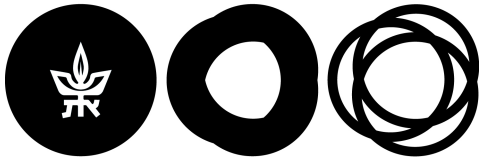

The Raymond and Beverly Sackler Faculty of Exact Sciences
The Blavatnik school of computer science

Simplifying the BCG pseudo-random pseudo-distribution for space bounded computation

Thesis submitted for the degree of
Master of Sciences
by
Daniel Kozlov

This work was carried out under the supervision of
Professor Amnon Ta Shma
February 2020

Abstract

The [BCG17] paper introduced a new primitive which can be used to derandomize two-sided error algorithms, *Pseudo-random pseudo-distributions*(PRPDs). This is a generalization of pseudo-random generators such that one may assign negative or unbounded weights and not necessarily work with a probability distribution. In this thesis we aim to simplify and provide new analysis for the [BCG17] paper which shows a construction of a PRPD against (n, w)-ROBP with seed length $\tilde{O}(\log^2(n) + \log(1/\epsilon) + \log(w) \cdot \log(n))$. The best known PRG in the general unrestricted case has seed-length $O(\log(n) \cdot \log(\frac{n \cdot w}{\varepsilon}))$ ([INW94] [Nis92]), hence when $\log(1/\epsilon) > \log(n)$ the [BCG17] construction is strictly better.

Contents

1 Introduction

Randomized algorithms are algorithms that employ a degree of randomness as part of their logic, these algorithms are usually modelled as probabilistic turing machines and understanding how randomness affects computation is a major question in computational complexity theory. The study of derandomization asks under what circumstances and how could you 'derandomize' such randomized algorithms, that is make equivalent algorithms that use little or no randomness, and how this derandomization affects the algorithm's running time or the amount of space it uses.
It is widely believed that randomness does not add significant computational power to space-bounded or time-bounded machines. In this thesis we focus on the area of space-bounded derandomization, meaning that we aim to transform randomized algorithms which have some bound on the amount of space they use to deterministic (or at least less random) space bounded algorithms. A major open question in this field is whether $BPL = L$, where L is the set of all languages decidable by logarithmic space turing machines, and BPL is the set of all languages decidable by a probabilistic logarithmic space turing machine, note that $L \subseteq BPL$ is obvious, in the other direction [Nis92] proved that $BPL \subseteq DSpace[\log^2(n)]$ and more recently Saks and Zhou in [SZ99] showed: $BPL \subseteq DSpace[\log^{1.5}(n)]$ which is the strongest inclusion currently known.

1.1 Pseudo-random distributions

The study of derandomizing space bounded algorithms is usually done by considering the non-uniform model of read once branching programs (ROBP).

Definition 1.1. *(ROBP) an (n, w)-ROBP is a directed graph with $n + 1$ layers, each such layer consists of w vertices where one of the vertices of the last layer is the "accept" vertex, and one of the vertices of the first layer is the "start" vertex. Each vertex in the graph, except those in the last layer, has two outgoing edges to the next layer labeled by 0 and 1.*
Given input $x \in \{0,1\}^n$ an (n, w)-ROBP computes by starting from the "start" vertex and then proceeding by following the edges of the ROBP according to the bits of x. For a (n, w)-ROBP B and $x \in \{0,1\}^n$ we define $B(x)$ as the vertex the ROBP ends it's computation on, if $B(x) = accept$ we say the ROBP accepts x, otherwise we say it rejects x.

The connection to space bounded random turing machines stems from the fact that a space s random machine can be simulated by an (n, w)-ROBP with $n, w = 2^{O(s)}$, Thus derandomizing ROBPs yields a derandomization for space bounded turing machines.
An approach to derandomizing ROBPs is to find some distribution of small support that "looks random" to any such ROBP.

Definition 1.2. *(Pseudo-random distribution) an (n, w, ε)-pseudo-random distribution is a distribution D over $\{0,1\}^n$ such that for every length n, width w ROBP B, it holds that:*

$$|Pr_{x\sim U_n}[B(x) = accept] - Pr_{x\sim D}[B(x) = accept]| < \varepsilon$$

Notice that derandomizing space bounded machines using a pseudo-random distribution with small support is fairly easy, we can just go through any string in the support and simulate the machine, getting an ε approximation of the acceptance probability of the machine given a truly random string, hence if the machine is a BPL machine (for instance $BPL[\frac{1}{3}, \frac{2}{3}]$) and assuming $\varepsilon < \frac{1}{10}$ then this approximation is greater than $\frac{1}{2}$ if and only if the input is in the language.

1.2 INW generator

Before showing the INW generrator we consider a very naive construction of a pseudo-distribution to illustrate the intuition behind the INW construction. First we need to define the tensor product of two distributions:

Definition 1.3. *(Tensor product) For two distributions A, B over $\{0,1\}^n$ we define the distribution $A \otimes B$. This is a distribution over $\{0,1\}^{2n}$ and has support size $|support(B)| \cdot |support(A)|$, where for every $a_i \in support(A), b_j \in support(B)$ the distribution gives $a_i \circ b_j$ (here $\circ$ stands for concatenation) with probability $\alpha_i \cdot \beta_j$ where α_i, β_j are the probabilities of a_i and b_j accordingly.*

The construction is then given by using this product inductively, we start from $P^{(0)}$ - the uniform distribution over $\{0,1\}$, and define: $P^{(\ell+1)} = P^{(\ell)} \otimes P^{(\ell)}$. Note that $P^{(\ell)}$ is actually the uniform distribution over $\{0,1\}^{2^\ell}$ and in particular $P^{(\log(n))}$ is an (n, w, ε)-pseudo distribution with $\varepsilon = 0$. The idea behind the INW construction is to derandomize this tensor product using samplers, and then using this derandomized tensoring recursively just as here.
We now show an equivalent variation of the INW generator constructed with samplers, for the original construction see [INW94]. Throughout this subsection we will assume the reader is familiar with samplers, for the formal definition see section 3 in this thesis, more on samplers can be found in [Gol11].

Definition 1.4. *(Derandomized tensoring) For two distributions A, B over $\{0,1\}^n$ and a left regular bipartite graph $G = ([|Support(A)|], [|Support(B)|], E)$ with degree D, we define the distribution $A \circ_G B$. This is a distribution over $\{0,1\}^{2n}$ and has support size $|E| = D \cdot |support(A)|$, where for every edge $e = (i, j)$ the distribution gives $a_i \circ b_j$ (here $\circ$ stands for concatenation) with probability $\frac{\alpha_i}{D}$ where α_i is the probability of getting a_i from the distribution A. Note that when G is the complete bipartite graph and B is uniform this is exactly the same as tensor product.*

Now we'll construct the pseudo-random distribution inductively. As earlier define $P^{(0)}$ as the uniform distribution over $\{0,1\}$, and $P^{(\ell+1)} = P^{(\ell)} \circ_G P^{(\ell)}$ where G is some (ε, δ)-sampler, note that $P^{(\ell)}$ is a distribution over $\{0,1\}^{2^\ell}$. The intuition behind the product $P^{(\ell)} \circ_G P^{(\ell)}$ is that due to the properties of G this distribution will be very close to $P^{(\ell)} \otimes P^{(\ell)}$, while having support $D \cdot |Support(P^{(\ell)})|$ instead of $|Support(P^{(\ell)})|^2$. This allows the resulting distribution have much smaller support compared to the uniform distribution from earlier, yet still be a pseudo-random distribution (instead of a completely uniform one).
By theorem 3.3 we know there exists an (ε, δ)-sampler with degree $d = O(\delta^{-1} \cdot \varepsilon^{-2})$, we will choose $\delta = \varepsilon$ throughout this construction. Straightforward error analysis shows that the error of this construction at the ℓ'th level is $\varepsilon(\ell) = O(2^\ell \cdot w \cdot \delta)$, hence choosing $\delta = O(\frac{\varepsilon'}{nw})$ implies that $P^{\log(n)}$ is an (n, w, ε')-pseudo-random distribution. For the size analysis, denote $s(\ell) = |Support(P^\ell)|$, we get $s(\ell) = s(\ell-1) \cdot O(\delta^{-3})$, hence $s(\ell) = O(\delta^{-3\ell})$, so finally we get: $s(\log(n)) = O((\frac{nw}{\varepsilon'})^{\log(n)})$ and the seed length is: $O(\log(n) \cdot \log(\frac{nw}{\varepsilon'}))$.
In fact this is the best known pseudo-random distribution (in terms of seed-length) in the general case, setting $\varepsilon < \frac{1}{10}$ in this pseudo-random distribution implies $BPL \subseteq Dspace[\log^2(n)]$ via the naive derandomization method presented earlier.

1.3 Pseudo-random pseudo distributions

One key aspect of the [BCG17] construction is that instead of constructing a pseudo-random distribution, they relaxed the requirement of being a distribution and introduced a new object called pseudo-random pseudo-distribution. this construction has some similarities with the INW construction we mentioned but it is much more complex. One of the main ideas in the [BCG17] construction was that instead of recursively using the same "expensive" sampler with low error (like in INW), they broke it down and used a series of samplers with growing precision $G_1, G_2, \ldots, G_k$, where G_k is the most expansive one. and instead of simply computing $P^{(\ell)} \circ_{G_k} P^{(\ell)}$ they broke it down to a telescopic sum:

$$P^{(\ell)} \circ_{G_k} P^{(\ell)} = P^{(\ell)} \circ_{G_1} P^{(\ell)} + P^{(\ell)} \circ_{G_2} P^{(\ell)} - P^{(\ell)} \circ_{G_1} P^{(\ell)} + \ldots P^{(\ell)} \circ_{G_k} P^{(\ell)} - P^{(\ell)} \circ_{G_{k-1}} P^{(\ell)}$$

Here you can think of $P^{(\ell)} \circ_{G_1} P^{(\ell)}$ as the first fragment of the distribution serving as a baseline approximation of $P^{(\ell)} \circ_{G_k} P^{(\ell)}$. And for $r > 1$, the r'th fragment is $P_i \circ_{G_r} P_j - P_i \circ_{G_{r-1}} P_j$ serving as a correcting term bringing the approximation closer to $P^{(\ell)} \circ_{G_k} P^{(\ell)}$, with each correcting term getting less and less significant. This allows to store the distribution in fragments such that their combination gives the distribution but you maintain them ordered by their significance. Note that in this telescopic representation some distributions come with a minus sign, and this forces us to deal with distributions that have "negative probabilities" in some sense, That's where pseudo-distribution come in and allow to take advantage of this telescopic presentation.

Definition 1.5. *((n, w, ε)-pseudo-random pseudo distribution) A* pseudo-distribution *over Λ with support C is $A = \{(\alpha_c, x_c)\}_{c \in C}$ where $\alpha_c \in \mathbb{R}$ and $x_c \in \Lambda$. We say a pseudo-distribution is (n, w, ε)-pseudo-random if for every (n, w)-ROBP B, it holds:*

$$|\sum_{c \in C} \alpha_c \cdot 1_{B(x_c)} - Pr_{x \sim U_n}[B(x) = accept| < \varepsilon$$

Where $1_{B(x_c)}$ is the indicator function to whether $B(x_c) = accept$.

Note that pseudo-random pseudo distributions are as good as classical distributions for the naive derandomization method. As we may simply sum the weights of the accepted words and get an ε approximation of the probability of the machine accepting given a truly random string. The main result The [BCG17] paper achieves is:

Theorem 1.6. *(Main Result) For every integers $n, w \geq 1$ and $0 < \varepsilon < 1/n$, there exists a (n, w, ε)-Pseudo-random pseudo distribution with seed-length:*

$$\tilde{O}(\log^2(n) + \log(1/\epsilon) + \log(w) \cdot \log(n))$$

Note that this is strictly better than the [INW] result when $\log(\frac{1}{\epsilon}) = \omega(\log(n))$. The goal in this thesis is to show the [BCG17] construction in a relatively clean and simple way.

1.4 The BCG construction

In this subsection we will briefly discuss the [BCG17] construction and explain what was done in this thesis in comparison. this construction is incredibly complicated and we will omit and over-simplify a lot of details in this subsection. The main object we focus on will be sequences

of $A+1$-pseudo distributions indexed by $\{0,1,\ldots,A\}$, the idea is that these pseudo-distributions will be ordered by decreasing significance to the resulting pseudo-distribution. This construction will be similar in spirit to INW as we will define some product rule for those sequences of pseudo-distributions and use it inductively to build $P^{(\ell)} = \{P_0^{(\ell)}, P_1^{(\ell)}, \ldots, P_A^{(\ell)}\}$. The heart of the construction is how this product was defined, namely how to define $P^{(\ell+1)}$ out of $P^{(\ell)}$.

- Assume that for every $i,j \in \{0,1,\ldots,A\}$ we have a sequence of samplers $G_{i,j}^1, G_{i,j}^2 \ldots, G_{i,j}^k$ (note that k might be dependant on i and j) with increasing precision.

- For every $i,j \in \{0,1,\ldots,A\}$ look at the telescopic representation of $P_i^{(\ell)} \circ_{G_k} P_j^{(\ell)}$ as mentioned earlier. This gives us k pseudo distribution whose union is exactly $P_i^{(\ell)} \circ_{G_k} P_j^{(\ell)}$, the first of those is $P_i^{(\ell)} \circ_{G_1} P_j^{(\ell)}$ and for $r>1$ the r'th one is:

  $$P_i^{(\ell)} \circ_{G_r} P_j^{(\ell)} - P_i^{(\ell)} \circ_{G_{r-1}} P_j^{(\ell)}.$$

  Repeating this process for every $i,j \in [A]$ we get a large pool of pseudo-distributions, let us denote them by: $P_{i,j,r}$, we will focus on this set of pseudo-distributions.

- A naive way to define the product is to simply ignore that $P^{(\ell)}$ is ordered by significance and group all these pseudo-distributions together. Note that since for every i,j the combination of all the pseudo-distributions created from $P_i^{(\ell)}$ and $P_j^{(\ell)}$ will be exactly $P_i^{(\ell)} \circ_{G_k} P_j^{(\ell)}$, This will be equivalent to the INW generator.

