On the Multidimensional Random Subset Sum Problem

Luca Becchetti, Arthur Carvalho Walraven da Cuhna, Andrea Clementi, Francesco d'Amore, Hicham Lesfari, Emanuele Natale, Luca Trevisan

On the Multidimensional Random Subset Sum Problem

Luca Becchetti[1], Arthur C. W. da Cunha[2], Andrea Clementi[3], Francesco d'Amore[2], Hicham Lesfari[2], Emanuele Natale[2], and Luca Trevisan[4]

[1]University of Rome La Sapienza
[2]INRIA, CNRS, Université Côte d'Azur
[3]University of Rome Tor Vergata
[4]Bocconi University

Abstract

In the Random Subset Sum Problem, given n i.i.d. random variables $X_1, ..., X_n$, we wish to approximate any point $z \in [-1, 1]$ as the sum of a suitable subset $X_{i_1(z)}, ..., X_{i_s(z)}$ of them, up to error ε. Despite its simple statement, this problem is of fundamental interest to both theoretical computer science and statistical mechanics. More recently, it gained renewed attention for its implications in the theory of Artificial Neural Networks. An obvious multidimensional generalisation of the problem is to consider n i.i.d. d-dimensional random vectors, with the objective of approximating every point $\mathbf{z} \in [-1, 1]^d$. Rather surprisingly, after Lueker's 1998 proof that, in the one-dimensional setting, $n = O(\log \frac{1}{\varepsilon})$ samples guarantee the approximation property with high probability, little progress has been made on achieving the above generalisation.

In this work, we prove that, in d dimensions, $n = O(d^3 \log \frac{1}{\varepsilon} \cdot (\log \frac{1}{\varepsilon} + \log d))$ samples suffice for the approximation property to hold with high probability. As an application highlighting the potential interest of this result, we prove that a recently proposed neural network model exhibits *universality*: with high probability, the model can approximate any neural network within a polynomial overhead in the number of parameters.

1 Introduction

In the *Random Subset Sum Problem (RSSP)*, given a target value z, an error parameter $\varepsilon \in \mathbb{R}_{>0}$ and n independent random variables $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$, one is interested in estimating the probability that there exists a subset $S \subseteq [n]$ for which

$$\left| z - \sum_{i \in S} \mathrm{X}_i \right| \leq \varepsilon.$$

Historically, the analysis of this problem was mainly motivated by research on the average case of its deterministic counterpart, the classic Subset Sum Problem, and the equivalent Number Partition Problem. These investigations lead to a number of insightful results, mostly in the 80s and 90s [Lue82, KKLO86, Lue98]. In addition, research on the phase transition of the problem extended to the early 2000s, with interesting applications in statistical physics [MM09, BCP01, BCMP04].

More recently, one of the results on the RSSP has attracted quite some attention. A simplified statement for it would be

Theorem 1 (Lueker, [Lue98]). *Let $\mathrm{X}_1, \ldots, \mathrm{X}_n$ be i.i.d. uniform random variables over $[-1, 1]$, and let $\varepsilon \in (0, 1)$. There exists a universal constant $C > 0$ such that, if $n \geq C \log_2 \frac{1}{\varepsilon}$, then, with high probability, for all $z \in [-1, 1]$ there exists a subset $S_z \subseteq [n]$ for which*

$$\left| z - \sum_{i \in S_z} \mathrm{X}_i \right| \leq \varepsilon.$$

That is, a rather small number (of the order of $\log \frac{1}{\varepsilon}$) of random variables suffices to have a high probability of approximating not only a single target z, but all values in an interval. In fact, this result is asymptotically optimal, since each of the 2^n subsets can cover at most one of two values more than 2ε apart and, hence, we must have $n = \Omega(\log \frac{1}{\varepsilon})$. Also, the original work generalises the result to a wide class of distributions.

Those features allowed Theorem 1 to be quite successful in applications. In the field of Machine Learning, particularly, many recent works, such as [PRN+20, dCNV21, FB21, BLMG22, FTGB22, WDM+21], leverage this result. We discuss those contributions in more detail in Section 2.

In this paper, we investigate a natural multidimensional generalisation of Theorem 1. Mainly, we prove

Theorem 2 (Main Theorem). *Given $\varepsilon \in (0, 1)$ and $d, n \in \mathbb{N}$, consider n independent d-dimensional standard normal random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$. There exists a universal constant $C > 0$ for which, if*

$$n \geq C d^3 \log_2 \frac{1}{\varepsilon} \cdot \left(\log_2 \frac{1}{\varepsilon} + \log_2 d \right),$$

then, with high probability, for all $\boldsymbol{z} \in [-1, 1]^d$ there exists a subset $S_{\boldsymbol{z}} \subseteq [n]$ for which

$$\left\| \boldsymbol{z} - \sum_{i \in S_{\boldsymbol{z}}} \mathbf{X}_i \right\|_\infty \leq \varepsilon.$$

Moreover, the approximations can be achieved with subsets of size $\frac{n}{6\sqrt{d}}$.

We believe many promising applications of the RSSP can become feasible with this extension of Theorem 1 to multiple dimensions. To illustrate this, we consider the *Neural Network Evolution (NNE)* model recently introduced by [GLP+19]. It is natural to wonder whether their model is *universal*, in the sense that, with high probability, it can approximate any dense feed-forward neural network. While applying Theorem 1 to this end would yield exponential bounds on the required overparameterization, in Section 6 we prove the universality of the model within polynomial bounds. To broaden the scope of our result, we additionally provide some useful generalisations in Appendix C. In particular, we extend it to a wide class of distributions, proving an analogous extension to the one [Lue98] given for Theorem 1. Finally, in Appendices D and E we discuss a discretization of our result and potential applications in the context of nondeterministic random walks.

Organisation of the paper. After discussing related works in Section 2, we present a high level overview of the difficulties posed by the problem and of our proof of Theorem 2 (Section 3). We then introduce our notation in Section 4 in preparation for the presentation of our analysis in Section 5. We follow up with an application of our result to the NNE model [GLP+19] and conclude with some notes on the tightness of our analysis in Section 7. Finally, generalisations of our results, further extensions, as well as all omitted proofs can be found in the Appendix.

2 Related work

As remarked in the Introduction, the first studies of the RSSP were mainly motivated by average-case analyses of the classic Subset Sum and Number Partition problems [KKLO86, Lue82, Lue98]. Both can be efficiently solved if the precision of the values considered is sufficiently low relative to the size of the input set. In particular, [Mer98] applies methods from statistical physics to indicate that this is a fundamental property of the problem: the amount of exact solutions for the randomised version exhibits a phase transition when the precision increases relative to the sample size. The work [BCP01] later confirmed formally the existence of a phase transition. [Lue82] shows that the median of the minimum error in the RSSP is exponentially small when the target is near the expected sum of the random variables. This work was followed by [Lue98], which proves Theorem 1. Recently, [dCdG$^+$22] provided a simpler alternative to the original proof. The discrete setting of a variant of RSSP has also been recently studied in [CJRS22] which proves that an integral linear combination (with coefficients in $\{-1, 0, 1\}$) of the sample variables can approximate a range of target values.

In the last few years, Theorem 1 has been very useful in studying the *Strong Lottery Ticket Hypothesis*, which states that Artificial Neural Networks (ANN) with random weights are likely to contain an approximation of any sufficiently smaller ANN as a subnetwork. In particular, such claim poses the deletion of connections (pruning) as a theoretically solid alternative to careful calibration of their weights (training). [PRN$^+$20] uses Theorem 1 to prove the hypothesis under optimal overparameterization for dense ReLU neural networks. [dCNV21] extends this result to convolutional networks and [FTGB22] further extends the latter to the class of equivariant networks. Also, [BLMG22] applies Theorem 1 to construct neural networks that can be adapted to a variety of tasks with minimal retraining.

3 Overview of our analysis

3.1 Insights on the difficulty of the problem

In d dimensions, since we need $2^{\Theta(d \log \frac{1}{\varepsilon})}$ hypercubes of radius ε to cover the set $[-1, 1]^d$, we need a sample of $\Omega(d \log \frac{1}{\varepsilon})$ vectors to be able to approximate (up to error ε) every vector in $[-1, 1]^d$.

On the other hand, having $n = \mathcal{O}(d \log \frac{1}{\varepsilon})$ vectors is enough in expectation. To see it, it is sufficient to consider subsets of the sample with $\frac{n}{2}$ vectors. There are $\binom{n}{n/2} \approx 2^{n-o(n)}$ such subsets, each summing to a random vector distributed as $\mathcal{N}(\mathbf{0}, \frac{n}{2} \cdot \boldsymbol{I}_d)$. Thus, given any $\boldsymbol{z} \in [-1, 1]^d$, each of those sums has probability approximately $\varepsilon^d (\frac{n}{2})^{-\frac{d}{2}} = 2^{-d \log \frac{1}{\varepsilon} - \frac{d}{2} \log \frac{n}{2}}$ of being at most ε far from $\boldsymbol{z}$. We can then conclude that the expected number of approximations is $2^{n-o(n)} \cdot 2^{-d \log \frac{1}{\varepsilon} - \frac{d}{2} \log \frac{n}{2}}$, which is still of order $2^{n-o(n)}$ provided that $n \geq Cd \log \frac{1}{\varepsilon}$ for a sufficiently large constant C.

It would thus suffice to prove concentration bounds on the expectation. The technical challenge is handling the stochastic dependency between subsets of the sample, as pairs of those typically intersect, with many random variables thus appearing for both resulting sums. The original proof of Theorem 1 [Lue98] and the simplified one [dCdG$^+$22] address dependencies in similar ways. Both keep track of the fraction of values in $[-1, 1]$ that can be approximated by a sum of a subset of the first i random variables, $\mathrm{X}_1, \ldots, \mathrm{X}_i$. Their core goal is to bound the proportional increase in this fraction when an additional random variable X_{i+1} is considered. As it turns out, the *conditional expectation* of this increment can be bounded by a constant factor, regardless of the values of $\mathrm{X}_1, \ldots, \mathrm{X}_i$. Unfortunately, naively extending those ideas to d dimensions leads to an estimation of this increment that is exponentially small in d. It is not clear to the authors how to make the

estimation depend polynomially on d without leveraging some knowledge of the actual values of $\mathrm{X}_1, \ldots, \mathrm{X}_i$. In fact, even which kind of assumption on the previous samples could work in this sense is not totally clear.

As for other classical concentration techniques that might appear suitable at first, we remark our failed attempts to leverage an average bounded differences argument [War16]. Specifically, we could not identify any natural function related to the fraction of values that can be approximated, which was also Lipschitz relative to the sample vectors. Moreover, both Janson's variant of Chernoff bound [Jan04] and a recent refinement of it [WRG17] seem to capture the stochastic dependence of the subset sums too loosely for our needs.

3.2 Our approach

Our strategy to overcome the difficulties highlighted in the previous subsection consists in a second-moment approach.

Unlike the proofs for the single dimensional case, our argument, at first, analyses the probability of approximating a single target value $\boldsymbol{z} \in [-1, 1]^d$. To this end, consider a sample of n independent random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and a family $\mathcal{C}$ of subsets of the sample. Let Y be the number of subsets in $\mathcal{C}$ whose sum approximates $\boldsymbol{z}$ up to error ε.

For a single subset, it is not hard to estimate the probability with which a subset-sum $\sum_{i \in S} \mathbf{X}_i$ lies close to $\boldsymbol{z}$. This allows us to easily obtain good bounds on $\mathbb{E}[\mathrm{Y}]$.

We, then, proceed to estimate the variance of Y, circumventing the obstacles mentioned in the previous section by restricting the analysis to families of subsets with sufficiently small pairwise intersections. While this restriction limits the maximum amount of subsets that are available, a standard probabilistic argument allows us to prove the existence of large families of subsets with the desired property, ensuring that $\mathbb{E}[\mathrm{Y}]$ can be large enough for our purposes.

For each pair of subsets, S and T, we leverage the hypothesis on the size of intersections to consider partitions $S = S_\mathrm{A} \cup S_\mathrm{B}$ and $T = T_\mathrm{C} \cup T_\mathrm{B}$, with S_A and T_C being large, stochastically independent parts, and the smaller parts S_B and T_B containing $S \cap T$. The bulk of our analysis then consists in deriving careful bounds on their reciprocal dependencies and consequent contributions to the second moment of Y.

The resulting estimate allows us to apply Chebyshev's inequality to Y, obtaining a constant lower bound on $\Pr[\mathrm{Y} \geq 1]$. That is, with conclude that with at least some constant probability at least one of the subsets yields a suitable approximation of $\boldsymbol{z}$. Finally, we employ a probability-amplification argument in order to apply a union bound over all possible target values in $[-1, 1]^d$.

4 Preliminaries

Notation Throughout the text we identify the different types of objects by writing their symbols in different styles. This applies to scalars (e.g. x), real random variables (e.g. X), vectors (e.g. $\boldsymbol{x}$), random vectors (e.g. $\mathbf{X}$), matrices (e.g. $\boldsymbol{X}$). and tensors (e.g. $\mathbf{X}$). In particular, for $d \in \mathbb{N}$, the symbol $\boldsymbol{I}_d$ represents the d-dimensional identity matrix, where $\mathbb{N}$ refers to the set of positive integers. Let $n \in \mathbb{N}$. We denote the set $\{1, \ldots, n\}$ by $[n]$, and given a set S employ the notation $\binom{S}{n}$ to refer to the family of all subsets of S containing exactly n elements of S. Let $\boldsymbol{x} \in \mathbb{R}^d$. The notation $\|\boldsymbol{x}\|_2$ represents the euclidean norm of $\boldsymbol{x}$ while $\|\boldsymbol{x}\|_\infty$ denotes its maximum-norm. Moreover, given $r \in \mathbb{R}_{>0}$ we denote the set $\{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y} - \boldsymbol{x}\|_\infty \leq r\}$ by $\mathcal{B}^d_\infty(\boldsymbol{x}, r)$. We represent the variance of an arbitrary random variable X by σ^2_X and its density function by φ_X. Finally, the notation $\log(\cdot)$ refers to the binary logarithm. Let $d, n \in \mathbb{N}$ and $\varepsilon \in \mathbb{R}_{>0}$, and consider $\boldsymbol{z} \in [-1, 1]^d$

and n independent standard normal d-dimensional random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$. Given $S \subseteq [n]$ we define the random variable

$$\mathrm{Y}_{S,\varepsilon,\boldsymbol{z},\mathbf{X}_1,\dots,\mathbf{X}_n} = \begin{cases} 1 & \text{if } \|\boldsymbol{z} - \sum_{i\in S} \mathbf{X}_i\|_\infty \leq \varepsilon, \\ 0 & \text{otherwise,} \end{cases}$$

that we represent simply by Y_S when the other parameters are clear from context. Since we are interested in studying families of subsets, we also define, for $\mathcal{C}$ contained in the power set of $[n]$, the random variable

$$\mathrm{Y}_{\mathcal{C},\varepsilon,\boldsymbol{z},\mathbf{X}_1,\dots,\mathbf{X}_n} = \sum_{S\in\mathcal{C}} \mathrm{Y}_S,$$

which we represent simply as Y.

