B})} + \mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i)\right]. \tag{6.13}$$

As $\|\mathrm{A}_i\|_\infty^2 \le 2^{\mu(\mathcal{A})}$ and since $|\varepsilon(i)| \le 2^{\mu(\mathcal{B})}$ for all $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$,

$$\mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i)\right] \le \mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i) \;\middle|\; i \in S\right] + 2^{\mu(\mathcal{A})+\mu(\mathcal{B})}\, \mathbf{Pr}[i \notin S]$$
$$\le \varepsilon 2^{\mu(\mathcal{B})} \mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2 \;\middle|\; i \in S\right] + 2^{\mu(\mathcal{A})+\mu(\mathcal{B})}\delta.$$

As $\delta \le 1/2$, $\mathbf{Pr}[i \in S] = 1 - \delta \ge 1/2$ and since $\|\mathrm{A}_i\|_\infty^2 \ge 0$,

$$\mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2 \;\middle|\; i \in S\right] \le \frac{1}{\mathbf{Pr}[i \in S]} \mathop{\mathbf{E}}_{i}\left[\|\mathrm{A}_i\|_\infty^2\right]$$
$$\le 2^{-\sigma(\mathcal{A})+1}.$$

Hence, $\mathbf{E}_i\left[\|\mathrm{A}_i\|_\infty^2 \varepsilon(i)\right] \le 2^{\mu(\mathcal{B})-\sigma(\mathcal{A})+1}\varepsilon + 2^{\mu(\mathcal{A})+\mu(\mathcal{B})}\delta$. Plugging this to Equation (6.13), we get

$$2^{-\sigma(\mathcal{C})} \le 2^{-\sigma(\mathcal{A})-\sigma(\mathcal{B})} + 2^{\mu(\mathcal{B})-\sigma(\mathcal{A})+1}\varepsilon + 2^{\mu(\mathcal{A})+\mu(\mathcal{B})}\delta.$$

Substituting for ε, δ we conclude

$$2^{-\sigma(\mathcal{C})} \le \left(1 + \frac{3\tau}{8}\right) 2^{-\lambda(\mathcal{A})-\lambda(\mathcal{B})} \le 2^{-\lambda(\mathcal{A})-\lambda(\mathcal{B})+\tau},$$

where, for the last inequality we used the fact that $1 + x \le e^x$ for all x.

We move to analyze the magnitude. As $\|\cdot\|_\infty$ is sub-multiplicative (and sub-additive), for every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$,

$$\|\mathrm{C}_i\|_\infty^2 \le \|\mathrm{A}_i\|_\infty^2 \left\| \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \langle \mathrm{B}_j \rangle \right\|_\infty^2$$
$$\le \|\mathrm{A}_i\|_\infty^2 \mathop{\mathbf{E}}_{j\sim\Gamma(i)} \|\mathrm{B}_j\|_\infty^2$$
$$\le 2^{\mu(\mathcal{A})+\mu(\mathcal{B})},$$

which implies that $\mu(\mathcal{C}) \le \mu(\mathcal{A}) + \mu(\mathcal{B})$. □

The proof of Lemma 6.14, which considers the product $\overset{\rightarrow}{\bullet}$ can be adapted to prove the same lemma for $\overset{\leftarrow}{\bullet}$, which is given by the following lemma.

Lemma 6.15. *Let $\mathcal{A}$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS and $\mathcal{B}$ a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS. Let $\tau \in (0,1]$ and $G = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E)$ be an (ε,δ)-sampler for which Equation (6.12) holds. Then, $\sigma(\mathcal{A} \overset{\leftarrow}{\bullet}_G \mathcal{B}) \ge \lambda(\mathcal{A}) + \lambda(\mathcal{B}) - \tau$ and $\mu(\mathcal{A} \overset{\leftarrow}{\bullet}_G \mathcal{B}) \le \mu(\mathcal{A}) + \mu(\mathcal{B})$.*

6.3 The multiplication rules $\overrightarrow{\bullet}, \overleftarrow{\bullet}$ parameterized by delta of samplers

In this section we define multiplication rules that are parameterized by the difference, or delta, between two samplers.

Definition 6.16. *Let $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ be a $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$-MBS, where*

$$\mathrm{A}_i = \left(((\alpha_i)_1, (A_i)_1), \ldots, ((\alpha_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}, (A_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}})\right).$$

Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ be a $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$-MBS, where

$$\mathrm{B}_i = \left(((\beta_i)_1, (B_i)_1), \ldots, ((\beta_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}, (B_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}})\right).$$

Let $D \geq d \geq 1$ be integers. Let $G_D = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_D)$ be a left-regular bipartite graph with left-degree 2^D and $G_d = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_d)$ a left-regular bipartite graph with left-degree 2^d. We define the $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + D + 1, w)$-MBS $\mathcal{C} = \mathcal{A} \overrightarrow{\bullet}_{G_D - G_d} \mathcal{B}$ as follows: For $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$, $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$, and $j \in [2^D]$, define

$$(\mathrm{C}_i)^D_{j,k,\ell} = (2^{-D}(\alpha_i)_k(\beta_{\Gamma_{G_D}(i,j)})_\ell, (A_i)_k(B_{\Gamma_{G_D}(i,j)})_\ell).$$

For $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$, $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$, and $j \in [2^d]$, define

$$(\mathrm{C}_i)^d_{j,k,\ell} = (-2^{-d}(\alpha_i)_k(\beta_{\Gamma_{G_d}(i,j)})_\ell, (A_i)_k(B_{\Gamma_{G_d}(i,j)})_\ell).$$

Finally, $\mathcal{C} = (\mathrm{C}_i)_{i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]}$ where C_i is the concatenation of the sequences $\mathrm{C}_i^D, \mathrm{C}_i^d$.

Note that $\mathcal{C}$ is an MBS as the stochastic property is preserved. Further, by definition,

$$\begin{aligned} 2^{d_{\mathsf{in}}(\mathcal{C})} &= 2^{d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B})} \cdot \left(2^D + 2^d\right) \\ &\leq 2^{d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B})} \cdot 2^{D+1}, \end{aligned}$$

and so we indeed may regard $\mathcal{C}$ as having the stated d_{in} (see the remark regarding the monotonicity of d_{in} in Section 5.2). We have the following claim.

Claim 6.17. *With the notation of Definition 6.16, for every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$,*

$$\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \left(\mathop{\mathbf{E}}_{j \sim \Gamma_{G_D}(i)} \langle \mathrm{B}_j \rangle - \mathop{\mathbf{E}}_{j \sim \Gamma_{G_d}(i)} \langle \mathrm{B}_j \rangle \right).$$

Proof. Let $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$. As C_i is the concatenation of $(\mathrm{C}_i)^D$ and $(\mathrm{C}_i)^d$, $\langle \mathrm{C}_i \rangle = \langle (\mathrm{C}_i)^D \rangle + \langle (\mathrm{C}_i)^d \rangle$. Thus,

$$\langle \mathrm{C}_i \rangle = \sum_{k,\ell} \sum_{j \in [2^D]} (\mathrm{C}_i)^D_{j,k,\ell} + \sum_{k,\ell} \sum_{j \in [2^d]} (\mathrm{C}_i)^d_{j,k,\ell}$$

The first summand in the RHS of the above equation equals to

$$\begin{aligned}\sum_{k,\ell}\sum_{j\in[2^D]}(\mathrm{C}_i)^D_{j,k,\ell} &= \sum_{k,\ell}\sum_{j\in[2^D]} 2^{-D}(\alpha_i)_k(\beta_{\Gamma_{G_D}(i,j)})_\ell(A_i)_k(B_{\Gamma_{G_D}(i,j)})_\ell\\ &= \sum_k(\alpha_i)_k(A_i)_k \mathop{\mathbf{E}}_{j\sim\Gamma_{G_D}(i)}\sum_\ell(\beta_j)_\ell(B_j)_\ell\\ &= \langle \mathrm{A}_i\rangle \mathop{\mathbf{E}}_{j\sim\Gamma_{G_D}(i)}\langle \mathrm{B}_j\rangle.\end{aligned}$$

As for the second summand,

$$\begin{aligned}\sum_{k,\ell}\sum_{j\in[2^d]}(\mathrm{C}_i)^d_{j,k,\ell} &= \sum_{k,\ell}\sum_{j\in[2^d]} -2^{-d}(\alpha_i)_k(\beta_{\Gamma_{G_d}(i,j)})_\ell(A_i)_k(B_{\Gamma_{G_d}(i,j)})_\ell\\ &= -\sum_k(\alpha_i)_k(A_i)_k \mathop{\mathbf{E}}_{j\sim\Gamma_{G_d}(i)}\sum_\ell(\beta_j)_\ell(B_j)_\ell\\ &= -\langle \mathrm{A}_i\rangle \mathop{\mathbf{E}}_{j\sim\Gamma_{G_d}(i)}\langle \mathrm{B}_j\rangle,\end{aligned}$$

which completes the proof. □

Claim 6.17, together with Equation (6.5), readily yields

$$\langle \mathcal{A} \vec{\bullet}_{G_D-G_d} \mathcal{B}\rangle = \langle \mathcal{A} \vec{\bullet}_{G_D} \mathcal{B}\rangle - \langle \mathcal{A} \vec{\bullet}_{G_d} \mathcal{B}\rangle. \tag{6.14}$$

We refer to this property as the *linearity of* $\vec{\bullet}$.

Lemma 6.18. *Let* $\mathcal{A} = (\mathrm{A}_1,\ldots,\mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$*-MBS and* $\mathcal{B} = (\mathrm{B}_1,\ldots,\mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ *a* $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$*-MBS. Let* $G_D = ([2^{d_{\mathsf{out}}(\mathcal{A})}],[2^{d_{\mathsf{out}}(\mathcal{B})}],E_D)$ *be an* (ε_1,δ_1)*-sampler and* $G_d = ([2^{d_{\mathsf{out}}(\mathcal{A})}],[2^{d_{\mathsf{out}}(\mathcal{B})}],E_d)$ *an* (ε_2,δ_2)*-sampler. Assume that* $0<\varepsilon_1\le\varepsilon_2<1$ *and* $0<\delta_1\le\delta_2\le 1/(4w^2)$*. Denote the degrees of* G_D, G_d *by* $2^D, 2^d$*, respectively, and assume that* $D\ge d$*. Then,*

$$\begin{aligned}d_{\mathsf{in}}(\mathcal{A} \vec{\bullet}_{G_D-G_d} \mathcal{B}) &\le d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + D + 1;\\ d_{\mathsf{out}}(\mathcal{A} \vec{\bullet}_{G_D-G_d} \mathcal{B}) &= d_{\mathsf{out}}(\mathcal{A});\\ \sigma(\mathcal{A} \vec{\bullet}_{G_D-G_d} \mathcal{B}) &\ge \min\left(2\log\left(\frac{1}{\varepsilon_2}\right)+\sigma(\mathcal{A}),\ \log\left(\frac{1}{\delta_2}\right)-\mu(\mathcal{A})\right) - \mu(\mathcal{B}) - 2\log w - 6;\\ \mu(\mathcal{A} \vec{\bullet}_{G_D-G_d} \mathcal{B}) &\le \mu(\mathcal{A})+\mu(\mathcal{B})+2.\end{aligned}$$

Proof. The assertions regarding $d_{\mathsf{in}}, d_{\mathsf{out}}$ readily follow by Definition 6.16 as since we assume $D\ge d$. We turn to analyze the smallness of the product. Write $\mathcal{C} = \mathcal{A} \vec{\bullet}_{G_D-G_d} \mathcal{B} = (\mathrm{C}_i)_{i=1}^{2^{d_{\mathsf{out}}(\mathcal{A})}}$. Let $\Gamma_1\colon [2^{d_{\mathsf{out}}(\mathcal{A})}]\times[2^D]\to[2^{d_{\mathsf{out}}(\mathcal{B})}]$ be the neighborhood function of G_D and $\Gamma_2\colon [2^{d_{\mathsf{out}}(\mathcal{A})}]\times[2^d]\to[2^{d_{\mathsf{out}}(\mathcal{B})}]$ the neighborhood function of G_d. By Claim 6.17, for all $i\in[2^{d_{\mathsf{out}}(\mathcal{A})}]$,

$$\langle \mathrm{C}_i\rangle = \langle \mathrm{A}_i\rangle\left(\mathop{\mathbf{E}}_{j\sim\Gamma_1(i)}\langle \mathrm{B}_j\rangle - \mathop{\mathbf{E}}_{j\sim\Gamma_2(i)}\langle \mathrm{B}_j\rangle\right),$$

and so, using the fact that $\|\cdot\|_\infty$ is sub-multiplicative,

$$\|\mathrm{C}_i\|_\infty \le \|\mathrm{A}_i\|_\infty \Big\| \mathop{\mathbf{E}}_{j\sim\Gamma_1(i)} \langle \mathrm{B}_j\rangle - \mathop{\mathbf{E}}_{j\sim\Gamma_2(i)} \langle \mathrm{B}_j\rangle \Big\|_\infty. \tag{6.15}$$

By standard norm inequalities (see Claim 3.6),

$$\begin{aligned}\Big\| \mathop{\mathbf{E}}_{j\sim\Gamma_1(i)} \langle \mathrm{B}_j\rangle - \mathop{\mathbf{E}}_{j\sim\Gamma_2(i)} \langle \mathrm{B}_j\rangle \Big\|_\infty &\le w \Big\| \mathop{\mathbf{E}}_{j\sim\Gamma_1(i)} \langle \mathrm{B}_j\rangle - \mathop{\mathbf{E}}_{j\sim\Gamma_2(i)} \langle \mathrm{B}_j\rangle \Big\|_{\max} \\ &\le w \left(\Big\| \mathop{\mathbf{E}}_{j\sim\Gamma_1(i)} \langle \mathrm{B}_j\rangle - \langle \mathcal{B}\rangle \Big\|_{\max} + \Big\| \mathop{\mathbf{E}}_{j\sim\Gamma_2(i)} \langle \mathrm{B}_j\rangle - \langle \mathcal{B}\rangle \Big\|_{\max} \right).\end{aligned} \tag{6.16}$$

Fix $\alpha,\beta \in [w]$. For $s \in \{1,2\}$ and $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, define

$$\varepsilon_s^{\alpha,\beta}(i) = \mathop{\mathbf{E}}_{j\sim\Gamma_s(i)} \langle \mathrm{B}_j\rangle_{\alpha,\beta} - \langle \mathcal{B}\rangle_{\alpha,\beta}.$$

Note that $\langle \mathcal{B}\rangle_{\alpha,\beta} = \mathbf{E}_{j\sim[2^{d_{\mathsf{out}}(\mathcal{B})}]} \langle \mathrm{B}_j\rangle_{\alpha,\beta}$. Thus, as G_s is an (ε_s,δ_s)-sampler (here, we refer to G_D by G_1 and G_d by G_2), and since $|\langle \mathrm{B}_j\rangle_{\alpha,\beta}| \le 2^{\mu(\mathcal{B})/2}$ for all $j \in [2^{d_{\mathsf{out}}(\mathcal{B})}]$, there exists a set $S_s^{\alpha,\beta} \subseteq [2^{d_{\mathsf{out}}(\mathcal{A})}]$ of size $|S_s^{\alpha,\beta}| \ge (1-\delta_s)2^{d_{\mathsf{out}}(\mathcal{A})}$ such that for every $i \in S_s^{\alpha,\beta}$, $|\varepsilon_s^{\alpha,\beta}(i)| \le 2^{\mu(\mathcal{B})/2+1}\varepsilon_s$. Moreover, for every $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, $|\varepsilon_s^{\alpha,\beta}(i)| \le 2^{\mu(\mathcal{B})/2+1}$. For $s \in \{1,2\}$ and $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$, define

$$\varepsilon_s(i) = \max_{\alpha,\beta\in[w]} \left|\varepsilon_s^{\alpha,\beta}(i)\right|.$$

By Equation (6.16),

$$\Big\| \mathop{\mathbf{E}}_{j\sim\Gamma_1(i)} \langle \mathrm{B}_j\rangle - \mathop{\mathbf{E}}_{j\sim\Gamma_2(i)} \langle \mathrm{B}_j\rangle \Big\|_\infty \le w\left(\varepsilon_1(i) + \varepsilon_2(i)\right).$$

