Revisiting the Random Subset Sum problem

Arthur da Cunha, Francesco d'Amore, Frédéric Giroire, Hicham Lesfari, Emanuele Natale, Laurent Viennot

Revisiting the Random Subset Sum problem

Arthur Carvalho Walraven Da Cunha
Université Côte d'Azur, Inria Sophia Antipolis, CNRS

Francesco D'Amore
Université Côte d'Azur, Inria Sophia Antipolis, CNRS

Frédéric Giroire
Université Côte d'Azur, Inria Sophia Antipolis, CNRS

Hicham Lesfari
Université Côte d'Azur, Inria Sophia Antipolis, CNRS

Emanuele Natale
Université Côte d'Azur, Inria Sophia Antipolis, CNRS

Laurent Viennot
Inria Paris, IRIF

Abstract

The average properties of the well-known *Subset Sum Problem* can be studied by the means of its randomised version, where we are given a target value z, random variables $X_1, \ldots, X_n$, and an error parameter $\varepsilon > 0$, and we seek a subset of the X_i's whose sum approximates z up to error ε. In this setup, it has been shown that, under mild assumptions on the distribution of the random variables, a sample of size $\mathcal{O}\left(\log(1/\varepsilon)\right)$ suffices to obtain, with high probability, approximations for all values in $[-1/2, 1/2]$. Recently, this result has been rediscovered outside the algorithms community, enabling meaningful progress in other fields. In this work we present an alternative proof for this theorem, with a more direct approach and resourcing to more elementary tools, in the hope of disseminating it even further.

2012 ACM Subject Classification

Keywords and phrases

1 Introduction

In the *Subset Sum Problem (SSP)*, one is given as input a set of n integers $X = \{x_1, x_2, \ldots, x_n\}$ and a target value z, and wishes to decide if there exists a subset of X that sums to z. That is, one is to reason about a subset $S \subseteq [n]$ such that

$$\sum_{i \in S} x_i = z.$$

The special case where z is half of the sum of X is known as the *Number Partition Problem (NPP)*. The converse reduction is also rather immediate.[1]

Be it in either of these forms, the SSP finds applications in a variety of fields, ranging from combinatorial number theory [24] to cryptography [12, 18]. In complexity theory, the SSP is a well-known NP-complete problem, being a common base for NP-completeness proofs. In fact, the NPP version figures among Garey and Johnson's six basic NP-hard problems [11]. Under certain circumstances, the SSP can be challenging even for heuristics that perform well for other NP-hard problems [17, 23], and a variety of dedicated algorithms have been

[1] To find a subset of X summing to z, one only needs to solve the NPP for the set $X \cup \{2z, \sum_{i \in [n]} x_i\}$. By doing so, one of the parts will consist of the element $\sum_{i \in [n]} x_i$ alongside the desired subset.

proposed to solve it [5, 9, 14–16], even using quantum computing [3, 14, 19]. Nonetheless, it is not hard to solve it in polynomial time if we restrict the input integers to a fixed range [2]. It suffices to recursively list all achievable sums using the first i integers: we start with $A_0 = \{0\}$ and compute A_{i+1} as $A_i \cup \{a + x_{i+1} \mid a \in A_i\}$. For integers in the range $[0, R]$, the search space has size $\mathcal{O}(nR)$.

Studying how the problem becomes hard as we consider larger ranges of integers (relative to n) requires a randomised version of the problem, the *Random Subset Sum Problem (RSSP)*, where the input values are taken as independently and identically distributed random variables. In this setup, the work [4] proved that the problem experiences a phase transition in its average complexity when the range of integers increases.

The result we approach in this work comes from related studies on the typical properties of the problem. In [20] the author proves that, under fairly general conditions, the expected minimal distance between a subset sum and the target value is exponentially small. More specifically, they show the following result.