We control the stochastic dependency among subsets by restricting to families of subsets with small pairwise intersection. While this reduces how many subsets we can be considered, we can use the probabilistic method to prove that large families are still available.

Lemma 3. *For all $n \in \mathbb{N}$ and $\alpha \in (0, \frac{1}{2})$, there exists $\mathcal{C} \subseteq \binom{[n]}{\alpha n}$ with $|\mathcal{C}| \geq 2^{\frac{\alpha^2 n}{6}}$ such that for all $S, T \in \mathcal{C}$, if $S \neq T$, then*

$$|S \cap T| \leq 2\alpha^2 n.$$

Notice that, while this amount is still exponential, it already imposes $n = \mathcal{O}(\frac{d}{\alpha^2} \log \frac{1}{\varepsilon})$ if we are to approximate all points in $[-1,1]^d$ up to error ε.

5 Proof of the main result

As we frequently consider values relatively close to the origin, approximation of the normal distribution by a uniform one is sufficient for many of our estimations.

Lemma 4. *Let $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $\sigma \in \mathbb{R}_{>0}$, and $\boldsymbol{z} \in [-1,1]^d$. If $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \cdot \boldsymbol{I}_d)$, then*

$$e^{-\frac{2d}{\sigma^2}} \cdot \frac{(2\varepsilon)^d}{\left(2\pi\sigma^2\right)^{\frac{d}{2}}} \leq \Pr\left[\mathbf{X} \in \mathcal{B}_\infty^d(\boldsymbol{z}, \varepsilon)\right] \leq \frac{(2\varepsilon)^d}{\left(2\pi\sigma^2\right)^{\frac{d}{2}}}.$$

As a corollary, we bound the first moment of the random variable Y.

Corollary 5. *Given $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\alpha \in (0, \frac{1}{2})$, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent standard normal d-dimensional random vectors. Then, for all $\boldsymbol{z} \in [-1,1]^d$ and $\mathcal{C} \subseteq \binom{[n]}{\alpha n}$, it holds that*

$$e^{-\frac{2d}{\alpha n}} \frac{(2\varepsilon)^d |\mathcal{C}|}{(2\pi\alpha n)^{\frac{d}{2}}} \leq \mathbb{E}[\mathrm{Y}] \leq \frac{(2\varepsilon)^d |\mathcal{C}|}{(2\pi\alpha n)^{\frac{d}{2}}}.$$

Proof. Let $S \in \mathcal{C}$ and, hence, $|S| = \alpha n$. Since $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$ for all $i \in [n]$, we have that $\sum_{i\in S} \mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \alpha n \cdot \boldsymbol{I}_d)$. Therefore, as $\Pr[\mathrm{Y}_S = 1] = \Pr\left[\sum_{i\in S} \mathbf{X}_i \in \mathcal{B}_\infty^d(\boldsymbol{z}, \varepsilon)\right]$, by Lemma 4, we have that

$$e^{-\frac{2d}{\alpha n}} \frac{(2\varepsilon)^d}{(2\pi\alpha n)^{\frac{d}{2}}} \leq \Pr[\mathrm{Y}_S = 1] \leq \frac{(2\varepsilon)^d}{(2\pi\alpha n)^{\frac{d}{2}}},$$

and we can conclude the thesis by noting that $\mathbb{E}[\mathrm{Y}] = \sum_{S\in\mathcal{C}} \Pr[\mathrm{Y}_S = 1]$. □

We proceed by estimating the second moment of Y.

Lemma 6. *Given $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\alpha \in (0, \frac{1}{6}]$, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent d-dimensional standard normal random vectors, $z \in [-1,1]^d$, and $\mathcal{C} \subseteq \binom{[n]}{\alpha n}$. If $n \geq \frac{81}{\alpha(1-2\alpha)}$ and any two subsets in $\mathcal{C}$ intersect in at most $2\alpha^2 n$ elements, then*

$$\operatorname{Var}[\mathrm{Y}] \leq \frac{(2\varepsilon)^{2d}\,|\mathcal{C}|^2}{(2\pi\alpha n)^d} \cdot \left[(1-4\alpha^2)^{-\frac{d}{2}} - e^{-\frac{4d}{\alpha n}}\right] + \frac{(2\varepsilon)^d\,|\mathcal{C}|}{(2\pi\alpha n)^{\frac{d}{2}}}.$$

Proof. We have

$$\begin{aligned}
\operatorname{Var}[\mathrm{Y}] &= \sum_{S,T\in\mathcal{C}} \operatorname{Cov}[\mathrm{Y}_S, \mathrm{Y}_T] \\
&= \sum_{S,T\in\mathcal{C}} \left(\mathbb{E}[\mathrm{Y}_S \cdot \mathrm{Y}_T] - \mathbb{E}[\mathrm{Y}_S]\,\mathbb{E}[\mathrm{Y}_T]\right) \\
&= \sum_{S,T\in\mathcal{C}} \left(\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1] - \Pr[\mathrm{Y}_S = 1]\Pr[\mathrm{Y}_T = 1]\right) \\
&= \sum_{S\neq T\in\mathcal{C}} \left(\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1] - \Pr[\mathrm{Y}_S = 1]^2\right) + \sum_{S\in\mathcal{C}} \Pr[\mathrm{Y}_S = 1]\left(1 - \Pr[\mathrm{Y}_S = 1]\right).
\end{aligned}$$

We shall use Lemma 4 to estimate $\Pr[\mathrm{Y}_S = 1]$, thus, the core of our argument is to bound the joint probability $\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1]$. To this end, since $\operatorname{Cov}[\mathrm{Y}_S, \mathrm{Y}_T]$ increases monotonically with $|S \cap T|$, we fix $S, T \in \mathcal{C}$ with $|S \cap T| = 2\alpha^2 n$. Moreover, since Y_S is defined in terms of the max-norm, we can analyse the associated event for each coordinate independently. So, we let $\mathrm{X}_1, \cdots, \mathrm{X}_n \sim \mathcal{N}(0,1)$ and $z \in [-1,1]$.

Consider the partitions $S = S_{\mathrm{A}} \cup S_{\mathrm{B}}$ and $T = T_{\mathrm{C}} \cup T_{\mathrm{B}}$, with $S_{\mathrm{B}} = T_{\mathrm{B}} = S \cap T$, and let

$$\mathrm{A} = \sum_{i\in S_{\mathrm{A}}} \mathrm{X}_i, \qquad \mathrm{C} = \sum_{i\in T_{\mathrm{C}}} \mathrm{X}_i, \qquad \mathrm{B} = \sum_{i\in S\cap T} \mathrm{X}_i.$$

In this way, we have $\sum_{i\in S} \mathrm{X}_i = \mathrm{A} + \mathrm{B}$ and $\sum_{i\in T} \mathrm{X}_i = \mathrm{C} + \mathrm{B}$, with A, C independent random variables distributed as $\mathcal{N}(0, \sigma_{\mathrm{A}}^2)$ and $\mathrm{B} \sim \mathcal{N}(0, \sigma_{\mathrm{B}}^2)$, where $\sigma_{\mathrm{A}}^2 = \alpha n(1-2\alpha)$ and $\sigma_{\mathrm{B}}^2 = 2\alpha^2 n$.

With this setup, we have,

$$\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1] = \Big(\Pr\left[\mathrm{A} + \mathrm{B} \in (z-\varepsilon, z+\varepsilon),\, \mathrm{C} + \mathrm{B} \in (z-\varepsilon, z+\varepsilon)\right]\Big)^d.$$

From the law of total probability, it holds that

$$\begin{aligned}
&\Pr\left[\mathrm{A} + \mathrm{B} \in (z-\varepsilon, z+\varepsilon),\, \mathrm{C} + \mathrm{B} \in (z-\varepsilon, z+\varepsilon)\right] \\
&\quad= \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \Pr\left[\mathrm{A} + x \in (z-\varepsilon, z+\varepsilon),\, \mathrm{C} + x \in (z-\varepsilon, z+\varepsilon)\right] \mathrm{d}x \\
&\quad= \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon),\, \mathrm{C} \in (z-x-\varepsilon, z-x+\varepsilon)\right] \mathrm{d}x \\
&\quad= \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \Big(\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\Big)^2 \mathrm{d}x, \qquad (1)
\end{aligned}$$

where the last equality follows from the independence of A and C.

Since A is a normal random variable with 0 average, by Claim 17, we have that

$$\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \Big(\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\Big)^2 \,\mathrm{d}x \le \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \Big(\Pr\left[\mathrm{A} \in (x-\varepsilon, x+\varepsilon)\right]\Big)^2 \,\mathrm{d}x$$
$$= \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \left(\int_{x-\varepsilon}^{x+\varepsilon} \varphi_{\mathrm{A}}(y)\,\mathrm{d}y\right)^2 \mathrm{d}x.$$

The hypothesis on n implies that $2\sigma_a^2 \ge 162$, so, by Claim 18,

$$\left(\int_{x-\varepsilon}^{x+\varepsilon} \varphi_{\mathrm{A}}(y)\,\mathrm{d}y\right)^2 \le \left[\int_{x-\varepsilon}^{x+\varepsilon} \frac{\exp\left(-\frac{(x+\varepsilon)^2}{2\sigma_{\mathrm{A}}^2}\right) + \exp\left(-\frac{(x-\varepsilon)^2}{2\sigma_{\mathrm{A}}^2}\right)}{2\sqrt{2\pi\sigma_{\mathrm{A}}^2}} \cdot \exp\!\left(\frac{\varepsilon^2}{2\sigma_{\mathrm{A}}^2}\right) \mathrm{d}y\right]^2$$
$$= \frac{(2\varepsilon)^2}{2\pi\sigma_{\mathrm{A}}^2} \cdot \frac{\exp\left(-\frac{(x+\varepsilon)^2}{\sigma_{\mathrm{A}}^2}\right) + \exp\left(-\frac{(x-\varepsilon)^2}{\sigma_{\mathrm{A}}^2}\right) + 2\exp\left(-\frac{x^2+\varepsilon^2}{\sigma_{\mathrm{A}}^2}\right)}{4} \cdot \exp\left(\frac{\varepsilon^2}{\sigma_{\mathrm{A}}^2}\right)$$
$$= e^{\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \frac{1}{\sqrt{2}} \cdot \frac{(2\varepsilon)^2}{\sqrt{2\pi\sigma_{\mathrm{A}}^2}} \cdot \frac{\varphi_{\mathrm{A}/\sqrt{2}}(x+\varepsilon) + \varphi_{\mathrm{A}/\sqrt{2}}(x-\varepsilon) + 2e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \varphi_{\mathrm{A}/\sqrt{2}}(x)}{4}.$$

Moreover, it holds that

$$\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \left[\varphi_{\mathrm{A}/\sqrt{2}}(x+\varepsilon) + \varphi_{\mathrm{A}/\sqrt{2}}(x-\varepsilon) + 2e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \varphi_{\mathrm{A}/\sqrt{2}}(x)\right] \mathrm{d}x$$
$$= (\varphi_{\mathrm{B}} * \varphi_{\mathrm{A}/\sqrt{2}})(\varepsilon) + (\varphi_{\mathrm{B}} * \varphi_{\mathrm{A}/\sqrt{2}})(-\varepsilon) + 2e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot (\varphi_{\mathrm{B}} * \varphi_{\mathrm{A}/\sqrt{2}})(0)$$
$$= \varphi_{\mathrm{B+A}/\sqrt{2}}(\varepsilon) + \varphi_{\mathrm{B+A}/\sqrt{2}}(-\varepsilon) + 2e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \varphi_{\mathrm{B+A}/\sqrt{2}}(0)$$
$$= \frac{2e^{-\varepsilon^2/\sigma^2_{\mathrm{B+A}/\sqrt{2}}} + 2e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2}}{\sqrt{2\pi\sigma^2_{\mathrm{B+A}/\sqrt{2}}}}$$
$$\le 4 \cdot \frac{e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2}}{\sqrt{2\pi\sigma^2_{\mathrm{B+A}/\sqrt{2}}}},$$

here $*$ denotes the convolution operation, and the last inequality comes from the hypothesis $\alpha \le \frac{1}{6}$, which implies that $\sigma^2_{\mathrm{B+A}/\sqrt{2}} \le \sigma_{\mathrm{A}}^2$.

Altogether, we have

$$\Pr\left[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1\right] \le \left(e^{\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \frac{1}{\sqrt{2}} \cdot \frac{(2\varepsilon)^2}{\sqrt{2\pi\sigma_{\mathrm{A}}^2}} \cdot \frac{e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2}}{\sqrt{2\pi\sigma^2_{\mathrm{B+A}/\sqrt{2}}}}\right)^d \tag{2}$$
$$= \left(\frac{(2\varepsilon)^2}{2\pi} \cdot \frac{1}{\sqrt{2\sigma_{\mathrm{A}}^2\sigma^2_{\mathrm{B+A}/\sqrt{2}}}}\right)^d$$
$$= \frac{(2\varepsilon)^{2d}}{(2\pi\alpha n)^d} \cdot (1-4\alpha^2)^{-\frac{d}{2}},$$

where the last equality follows from recalling that $\sigma_{\mathrm{B}}^2 = 2\alpha^2 n$ and $\sigma_{\mathrm{A}}^2 = \alpha n\,(1-2\alpha)$, and, thus, $\sigma_{\mathrm{B}+\mathrm{A}/\sqrt{2}}^2 = 2\alpha^2 n + \frac{\alpha t}{2}\,(1-2\alpha)$.