- Instead we will use the fact that $P^{(\ell)}$ is ordered by significance. Moreover for every $i \leq A$ we will assume some bound on the significance of $P_i^{(\ell)}$ which depends on i. This means we would have to order the pseudo distributions $P_{i,j,r}$ by their significance in the resulting distribution. As a result we have to fix a significance function $s(i,j,r) \geqslant 0$ which tells us how significant $P_{i,j,r}$ is.

- This significance function dictates the construction, for every $a \leq A$ we form $P_a^{(\ell+1)}$ by grouping together all the $P_{i,j,r}$ such that $s(i,j,r) = a$. Note that pseudo-distributions $P_{i,j,r}$ with $s(i,j,r) > A$ are thrown away at this stage as they are too insignificant.

- Note that the significance function $s(i,j,r)$ has to uphold that the significance of $P_{i,j,r}$ is indeed bounded as $s(i,j,r)$ implies it is. Of course this significance heavily depends on the parameters of the relevant samplers. To prove that this is indeed the case is where we need to use that the significance of $P_i^{(\ell)}$ and $P_j^{(\ell)}$ is bounded.

Actually sequences of pseudo-distributions aren't subtle enough and the actual objects used in the construction are sequences of sequences of pseudo-distributions (we will view them as sequences of distributions over pseudo-distributions - or d.p.ds for short). And in fact we measure the significance of d.p.d's and not simply pseudo-distributions. Note that there are many "moving parts" in this construction, namely the function $s(i,j,r)$ should be chosen as well as the parameters of the samplers used throughout and the parameter A. These parameters interact and affect each other in complicated manner making the analysis of the construction fairly lengthy and technical.

1.5 What was done in this thesis

The analysis in [BCG17] goes roughly as follows. First they fix some family of samplers to use throughout the construction and then determine what should $s(i,j,r)$ be to maintain some bound on the significance. This leads to a very technical analysis and results in the rather complicated expression:

$$s(i,j,r) = \begin{cases} i+j & \text{If } r=0 \\ \max(i,j)+2^r & \text{If } r>0 \text{ and } \min\{i,j\}=0 \\ 2^{r+2}\cdot\max(i,j) & \text{If } r>0 \text{ and } \lfloor\log i\rfloor = \lfloor\log j\rfloor \\ \max(i,j)+(2^{r+3})\cdot\min(i,j) & \text{If } r>0 \text{ and } \lfloor\log i\rfloor \neq \lfloor\log j\rfloor. \end{cases}$$

This complicates the [BCG17] paper making it more difficult to follow. Moreover, this hardly tells anything about how the sampler quality (as measured by ε and δ) affects the construction.

In this thesis we aim to analyse the construction as cleanly and simply as possible, trying to illuminate how the different parameters affect each other. Before we begin we:

- We define the significance measure. Conceptually, this is done in an almost identical way to [BCG17]. For a d.p.d $\mathbf{A}$ - a distribution over 2^r pseudo-distributions each of them we identify with their corresponding transition matrix. We define the significance of $\mathbf{A}$ as: $\mathbb{E}_{i\in[2^r]} \| A_i \|_\infty$. For comparison in [BCG17] the significance measure used is: $-\log(\mathbb{E}_{i\in[2^r]}[\| A_i \|_\infty^2])$
- We choose the function $s(i,j,r)$. We choose what we believe is a natural function:

  $$s(i,j,r) = i+j+r.$$

  As mentioned, the analysis in [BCG17] leads to a function $s(i,j,r)$ that grows exponentially with r which might seem quite arbitrary, this is essentially a byproduct of the (ε,δ) parmeters they fixed for the samplers. Note that in both our function and in the [BCG17] paper it holds that $s(i,j,0) = i+j$.

As $s(i,j,r)$ is a measure of significance, the above two choices imply constraints on the required quality (ε,δ) of the different samplers, so that we indeed have: $significance(P_a^{(\ell)}) \leq \Delta^a$ for every ℓ, where $\Delta < \frac{1}{2}$ is some parameter we discuss soon. This is where most of the technical work is done. The constraints we get are roughly:

$$\varepsilon(i,j,r) \leq \frac{\Delta^{s(i,j,r)-\max(i,j)}}{4w^2\cdot 2^j\cdot 20A}.$$

and:

$$\delta(i,j,r) \leq \frac{\Delta^{s(i,j,r)-\max(i,j)}}{4w^2\cdot 2^i\cdot 2^j\cdot 20A}.$$

We show (in Section 8) that using samplers with parameters within those constraints leads to getting that $P^{(\log(n))}$ is a (n,w)-pseudo-random pseudo-distribution with error $w^3\cdot(\mathrm{A}+1)^{7\mathrm{A}\log(n)}\cdot\Delta^{\mathrm{A}/4}$.

Now, we can determine the value of the mystery parameter Δ. To get error ϵ we:

- Since $A \geq 1$ we get that the error is at least $w^3\cdot 2^{7A\log(n)-\log(1/\Delta)\cdot A/4}$ This implies that $\log(1/\Delta) = \Omega(\log(n))$. More precisely we choose $\log(1/\Delta)$ to be $O(\log(n)\cdot\log(\log(w/\varepsilon))+\log(w))$, and,

- to get error ϵ we choose $A = O(\frac{\log(w/\varepsilon)}{\log(1/\Delta)})$.

The size analysis now gives that the seed-length of $P^{(\log(n))}$ is:

$$\tilde{O}(A \cdot \log(1/\Delta) + \log(n) \cdot \log(1/\Delta)) = \tilde{O}(\log^2(n) + \log(1/\varepsilon) + \log(w) \cdot \log(n)).$$

the same as in the [BCG17] paper.

We now explain the relationship between the parameter Δ we have, and a somewhat mysterious parameter g in the [BCG17] construction. The main object in our construction is what we call a *cube*. A cube is a sequence of $A + 1$ "d.p.d"s (distributions over pseudo-distributions) ordered by decreasing significance with the i'th d.p.d. having at most approximately Δ^i significance. The main object of the [BCG17] construction is a (k, g, w)-leveled matrix representation (LMR). These (k, g, w)-LMRs are sequences of $k + 1$ d.p.ds such that the i'th d.p.d. is zero if $i \nmid g$, and otherwise has insignificance at least approximately i.

As explained before, significance k in our measure, translates to about significance $-\log(k)$ insignificance in the [BCG17] measure. Thus, a d.p.d with significance Δ^i in this thesis translates roughly to one with insignificance $i \cdot \log(1/\Delta)$ and vice versa. The [BCG17] paper insignificance rises only in jumps of g. Since the $i \cdot g$'th d.p.d in the (k, g, w)-LMR has insignificance roughly $i \cdot g$ we essentially see that $\log(1/\Delta) = g$. Thus, we replace the (k, g, w)-LMRs with sequences of $k/g + 1$ d.p.ds and we do not need any more the requirement that insignificance rises in jumps of g.

Thus, the mysterious g is equivalent to the parameter Δ we have. We need to choose Δ so that the error is satisfied, and before we saw that Δ has to be at least $\Omega(\log(n))$. This forces the final output length to be $\Omega(\log^2 n)$. Finding a way to relax this parameter will probably lead to the seminal result of an explicit pseudo-random pseudo-distribution with seed length $o(\log^2 n)$.

2 Preliminaries

For a $w \times w$ matrix B:

$$||B|| = \| B \|_\infty = \max_{i \in [w]} (\sum_{j=1}^{w} |B_{i,j}|),$$

and $||B||_{max} = \max_{i,j \in [w]}(|B_{i,j}|)$. All sets are *multi-sets*, i.e., elements may appear with multiplicity. For an integer n, $[n] = \{1, \ldots, n\}$.

3 Samplers

Definition 3.1. *Let $0 < \epsilon, \delta < 1$, A left-regular bipartite graph $G = (L, R, E)$ is an (ϵ, δ)-sampler if for every $f : R \to [0, 1]$ for all but δ-fraction of vertices $v \in L$ it holds that:*

$$\left\| \mathbb{E}_{w \in \Gamma(v)} f(w) - E_{w \in R} f(w) \right\| \leq \varepsilon.$$

We first check that samplers for functions $f : R \to [0, 1]$ give also good samplers for arbitrary functions $f : R \to [a, b]$:

Claim 3.2. *Let $a \geq b \geq 0$. Let $G = (L, R, E)$ be an (ϵ, δ)-sampler. Then for every function $f : R \to [a, b]$ and for all but δ-fraction of vertices $v \in L$, it holds that:*

$$|\mathbb{E}_{w \in \Gamma(v)} f(w) - E_{w \in R} f(w)| \leq \epsilon \cdot (b - a).$$

Proof: Define a function $h : R \to [0, 1]$ by $h(w) = \frac{f(w)-a}{b-a}$. As G is an (ϵ, δ)-sampler, for all but δ-fraction of vertices $v \in L$ it holds that:

$$|\mathbb{E}_{w \in \Gamma(v)} h(w) - E_{w \in R} h(w)| \leq \epsilon.$$

Therefore, for all but δ-fraction of vertices $v \in L$ we have:

$$|\mathbb{E}_{w \in \Gamma(v)} \frac{f(w) - a}{b - a} - E_{w \in R} \frac{f(w) - a}{b - a}| \leq \epsilon,$$

as desired. ■

Theorem 3.3. *([GW97]) For every integer n and all $\epsilon, \delta > 0$, there exists an (ϵ, δ)-sampler $BSamp(n, \epsilon, \delta) = (L, R, E)$, with $|L| = |R| = n$, having degree $d = O(\delta^{-1} \cdot \epsilon^{-2})$.*

Theorem 3.4. *([Zuc06]) There exists a universal constant $c \geq 1$ such that the following holds. For all integers l, r and all $\epsilon, \delta > 0$ for which $l \geq r/\delta^2$, there exists an (ϵ, δ)-sampler $UBSamp(l, r, \epsilon, \delta) = ([l], [r], E)$, with degree $d = (\log(\delta^{-1}) \cdot \epsilon^{-1})^c$.*

4 Pseudo-distributions, d.p.ds and cubes

Definition 4.1. *(Pseudo-distribution) A* pseudo-distribution *over Λ with support C is $A = \{(\alpha_c, x_c)\}_{c \in C}$ where $\alpha_c \in \mathbb{R}$ and $x_c \in \Lambda$. The support size is $|C|$. If α is clear from the context then we omit it. If $\alpha_c \geq 0$ for every $c \in C$ and $\sum_{c \in C} \alpha_c = 1$, we say A is a distribution.*

Definition 4.2. *(d.p.d) $\mathbf{A}_{r,C} = \{A_i\}_{i \in [2^r]}$ is a d.p.d. over Λ with $|2^r|$ rows and $|C|$ columns, if each A_i is a pseudo-distribution over Λ with support C. We denote $Rows(\mathbf{A}) = |2^r|$ and $Col(\mathbf{A}) = |C|$.*

The acronym d.p.d. stands for distribution over pseudo distributions. We adopt the convention that we denote elements in Λ by small letters, pseudo-distributions by capital letters and d.p.d.s by **bold** capital letters.

4.1 Operations on d.p.ds and pseudo-distributions

Next we define union and tensor of pseudo-distributions over Λ.

Definition 4.3. *(Operations on pseudo-distributions) Suppose $b \in \mathbb{R}$ and A, A' are two pseudo-distributions over Λ with support C, C' respectively, i.e.: $A = \{(\alpha_c, x_c)\}_{c \in C}$, $A' = \left\{(\alpha'_{c'}, x'_{c'})\right\}_{c' \in C'}$. We define:*

$$\begin{aligned} b \cdot A &= \{(b\alpha_c, x_c)\}_{c \in C} \\ A + A' &= \{(\alpha_c, x_c)\}_{c \in C} \cup \left\{(\alpha'_{c'}, x'_{c'})\right\}_{c' \in C'} \\ A \otimes A' &= \left\{(\alpha_c \cdot \alpha'_{c'}, x_c \cdot x'_{c'})\right\}_{c, c' \in C \times C'}, \end{aligned}$$

where $\cdot$ is an associative multiplication (e.g., if Λ is a set of strings, $\cdot$ may be string concatenation). Notice that $b \cdot A$ is a pseudo-distribution over Λ with support C, $A + A'$ over Λ with support $C \cup C'$ and $A \otimes A'$ over $\Lambda \cdot \Lambda = \{x \cdot x' \mid x, x' \in \Lambda\}$ with support $C \times C'$.

We also let $A - A'$ denote $A + ((-1) \cdot A')$. Notice that the tensor of two distributions is, again, a distribution, but this is not the case with addition or subtraction.

Definition 4.4. *(Operations on d.p.ds) Suppose $\mathbf{A}_{r,C} = \{A_i\}_{i\in[2^r]}$ and $\mathbf{B}_{r',C'} = \{B_i\}_{i\in[2^{r'}]}$ are two d.p.d.s over Λ and $b \in \mathbb{R}$. Define:*

- $(b\mathbf{A})_{r,C} = \{b \cdot A_i\}_{i\in[2^r]}$.
- $(\mathbf{A} \otimes \mathbf{B})_{r+r',C\times C'} = \{A_i \otimes B_{i'}\}_{i\in[2^r],i'\in[2^{r'}]}$.
- *When $r = r'$, $(\mathbf{A} +_{cols} \mathbf{B})_{r,C\cup C'} = \{A_i + B_i\}_{i\in[2^r]}$. If it is clear from the context we denote $+_{cols}$ by $+$. If we add several d.p.ds we use the Σ symbol.*

Definition 4.5. *(Row summation) Let $\{\mathbf{A}_i\}_{i=1}^T$ be T d.p.ds, where $\mathbf{A}_i$ has 2^{r_i} rows and C columns. Suppose*

$$r_{max} = \max_i r_i.$$

We define $\sum_i \mathbf{A}_i$ to be a d.p.d. with $T \cdot 2^{r_{max}}$ rows and C columns, where each d.p.d. $\mathbf{A}_i$ is repeated $2^{r_{max}-r_i}$ times (thus each d.p.d. $\mathbf{A}_i$ is allocated $2^{r_{max}}$ rows).