Let

$$S = \bigcap_{\alpha,\beta=1}^{w} \left(S_1^{\alpha,\beta} \cap S_2^{\alpha,\beta} \right).$$

Note that

$$|S| \ge \left(1 - (\delta_1+\delta_2)w^2\right) 2^{d_{\mathsf{out}}(\mathcal{A})} \ge \left(1 - 2\delta_2 w^2\right) 2^{d_{\mathsf{out}}(\mathcal{A})}. \tag{6.17}$$

Moreover, for every $i \in S$,

$$\varepsilon_1(i) + \varepsilon_2(i) \le (\varepsilon_1+\varepsilon_2)2^{\mu(\mathcal{B})/2+1} \le \varepsilon_2 2^{\mu(\mathcal{B})/2+2}. \tag{6.18}$$

By Equation (6.15),

$$\|\mathrm{C}_i\|_\infty^2 \le \|\mathrm{A}_i\|_\infty^2 w^2 \left(\varepsilon_1(i) + \varepsilon_2(i)\right)^2.$$

Taking expectation over $i \sim [2^{d_{\mathsf{out}}(\mathcal{A})}]$, we get

$$\begin{aligned}
2^{-\sigma(\mathcal{C})} &= \mathop{\mathbf{E}}_i \|\mathrm{C}_i\|_\infty^2 \\
&\le w^2 \mathop{\mathbf{E}}_i \left[\|\mathrm{A}_i\|_\infty^2 \left(\varepsilon_1(i) + \varepsilon_2(i)\right)^2\right] \\
&\le w^2 \mathop{\mathbf{E}}_i \left[\|\mathrm{A}_i\|_\infty^2 \left(\varepsilon_1(i) + \varepsilon_2(i)\right)^2 \,\middle|\, i \in S\right] + 2^{\mu(\mathcal{A})+\mu(\mathcal{B})+4}\, \mathbf{Pr}[i \notin S] \\
&\le w^2 \varepsilon_2^2 2^{\mu(\mathcal{B})+4} \mathop{\mathbf{E}}_i \left[\|\mathrm{A}_i\|_\infty^2 \,\middle|\, i \in S\right] + 2^{\mu(\mathcal{A})+\mu(\mathcal{B})+4}\, \mathbf{Pr}[i \notin S], \qquad (6.19)
\end{aligned}$$

where, for the penultimate inequality we used the fact that $\|\mathrm{A}_i\|_\infty^2 \le 2^{\mu(\mathcal{A})}$ and $\varepsilon_s(i) \le 2^{\mu(\mathcal{B})/2+1}$ for all i, and the last inequality follows by Equation (6.18). By Equation (6.17),

$$\mathbf{Pr}[i \notin S] \le 2\delta_2 w^2. \qquad (6.20)$$

In particular, $\mathbf{Pr}[i \in S] \ge 1/2$ per our assumption on δ_2. Using the fact that $\|\mathrm{A}_i\|_\infty^2 \ge 0$,

$$\mathop{\mathbf{E}}_i \left[\|\mathrm{A}_i\|_\infty^2 \,\middle|\, i \in S\right] \le \frac{\mathbf{E}_i \left[\|\mathrm{A}_i\|_\infty^2\right]}{\mathbf{Pr}[i \in S]} \le 2 \mathop{\mathbf{E}}_i \left[\|\mathrm{A}_i\|_\infty^2\right] = 2^{-\sigma(\mathcal{A})+1}. \qquad (6.21)$$

Equations (6.19), (6.20),(6.21) then imply $2^{-\sigma(\mathcal{C})} \le 2^{\mu(\mathcal{B})+5} w^2 \left(\varepsilon_2^2 2^{-\sigma(\mathcal{A})} + 2^{\mu(\mathcal{A})}\delta_2\right)$, which concludes the proof regarding the smallness of $\mathcal{C}$.

As for the magnitude, by Claim (6.17),

$$\langle \mathrm{C}_i \rangle = \langle \mathrm{A}_i \rangle \left(\mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \langle \mathrm{B}_j \rangle - \mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \langle \mathrm{B}_j \rangle \right),$$

and so, as $\|\cdot\|_\infty$ is sub-multiplicative (and sub-additive),

$$\begin{aligned}
\|\mathrm{C}_i\|_\infty &\le \|\mathrm{A}_i\|_\infty \left\| \mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \langle \mathrm{B}_j \rangle - \mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \langle \mathrm{B}_j \rangle \right\|_\infty \\
&\le \|\mathrm{A}_i\|_\infty \left(\left\| \mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \langle \mathrm{B}_j \rangle \right\|_\infty + \left\| \mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \langle \mathrm{B}_j \rangle \right\|_\infty \right) \\
&\le \|\mathrm{A}_i\|_\infty \left(\mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \|\mathrm{B}_j\|_\infty + \mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \|\mathrm{B}_j\|_\infty \right).
\end{aligned}$$

Hence, by Jensen's inequality,

$$\begin{aligned}
\|\mathrm{C}_i\|_\infty^2 &\le \|\mathrm{A}_i\|_\infty^2 \left(\mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \|\mathrm{B}_j\|_\infty + \mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \|\mathrm{B}_j\|_\infty \right)^2 \\
&\le \|\mathrm{A}_i\|_\infty^2 \cdot 2 \left(\left(\mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \|\mathrm{B}_j\|_\infty \right)^2 + \left(\mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \|\mathrm{B}_j\|_\infty \right)^2 \right) \\
&\le \|\mathrm{A}_i\|_\infty^2 \cdot 2 \left(\mathop{\mathbf{E}}_{j \sim \Gamma_1(i)} \|\mathrm{B}_j\|_\infty^2 + \mathop{\mathbf{E}}_{j \sim \Gamma_2(i)} \|\mathrm{B}_j\|_\infty^2 \right) \\
&\le 4 \cdot 2^{\mu(\mathcal{A})+\mu(\mathcal{B})}.
\end{aligned}$$

As this holds for all i, $\mu(\mathcal{C}) \le \mu(\mathcal{A}) + \mu(\mathcal{B}) + 2$, as claimed. □

Definition 6.19. *Let* $\mathcal{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_{2^{d_{\mathsf{out}}(\mathcal{A})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$*-MBS, where*

$$\mathrm{A}_i = \left(((\alpha_i)_1, (A_i)_1)\,, \ldots, ((\alpha_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}}, (A_i)_{2^{d_{\mathsf{in}}(\mathcal{A})}})\right).$$

Let $\mathcal{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_{2^{d_{\mathsf{out}}(\mathcal{B})}})$ *be a* $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$*-MBS, where*

$$\mathrm{B}_i = \left(((\beta_i)_1, (B_i)_1)\,, \ldots, ((\beta_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}}, (B_i)_{2^{d_{\mathsf{in}}(\mathcal{B})}})\right).$$

Let $D \geq d \geq 1$ *be integers. Let* $G_D = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_D)$ *be a left-regular bipartite graph with left-degree* 2^D *and* $G_d = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_d)$ *a left-regular bipartite graph with left-degree* 2^d*. We define the* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + D + 1, w)$*-MBS* $\mathcal{C} = \mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B}$ *as follows: For* $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$*,* $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$*,* $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$*, and* $j \in [2^D]$*, define*

$$(\mathrm{C}_i)^D_{j,k,\ell} = (2^{-D}(\alpha_i)_k(\beta_{\Gamma_{G_D}(i,j)})_\ell, (B_{\Gamma_{G_D}(i,j)})_\ell(A_i)_k).$$

For $i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]$*,* $k \in [2^{d_{\mathsf{in}}(\mathcal{A})}]$*,* $\ell \in [2^{d_{\mathsf{in}}(\mathcal{B})}]$*, and* $j \in [2^d]$*, define*

$$(\mathrm{C}_i)^d_{j,k,\ell} = (-2^{-d}(\alpha_i)_k(\beta_{\Gamma_{G_d}(i,j)})_\ell, (B_{\Gamma_{G_d}(i,j)})_\ell(A_i)_k).$$

Finally, $\mathcal{C} = (\mathrm{C}_i)_{i \in [2^{d_{\mathsf{out}}(\mathcal{A})}]}$ *where* C_i *is the concatenation of the sequences* $\mathrm{C}_i^D, \mathrm{C}_i^d$*.*

Similarly to the product $\overrightarrow{\bullet}$, one can show that

$$\langle \mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B} \rangle = \mathop{\mathbf{E}}_i \left[\left(\mathop{\mathbf{E}}_{j \sim \Gamma_{G_D}(i)} \langle \mathrm{B}_j \rangle - \mathop{\mathbf{E}}_{j \sim \Gamma_{G_d}(i)} \langle \mathrm{B}_j \rangle \right) \langle \mathrm{A}_i \rangle \right],$$

and that $\langle \mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B} \rangle = \langle \mathcal{A} \overleftarrow{\bullet}_{G_D} \mathcal{B} \rangle - \langle \mathcal{A} \overleftarrow{\bullet}_{G_d} \mathcal{B} \rangle$. The following lemma follows by similar arguments to those used to prove Lemma 6.18.

Lemma 6.20. *Let* $\mathcal{A}$ *be a* $(d_{\mathsf{out}}(\mathcal{A}), d_{\mathsf{in}}(\mathcal{A}), w)$*-MBS and* $\mathcal{B}$ *a* $(d_{\mathsf{out}}(\mathcal{B}), d_{\mathsf{in}}(\mathcal{B}), w)$*-MBS. Let* $G_D = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_D)$ *be an* $(\varepsilon_1, \delta_1)$*-sampler and* $G_d = ([2^{d_{\mathsf{out}}(\mathcal{A})}], [2^{d_{\mathsf{out}}(\mathcal{B})}], E_d)$ *an* $(\varepsilon_2, \delta_2)$*-sampler. Assume that* $0 < \varepsilon_1 \leq \varepsilon_2 < 1$ *and* $0 < \delta_1 \leq \delta_2 \leq 1/(4w^2)$*. Denote the degrees of* G_d, G_d *by* $2^D, 2^d$*, respectively, and assume that* $D \geq d$*. Then,*

$$d_{\mathsf{in}}(\mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B}) \leq d_{\mathsf{in}}(\mathcal{A}) + d_{\mathsf{in}}(\mathcal{B}) + D + 1;$$

$$d_{\mathsf{out}}(\mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B}) = d_{\mathsf{out}}(\mathcal{A});$$

$$\sigma(\mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B}) \geq \min\left(2\log\left(\frac{1}{\varepsilon_2}\right) + \sigma(\mathcal{A}),\, \log\left(\frac{1}{\delta_2}\right) - \mu(\mathcal{A})\right) - \mu(\mathcal{B}) - 2\log w - 6;$$

$$\mu(\mathcal{A} \overleftarrow{\bullet}_{G_D - G_d} \mathcal{B}) \leq \mu(\mathcal{A}) + \mu(\mathcal{B}) + 2.$$

7 Leveled Matrix Representations

Definition 7.1. *A* (k, w)-matrix representation *is a sequence* $\mathbf{A} = ((a_0, \mathcal{A}_0), \ldots, (a_k, \mathcal{A}_k))$ *where:*

- $a_i \geq 0$ *are real numbers and* $\mathcal{A}_i$ *are MBSs; and*
- *for every* $i \geq 1$*,* $\sigma(\mathcal{A}_i) \geq i - (i-1)\tau$*, where* $\tau = 1/(10k^2)$*.*

The matrix that is realized *by* $\mathbf{A}$ *is defined by* $\langle \mathbf{A} \rangle = \sum_{i=0}^k a_i \langle \mathcal{A}_i \rangle$*. We define the* weight *of* $\mathbf{A}$ *by* $\vartheta(\mathbf{A}) = \sum_i a_i$*.*

Remark regarding τ. Ideally, the property $\sigma(\mathcal{A}_i) \geq i-(i-1)\tau$ would have been replaced by $\sigma(\mathcal{A}_i) \geq i$ which captures in a cleaner way the fact that the smallness, or more precisely, the bound we can guarantee on the smallness, increases with i. However, the machinery we developed in Section 6 does not allow us to maintain such invariant. Thus, we are forced to introduce and work with this small relaxation.

Matrix representations capture the way in which we represent matrices. However, we will require, and maintain, some more structure. We find it useful to define this extra structure "on top" of the basic definition rather than mix them into one. We start with some preparations. For integers $k \geq g \geq 1$, define the function $\mathsf{level}_{k,g} \colon \{0, g, g+1, \ldots, k\} \to \mathbb{N}$ by

$$\mathsf{level}_{k,g}(i) = \begin{cases} 0, & i = 0; \\ 1 + \left\lfloor \log\left(\frac{i}{g}\right) \right\rfloor, & i \geq 1. \end{cases}$$

When k, g are clear from context, we omit them from the subscript and simply write $\mathsf{level}(i)$. Note that if $i, j > 0$ are such that $\mathsf{level}(i) = \mathsf{level}(j)$ then $i/2 \leq j \leq 2i$. From this point on, for simplicity, we assume that g divides k.

For ease of readability, from this point on we define the function $\omega(w) = 2\log w + 6$. When w is clear from the context, we omit it and write ω instead of $\omega(w)$. We remind the reader that all matrices considered are of order $w \times w$.

Definition 7.2. *Let k, g, w be integers such that*

$$k \geq g \geq 10(\omega + \log k). \tag{7.1}$$

*A (k, g, w)-*leveled matrix representation (LMR for short) $\mathbf{A}$ *is a (k, w)-matrix representation $\mathbf{A} = ((a_0, \mathcal{A}_0), \ldots, (a_k, \mathcal{A}_k))$ such that*

- *$\mathcal{A}_0$ is thin and $a_0 = 1$;*
- *$a_i = 0$ for all i such that $g \not| i$; and*
- *$\mu(\mathcal{A}_i) \leq i$.*

Moreover, for every $i, j \in \{0, g, \ldots, k\}$,

- *If $\mathsf{level}(i) = \mathsf{level}(j)$ then $d_{\mathsf{out}}(\mathcal{A}_i) = d_{\mathsf{out}}(\mathcal{A}_j)$; and*
- *If $\mathsf{level}(i) > \mathsf{level}(j)$ then $d_{\mathsf{out}}(\mathcal{A}_i) \geq d_{\mathsf{out}}(\mathcal{A}_j) + 10k$.*

As we care mainly about $\langle \mathbf{A} \rangle$, the matrix that is realized by $\mathbf{A}$, whenever $a_i = 0$ we also write $\mathcal{A}_i = \emptyset$.

Definition 7.3. *Let $\delta_{\mathsf{out}}, \delta_{\mathsf{in}}, \mu', \vartheta \colon \mathbb{R} \to \mathbb{R}$ be monotone non-decreasing functions. Let $\mathbf{A} = ((1, \mathcal{A}_0), \ldots, (a_k, \mathcal{A}_k))$ be a (k, g, w)-LMR. We say that:*

- $\mathbf{A}$ respects the out-function δ_{out} *if $d_{\mathsf{out}}(\mathcal{A}_i) \leq \delta_{\mathsf{out}}(i)$ for all $i \geq 0$;*

- $\mathbf{A}$ respects the in-function δ_{in} *if* $d_{\mathsf{in}}(\mathcal{A}_i) \le \delta_{\mathsf{in}}(i)$ *for all* $i > 0$*;*
- $\mathbf{A}$ respects the magnitude-function μ' *if* $\mu(\mathcal{A}_i) \le \mu'(i)$ *for all* $i > 0$*;*
- $\mathbf{A}$ respects the weight-function ϑ *if* $a_i \le \vartheta(i)$ *for all* $i > 0$*;*
- $\mathbf{A}$ *respects* $(\delta_{\mathsf{out}}, \delta_{\mathsf{in}}, \mu', \vartheta)$ *if* $\mathbf{A}$ *respects the out-function* δ_{out}*, the in-function* δ_{in}*, the magnitude-function* μ'*, and the weight-function* ϑ*.*

Remark. Note that we do not make any requirement of d_{in}, μ and ϑ for $i = 0$. This is because in some cases the functions δ_{in}, μ', ϑ that we work with are not well-defined at $i = 0$. While one can always tweak the functions appropriately, it is cumbersome and in any case, as $\mathbf{A}$ is an LMR, $a_0 = 1$ and $\mathcal{A}_0$ is thin, and so $d_{\mathsf{in}}(\mathcal{A}_0) = \mu(\mathcal{A}_0) = 0$.