Finally, from this bound and from Lemma 4 we can conclude that

$$\begin{aligned}
\mathrm{Var}\left[\mathrm{Y}\right] &= \sum_{S\neq T\in\mathcal{C}} \Big(\Pr\left[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1\right] - \Pr\left[\mathrm{Y}_S = 1\right]^2\Big) + \sum_{S\in\mathcal{C}} \Pr\left[\mathrm{Y}_S = 1\right]\big(1 - \Pr\left[\mathrm{Y}_S = 1\right]\big) \\
&\leq \sum_{S\neq T\in\mathcal{C}} \left[\frac{(2\varepsilon)^{2d}}{(2\pi\alpha n)^d}\cdot(1-4\alpha^2)^{-\frac{d}{2}} - \frac{(2\varepsilon)^{2d}}{(2\pi\alpha n)^d}\cdot e^{-\frac{4d}{\alpha n}}\right] + \sum_{S\in\mathcal{C}} \frac{(2\varepsilon)^d}{(2\pi\alpha n)^{\frac{d}{2}}}\left[1 - e^{-\frac{2d}{\alpha n}}\cdot\frac{(2\varepsilon)^d}{(2\pi\alpha n)^{\frac{d}{2}}}\right] \\
&\leq \frac{(2\varepsilon)^{2d}\left|\mathcal{C}\right|^2}{(2\pi\alpha n)^d}\cdot\left[(1-4\alpha^2)^{-\frac{d}{2}} - e^{-\frac{4d}{\alpha n}}\right] + \frac{(2\varepsilon)^d\left|\mathcal{C}\right|}{(2\pi\alpha n)^{\frac{d}{2}}}.
\end{aligned}$$

□

Remark 7. *In the proof of Lemma 6, after applying the law of total probability it is possible to employ Lemma 4 to estimate the joint probability. While this simplifies the argument, doing so would ultimately weaken the bound in Theorem 9 by a factor of d. In fact, in Section 7 we argue that the estimation we provide is essentially optimal.*

For our next result, recall that the existence of a suitable family of subsets is ensured by Lemma 3.

Lemma 8. *Given $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\alpha \in (0, \frac{1}{6}]$, let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent d-dimensional standard normal random vectors, $z \in [-1,1]^d$, and $\mathcal{C} \subseteq \binom{[n]}{\alpha n}$ with $|\mathcal{C}| \geq 2^{\frac{\alpha^2 n}{6}}$. If any two subsets in $\mathcal{C}$ intersect in at most $2\alpha^2 n$ elements, $\alpha \leq \frac{1}{6\sqrt{d}}$, and*

$$n \geq \frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right),$$

then

$$\Pr\left[\mathrm{Y} \geq 1\right] \geq \frac{1}{3}.$$

Proof. By Chebyshev's inequality, it holds that

$$\begin{aligned}
\Pr\left[\mathrm{Y} \geq 1\right] &\geq \Pr\left[\left|\mathrm{Y} - \mathbb{E}\left[\mathrm{Y}\right]\right| < \frac{\mathbb{E}\left[\mathrm{Y}\right]}{2}\right] \\
&\geq 1 - \frac{4\cdot\mathrm{Var}\left[\mathrm{Y}\right]}{\mathbb{E}\left[\mathrm{Y}\right]^2}.
\end{aligned}$$

Applying Corollary 5 and Lemma 6, we get that

$$\begin{aligned}
\frac{4\cdot\mathrm{Var}\left[\mathrm{Y}\right]}{\mathbb{E}\left[\mathrm{Y}\right]^2} &\leq 4\cdot\frac{e^{\frac{4d}{\alpha n}}\cdot(2\pi\alpha n)^d}{(2\varepsilon)^{2d}\left|\mathcal{C}\right|^2}\cdot\left[\frac{(2\varepsilon)^{2d}\left|\mathcal{C}\right|^2}{(2\pi\alpha n)^d}\cdot\left[(1-4\alpha^2)^{-\frac{d}{2}} - e^{-\frac{4d}{\alpha n}}\right] + \frac{(2\varepsilon)^d\left|\mathcal{C}\right|}{(2\pi\alpha n)^{\frac{d}{2}}}\right] \\
&= 4\cdot\left(\frac{e^{\frac{4d}{\alpha n}}}{(1-4\alpha^2)^{\frac{d}{2}}} - 1\right) + \frac{4e^{\frac{4d}{\alpha n}}\cdot(2\pi\alpha n)^{\frac{d}{2}}}{(2\varepsilon)^d\left|\mathcal{C}\right|}.
\end{aligned}$$

Since $n \geq \frac{68d}{\alpha}$ and $\alpha \leq \frac{1}{6\sqrt{d}}$, by Claim 15

$$4 \cdot \left(\frac{e^{\frac{4d}{\alpha n}}}{(1-4\alpha^2)^{\frac{d}{2}}} - 1 \right) \leq \frac{1}{2}.$$

Furthermore, as $n \geq \frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)$, $|\mathcal{C}| \geq 2^{\frac{\alpha^2 n}{6}}$, and $\alpha \leq \frac{1}{6}$, by Claim 16,

$$\frac{4e^{\frac{4d}{\alpha n}} \cdot (2\pi\alpha n)^{\frac{d}{2}}}{(2\varepsilon)^d |\mathcal{C}|} \leq \varepsilon.$$

□

Applying an union bound, we amplify the last lemma to get our main result.

Theorem 9. *Let $\varepsilon \in (0,1)$ and given $d, n \in \mathbb{N}$ let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent standard normal d-dimensional random vectors and let $\alpha \in \left(0, \frac{1}{6\sqrt{d}}\right]$. There exists a universal constant $C > 0$ such that, if*

$$n \geq C\frac{d^2}{\alpha^2}\log\frac{1}{\varepsilon} \cdot \left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right),$$

then, with probability

$$1 - \exp\left[-\ln 2 \cdot \left(\frac{n}{C\frac{d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)} - d\log\frac{1}{\varepsilon}\right)\right],$$

for all $\boldsymbol{z} \in [-1,1]^d$ there exists a subset $S_{\boldsymbol{z}} \subseteq [n]$ for which

$$\left\| \boldsymbol{z} - \sum_{i \in S_{\boldsymbol{z}}} \mathbf{X}_i \right\|_\infty \leq \varepsilon.$$

Moreover, this remains true even when restricted to subsets of size αn.

Theorem 2 follows from Theorem 9 by setting $\alpha = \frac{1}{6\sqrt{d}}$.

6 Application to Neural Net Evolution

In this section, we present an application of our main result on the multidimensional RSSP (see Theorem 2) to a neural network model recently introduced in [GLP+19].

We first provide a description of their model in a setting relevant to our application. Then, we prove that their model exhibits *universality*; namely, with high probability, it can approximate any neural network within a polynomial overhead in the number of parameters.

6.1 The NNE model

The *Neural Net Evolution* (NNE) model [GLP+19] has been recently introduced as an alternative approach to train neural networks, based on evolutionary methods. The aim is to provide a biologically inspired alternative to the backpropagation process behind ANNs [RHW86, GBC16], which happens in evolutionary time, instead of lifetime.

The NNE model is inspired by a standard update rule in population genetics and, in [GLP+19], it is shown to succeed in creating neural networks that can learn linear classification problems reasonably well with no explicit backpropagation.

To define the NNE model, we first need to define random genotypes. Given a vector $\boldsymbol{p} \in [0,1]^n$, a random *genotype* $\boldsymbol{x} \in \{0,1\}^n$ is sampled by setting $x_i = 1$ with probability p_i, independently for each i. Each entry x_i indicates whether or not a *gene* is active.

Then, for each i, a random tensor $\boldsymbol{\Theta}^{(i)} \in \mathbb{R}^{\ell \times d \times d}$ is sampled. In the original version of the model [GLP+19], each entry of the tensor is chosen independently and uniformly at random from $[-1,1]$ with probability β, while is set to 0, otherwise. For the sake of our application, we here consider a natural variant where the entries of the tensor are independently drawn from a standard normal distribution.

Now, given a genotype $\boldsymbol{x} \in \{0,1\}^n$, we define

$$\boldsymbol{\Theta}_{\boldsymbol{x}} = \sum_{i\,:\,x_i=1} \boldsymbol{\Theta}^{(i)}. \tag{3}$$

Each genotype is then associated with a *feed-forward neural network*, represented by a weighted complete multipartite directed graph. The graph is formed by layers $\{L_i\}_{i=0}^{\ell}$ of d nodes and two consecutive layers are fully connected via a biclique whose edge weights are determined by the tensor $\boldsymbol{\Theta}_{\boldsymbol{x}}$ in the following manner: for every $i \in [\ell]$, the edge between the j-th node of layer L_{i-1} and the k-th node of layer L_i has weight $(\boldsymbol{\Theta}_{\boldsymbol{x}})_{ijk}$.

Eq. (3) tells us that if a gene is active then it gives a random contribution to each weight of the genotype network.

The learning process in the NNE model works by updating the genotype probabilities $\boldsymbol{p}$ according to some standard population genetics equations [Bür00, CLPV14]. In [GLP+19], it is proved that the adopted update rule indirectly performs backpropagation and enables to decrease the loss function of the networks.

6.2 Universality and RSSP

Let $f\colon \mathbb{R}^d \to \mathbb{R}^d$ be a feed-forward neural network of the form

$$f(\boldsymbol{y}) = \mathbf{W}_\ell\, \sigma(\mathbf{W}_{\ell-1} \ldots \sigma(\mathbf{W}_1\, \boldsymbol{y})), \tag{4}$$

where $\mathbf{W}_i \in \mathbb{R}^{d\times d}$ is a weight matrix and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^d$ is the ReLU (Rectified Linear Unit) activation function that converts each coordinate y_i of a given vector $\boldsymbol{y} \in \mathbb{R}^d$ to $\max(0, y_i)$.

The restrictions on the weight matrix sizes $d \times d$ aim only to ease presentation and can be adapted to any arbitrary dimensions.

Let us construct a third-order tensor $\boldsymbol{\Theta}_f \in \mathbb{R}^{\ell\times d\times d}$ by stacking the weight matrices $\mathbf{W}_1, \ldots, \mathbf{W}_\ell$. We correspondingly denote f by $f_{\boldsymbol{\Theta}}$. Conversely, every tensor $\boldsymbol{\Theta} \in \mathbb{R}^{\ell\times d\times d}$ is associated with a neural network $f_{\boldsymbol{\Theta}}$ in the form of Eq. (4) whose corresponding weight matrices are the tensor slices, that is, $\mathbf{W}_m = (\boldsymbol{\Theta})_{\substack{i=m \\ j,k\in[d]}}$ for every $m \in [\ell]$.

We can use Theorem 2 to prove a notion of universality for the NNE model.

Theorem 10. *Let $\varepsilon > 0$ and $n, d, \ell \in \mathbb{N}$. Let $\mathcal{F}$ be the class of neural networks $f\colon \mathbb{R}^d \to \mathbb{R}^d$ of the form given in Eq. (4) such that their corresponding tensor satisfies $\max_{ijk}|(\Theta_f)_{ijk}| < 1$. A constant $C > 0$ exists such that, if $n \geq C(\ell \cdot d \cdot d)^3 \log \frac{1}{\varepsilon} \cdot \left(\log \frac{1}{\varepsilon} + \log(\ell \cdot d \cdot d)\right)$, then, with high probability, the tensors $\Theta^{(1)}, \ldots, \Theta^{(n)}$ associated to each gene are such that, for any $f \in \mathcal{F}$, there is a genotype $x \in \{0,1\}^n$ which satisfies*

$$\max_{\substack{i\in[\ell]\\ j,k\in[d]}} \left|(\Theta_f)_{ijk} - (\Theta_x)_{ijk}\right| < \varepsilon.$$

We note that standard techniques (e.g., [PRN+20, dCNV21]) can be used to provide bounds on the approximation of the output of neural networks, as well as translating Theorem 10 for general network architectures (e.g., convolutional neural networks).

7 Tightness of analysis

In Lemma 3 we prove the existence of a suitable family of subsets via a probabilistic argument, sampling their elements uniformly at random. The same argument also implies that the pairwise intersections of almost all subsets is at least $\frac{\alpha^2 n}{2}$. In the next result, we assume such lower bound and prove that our estimation of the joint probability $\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1]$ in Lemma 6 (specifically, in Eq. 2), is essentially tight. Namely, the next lemma implies that it is not possible to obtain a high-probability bound on Y in Lemma 8.

Lemma 11. *Given $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\alpha \in (0, \frac{1}{2})$, let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent standard normal d-dimensional random vectors and $z \in [-1,1]^d$. If any two subsets in $\mathcal{C}$ intersect in at least $\frac{\alpha^2 n}{2}$ elements and $n \geq \frac{10}{\alpha(2-\alpha)}$, then*

$$\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1] \geq \frac{(2\varepsilon)^{2d}}{(2\pi\alpha n)^d} \cdot \left(1 - \frac{\alpha^2}{4}\right)^{-\frac{d}{2}} \cdot \exp\left(-\frac{3d}{\alpha n}\right).$$

We can extend the above result by letting z lie in a wider range. This will be useful for the generalisation section Appendix C.