4.2 Derandomized Tensoring

Definition 4.6. *Suppose $\mathbf{A}_{r,C}$ and $\mathbf{B}_{r',C'}$ are two d.p.d.s. Let $G = ([2^r], [2^{r'}], E)$ be a bipartite graph. We let $\mathbf{A}\circ_G\mathbf{B}$ be the d.p.d. $\mathbf{C}_{[E],C\times C'}$ where for $e = (i, i') \in E$ we have $\mathbf{C}_e = A_i \otimes B_{i'}$.*

Definition 4.7. *Suppose $\mathbf{A}_{r,C}$ and $\mathbf{B}_{r',C'}$ are two d.p.d.s. with $r \geq r'$. Let $G = ([2^r], [2^{r'}], E)$ be a bipartite graph with left regular degree D where $\Gamma(i, d)$ is the d neighbour of i in E. We let $\mathbf{A}\bullet_G\mathbf{B}$ be the d.p.d. $\mathbf{C}_{r,C\times C'\times[D]}$ where $\mathbf{C}_i = \frac{1}{D} \cdot \sum_{d\in[D]} A_i \otimes B_{\Gamma(i,d)}$. The sum here is many $+_{cols}$ operations.*

Definition 4.8. *Suppose $\mathbf{A}_{r,C}$ and $\mathbf{B}_{r',C'}$ are two d.p.d.s. with $r < r'$. Let $G = ([2^r], [2^{r'}], E)$ be a bipartite graph with right regular degree D' where $\Gamma(r', d')$ is the d' neighbour of r' in E. We let $\mathbf{A}\bullet_G\mathbf{B}$ be the d.p.d. $\mathbf{C}_{r',C\times C'\times[D']}$ where $\mathbf{C}_{i'} = \frac{1}{D'} \cdot \sum_{d'\in[D']} A_{\Gamma(i',d')} \otimes B_{i'}$.*

We assume for every two d.p.d.s $\mathbf{A}$ and $\mathbf{B}$ and every $\varepsilon', \delta' > 0$, we fixed a sampler $G = (L = Rows(\mathbf{A}), R = Rows(\mathbf{B}), E)$ that is an (ε', δ') sampler over M. We let $\mathbf{A} \circ_{\varepsilon',\delta'} \mathbf{B}$ denote $\mathbf{A} \circ_G \mathbf{B}$, similarly $\mathbf{A} \bullet_{\varepsilon',\delta'} \mathbf{B}$ denotes $\mathbf{A} \bullet_G \mathbf{B}$.

4.3 Telescopic tensoring

Definition 4.9. *Suppose $\mathbf{A}_{r,C}$ and $\mathbf{B}_{r',C'}$ are two d.p.d.s. Let $G_1 = ([2^r], [2^{r'}], E_1), G_2 = ([2^r], [2^{r'}], E_2)$ be two bipartite graphs on $2^r \times 2^{r'}$, G_b with regular left-degree D_b (for $b = 1, 2$). Notice that $(\mathbf{A} \bullet_{G_1} \mathbf{B})_{r,C\times C'\times[D_1]}$ and $(\mathbf{A} \bullet_{G_2} \mathbf{B})_{r,C\times C'\times[D_2]}$. We define the d.p.d $(\mathbf{A}\bullet_{G_1-G_2}\mathbf{B})_{r,C\times C'\times(D_1\cup D_2)}$ by*

$$\mathbf{A}\bullet_{G_1-G_2}\mathbf{B} = \mathbf{A} \bullet_{G_1} \mathbf{B} \ +_{cols} \ (-\mathbf{A} \bullet_{G_2} \mathbf{B}).$$

Remark. *Suppose $\mathbf{A}_{r,C}$ is a d.p.d — note that by definition this d.p.d has 2^r rows — but for every $r' \geq r$ we may assume, without loss of generality, that it has $2^{r'}$ rows. This is because we may duplicate each row $2^{r'-r}$ times. We will do this transformation implicitly throughout the thesis (for instance, when doing derandomized tensoring we might use graphs with more vertices on each side than rows in the d.p.d).*

4.4 Derandomized product of Cubes

Definition 4.10. *(Cube) A cube is a sequence of d.p.ds and positive real coefficients. We denote it by $\mathcal{C} = (\alpha_i, \mathbf{C}_i)_{i=0}^{A}$ where each $\mathcal{C}_i$ is a d.p.d. and $\alpha_i \geqslant 0$ is a real number, in which case we say that the cube $\mathcal{C}$ has $A+1$ layers.*

Given two cubes: $\mathcal{A} = (\alpha_a, \mathbf{A}_a)_{a=0}^{\mathrm{A}}$ and $\mathcal{B} = (\beta_b, \mathbf{B}_b)_{b=0}^{\mathrm{A}}$,
, two functions:

$$\varepsilon, \delta : [\mathrm{A}]^3 \to R_{\geq 0}.$$

We define the cube:

$$\mathcal{A} \cdot_{\varepsilon,\delta} \mathcal{B} = (\gamma_c, \mathbf{C}_c)_{c=0}^{\mathrm{A}}$$

So that their derandomized product is as follows:

- We define $s : [\mathrm{A}]^3 \to \mathbb{N}$ by:

$$s(a,b,k) \quad = \quad a+b+k$$

- Given $0 \leq c \leq \mathrm{A}$ we define:

$$\begin{aligned} N_c &= \{(a,b,k) \mid a,b,k \geq 0, s(a,b,k) = c\}, \\ \circ_c &= \{(a,b,k) \in N_c \mid k = 0 \; and \; \min\{a,b\} = 0\}, \\ \bullet_c &= \{(a,b,k) \in N_c \mid k > 0 \; or \; \min\{a,b\} > 0\}. \end{aligned}$$

Definition 4.11. *For every $a, b \in [\mathrm{A}]$ we denote $K_{a,b} = \max\{k | s(a,b,k) \leq \mathrm{A}\}$.*

- For every $(a,b,k) \in N_c$ define:

$$\mathbf{C}_{a,b,k} \quad = \quad \begin{cases} \mathbf{A}_a \circ_{\varepsilon(a,b,k),\delta(a,b,k)} \mathbf{B}_b & \text{If } (a,b,k) \in \circ_c \\ \mathbf{A}_a \bullet_{\varepsilon(a,b,k),\delta(a,b,k)} \mathbf{B}_b & \text{If } (a,b,k) \in \bullet_c,\ k = 0 \\ \mathbf{A}_a \bullet_{\varepsilon(a,b,k),\delta(a,b,k)} \mathbf{B}_b -_{cols} \mathbf{A}_a \bullet_{\varepsilon(a,b,k-1),\delta(a,b,k-1)} \mathbf{B}_b & \text{If } (a,b,k) \in \bullet_c,\ k \geq 1. \end{cases}$$

With this notation, for $c \in [\mathrm{A}]$ we let $m_c = \max_{(a,b,k) \in N_c}(\alpha_a \cdot \beta_b)$ and define

$$\begin{aligned} \mathbf{C}_c &= \sum_{(a,b,k) \in N_c} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot \mathbf{C}_{a,b,k}, \\ \gamma_c &= |N_c| \cdot m_c \end{aligned}$$

Where the sum here is row summation

5 Value, Average-norm and Max-norm

We fix a (possibly) non-commutative ring M over $\mathbb{R}$ equipped with a sub-multiplicative norm $\|\cdot\|$, and a mapping $val : \Lambda \to \mathrm{M}$. We assume $\max_{a \in \Lambda} \| val(a) \| = 1$. Note that after fixing some (n,w)-ROBP B we can for every $\sigma \in \Sigma^i$ (for $i \leq n$) think of the $w \times w$ transition matrix which corresponds to σ with regard to B and the norm will be the infinity norm.

5.1 Value

Definition 5.1. *(Value) The* value *of a pseudo-distribution $A = \{(\alpha_c, x_c)\}_{c \in C}$ is*

$$val(A) = \sum_{c \in C} \alpha_c \cdot val(x_c).$$

The value *of a d.p.d. $\mathbf{A} = \{A_i\}_{i \in [2^r]}$ is*

$$val(\mathbf{A}) = \mathbb{E}_{i \in [2^r]}\, val(A_i).$$

Also define $sum(\mathbf{A}) = \sum_i\, val(A_i)$, i.e., $sum(\mathbf{A}) = Rows(\mathbf{A}) \cdot val(\mathbf{A})$.

Claim 5.2. *Let A, B be two pseudo-distributions. Then:*

$$\begin{aligned} val(cA) &= c \cdot val(A), \\ val(A + B) &= val(A) + val(B), and, \\ val(A \otimes B) &= val(A) \cdot val(B). \end{aligned}$$

Claim 5.3. *Let $\mathbf{A}_{r_1,C_1}, \mathbf{B}_{r_2,C_2}$ be two d.p.d's, we have:*

$$\begin{aligned} val(c\mathbf{A}) &= c \cdot val(\mathbf{A}), \\ val(\mathbf{A} \otimes \mathbf{B}) &= val(\mathbf{A}) \cdot val(\mathbf{B}), \\ val(\mathbf{A} +_{cols} \mathbf{B}) &= \mathbb{E}_i[val(A_i + B_i)] = val(\mathbf{A}) + val(\mathbf{B}), \\ val(\sum_{i=1}^{T} \mathbf{A}_i) &= \frac{1}{T} \cdot \sum_{i=1}^{T} val(\mathbf{A}_i). \end{aligned}$$

Proof: We prove the last item. Let $R_i = Rows(\mathbf{A}_i)$ and $R_{max} = \max_i R_i$. Then, $Rows(\sum_i \mathbf{A}_i) = T \cdot R_{max}$ and

$$val(\sum_i \mathbf{A}_i) = \quad = \quad \frac{1}{Rows(\sum_i \mathbf{A}_i)} \cdot \sum_i \frac{R_{max}}{R_i} sum(\mathbf{A}_i) = \frac{1}{T} \cdot \sum_{i=1}^{T} val(\mathbf{A}_i).$$

■

Claim 5.4.

$$\begin{aligned} val(\mathbf{A} \circ_G \mathbf{B}) &= val(\mathbf{A} \bullet_G \mathbf{B}) = \mathbb{E}_{(i,i') \in E}[val(A_i) \cdot val(B_{i'})] \\ val(\mathbf{A} \bullet_{G_1 - G_2} \mathbf{B}) &= val(\mathbf{A} \bullet_{G_1} \mathbf{B}) - val(\mathbf{A} \bullet_{G_2} \mathbf{B}). \end{aligned}$$

Proof: We have

$$\begin{aligned} val(\mathbf{A} \circ_G \mathbf{B}) &= \mathbb{E}_{(i,i') \in E} val(A_i \otimes B_{i'}) \;=\; \mathbb{E}_{(i,i') \in E}[val(A_i) \cdot val(B_{i'})] \\ val(\mathbf{A} \bullet_G \mathbf{B}) &= \mathbb{E}_{i \in V} \frac{1}{D} \cdot val(\cup_{d \in [D]} A_i \otimes B_{\Gamma(i,d)}) = \mathbb{E}_{(i,i') \in E}[val(A_i) \cdot val(B_{i'})] \end{aligned}$$

■

Definition 5.5. *Let $\mathcal{C} = (\gamma_a, \mathbf{C}_a)_{a=0}^{A}$ be a cube. The* value *of the cube is $val(\mathcal{C}) = \sum_{a=0}^{\mathrm{A}} \gamma_a val(\mathbf{C}_a)$.*

Lemma 5.6. *Let $\mathcal{A} = (\alpha_a, \mathbf{A}_a)_{a=0}^{\mathrm{A}}$, $\mathcal{B} = (\beta_b, \mathbf{B}_b)_{b=0}^{\mathrm{A}}$ be cubes. Let $\mathcal{C} = \mathcal{A} \cdot_{\varepsilon,\delta} \mathcal{B} = (\gamma_c, \mathbf{C}_c)_{c=0}^{\mathrm{A}}$. Then,*

$$val(\mathcal{C}) = \sum_{a,b:a+b\leq \mathrm{A}} \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a \bullet_{K_{a,b}} \mathbf{B}_b).$$

Proof: Keeping in mind that:

$$\mathbf{C}_c \quad = \quad \sum_{(a,b,k)\in N_c} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot \mathbf{C}_{a,b,k},$$

Note that the sum here stands for row summation, and hence by definition each d.p.d in this sum is duplicated such that every one of them has the same amount of rows,
hence we get by Claim 5.3:

$$val(\mathbf{C}_c) \quad = \quad \frac{1}{|N_c|} \cdot \sum_{(a,b,k)\in N_c} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot val(\mathbf{C}_{a,b,k})$$

And we continue:

$$\begin{aligned} val(\mathbf{C}_c) \quad &= \quad \frac{1}{|N_c|} \cdot \sum_{(a,b,k)\in N_c} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot val(\mathbf{C}_{a,b,k}) \\ &= \quad \frac{1}{|N_c|}\Big(\sum_{(a,b,0)\in \circ_c} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot val(\mathbf{A}_a \circ_0 \mathbf{B}_b) + \sum_{(a,b,0)\in \bullet_c} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot val(\mathbf{A}_a \bullet_0 \mathbf{B}_b) + \\ &\qquad \sum_{(a,b,k)\in \bullet_c, k>0} \frac{\alpha_a \cdot \beta_b}{m_c} \cdot val(\mathbf{A}_a \bullet_k \mathbf{B}_b) - \frac{\alpha_a \cdot \beta_b}{m_c} \cdot val(\mathbf{A}_a \bullet_{k-1} \mathbf{B}_b)\Big). \end{aligned}$$

It therefore follows that

$$val(\mathcal{C}) = \sum_{c=0}^{\mathrm{A}} \gamma_c \cdot val(\mathbf{C}_c) \quad = \quad \sum_{a,b:a+b\leq \mathrm{A}} \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a \bullet_{K_{a,b}} \mathbf{B}_b),$$

where we have used that $val(\mathbf{A}_a \circ_k \mathbf{B}_b) = val(\mathbf{A}_a \bullet_k \mathbf{B}_b)$. ■