We sometimes abuse notation and also use $d_{\mathsf{out}}, d_{\mathsf{in}}, \mu$ instead of introducing the notation $\delta_{\mathsf{out}}, \delta_{\mathsf{in}}, \mu'$. The meaning will always be clear from context.

8 The Family $\mathcal{F}(\mathbf{A}, \mathbf{B})$

From this point, given an integer k, we set

$$\delta = 2^{-5k}. \tag{8.1}$$

For integers n, d, let $\mathsf{BS}(n, d)$ be the balanced sampler $\mathsf{BSamp}(n, 2^{-d}, 2^{-d}) = ([n], [n], E)$ that is given by Theorem 3.9. By Theorem 3.9, the degree of $\mathsf{BS}(n, d)$ is $O(2^{3d})$. For ease of readability, we omit n and write $\mathsf{BS}(d)$ whenever n is clear from context. For integers ℓ, r, d for which $\ell \ge r/\delta^2$ let $\mathsf{US}(\ell, r, d)$ be the sampler $\mathsf{UBSamp}(\ell, r, 2^{-d}, \delta) = ([\ell], [r], E)$ that is given by Theorem 3.11. By Theorem 3.11, the degree of $\mathsf{US}(\ell, r, d)$ is $O((2^d \cdot 5k)^{c_{\mathsf{samp}}}) = O((2^d \cdot k)^{c_{\mathsf{samp}}})$. When ℓ, r are clear from context we omit them and write $\mathsf{US}(d)$.

Definition 8.1. *Let* $\mathbf{A} = ((1, \mathcal{A}_0), \ldots, (a_k, \mathcal{A}_k))$*,* $\mathbf{B} = ((1, \mathcal{B}_0), \ldots, (b_k, \mathcal{B}_k))$ *be a pair of* (k, g, w)*-LMRs. Assume that* $d_{\mathsf{out}}(\mathcal{A}_i) = d_{\mathsf{out}}(\mathcal{B}_i)$ *for all* i*. Define* $\mathcal{F}(\mathbf{A}, \mathbf{B})$ *to be the following collection of MBSs:*

1.
$$\left\{\mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0\right\} \cup \left\{\mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g) - \mathsf{BS}(2^r g)} \mathcal{B}_0 \;\middle|\; r = 1, \ldots, \log(k/g)\right\};$$

2. For every $j \in \{g, 2g, \ldots, k\}$*,*
$$\left\{\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0\right\} \cup \left\{\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g) - \mathsf{US}(2^r g)} \mathcal{A}_0 \;\middle|\; r = 0, 1, \ldots, \log(k/g)\right\};$$

3. For every $i \in \{g, 2g, \ldots, k\}$*,*
$$\left\{\mathcal{A}_i \overset{\rightarrow}{\circ}_{\mathsf{US}(g)} \mathcal{B}_0\right\} \cup \left\{\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}g) - \mathsf{US}(2^r g)} \mathcal{B}_0 \;\middle|\; r = 0, 1, \ldots, \log(k/g)\right\};$$

4. *For every* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(i) = \mathsf{level}(j)$,

$$\left\{\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j\right\} \cup \left\{\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j \;\middle|\; r = 0, 1, \ldots, \log(k/i)\right\};$$

5. *For* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(i) > \mathsf{level}(j)$,

$$\left\{\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j\right\} \cup \left\{\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j)-\mathsf{US}(2^r\cdot 8j)} \mathcal{B}_j, \;\middle|\; r = 0, 1, \ldots, \log(k/j)\right\};$$

6. *For* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(j) > \mathsf{level}(i)$,

$$\left\{\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(8i)} \mathcal{A}_i\right\} \cup \left\{\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8i)-\mathsf{US}(2^r\cdot 8i)} \mathcal{A}_i, \;\middle|\; r = 0, 1, \ldots, \log(k/i)\right\}.$$

Remark on the validity of Definition 8.1. The MBSs listed in Definition 8.1 are obtained by multiplying MBSs where the product is parameterized by a balanced or an unbalanced sampler (or delta of such). Therefore, one must verify that the MBSs that are being multiplied have d_{out} as required by the corresponding sampler. This indeed holds for all MBSs listed in Definition 8.1. Indeed,

- For all products that are parameterized by an unbalanced sampler (or by the delta of such), the requirement regarding the ratio between the sides of the sampler holds. Indeed, by the hypothesis, and since $\mathbf{A}, \mathbf{B}$ are LMRs, for every $i, j \in \{0, g, \ldots, k\}$ with $\mathsf{level}(i) > \mathsf{level}(j)$ it holds that

$$d_{\mathsf{out}}(\mathcal{A}_i) \geq d_{\mathsf{out}}(\mathcal{A}_j) + 10k = d_{\mathsf{out}}(\mathcal{B}_j) + 10k$$

  (and similarly, $d_{\mathsf{out}}(\mathcal{B}_i) \geq d_{\mathsf{out}}(\mathcal{A}_j) + 10k$). Hence, the ratio between the two sides of the sampler is bounded below by $2^{10k} = \delta^{-2}$, per Equation (8.1), as required by Theorem 3.11.

- When taking a product with balanced samplers (or the delta of such), the two sides of the samplers are of equal size, as for i, j with $\mathsf{level}(i) = \mathsf{level}(j)$ it holds that $d_{\mathsf{out}}(\mathcal{A}_i) = d_{\mathsf{out}}(\mathcal{A}_j) = d_{\mathsf{out}}(\mathcal{B}_j)$.

We set some useful notation. Let $\mathbf{A}, \mathbf{B}$ be a pair of (k, g, w)-LMRs. For $i, j \in \{0, g, 2g, \ldots, k\}$ we let $S_{i,j}$ be the sum of all matrices that are realized by MBSs in the corresponding item of Definition 8.1. Let $\mathcal{C} \in \mathcal{F}(\mathbf{A}, \mathbf{B})$ and let $i, j \in \{0, g, 2g, \ldots, k\}$ be such that $\mathcal{C}$ is obtained by taking the product of $\mathcal{A}_i$ and $\mathcal{B}_j$ when parameterized by some sampler or delta of such. We denote this corresponding indices by $i(\mathcal{C}), j(\mathcal{C})$[9].

The following claim states that the sum of all MBSs in $\mathcal{F}(\mathbf{A}, \mathbf{B})$, when weighted properly, approximates the product $\langle \mathbf{A} \rangle \langle \mathbf{B} \rangle$.

[9] By just looking at an MBS $\mathcal{C}$, $i(\mathcal{C})$ and $j(\mathcal{C})$ are not well defined but given the context of $\mathcal{C}$ belonging to $\mathcal{F}(\mathbf{A}, \mathbf{B})$, these quantities are well defined and we will use the notation only in such context.

Claim 8.2. *Let* $\mathbf{A}, \mathbf{B}$ *be a pair of* (k, g, w)*-LMRs. Then,*

$$\Big\| \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle - \sum_{i,j=0}^{k} a_i b_j S_{i,j} \Big\|_{\max} \leq 8w\vartheta(\mathbf{A})\vartheta(\mathbf{B})2^{-k}.$$

Proof. For $i = j = 0$ we have,

$$\begin{aligned} S_{0,0} &= \langle \mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0 \rangle + \sum_{r=1}^{\log(k/g)} \langle \mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g) - \mathsf{BS}(2^r g)} \mathcal{B}_0 \rangle \\ &= \langle \mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2k)} \mathcal{B}_0 \rangle, \end{aligned}$$

where the last equality follows by Claim 6.12 and by the linearity of $\overset{\rightarrow}{\bullet}$ (see Equation (6.14)). As $\mathsf{BS}(2k)$ is a $(2^{-2k}, 2^{-2k})$-sampler, Lemma 6.13 implies that $\|\langle \mathcal{A}_0 \rangle \langle \mathcal{B}_0 \rangle - S_{0,0}\|_{\max} \leq 8w2^{-2k} \leq 8w2^{-k}$.

Similarly, for every $j \in \{g, 2g, \ldots, k\}$,

$$\begin{aligned} S_{0,j} &= \langle \mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0 \rangle + \sum_{r=0}^{\log(k/g)} \langle \mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g) - \mathsf{US}(2^r g)} \mathcal{A}_0 \rangle \\ &= \langle \mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2k)} \mathcal{A}_0 \rangle. \end{aligned}$$

As $\mathsf{US}(2k)$ is a $(2^{-2k}, \delta)$-sampler, Lemma 6.13 yields $\|\langle \mathcal{A}_0 \rangle \langle \mathcal{B}_j \rangle - S_{0,j}\|_{\max} \leq 8w2^{-k}$ (Assuming that $\mu(\mathcal{B}_j) \leq j \leq k$). In the same way one can show that for $i \in \{g, 2g, \ldots, k\}$, $\|\langle \mathcal{A}_i \rangle \langle \mathcal{B}_0 \rangle - S_{i,0}\|_{\max} \leq 8w2^{-k}$. Consider $i, j \in \{g, 2g, \ldots, k\}$ with $\mathsf{level}(i) = \mathsf{level}(j)$, namely, MBSs from Item 4 of Definition 8.1. By the linearity of $\overset{\rightarrow}{\bullet}$, $S_{i,j} = \langle \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(16k)} \mathcal{B}_j \rangle$ (we assumed that $d_{\mathsf{out}}(\mathcal{A}_i) = d_{\mathsf{out}}(\mathcal{A}_j)$ whenever $\mathsf{level}(i) = \mathsf{level}(j)$) and so, by Lemma 6.13,

$$\|\langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle - S_{i,j}\|_{\max} \leq 4w2^{\frac{\mu(\mathcal{B}_j)}{2}} \cdot \left(2^{\frac{\mu(\mathcal{A}_i)}{2} - 16k} + 2^{-\frac{\sigma(\mathcal{A}_i)}{2} - 16k}\right) \leq 8w2^{-k}.$$

The same bound can be shown to hold for MBSs from Items 5,6 of Definition 8.1. We show here for Item 5. By the linearity of $\overset{\rightarrow}{\bullet}$, $S_{i,j} = \langle \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(16k)} \mathcal{B}_j \rangle$ and so, by Lemma 6.13,

$$\|\langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle - S_{i,j}\|_{\max} \leq 4w2^{\frac{\mu(\mathcal{B}_j)}{2}} \cdot \left(2^{\frac{\mu(\mathcal{A}_i)}{2} - 5k} + 2^{-\frac{\sigma(\mathcal{A}_i)}{2} - 16k}\right) \leq 8w2^{-k}.$$

Thus, altogether we established that for every $i, j \in \{0, g, 2g, \ldots, k\}$,

$$\|\langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle - S_{i,j}\|_{\max} \leq 8w2^{-k}. \tag{8.2}$$

Now,

$$\begin{aligned} \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle &= \left(\sum_{i=0}^{k} a_i \langle \mathcal{A}_i \rangle \right) \left(\sum_{j=0}^{k} b_j \langle \mathcal{B}_j \rangle \right) \\ &= \sum_{i,j=0}^{k} a_i b_j \langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle. \end{aligned}$$

Equation (8.2) together with the triangle inequality then implies

$$\Big\| \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle - \sum_{i,j=0}^{k} a_i b_j S_{i,j} \Big\|_{\max} \le \sum_{i,j=0}^{k} a_i b_j \| \langle \mathcal{A}_i \rangle \langle \mathcal{B}_j \rangle - S_{i,j} \|_{\max}$$
$$\le 8 w \vartheta(\mathbf{A}) \vartheta(\mathbf{B}) 2^{-k}.$$

□

8.1 Basic properties of the MBSs in $\mathcal{F}(\mathbf{A},\mathbf{B})$: d_{out}, d_{in}, μ, σ

In this section we give a series of claims that analyze the MBSs in $\mathcal{F}(\mathbf{A},\mathbf{B})$ in terms of their $d_{\mathsf{out}}, d_{\mathsf{in}}$, magnitude μ, and smallness σ. Throughout this section, $\mathbf{A},\mathbf{B}$ is a pair of (k,g,w)-LMRs as in Definition 8.1. We further recall that $\delta = 2^{-5k}$ per Equation (8.1) and that $\tau = 1/(10k^2)$ per Definition 7.2. We start by considering the MBSs that are given in Item 1 of Definition 8.1.

Claim 8.3. *The MBS $\mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0$ is thin and $d_{\mathsf{out}}(\mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0) \le d_{\mathsf{out}}(\mathcal{A}_0) + 7g$. Moreover, for every $r \in \{1,\dots,\log(k/g)\}$,*

$$d_{\mathsf{in}}\left(\mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0\right) \le 2^{r+3} g;$$
$$d_{\mathsf{out}}\left(\mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0\right) = d_{\mathsf{out}}(\mathcal{A}_0);$$
$$\sigma\left(\mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0\right) \ge 2^r g - \omega;$$
$$\mu\left(\mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0\right) \le 2.$$

Proof. As both $\mathcal{A}_0, \mathcal{B}_0$ are thin, Claim 6.7 implies that $\mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0$ is thin. As the sampler $\mathsf{BS}(2g)$ has degree $O(2^{6g})$ which we assume is bounded by 2^{7g}, Lemma 6.5 implies that $d_{\mathsf{out}}(\mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0) \le d_{\mathsf{out}}(\mathcal{A}_0) + 7g$, as stated. Moving to the moreover part, fix $r \in \{1,\dots,\log(k/g)\}$ and write $\mathcal{C} = \mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0$. By Lemma 6.18, whose hypothesis is satisfied per Equation (7.1), and using the fact that $\mathcal{A}_0, \mathcal{B}_0$ are thin, we get $d_{\mathsf{in}}(\mathcal{C}) = 3 \cdot 2^{r+1} g + O(1)$, which yields the stated bound. The assertion regarding $d_{\mathsf{out}}(\mathcal{C})$ readily follows by Definition 6.16. As for the smallness, Lemma 6.18 implies that $\sigma(\mathcal{C}) \ge 2^r g - \omega$. Lastly, as $\mu(\mathcal{A}_0) = \mu(\mathcal{B}_0) = 0$, Lemma 6.18 implies that $\mu(\mathcal{C}) \le 2$. □

Claim 8.4. *For every $j \in \{g, 2g, \dots, k\}$,*

$$d_{\mathsf{in}}\left(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0\right) = d_{\mathsf{in}}(\mathcal{B}_j);$$
$$d_{\mathsf{out}}\left(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0\right) \le d_{\mathsf{out}}(\mathcal{B}_j) + 2c_{\mathsf{samp}} g;$$
$$\sigma\left(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0\right) \ge \sigma(\mathcal{B}_j);$$
$$\mu\left(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0\right) \le \mu(\mathcal{B}_j).$$

Moreover, for every $r \in \{0, 1, \ldots, \log(k/g)\}$,

$$d_{\mathsf{in}}\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0\right) \le d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+2} g;$$
$$d_{\mathsf{out}}\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0\right) = d_{\mathsf{out}}(\mathcal{B}_j);$$
$$\sigma\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0\right) \ge \min\left(\sigma(\mathcal{B}_j) + 2^r g,\, k+1\right);$$
$$\mu\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0\right) \le \mu(\mathcal{B}_j) + 2.$$

Proof. As $\mathcal{A}_0$ is thin, Lemma 6.6 implies the assertions regarding $d_{\mathsf{in}}(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0)$, $\sigma(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0)$ and $\mu(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0)$. The bound $d_{\mathsf{out}}(\mathcal{B}_j \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0) \le d_{\mathsf{out}}(\mathcal{B}_j) + 2c_{\mathsf{samp}} g$ follows as the degree of $\mathsf{US}(g)$ is $O((2^g \cdot k)^{c_{\mathsf{samp}}}) \le 2^{2c_{\mathsf{samp}} g}$, where we used the fact that $g \ge 10 \log k$.