Remark 12. *If $\lambda > 1$ and $z \in [-\lambda\sqrt{n}, \lambda\sqrt{n}]^d$, then we have*

$$\Pr[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1] \geq \frac{(2\varepsilon)^{2d}}{(2\pi\alpha n)^d} \cdot \left(1 - \frac{\alpha^2}{4}\right)^{-\frac{d}{2}} \cdot \exp\left(-\frac{3\lambda^2 d}{\alpha}\right).$$

8 Acknowledgements

We would like to thank Bianca C. Araújo and Paulo B. S. Serafim for their feedback on preliminary versions of the manuscript, and Michele Salvi for the helpful discussions about the problem.

This work has been supported by the AID INRIA-DGA agreement n°2019650072, by the Visiting 2021 funding (Decreto n. 1589/2021 2/7/2021) of University of Rome "Tor Vergata", and partially supported by the ERC Advanced Grant 788893 AMDROMA "Algorithmic and Mechanism Design Research in Online Markets", the EC H2020RIA project "SoBigData++" (871042), the MIUR PRIN project ALGADIMAR "Algorithms, Games, and Digital Markets, and the French government through the UCA JEDI (ANR-15-IDEX-01) and EUR DS4H (ANR-17-EURE-004) Investments in the Future projects.

References

[BCMP04] C. Borgs, J. T. Chayes, S. Mertens, and B. Pittel. Phase diagram for the constrained integer partitioning problem. *Random Structures & Algorithms*, 24(3):315–380, 2004. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20001. [Cited on 1.]

[BCP01] Christian Borgs, Jennifer T. Chayes, and Boris G. Pittel. Phase transition and finite-size scaling for the integer partitioning problem. *Random Struct. Algorithms*, 19(3-4):247–288, 2001. [Cited on 1 and 2.]

[BLMG22] Rebekka Burkholz, Nilanjana Laha, Rajarshi Mukherjee, and Alkis Gotovos. On the existence of universal lottery tickets. In *International Conference on Learning Representations*, 2022. [Cited on 1 and 2.]

[Bür00] Reinhard Bürger. *The mathematical theory of selection, recombination, and mutation.* John Wiley & Sons, 2000. [Cited on 6.1.]

[CJRS22] Xi Chen, Yaonan Jin, Tim Randolph, and Rocco A. Servedio. Average-case subset balancing problems. In Joseph (Seffi) Naor and Niv Buchbinder, editors, *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 743–778. SIAM, 2022. [Cited on 2.]

[CLPV14] Erick Chastain, Adi Livnat, Christos Papadimitriou, and Umesh Vazirani. Algorithms, games, and evolution. *Proceedings of the National Academy of Sciences*, 111(29):10620–10623, 2014. [Cited on 6.1.]

[dCdG+22] Arthur da Cunha, Francesco d'Amore, Frédéric Giroire, Hicham Lesfari, Emanuele Natale, and Laurent Viennot. Revisiting the Random Subset Sum problem, April 2022. Number: arXiv:2204.13929 arXiv:2204.13929 [math]. [Cited on 2 and 3.1.]

[dCNV21] Arthur da Cunha, Emanuele Natale, and Laurent Viennot. Proving the Lottery Ticket Hypothesis for Convolutional Neural Networks. In *International Conference on Learning Representations (ICLR)*, September 2021. [Cited on 1, 2, and 6.2.]

[DP09] Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009. [Cited on 14.]

[FB21] Jonas Fischer and Rebekka Burkholz. Towards strong pruning for lottery tickets with non-zero biases. *CoRR*, abs/2110.11150, 2021. [Cited on 1.]

[FTGB22] Damien Ferbach, Christos Tsirigotis, Gauthier Gidel, and Avishek Joey Bose. A general framework for proving the equivariant strong lottery ticket hypothesis. *CoRR*, abs/2206.04270, 2022. [Cited on 1 and 2.]

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning.* MIT press, 2016. [Cited on 6.1.]

[GLP+19] Sruthi Gorantla, Anand Louis, Christos H. Papadimitriou, Santosh Vempala, and Naganand Yadati. Biologically Plausible Neural Networks via Evolutionary Dynamics and Dopaminergic Plasticity. In *International Conference on Learning Representations (ICLR)*, 2019. [Cited on 1, 1, 6, 6.1, and 6.1.]

[Jan04] Svante Janson. Large deviations for sums of partly dependent random variables: Large Deviations for Dependent Random Variables. *Random Structures & Algorithms*, 24(3):234–248, May 2004. [Cited on 3.1.]

[KKLO86] Narendra Karmarkar, Richard M Karp, George S Lueker, and Andrew M Odlyzko. Probabilistic analysis of optimum partitioning. *Journal of Applied probability*, 23(3):626–645, 1986. [Cited on 1 and 2.]

[Lue82] George S Lueker. On the average difference between the solutions to linear and integer knapsack problems. In *Applied Probability-Computer Science: The Interface Volume 1*, pages 489–504. Springer, 1982. [Cited on 1 and 2.]

[Lue98] George S. Lueker. Exponentially small bounds on the expected optimum of the partition and subset sum problems. *Random Structures & Algorithms*, 12(1):51–62, 1998. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291098-2418%28199801%2912%3A1%3C51%3A%3AAID-RSA3%3E3.0.CO%3B2-S. [Cited on 1, 1, 1, 2, 3.1, and C.]

[Mer98] Stephan Mertens. Phase transition in the number partitioning problem. *Physical Review Letters*, 81:4281–4284, 1998. [Cited on 2.]

[MM09] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford graduate texts. Oxford University Press, Oxford ; New York, 2009. OCLC: ocn234430714. [Cited on 1.]

[MP10] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010. [Cited on E.]

[PLW19] Élie de Panafieu, Mohamed Lamine Lamali, and Michael Wallner. Combinatorics of nondeterministic walks of the dyck and motzkin type. In *2019 Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 1–12. SIAM, 2019. [Cited on E.]

[PRN+20] Ankit Pensia, Shashank Rajput, Alliot Nagle, Harit Vishwakarma, and Dimitris S. Papailiopoulos. Optimal lottery tickets via subset sum: Logarithmic over-parameterization is sufficient. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. [Cited on 1, 2, and 6.2.]

[RHW86] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. [Cited on 6.1.]

[War16] Lutz Warnke. On the method of typical bounded differences. *Combinatorics, Probability and Computing*, 25(2):269–299, 2016. [Cited on 3.1.]

[WDM+21] Chenghong Wang, Jieren Deng, Xianrui Meng, Yijue Wang, Ji Li, Sheng Lin, Shuo Han, Fei Miao, Sanguthevar Rajasekaran, and Caiwen Ding. A secure and efficient federated learning framework for NLP. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 7676–7682. Association for Computational Linguistics, 2021. [Cited on 1.]

[WRG17] Yuyi Wang, Jan Ramon, and Zheng-Chu Guo. Learning from networked examples. *arXiv:1405.2600 [cs, stat]*, June 2017. arXiv: 1405.2600. [Cited on 3.1.]

A Tools

Below we list some standard tools we use, and prove some inequalities.

A.1 Concentration bounds

Theorem 13 (Chebyshev's inequality)**.** *Let* X *be a random variable with finite expected value* μ *and finite non-zero variance* σ^2*. Then for any real number* $k > 0$*, it holds that*

$$\Pr\left[|\mathrm{X} - \mu| \geq k\right] \leq \frac{\sigma^2}{k^2}.$$

Lemma 14 (Chernoff-Hoeffding bounds [DP09])**.** *Let* $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ *be independent random variables such that*

$$\Pr\left[0 \leq \mathrm{X}_i \leq 1\right] = 1$$

for all $i \in [n]$*. Let* $\mathrm{X} = \sum_{i=1}^{n} \mathrm{X}_i$ *and* $\mathbb{E}\left[\mathrm{X}\right] = \mu$*. Then, for any* $\delta \in (0, 1)$ *the following holds:*

1. *if* $\mu \leq \mu_+$*, then* $\Pr\left[\mathrm{X} \geq (1+\delta)\mu_+\right] \leq \exp\left(-\frac{\delta^2 \mu_+}{3}\right)$;
2. *if* $0 \leq \mu_- \leq \mu$*, then* $\Pr\left[\mathrm{X} \leq (1-\delta)\mu_-\right] \leq \exp\left(-\frac{\delta^2 \mu_+}{2}\right)$.

A.2 Claims

Claim 15. *Let* $d, n \in \mathbb{N}$ *and* $\alpha \in \mathbb{R}_{>0}$*. If* $n \geq \frac{68d}{\alpha}$ *and* $\alpha \leq \frac{1}{6\sqrt{d}}$*, then*

$$e^{\frac{4d}{\alpha n}} \cdot \frac{1}{\left(1 - 4\alpha^2\right)^{\frac{d}{2}}} \leq 1 + \frac{1}{8}.$$

Proof. Since $e^x \leq (1-x)^{-1}$ for $x \leq 1$, for $n \geq \frac{4d}{\alpha}$, it holds that

$$e^{\frac{4d}{\alpha n}} \leq \frac{1}{1 - \frac{4d}{\alpha n}} = 1 + \frac{4d}{\alpha n - 4d}.$$

Thus, having $n \geq \frac{68d}{\alpha}$ implies that

$$e^{\frac{4d}{\alpha n}} \leq 1 + \frac{1}{16}.$$

Moreover, by Bernoulli's inequality, since $\alpha < \frac{1}{2}$, it holds that,

$$\frac{1}{\left(1 - 4\alpha^2\right)^{\frac{d}{2}}} \leq \frac{1}{1 - 2d\alpha^2}.$$

Altogether, we need that

$$\frac{1 + \frac{1}{16}}{1 - 2d\alpha^2} \leq 1 + \frac{1}{8},$$

which holds for $\alpha \leq \frac{1}{6\sqrt{d}}$. □

Claim 16. *Let $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\alpha \in (0, \frac{1}{6})$. If $n \geq \frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)$, then*

$$\frac{4e^{\frac{4d}{\alpha n}}}{2^{\frac{\alpha^2 n}{6}}} \cdot \left(\frac{\pi\alpha n}{2\varepsilon^2}\right)^{\frac{d}{2}} \leq \varepsilon.$$

Proof. Consider the function

$$f(n) = \frac{n^d}{2^{\frac{\alpha^2 n}{6}}}.$$

We have that

$$\begin{aligned} f'(n) &= \frac{dn^{d-1}2^{\frac{\alpha^2 n}{6}} - \frac{\alpha^2 \ln 2}{6} \cdot n^d 2^{\frac{\alpha^2 n}{6}}}{2^{\frac{\alpha^2 n}{3}}} \\ &= \frac{n^{d-1}2^{\frac{\alpha^2 n}{6}}}{2^{\frac{\alpha^2 n}{3}}} \cdot \left(d - \frac{\alpha^2 n \ln 2}{6}\right), \end{aligned}$$

and, hence, f is non-increasing for $n \geq \frac{6d}{\alpha^2 \ln 2}$. Thus, since $f\left(\frac{6d}{\alpha^2 \ln 2}\right) = \left(\frac{6d}{e\alpha^2 \ln 2}\right)^d$, it holds that

$$\begin{aligned} \frac{4e^{\frac{4d}{\alpha n}}}{2^{\frac{\alpha^2 n}{6}}} \cdot \left(\frac{\pi\alpha n}{2\varepsilon^2}\right)^{\frac{d}{2}} &= \frac{4e^{\frac{4d}{\alpha n}}}{2^{\frac{\alpha^2 n}{12}}} \cdot \left(\frac{\pi\alpha}{2\varepsilon^2}\right)^{\frac{d}{2}} \sqrt{\frac{n^d}{2^{\frac{\alpha^2 n}{6}}}} \\ &\leq \frac{4e^{\frac{4d}{\alpha n}}}{2^{\frac{\alpha^2 n}{12}}} \cdot \left(\frac{\pi\alpha}{2\varepsilon^2}\right)^{\frac{d}{2}} \sqrt{\left(\frac{6d}{e\alpha^2 \ln 2}\right)^d} \\ &= \frac{4e^{\frac{4d}{\alpha n}}}{2^{\frac{\alpha^2 n}{12}}} \cdot \left(\frac{3\pi d}{\varepsilon^2 e\alpha \ln 2}\right)^{\frac{d}{2}} \\ &< \frac{8}{2^{\frac{\alpha^2 n}{12}}} \cdot \left(\frac{6d}{\varepsilon^2\alpha}\right)^{\frac{d}{2}} \end{aligned}$$

where the last inequality comes from noting that $6 > \frac{3\pi}{e \ln 2}$ and that $n \geq \frac{8d}{\alpha}$ implies $e^{\frac{4d}{\alpha n}} < 2$. This is at most ε if

$$\frac{8}{\varepsilon^{d+1}} \cdot \left(\frac{6d}{\alpha}\right)^{\frac{d}{2}} \leq 2^{\frac{\alpha^2 n}{12}},$$

or, equivalently,

$$n \geq \frac{12}{\alpha^2}\left(\log 8 + (d+1)\log\frac{1}{\varepsilon} + \frac{d}{2}\log\frac{1}{\alpha} + \frac{d}{2}\log 6d\right).$$

The thesis follows from the bounds $d, n \geq 1$, $\varepsilon \in (0,1)$, and $\alpha < \frac{1}{6}$. □

Claim 17. *Let* A, B *be two centred normal random variables, and let* $\varphi_{\mathrm{B}}(x)$ *be the density function of* B*. Then, for any* $z \in \mathbb{R}$*, for any* $\varepsilon > 0$*, it holds that*

$$\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\left[\Pr\left[\mathrm{A} \in (z - x - \varepsilon, z - x + \varepsilon)\right]\right]^2 \mathrm{d}x \leq \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\left[\Pr\left[\mathrm{A} \in (-x - \varepsilon, -x + \varepsilon)\right]\right]^2 \mathrm{d}x.$$

Proof. For any $x, z \in \mathbb{R}$, let

$$h(x,z) = \varphi_{\mathrm{B}}(x) \left[\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\right]^2 \mathrm{d}x,$$

and let

$$H(z) = \int_{\mathbb{R}} h(x,z)\,\mathrm{d}x.$$

Let $\varphi_{\mathrm{A}}(x)$ be the density function of a. Since

$$\begin{aligned}\left|\frac{\partial h(x,z)}{\partial z}\right| &= 2\left|\varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon) - \varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\right| \\ &\leq 2\varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon) + \varphi_{\mathrm{A}}(z-x-\varepsilon)\right),\end{aligned}$$

$h(x,z)$ meets the hypothesis of the Leibniz integral rule and we can write

$$\begin{aligned}\frac{\mathrm{d}H(z)}{\mathrm{d}z} &= \int_{\mathbb{R}} \frac{\partial h(x,z)}{\partial z}\,\mathrm{d}x \\ &= 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon) - \varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\,\mathrm{d}x\end{aligned}$$

If we prove that such a function is zero in $z = 0$, positive for $z < 0$ and negative for $z > 0$, then we have that the maximum of H is reached in $z = 0$.