▶ **Theorem 1** (Lueker, [20]). *Let $X_1, \ldots, X_n$ be i.i.d. uniform random variables over $[-1, 1]$, and let $\varepsilon \in (0, 1/2)$. There exists a universal constant $C > 0$ such that, if $n \geq 2C \log(1/\varepsilon)$, then*

$$\Pr\left[\forall z \in \left[-\frac{1}{2}, \frac{1}{2}\right], \exists S \subseteq [n] : \left|z - \sum_{i \in S} X_i\right| \leq 2\varepsilon\right] \geq 1 - \exp\left(-\frac{\left(\frac{n}{2} - C \log \frac{1}{\varepsilon}\right)^2}{2n}\right).$$

Even though Theorem 1 is stated and proved for uniform random variables over $[-1, 1]$, it is not hard to extend the result to distributions whose probability density function f has bounded support and satisfies $f(x) \geq b$ for all $x \in [-a, a]$, for some constants $a, b > 0$ (see Corollary 3.3 from [20]). With this added generality, the theorem has recently enabled progress in the field of Machine Learning, taking part in a proof of the Strong Lottery Ticket Hypothesis [22], in subsequent related works [6, 7, 10], and in Federated Learning [25].

The simplicity and ubiquity of SSP has granted the related results a special didactic place. Be it as a first example of NP-complete problem [11], a path to science communication [13], or a demonstration of advanced techniques [21], it has been a tool to make important, but sometimes complicated, ideas easier to communicate. We try to recover some of this property by proposing an alternative proof for it, that not only attains itself to more accessible tools, but also preserves much of the intuition behind the theorem. We believe our argument to be accessible enough for an undergraduate course on randomised algorithms.

1.1 Former work

The work [20] tackles Theorem 1 by considering the random variable associated to the proportion of the values in the interval $[-1/2, 1/2]$ that can be approximated up to error ε by the sum of some subset of the first t variables, $X_1, \ldots, X_t$. They proceed by evaluating the expected per-round growth of such variable, conditioned on the outcomes of $X_1, \ldots, X_t$, and then applying a nonlinear transformation to such expected value in order to make it amenable to analysis by martingale theory. This is only possible by restricting to approximations realised by subsets for which all partial sums lie in $[-1/2, 1/2]$. The proof follows by employing the previous evaluations in the construction of a suitable martingale, for which we could not find an intuitive description. A subsequent application of the Azuma-Hoeffding bound [1] followed by a case analysis leads to the result.

In this work, we propose a simplified strategy, with fewer assumptions than the original. This allows us to maintain the intuition of what our measures and variables describe. During the analysis, we highlight the simplifications we make and provide details about them.

2 Our result

In this section, we provide an alternative argument for Theorem 1. Our approach yields a slightly different thesis, yet preserving all the essence of the result. Most notably, our proof directly[2] ensures approximations for all values in $[-1, 1]$, which is arguably more natural than the original $[-1/2, 1/2]$ range. We prove

▶ **Theorem 2.** *Let $X_1, \dots, X_n$ be independent uniform random variables over $[-1, 1]$, and let $\varepsilon \in (0, 1/3)$. There exist constants $C > 0$ and $\kappa > 0$ such that, if $n \geq C \log(1/2\varepsilon)$, then*

$$\Pr\left[\forall z \in [-1,1],\, \exists S \subseteq [n] : \left|z - \sum_{i \in S} X_i\right| \leq 2\varepsilon\right] \geq 1 - \exp\left[-\frac{\left(n - C\log\frac{1}{2\varepsilon}\right)^2}{\kappa n}\right].$$

Our argument takes shape much like the pseudo-polynomial algorithm we described in the introduction. Leveraging the recursive nature of the problem, we construct a process which, at time t, describes the proportion of the interval $[-1, 1]$ that can be approximated by some subset of the first t variables.

We will show that after some time (proportional to $\log(1/\varepsilon)$) a factor of $1 - \varepsilon/2$ of the values in $[-1, 1]$ can be approximated up to error ε.