Moving to the moreover part of the claim, denote $\mathcal{C} = \mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0$. Recall that the degree of $\mathsf{US}(2^{r+1}g)$ is $O((2^{2^{r+1}g} \cdot k)^{c_{\mathsf{samp}}}) \le 2^{c_{\mathsf{samp}} 2^{r+2} g}$[10] per our assumption $g \ge 10 \log k$. Lemma 6.20 then implies that $d_{\mathsf{in}}(\mathcal{C}) \le d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+2} g$. The assertion regarding $d_{\mathsf{out}}(\mathcal{C})$ follows by definition, and the bound on the magnitude follows by Lemma 6.20 and since $\mathcal{A}_0$ is thin. As for the smallness, by Lemma 6.20,

$$\begin{aligned}\sigma(\mathcal{C}) &\ge \min\left(2^{r+1} g + \sigma(\mathcal{B}_j),\, 5k - \mu(\mathcal{B}_j)\right) - \omega \\ &\ge \min\left(\sigma(\mathcal{B}_j) + 2^r g,\, k+1\right),\end{aligned}$$

where in the above inequality we used the hypothesis $g \ge \omega$ and that $\mu(\mathcal{B}_j) + \omega \le j + \omega \le 2k$. □

Claim 8.5. *For every* $i \in \{g, 2g, \ldots, k\}$,

$$d_{\mathsf{in}}\left(\mathcal{A}_i \overset{\rightarrow}{\circ}_{\mathsf{US}(g)} \mathcal{B}_0\right) = d_{\mathsf{in}}(\mathcal{A}_i);$$
$$d_{\mathsf{out}}\left(\mathcal{A}_i \overset{\rightarrow}{\circ}_{\mathsf{US}(g)} \mathcal{B}_0\right) \le d_{\mathsf{out}}(\mathcal{A}_i) + 2c_{\mathsf{samp}} g;$$
$$\sigma\left(\mathcal{A}_i \overset{\rightarrow}{\circ}_{\mathsf{US}(g)} \mathcal{B}_0\right) \ge \sigma(\mathcal{A}_i);$$
$$\mu\left(\mathcal{A}_i \overset{\rightarrow}{\circ}_{\mathsf{US}(g)} \mathcal{B}_0\right) \le \mu(\mathcal{A}_i).$$

Moreover, for every $r \in \{0, 1, \ldots, \log(k/g)\}$,

$$d_{\mathsf{in}}\left(\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{B}_0\right) \le d_{\mathsf{in}}(\mathcal{A}_i) + c_{\mathsf{samp}} 2^{r+2} g;$$
$$d_{\mathsf{out}}\left(\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{B}_0\right) = d_{\mathsf{out}}(\mathcal{A}_i);$$
$$\sigma\left(\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{B}_0\right) \ge \min\left(\sigma(\mathcal{A}_i) + 2^r g,\, k+1\right);$$
$$\mu\left(\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{B}_0\right) \le \mu(\mathcal{A}_i) + 2.$$

[10] g is larger than a large enough constant.

The proof of Claim 8.5 is similar to the proof of Claim 8.4 and we omit it.

Claim 8.6. *For every* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(i) = \mathsf{level}(j)$*,*

$$\begin{aligned} d_{\mathsf{in}}\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j\right) &\le d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 25i; \\ d_{\mathsf{out}}\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j\right) &= d_{\mathsf{out}}(\mathcal{A}_i); \\ \sigma\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j\right) &\ge \min(\sigma(\mathcal{A}_i), i) + \min(\sigma(\mathcal{B}_j), j) - \tau; \\ \mu\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j\right) &\le \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j). \end{aligned}$$

Moreover, for every $r \in \{0, 1, \ldots, \log(k/i)\}$*,*

$$\begin{aligned} d_{\mathsf{in}}\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j\right) &\le d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 50i \cdot 2^r; \\ d_{\mathsf{out}}\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j\right) &= d_{\mathsf{out}}(\mathcal{A}_i); \\ \sigma\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j\right) &\ge 2^{r+2}i; \\ \mu\left(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j\right) &\le \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2. \end{aligned}$$

Proof. We wish to invoke Lemma 6.14. Thus, we first must verify that Equation (6.12) holds for $\lambda(\mathcal{A}_i) = \min(\sigma(\mathcal{A}_i), i)$ and $\lambda(\mathcal{B}_j) = \min(\sigma(\mathcal{B}_j), j)$. As $\mathsf{BS}(8i)$ is a $(2^{-8i}, 2^{-8i})$-sampler, it suffices to check that

$$8i \ge i + j + \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + \log(1/\tau) + 3.$$

As $\mathsf{level}(i) = \mathsf{level}(j)$ we have $j \le 2i$. Since $\mu(\mathcal{A}_i) \le i$ and $\mu(\mathcal{B}_j) \le j$ it holds that

$$i + j + \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + \log(1/\tau) + 3 \le 6i + 2\log k + 7,$$

where we have used the remark regarding σ that appears after Definition 7.1. As $i \ge g \ge 10\log k$, the RHS is indeed bounded by $8i$. Lemma 6.14 then implies the assertion regarding the smallness and magnitude of $\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$. The assertion regarding $d_{\mathsf{out}}(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j)$ follows by Definition 6.8. Since the degree of $\mathsf{BS}(8i)$ is $O(2^{24i})$ which we assume is bounded by 2^{25i}, the bound on $d_{\mathsf{in}}(\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j)$ follows.

Fix $r \in \{0, 1, \ldots, \log(k/i)\}$ and write $\mathcal{C} = \mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j$. Recall that the degree of $\mathsf{BS}(2^{r+1} \cdot 8i)$ is $O(2^{3\cdot 2^{r+1}\cdot 8i}) \le 2^{49i\cdot 2^r}$. Therefore, Lemma 6.18 implies the asserted bound on $d_{\mathsf{in}}(\mathcal{C})$. The bound on $d_{\mathsf{out}}(\mathcal{C})$ follows by Definition 6.16, and the bound on $\mu(\mathcal{C})$ readily follows by Lemma 6.18. As for the smallness, by Lemma 6.18,

$$\begin{aligned} \sigma(\mathcal{C}) &\ge \min\left(\sigma(\mathcal{A}_i) + 2^{r+4}i,\, 2^{r+3}i - \mu(\mathcal{A}_i)\right) - \mu(\mathcal{B}_j) - \omega \\ &= 2^{r+3}i - \mu(\mathcal{A}_i) - \mu(\mathcal{B}_j) - \omega \\ &\ge 2^{r+3}i - 4i \\ &\ge 2^{r+2}i, \end{aligned}$$

where we used the fact that $\sigma(\mathcal{A}_i) \geq i-(i-1)\tau$, $\mu(\mathcal{A}_i) \leq i$, $\mu(\mathcal{B}_j) \leq j \leq 2i$ which follows as $\mathsf{level}(i) = \mathsf{level}(j)$, and that $i \geq g \geq \omega$. $\square$

Claim 8.7. *For every $i, j \in \{g, 2g, \ldots, k\}$ such that $\mathsf{level}(i) > \mathsf{level}(j)$,*

$$\begin{aligned} d_{\mathsf{in}}\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j\right) &\leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 9c_{\mathsf{samp}} j; \\ d_{\mathsf{out}}\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j\right) &= d_{\mathsf{out}}(\mathcal{A}_i); \\ \sigma\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j\right) &\geq \min(\sigma(\mathcal{A}_i), i) + \min(\sigma(\mathcal{B}_j), j) - \tau; \\ \mu\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j\right) &\leq \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j). \end{aligned}$$

Moreover, for every $r \in \{0, 1, \ldots, \log(k/j)\}$,

$$\begin{aligned} d_{\mathsf{in}}\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j) - \mathsf{US}(2^r \cdot 8j)} \mathcal{B}_j\right) &\leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+5} j; \\ d_{\mathsf{out}}\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j) - \mathsf{US}(2^r \cdot 8j)} \mathcal{B}_j\right) &= d_{\mathsf{out}}(\mathcal{A}_i); \\ \sigma\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j) - \mathsf{US}(2^r \cdot 8j)} \mathcal{B}_j\right) &\geq \min\left(\sigma(\mathcal{A}_i) + 2^{r+3} j,\ k+1\right); \\ \mu\left(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j) - \mathsf{US}(2^r \cdot 8j)} \mathcal{B}_j\right) &\leq \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2. \end{aligned}$$

Proof. Recall that $\mathsf{US}(8j)$ is a $(2^{-8j}, \delta)$-sampler where $\delta = 2^{-5k}$. To invoke Lemma 6.14, we must first verify that Equation (6.12) holds for $\lambda(\mathcal{A}_i) = \min(\sigma(\mathcal{A}_i), i)$ and $\lambda(\mathcal{B}_j) = \min(\sigma(\mathcal{B}_j), j)$, namely,

$$\begin{aligned} &8j \geq j + \mu(\mathcal{B}_j) + \log(1/\tau) + 3, \\ &5k \geq i + j + \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + \log(1/\tau) + 3. \end{aligned}$$

The first inequality holds as

$$j + \mu(\mathcal{B}_j) + \log(1/\tau) + 3 \leq 2j + \log(10k^2) + 3,$$

which is indeed bounded above by $8j$ as $j \geq g \geq 10 \log k$ (see the remark regarding σ that appears after Definition 7.1). As for the second inequality,

$$\begin{aligned} i + j + \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + \log(1/\tau) + 3 &\leq 2i + 2j + \log(1/\tau) + 3 \\ &\leq 4k + \log(10k^2) + 3 \\ &\leq 5k. \end{aligned}$$

Thus, the asserted bounds regarding the smallness and magnitude of $\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j$ follow by Lemma 6.14. That $d_{\mathsf{out}}(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j) = d_{\mathsf{out}}(\mathcal{A}_i)$ follows by Definition 6.8. As for $d_{\mathsf{in}}(\mathcal{A}_i \vec{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j)$, recall that the degree of the sampler $\mathsf{US}(8j)$ is $O((2^{8j} \cdot k)^{c_{\mathsf{samp}}}) \leq 2^{9c_{\mathsf{samp}} j}$, where the inequality follows as $j \geq g \geq 10 \log k$. The bound on $d_{\mathsf{in}}(\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j)$ then follows by Definition 6.8.

Write $\mathcal{C} = \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1} \cdot 8j) - \mathsf{US}(2^r \cdot 8j)} \mathcal{B}_j$. The bounds on $d_{\mathsf{out}}(\mathcal{C}), \mu(\mathcal{C})$ readily follow by Lemma 6.18. As $\mathsf{US}(2^{r+1} \cdot 8j)$ has degree $O((2^{2^{r+1} \cdot 8j} \cdot k)^{c_{\mathsf{samp}}})$, Lemma 6.18 implies the stated bound on $d_{\mathsf{in}}(\mathcal{C})$. As for $\sigma(\mathcal{C})$, by Lemma 6.18,

$$
\begin{aligned}
\sigma(\mathcal{C}) &\geq \min\left(2^{r+4} j + \sigma(\mathcal{A}_i),\, 5k - \mu(\mathcal{A}_i)\right) - \mu(\mathcal{B}_j) - \omega \\
&\geq \min\left(\sigma(\mathcal{A}_i) + 2^{r+3} j,\, k+1\right),
\end{aligned}
$$

which completes the proof. □

Claim 8.8. *For every $i, j \geq g$ such that $\mathsf{level}(i) < \mathsf{level}(j)$,*

$$
\begin{aligned}
d_{\mathsf{in}}\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(8i)} \mathcal{A}_i\right) &\leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 9 c_{\mathsf{samp}} i; \\
d_{\mathsf{out}}\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(8i)} \mathcal{A}_i\right) &= d_{\mathsf{out}}(\mathcal{B}_j); \\
\sigma\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(8i)} \mathcal{A}_i\right) &\geq \min(\sigma(\mathcal{A}_i), i) + \min(\sigma(\mathcal{B}_j), j) - \tau; \\
\mu\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(8i)} \mathcal{A}_i\right) &\leq \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j).
\end{aligned}
$$

Moreover, for every $r \in \{0, 1, \ldots, \log(k/i)\}$,

$$
\begin{aligned}
d_{\mathsf{in}}\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1} \cdot 8i) - \mathsf{US}(2^r \cdot 8i)} \mathcal{A}_i\right) &\leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+5} i; \\
d_{\mathsf{out}}\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1} \cdot 8i) - \mathsf{US}(2^r \cdot 8i)} \mathcal{A}_i\right) &= d_{\mathsf{out}}(\mathcal{B}_j); \\
\sigma\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1} \cdot 8i) - \mathsf{US}(2^r \cdot 8i)} \mathcal{A}_i\right) &\geq \min\left(\sigma(\mathcal{B}_j) + 2^{r+3} i,\, k+1\right); \\
\mu\left(\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1} \cdot 8i) - \mathsf{US}(2^r \cdot 8i)} \mathcal{A}_i\right) &\leq \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2.
\end{aligned}
$$

The proof of Claim 8.8 is similar to the proof of Claim 8.7 and we omit the details.

8.2 The slices of $\mathcal{F}(\mathbf{A}, \mathbf{B})$

In this section we define the s-slice of $\mathcal{F}(\mathbf{A}, \mathbf{B})$ that, roughly speaking, consists of all MBSs $\mathcal{C} \in \mathcal{F}(\mathbf{A}, \mathbf{B})$ for which s is the best lower bound we can give on the $\sigma(\mathcal{C})$.

Definition 8.9. *Let $\mathbf{A} = ((1, \mathcal{A}_0), \ldots, (a_k, \mathcal{A}_k))$, $\mathbf{B} = ((1, \mathcal{B}_0), \ldots, (b_k, \mathcal{B}_k))$ be a pair of (k, g, w)-LMRs. Let $s \in \{0, 1, \ldots, k\}$. Define $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$ to be the following collection of MBSs:*

1. *$\mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0$ if $s = 0$, and $\mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1} g) - \mathsf{BS}(2^r g)} \mathcal{B}_0$ if there is $r \in \{1, \ldots, \log(k/g)\}$ such that $s = (2^r - 1) g$;*
2. *$\mathcal{B}_s \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0$, and $\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1} g) - \mathsf{US}(2^r g)} \mathcal{A}_0$ for all $r \in \{0, 1, \ldots, \log(k/g)\}$ and $j \in \{g, 2g, \ldots, k\}$ such that $j + 2^r g = s$;*

3. $\mathcal{A}_s \overset{\rightarrow}{\circ}_{\mathsf{US}(g)} \mathcal{B}_0$, *and* $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{B}_0$ *for all* $r \in \{0,1,\ldots,\log(k/g)\}$ *and* $i \in \{g, 2g, \ldots, k\}$ *such that* $i + 2^r g = s$*;*

4. $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$ *for every* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(i) = \mathsf{level}(j)$ *and* $i + j = s$*, as well as* $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r\cdot 8i)} \mathcal{B}_j$ *for every* $i, j \in \{g, 2g, \ldots, k\}$ *and* $r \in \{0,1,\ldots,\log(k/i)\}$ *such that* $\mathsf{level}(i) = \mathsf{level}(j)$ *and* $2^{r+2} i = s$*.*

5. $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j$ *for every* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(i) > \mathsf{level}(j)$ *and* $i + j = s$*, as well as* $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j)-\mathsf{US}(2^r\cdot 8j)} \mathcal{B}_j$ *for every* $i, j \in \{g, 2g, \ldots, k\}$ *and* $r \in \{0,1,\ldots,\log(k/j)\}$ *such that* $\mathsf{level}(i) > \mathsf{level}(j)$ *and* $i + 2^{r+3} j = s$*.*

6. $\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(8i)} \mathcal{A}_i$ *for every* $i, j \in \{g, 2g, \ldots, k\}$ *such that* $\mathsf{level}(j) > \mathsf{level}(i)$ *and* $i + j = s$*, as well as* $\mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8i)-\mathsf{US}(2^r\cdot 8i)} \mathcal{A}_i$ *for every* $i, j \in \{g, 2g, \ldots, k\}$ *and* $r \in \{0,1,\ldots,\log(k/i)\}$ *such that* $\mathsf{level}(j) > \mathsf{level}(i)$ *and* $j + 2^{r+3} i = s$*.*

We start by analyzing the slices of $\mathcal{F}(\mathbf{A}, \mathbf{B})$.