First case: $z = 0$. Then

$$\begin{aligned}\frac{\mathrm{d}H(0)}{\mathrm{d}z} &= 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (x-\varepsilon, x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(x-\varepsilon) - \varphi_{\mathrm{A}}(x+\varepsilon)\right)\,\mathrm{d}x \qquad (5) \\ &= 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (x-\varepsilon, x+\varepsilon)\right]\varphi_{\mathrm{A}}(x-\varepsilon)\,\mathrm{d}x \\ &\quad - 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (x-\varepsilon, x+\varepsilon)\right]\varphi_{\mathrm{A}}(x+\varepsilon)\,\mathrm{d}x \\ &= 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (x-\varepsilon, x+\varepsilon)\right]\varphi_{\mathrm{A}}(x-\varepsilon)\,\mathrm{d}x \\ &\quad - 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(y)\Pr\left[\mathrm{A} \in (y-\varepsilon, y+\varepsilon)\right]\varphi_{\mathrm{A}}(y-\varepsilon)\,\mathrm{d}x \qquad (6) \\ &= 0,\end{aligned}$$

where in Eq. (5) we exploited the symmetry of the integrand functions, Eq. (6) we substituted in the second integral $y = -x$ and used again symmetry.

Second case: $z > 0$. Then

$$
\begin{aligned}
&\frac{\mathrm{d}H(z)}{\mathrm{d}z} \\
&= 2\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&= 2\int_{-\infty}^{-z} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&\quad + 2\int_{-z}^{+z} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&\quad + 2\int_{+z}^{+\infty} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&= 2\int_{+z}^{+\infty} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z+x-\varepsilon, z+x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z+x+\varepsilon)-\varphi_{\mathrm{A}}(z+x-\varepsilon)\right)\mathrm{d}x \qquad (7) \\
&\quad + 2\int_{+3z}^{+\infty} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&\quad + 2\int_{+z}^{+3z} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&\quad + 2\int_{-z}^{+z} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&= 2\int_{+z}^{+\infty} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z+x-\varepsilon, z+x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z+x+\varepsilon)-\varphi_{\mathrm{A}}(z+x-\varepsilon)\right)\mathrm{d}x \\
&\quad - 2\int_{+z}^{+\infty} \varphi_{\mathrm{B}}(2z+x)\Pr\left[\mathrm{A} \in (z+x-\varepsilon, z+x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z+x+\varepsilon)-\varphi_{\mathrm{A}}(z+x-\varepsilon)\right)\mathrm{d}x \qquad (8) \\
&\quad - 2\int_{-z}^{+z} \varphi_{\mathrm{B}}(x-2z)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \qquad (9) \\
&\quad + 2\int_{-z}^{+z} \varphi_{\mathrm{B}}(x)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x \\
&= 2\int_{+z}^{+\infty} \left(\varphi_{\mathrm{B}}(x)-\varphi_{\mathrm{B}}(2z+x)\right)\Pr\left[\mathrm{A} \in (z+x-\varepsilon, z+x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z+x+\varepsilon)-\varphi_{\mathrm{A}}(z+x-\varepsilon)\right)\mathrm{d}x \\
&\qquad (10) \\
&\quad + 2\int_{-z}^{+z} \left(\varphi_{\mathrm{B}}(x)-\varphi_{\mathrm{B}}(x-2z)\right)\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\left(\varphi_{\mathrm{A}}(z-x+\varepsilon)-\varphi_{\mathrm{A}}(z-x-\varepsilon)\right)\mathrm{d}x, \\
&\qquad (11)
\end{aligned}
$$

where in Eq. (7) we substituted $x' = -x$ and used the symmetry of the integrand functions, in Eqs. (8) and (9) we substituted $x' = x - 2z$ and $x' = 2z - x$, respectively, and used again the symmetry. The expression in Eq. (10) is negative as $\varphi_{\mathrm{B}}(x) > \varphi_{\mathrm{B}}(2z+x)$ and $\varphi_{\mathrm{A}}(z+x+\varepsilon) < \varphi_{\mathrm{A}}(z+x-\varepsilon)$ for $x \geq z$; the expression in Eq. (11) is negative as $\varphi_{\mathrm{B}}(x) > \varphi_{\mathrm{B}}(x-2z)$ and $\varphi_{\mathrm{A}}(z-x+\varepsilon) < \varphi_{\mathrm{A}}(z-x-\varepsilon)$ for $x \in (-z, z)$.

Third case: $z < 0$**.** This case is similar to the previous one: with the same arguments, we obtain

$$\frac{\mathrm{d}H(z)}{\mathrm{d}z}$$
$$= 2\int_{-\infty}^{+z} \big(\varphi_{\mathrm{B}}(x) - \varphi_{\mathrm{B}}(2z+x)\big) \Pr\left[\mathrm{A} \in (z+x-\varepsilon, z+x+\varepsilon)\right] \big(\varphi_{\mathrm{A}}(z+x+\varepsilon) - \varphi_{\mathrm{A}}(z+x-\varepsilon)\big)\,\mathrm{d}x \tag{12}$$
$$+ 2\int_{+z}^{-z} \big(\varphi_{\mathrm{B}}(x) - \varphi_{\mathrm{B}}(x-2z)\big) \Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right] \big(\varphi_{\mathrm{A}}(z-x+\varepsilon) - \varphi_{\mathrm{A}}(z-x-\varepsilon)\big)\,\mathrm{d}x. \tag{13}$$

The expression in Eq. (12) is positive as $\varphi_{\mathrm{B}}(x) > \varphi_{\mathrm{B}}(2z+x)$ and $\varphi_{\mathrm{A}}(z+x+\varepsilon) > \varphi_{\mathrm{A}}(z+x-\varepsilon)$ for $x \le z$; the expression in Eq. (13) is positive as $\varphi_{\mathrm{B}}(x) > \varphi_{\mathrm{B}}(x-2z)$ and $\varphi_{\mathrm{A}}(z-x+\varepsilon) < \varphi_{\mathrm{A}}(z-x-\varepsilon)$ for $x \in (z, -z)$. □

Claim 18. *For all* $x \in \mathbb{R}$, $c \in \left(0, \frac{1}{162}\right)$, *and* $\varepsilon \in (0,1)$, *it holds that*

$$\left(\int_{-\varepsilon}^{\varepsilon} e^{-c(x+s)^2}\,\mathrm{d}s\right)^2 \le \left(\int_{-\varepsilon}^{\varepsilon} \frac{e^{-c(x+\varepsilon)^2} + e^{-c(x-\varepsilon)^2}}{2} e^{c\varepsilon^2}\,\mathrm{d}s\right)^2.$$

Proof. Let

$$f_x(s) = e^{-c(x+s)^2}.$$

Since

$$\int_{-\varepsilon}^{\varepsilon} \frac{e^{-c(x+\varepsilon)^2} + e^{-c(x-\varepsilon)^2}}{2} e^{c\varepsilon^2}\,\mathrm{d}s = \int_{-\varepsilon}^{\varepsilon} ms + \frac{e^{-c(x+\varepsilon)^2} + e^{-c(x-\varepsilon)^2}}{2} e^{c\varepsilon^2}\,\mathrm{d}s$$

for any $m \in \mathbb{R}$, we choose it to be the angular coefficient of the line passing through $f_x(-\varepsilon)$ and $f_x(\varepsilon)$, and prove the stronger result

$$e^{-c(x+s)^2} \le \frac{e^{-c(x+\varepsilon)^2} - e^{-c(x-\varepsilon)^2}}{2\varepsilon} s + \frac{e^{-c(x+\varepsilon)^2} + e^{-c(x-\varepsilon)^2}}{2} e^{c\varepsilon^2} \tag{14}$$

for all $s \in (-\varepsilon, \varepsilon)$. In fact, the right hand side of Eq. (14) is the equation for the line passing by the extrema of f_x in $(-\varepsilon, \varepsilon)$ lifted by a factor of $e^{c\varepsilon^2}$. Therefore, the results holds trivially if f_x is convex in the entire range $(-\varepsilon, \varepsilon)$, which is true when $|x| > 1 + \frac{1}{\sqrt{2c}}$. Moreover, the factor $e^{c\varepsilon^2}$ ensures the result for $x = 0$, so, we follow with the analysis of the case $x \in \left(0, 1 + \frac{1}{\sqrt{2c}}\right]$ and the remaining case $x \in \left[-1 - \frac{1}{\sqrt{2c}}, 0\right)$ follows by symmetry.

Dividing both side of Eq. (14) by $e^{-c(x+s)^2}$, we obtain

$$1 \le e^{2csx+cs^2}\left[\frac{e^{-c\varepsilon^2}s}{\varepsilon} \cdot \frac{e^{-2c\varepsilon x} - e^{2c\varepsilon x}}{2} + \frac{e^{-2c\varepsilon x} + e^{2c\varepsilon x}}{2}\right] \tag{15}$$
$$= e^{2csx+cs^2}\left[-\frac{e^{-c\varepsilon^2}s}{\varepsilon}\sinh(2c\varepsilon x) + \cosh(2c\varepsilon x)\right].$$

Let $g(x)$ be the right hand side of this inequality. Then

$$\begin{aligned} g'(x) &= 2csg(x) + 2c\varepsilon e^{2csx+cs^2}\left[-\frac{e^{-c\varepsilon^2}s}{\varepsilon}\cosh(2c\varepsilon x) + \sinh(2c\varepsilon x)\right] \\ &= 2ce^{2csx+cs^2}\left[\cosh(2c\varepsilon x)\left(s - se^{-c\varepsilon^2}\right) + \sinh(2c\varepsilon x)\left(\varepsilon - \frac{s^2}{\varepsilon}e^{-c\varepsilon^2}\right)\right]. \end{aligned}$$

If $s \in [0,\varepsilon)$, then $s \geq se^{-c\varepsilon^2}$ and $\varepsilon \geq \frac{\varepsilon^2}{\varepsilon}e^{-c\varepsilon^2} \geq \frac{s^2}{\varepsilon}e^{-c\varepsilon^2}$, hence $g'(x) \geq 0$. Since $g(0) \geq 1$, this ensures Eq. (15).

The sub-case $s \in (-\varepsilon, 0)$ offers much more resistance. To analyse it we exploit that $x \in \left(0, 1 + \frac{1}{\sqrt{2c}}\right)$ implies that $cx \leq \sqrt{2c}$ for $c < \frac{1}{2}$ and make extensive use of Taylor's theorem to approximate the exponential functions.

We start by rewriting Eq. (15) as

$$\varepsilon e^{-2csx-cs^2} \leq e^{2c\varepsilon x}\left(\frac{\varepsilon}{2} - \frac{s}{2}e^{-c\varepsilon^2}\right) + e^{-2c\varepsilon x}\left(\frac{\varepsilon}{2} + \frac{s}{2}e^{-c\varepsilon^2}\right). \tag{16}$$

By Taylor's theorem, there exist $\lambda_1, \lambda_2 \in [0, 2c\varepsilon x] \subseteq [0, 2\sqrt{2c}\varepsilon]$, $\lambda_3 \in [0, -2csx] \subseteq [0, 2\sqrt{2c}\varepsilon]$, $\lambda_4 \in [0, cs^2]$, $\lambda_5 \in [0, c\varepsilon^2]$ such that

$$\begin{aligned} e^{+2c\varepsilon x} &= 1 + 2c\varepsilon x + 2c^2\varepsilon^2x^2 + \frac{4}{3}c^3\varepsilon^3x^3e^{\lambda_1}, \\ e^{-2c\varepsilon x} &= 1 - 2c\varepsilon x + 2c^2\varepsilon^2x^2 - \frac{4}{3}c^3\varepsilon^3x^3e^{\lambda_2}, \\ e^{-2csx} &= 1 - 2csx + 2c^2s^2x^2 - \frac{4}{3}c^3s^3x^3e^{\lambda_3}, \\ e^{-cs^2} &= 1 - cs^2 + \frac{c^2s^4}{2}e^{-\lambda_4}, \qquad (17) \\ e^{-c\varepsilon^2} &= 1 - c\varepsilon^2 + \frac{c^2\varepsilon^4}{2}e^{-\lambda_5}, \qquad (18) \end{aligned}$$

where we used second order approximations for the first three terms and first order approximations for the last two. Plugging those in Eq. (16) we obtain

$$\begin{aligned} \varepsilon e^{-cs^2}\left(1 - 2csx + 2c^2s^2x^2 - \frac{4}{3}c^3s^3x^3e^{\lambda_3}\right) \leq \left(1 + 2c\varepsilon x + 2c^2\varepsilon^2x^2 + \frac{4}{3}c^3\varepsilon^3x^3e^{\lambda_1}\right)\left(\frac{\varepsilon}{2} - \frac{s}{2}e^{-c\varepsilon^2}\right) \\ + \left(1 - 2c\varepsilon x + 2c^2\varepsilon^2x^2 - \frac{4}{3}c^3\varepsilon^3x^3e^{\lambda_2}\right)\left(\frac{\varepsilon}{2} + \frac{s}{2}e^{-c\varepsilon^2}\right). \end{aligned}$$