5.2 Max norm

Definition 5.7. *The* max norm *of a d.p.d. $\mathbf{A} = \{A_i\}_{i\in[2^r]}$ is*

$$mn(\mathbf{A}) = \max_{i\in[2^r]} \| \, val(A_r) \, \| \, .$$

Claim 5.8. *(Max norm and basic d.p.d operations) Let $\mathbf{A}_{r_1,C_1}, \mathbf{B}_{r_2,C_2}$ be two d.p.d's. We have:*

$$\begin{aligned} mn(c\mathbf{A}) &= c \cdot mn(\mathbf{A}), \\ mn(\mathbf{A} \otimes \mathbf{B}) &= mn(\mathbf{A}) \cdot mn(\mathbf{B}), \\ mn(\mathbf{A} +_{cols} \mathbf{B}) &\leq mn(\mathbf{A}) + mn(\mathbf{B}), \\ mn(\mathbf{A} +_{rows} \mathbf{B}) &= \max\{mn(\mathbf{A}), mn(\mathbf{B})\}, \\ mn(\mathbf{A} \circ_G \mathbf{B}) &\leq mn(\mathbf{A}) \cdot mn(\mathbf{B}), \\ mn(\mathbf{A} \bullet_G \mathbf{B}) &\leq mn(\mathbf{A}) \cdot mn(\mathbf{B}) \\ mn(\mathbf{A} \bullet_{G_2-G_1} \mathbf{B}) &\leq 2 \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}). \end{aligned}$$

Proof: The first claims are almost immediate. For the last two claims:

$$mn(\mathbf{A} \circ_G \mathbf{B}) = \max_{e=(i,i')} \| val(A_i) \cdot val(B_{i'}) \| \leq mn(\mathbf{A}) \cdot mn(\mathbf{B})$$

Similarly, $mn(\mathcal{A} \bullet_G \mathcal{B}) = \max_r \left\| \frac{1}{D} \cdot val(A_r) \cdot \sum_{r' \in \Gamma(r)} val(B_{r'}) \right\| \leq mn(\mathbf{A}) \cdot mn(\mathbf{B})$.
For the last claim denote $\mathbf{C} = \mathbf{A} \bullet_{G_1-G_2} \mathbf{B}$, since the norm is sub-multiplicative we have for every $i \in Rows(\mathbf{C})$:

$$\begin{aligned} \| val(C_i) \| &\leq \| val(A_i) \| \, \| \mathbb{E}_{j \sim \Gamma_1} val(B_j) - \mathbb{E}_{j \sim \Gamma_2} val(B_j) \| \\ &\leq \| val(A_i) \| \, (\| \mathbb{E}_{j \sim \Gamma_1} val(B_j) \| + \| \mathbb{E}_{j \sim \Gamma_2} val(B_j) \|) \\ &\leq mn(\mathbf{A}) \cdot 2 \cdot mn(\mathbf{B}) \end{aligned}$$

■

5.3 Average Norm

Definition 5.9. *(Average) The* average norm *of a d.p.d. $\mathbf{A} = \{A_i\}_{i \in [2^r]}$ is*

$$an(\mathbf{A}) = \mathbb{E}_{i \in [2^r]} \| val(A_i) \| .$$

By the triangle inequality:

Fact 5.10. *Let $\mathbf{A} = \{A_i\}_{i \in [2^r]}$ be a d.p.d. Then, $\| val(\mathbf{A}) \| = \| \mathbb{E}_i val(A_i) \| \leq E_i \| val(A_i) \| = an(\mathbf{A})$.*

Claim 5.11. *Let $\mathbf{A}_{r_1,C_1}, \mathbf{B}_{r_2,C_2}$ be two d.p.d's. We have:*

$$\begin{aligned} an(c\mathbf{A}) &= c \cdot an(\mathbf{A}), \\ an(\mathbf{A} \otimes \mathbf{B}) &\leq an(\mathbf{A}) \cdot an(\mathbf{B}), \\ an(\mathbf{A} +_{cols} \mathbf{B}) &\leq an(\mathbf{A}) + an(\mathbf{B}), \\ an(\sum_{i=1}^{T} \mathbf{A}_i) &= \frac{1}{T} \sum_{i=1}^{T} an(\mathbf{A}_i). \end{aligned}$$

Proof: For the second item:

$$\begin{aligned} an(\mathbf{A} \otimes \mathbf{B}) &= \mathbb{E}_{i,i'} \left\| val((A \otimes B)_{i,i'}) \right\| = \mathbb{E}_{i,i'} \| val(A_i) \cdot val(B_{i'}) \| \\ &\leq \mathbb{E}_i \| val(A_i) \| \cdot \mathbb{E}_{i'} \| val(B_{i'}) \| = an(\mathbf{A}) \cdot an(\mathbf{B}). \end{aligned}$$

For the third line,

$$an(\mathbf{A} +_{cols} \mathbf{B}) \quad = \quad \mathbb{E}_i \| val(A_i + B_i) \| \leq \mathbb{E}_i \| val(A_i) \| + \| val(B_i) \| = an(\mathbf{A}) + an(\mathbf{B}).$$

For the last item:

$$\begin{aligned} an(\sum_{i=1}^{T} \mathbf{A}_i) &= \mathbb{E}_{i \in [T \cdot 2^{r_{max}}]} \| val(A_i) \| \\ &= \mathbb{E}_{i \in [T]} [\mathbb{E}_{j \in [2^{r_{max}}]} \| val((A_i)_j) \|] \\ &= \frac{1}{T} \sum_{i=1}^{T} an(\mathbf{A}_i) \end{aligned}$$

■

Claim 5.12. *Let $\mathbf{A}_{r,C_1}, \mathbf{B}_{r',C_2}$ be two d.p.d's and $\epsilon, \delta > 0$. Then,*

$$an(\mathbf{A} \bullet_{\varepsilon,\delta} \mathbf{B}) \quad \leq \quad an(\mathbf{A}) \cdot an(\mathbf{B}) + \delta \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}) + \varepsilon \cdot mn(\mathbf{B}) \cdot an(\mathbf{A}).$$

The same holds for $an(\mathbf{A} \circ_{\varepsilon,\delta} \mathbf{B})$.

Proof: First note that:

$$an(\mathbf{A} \bullet_{\varepsilon,\delta} \mathbf{B}) \quad = \quad \mathbb{E}_i \left\| \frac{1}{D} \sum_{i' \in \Gamma(i)} val(A_i \otimes B_{i'}) \right\| \leq \mathbb{E}_i \| val(A_i) \| \cdot \mathbb{E}_{i' \in \Gamma(i)} \| val(B_{i'}) \|$$

The same also holds for $A \circ_{\varepsilon,\delta} B$ as:

$$an(\mathbf{A} \circ_{\varepsilon,\delta} \mathbf{B}) = \mathbb{E}_{(i,i') \in E} \| val(A_i) \cdot val(B_{i'}) \| \leq \mathbb{E}_i \| val(A_i) \| \cdot \mathbb{E}_{i' \in \Gamma(i)} \| val(B_{i'}) \|$$

Define $f : [2^{r'}] \to [0, mn(\mathbf{B})]$ by $f(i') = \| val(B_{i'}) \|$. Since G is an (ε, δ) sampler we get that except for a set V of size at most $\delta \cdot 2^r$, for every vertex $i \in [2^r] \setminus V$:

$$|\mathbb{E}_{i' \in \Gamma(i)} f(i') - \mathbb{E}_{i' \in [2^{r'}]} f(i')] \leq \epsilon \cdot mn(\mathbf{B}).$$

For $i \in V$ we have the trivial bound $|\mathbb{E}_{i' \in \Gamma(r)} f(i')| \leq mn(\mathbf{B})$. Altogether,

$$\begin{aligned} \mathbb{E}_i \| val(A_i) \| \cdot E_{i' \in \Gamma(i)} f(i') &\leq \frac{|V|}{2^r} \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}) + \frac{1}{2^r} \sum_{i \notin V} \| val(A_i) \| \cdot (\mathbb{E}_{i' \in [2^{r'}]} f(r') + \varepsilon \cdot mn(\mathbf{B})) \\ &\leq \delta \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}) + \mathbb{E}_{i,i'} \| val(A_i) \| \cdot \| val(B_{i'}) \| + \varepsilon \cdot mn(\mathbf{B}) \cdot \mathbb{E}_i \| val(A_i) \| \\ &= an(\mathbf{A}) \cdot an(\mathbf{B}) + \delta \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}) + \varepsilon \cdot mn(\mathbf{B}) \cdot an(\mathbf{A}). \end{aligned}$$

■

Lemma 5.13. *Let $\mathbf{A}, \mathbf{B}$ be two d.p.d's, let $\tau \in (0, 1]$ and:*

$$\begin{aligned} \epsilon &\leq \frac{an(\mathbf{B}) \cdot \tau}{8mn(\mathbf{B})} \\ \delta &\leq \frac{an(\mathbf{A}) \cdot an(\mathbf{B}) \cdot \tau}{8mn(\mathbf{A}) \cdot mn(\mathbf{B})} \end{aligned}$$

then:

$$an(\mathbf{A} \bullet_{\varepsilon,\delta} \mathbf{B}) \leq an(\mathbf{A} \circ_{\varepsilon,\delta} \mathbf{B}) \quad \leq \quad an(\mathbf{A}) \cdot an(\mathbf{B}) \cdot 2^{\tau}$$

Proof: Substituting ϵ and δ in the previous lemma we get:

$$\begin{aligned} an(\mathbf{A} \bullet_{\varepsilon,\delta} \mathbf{B}) \leq an(\mathbf{A} \circ_{\varepsilon,\delta} \mathbf{B}) &\leq an(\mathbf{A}) \cdot an(\mathbf{B}) + \delta \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}) + \varepsilon \cdot mn(\mathbf{B}) \cdot an(\mathbf{A}) \\ &\leq an(\mathbf{A}) \cdot an(\mathbf{B})(1 + 2\tau/8) \leq an(\mathbf{A}) \cdot an(\mathbf{B}) \cdot e^{2\tau/8} \\ &\leq an(\mathbf{A}) \cdot an(\mathbf{B}) \cdot 2^{\tau} \end{aligned}$$

■

Lemma 5.14. *Let $\mathbf{A}, \mathbf{B}$ be two d.p.d's such that $Col(\mathbf{B}) = 1$ and all of his coefficients are 1 then:*

$$an(\mathbf{A} \circ_{\epsilon,\delta} \mathbf{B}) \leq an(\mathbf{A})$$

Proof:

$$an(\mathbf{A} \circ_{\epsilon,\delta} \mathbf{B}) = E_{i,i'} \| val(A_i \otimes B_{i'}) \| \leq E_i \| val(A_i) \| = an(\mathbf{A})$$

■

Lemma 5.15. *Let $\mathbf{A}, \mathbf{B}$ be two d.p.d's and G_1, G_2 be $(\epsilon_1, \delta_1), (\epsilon_2, \delta_2)$-samplers correspondingly, such that $\epsilon_1 \leqslant \epsilon_2$ and $\delta_1 \leqslant \delta_2$ then:*

$$an(\mathbf{A} \bullet_{G_1 - G_2} \mathbf{B}) \leq 4w^2 \cdot mn(\mathbf{B})(\epsilon_2 \cdot an(\mathbf{A}) + mn(\mathbf{A}) \cdot \delta_2)$$

This is done assuming that the value function gives for every $\sigma \in \Sigma^i$ a real $w \times w$ stochastic matrix.

Proof: First, note that for $r \in [Rows(\mathbf{A})]$:

$$\| val((\mathbf{A} \bullet_{G_1 - G_2} \mathbf{B})_r) \| \leq \| val(A_r) \| \left\| \mathbb{E}_{j \sim \Gamma_1(r)}[val(B_j)] - \mathbb{E}_{j \sim \Gamma_2(r)}[val(B_j)] \right\|$$

now note that:

$$\begin{aligned} \left\| \mathbb{E}_{j \sim \Gamma_1(r)}[val(B_j)] - \mathbb{E}_{j \sim \Gamma_2(r)}[val(B_j)] \right\| &\leq w \left\| \mathbb{E}_{j \sim \Gamma_1(r)}[val(B_j)] - \mathbb{E}_{j \sim \Gamma_2(r)}[val(B_j)] \right\|_{max} \\ &\leq w(\left\| \mathbb{E}_{j \sim \Gamma_1(r)}[val(B_j)] - val(\mathbf{B}) \right\|_{max} \\ &+ \left\| \mathbb{E}_{j \sim \Gamma_2(r)}[val(B_j)] - val(\mathbf{B}) \right\|_{max}) \end{aligned}$$

now for every $\alpha, \beta \in [w], i \in \{1, 2\}, r \in [Rows(\mathbf{A})]$ define:

$$\epsilon_i^{\alpha,\beta}(r) = |\mathbb{E}_{j \sim \Gamma_i(r)}[val(B_j)]_{\alpha,\beta} - val(\mathbf{B})_{\alpha,\beta}|$$

and also: $\epsilon_i(r) = max_{\alpha,\beta}(\epsilon_i^{\alpha,\beta}(r))$
so by the previous inequality we get:

$$\left\| \mathbb{E}_{j \sim \Gamma_1(r)}[val(B_j)] - \mathbb{E}_{j \sim \Gamma_2(r)}[val(B_j)] \right\| \leq w(\epsilon_1(r) + \epsilon_2(r))$$

now we define the function $f_{\alpha,\beta} : Rows(\mathbf{B}) \to [mn(\mathbf{B})]$, in the following way:

$$f_{\alpha,\beta}(r') = val(B_{r'})_{\alpha,\beta}$$

now since G_i is an (ϵ_i, δ_i)-sampler there exists a set $S_i^{\alpha,\beta} \subset [Rows(\mathbf{A})]$ of size $|S_i^{\alpha,\beta}| \geqslant (1 - \delta_i)Rows(\mathbf{A})$ such that for every $r \in S_i^{\alpha,\beta}$ it holds that:

$$|\mathbb{E}_{j\sim\Gamma_i(r)}[val(B_j)_{\alpha,\beta}] - E_{j\in Rows(\mathbf{B})}[val(B_j)_{\alpha,\beta}]| = \epsilon_i^{\alpha,\beta}(r) \leqslant \epsilon_i \cdot mn(\mathbf{B})$$

now we'll define the set $S = \bigcap_{\alpha,\beta=1}^{w}(S_1^{\alpha,\beta} \cap S_2^{\alpha,\beta})$, note that:

$$|S| \geqslant (1 - (\delta_1 + \delta_2)w^2)Rows(\mathbf{A}) \geqslant (1 - 2\delta_2 w^2)Rows(\mathbf{A})$$

and for every $r \in S$:

$$\epsilon_1(r) + \epsilon_2(r) \leqslant (\epsilon_1 + \epsilon_2) \cdot mn(\mathbf{B})$$

combining we get:

$$\begin{aligned}
an(\mathbf{A} \bullet_{G_1-G_2} \mathbf{B}) &= \mathbb{E}_{r\in Rows(\mathbf{A})} \| \, val((\mathbf{A} \bullet_{G_1-G_2} \mathbf{B})_r) \, \| \\
&\leq \mathbb{E}_{r\in Rows(\mathbf{A})}[\| \, val(A_r) \, \| \cdot w(\epsilon_1(r) + \epsilon_2(r))] \\
&\leq w(\mathbb{E}_{r\in Rows(\mathbf{A})}[\| \, val(A_r) \, \| \cdot (\epsilon_1(r) + \epsilon_2(r))|r \in S] + 2mn(\mathbf{A}) \cdot mn(\mathbf{B}) \cdot Pr[r \notin S]) \\
&\leq w((\epsilon_1 + \epsilon_2) \cdot mn(\mathbf{B}) \cdot 2 \cdot an(\mathbf{A}) + 2mn(\mathbf{A}) \cdot mn(\mathbf{B}) \cdot (2\delta_2 w^2)) \\
&\leq 4w^2 \cdot mn(\mathbf{B})(\epsilon_2 \cdot an(\mathbf{A}) + \cdot mn(\mathbf{A}) \cdot \delta_2)
\end{aligned}$$