Claim 8.10. *Let* $\mathbf{A}, \mathbf{B}$ *be a pair of* (k, g, w)*-LMRs. Then, for every* $s : g \not| s$*,* $\mathcal{F}_s(\mathbf{A}, \mathbf{B}) = \emptyset$*.*

Proof. By inspecting the MBSs in Definition 8.9, one can readily see that the MBSs in $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$ are products of MBSs $\mathcal{A}_i$, $\mathcal{B}_j$ such that $ai + bj + cg = s$ for some integers a, b, c. As $\mathbf{A}, \mathbf{B}$ are (k, g, w)-LMRs, both i, j are divisible by g and so s is also divisible by g. Put differently, for s not divisible by g, the collection $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$ is empty. □

Claim 8.11. *Let* $\mathbf{A}, \mathbf{B}$ *be a pair of* (k, g, w)*-LMRs. Then, for every* $s \in \{g, 2g, \ldots, k\}$ *and* $\mathcal{C} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$*, it holds that*

$$\sigma(\mathcal{C}) \geq s - (s-1)\tau.$$

Moreover,

$$\{\mathcal{C} \in \mathcal{F}(\mathbf{A}, \mathbf{B}) \mid \sigma(\mathcal{C}) \leq k\} \subseteq \bigcup_{s=\{0,g,2g,\ldots,k\}} \mathcal{F}_s(\mathbf{A}, \mathbf{B}). \tag{8.3}$$

Proof. Consider the MBS $\mathcal{C} = \mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0$ where $r \in \{1, \ldots, \log(k/g)\}$ is such that $s = (2^r - 1)g$. By Claim 8.3, $\sigma(\mathcal{C}) \geq s$ as desired. By Claim 8.4, $\sigma(\mathcal{B}_s \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0) \geq \sigma(\mathcal{B}_s)$. As $\mathbf{B}$ is an LMR, $\sigma(\mathcal{B}_s) \geq s - (s-1)\tau$, as desired. Consider the MBS $\mathcal{C} = \mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0$ for $r \in \{0, 1, \ldots, \log(k/g)\}$ and $j \in \{g, 2g, \ldots, k\}$ such that $j + 2^r g = s$. By Claim 8.4, $\sigma(\mathcal{C}) \geq \min(\sigma(\mathcal{B}_j) + 2^r g,\ k+1)$. If $\sigma(\mathcal{B}_j) + 2^r g > k + 1$ then $\sigma(\mathcal{C}) > k > s$ and we are done. Otherwise, using that $\mathbf{B}$ is an LMR,

$$\begin{aligned}\sigma(\mathcal{C}) &\geq \sigma(\mathcal{B}_j) + 2^r g \\ &\geq j - (j-1)\tau + 2^r g \\ &\geq s - (s-1)\tau.\end{aligned}$$

A similar argument can be used to prove the assertion for MBSs from Item 3 of Definition 8.9 and we omit the details.

Consider now the MBS $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$ where $i, j \in \{g, 2g, \ldots, k\}$ are such that $\mathsf{level}(i) = \mathsf{level}(j)$ and $i + j = s$. By Claim 8.6 and since $\mathbf{A}, \mathbf{B}$ are LMRs,

$$\begin{aligned}\sigma(\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j) &\geq \min(\sigma(\mathcal{A}_i), i) + \min(\sigma(\mathcal{B}_j), j) - \tau \\ &\geq i - (i-1)\tau + j - (j-1)\tau - \tau \\ &= s - (s-1)\tau,\end{aligned}$$

as stated. Let $\mathcal{C} = \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1} \cdot 8i) - \mathsf{BS}(2^r \cdot 8i)} \mathcal{B}_j$ where $i, j \in \{g, 2g, \ldots, k\}$ and $r \in \{0, 1, \ldots, \log(k/i)\}$ are such that $\mathsf{level}(i) = \mathsf{level}(j)$ and $2^{r+2} i = s$. By Claim 8.6, $\sigma(\mathcal{C}) \geq 2^{r+2} i = s$, as desired. A similar argument can be used for the remaining MBSs in $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$, for which $\mathsf{level}(i) \neq \mathsf{level}(j)$, and we omit the details.

Moving to the moreover part, a careful inspection of Definition 8.1, Definition 8.9 and the claims in Section 8.1 yields that we did not "leave out" any MBS of smallness not larger than k in Definition 8.9. This, together with the fact that $\sigma(\mathcal{C}) \geq s - (s-1)\tau$ for all $s \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$, yields

$$\{\mathcal{C} \in \mathcal{F}(\mathbf{A}, \mathbf{B}) \mid \sigma(\mathcal{C}) \leq k\} \subseteq \bigcup_{s=0}^{k} \mathcal{F}_s(\mathbf{A}, \mathbf{B}).$$

We omit the details of the proof. Equation (8.3) then follows by Claim 8.10. □

Claim 8.12. *Let $\mathbf{A}, \mathbf{B}$ be a pair of (k, g, w)-LMRs. Then, for every $s \in \{g, 2g, \ldots, k\}$, $|\mathcal{F}_s(\mathbf{A}, \mathbf{B})| = O((s/g)^2)$.*

Proof. Clearly, Item 1 in Definition 8.9 contributes at most one MBS to $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$. As for Item 2, for every fixed j, the number of MBSs contributed is one. As $\mathbf{B}$ is an LMR, we only need to consider j that is divisible by g and so the total number of MBSs contributed by Item 2 is $O((s/g))$. As $\mathbf{A}$ is also an LMR, a similar argument gives the same bound on the number of MBSs coming from Item 3.

Moving on to Item 4, the number of MBSs of the form $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$ is equal to the number of solutions to $i + j = s$. As i, j are divisible by g, the number of solutions is $O(s/g)$. The remaining MBSs in Item 4 are of the form $\mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1} \cdot 8i) - \mathsf{BS}(2^r \cdot 8i)} \mathcal{B}_j$ where $i, j \in \{g, 2g, \ldots, k\}$, $\mathsf{level}(i) = \mathsf{level}(j)$, and $r \in \{0, 1, \ldots, \log(k/i)\}$ is such that $2^{r+2} i = s$. As $i \geq g$ and i is divisible by g, the number of (i, r) pairs the satisfy the latter equation is $O(s/g)$. For every such (i, r) pair, the number of j's for which $\mathsf{level}(i) = \mathsf{level}(j)$ is $O(i/g)$. Indeed the latter constraint implies that $i/2 \leq j \leq 2i$, and j is divisible by g. Summing over all these values, we conclude that the total number of MBSs of the latter form is $O((s/g)^2)$. Similar arguments can be used to bound the number of MBSs from Item 5 and Item 6 by $O((s/g)^2)$ and we omit the details. □

8.3 Analyzing d_{out}, d_{in}, μ of the slices of $\mathcal{F}(\mathbf{A}, \mathbf{B})$

In this section we further analyze the MBSs in $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$ based on the calculations done in Section 8.1. We start by analyzing $d_{\mathsf{in}}(\mathcal{C})$ for MBSs $\mathcal{C} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$. Then, in Claim 8.14 and Claim 8.15, we analyze $d_{\mathsf{out}}(\mathcal{C})$ and $\mu(\mathcal{C})$, respectively.

Claim 8.13. *Let $c_{\mathsf{in}} = 100c_{\mathsf{samp}}$, where $c_{\mathsf{samp}} \geq 1$ is the constant from Theorem 3.11. Assume that $\mathbf{A}, \mathbf{B}$ is a pair of (k, g, w)-LMRs that respect the in-function $d_{\mathsf{in}}(i) = c_{\mathsf{in}} i \log i$. Then, for every $s \in \{g, 2g, \ldots, k\}$ and $\mathcal{C} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$, $d_{\mathsf{in}}(\mathcal{C}) \leq c_{\mathsf{in}} s \log s$.*

Proof. Consider the MBS $\mathcal{C} = \mathcal{A}_0 \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0$ with $s = (2^r - 1)g$, assuming such r exists, as defined in Item 1 of Definition 8.9. By Claim 8.3, $d_{\mathsf{in}}(\mathcal{C}) \leq 2^{r+3}g$. It is therefore suffices to show that

$$2^{r+3}g \leq c_{\mathsf{in}}(2^r - 1)g \log((2^r - 1)g),$$

which holds as $c_{\mathsf{in}} \geq 16$.

Moving to Item 2 of Definition 8.9, consider the MBS $\mathcal{B}_s \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0$. By Claim 8.4, $d_{\mathsf{in}}(\mathcal{B}_s \overset{\leftarrow}{\circ}_{\mathsf{US}(g)} \mathcal{A}_0) = d_{\mathsf{in}}(\mathcal{B}_s)$ which by the hypothesis is bounded above by $c_{\mathsf{in}} s \log s$, as desired. Now, let $\mathcal{C} = \mathcal{B}_j \overset{\leftarrow}{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0$ where $r \in \{0, 1, \ldots, \log(k/g)\}$ and j are such that $s = j + 2^r g$. By Claim 8.4, $d_{\mathsf{in}}(\mathcal{C}) \leq d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+2} g$. It is therefore suffices to prove that

$$d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+2} g \leq c_{\mathsf{in}}(j + 2^r g) \log(j + 2^r g).$$

As $\mathbf{B}$ respects the in-function $d_{\mathsf{in}}(j) = c_{\mathsf{in}} j \log j$, it suffices to show that $c_{\mathsf{samp}} 2^{r+2} g \leq c_{\mathsf{in}} 2^r g$, which holds by our choice of c_{in}. A similar calculation, using Claim 8.5, can be applied for analyzing the MBSs that are given by Item 3 of Definition 8.1. We omit the details.

Take $i, j \in \{g, 2g, \ldots, k\}$ with $\mathsf{level}(i) = \mathsf{level}(j)$ such that $i + j = s$. Consider the MBS $\mathcal{C} = \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$. By Claim 8.6, $d_{\mathsf{in}}(\mathcal{C}) \leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 25i$, and so we ought to show that

$$c_{\mathsf{in}} i \log i + c_{\mathsf{in}} j \log j + 25i \leq c_{\mathsf{in}}(i + j) \log(i + j).$$

Observe that it suffices to prove that the above equation holds for $i \geq j$. Rearranging, and using the fact that $j \leq i \leq k$, it suffices to verify that

$$25i \leq c_{\mathsf{in}} i \log\left(1 + \frac{j}{i}\right).$$

As $\mathsf{level}(i) = \mathsf{level}(j)$, $j \geq i/2$ and so one only needs to verify that $25i \leq c_{\mathsf{in}} i/2$, which holds as $c_{\mathsf{in}} \geq 50$.

Consider an MBS of the form $\mathcal{C} = \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{BS}(2^{r+1} \cdot 8i)-\mathsf{BS}(2^r \cdot 8i)} \mathcal{B}_j$ where $i, j \in \{g, 2g, \ldots, k\}$ and $r \in \{0, 1, \ldots, \log(k/i)\}$ are such that $\mathsf{level}(i) = \mathsf{level}(j)$ and $2^{r+2} i = s$. By Claim 8.6, $d_{\mathsf{in}}(\mathcal{C}) \leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 50i \cdot 2^r$. Therefore, we ought to prove that

$$c_{\mathsf{in}} i \log i + c_{\mathsf{in}} j \log j + 50i \cdot 2^r \leq c_{\mathsf{in}} 2^{r+2} i \log(2^{r+2} i).$$

As $\mathsf{level}(i) = \mathsf{level}(j)$, $j \leq 2i$, and so it suffices to verify that

$$3c_{\mathsf{in}} i \log(2i) + 50i \cdot 2^r \leq c_{\mathsf{in}} 2^{r+2} i \log i$$

which holds since $c_{\mathsf{in}} \geq 50$ and $r \geq 0$.

Moving on to Item 5 of Definition 8.9, consider $i, j \in \{g, 2g, \ldots, k\}$ such that $\mathsf{level}(i) > \mathsf{level}(j)$ and $i+j = s$. Let $\mathcal{C} = \mathcal{A}_i \overset{\to}{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j$. By Claim 8.7, $d_{\mathsf{in}}(\mathcal{C}) \leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + 9c_{\mathsf{samp}}j$. It is therefore suffices to show that

$$c_{\mathsf{in}} i \log i + c_{\mathsf{in}} j \log j + 9c_{\mathsf{samp}} j \leq c_{\mathsf{in}} (i+j) \log(i+j).$$

Rearranging, it suffices to verify that

$$9c_{\mathsf{samp}} j \leq c_{\mathsf{in}} i \log \left(1 + \frac{j}{i}\right).$$

Using the inequality $\log_2(1+x) \geq x/(1+x)$ which holds for all $x \geq 0$, it suffices to prove that

$$9c_{\mathsf{samp}} j \leq c_{\mathsf{in}} \frac{ij}{i+j}.$$

The above inequality holds as $i \geq j$ and $c_{\mathsf{in}} \geq 18c_{\mathsf{samp}}$.

Consider now an MBS of the form $\mathcal{C} = \mathcal{A}_i \overset{\to}{\bullet}_{\mathsf{US}(2^{r+1} \cdot 8j) - \mathsf{US}(2^r \cdot 8j)} \mathcal{B}_j$ where $i, j \in \{g, 2g, \ldots, k\}$ and $r \in \{0, 1, \ldots, \log(k/j)\}$ are such that $\mathsf{level}(i) > \mathsf{level}(j)$ and $i + 2^{r+3}j = s$. By Claim 8.7, $d_{\mathsf{in}}(\mathcal{C}) \leq d_{\mathsf{in}}(\mathcal{A}_i) + d_{\mathsf{in}}(\mathcal{B}_j) + c_{\mathsf{samp}} 2^{r+5} j$. Hence, we ought to prove that

$$c_{\mathsf{in}} i \log i + c_{\mathsf{in}} j \log j + c_{\mathsf{samp}} 2^{r+5} j \leq c_{\mathsf{in}} (i + 2^{r+3} j) \log\left(i + 2^{r+3} j\right).$$

Rearranging, it is sufficient to show that

$$c_{\mathsf{in}} j \log j + c_{\mathsf{samp}} 2^{r+5} j \leq c_{\mathsf{in}} 2^{r+3} j \log i.$$

which readily follows. The remaining MBSs in $\mathcal{F}_s(\mathbf{A}, \mathbf{B})$, given by Item 6, follow a similar analysis and we omit the details. □

In the following claim we turn to analyze $d_{\mathsf{out}}(\mathcal{C})$ for MBSs $\mathcal{C} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$.

Claim 8.14. *Let $\mathbf{A}, \mathbf{B}$ be a pair of (k, g, w)-LMRs that respect the out-function $d_{\mathsf{out}}(i) = 10k \cdot \mathsf{level}(i) + d$ for some integer d. Then, for every $s \in \{0, g, 2g, \ldots, k\}$ and MBS $\mathcal{C} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$,*

$$d_{\mathsf{out}}(\mathcal{C}) \leq 10k \cdot \mathsf{level}(s) + d + 7c_{\mathsf{samp}} g.$$

Proof. By inspecting the claims in Section 8.1, one can verify that if $\mathcal{C} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$ is such that both $i(\mathcal{C}), j(\mathcal{C})$ are non-zero then $d_{\mathsf{out}}(\mathcal{C}) = \max\left(d_{\mathsf{out}}(\mathcal{A}_i), d_{\mathsf{out}}(\mathcal{B}_j)\right)$ whereas $s \geq \max(i, j)$ in which case the proof readily follows. Hence, we only need to consider $\mathcal{C}$ such that at least one of $i(\mathcal{C}), j(\mathcal{C})$ is zero. Consider the MBS $\mathcal{A}_0 \overset{\to}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0$. By Claim 8.3,

$$d_{\mathsf{out}}(\mathcal{A}_0 \overset{\to}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0) \leq d_{\mathsf{out}}(\mathcal{A}_0) + 7g \leq d + 7g.$$

As $c_{\mathsf{samp}} \geq 1$ and $\mathsf{level}(0) = 0$, the proof for this MBS follows. The assertion for MBSs of the form $\mathcal{C} = \mathcal{A}_0 \overset{\to}{\bullet}_{\mathsf{BS}(2^{r+1}g) - \mathsf{BS}(2^r g)} \mathcal{B}_0$ readily follows as by Claim 8.3, $d_{\mathsf{out}}(\mathcal{C}) = d_{\mathsf{out}}(\mathcal{A}_0) \leq d$.