The latter becomes

$$\begin{aligned} &\varepsilon\left(1 - e^{-cs^2}\right) + 2cs\varepsilon x\left(e^{-cs^2} - e^{-c\varepsilon^2}\right) + 2c^2\varepsilon x^2\left(\varepsilon^2 - s^2e^{-cs^2}\right) \\ &+ \frac{4}{3}c^3\varepsilon x^3\left(\varepsilon^2e^{\lambda_1}\left(\frac{\varepsilon}{2} - \frac{s}{2}e^{-c\varepsilon^2}\right) - \varepsilon^2e^{-\lambda_2}\left(\frac{\varepsilon}{2} + \frac{s}{2}e^{-c\varepsilon^2}\right) + s^3e^{\lambda_3 - cs^2}\right) \geq 0 \end{aligned}$$

Now, notice that

$$\varepsilon^2e^{\lambda_1}\left(\frac{\varepsilon}{2} - \frac{s}{2}e^{-c\varepsilon^2}\right) - \varepsilon^2e^{-\lambda_2}\left(\frac{\varepsilon}{2} + \frac{s}{2}e^{-c\varepsilon^2}\right) \geq 0,$$

as $\frac{\varepsilon}{2} - \frac{s}{2}e^{-c\varepsilon^2} \geq \frac{\varepsilon}{2} + \frac{s}{2}e^{-c\varepsilon^2}$ since $-\varepsilon \leq s < 0$, and $\varepsilon^2 e^{\lambda_1} \geq \varepsilon^2 \geq \varepsilon^2 e^{-\lambda_2}$. Furthermore, observe that $s^3 e^{\lambda_3 - cs^2} \geq 2s^3$ as $s < 0$ and $\lambda_3 \leq 2\sqrt{2c}\varepsilon \leq \frac{1}{2}$ if $c \leq \frac{1}{32}$. Thus, the inequality is true if

$$\varepsilon\left(1 - e^{-cs^2}\right) + 2cs\varepsilon x\left(e^{-cs^2} - e^{-c\varepsilon^2}\right) + 2c^2\varepsilon x^2\left(\varepsilon^2 - s^2 e^{-cs^2}\right) + \frac{8}{3}c^3 s^3 \varepsilon x^3 \geq 0.$$

Applying Eqs. (17) and (18), the latter inequality yields that

$$\begin{aligned}
&\varepsilon\left(cs^2 - \frac{c^2 s^4}{2}e^{-\lambda_4}\right) + 2cs\varepsilon x\left(c\varepsilon^2 - cs^2 - \frac{c^2\varepsilon^4}{2}e^{-\lambda_5} + \frac{c^2 s^4}{2}e^{-\lambda_4}\right) \\
&\qquad + 2c^2\varepsilon x^2\left(\varepsilon^2 - s^2 + cs^4 - \frac{c^2 s^6}{2}e^{-\lambda_4}\right) + \frac{8}{3}c^3 s^3 \varepsilon x^3 \\
&\quad = \varepsilon c s^2 - \frac{c^2 s^4 \varepsilon}{2}e^{-\lambda_4} - c^3 s\varepsilon^5 x e^{-\lambda_5} + c^3 s^5 \varepsilon x e^{-\lambda_4} + \left(2c^3 s^4 \varepsilon x^2 - c^4 s^6 \varepsilon x^2 e^{-\lambda_4}\right) + \frac{8}{3}c^3 s^3 \varepsilon x^3 \\
&\qquad + 2c^2\varepsilon x\left(\varepsilon^2 - s^2\right)(x + s).
\end{aligned}$$

Now observe that

$$\left(2c^3 s^4 \varepsilon x^2 - c^4 s^6 \varepsilon x^2 e^{-\lambda_4}\right) \geq 0$$

as $c < 1, s \leq \varepsilon \leq 1, e^{-\lambda_4} < 1$; $-c^3 s\varepsilon^5 x e^{-\lambda_5} > 0$ as $s < 0$;

$$\begin{aligned}
\varepsilon c s^2 - \frac{c^2 s^4 \varepsilon}{2}e^{-\lambda_4} + c^3 s^5 \varepsilon x e^{-\lambda_4} + \frac{8}{3}c^3 s^3 \varepsilon x^3 &\geq cs^2\varepsilon - \frac{c^2 s^2 \varepsilon^3}{2} - c^2\sqrt{2c}s^2\varepsilon^4 - \frac{8}{3}c^3 s^2 \varepsilon^2 x^3 \\
&> cs^2\varepsilon - \frac{c^2 s^2 \varepsilon^3}{2} - c^2\sqrt{2c}s^2\varepsilon^4 - 6c\sqrt{2c}s^2\varepsilon^2 \qquad (19) \\
&= cs^2\varepsilon\left(1 - \frac{c\varepsilon^2}{2} - c^2\sqrt{2c}\varepsilon^3 - 6\sqrt{2c}\varepsilon\right) \\
&\geq cs^2\varepsilon\left(1 - \frac{c}{2} - c^2\sqrt{2c} - 6\sqrt{2c}\right) \qquad (20) \\
&\geq \frac{cs^2\varepsilon}{5}, \qquad (21)
\end{aligned}$$

where in Eq. (19) we used that $cx \leq \sqrt{2c}$, in Eq. (20) that $\varepsilon \leq 1$, and in Eq. (21) we used i) $c^2\sqrt{2c} \leq \frac{c}{2}$ when $c \leq \frac{1}{\sqrt[3]{2}}$, ii) $c < \sqrt{2c}$ since $c < 1$ and iii) $1 - 7\sqrt{2c} \geq \frac{1}{5}$, whenever $c \leq \frac{1}{162}$.

Going back to the inequality, we now have that

$$\frac{cs^2\varepsilon}{5} + 2c^2\varepsilon x\left(\varepsilon^2 - s^2\right)(x + s).$$

If $x \geq |s|$ the latter is positive, otherwise it becomes

$$\begin{aligned}
\frac{cs^2\varepsilon}{5} + 2c^2\varepsilon x\left(\varepsilon^2 - s^2\right)(x + s) &\geq \frac{cs^2\varepsilon}{5} + 2c^2\varepsilon x^2\left(\varepsilon^2 - s^2\right) - 2c^2\varepsilon s^2\left(\varepsilon^2 - s^2\right) \\
&\geq \frac{cs^2\varepsilon}{5} - 2c^2\varepsilon s^2 + 2c^2\varepsilon x^2\left(\varepsilon^2 - s^2\right) \\
&\geq cs^2\varepsilon\left(\frac{1}{5} - 2c\right),
\end{aligned}$$

which is positive for $c < \frac{1}{10}$. □

Claim 19. *For all $x \in \mathbb{R}$, $c \in \left(0, \frac{1}{10}\right)$, and $\varepsilon \in (0,1)$, it holds that*

$$\left(\int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-cy^2\right) \mathrm{d}y\right)^2 \geq \int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-c(x-\varepsilon)^2\right) \mathrm{d}y \cdot \int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-c(x+\varepsilon)^2\right) \mathrm{d}y. \tag{22}$$

Proof. We can express Eq. (22) as

$$\begin{aligned}
&\left[\int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-cy^2\right) \mathrm{d}y\right]^2 - \left[\int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-c(x^2+\varepsilon^2)\right) \mathrm{d}y\right]^2 \\
&\quad = \left[\int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-cy^2\right) - \exp\left(-c(x^2+\varepsilon^2)\right) \mathrm{d}y\right] \cdot \left[\int_{x-\varepsilon}^{x+\varepsilon} \exp\left(-cy^2\right) + \exp\left(-c(x^2+\varepsilon^2)\right) \mathrm{d}y\right] \\
&\quad \geq 0,
\end{aligned}$$

which holds if and only if

$$\int_{-\varepsilon}^{+\varepsilon} \exp\left(-c(x+s)^2\right) \mathrm{d}s \geq \int_{-\varepsilon}^{+\varepsilon} \exp\left(-c(x^2+\varepsilon^2)\right) \mathrm{d}s. \tag{23}$$

The result is immediate for $x = 0$, so we assume $x > 0$ and the claim follows by symmetry. Let

$$f_x(s) = \exp(-c(x+s)^2).$$

We provide distinct arguments depending on whether x is small or large.

Case $x \in (0,1)$. Since we assume $c < \frac{1}{8}$ and $\varepsilon < 1$, we have for any $x \leq 1$ that f_x is concave in $(-\varepsilon, \varepsilon)$. That is,

$$f_x(s) \geq \frac{f_x(\varepsilon) - f_x(-\varepsilon)}{2\varepsilon} s + \frac{f_x(\varepsilon) + f_x(\varepsilon)}{2},$$

for all $s \in (-\varepsilon, \varepsilon)$. Thus,

$$\begin{aligned}
\int_{-\varepsilon}^{\varepsilon} f_x(s) \, \mathrm{d}s &\geq \int_{-\varepsilon}^{\varepsilon} \frac{f_x(\varepsilon) - f_x(-\varepsilon)}{2\varepsilon} s + \frac{f_x(\varepsilon) + f_x(-\varepsilon)}{2} \, \mathrm{d}s \\
&= \int_{-\varepsilon}^{\varepsilon} \frac{f_x(\varepsilon) + f_x(-\varepsilon)}{2} \, \mathrm{d}s \\
&= \int_{-\varepsilon}^{\varepsilon} \exp\left(-c(x^2+\varepsilon^2)\right) \cdot \frac{\exp\left(-2cx\varepsilon\right) + \exp\left(2cx\varepsilon\right)}{2} \, \mathrm{d}s \\
&\geq \int_{-\varepsilon}^{\varepsilon} \exp\left(-c(x^2+\varepsilon^2)\right) \mathrm{d}s.
\end{aligned}$$

Case $x \geq 1$. The integral on the right hand side of Eq. (23) has the same value for any affine integrand r_x for which $r_x(0) = \exp\left(-c(x^2+\varepsilon^2)\right)$. Thus, proving that $f_x(s) \geq r_x(s)$, for all $s \in (-\varepsilon, \varepsilon)$, concludes the proof.

In particular, we can choose

$$r_x(s) = f_x'(0) \cdot s + \exp\left(-c(x^2+\varepsilon^2)\right).$$

Since

$$f'_x(s) = -2c(x+s)\exp\left(-c(x+s)^2\right),$$

we aim to show that

$$\exp\left(-c(x+s)^2\right) \geq -2csx\exp\left(-cx^2\right) + \exp\left(-c(x^2+\varepsilon^2)\right)$$

for $s \in (-\varepsilon, \varepsilon)$. Dividing by $\exp(-c(x^2+s^2))$ and rearranging, we obtain

$$\exp(-2csx) + 2csx\exp\left(cs^2\right) - \exp\left(-c\left(\varepsilon^2 - s^2\right)\right) \geq 0. \tag{24}$$

Now, if $s \geq 0$, we have that

$$\begin{aligned}\exp(-2csx) + 2csx\exp(cs^2) - \exp(-c(\varepsilon^2 - s^2)) &\geq 1 - 2csx + 2csx(1+cs^2) - \exp(-c\varepsilon^2) \qquad (25)\\ &= 1 + 2c^2s^3x - \exp(-c\varepsilon^2)\\ &\geq 2c^2s^3x\\ &\geq 0,\end{aligned}$$

where in Eq. (25) we used that $e^y \geq 1+y$.

Now consider the sub-case $s < 0$. By Taylor's theorem,

$$\exp(y) = 1 + y + \frac{y^2}{2} + \frac{\exp(\xi_1)\cdot y^3}{6}$$

and

$$\exp(y) = 1 + y + \frac{\exp(\xi_2)\cdot y^2}{2},$$

for some $\xi_1, \xi_2 \in [0, y]$. Letting $\ell = -s \in (0,1)$, we have

$$\exp(2c\ell x) \geq 1 + 2c\ell x + \frac{(2c\ell x)^2}{2} + \frac{(2c\ell x)^3}{6}$$

and

$$\begin{aligned}\exp(c\ell^2) &\leq 1 + c\ell^2 + \frac{\exp(c\ell^2)(c\ell^2)^2}{2}\\ &\leq 1 + c\ell^2 + \frac{\left(1+3(c\ell^2)\right)(c\ell^2)^2}{2}.\end{aligned}$$

since $e^y \leq 1+3y$ for $0 \leq y \leq 1$. Finally, applying this to Eq. (24), we have

$$\begin{aligned}&\exp(-2csx) + 2csx\exp\left(cs^2\right) - \exp\left(-c\left(\varepsilon^2 - s^2\right)\right)\\ &\quad\geq \exp(2c\ell x) - 2c\ell x\exp\left(c\ell^2\right) - 1\\ &\quad\geq 1 + 2c\ell x + \frac{(2c\ell x)^2}{2} + \frac{(2c\ell x)^3}{6} - 2c\ell x\left(1 + c\ell^2 + \frac{c^2\ell^4(1+3c\ell^2)}{2}\right) - 1\\ &\quad= 2c^2\ell^2x^2 + \frac{4}{3}c^3\ell^3x^3 - 2c\ell x\left(c\ell^2 + \frac{c^2\ell^4(1+3c\ell^2)}{2}\right)\\ &\quad= 2c^2\ell^2x(x-\ell) + c^3\ell^3x\left(\frac{4}{3}x^2 - \ell^2(1+3c\ell^2)\right).\end{aligned}$$

The latter is non negative for $x \geq 1$ and $c \leq \frac{1}{9}$, since $\ell = -s \leq \varepsilon < 1$, so that $\frac{4}{3}x^2 - \ell^2(1 + 3c\ell^2) \geq \frac{4}{3} - 1 - \frac{1}{3} = 0$. □

B Proofs omitted

B.1 Proof of Lemma 3

If $\alpha \geq \frac{1}{2}$, then $2\alpha^2 n \geq \alpha n$, and the result holds trivially. So we assume $\alpha < \frac{1}{2}$.