■

5.4 Error in one derandomization step

Lemma 5.16. *Let $\mathbf{A}_{r_1,C_1}$ and $\mathbf{B}_{r_2,C_2}$ be two d.p.d's over Λ and let $G = (R_1, R_2, E)$ be an (ϵ, δ)-sampler over M, then:*

$$\| \, val(\mathbf{A} \bullet_G \mathbf{B}) - val(\mathbf{A}) \cdot val(\mathbf{B}) \, \| \leq w \cdot (\epsilon \cdot mn(\mathbf{B}) \cdot an(\mathbf{A}) + 2\delta \cdot w^2 \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B}))$$

Proof: $val(\mathbf{A} \circ_G \mathbf{B}) = val(\mathbf{A} \bullet_G \mathbf{B}) = \mathbb{E}_{(i,i')\in E} val(A_i) \cdot val(B_{i'}) = \mathbb{E}_{i\in[2^{r_1}]} val(A_i) \cdot \mathbb{E}_{i'\in\Gamma(i)} val(B_{i'})$. Thus,

$$\begin{aligned}
\| \, val(\mathbf{A} \bullet_G \mathbf{B}) - val(\mathbf{A})val(\mathbf{B}) \, \| &= \left\| \, \mathbb{E}_{i\in[2^{r_1}]} val(A_i)\mathbb{E}_{i'\in\Gamma(v)} val(B_{i'}) - \mathbb{E}_{i\in[2^{r_1}]} val(A_i) \cdot \mathbb{E}_{i'\in[2^{r_2}]} val(B_{i'}) \, \right\| \\
&\leq \mathbb{E}_{i\in[2^{r_1}]} \| \, val(A_i) \, \| \cdot \left\| \, \mathbb{E}_{i'\in\Gamma(i)} val(B_{i'}) - \mathbb{E}_{i'\in[2^{r_2}]} val(B_{i'}) \, \right\|
\end{aligned}$$

now note that for every $i \in [2^{r_1}]$:

$$\begin{aligned}
\left\| \, \mathbb{E}_{j\sim\Gamma(i)}[val(B_j)] - \mathbb{E}_{j\sim[2^{r_2}]}[val(B_j)] \, \right\| &\leq w \left\| \, \mathbb{E}_{j\sim\Gamma(i)}[val(B_j)] - \mathbb{E}_{j\sim[2^{r_2}]}[val(B_j)] \, \right\|_{max} \\
&\leq w(\left\| \, \mathbb{E}_{j\sim\Gamma(i)}[val(B_j)] - val(\mathbf{B}) \, \right\|_{max}
\end{aligned}$$

now we define the function $f_{\alpha,\beta} : Rows(\mathbf{B}) \to [mn(\mathbf{B})]$, in the following way:

$$f_{\alpha,\beta}(i') = val(B_{i'})_{\alpha,\beta}$$

now since G is an (ϵ, δ)-sampler there exists a set $S^{\alpha,\beta} \subset [Rows(\mathbf{A})]$ of size $|S^{\alpha,\beta}| \geqslant (1-\delta)Rows(\mathbf{A})$ such that for every $i \in S^{\alpha,\beta}$ it holds that:

$$|\mathbb{E}_{j\sim\Gamma(i)}[val(B_j)_{\alpha,\beta}] - E_{j\in Rows(\mathbf{B})}[val(B_j)_{\alpha,\beta}]| \leqslant \epsilon \cdot mn(\mathbf{B})$$

. now we'll define the set $S = \bigcap_{\alpha,\beta=1}^{w} S^{\alpha,\beta}$, note that:

$$|S| \geqslant (1 - \delta w^2) Rows(\mathbf{A})$$

Combining everything we get:

$$\begin{aligned} \| \, val(\mathbf{A} \bullet_G \mathbf{B}) - val(\mathbf{A}) val(\mathbf{B}) \, \| &\leq \mathbb{E}_{i \in [2^{r_1}]} [\| \, val(A_i) \, \| \cdot w \left\| \, \mathbb{E}_{j \sim \Gamma(i)} [val(B_j)] - val(\mathbf{B}) \, \right\|_{max}] \\ &\leq w \cdot \frac{1}{2^{r_1}} (\sum_{i \in S} \| \, val(A_i) \, \| \cdot \epsilon \cdot mn(\mathbf{B}) + \sum_{i \notin S} mn(\mathbf{A}) \cdot 2mn(\mathbf{B})) \\ &\leq w \cdot (\epsilon \cdot mn(\mathbf{B}) \cdot an(\mathbf{A}) + 2\delta \cdot w^2 \cdot mn(\mathbf{A}) \cdot mn(\mathbf{B})) \end{aligned}$$

■

Lemma 5.17. *Let $\mathcal{A} = (\alpha_a, \mathbf{A}_a)_{a=0}^{\mathrm{A}}, \mathcal{B} = (\beta_a, \mathbf{B}_a)_{a=0}^{\mathrm{A}}$ be two cubes , $\varepsilon, \delta : \{0, \ldots, \mathrm{A}\}^3 \to \mathbb{R}_{\geq 0}$. Then:*

$$\begin{aligned} \| \, val(\mathcal{A} \cdot_{\varepsilon,\delta} \mathcal{B}) - val(\mathcal{A}) \cdot val(\mathcal{B}) \, \| \leq & \sum_{a,b:a+b \leq A} \alpha_a \cdot \beta_b \cdot w^3 \cdot 2\delta(a,b,K_{a,b}) \cdot mn(\mathbf{A}_a) \cdot mn(\mathbf{B}_b) \\ + & \sum_{a,b:a+b \leq A} \alpha_a \cdot \beta_b \cdot w \cdot \varepsilon(a,b,K_{a,b}) \cdot mn(\mathbf{B}_b) \cdot an(\mathbf{A}_a) \\ + & \sum_{a,b:a,b \leq \mathrm{A}, a+b > \mathrm{A}} \alpha_a \cdot \beta_b \cdot an(\mathbf{A}_a) \cdot an(\mathbf{B}_b). \end{aligned}$$

Proof: By Lemma 5.6

$$val(\mathcal{A} \cdot \mathcal{B}) = \sum_{a,b:a+b \leq \mathrm{A}} \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a \bullet_{K_{a,b}} \mathbf{B}_b)$$

Hence:

$$\begin{aligned} & \left\| val(\mathcal{A} \cdot \mathcal{B}) - \sum_{a,b:a+b \leq A} \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a) \cdot val(\mathbf{B}_b) \right\| \leq \\ \leq & \sum_{a,b:a+b \leq A} \left\| \, \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a \bullet_{K_{a,b}} \mathbf{B}_b) - \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a) \cdot val(\mathbf{B}_b) \, \right\| \\ \leq & \sum_{a,b:a+b \leq A} \alpha_a \cdot \beta_b \cdot w^3 \cdot 2\delta(a,b,K_{a,b}) \cdot mn(\mathbf{A}_a) \cdot mn(\mathbf{B}_b) \\ + & \sum_{a,b:a+b \leq A} \alpha_a \cdot \beta_b \cdot w \cdot \varepsilon(a,b,K_{a,b}) \cdot mn(\mathbf{B}_b) \cdot an(\mathbf{A}_a) \end{aligned}$$

where we have used Lemma 5.16. Also, by definition, $val(\mathcal{A}) = \sum_{a=0}^{\mathrm{A}} \alpha_a \cdot val(\mathbf{A}_a)$ and $val(\mathcal{B}) = \sum_{b=0}^{\mathrm{A}} \beta_b \cdot val(\mathbf{B}_b)$ so

$$val(\mathcal{A}) \cdot val(\mathcal{B}) = \sum_{a,b=0}^{\mathrm{A}} \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a) \cdot val(\mathbf{B}_b).$$

So:

$$\begin{aligned}\left\| val(\mathcal{A}) \cdot val(\mathcal{B}) - \sum_{a,b:a+b\leq A} \alpha_a \cdot \beta_b \cdot val(\mathbf{A}_a) \cdot val(\mathbf{B}_b) \right\| &\leq \sum_{a,b:a,b\leq \mathrm{A}, a+b>\mathrm{A}} \alpha_a \cdot \beta_b \cdot \| val(\mathbf{A}_a) \| \cdot \| val(\mathbf{B}_b) \| \\ &\leq \sum_{a,b:a,b\leq \mathrm{A}, a+b>\mathrm{A}} \alpha_a \cdot \beta_b \cdot an(\mathbf{A}_a) \cdot an(\mathbf{B}_b)\end{aligned}$$

where we have used Fact 5.10 in the last inequality. ■

6 The Pseudo-random Pseudo-distribution (PRPD)

In this section, using the notion of derandomized cube tensoring, we will state the construction of a PRPD over Σ^n given Σ and n as inputs. Throughout said construction we will make use of the parameters A and Δ which we will fix explicitly only at the end of the thesis.

- We define
$$\varepsilon, \delta : \bigcup_{c=0}^{\mathrm{A}} N_c \to R_{\geq 0},$$
as follows:
$$\begin{aligned}\varepsilon(a,b,k) &= \begin{cases} \Delta^9 & \text{If } \min(a,b) = k = 0 \\ \Delta^{9\cdot(\min(a,b)+k)} & \text{Otherwise.}\end{cases} \\ \delta(a,b,k) &= \begin{cases} \varepsilon(a,b,k) & \text{If } \lfloor \log a \rfloor = \lfloor \log b \rfloor \\ \Delta^{10\mathrm{A}} & \text{Otherwise.}\end{cases}\end{aligned}$$
This somewhat arbitrary choice is justified by the bounds in claim 8.4.
- Given ℓ we construct the cube $\mathcal{P}^{(\ell)}$ with A layers where the a'th layer (for $0 \leq a \leq \mathrm{A}$) is a d.p.d denoted $(\mathbf{P}_a^{(\ell)})_{R_a^{(\ell)}, C_a^{(\ell)}}$ over $\Lambda_\ell = \{0,1\}^{2^\ell}$ with $R_a^{(\ell)} = Rows(\mathbf{P}_a^{(\ell)})$ rows and $Cols_a^{(\ell)} = C(\mathbf{P}_a^{(\ell)})$ columns. We construct it inductively on ℓ:
- In the base case $\ell = 0$. $\mathcal{P}^{(0)}$ is the following cube with A layers:
  - For $a = 0$, the layer $(\mathbf{P}_0^{(0)})_{\log(\Sigma),1}$ is the d.p.d having Σ rows, where row $\sigma \in \Sigma$ has a single column $(1, b)$.
  - For all $a > 0$, $\mathbf{P}_a^{(0)}$ is empty with no rows or columns.

  Thus,
$$R_a^{(\ell=0)} = \begin{cases} \Sigma & \text{If } a = 0 \\ 0 & \text{If } a > 0 \end{cases}$$
and
$$C_a^{(\ell=0)} = \begin{cases} 1 & \text{If } a = 0 \\ 0 & \text{If } a > 0 \end{cases}$$

- Now assume we have constructed $\mathbf{P}_a^{(\ell)}$ for $0 \leq a < \mathrm{A}$. $\mathbf{P}_a^{(\ell)}$ has $R_a^{(\ell)}$ rows and $Cols_a^{(\ell)}$ columns.

  We let:
  $$\mathcal{P}^{(\ell+1)} = \mathcal{P}^{(\ell)} \cdot_{\varepsilon,\delta} \mathcal{P}^{(\ell)}.$$
  and our desired pseudo-random pseudo-distribution will be $\mathcal{P}^{(\log(n))}$, where the graphs we use are taken from the family $G_{a,b,k,\ell}$ which we will fix next.

- We define the graph family $G_{a,b,k,\ell}$ which is used throughout the construction. This family will be composed of $(\epsilon(a,b,k), \delta(a,b,k))$-samplers of two types, balanced and unbalanced samplers. In the next section we will prove that $R_a^{(\ell)} \leq \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log a\rfloor+\ell)}$ for some constant $c \geq 2$, keeping that in mind we will choose the samplers accordingly:

  - When $\lfloor\log a\rfloor = \lfloor\log b\rfloor$ we will use the balanced samplers from Theorem 3.3, namely $G_{a,b,k,\ell}$ will be $BSamp(\Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log a\rfloor+\ell)}, \varepsilon(a,b,k), \delta(a,b,k))$.
  - Otherwise (assuming W.L.O.G $\lfloor\log a\rfloor > \lfloor\log b\rfloor$) we will use the unbalanced samplers from theorem 3.4, namely $G_{a,b,k,\ell}$ will be:
    $$UBSamp(\Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log a\rfloor+\ell)}, \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log b\rfloor+\ell)}, \varepsilon(a,b,k), \delta(a,b,k))$$
    .