Moving to Item 2 of Definition 8.9, consider the MBS $\mathcal{B}_s \overleftarrow{\circ}_{\mathsf{US}(g)} \mathcal{A}_0$. By Claim 8.4,

$$\begin{aligned} d_{\mathsf{out}}(\mathcal{B}_s \overleftarrow{\circ}_{\mathsf{US}(g)} \mathcal{A}_0) &\le d_{\mathsf{out}}(\mathcal{B}_s) + 2c_{\mathsf{samp}}g \\ &\le 10k \cdot \mathsf{level}(s) + d + 2c_{\mathsf{samp}}g, \end{aligned}$$

as desired. Let $\mathcal{C} = \mathcal{B}_j \overleftarrow{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0$ where $r \in \{0,1,\dots,\log(k/g)\}$ and $j \in \{g,2g,\dots,k\}$ are such that $j + 2^r g = s$. By Claim 8.4, $d_{\mathsf{out}}(\mathcal{C}) = d_{\mathsf{out}}(\mathcal{B}_j)$ which together with the fact that $s \ge j$, completes the proof for $\mathcal{C}$. A similar argument proves the claim for MBSs from Item 3 of Definition 8.9 and we omit the details. □

Claim 8.15. *Let $\mathbf{A}, \mathbf{B}$ be a pair of (k,g,w)-LMRs that respect the magnitude-function $\mu(i) = 2i/g$. Then, for every $s \in \{0,g,2g,\dots,k\}$ and MBS $\mathcal{C} \in \mathcal{F}_s(\mathbf{A},\mathbf{B})$, it holds that $\mu(\mathcal{C}) \le 2s/g$.*

Proof. By Claim 8.3, the MBS $\mathcal{A}_0 \overrightarrow{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0$ is thin and so the assertion readily follows for it. Consider the MBS $\mathcal{C} = \mathcal{A}_0 \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}g)-\mathsf{BS}(2^r g)} \mathcal{B}_0$ where $r \in \{1,\dots,\log(k/g)\}$ is such that $s = (2^r - 1)g$. The assertion for $\mathcal{C}$ follows as by Claim 8.3, $\mu(\mathcal{C}) \le 2$.

By Claim 8.4, $\mu(\mathcal{B}_s \overleftarrow{\circ}_{\mathsf{US}(g)} \mathcal{A}_0) \le \mu(\mathcal{B}_s)$ and so the claim readily follows in this case. Now, consider the MBS $\mathcal{C} = \mathcal{B}_j \overleftarrow{\bullet}_{\mathsf{US}(2^{r+1}g)-\mathsf{US}(2^r g)} \mathcal{A}_0$ where $r \in \{0,1,\dots,\log(k/g)\}$ and $j \in \{g,2g,\dots,k\}$ are such that $j + 2^r g = s$. By Claim 8.4,

$$\mu(\mathcal{C}) \le \mu(\mathcal{B}_j) + 2 \le \frac{2j}{g} + 2 \le \frac{2s}{g},$$

where the last inequality holds as $s \ge j + g$. A similar argument, based on Claim 8.5, can be used to analyze MBSs from Item 3 of Definition 8.9. Let $i, j \in \{g,2g,\dots,k\}$ be such that $\mathsf{level}(i) = \mathsf{level}(j)$ and $i + j = s$. Denote $\mathcal{C} = \mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$. By Claim 8.6, $\mu(\mathcal{C}) \le \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j)$ and so, it suffices to verify that $\mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) \le 2(i+j)/g$, which readily holds by the hypothesis.

Denote $\mathcal{C} = \mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(2^{r+1}\cdot 8i)-\mathsf{BS}(2^r \cdot 8i)} \mathcal{B}_j$ where $i, j \in \{g,2g,\dots,k\}$ and $r \in \{0,1,\dots,\log(k/i)\}$ are such that $\mathsf{level}(i) = \mathsf{level}(j)$ and $2^{r+2}i = s$. By Claim 8.6, $\mu(\mathcal{C}) \le \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2$. Hence, it suffices to prove that

$$\mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2 \le \frac{2^{r+3}i}{g}.$$

As $\mathsf{level}(i) = \mathsf{level}(j)$, $j \le 2i$ and so, using the hypothesis, it suffices to show that

$$\frac{6i}{g} + 2 \le \frac{2^{r+3}i}{g},$$

which holds as $r \ge 0$ and $i \ge g$.

Consider the MBS $\mathcal{C} = \mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{US}(8j)} \mathcal{B}_j$ where $i, j \in \{g,2g,\dots,k\}$ are such that $\mathsf{level}(i) > \mathsf{level}(j)$ and $i + j = s$. By Claim 8.7, we have the same bound on $\mu(\mathcal{C})$ as we have for the MBS $\mathcal{A}_i \overrightarrow{\bullet}_{\mathsf{BS}(8i)} \mathcal{B}_j$ which we analyzed above, and so the exact same analysis can be used for it. Now, consider the MBS $\mathcal{C} = \mathcal{A}_i \overset{\rightarrow}{\bullet}_{\mathsf{US}(2^{r+1}\cdot 8j)-\mathsf{US}(2^r\cdot 8j)} \mathcal{B}_j$ where $i, j \in \{g, 2g, \ldots, k\}$ and $r \in \{0, 1, \ldots, \log(k/j)\}$ are such that $\mathsf{level}(i) > \mathsf{level}(j)$ and $i + 2^{r+3}j = s$. By Claim 8.7, $\mu(\mathcal{C}) \le \mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2$. Therefore, it suffices to prove that

$$\mu(\mathcal{A}_i) + \mu(\mathcal{B}_j) + 2 \le \frac{2(i + 2^{r+3}j)}{g}.$$

Using the hypothesis, it suffices to verify that

$$\frac{2j}{g} + 2 \le \frac{2^{r+4}j}{g}$$

which holds as $j \ge g$ and $r \ge 0$. MBSs from Item 6 of Definition 8.9 follow a similar analysis. We omit the details. □

9 The Multiplication Rule for Leveled Matrix Representations

In this section we define a product rule between a pair of LMRs $\mathbf{A}, \mathbf{B}$, which we denote by $\mathbf{A} \cdot \mathbf{B}$, based on the definition of $\mathcal{F}(\mathbf{A}, \mathbf{B})$ and its slices. Following the definition of $\mathbf{A} \cdot \mathbf{B}$, we prove in Claim 9.2 that the product is indeed an LMR and show that it respects certain out-function and magnitude function. In Claim 9.3 we prove that $\langle \mathbf{A} \cdot \mathbf{B} \rangle$ approximates $\langle \mathbf{A} \rangle \langle \mathbf{B} \rangle$. The weight function of $\mathbf{A} \cdot \mathbf{B}$ is analyzed in Claim 9.4. Lastly, we collect all the results in Proposition 9.5.

Definition 9.1. *Let $\mathbf{A} = ((1, \mathcal{A}_0), \ldots, (a_k, \mathcal{A}_k))$, $\mathbf{B} = ((1, \mathcal{B}_0), \ldots, (b_k, \mathcal{B}_k))$ be a pair of (k, g, w)-LMRs. We define the sequence $\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = ((c_0, \mathcal{C}_0), \ldots, (c_k, \mathcal{C}_k))$, where $c_i \in \mathbb{R}$ and $\mathcal{C}_i$ MBSs, as follows. For $s \in \{0, g, 2g, \ldots, k\}$ let*

$$m_s = \max\left(a_{i(\mathcal{D})} b_{j(\mathcal{D})} \mid \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right).$$

Define

$$\mathcal{C}_s = \mathsf{glue}\left(\frac{a_{i(\mathcal{D})} b_{j(\mathcal{D})}}{m_s} \mathcal{D} \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right)$$

and $c_s = |\mathcal{F}_s(\mathbf{A}, \mathbf{B})| \cdot m_s$.

For the glue operation to be defined above, we assume that $\forall \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$, we pad to make all the d_{out}'s and d_{in}'s to be equal to the maximum.

Claim 9.2. *Let $\mathbf{A}, \mathbf{B}$ be a pair of (k, g, w)-LMRs that respect the magnitude-function $\mu(i) = 2i/g$ and the out-function $d_{\mathsf{out}}(i) = 10k \cdot \mathsf{level}(i) + d$ for some integer d. Then, the sequence $\mathbf{C}$ is a (k, g, w)-LMR. Furthermore, $\mathbf{C}$ respects the out-function $d'_{\mathsf{out}}(i) = d_{\mathsf{out}}(i) + 8c_{\mathsf{samp}}g$ and the same magnitude-function $\mu(i) = 2i/g$.*

Proof. We start by proving that $\mathbf{C}$ is a (k, w)-matrix representation. First, by definition, $c_s \geq 0$ for all s. Second, we ought to show that for all $s \geq 1$, $\sigma(\mathcal{C}_s) \geq s - (s-1)\tau$. By Claim 8.11 for every $\mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$, $\sigma(\mathcal{D}) \geq s - (s-1)\tau$. Claim 5.11 and Claim 5.8 then imply that

$$\begin{aligned}
\sigma(\mathcal{C}_s) &= \sigma\left(\mathsf{glue}\left(\frac{a_{i(\mathcal{D})}b_{j(\mathcal{D})}}{m_s}\mathcal{D} \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right)\right) \\
&\geq \min\left(\sigma\left(\frac{a_{i(\mathcal{D})}b_{j(\mathcal{D})}}{m_s}\mathcal{D}\right) \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right) \\
&= \min\left(\sigma(\mathcal{D}) + 2\log\left(\frac{m_s}{a_{i(\mathcal{D})}b_{j(\mathcal{D})}}\right) \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right) \\
&\geq \min\left(\sigma(\mathcal{D}) \mid \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right) \\
&\geq s - (s-1)\tau.
\end{aligned}$$

This proves that $\mathbf{C}$ is a (k, w)-matrix representation.

We turn to show that $\mathbf{C}$ is in fact a (k, g, w)-LMR. To this end, note that by Definition 8.9, $\mathcal{C}_0 = \mathcal{A}_0 \overset{\rightarrow}{\circ}_{\mathsf{BS}(2g)} \mathcal{B}_0$. Hence, by Claim 8.3, $\mathcal{C}_0$ is thin. Now, as $c_0 = a_0 b_0$ and since $\mathbf{A}, \mathbf{B}$ are LMRs, we have that $c_0 = 1$. Moreover, by Claim 8.10, for every s not divisible by g, $c_s = 0$. Next, we ought to show that $\mu(\mathcal{C}_s) \leq s$ for all $s \geq 0$. This clearly holds for $s = 0$ as $\mathcal{C}_0$ is thin. Consider $s \geq g$. By Claim 8.15 and by the hypothesis, for every $\mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$, $\mu(\mathcal{D}) \leq 2s/g \leq s$. Therefore, by Claim 5.11 and Claim 5.8,

$$\begin{aligned}
\mu(\mathcal{C}_s) &= \mu\left(\mathsf{glue}\left(\frac{a_{i(\mathcal{D})}b_{j(\mathcal{D})}}{m_s}\mathcal{D} \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right)\right) \\
&\leq \max\left(\mu\left(\frac{a_{i(\mathcal{D})}b_{j(\mathcal{D})}}{m_s}\mathcal{D}\right) \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right) \\
&\leq \max\left(\mu(\mathcal{D}) - 2\log\left(\frac{m_s}{a_{i(\mathcal{D})}b_{j(\mathcal{D})}}\right) \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right) \\
&\leq \max\left(\mu(\mathcal{D}) \mid \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right) \\
&\leq 2s/g,
\end{aligned}$$

which is bounded by s, as desired. The above equation also proves that $\mathbf{C}$ respects the magnitude-function $\mu(s) = 2s/g$. By Claim 8.14 and by the hypothesis, for every $s \in \{0, g, 2g, \ldots, k\}$ and MBS $\mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$,

$$d_{\mathsf{out}}(\mathcal{D}) \leq 10k \cdot \mathsf{level}(s) + d + 7c_{\mathsf{samp}}g.$$

Claim 5.11 and Claim 8.12, together with the hypothesis $g \geq 10\log k$, then imply that

$$\begin{aligned}
d_{\mathsf{out}}(\mathcal{C}_s) &\leq 10k \cdot \mathsf{level}(s) + d + 7c_{\mathsf{samp}}g + \log|\mathcal{F}_s(\mathbf{A}, \mathbf{B})| \\
&\leq 10k \cdot \mathsf{level}(s) + d + 7c_{\mathsf{samp}}g + 4\log k \\
&\leq 10k \cdot \mathsf{level}(s) + d + 8c_{\mathsf{samp}}g.
\end{aligned}$$

Here, we used that $|\mathcal{F}_s(\mathbf{A},\mathbf{B})| = O((s/g)^2) \le k^2$[11]. By the remark in Section 5.2, we may assume that the above holds with equality, namely,

$$d_{\mathsf{out}}(\mathcal{C}_s) = 10k \cdot \mathsf{level}(s) + d + 8c_{\mathsf{samp}}g. \tag{9.1}$$

Thus, for every $i, j \in \{0, g, 2g, \dots, k\}$, if $\mathsf{level}(i) = \mathsf{level}(j)$ then $d_{\mathsf{out}}(\mathcal{C}_i) = d_{\mathsf{out}}(\mathcal{C}_j)$. Furthermore, if $\mathsf{level}(i) > \mathsf{level}(j)$ then $d_{\mathsf{out}}(\mathcal{C}_i) \ge d_{\mathsf{out}}(\mathcal{C}_j) + 10k$. To complete the proof, note that Equation (9.1) implies that $\mathbf{C}$ respects the out-function $d'_{\mathsf{out}}(i) = d_{\mathsf{out}}(i) + 8c_{\mathsf{samp}}g$. ☐

Claim 9.3. *For every pair* $\mathbf{A}, \mathbf{B}$ *of* (k, g, w)*-LMRs,*

$$\|\langle \mathbf{A} \cdot \mathbf{B} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle\|_{\max} \le (k^3 + 8w)2^{-k/2}\vartheta(\mathbf{A})\vartheta(\mathbf{B}).$$

Proof. Write $\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = ((1, \mathcal{C}_0), (c_g, \mathcal{C}_g), \dots, (c_k, \mathcal{C}_k))$. By Claim 5.11 and Claim 5.8, for every s for which $\mathcal{F}_s(\mathbf{A},\mathbf{B}) \ne \emptyset$,

$$\begin{aligned}
\langle \mathcal{C}_s \rangle &= \left\langle \mathsf{glue}\left(\frac{a_{i(\mathcal{D})} b_{j(\mathcal{D})}}{m_s} \mathcal{D} \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A},\mathbf{B}) \right) \right\rangle \\
&= \frac{1}{|\mathcal{F}_s(\mathbf{A},\mathbf{B})|} \sum_{\mathcal{D} \in \mathcal{F}_s(\mathbf{A},\mathbf{B})} \left\langle \frac{a_{i(\mathcal{D})} b_{j(\mathcal{D})}}{m_s} \mathcal{D} \right\rangle \\
&= \frac{1}{|\mathcal{F}_s(\mathbf{A},\mathbf{B})|} \sum_{\mathcal{D} \in \mathcal{F}_s(\mathbf{A},\mathbf{B})} \frac{a_{i(\mathcal{D})} b_{j(\mathcal{D})}}{m_s} \langle \mathcal{D} \rangle.
\end{aligned}$$

Recall that $c_s = |\mathcal{F}_s(\mathbf{A},\mathbf{B})| \cdot m_s$ and so

$$c_s \langle \mathcal{C}_s \rangle = \sum_{\mathcal{D} \in \mathcal{F}_s(\mathbf{A},\mathbf{B})} a_{i(\mathcal{D})} b_{j(\mathcal{D})} \langle \mathcal{D} \rangle.$$

Thus, if we denote $\mathcal{F}_{\le k}(\mathbf{A},\mathbf{B}) = \cup_{s=0}^{k} \mathcal{F}_s(\mathbf{A},\mathbf{B})$ then

$$\langle \mathbf{C} \rangle = \sum_{s=0}^{k} c_s \langle \mathcal{C}_s \rangle = \sum_{\mathcal{D} \in \mathcal{F}_{\le k}(\mathbf{A},\mathbf{B})} a_{i(\mathcal{D})} b_{j(\mathcal{D})} \langle \mathcal{D} \rangle.$$

Note that, by linearity,

$$\sum_{\mathcal{D} \in \mathcal{F}(\mathbf{A},\mathbf{B})} a_{i(\mathcal{D})} b_{j(\mathcal{D})} \langle \mathcal{D} \rangle = \sum_{i,j} a_i b_j S_{i,j},$$

and so, if we denote $\mathcal{F}_{>k}(\mathbf{A},\mathbf{B}) = \mathcal{F}(\mathbf{A},\mathbf{B}) \setminus \mathcal{F}_{\le k}(\mathbf{A},\mathbf{B})$ then

$$\langle \mathbf{C} \rangle - \sum_{i,j} a_i b_j S_{i,j} = \sum_{\mathcal{D} \in \mathcal{F}_{>k}(\mathbf{A},\mathbf{B})} a_{i(\mathcal{D})} b_{j(\mathcal{D})} \langle \mathcal{D} \rangle.$$

[11] g is large enough.