Let k be any integer, and let $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$, with each C_i drawn uniformly from the collection of subsets of $[n]$ with size αn. Given $i, j \in [k]$, if $i \neq j$, then

$$\begin{aligned}
\mathbb{E}\left[|C_i \cap C_j|\right] &= \sum_{a \in [n]} \Pr\left[a \in C_i \cap C_j\right] \\
&= \sum_{a \in [n]} \Pr\left[a \in C_i\right] \cdot \Pr\left[a \in C_j\right] \\
&= \alpha^2 n.
\end{aligned}$$

By the multiplicative form of Chernoff bounds (see Lemma 14), it holds that

$$\Pr\left[|C_i \cap C_j| > 2\alpha^2 n\right] \leq \exp\left(-\frac{\alpha^2 n}{3}\right).$$

Finally, for the event of interest, we have that

$$\begin{aligned}
\Pr\left[\bigcap_{i \neq j \in [k]} \left\{|C_i \cap C_j| \leq 2\alpha^2 n\right\}\right] &= 1 - \Pr\left[\bigcup_{i \neq j \in [k]} \left\{|C_i \cap C_j| > 2\alpha^2 n\right\}\right] \\
&\geq 1 - \binom{k}{2} \exp\left(-\frac{\alpha^2 n}{3}\right) \\
&\geq 1 - 2^{\frac{\alpha^2 t}{3}} \cdot \exp\left(-\frac{\alpha^2 n}{3}\right) \qquad (26) \\
&> 0,
\end{aligned}$$

where in Eq. (26) we have chosen $k = 2^{\frac{\alpha^2 n}{6}}$.

B.2 Proof of Lemma 4

By the distribution of $\mathbf{X}$,

$$\Pr\left[\mathbf{X} \in \mathcal{B}_\infty^d(z, \varepsilon)\right] = \int_{\mathcal{B}_\infty^d(z,\varepsilon)} \frac{1}{\left(2\pi\sigma^2\right)^{\frac{d}{2}}} \cdot \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right) \mathrm{d}x.$$

Since $\mathcal{B}_\infty^d(z, \varepsilon) \subseteq \mathcal{B}_\infty^d(\mathbf{0}, 2)$ and for all $x \in \mathbb{R}^d$ it holds that $\|x\|_2 \leq \sqrt{d} \cdot \|x\|_\infty$, and, thus,

$$\exp\left(-\frac{2d}{\sigma^2}\right) \leq \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right) \leq 1.$$

The thesis follows by noting that the hypercube $\mathcal{B}_\infty^d(z, \varepsilon)$ has measure $(2\varepsilon)^d$.

B.3 Proof of Lemma 11

Inheriting the setup from the proof of Lemma 6 and proceeding analogously we obtain that $\sigma_{\mathrm{A}}^2 = \alpha n\left(1-\frac{\alpha}{2}\right)$ and $\sigma_{\mathrm{B}}^2 = \frac{\alpha^2 n}{2}$. We diverge from that argument after Eq. (1). Preserving equality for a bit longer, we have that

$$\begin{aligned}\left(\Pr\left[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1\right]\right)^{\frac{1}{d}} &= \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \left(\Pr\left[\mathrm{A} \in (z-x-\varepsilon, z-x+\varepsilon)\right]\right)^2 \mathrm{d}x \\ &= \int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \left(\int_{z-x-\varepsilon}^{z-x+\varepsilon} \varphi_{\mathrm{A}}(y)\,\mathrm{d}y\right)^2 \mathrm{d}x.\end{aligned}$$

The hypothesis on n implies that $2\sigma_a^2 \geq 10$, so, by Claim 19,

$$\begin{aligned}\left(\int_{z-x-\varepsilon}^{z-x+\varepsilon} \varphi_{\mathrm{A}}(y)\,\mathrm{d}y\right)^2 &\geq (2\varepsilon)^2 \cdot \varphi_{\mathrm{A}}(z-x-\varepsilon) \cdot \varphi_{\mathrm{A}}(z-x+\varepsilon) \\ &= \frac{(2\varepsilon)^2}{2\pi\sigma_{\mathrm{A}}^2} \cdot \exp\left(-\frac{(z-x-\varepsilon)^2}{2\sigma_{\mathrm{A}}^2}\right) \cdot \exp\left(-\frac{(z-x+\varepsilon)^2}{2\sigma_{\mathrm{A}}^2}\right) \\ &= e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \frac{1}{\sqrt{2}} \cdot \frac{(2\varepsilon)^2}{\sqrt{2\pi\sigma_{\mathrm{A}}^2}} \cdot \frac{1}{\sqrt{\pi\sigma_{\mathrm{A}}^2}} \cdot \exp\left(-\frac{(z-x)^2}{\sigma_{\mathrm{A}}^2}\right) \\ &= e^{-\varepsilon^2/\sigma_{\mathrm{A}}^2} \cdot \frac{1}{\sqrt{2}} \cdot \frac{(2\varepsilon)^2}{\sqrt{2\pi\sigma_{\mathrm{A}}^2}} \cdot \varphi_{\mathrm{A}/\sqrt{2}}(z-x).\end{aligned}$$

Then, as before, we can reduce the main integral to a convolution. Namely, it holds that

$$\begin{aligned}\int_{\mathbb{R}} \varphi_{\mathrm{B}}(x) \cdot \varphi_{\mathrm{A}/\sqrt{2}}(z-x)\,\mathrm{d}x &= \varphi_{\mathrm{B+A}/\sqrt{2}}(z) \\ &= \frac{1}{\sqrt{2\pi\sigma_{\mathrm{B+A}/\sqrt{2}}^2}} \cdot \exp\left(-\frac{z^2}{2\sigma_{\mathrm{B+A}/\sqrt{2}}^2}\right).\end{aligned}$$

Altogether, we have that

$$\begin{aligned}\left(\Pr\left[\mathrm{Y}_S = 1, \mathrm{Y}_T = 1\right]\right)^{\frac{1}{d}} &\geq \frac{(2\varepsilon)^2}{2\pi} \cdot \frac{1}{\sqrt{2\sigma_{\mathrm{A}}^2\sigma_{\mathrm{B+A}/\sqrt{2}}^2}} \cdot \exp\left(-\frac{\varepsilon^2}{\sigma_{\mathrm{A}}^2} - \frac{z^2}{2\sigma_{\mathrm{B+A}/\sqrt{2}}^2}\right) \\ &= \frac{(2\varepsilon)^2}{2\pi\alpha n} \cdot \frac{1}{\sqrt{1-\frac{\alpha^2}{4}}} \cdot \exp\left(-\frac{1}{\alpha n} \cdot \left(\frac{2\varepsilon^2}{2-\alpha} + \frac{2z^2}{2+\alpha}\right)\right).\end{aligned}$$

where the last equality follows from recalling that $\sigma_{\mathrm{B}}^2 = \frac{\alpha^2 n}{2}$ and $\sigma_{\mathrm{A}}^2 = \alpha n\left(1-\frac{\alpha}{2}\right)$, which implies that $\sigma_{\mathrm{B+A}/\sqrt{2}}^2 = \frac{\alpha^2 n}{2} + \frac{\alpha n}{2}\left(1-\frac{\alpha}{2}\right)$. Finally, the hypotheses $z \in [-1,1]$, $\varepsilon \in (0,1)$, and $\alpha \in \left(0, \frac{1}{2}\right)$ imply that $\frac{2\varepsilon^2}{2-\alpha} + \frac{2z^2}{2+\alpha} < 3$.

B.4 Proof of Theorem 9

Let $n = k \cdot \frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)$ with $k \in \mathbb{N}$. By Lemma 8, for any $z \in [-1,1]^d$, the probability than no subset-sum is sufficiently close to z is at most $\left(\frac{2}{3}\right)^k$. Leveraging the fact that it is possible to cover $[-1,1]^d$ by $\frac{1}{\varepsilon^d}$ hypercubes of radius ε, we can ensure that the probability of failing to approximate any z is, by the union bound, at most

$$\begin{aligned}\frac{1}{\varepsilon^d}\cdot\left(\frac{2}{3}\right)^k &= 2^{-k\log\frac{3}{2}+d\log\frac{1}{\varepsilon}} \\ &= \exp\left[-\ln 2\cdot\frac{n-\frac{144d^2}{\alpha^2\log\frac{3}{2}}\log\frac{1}{\varepsilon}\left(\log\frac{1}{\varepsilon}+\log d+\log\frac{1}{\alpha}\right)}{\frac{144d}{\alpha^2\log\frac{3}{2}}\left(\log\frac{1}{\varepsilon}+\log d+\log\frac{1}{\alpha}\right)}\right].\end{aligned}$$

Thus, we can conclude the result for

$$n \geq \frac{144}{\log\frac{3}{2}}\cdot\frac{d^2}{\alpha^2}\log\frac{1}{\varepsilon}\cdot\left(\log\frac{1}{\varepsilon}+\log d+\log\frac{1}{\alpha}\right).$$

C Generalisation of our result

If the target value z lies in the hypercube $[-\lambda\sqrt{n},\lambda\sqrt{n}]^d$, for some $\lambda > \frac{1}{\sqrt{n}}$, we have slightly different bounds for the expectation and for the variance of Y. In particular, Corollary 5 would give

$$e^{-\frac{2\lambda^2 d}{\alpha}}\frac{(2\varepsilon)^d|\mathcal{C}|}{(2\pi\alpha n)^{\frac{d}{2}}} \leq \mathbb{E}[\mathrm{Y}] \leq \frac{(2\varepsilon)^d|\mathcal{C}|}{(2\pi\alpha n)^{\frac{d}{2}}}. \tag{27}$$

On the other hand, as the proof of Lemma 6 never uses that $z \in [-1,1]^d$ but only exploits the bound on the expectation, it would yield

$$\mathrm{Var}[\mathrm{Y}] \leq \frac{(2\varepsilon)^{2d}|\mathcal{C}|^2}{(2\pi\alpha n)^d}\left[(1-4\alpha^2)^{-\frac{d}{2}-e^{-\frac{4\lambda^2 d}{\alpha}}}\right] + \frac{(2\varepsilon)^d|\mathcal{C}|}{(2\pi\alpha n)^{\frac{d}{2}}}. \tag{28}$$

We focus on the case $\lambda = \frac{1}{2}\sqrt{\frac{\alpha}{17d}}$ when $n > \frac{68d}{\alpha}$ (which implies $\lambda\sqrt{n} > 1$). Thus, we have a new estimation for the probability to hit a single value.

Lemma 20. *Given $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\alpha \in (0, \frac{1}{6}]$, let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ i.d.d. following $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$, $z \in [-\lambda\sqrt{n},\lambda\sqrt{n}]^d$, with $\lambda = \frac{1}{2}\sqrt{\frac{\alpha}{17d}}$, and $\mathcal{C} \subseteq \binom{[n]}{\alpha n}$. If any two subsets in $\mathcal{C}$ intersect in at most $2\alpha^2 n$ elements, $\alpha \leq \frac{1}{6\sqrt{d}}$, and*

$$n \geq \frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon}+\log d+\log\frac{1}{\alpha}\right),$$

then

$$\Pr[\mathrm{Y} \geq 1] \geq \frac{1}{3}.$$

Proof. By Chebyshev's inequality, it holds that

$$\begin{aligned}\Pr\left[\mathrm{Y} \geq 1\right] &\geq \Pr\left[\left|\mathrm{Y} - \mathbb{E}\left[\mathrm{Y}\right]\right| < \frac{\mathbb{E}\left[\mathrm{Y}\right]}{2}\right] \\ &\geq 1 - \frac{4 \cdot \mathrm{Var}\left[\mathrm{Y}\right]}{\mathbb{E}\left[\mathrm{Y}\right]^2}.\end{aligned}$$

Notice that $\frac{4\lambda^2 d}{\alpha} = \frac{1}{17}$. Hence, using Eqs. (27) and (28), we get that

$$\begin{aligned}\frac{4 \cdot \mathrm{Var}\left[\mathrm{Y}\right]}{\mathbb{E}\left[\mathrm{Y}\right]^2} &\leq 4 \cdot \frac{e^{\frac{1}{17}} \cdot (2\pi\alpha n)^d}{(2\varepsilon)^{2d} \left|\mathcal{C}\right|^2} \cdot \left[\frac{(2\varepsilon)^{2d} \left|\mathcal{C}\right|^2}{(2\pi\alpha n)^d} \cdot \left[(1-4\alpha^2)^{-\frac{d}{2}} - e^{-\frac{1}{17}}\right] + \frac{(2\varepsilon)^d \left|\mathcal{C}\right|}{(2\pi\alpha n)^{\frac{d}{2}}}\right] \\ &= 4 \cdot \left(\frac{e^{\frac{1}{17}}}{(1-4\alpha^2)^{\frac{d}{2}}} - 1\right) + \frac{4e^{\frac{1}{17}} \cdot (2\pi\alpha n)^{\frac{d}{2}}}{(2\varepsilon)^d \left|\mathcal{C}\right|}.\end{aligned}$$

Note that Claim 15 holds exactly as it is for the ratio

$$\frac{e^{\frac{1}{17}}}{(1-4\alpha^2)^{\frac{d}{2}}}$$

obtaining the same bound for $n \geq \frac{68d}{\alpha}$ and $\alpha \leq \frac{1}{6\sqrt{d}}$, which yields

$$4 \cdot \left(\frac{e^{\frac{1}{17}}}{(1-4\alpha^2)^{\frac{d}{2}}} - 1\right) \leq \frac{1}{2}.$$

Furthermore, also Claim 16 is true replacing $e^{\frac{4d}{\alpha n}}$ by $e^{\frac{1}{17}}$. Thus, as $n \geq \frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)$ and $\alpha \leq \frac{1}{6}$, Claim 16 implies that

$$\frac{4e^{\frac{1}{17}} \cdot (2\pi\alpha n)^{\frac{d}{2}}}{(2\varepsilon)^d \left|\mathcal{C}\right|} \leq \varepsilon.$$

□

We remark that we cannot let λ be asymptotically greater than $\sqrt{\frac{\alpha}{d}}$ otherwise our method fails. Indeed, by Remark 12, the term $\frac{4\mathrm{Var}[\mathrm{Y}]}{\mathbb{E}[\mathrm{Y}]^2}$ is at least

$$4 \cdot \left(\frac{e^{\frac{4\lambda^2 d}{\alpha} - \frac{3\lambda^2 d}{\alpha}}}{\left(1 - \frac{\alpha^2}{4}\right)^{\frac{d}{2}}} - 1\right).$$

The latter is greater than or equal to 1 if $\lambda \geq \sqrt{\frac{\alpha}{d}}$ since $e^{\frac{\lambda^2 d}{\alpha}} \geq 1 + \frac{\lambda^2 d}{\alpha}$.