    It's important to note that these samplers come with a condition, it needs to hold that:
    $$\Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log a\rfloor+\ell)} \geq \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log b\rfloor+\ell)}/(\delta(a,b,k)^2).$$
    This indeed holds as:
    $$\begin{aligned}\Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log b\rfloor+\ell)}\cdot(1/\delta(a,b,k)^2) &\leq \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log b\rfloor+\ell)-20\mathrm{A}}\\ &\leq \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log b\rfloor+10\mathrm{A}+\ell)}\\ &\leq \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log a\rfloor+\ell)}\end{aligned}$$

  We denote the degree of $G_{a,b,k,\ell}$ by $Deg_{a,b,k}$ — we omit the ℓ since in both cases the degree depends only on $\epsilon(a,b,k)$ and $\delta(a,b,k)$

- From the construction we have the following recursions:
  $$R_c^{(\ell+1)} = \sum_{(a,b,k)\in\circ_c} R_{\max\{a,b\}}^{(\ell)}\cdot DEG_{a,b,k}^{(\ell)} + \sum_{(a,b,k)\in\bullet_c} R_{\max\{a,b\}}^{(\ell)}.$$
  and
  $$C_c^{(\ell+1)} = \max\left\{\max_{(a,b,k)\in\circ_c}\left\{C_a^{(\ell)}\cdot C_b^{(\ell)}\right\}, \max_{(a,b,k)\in\bullet_c}\left\{2\cdot C_a^{(\ell)}\cdot C_b^{(\ell)}\cdot DEG_{a,b,k}^{(\ell)}\right\}\right\}.$$

  Using this we will analyze now the size of the construction.

7 Size analysis

To bound $R_a^{(\ell)}$ and $C_a^{(\ell)}$ we first need to bound the degrees of the samplers used in this construction, these bounds follow the choices we made for $\varepsilon(a,b,k)$ and $\delta(a,b,k)$.

Lemma 7.1. *There exists a constant c_{deg} such that for every $(a,b,k) \in \bigcup_{c=0}^{\mathrm{A}} N_c$ it holds that:*

$$\frac{\log(1/\delta(a,b,k))}{\varepsilon(a,b,k)} \leq (\frac{1}{\varepsilon(a,b,k)})^{c_{deg}}$$

Proof: Firstly note that for every $a,b,k \in [\mathrm{A}]$ we have:

$$1/\delta(a,b,k) < \Delta^{-10\mathrm{A}}$$

and on the other hand:

$$1/\varepsilon(a,b,k) > \Delta^{-9}$$

hence we get:

$$\begin{aligned} \log(1/\delta(a,b,k)) &\leq \log(\Delta^{-10\mathrm{A}}) \\ &\leq 10\mathrm{A} \cdot \log(1/\Delta) \\ &\leq 1/\Delta \leq 1/\varepsilon(a,b,k) \end{aligned}$$

□

Claim 7.2. *There exists a constant $c_{deg} \geq 1$ such that for all $(i,j,k) \in \bigcup_{c=0}^{\mathrm{A}} N_c$:*

$$DEG(i,j,k) = \begin{cases} \Delta^{-9 \cdot c_{deg}} & \text{If } \min(a,b) = k = 0 \\ \Delta^{-9 \cdot (\min(a,b)+k) \cdot c_{deg}} & \text{Otherwise.} \end{cases}$$

Proof: First note that we chose:

$$\varepsilon(a,b,k) = \begin{cases} \Delta^{9} & \text{If } \min(a,b) = k = 0 \\ \Delta^{9 \cdot (\min(a,b)+k)} & \text{Otherwise.} \end{cases}$$

By theorem 3.3 we know that the balanced samplers we used throughout the construction have degree $O(\varepsilon(a,b,k)^{-3})$ since we chose in those cases $\delta(a,b,k) = \varepsilon(a,b,k)$, hence there exists such a constant c_{deg} for the degrees of the balanced samplers.
For the unbalanced samplers we know by theorem 3.4 that their degree is bounded by $(\frac{\log(1/\delta(a,b,k))}{\varepsilon(a,b,k)})^c$ for some constant $c \geq 1$, hence by lemma 7.1 we get that the claim is true. □

Now we move to bounding $R_a^{(\ell)}$ and $C_a^{(\ell)}$.

Claim 7.3. *For every $\ell \geq 0$ we have $C_0^{(\ell)} = 1$.*

Proof: We prove the claim by induction on ℓ. The case $\ell = 0$ is clear. Assume for ℓ we prove for $\ell + 1$. For accuracy $a = 0$ the only suitable $i,j,r \in N_0$ are those such that $i = j = r = 0$. Clearly, $(0,0,0) \in \circ_0$. Hence, by induction, $C_0^{(\ell+1)} = C_0^{(\ell)} \cdot C_0^{(\ell)} = 1$. ■

Claim 7.4. *There exist constants $c, c_1 \geq 20c_{deg}$, such that for every $\ell \geq 0$ and $0 \leq a < \mathrm{A}$ we have*

$$\begin{aligned} C_a^{(\ell)} &\leq (\Delta)^{-c_1 \cdot a \cdot \log(a)}, \text{ and,} \\ R_a^{(\ell)} &\leq \Sigma \cdot \Delta^{-c \cdot (\frac{\log(1/\delta_{min})}{\log(1/\Delta)} \cdot \lfloor \log a \rfloor + \ell)} = \Sigma \cdot \Delta^{-c \cdot (10\mathrm{A} \cdot \lfloor \log a \rfloor + \ell)}. \end{aligned}$$

Proof: We prove by induction on ℓ. Assume for ℓ, let us prove for $\ell + 1$.

For $C_a^{(\ell+1)}$ let us divide for four cases:

1. First we consider the case where $(i, j, k) \in \circ_a$, i.e., $k = \min\{i, j\} = 0$. W.l.o.g. assume $i = 0$. Then $C_i^{(\ell)} \cdot C_j^{(\ell)} \leq C_j^{(\ell)} \leq \Delta^{-c_1 \cdot j \cdot \log(j)}$.

2. Next we consider the case where $(i, j, k) \in \bullet_a$ and $k = 0$ (this implies $i, j > 0$), W.L.O.G assume $j \geq i$ and we get:

$$C_i^{(\ell)} \cdot C_j^{(\ell)} \cdot DEG_{i,j,k} \leq \Delta^{-c_1 \cdot i \cdot \log(i)} \cdot \Delta^{-c_1 \cdot j \cdot \log(j)} \cdot \Delta^{-9c_{deg} \cdot i} \leq \Delta^{-c_1 \cdot a \cdot \log(a)}$$

   Note that for the last inequality to hold it suffices to show:

$$c_1 \cdot i \cdot \log(i) + c_1 \cdot j \cdot \log(j) + 9c_{deg} \cdot i \leq c_1 \cdot (i + j) \cdot \log(i + j)$$

   Re-arranging we get:

$$9c_{deg} \cdot i \leq c_1 \cdot (i \cdot \log(1 + \frac{j}{i}) + j \cdot \log(1 + \frac{i}{j}))$$

   which holds for $c_1 \geq 20c_{deg}$

3. Next we consider the case where $(i, j, k) \in \bullet_a$ and $i = j = 0$ (this implies $k > 0$), assuming $j \geq i$ we get:

$$2 \cdot C_i^{(\ell)} \cdot C_j^{(\ell)} \cdot DEG_{i,j,k} \leq 2 \cdot \Delta^{-9c_{deg} \cdot k} \leq \Delta^{-c_1 \cdot a \cdot \log(a)}.$$

   Note that for the last inequality to hold it suffices to show:

$$9c_{deg} \cdot k + 1 \leq c_1 \cdot k \log(k)$$

4. Next we consider the case where $(i, j, k) \in \bullet_a$ and only $\min(i, j) = 0$. W.l.o.g. assume $i = 0$. Then we get

$$2 \cdot C_i^{(\ell)} \cdot C_j^{(\ell)} \cdot DEG_{i,j,k} \leq 2 \cdot \Delta^{-c_1 \cdot j \cdot \log(j)} \cdot \Delta^{-9c_{deg} \cdot k} \leq \Delta^{-c_1 \cdot a \cdot \log(a)}.$$

   Note that for the last inequality to hold it suffices to show:

$$c_1 \cdot j \cdot \log(j) + 9c_{deg} \cdot k + 1 \leq c_1 \cdot (j + k) \log(j + k)$$

   Re-arranging we get:

$$9c_{deg} \cdot k + 1 \leq c_1 \cdot (j \cdot \log(1 + \frac{k}{j}) + k \cdot \log(j + k))$$

   which holds for $c_1 \geq 20c_{deg}$

5. Finally, consider the case where $(i,j,k) \in \bullet_a$ and all i, j and k are not equal to 0. W.l.o.g. assume $j \geqslant i$, hence we get:

$$\begin{aligned} 2 \cdot C_j^{(\ell)} \cdot C_i^{(\ell)} \cdot DEG_{i,j,k} &\leq 2 \cdot \Delta^{-c_1 \cdot j \cdot \log(j)} \cdot \Delta^{-c_1 \cdot i \cdot \log(i)} \cdot \Delta^{-9c_{deg} \cdot (i+k)} \\ &\leq \Delta^{-c_1 \cdot a \cdot \log(a)}. \end{aligned}$$

Note that for the last inequality to hold it suffices to show:

$$c_1 \cdot j \cdot \log(j) + c_1 \cdot i \cdot \log(i) + 9c_{deg} \cdot (i+k) + 1 \leq c_1 \cdot (i+j+k) \log(i+j+k)$$

Re-arranging we get:

$$9c_{deg} \cdot (i+k) + 1 \leq c_1 \cdot (i \cdot \log(1 + \frac{j+k}{i}) + j \cdot \log(1 + \frac{i+k}{j}) + k \cdot \log(i+j+k))$$

which holds for $c_1 \geq 20c_{deg}$

Hence in conclusion we get:

$$C_a^{(\ell+1)} \leq \Delta^{-f(a)}$$

For $R_a^{(\ell)}$ we have that:

$$\begin{aligned} R_a^{(\ell+1)} &= 2 \sum_{(i,j,k) \in \bullet_a, i \geq j} R_i^{(\ell)} + 2 \sum_{(i,j,k) \in \circ_a, i \geq j} R_i^{(\ell)} \cdot DEG_{i,j,k} \\ &\leq \mathrm{A}^3 \cdot \Sigma \cdot \Delta^{-c \cdot (\frac{\log(1/\delta_{min})}{\log(1/\Delta)} \cdot \lfloor \log a \rfloor + \ell)} + 2 \cdot \Sigma \cdot \Delta^{-c \cdot (\frac{\log(1/\delta_{min})}{\log(1/\Delta)} \cdot \lfloor \log a \rfloor + \ell)} \cdot \Delta^{-9c_{deg} \cdot a/2} \\ &\leq 3 \cdot \Sigma \cdot \Delta^{-c \cdot (\frac{\log(1/\delta_{min})}{\log(1/\Delta)} \cdot \lfloor \log a \rfloor + \ell)} \cdot \Delta^{-9c_{deg} \cdot a/2} \\ &\leq \Sigma \cdot \Delta^{-c \cdot (\frac{\log(1/\delta_{min})}{\log(1/\Delta)} \cdot \lfloor \log a \rfloor + \ell + 1)}, \end{aligned}$$

because $\Delta^{-9c_{deg}} \geq \mathrm{A}^3$, and $c \geq 20c_{deg}$. ■

8 Error Analysis

8.1 Max-norm

Claim 8.1. *For every ℓ, $mn(\mathbf{P}_0^{(\ell)}) \leq 1$.*

Proof: We prove by induction. For $\ell = 0$, we have $mn(\mathbf{P}_0^{(0)}) \leq 1$. Assume for ℓ, and let us prove for $\ell + 1$. We have $\mathbf{P}_0^{(\ell+1)} = \mathbf{P}_0^{(\ell)} \circ_{\varepsilon,\delta} \mathbf{P}_0^{(\ell)}$ and by what we have proved before $mn(\mathbf{P}_0^{(\ell+1)}) \leq mn(\mathbf{P}_0^{(\ell)}) \cdot mn(\mathbf{P}_0^{(\ell)}) \leq 1$. ■

Claim 8.2. *For every $0 \leq c < \mathrm{A}$, $mn(\mathbf{P}_c^{(\ell)}) \leq 2^c$*

Proof: For $\ell = 0$, we have $mn(\mathbf{P}_0^{(0)}) \leq 1$ and $mn(\mathbf{P}_c^{(0)}) = 0$ for $c > 0$. Assume for ℓ and let us prove for $\ell+1$. Fix $c > 0$. We examine the pieces of $mn(\mathbf{P}_c^{(\ell+1)})$ one by one. The pieces are indexed by a, b, k such that $s(a,b,k) = c$.

- First, if $k = \min\{a, b\} = 0$. W.l.o.g. assume that $a = c$ and $b = 0$. Then,

$$mn(\frac{\alpha_a \cdot \beta_b}{m_c} \cdot \mathbf{P}_c^{(\ell)} \circ_{\varepsilon,\delta} \mathbf{P}_0^{(\ell)}) \leq mn(\mathbf{P}_c^{(\ell)}) \cdot mn(\mathbf{P}_0^{(\ell)}) \leq 2^c.$$

- Next we consider the case where $k = 0$ and $a + b = c$ and $a, b \geq 1$. Then:

$$\begin{aligned} mn(\frac{\alpha_a \cdot \beta_b}{m_c} \cdot \mathbf{P}_a^{(\ell)} \bullet_{\varepsilon,\delta} \mathbf{P}_b^{(\ell)}) &\leq mn(\mathbf{P}_a^{(\ell)}) \cdot mn(\mathbf{P}_b^{(\ell)}) \\ &\leq 2^a \cdot 2^b = 2^c \end{aligned}$$

- Next we consider $k \geq 1$, note that here $c \geqslant a + b + 1$:

$$\begin{aligned} mn(\frac{\alpha_a \cdot \beta_b}{m_c}[\mathbf{P}_a^{(\ell)} \bullet_k \mathbf{P}_b^{(\ell)} - \mathbf{P}_a^{(\ell)} \bullet_{k-1} \mathbf{P}_b^{(\ell)}]) &\leq 2 \cdot mn(\mathbf{P}_a^{(\ell)}) \cdot mn(\mathbf{P}_b^{(\ell)}) \\ &\leq 2^{a+b+1} \leq 2^c \end{aligned}$$

■

8.2 Average norm

We now want to bound $an(\mathbf{P}_c^{(\ell)})$. Throughout this subsection we will show bounds on the ϵ, δ parameters of the samplers, these bounds are the reason (or justification) for the choices we made for these parameters in section 6.