As $|\mathcal{F}_{>k}(\mathbf{A},\mathbf{B})| \le |\mathcal{F}(\mathbf{A},\mathbf{B})| \le k^3$ (because there are k different values of s and as we saw before $|\mathcal{F}_s(\mathbf{A},\mathbf{B})| \le k^2$) and since, by Claim 5.6, $\|\mathcal{D}\|_{\max} \le \|\mathcal{D}\|_\infty \le 2^{-k/2}$ for every $\mathcal{D}$ with $\sigma(\mathcal{D}) > k$, we have that

$$\begin{aligned}\Big\|\langle \mathbf{C}\rangle - \sum_{i,j} a_i b_j S_{i,j}\Big\|_{\max} &\le \Big\| \sum_{\mathcal{D}\in\mathcal{F}_{>k}(\mathbf{A},\mathbf{B})} a_{i(\mathcal{D})} b_{j(\mathcal{D})} \langle \mathcal{D}\rangle \Big\|_{\max} \\ &\le \sum_{\mathcal{D}\in\mathcal{F}_{>k}(\mathbf{A},\mathbf{B})} a_{i(\mathcal{D})} b_{j(\mathcal{D})} \|\mathcal{D}\|_{\max} \\ &\le k^3 \vartheta(\mathbf{A})\vartheta(\mathbf{B}) 2^{-k/2}.\end{aligned}$$

The proof then follows as by Claim 8.2,

$$\Big\|\langle \mathbf{A}\rangle\langle \mathbf{B}\rangle - \sum_{i,j} a_i b_j S_{i,j}\Big\|_{\max} \le 8w\vartheta(\mathbf{A})\vartheta(\mathbf{B})2^{-k}.$$

□

Claim 9.4. *Let $\mathbf{A} = ((1,\mathcal{A}_0),\ldots,(a_k,\mathcal{A}_k))$, $\mathbf{B} = ((1,\mathcal{B}_0),\ldots,(b_k,\mathcal{B}_k))$ be a pair of (k,g,w)-LMRs that respect the weight-function $\vartheta(s) = (s/g)^{(3s/g)t}$[12] for some $t \ge 0$. Then, $\mathbf{A}\cdot\mathbf{B}$ respects the weight-function $c'(s/g)^{(3s/g)(t+1)})$, where c' is a large enough constant.*

Proof. Write $\mathbf{C} = \mathbf{A}\cdot\mathbf{B} = ((c_0,\mathcal{C}_0),\ldots,(c_k,\mathcal{C}_k))$ $(c_0 = 1)$. Let $s \ge g$. Recall that

$$c_s = |\mathcal{F}_s(\mathbf{A},\mathbf{B})| \cdot \max\left(a_{i(\mathcal{D})} b_{j(\mathcal{D})} \mid \mathcal{D} \in \mathcal{F}_s(\mathbf{A},\mathbf{B})\right).$$

By inspecting the MBSs in Definition 8.9, one can see that $i(\mathcal{D}) + j(\mathcal{D}) \le s$ for every $\mathcal{D} \in \mathcal{F}_s(\mathbf{A},\mathbf{B})$. Moreover, by Claim 8.12, $|\mathcal{F}_s(\mathbf{A},\mathbf{B})| = O((s/g)^2)$. We assume, for simplicity, that the bound is $c'(s/g)^3$ where c' is a large enough constant. Thus,

$$\begin{aligned} c_s &\le c'(s/g)^3 \max\left(\vartheta(i,t)\vartheta(j,t) \mid i+j \le s\right) \\ &\le c'(s/g)^3 \max\left((i/g)^{(3i/g)t}(j/g)^{(3j/g)t} \mid i+j\le s\right) \\ &\le c'(s/g)^3 \max\left((s/g)^{(3i/g)t}(s/g)^{(3j/g)t} \mid i+j\le s\right) \\ &= c'(s/g)^3 (s/g)^{(3s/g)t} \\ &\le c'(s/g)^{3(s/g)(t+1)}, \end{aligned}$$

where for the last inequality we used the fact that $s \ge g$. □

We summarize the results obtained so far in the following proposition.

Proposition 9.5. *Let k, g, w be integers where $k \ge g \ge 10(\omega + \log k)$[13]. Let $\mathbf{A} = ((1,\mathcal{A}_0),\ldots,(a_k,\mathcal{A}_k))$, $\mathbf{B} = ((1,\mathcal{B}_0),\ldots,(b_k,\mathcal{B}_k))$ be a pair of (k,g,w)-LMRs. Assume that*

[12]Implicitly we are assuming that $\vartheta(0) = 1$.

[13]Actually, we assume $k \ge c\cdot g$, where c is a large enough constant.

both $\mathbf{A}, \mathbf{B}$ *respect* $(d_{\mathsf{out}}, d_{\mathsf{in}}, \mu, \vartheta)$*, where*

$$\begin{aligned} d_{\mathsf{out}}(s) &= 10k \cdot \mathsf{level}(s) + d, \\ d_{\mathsf{in}}(s) &= c_{\mathsf{in}} s \log s, \\ \mu(s) &= 2s/g, \\ \vartheta(s) &= c'^t (s/g)^{(3s/g)t}. \end{aligned}$$

for some $d, t \geq 0$ *and the constant* c_{in} *is as defined in Claim 8.13. Then,* $\mathbf{A} \cdot \mathbf{B}$ *is a* (k, g, w)*-LMR that respects* $(d'_{\mathsf{out}}, d_{\mathsf{in}}, \mu, \vartheta')$ *where*

$$\begin{aligned} d'_{\mathsf{out}}(s) &= d_{\mathsf{out}}(s) + 8c_{\mathsf{samp}} g, \\ \vartheta'(s) &= c'^{t+1} (s/g)^{(3s/g)(t+1)}. \end{aligned}$$

Moreover,

$$\|\langle \mathbf{A} \cdot \mathbf{B} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle\|_{\max} \leq (k^3 + w)(k/g)^{(8k/g)(t+1)} 2^{-k/2}.$$

Proof. Write $\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = ((1, \mathcal{C}_0), \ldots, (c_k, \mathcal{C}_k))$. As d_{out}, μ satisfy the hypothesis of Claim 9.2, the fact that $\mathbf{A}, \mathbf{B}$ are LMRs implies that $\mathbf{C}$ is also an LMR, and that $\mathbf{C}$ respects the out-function d'_{out} and the magnitude-function μ. By Claim 8.13, for every $\mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})$ it holds that $d_{\mathsf{in}}(\mathcal{D}) \leq d_{\mathsf{in}}(s)$. Recall that

$$\mathcal{C}_s = \mathsf{glue}\left(\frac{a_{i(\mathcal{D})} b_{j(\mathcal{D})}}{m_s} \mathcal{D} \;\middle|\; \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B}) \right).$$

Therefore, by Claim 5.11, $d_{\mathsf{in}}(\mathcal{C}_s) = \max\left(d_{\mathsf{in}}(\mathcal{D}) \,|\, \mathcal{D} \in \mathcal{F}_s(\mathbf{A}, \mathbf{B})\right)$, which is bounded above by $d_{\mathsf{in}}(s)$, as desired. The assertion that $\mathbf{C}$ respects the weight-function ϑ' readily follows by Claim 9.4. Lastly, by Claim 9.3 (and assuming $k/g \geq c'$),

$$\begin{aligned} \|\langle \mathbf{C} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle\|_{\max} &\leq (k^3 + 8w) \vartheta(\mathbf{A}) \vartheta(\mathbf{B}) 2^{-k/2} \\ &\leq (k^3 + w) k^{(8k/g)(t+1)} 2^{-k/2}. \end{aligned}$$

□

9.1 Multiplying a sequence of LMRs

We start by introducing some notation. Let $\widetilde{A}$ be a $w \times w$ stochastic matrix. Let A be a stochastic matrix approximating $\widetilde{A}$ such that $\|A - \widetilde{A}\|_\infty \leq \frac{\varepsilon}{2n}$ and $A = \mathbf{E}_{j \sim [\mathrm{poly}(wn/\varepsilon)]} A^j$ where each A^j is a 0-1 stochastic matrix[14]. We define a sequence of $\mathrm{poly}(wn/\varepsilon)$ $(0, w)$-matrix bundles $\mathrm{A}^j = ((1, A^j))$; the $(O(\log(wn/\varepsilon), 0, w)$-MBS $\mathcal{A} = (\mathrm{A}^1, \mathrm{A}^2, \ldots)$ and the matrix representation $\mathsf{canon}(A) = ((1, \mathcal{A}))$. Note that $\mathcal{A}$ is thin and $\langle \mathsf{canon}(A) \rangle = A$. Moreover, we may regard $\mathsf{canon}(A)$ as a (k, g, w)-LMR for any $k, g \geq 1$.

[14] Such an approximation can be easily found by truncating each entry of $\widetilde{A}$ to $\mathrm{O}(\log(wn/\varepsilon))$ bits after the decimal point.

Let $h \geq 0$ be an integer and write $n = 2^h$. We want to approximate the product of n stochastic matrices $\widetilde{A}_1, \widetilde{A}_2, ..., \widetilde{A}_n$. Let $A_1, \ldots, A_n$ be the corresponding sequence of $w \times w$ stochastic matrices as defined above. Firstly, $\| \prod_{i=1}^n A_i - \prod_{i=1}^n \widetilde{A}_i \|_{\max} \leq \varepsilon/2$. Next, we approximate $\prod_{i=1}^n A_i$. Let $\mathcal{T}$ be the complete rooted binary tree of depth h. We label every node u of $\mathcal{T}$ by a matrix representation, which we denote by $\mathbf{A}_u$. The i'th leaf of the tree, counting from the left, is labeled by $\mathsf{canon}(A_i)$. Then, inductively over the depth, if u is the parent of the nodes v, w, we define $\mathbf{A}_u = \mathbf{A}_v \cdot \mathbf{A}_w$. For a node u in $\mathcal{T}$, define A_u to be the product of all matrices that correspond to the matrices associated to the leaves in the subtree rooted by u.

Claim 9.6. *For every $\ell \geq 0$ and every node u of height ℓ in $\mathcal{T}$ it holds that*

$$\|\langle \mathbf{A}_u \rangle - A_u\|_{\max} \leq (k^3 + w)(k/g)^{(8k/g)(\ell+1)} 2^{-k/2}.$$

Moreover, $\mathbf{A}_u$ is an LMR that respects $(d_{\mathsf{out}}, d_{\mathsf{in}}, \mu, \vartheta)$ where

$$\begin{aligned} d_{\mathsf{out}}(s) &= 10k \cdot \mathsf{level}(s) + 8c_{\mathsf{samp}} g\ell; \\ d_{\mathsf{in}}(s) &= c_{\mathsf{in}} s \log s; \\ \mu(s) &= 2s/g; \\ \vartheta(s) &= c'^{l} (s/g)^{(3s/g)\ell}. \end{aligned}$$

Proof. The proof is by a straightforward induction. The base case $\ell = 0$ readily holds (as $k \geq c_4 \log(wn/\varepsilon)$ for large enough constant c_4). As for the inductive step, the fact that the respective matrix representation is an LMR that respects $(d_{\mathsf{out}}, d_{\mathsf{in}}, \mu, \vartheta)$ as defined above readily follows by the inductive hypothesis and by Proposition 9.5. For a node u, let $\varepsilon(u) = \|\langle \mathbf{A}_u \rangle - A_u\|_{\max}$. Let u be a node in level $\ell > 0$ and v, w its left and right children, respectively. Then,

$$\begin{aligned} \varepsilon(u) &= \|\langle \mathbf{A}_u \rangle - A_u\|_{\max} \\ &= \|\langle \mathbf{A}_v \cdot \mathbf{A}_w \rangle - A_v A_w\|_{\max} \\ &\leq \|\langle \mathbf{A}_v \cdot \mathbf{A}_w \rangle - \langle \mathbf{A}_v \rangle \langle \mathbf{A}_w \rangle\|_{\max} + \|\langle \mathbf{A}_v \rangle \langle \mathbf{A}_w \rangle - A_v A_w\|_{\max} \\ &\leq (k^3 + w)(k/g)^{(8k/g)(\ell)} 2^{-k/2} + \|\langle \mathbf{A}_v \rangle \langle \mathbf{A}_w \rangle - A_v A_w\|_{\max}, \end{aligned} \tag{9.2}$$

where the last inequality follows by Proposition 9.5 and by the induction hypothesis. As for the second summand in Equation (9.2),

$$\begin{aligned} \|\langle \mathbf{A}_v \rangle \langle \mathbf{A}_w \rangle - A_v A_w\|_{\max} &\leq \|\langle \mathbf{A}_v \rangle \langle \mathbf{A}_w \rangle - A_v A_w\|_{\infty} \\ &\leq \|\langle \mathbf{A}_v \rangle \langle \mathbf{A}_w \rangle - \langle \mathbf{A}_v \rangle A_w\|_{\infty} + \|\langle \mathbf{A}_v \rangle A_w - A_v A_w\|_{\infty} \\ &\leq \|\mathbf{A}_v\|_{\infty} \|\langle \mathbf{A}_w \rangle - A_w\|_{\infty} + \|A_w\|_{\infty} \|\langle \mathbf{A}_v \rangle - A_v\|_{\infty}. \end{aligned} \tag{9.3}$$

Consider the first summand. As A_v is stochastic, we have that

$$\begin{aligned} \|\mathbf{A}_v\|_{\infty} \|\langle \mathbf{A}_w \rangle - A_w\|_{\infty} &= \|\langle \mathbf{A}_v \rangle - A_v + A_v\|_{\infty} \|\langle \mathbf{A}_w \rangle - A_w\|_{\infty} \\ &\leq (\|\langle \mathbf{A}_v \rangle - A_v\|_{\infty} + \|A_v\|_{\infty}) \|\langle \mathbf{A}_w \rangle - A_w\|_{\infty} \\ &= (\|\langle \mathbf{A}_v \rangle - A_v\|_{\infty} + 1) \|\langle \mathbf{A}_w \rangle - A_w\|_{\infty} \\ &= (\varepsilon(v) + 1)\varepsilon(w). \end{aligned}$$

As A_w is stochastic, the second summand on the right hand side of Equation (9.3) is bounded above by $\varepsilon(v)$. Thus,

$$\begin{aligned}\|\langle \mathbf{A}_v\rangle\langle \mathbf{A}_w\rangle - A_v A_w\|_{\max} &\le (\varepsilon(v)+1)\varepsilon(w)+\varepsilon(v)\\ &\le 2(\varepsilon(v)+\varepsilon(w)).\end{aligned}$$

Plugging this to Equation (9.2), and using the induction hypothesis, we get

$$\begin{aligned}\varepsilon(u) &\le 2(\varepsilon(v)+\varepsilon(w)) + (k^3+w)(k/g)^{(8k/g)(\ell)}2^{-k/2}\\ &\le 5(k^3+w)(k/g)^{(8k/g)(\ell)}2^{-k/2}\\ &\le (k^3+w)(k/g)^{(8k/g)(\ell+1)}2^{-k/2},\end{aligned}$$

where the last inequality holds as $k \ge 2g$. □

As a corollary of Claim 9.6 we get that

Corollary 9.7. *There exist universal constants $c_1, c_2 \ge 1$ such that the following holds. Let n, w be integers and $\varepsilon > 0$ such that $\varepsilon < 1/n^2$. Set*

$$\begin{aligned} g &= c_1\left(\log(n)\cdot\log\left(\frac{\log(1/\varepsilon)}{\log n}\right) + \log w + \log\log(1/\varepsilon)\right)\\ k &= c_2\left(g + \log(w/\varepsilon)\right).\end{aligned}$$

Let r be the root of $\mathcal{T}$. Then,

$$\left\|\langle \mathbf{A}_r\rangle - \prod_{i=1}^{n} A_i\right\|_{\max} \le \varepsilon/2.$$

Moreover, write $\mathbf{A}_r = ((1,\mathcal{A}_0),(a_g,\mathcal{A}_g),\ldots,(a_k,\mathcal{A}_k))$. Then, for every $s \in \{0, g, \ldots, k\}$,

$$d_{\mathsf{out}}(\mathcal{A}_s) + d_{\mathsf{in}}(\mathcal{A}_s) = O\left(\log(w/\varepsilon)\log\log(w/\varepsilon) + \log^2(n)\cdot\log\left(\frac{\log(1/\varepsilon)}{\log n}\right) + \log n\cdot\log w\right).$$

Proof. First, we show that Equation (7.1) is satisfied by our choice of k, g. Indeed, by taking any $c_2 \ge 1$, we get $k \ge g$. Furthermore, by taking $c_1 \ge 40$, we get that $g \ge 20\omega$. Therefore, it suffices to verify that $g \ge 20\log k$ which is guaranteed to holds assuming $c_1 \ge 40$.