We are ready to state our first generalised version of Theorem 9.

Theorem 21. *For given d and $\varepsilon \in (0,1)$, let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n independent standard normal d-dimensional random vectors and let $\alpha \in (0, \frac{1}{6\sqrt{d}}]$. There exist two universal constants $C > \delta > 0$ such that, if*

$$n \geq C\frac{d^2}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)^2,$$

the following holds with probability at least

$$1 - \exp\left[-\ln 2 \cdot \left(\frac{n}{\delta\frac{d}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)} - d\log\frac{1}{\varepsilon}\right)\right]:$$

for all $\boldsymbol{z} \in \left[-\lambda\sqrt{n}, \lambda\sqrt{n}\right]^d$, with $\lambda = \frac{1}{2}\sqrt{\frac{\alpha}{17d}}$, there exists a subset $S_{\boldsymbol{z}} \subseteq [n]$, such that

$$\left\|\boldsymbol{z} - \sum_{i \in S_{\boldsymbol{z}}} \mathbf{X}_i\right\|_\infty \leq \varepsilon.$$

Moreover, the property above remains true even if we restrict to subsets of size αn.

Proof. Let $\frac{n}{\frac{144d}{\alpha^2}\left(\log\frac{1}{\varepsilon}+\log d+\log\frac{1}{\alpha}\right)} = k \geq 1$ with $k \in \mathbb{N}$. By Lemma 20, for any $\boldsymbol{z} \in [-\lambda\sqrt{n}, \lambda\sqrt{n}]^d$, the probability than no subset-sum is sufficiently close to $\boldsymbol{z}$ is at most $\left(\frac{2}{3}\right)^k$. Leveraging the fact that it is possible to cover $[-\lambda\sqrt{n}, \lambda\sqrt{n}]^d$ by $\left(\frac{\lambda\sqrt{n}}{\varepsilon}\right)^d$ hypercubes of radius ε, we can ensure that the probability of failing to approximate any $\boldsymbol{z}$ is, by the union bound, at most

$$\begin{aligned}
\left(\frac{\lambda\sqrt{n}}{\varepsilon}\right)^d \cdot \left(\frac{2}{3}\right)^k &= 2^{-k\log\frac{3}{2} + d\left(\log\frac{1}{\varepsilon} + \frac{1}{2}\log n + \log\lambda\right)} \\
&= \exp\left[-\ln 2 \cdot \frac{n - \frac{144d^2}{\alpha^2\log\frac{3}{2}}\left(\log\frac{1}{\varepsilon} + \frac{1}{2}\log n + \log\lambda\right)\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)}{\frac{144d}{\alpha^2\log\frac{3}{2}}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)}\right] \\
&\leq \exp\left[-\ln 2 \cdot \frac{n - \frac{144d^2}{\alpha^2\log\frac{3}{2}}\left(\log\frac{1}{\varepsilon} + \frac{1}{2}\log n\right)\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)}{\frac{144d}{\alpha^2\log\frac{3}{2}}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)}\right] \qquad (29)
\end{aligned}$$

since $\lambda < 1$. Consider $\frac{n}{2} - \frac{144d^2}{2\alpha^2\log\frac{3}{2}}\log n\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)$. Let $k = k'\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)$,

which means that $n = \frac{144k'd}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)^2$. Then

$$\begin{aligned}
&\frac{n}{2} - \frac{144d^2}{2\alpha^2\log\frac{3}{2}}\log n\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\\
&\quad= \frac{144d}{2\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\left[k'\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\right.\\
&\qquad\left.-\frac{d}{\log\frac{3}{2}}\left(\log\frac{144}{\log\frac{3}{2}} + \log k' + \log d + 2\log\frac{1}{\alpha} + 2\log\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\right)\right]\\
&\quad\geq \frac{144d}{2\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\left[k'\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\right.\\
&\qquad\left.-2d\left(8 + \log k' + \log d + 2\log\frac{1}{\alpha} + 2\log\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\right)\right]
\end{aligned}$$

If $k' = 17d$, we have that

$$\begin{aligned}
&k'\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right) - 2d\left(8 + \log k' + \log d + 2\log\frac{1}{\alpha} + 2\log\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\right)\\
&\quad\geq 4d\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha} - \log\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)\right) + 13d\log d + 13d\log\frac{1}{\alpha}\\
&\qquad- 16d - 2d\log c - 3d\log d - 4d\log\frac{1}{\alpha}\\
&\quad= 10d\log d + 9d\log\frac{1}{\alpha} - 16d - 2d\log 17 \geq 0,
\end{aligned}$$

as $\alpha \leq \frac{1}{6}$. Thus, for $n \geq \frac{17\cdot 144d^2}{\alpha^2}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)^2$, we have that the expression in Eq. (29) is at most

$$\exp\left[-\ln 2\cdot\frac{n - \frac{288d^2}{\alpha^2\log\frac{3}{2}}\log\frac{1}{\varepsilon}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)}{\frac{288d}{\alpha^2\log\frac{3}{2}}\left(\log\frac{1}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)}\right].$$

We have the thesis by setting $\delta = \frac{288}{\log\frac{3}{2}}$ and $C = 17\cdot 144$. □

Our analysis, that relies on fixed subset sizes, easily extends Theorem 21 for non-centred and non-unitary normal random vectors.

Corollary 22. *Let $\sigma > 0$ and $\varepsilon \in (0, \sigma)$. Given $d, n \in \mathbb{N}$ let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent normal d-dimensional random vectors with $\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{v}, \sigma^2 \cdot \boldsymbol{I}_d)$, for any vector $\boldsymbol{v} \in \mathbb{R}^d$. Furthermore, let $\alpha \in \left(0, \frac{1}{6\sqrt{d}}\right)$. There exist two universal constants $C > \delta > 0$ such that, if*

$$n \geq C\frac{d^2}{\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)^2,$$

then, with probability

$$1 - \exp\left[-\ln 2 \cdot \left(\frac{n}{\delta \frac{d}{\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)} - d\log\frac{\sigma}{\varepsilon}\right)\right],$$

for all $\boldsymbol{z} \in [-\sigma\lambda\sqrt{n}, \sigma\lambda\sqrt{n}]^d + \alpha n \boldsymbol{v}$*, with* $\lambda = \frac{1}{2}\sqrt{\frac{\alpha}{17d}}$*, there exists a subset* $S_{\boldsymbol{z}} \subseteq [n]$ *for which*

$$\left\|\boldsymbol{z} - \sum_{i \in S_{\boldsymbol{z}}} \mathbf{X}_i\right\|_\infty \leq \varepsilon.$$

Moreover, this remains true even when restricted to subsets of size αn*.*

Proof. Simply apply Theorem 21 to the random vectors $\frac{\mathbf{X}_i - \boldsymbol{v}}{\sigma}$ with error $\frac{\varepsilon}{\sigma}$. □

Following the line of [Lue98], we also observe that our results extend to a wider class of probability distributions.

Definition 23. *Consider any two random variables* X *and* Y *having the same codomain, and let* $\varphi_{\mathrm{X}}(\boldsymbol{x}), \varphi_{\mathrm{Y}}(\boldsymbol{x})$ *be their probability density functions. We say that* X contains Y *with probability* $\boldsymbol{p}$ *if a constant* $\boldsymbol{p} \in (0, 1]$ *exists such that* $\varphi_{\mathrm{X}}(\boldsymbol{x}) = \boldsymbol{p} \cdot \varphi_{\mathrm{Y}}(\boldsymbol{x}) + (1 - \boldsymbol{p})\boldsymbol{f}(\boldsymbol{x})$ *for any function* $\boldsymbol{f}(\boldsymbol{x})$*.*

If X contains Y with probability $\boldsymbol{p}$, we can describe the behaviour of X as follows: with probability $\boldsymbol{p}$, draw Y; with probability $1 - \boldsymbol{p}$, draw something else. An adapted version of our result holds for random variables containing Gaussian distributions.

Corollary 24. *Let* $\sigma > 0$*,* $\varepsilon \in (0, \sigma)$*, and let* $\boldsymbol{p} \in (0, 1]$ *be a constant. Given* $\boldsymbol{d}, \boldsymbol{n} \in \mathbb{N}$ *let* $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ *be independent* $\boldsymbol{d}$*-dimensional random vectors containing* $\boldsymbol{d}$*-dimensional normal random vectors* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{v}, \sigma^2 \cdot \boldsymbol{I}_d)$ *with probability* $\boldsymbol{p}$*, where* $\boldsymbol{v}$ *is any vector in* $\mathbb{R}^d$ *. Furthermore, let* $\boldsymbol{\alpha} \in \left(0, \frac{1}{6\sqrt{d}}\right)$*. There exist two universal constants* $C > \delta > 0$ *such that, if*

$$n \geq 2C\frac{d^2}{p\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)^2,$$

then, with probability

$$1 - 2\exp\left[-\ln 2 \cdot \left(\frac{pn}{2\delta \frac{d}{\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)} - d\log\frac{\sigma}{\varepsilon}\right)\right],$$

for all $\boldsymbol{z} \in \left[-\sigma\lambda\sqrt{\frac{pn}{2}}, \sigma\lambda\sqrt{\frac{pn}{2}}\right]^d + \frac{\alpha pn}{2}\boldsymbol{v}$*, with* $\lambda = \frac{1}{2}\sqrt{\frac{\alpha}{17d}}$*, there exists a subset* $S_{\boldsymbol{z}} \subseteq [n]$ *for which*

$$\left\|\boldsymbol{z} - \sum_{i \in S_{\boldsymbol{z}}} \mathbf{X}_i\right\|_\infty \leq \varepsilon.$$

Moreover, this remains true even when restricted to subsets of size $\frac{\alpha pn}{2}$*.*

Proof. With a simple application of the Chernoff bound, we have that at least $\frac{pn}{2}$ random vectors are normal random vectors with probability $1 - e^{-\frac{pn}{8}}$. Conditional on this event, we can apply Corollary 22 to the $\frac{pn}{2}$ normal random vectors. Since $\Pr[A, B] \geq \Pr[A \mid B] \Pr[B]$ for any two events A, B, and $2\delta\frac{d}{\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right) \geq 8$, the thesis holds with probability at least

$$1 - \exp\left[-\ln 2 \cdot \left(\frac{pn}{2\delta\frac{d}{\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)} - d\log\frac{\sigma}{\varepsilon}\right)\right] - \exp\left[-\frac{pn}{8}\right]$$

$$\geq 1 - 2\exp\left[-\ln 2 \cdot \left(\frac{pn}{2\delta\frac{d}{\alpha^2}\left(\log\frac{\sigma}{\varepsilon} + \log d + \log\frac{1}{\alpha}\right)} - d\log\frac{\sigma}{\varepsilon}\right)\right].$$

□

D Discrete setting

We believe that it should not be hard to adapt our proof to several discrete distributions, in order to obtain results similar to those discussed in the Related Work section. We also note that our Theorem 2 already implies an analogous discrete result. Suppose that we quantise our random vectors by truncating them to the $\lfloor\log\frac{1}{\delta}\rfloor$-th binary place, obtaining vectors $\hat{\mathbf{X}}_i$ such that $\|\hat{\mathbf{X}}_i - \mathbf{X}_i\|_\infty < \delta$. For any $\boldsymbol{z} \in [-1, 1]^d$, Theorem 2 guarantees that w.h.p. there is a subset of indices $I \subseteq [n]$ such that $\|\boldsymbol{z} - \sum_{i\in I}\mathbf{X}_i\|_\infty < \varepsilon$ and, hence, by the triangular inequality, $\|\boldsymbol{z} - \sum_{i\in I}\hat{\mathbf{X}}_i\|_\infty < n\delta + \varepsilon$. As a special case ($\delta = \varepsilon$), we have the following:

Corollary 25 (Discretization of Theorem 2)**.** *Given $d \in \mathbb{N}$, $\varepsilon \in (0, 1)$, let $\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_n$ be independent standard normal d-dimensional vectors truncated to the $\lfloor\log\frac{1}{\varepsilon}\rfloor$-th binary place. There exists a universal constant $C > 0$ such that, if $n \geq Cd^3\log\frac{1}{\varepsilon}\left(\log\frac{1}{\varepsilon} + \log d\right)$, then, with high probability, for all vectors $\hat{\boldsymbol{z}}$ with entries in $\{k\varepsilon\}_{\lceil-\frac{1}{\varepsilon}\rceil \leq k \leq \lfloor\frac{1}{\varepsilon}\rfloor}$ there exists a subset $S_{\boldsymbol{z}} \subseteq [n]$ for which*

$$\left\|\hat{\boldsymbol{z}} - \sum_{i\in S_{\boldsymbol{z}}}\hat{\mathbf{X}}_i\right\|_\infty \leq \varepsilon(n+1).$$

Moreover, the approximation can be achieved with subsets of size $\frac{n}{6\sqrt{d}}$.

E Connection with non-deterministic random walks

Consider a discrete-time stochastic process whose state space is $\mathbb{R}^d$ which starts at the origin. At the first step, the process "branches" in two processes, one of which keeps still, while the other moves by the vector $\mathbf{X}_1$. Recursively, given any time i and any process, at the next time step the process branches in two other processes, one of which keeps still, while the other moves by the vector $\mathbf{X}_{i+1}$. In this setting, when $\mathbf{X}_{i+1}$ are sampled from a standard multivariate normal distribution, our results imply that the resulting process is space filling: the process eventually gets arbitrarily close to each point in $\mathbb{R}^d$. This should be contrasted with the fact that a Brownian motion is transient in dimension $d \geq 3$ [MP10]. The above process can also be interpreted as a multi-dimensional version of nondeterministic walks as introduced in [PLW19] in the context of the analysis of encapsulations and decapsulations of network protocols, where the i-th N-step is $\{\mathbf{X}_i, \vec{0}\}$.