Claim 8.3. *For every ℓ, $an(\mathbf{P}_0^{(\ell)}) \leq 1$.*

Proof: We prove this directly, note that $an(\mathbf{A}) \leqslant mn(\mathbf{A})$ for every d.p.d $\mathbf{A}$, hence we get: $an(\mathbf{P}_0^{(\ell)}) \leq mn(\mathbf{P}_0^{(\ell)}) \leq 1$. ■

Claim 8.4. *For every $0 \leq c < \mathrm{A}$, $an(\mathbf{P}_c^{(\ell)}) \leq \Delta^c \cdot 2^{(c-1)/20\mathrm{A}}$*

Proof: For $\ell = 0$, we have $an(\mathbf{P}_0^{(0)}) \leq 1$ and $an(\mathbf{P}_c^{(0)}) = 0$ for $c > 0$. Assume for ℓ and let us prove for $\ell + 1$. Fix $c > 0$. We examine the pieces of $an(\mathbf{P}_c^{(\ell+1)})$ one by one. The pieces are indexed by a, b, k such that $s(a, b, k) = c$.

- First, if $k = \min\{a, b\} = 0$. W.l.o.g. assume that $a = c$ and $b = 0$. Then by lemma 5.14 we get:

$$an(\frac{\alpha_a \cdot \beta_b}{m_c} \cdot \mathbf{P}_c^{(\ell)} \circ_{\varepsilon,\delta} \mathbf{P}_0^{(\ell)}) \leq an(\mathbf{P}_c^{(\ell)}) \leq \Delta^c \cdot 2^{(c-1)/20\mathrm{A}}.$$

- Next we consider the case where $k = 0$ and $a + b = c$ and $a, b \geq 1$, we would like to use lemma 5.13 with $\tau = \frac{1}{20\mathrm{A}}$ and get:

$$\begin{aligned} an(\frac{\alpha_a \cdot \beta_b}{m_c} \mathbf{P}_a^{(\ell)} \bullet_{\varepsilon,\delta} \mathbf{P}_b^{(\ell)}) &\leq an(\mathbf{P}_a^{(\ell)}) \cdot an(\mathbf{P}_b^{(\ell)}) \cdot 2^{\tau} \\ &\leq \Delta^a \cdot 2^{(a-1)/20\mathrm{A}} \cdot \Delta^b \cdot 2^{(b-1)/20\mathrm{A}} \cdot 2^{1/20\mathrm{A}} \\ &= \Delta^c \cdot 2^{(c-2)/20\mathrm{A}} \cdot 2^{1/20\mathrm{A}} = \Delta^c \cdot 2^{(c-1)/20\mathrm{A}} \end{aligned}$$

Now we need to verify we can use this lemma: First we need:

$$\epsilon(a,b,0) \leq \frac{an(\mathbf{P}_b^{(\ell)}) \cdot \tau}{8 \cdot mn(\mathbf{P}_b^{(\ell)})} = \frac{\Delta^b \cdot 2^{(b-1)/20\mathrm{A}}}{160\mathrm{A} \cdot 2^b}$$

Our choice of ε holds (assuming $\frac{1}{160\mathrm{A}} \geq \Delta$) as:

$$\varepsilon(a,b,0) = \Delta^{9b} \leq \frac{\Delta^b}{160\mathrm{A} \cdot 2^b}$$

now for the δ we need:

$$\delta(a,b,0) \leqslant \frac{an(\mathbf{A}) \cdot an(\mathbf{B}) \cdot \tau}{8mn(\mathbf{A}) \cdot mn(\mathbf{B})} = \frac{\Delta^{a+b} \cdot 2^{(a+b-1)/20\mathrm{A}}}{160\mathrm{A} \cdot 2^{a+b}}$$

and by our choice of δ, if $\lfloor \log a \rfloor = \lfloor \log b \rfloor$ then:

$$\Delta^{9b} \leq \Delta^{6b+1} \leq \frac{\Delta^{a+b} \cdot 2^{(a+b-1)/20\mathrm{A}}}{160\mathrm{A} \cdot 2^{a+b}}$$

Otherwise:

$$\Delta^{10\mathrm{A}} \leq \Delta^{2\mathrm{A}+1} \leq \frac{\Delta^{a+b} \cdot 2^{(a+b-1)/20\mathrm{A}}}{160\mathrm{A} \cdot 2^{a+b}}$$

- Next we consider the cases where $k \geq 1$, by using lemma 5.15 we know that:

$$an(\mathbf{P}_a^{(\ell)} \bullet_{G_k - G_{k-1}} \mathbf{P}_b^{(\ell)}) \leq 4w^2 \cdot mn(\mathbf{P}_b^{(\ell)})(\epsilon_{a,b,k-1} \cdot an(\mathbf{P}_a^{(\ell)}) + mn(\mathbf{P}_a^{(\ell)}) \cdot \delta_{a,b,k-1})$$

Assuming $a \geq b$ we'll use this lemma with $\tau = \frac{c-a}{20\mathrm{A}}$, note that we have (assuming $\Delta \leq \frac{1}{4w^2}$):

$$\epsilon(a,b,k-1) = \Delta^{9(b+k-1)} \leq \frac{\Delta^{c-a} \cdot \tau}{4w^2 \cdot mn(\mathbf{P}_b^{(\ell)})}$$

and for δ assuming $\lfloor \log a \rfloor = \lfloor \log b \rfloor$:

$$\delta(a,b,k-1) = \Delta^{9(b+k-1)} \leq \Delta^{6b+k+2} \leq \frac{\Delta^c \cdot \tau}{4w^2 \cdot mn(\mathbf{P}_a^{(\ell)}) \cdot mn(\mathbf{P}_b^{(\ell)})}$$

now assuming otherwise:

$$\delta(a,b,k-1) = \Delta^{10\mathrm{A}} \leq \Delta^{2\mathrm{A}+2} \leq \frac{\Delta^c \cdot \tau}{4w^2 \cdot mn(\mathbf{P}_a^{(\ell)}) \cdot mn(\mathbf{P}_b^{(\ell)})}$$

Thus we get:

$$an(\mathbf{P}_a^{(\ell)} \bullet_{G_k - G_{k-1}} \mathbf{B}) \quad \leq \quad \Delta^c \cdot 2^{(a-1)/20\mathrm{A}} \cdot 2^{(c-a)/20\mathrm{A}} = \Delta^c \cdot 2^{(c-1)/20\mathrm{A}}$$

■

8.3 Cube Coefficients

we'll denote by $\gamma_c^{(\ell)}$ as the coefficient of the c'th d.p.d in $\mathcal{P}^{(\ell)}$

Claim 8.5.

$$\gamma_c^{(\ell)} \leq (c+1)^{3 \cdot c \cdot \ell}$$

Proof:

$$\begin{aligned}
\gamma_c^{\ell+1} &\leq |N_c| \cdot \max(\gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} | a+b \leq c) \\
&\leq (c+1)^3 \cdot \max((a+1)^{3 \cdot a \cdot \ell} \cdot (b+1)^{3 \cdot b \cdot \ell} | a+b \leq c) \\
&\leq (c+1)^3 \cdot (c+1)^{3 \cdot c \cdot \ell} \\
&\leq (c+1)^{3 \cdot c \cdot (\ell+1)}
\end{aligned}$$

■

8.4 Putting it together

Let us denote:

$$\begin{aligned}
an_a^{(\ell)} &= an(\mathbf{P}_a^{(\ell)}) \\
mn_a^{(\ell)} &= mn(\mathbf{P}_a^{(\ell)}).
\end{aligned}$$

Then,

Lemma 8.6.

$$\left\| val(\mathcal{P}^\ell \cdot_{\varepsilon,\delta} \mathcal{P}^\ell) - val(\mathcal{P}^\ell) \cdot val(\mathcal{P}^\ell) \right\| \leq 12 \cdot w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4}$$

Proof: by lemma 5.17 we know that:

$$\begin{aligned}
\left\| val(\mathcal{P}^\ell \cdot_{\varepsilon,\delta} \mathcal{P}^\ell) - val(\mathcal{P}^\ell) \cdot val(\mathcal{P}^\ell) \right\| \leq & \sum_{a,b:a+b \leq A} \gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot w^3 \cdot 2\delta(a,b,K_{a,b}) \cdot mn_a^{(\ell)} \cdot mn_b^{(\ell)} \\
+ & \sum_{a,b:a+b \leq A} \gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot w \cdot \varepsilon(a,b,K_{a,b}) \cdot mn_b^{(\ell)} \cdot an_a^{(\ell)} \\
+ & \sum_{a,b:a,b \leq \mathrm{A}, a+b > \mathrm{A}} \gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot an_a^{(\ell)} \cdot an_b^{(\ell)}.
\end{aligned}$$

we'll go through these sums one by one, starting with the third one:

$$\begin{aligned}
\sum_{a,b:a,b \leq \mathrm{A}, a+b > \mathrm{A}} \gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot an_a^{(\ell)} \cdot an_b^{(\ell)} &\leq \sum_{a,b:a,b \leq \mathrm{A}, a+b > \mathrm{A}} (a+1)^{3a\ell} \cdot (b+1)^{3b\ell} \cdot an_a^{(\ell)} \cdot an_b^{(\ell)} \\
&\leq \sum_{a,b:a,b \leq \mathrm{A}, a+b > \mathrm{A}} (\mathrm{A}+1)^{6\mathrm{A}\ell} \cdot 4\Delta^a \cdot \Delta^b \\
&\leq 4 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}}
\end{aligned}$$

now the second sum in the case $b \neq 0$:

$$\begin{aligned}\gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot w \cdot \varepsilon(a,b,K_{a,b}) \cdot mn_b^{(\ell)} \cdot an_a^{(\ell)} &\leq (a+1)^{3a\ell} \cdot (b+1)^{3b\ell} \cdot w \cdot \varepsilon(a,b,K_{a,b}) \cdot 2^b \cdot 2\Delta^a \\ &\leq (\mathrm{A}+1)^{6\mathrm{A}\ell} \cdot w \cdot 2\Delta^{\mathrm{A}/2}\end{aligned}$$

where we used $\varepsilon(a,b,K_{a,b}) \cdot 2^b \cdot \Delta^a \leq \Delta^{\mathrm{A}/2}$, which holds as by our choice of $\epsilon(a,b,k)$ we get:

$$\epsilon(a,b,K_{a,b}) \leq \Delta^{9(\min(a,b)+K_{a,b})} = \Delta^{9(\mathrm{A}-\max(a,b))} \leq \Delta^{\mathrm{A}/2-\max(a,b)+\min(a,b)}$$

now the second sum in the case $b = 0$:

$$\begin{aligned}\gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot w \cdot \varepsilon(a,b,K_{a,b}) \cdot mn_b^{(\ell)} \cdot an_a^{(\ell)} &\leq (a+1)^{3a\ell} \cdot (b+1)^{3b\ell} \cdot w \cdot \varepsilon(a,0,K_{a,0}) \cdot 2\Delta^a \\ &\leq (\mathrm{A}+1)^{6\mathrm{A}\ell} \cdot w \cdot 2\Delta^{\mathrm{A}/2}\end{aligned}$$

where we used $\varepsilon(a,0,K_{a,0}) \cdot \Delta^a \leq \Delta^{\mathrm{A}/2}$, which again holds by our choice of $\epsilon(a,b,k)$:

$$\varepsilon(a,0,K_{a,0}) = \Delta^{9K_{a,b}} = \Delta^{9(\mathrm{A}-a)} \leq \Delta^{\mathrm{A}/2-a}$$

and for the first sum:

$$\begin{aligned}\sum_{a,b:a+b\leq A} \gamma_a^{(\ell)} \cdot \gamma_b^{(\ell)} \cdot w^3 \cdot 2\delta(a,b,K_{a,b}) \cdot mn_a^{(\ell)} \cdot mn_b^{(\ell)} &\leq \sum_{a,b:a+b\leq A} (\mathrm{A}+1)^{6\mathrm{A}\ell} \cdot w^3 \cdot 2 \cdot \delta(K_{a,b}) \cdot 2^{a+b} \\ &\leq 2 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot w^3 \delta(K_{a,b}) \cdot 2^{\mathrm{A}} \\ &\leq 2w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4}\end{aligned}$$

where in the last inequality we used: $\delta(K_{a,b}) \cdot 2^{\mathrm{A}} \leq \Delta^{\mathrm{A}/4}$, which holds by our choice of δ, firstly when $\lfloor \log a \rfloor = \lfloor \log b \rfloor$:

$$\delta(a,b,K_{a,b}) = \Delta^{9(b+K_{a,b})} = \Delta^{9(\mathrm{A}-a)} \leqslant \Delta^{9\cdot\frac{\mathrm{A}}{4}} \leq \Delta^{\mathrm{A}+\mathrm{A}/4}$$

and otherwise: $\delta(a,b,K_{a,b}) = \Delta^{10\mathrm{A}} \leq \Delta^{\mathrm{A}+\mathrm{A}/4}$
hence in conclusion we get that:

$$\left\| val(\mathcal{P}^\ell \cdot_{\varepsilon,\delta} \mathcal{P}^\ell) - val(\mathcal{P}^\ell) \cdot val(\mathcal{P}^\ell) \right\| \leq 12 \cdot w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4}$$

■

Lemma 8.7. *Let $\mathcal{R}^{(\ell)}$ be the d.p.d with Σ^{2^ℓ} rows, where row $\sigma \in \Sigma^{2^\ell}$ has a single column $(1,\sigma)$. and we define $\epsilon(\ell) = \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\|$, Then:*

$$\epsilon(\ell) = \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| \leq w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4}$$

Proof: We prove by induction. It is clear that for the base case $\ell = 0$ there is an equality. Assume for ℓ and let us prove for $\ell+1$. We have $\mathcal{P}^{(\ell+1)} = \mathcal{P}^{(\ell)} \cdot_{\varepsilon,\delta} \mathcal{P}^{(\ell)}$. Also, $\mathcal{R}^{(\ell+1)} = \mathcal{R}^{(\ell)} \otimes \mathcal{R}^{(\ell)}$ and $val(\mathcal{R}^{(\ell+1)}) = val(\mathcal{R}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)})$. Hence,

$$\begin{aligned}\left\| val(\mathcal{P}^{(\ell+1)}) - val(\mathcal{R}^{(\ell+1)}) \right\| &= \left\| val(\mathcal{P}^{(\ell)} \cdot_{\varepsilon,\delta} \mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)}) \right\| \\ &\leq \left\| val(\mathcal{P}^{(\ell)} \cdot_{\varepsilon,\delta} \mathcal{P}^{(\ell)}) - val(\mathcal{P}^{(\ell)}) \cdot val(\mathcal{P}^{(\ell)}) \right\| \\ &+ \left\| val(\mathcal{P}^{(\ell)}) \cdot val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)}) \right\|.\end{aligned}$$