By Claim 9.6 applied to the root r of $\mathcal{T}$,

$$\left\|\langle \mathbf{A}_r\rangle - \prod_{i=1}^{n} A_i\right\|_{\max} \le (k^3+w)(k/g)^{(8k/g)(\log n+1)}2^{-k/2}.$$

First, we show that

$$(k/g)^{(8k/g)(\log n+1)} \le 2^{k/4}. \tag{9.4}$$

By rearranging, it suffices to show that

$$g \geq 32(\log(n)+1)\log(k/g). \tag{9.5}$$

Now,

$$\begin{aligned}\frac{k}{g} &= \frac{c_2(g+\log(w/\varepsilon))}{g} \\ &= c_2\left(1+\frac{\log(w/\varepsilon)}{g}\right).\end{aligned}$$

As $\varepsilon < 1/n^2$, $g \geq c_1(\log w + \log n)$, and so

$$\begin{aligned}\frac{k}{g} &\leq c_2\left(1+\frac{\log(w/\varepsilon)}{c_1(\log w+\log n)}\right) \\ &\leq c_2\left(1+\frac{\log(1/\varepsilon)}{\log n}\right) \\ &\leq \frac{2c_2\log(1/\varepsilon)}{\log n}.\end{aligned} \tag{9.6}$$

Hence, to prove Equation (9.5), it suffices to show that

$$g \geq 32(\log(n)+1)\log\left(\frac{2c_2\log(1/\varepsilon)}{\log n}\right).$$

The above equation holds assuming that

$$\frac{c_1}{32}\cdot\log\left(\frac{\log(1/\varepsilon)}{\log n}\right) \geq \log\left(\frac{2c_2\log(1/\varepsilon)}{\log n}\right),$$

which holds by choosing the constants c_1, c_2 such that $c_1 \geq 64 + 32\log c_2$, which is consistent with the restrictions imposed so far.

Now that we proved Equation (9.4), we have that

$$\left\|\langle \mathbf{A}_r\rangle - \prod_{i=1}^{n} A_i\right\|_{\max} \leq (k^3+w)2^{-k/4}.$$

For large enough k, the RHS is bounded by $w2^{-k/5}$. As $k \geq c_2\log(w/\varepsilon)$, by taking $c_2 \geq 10$ we get $w2^{-k/5} \leq \varepsilon/2$, as desired.

Moving to the moreover part, by Claim 9.6, for every $s \in \{0, g, 2g, \ldots, k\}$,

$$\begin{aligned}d_{\mathsf{out}}(\mathcal{A}_s) &= 10k\cdot\mathsf{level}(s) + 8c_{\mathsf{samp}}g\log n \\ &= O(k\log k + g\log n).\end{aligned}$$

Note that

$$\log(n)\cdot\log\left(\frac{\log(1/\varepsilon)}{\log n}\right) = O(\log(1/\varepsilon)),$$

and so $k = O(\log(w/\varepsilon))$, which yields

$$\begin{aligned} d_{\mathsf{out}}(\mathcal{A}_s) &= O(\log(w/\varepsilon)\log\log(w/\varepsilon) + g\log n) \\ &= O\left(\log(w/\varepsilon)\log\log(w/\varepsilon) + \log^2(n)\cdot\log\left(\frac{\log(1/\varepsilon)}{\log n}\right) + \log n\cdot\log w\right). \end{aligned}$$

Note that d_{in} is dominated by d_{out} as $d_{\mathsf{in}}(\mathcal{A}_s) = O(s\log s) = O(d_{\mathsf{out}}(\mathcal{A}_s))$. □

9.2 Proof of Theorem 4.3

In this section we deduce Theorem 4.3.

Proof of Theorem 4.3. The pseudo-distribution $\widetilde{\mathcal{D}}$ is induced in a natural way from the multiplication rule between LMRs. As in Section 2.1, for any width-w, length-n branching program P, we can represent the transition between a pair of consecutive layers P_t, P_{t+1} of the program by a $w \times w$ stochastic matrix M_t, which is an average of two 0-1 stochastic matrices M_t^0 and M_t^1 representing the transitions when the t^{th} bit is 0 and 1 respectively. And as sparsification of matrix product gave us a PRG for ROBPs, the above mentioned process of multiplying a sequence of LMRs gives us a PRPG for ROBPs.

To be more precise, let $\mathbf{A} = ((1, \mathcal{A}_0), \dots, (a_k, \mathcal{A}_k))$ be the final LMR at the root of the tree described in Section 9.1 while multiplying the matrices $M_1, ..., M_n$. $\forall\ i \in \{0, 1, ..., k\}$, let $\mathcal{A}_i = (\mathrm{A}_{i,1}, ..., \mathrm{A}_{i,2^{d_{\mathsf{out}}(\mathcal{A}_i)}})$ and $\forall\ j \in [2^{d_{\mathsf{out}}\mathcal{A}_i}]$, $\mathrm{A}_{i,j} = ((\alpha_{i,j,1}, A_{i,j,1}), ..., (\alpha_{i,j,2^{d_{\mathsf{in}}(\mathcal{A}_i)}}, A_{i,j,2^{d_{\mathsf{in}}(\mathcal{A}_i)}}))$. It's easy to see that because we started with 0-1 stochastic matrices M_t^0, M_t^1 in the matrix bundles at the leaves, $\forall\ i, j, m,\ A_{i,j,m}$ is a 0-1 stochastic matrix and corresponds to a single n-length path in the branching program, say, $p_{i,j,m}$. This can be seen by induction on the levels of the tree; the innermost matrices at level l corresponds to paths of length 2^l.

Thus, the PRPG is the sequence $((\rho_{i,j,m}, p_{i,j,m}))_{i\in\{0,1,...,k\}, j\in[2^{d_{\mathsf{out}}(\mathcal{A}_i)}], m\in[2^{d_{\mathsf{in}}(\mathcal{A}_i)}]}$ where $\rho_{i,j,m} = a_i \cdot \frac{1}{2^{d_{\mathsf{out}}(\mathcal{A}_i)}} \cdot \alpha_{i,j,m}$. By following the matrix products, it's easy to see that the weights and coefficients depend only on the types of products used (starting with all coefficients being 1 at the leaves) and not on the entries of M_t^0 and M_t^1 and hence, PRPG is input-oblivious and does not depend on the ROBP. Next, each bit of the string, corresponding to the path $p_{i,j,m}$, can be computed knowing the definitions of the matrix products and corresponding samplers and inductively going down the tree (the information can be calculated from the indices i, j, m).

As the samplers that we use for the product between LMRs are log-space computable (log in the size of the bipartite graph), one can see that the $\widetilde{\mathcal{D}}$ is $\widetilde{O}\left(\log(w/\varepsilon)\log\log(w/\varepsilon) + \log^2(n)\cdot\log\left(\frac{\log(1/\varepsilon)}{\log n}\right) + \log n\cdot\log w\right)$-space computable. The seed length, which is given by,

$$\log\left(\sum_{i=0}^{k} 2^{d_{\mathsf{in}}(\mathcal{A}_i)+d_{\mathsf{out}}(\mathcal{A}_i)}\right) \le d_{\mathsf{in}}(\mathcal{A}_k) + d_{\mathsf{out}}(\mathcal{A}_k) + \log k$$

readily follows by Corollary 9.7.

As for the bound on the weights of $\widetilde{\mathcal{D}}$, note that the ρ_i's in $\widetilde{\mathcal{D}}$ are obtained by multiplying the weights of **A** with the coefficients of the MBSs composing **A**. It is easy to verify that the coefficients are all bounded above by 1 in absolute value. Therefore, it suffices to bound the weights of **A**. By Claim 9.6, $\vartheta(k) \le c'^{\log(n)}(k/g)^{(3k/g)\log n}$, and so

$$\begin{aligned}\log \vartheta(k) &\le \frac{3k \log n}{g} \log(k/g) + c' \log(n) \\ &\le \left(\log n + \frac{\log n \cdot \log(w/\varepsilon)}{g}\right) 3c_2 \log(k/g) + c' \log(n).\end{aligned}$$

By Equation (9.6), $k/g \le 2c_2 \log(1/\varepsilon)/\log n$. Let $t = \log\left(\frac{2c_2 \log(1/\varepsilon)}{\log n}\right)$. Then,

$$\begin{aligned}\log \vartheta(k) &\le 3c_2 \left(\log n + \frac{\log n \cdot \log(w/\varepsilon)}{g}\right) t + c' \log(n) \\ &= O\left(\log(1/\varepsilon) + \frac{\log n \cdot \log(w/\varepsilon) \cdot t}{g}\right)\end{aligned}$$

As $g = \Omega(t \log n + \log w)$, we have that

$$\log \vartheta(k) = O(\log(1/\varepsilon) + \log(w/\varepsilon)) = O(\log(w/\varepsilon)),$$

which completes the proof. □

References

[AB09] S. Arora and B. Barak. *Computational Complexity - A Modern Approach.* Cambridge University Press, 2009.

[Arm98] Roy Armoni. On the derandomization of space-bounded computations. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 47–59. Springer, 1998.

[ATSWZ00] Roy Armoni, Amnon Ta-Shma, Avi Wigderson, and Shiyu Zhou. An $O(\log(n)^{4/3})$ space algorithm for (s, t) connectivity in undirected graphs. *Journal of the ACM*, 47(2):294–311, 2000.

[BCP83] Allan Borodin, Stephen Cook, and Nicholas Pippenger. Parallel computation for well-endowed rings and space-bounded probabilistic machines. *Information and Control*, 58(1-3):113–136, 1983.

[BPW11] A. Bogdanov, P. Papakonstaninou, and A. Wan. Pseudorandomness for read-once formulas. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 240–246. IEEE, 2011.

[BPW12] A. Bogdanov, P. Papakonstantinou, and A. Wan. Pseudorandomness for linear length branching programs and stack machines. In *APPROX-RANDOM*, pages 447–458. Springer, 2012.

[BR94] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994*, pages 276–287. IEEE, 1994.

[BRRY14] M. Braverman, A. Rao, R. Raz, and A. Yehudayoff. Pseudorandom generators for regular branching programs. *SIAM Journal on Computing*, 43(3):973–986, 2014.

[BV10] Joshua Brody and Elad Verbin. The coin problem and pseudorandomness for branching programs. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 30–39. IEEE, 2010.

[De11] A. De. Pseudorandomness for permutation and regular branching programs. In *2011 IEEE 26th Annual Conference on Computational Complexity (CCC)*, pages 221–231. IEEE, 2011.

[DSTS17] Dean Doron, Amir Sarid, and Amnon Ta-Shma. On approximating the eigenvalues of stochastic matrices in probabilistic logspace. *Computational Complexity*, 26(2):393–420, 2017.

[GMR+12] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 120–129. IEEE, 2012.

[GMRZ13] P. Gopalan, R. Meka, O. Reingold, and D. Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM Journal on Computing*, 42(3):1051–1076, 2013.

[Gol11] O. Goldreich. A sample of samplers: A computational perspective on sampling. In *Studies in Complexity and Cryptography*, volume 6650 of *Lecture Notes in Computer Science*. Springer, 2011.

[GV17] R. Gurjar and B. Volk. Pseudorandom bits for oblivious branching programs. *arXiv preprint arXiv:1708.02054*, 2017.

[GW97] O. Goldreich and A. Wigderson. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Struct. Algorithms*, 11(4):315–343, 1997.

[IKW02] R. Impagliazzo, V. Kabanets, and A. Wigderson. In search of an easy witness: Exponential time vs. probabilistic polynomial time. *Journal of Computer and System Sciences*, 65(4):672–694, 2002.

[IMZ12] R. Impagliazzo, R. Meka, and D. Zuckerman. Pseudorandomness from shrinkage. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 111–119. IEEE, 2012.

[INW94] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing, STOC 1994*, pages 356–364. ACM, 1994.

[KI04] V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

[KNP11] M. Kouckỳ, P. Nimbhorkar, and P. Pudlák. Pseudorandom generators for group products. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 263–272. ACM, 2011.

[MRSV17] Jack Murtagh, Omer Reingold, Aaron Sidford, and Salil Vadhan. Derandomization beyond connectivity: Undirected laplacian systems in nearly logarithmic space. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 801–812. IEEE, 2017.

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

[Nis94] N. Nisan. $\mathbf{RL} \subseteq \mathbf{SC}$. *Computational Complexity*, 4(1):1–11, 1994.

[NSW92] N. Nisan, E. Szemeredi, and A. Wigderson. Undirected connectivity in $O(\log(n)^{1.5})$ space. In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science, 1992*, pages 24–29. IEEE, 1992.

[NW94] N. Nisan and A. Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[NZ96] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[Rei08] O. Reingold. Undirected connectivity in log-space. *Journal of the ACM (JACM)*, 55(4):17, 2008.

[RR99] R. Raz and O. Reingold. On recycling the randomness of states in space bounded computation. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 159–168. ACM, 1999.

[RSV13] Omer Reingold, Thomas Steinke, and Salil Vadhan. Pseudorandomness for regular branching programs via fourier analysis. In *APPROX-RANDOM*, pages 655–670. Springer, 2013.

[RTV06] Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom walks on regular digraphs and the RL vs. L problem. In *STOC*, volume 6, pages 457–466, 2006.

[RV05] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *APPROX-RANDOM*, pages 436–447. Springer, 2005.

[RVW01] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 18, 2001. https://eccc.weizmann.ac.il/report/2001/018/.

[Sav70] W. J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of computer and system sciences*, 4(2):177–192, 1970.

[Ste12] T. Steinke. Pseudorandomness for permutation branching programs without the group theory. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 19, page 6, 2012.

[SVW14] Thomas Steinke, Salil Vadhan, and Andrew Wan. Pseudorandomness and fourier growth bounds for width-3 branching programs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

[SZ99] M. Saks and S. Zhou. $\mathbf{BP}_H\mathbf{SPACE}(s) \subseteq \mathbf{DSPACE}(s^{3/2})$. *Journal of computer and system sciences*, 58(2):376–403, 1999.

[ŠŽ11] Jiří Šíma and Stanislav Žák. Almost k-wise independent sets establish hitting sets for width-3 1-branching programs. In *International Computer Science Symposium in Russia*, pages 120–133. Springer, 2011.

[Tri08] Vladimir Trifonov. An $O(\log(n) \log\log(n))$ space algorithm for undirected st-connectivity. *SIAM Journal on Computing*, 38(2):449–483, 2008.

[Vad11] S. P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 2011.

[Zuc97] D. Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11(4):345–367, 1997.

[Zuc07] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3:103–128, 2007.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il