We bound the first term using Lemma 8.6
and for the second:

$$\begin{aligned}\left\| val(\mathcal{P}^{(\ell)}) \cdot val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)}) \right\| &\leq \left\| val(\mathcal{P}^{(\ell)}) \cdot val(\mathcal{P}^{(\ell)}) - val(\mathcal{P}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)}) \right\| \\ &+ \left\| val(\mathcal{P}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \cdot val(\mathcal{R}^{(\ell)}) \right\| \\ &\leq \left\| val(\mathcal{P}^{(\ell)}) \right\| \cdot \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| \\ &+ \left\| val(\mathcal{R}^{(\ell)}) \right\| \cdot \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\|\end{aligned}$$

now for the first summand here:

$$\begin{aligned}\left\| val(\mathcal{P}^{(\ell)}) \right\| \cdot \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| &\leq \left\| val(\mathcal{P}^{(\ell)}) + val(\mathcal{R}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| \cdot \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| \\ &\leq (\left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| + \left\| val(\mathcal{R}^{(\ell)}) \right\|) \cdot \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| \\ &\leq (\epsilon(\ell)+1) \cdot \epsilon(\ell)\end{aligned}$$

and for the second summand:

$$\left\| val(\mathcal{R}^{(\ell)}) \right\| \cdot \left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\| \leq \epsilon(\ell)$$

hence in conclusion we have:

$$\begin{aligned}\epsilon(\ell+1) &\leqslant 12 \cdot w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4} + (\epsilon(\ell)+1) \cdot \epsilon(\ell) + \epsilon(\ell) \\ &\leq 12 \cdot w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4} + 3\epsilon(\ell) \\ &\leq 15 \cdot w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4} \\ &\leq w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}(\ell+1)} \cdot \Delta^{\mathrm{A}/4}\end{aligned}$$

which together concludes the proof.
note that for the last inequality we need that $15 \leq (\mathrm{A}+1)^{7\mathrm{A}}$, which holds for $\mathrm{A} \geqslant 1$ ■

note that by the previous claim we have:

$$\left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\|_{max} \leq w^3 \cdot (\mathrm{A}+1)^{7\mathrm{A}\ell} \cdot \Delta^{\mathrm{A}/4}$$

Claim 8.8. *Denote* $B = 8\log(w^3/\epsilon)$*, if we'll take* $\log(1/\Delta) = 56\log(n) \cdot \log(B) + 8\log(w)$ *and* $\mathrm{A} = \frac{B}{\log(1/\Delta)}$ *then:*

$$\left\| val(\mathcal{P}^{(\log n)}) - val(\mathcal{R}^{(\log n)}) \right\|_{max} \leq \epsilon$$

Proof: First note that:

$$(\mathrm{A}+1)^{7\mathrm{A}\log(n)} \quad \leq \quad \Delta^{-\mathrm{A}/8}$$

By the previous claim:

$$\left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\|_{max} \quad \leq \quad w^3 \cdot \Delta^{\mathrm{A}/8}$$

and because of our choices for A and g we have: $\Delta^{\mathrm{A}/8} = \frac{\epsilon}{w^3}$, so we'll get:

$$\left\| val(\mathcal{P}^{(\ell)}) - val(\mathcal{R}^{(\ell)}) \right\|_{max} \quad \leq \quad \epsilon$$

■

Claim 8.9. *The seed-length of the resulting pseudo-distribution $\mathcal{P}^{\log(n)}$ is:* $\tilde{O}(\log^2(n)+\log(w/\epsilon)+\log(w)\cdot\log(n))$

Proof: The support size is bounded by:

$$\begin{aligned}\sum_{a=0}^{\mathrm{A}} \mathcal{C}_a^{(\ell)} \cdot \mathcal{R}_a^{(\ell)} \quad &\leq \quad \sum_{a=0}^{\mathrm{A}} \Delta^{-f(a)} \cdot \Sigma \cdot \Delta^{-c\cdot(\frac{\log(1/\delta_{min})}{\log(1/\Delta)}\cdot\lfloor\log a\rfloor+\log(n))} \\ &\leq \quad \mathrm{A} \cdot \Delta^{-c_1\cdot\mathrm{A}\log(\mathrm{A})} \cdot \Sigma \cdot \Delta^{-c\cdot(10\mathrm{A}\cdot\lfloor\log \mathrm{A}\rfloor+\log(n))}\end{aligned}$$

hence the seed length is bounded by:

$$\begin{aligned}& O(\mathrm{A}\log(1/\Delta)\log(\mathrm{A}) + \mathrm{A}\cdot\log(\mathrm{A})\cdot\log(1/\Delta) + \log(n)\cdot\log(1/\Delta)) = \\ = \quad & O(\log(w/\epsilon)\cdot\log(\mathrm{A}) + \log^2(n)\cdot\log\log(w/\epsilon) + \log(w)\log(n)) \\ = \quad & \tilde{O}(\log^2(n) + \log(w/\epsilon) + \log(w)\cdot\log(n))\end{aligned}$$

■

9 Further Discussion

9.1 Why we have to use $\circ_G$

Throughout this thesis, we defined and used two slightly different notions for derandomized tensoring of d.p.ds, we named these $\circ_G$ and $\bullet_G$.
In almost every instance we've preferred $\bullet_G$ over $\circ_G$ and opted to use $\circ_G$ only in the case where $\min(a,b) = k = 0$, we will now show why we had to use it in this case and could not use only $\bullet_G$ in every case.
Assume we only used $\bullet_G$ throughout the construction, note that this only affects the size-analysis section of the construction.
We'll analyze first $C_a^{(\ell)}$ in this new setting, note that there is a change only in the case where $(i,j,k) \in \circ_a$, and we get that:

$$\begin{aligned}C_a^{\ell+1} \quad &= \quad C_a^{(\ell)} \cdot DEG_{i,j,k} \\ &\leq \quad C_a^{(\ell)} \cdot \Delta^{-9c_{deg}}\end{aligned}$$

Hence we must have $C_a^{(\ell)} \leq (\Delta)^{-(f(a)+g(\ell))}$
On the other hand, in the case where $(i,j,k) \in \bullet_a$ and $k=0$ we would've liked to get:

$$\begin{aligned} C_a^{\ell+1} &= C_i^{(\ell)} \cdot C_j^{(\ell)} \cdot DEG_{i,j,k} \\ &\leq \Delta^{-(f(i)+f(j)+2\cdot g(\ell))} \cdot \Delta^{-9c_{deg}\cdot i} \leq \Delta^{-(f(a)+g(\ell+1))} \end{aligned}$$

Note that for the last inequality to hold, we have to show that:

$$f(i)+f(j)+9c_{deg}\cdot i + 2g(\ell) \leq f(a)+g(\ell+1)$$

which implies that g has to grow rapidly, but note that if g is even quadratic then the seed length will be:

$$\Omega(\log(1/\Delta)\cdot g(\log(n))) = \Omega(log^3(n))$$

So essentially, we're using $\circ_g$ to make sure $C_a^{(\ell)}$ won't be dependant on ℓ which will majorly increase the size of the construction.

9.2 Why the seed-length of the construction can't be lower then $\Omega(log^2(n))$

In Lemma 8.7, we showed that our construction yields an $(n,w,\varepsilon)-$ pseudo random pseudo-distribution with $\varepsilon = w^3\cdot(\mathrm{A}+1)^{7\mathrm{A}\log(n)}\cdot\Delta^{\mathrm{A}/4}$, note that since $\mathrm{A}\geq 1$ we must have $\log(1/\Delta) > \log(n)$ in order to keep $\varepsilon < 1$.
Now as mentioned earlier, in the cases when $\min(i,j)=k=0$ we must use $\circ_G$, this implies:

$$R_a^{(\ell+1)} \geq Rows(\mathbf{P}_a^{(\ell)} \circ_{G_{(a,0,0,\ell)}} \mathbf{P}_0^{(\ell)}) = R_a^{(\ell)}\cdot\Delta^{-9c_{deg}}$$

Hence $R_a^{(\ell)} = \Omega(\Delta^{-O(\ell)})$, it follows that for our $(n,w,\varepsilon)-$ PRPD, the seed length cannot go under $\Omega(\log(n)\cdot\log(1/\Delta)) = \Omega(\log^2(n))$

9.3 Why we have to use $\bullet_G$

We saw that we have to use $\circ_G$ in this construction and using $\bullet_G$ instead worsens the seed-length, this raises the question, could we use only $\circ_G$ throughout the thesis, it turns out that the biggest barrier towards this goal is to bound the average norm of telescopic tensoring.
For that purpose we have to define $\mathbf{A}\circ_{G_1-G_2}\mathbf{B} = \mathbf{A}\circ_{G_1}\mathbf{B} -_{col} \mathbf{A}\circ_{G_2}\mathbf{B}$, thus:

$$\begin{aligned} an(\mathbf{A}\circ_{G_1-G_2}\mathbf{B}) &= \mathbb{E}_{(i,d)\in Rows(\mathbf{A})\times Deg(G_1)}[\left\| val(A_i)\cdot val(B_{\Gamma_1(i,d)}) - val(A_i)\cdot val(B_{\Gamma_2(i,d \mod Deg(G_2))}) \right\|] \\ &\leq \mathbb{E}_{(i,d)\in Rows(\mathbf{A})\times Deg(G_1)}[\| val(A_i)\| \cdot \left\| val(B_{\Gamma_1(i,d)}) - val(B_{\Gamma_2(i,d \mod Deg(G_2))}) \right\|] \\ &= \mathbb{E}_{i\in Rows(\mathbf{A})}[\| val(A_i)\| \cdot \mathbb{E}_{d\in Deg(G_1)}\left\| val(B_{\Gamma_1(i,d)}) - val(B_{\Gamma_2(i,d \mod Deg(G_2))}) \right\|] \end{aligned}$$

Note that here the expectation over the neighbours of i is outside of $\|\cdot\|$ which prevents us from using the properties of the samplers to bound this average norm, compared to the $\mathbf{A}\bullet_{G_1-G_2}\mathbf{B}$ case where:

$$\begin{aligned} an(\mathbf{A}\bullet_{G_1-G_2}\mathbf{B}) &= \mathbb{E}_{i\in Rows(\mathbf{A})}[\left\| val(A_i)\cdot\mathbb{E}_{d\in Deg(G_1)}val(B_{\Gamma_1(i,d)}) - val(A_i)\cdot\mathbb{E}_{d\in Deg(G_2)}val(B_{\Gamma_2(i,d)}) \right\|] \\ &\leq \mathbb{E}_{i\in Rows(\mathbf{A})}[\| val(A_i)\| \cdot \left\| \mathbb{E}_{d\in Deg(G_1)}val(B_{\Gamma_1(i,d)}) - \mathbb{E}_{d\in Deg(G_2)}val(B_{\Gamma_2(i,d)}) \right\|] \\ &\leq \mathbb{E}_{i\in Rows(\mathbf{A})}[\| val(A_i)\| \cdot (\left\| \mathbb{E}_{d\in Deg(G_1)}val(B_{\Gamma_1(i,d)}) - \mathbb{E}_{i'\in Rows(\mathbf{B})}val(B_{i'}) \right\| \\ &+ \left\| \mathbb{E}_{i'\in Rows(\mathbf{B})}val(B_{i'}) - \mathbb{E}_{d\in Deg(G_2)}val(B_{\Gamma_2(i,d)}) \right\|)] \end{aligned}$$

Here in comparison the expectation is inside the $\|\cdot\|$ which allows us to bound this average norm as shown in lemma 5.15.

References

[BCG17] Mark Braverman, Gil Cohen, and Sumegha Garg. Hitting sets with near-optimal error for read-once branching programs. 2017.

[Gol11] Oded Goldreich. A sample of samplers: A computational perspective on sampling. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 302–332. Springer, 2011.

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

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

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

[SZ99] Michael E. Saks and Shiyu Zhou. $BP_HSPACE(S) \subseteq DSPACE(S^{3/2})$. *J. Comput. Syst. Sci.*, 58(2):376–403, 1999.

[Zuc06] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 681–690. ACM, 2006.

תקציר

המאמר [BCG17] הציג אובייקט חדש שניתן באמצעותו לבצע דה-ראנדומיזציה לאלגוריתמים אקראיים, **פסאודו-התפלגות פסאודו-אקראית**.
זוהי הכללה של מחוללים-פסאודו אקראיים כך שניתן להביא משקלים שליליים או לא חסומים ולא בהכרח לבנות התפלגות.
המטרה של התזה הזאת היא להציג את הבנייה של [BCG17] בגרסא פשוטה יותר ולהציע ניתוח מעט שונה לאותה בנייה.
במקרה הכללי המחולל הפסאודו-אקראי הטוב ביותר הידוע היה עם גודל גרעין: $O(log^2(n) + \log(n) \cdot \log(1/\varepsilon))$, לעומת זאת בבנייה הזאת מקבלים פסאודו-התפלגות פסאודו אקראית עם גודל גרעין: $\tilde{O}(\log^2(n) + \log(1/\varepsilon))$, לכן כאשר $\log(1/\varepsilon) > \log(n)$ הבנייה הזאת בהחלט טובה יותר.

אוניברסיטת תל אביב
TEL AVIV UNIVERSITY

הפקולטה למדעים מדויקים ע"ש ריימונד ובברלי סאקלר
בית הספר למדעי המחשב ע"ש בלבטניק

הפשטת בניית ה-BCG של פסאודו-התפלגות פסאודו-אקראית עבור חישוב מוגבל זיכרון

חיבור זה מוגש כחלק מהדרישות לקבל התואר
"מוסמך אוניברסיטה" באוניברסיטת תל אביב
ביה"ס למדעי המחשב
ע"י
דניאל קוזלוב

עבודה זו בוצעה בהנחיית
פרופסור אמנון תא-שמע

פברואר 2020