2.1 Preliminaries

Let $X_1, \dots, X_n$ be realisations of random variables as in Theorem 2, and, without loss of generality, fix $\varepsilon \in (0, 1/3)$. We say a value $z \in \mathbb{R}$ is *ε-approximated at time t* if and only if there exists $S \subseteq [t]$ such that

$$\left|z - \sum_{i \in S} X_i\right| < \varepsilon.$$

For $0 \leq t \leq n$, let $f_t \colon \mathbb{R} \to \{0, 1\}$ be the indicator function for the event "z is ε-approximated at time t". Therefore, we have $f_0 = \mathbb{1}_{(-\varepsilon,\varepsilon)}$, since only the interval $(-\varepsilon, \varepsilon)$ can be approximated by an empty set of values. From there, we can exploit the recurrent nature of the problem: a value z can be ε-approximated at time $t + 1$ if and only if either z or $z - X_{t+1}$ could already be approximated at time t. This implies that for all $z \in \mathbb{R}$ we have

$$f_{t+1}(z) = f_t(z) + \big(1 - f_t(z)\big)\, f_t(z - X_{t+1}). \tag{1}$$

Now, to keep track of the proportion of values in $[-1, 1]$ that can be ε-approximated at each step, we define, for each $0 \leq t \leq n$, the random variable

$$v_t = \frac{1}{2}\int_{-1}^{1} f_t(z)\,\mathrm{d}z.$$

For better readability, throughout the text we will refer to v_t simply as "the volume."

[2] The shorter range in Theorem 1 is a mere artefact of their proof and can be easily overcome.

As we mentioned, our goal is to show that in at most n steps enough of the interval can be approximated. More precisely, we aim to prove that $v_n \geq 1 - \varepsilon/2$, as this implies that every value in $[-1, 1]$ is either ε-approximated, or distant at most ε from an ε-approximated value, so it must be 2ε-approximated.

2.1.1 Expected behaviour

Our first lemma provides a lower bound on the expected value of v_t.

▶ **Lemma 3.** *For all $0 \leq t < n$, it holds that*

$$\mathbb{E}\left[v_{t+1} \,\middle|\, X_1, \ldots, X_t\right] \geq v_t \left[1 + \frac{1}{4}\left(1 - v_t\right)\right].$$

Proof. The definition of v_t and the recurrence in Eq. (1) give us that

$$\begin{aligned}
\mathbb{E}\left[v_{t+1} \,\middle|\, X_1, \ldots, X_t\right] &= \mathbb{E}\left[\frac{1}{2}\int_{-1}^{1} f_{t+1}(z)\,\mathrm{d}z \,\middle|\, X_1, \ldots, X_t\right] \\
&= \int_{-1}^{1} \frac{1}{2}\left(\frac{1}{2}\int_{-1}^{1} f_t(z) + \big(1 - f_t(z)\big)\, f_t(z - x)\,\mathrm{d}z\right)\mathrm{d}x \\
&= \frac{1}{2}\int_{-1}^{1} f_t(z)\,\mathrm{d}z \int_{-1}^{1}\frac{1}{2}\,\mathrm{d}x + \frac{1}{2}\int_{-1}^{1}\frac{1}{2}\int_{-1}^{1}\big(1 - f_t(z)\big)\, f_t(z - x)\,\mathrm{d}z\,\mathrm{d}x \\
&= v_t + \frac{1}{4}\int_{-1}^{1}\big(1 - f_t(z)\big)\int_{-1}^{1} f_t(z - x)\,\mathrm{d}x\,\mathrm{d}z \\
&= v_t + \frac{1}{4}\int_{-1}^{1}\big(1 - f_t(z)\big)\int_{z-1}^{z+1} f_t(y)\,\mathrm{d}y\,\mathrm{d}z,
\end{aligned}$$

where the last equality holds by substituting $y = z - x$. For the previous ones we apply basic properties of integrals and Fubini's theorem to change the order of integration.

We now look for a lower bound for the last integral in terms of v_t. To this end, we exploit that, since all integrands are non-negative, for all $u \in [-1/2, 1/2]$ we have

$$\begin{aligned}
\int_{-1}^{1}\big(1 - f_t(z)\big)\int_{z-1}^{z+1} f_t(y)\,\mathrm{d}y\,\mathrm{d}z &\geq \int_{u-\frac{1}{2}}^{u+\frac{1}{2}}\big(1 - f_t(z)\big)\int_{z-1}^{z+1} f_t(y)\,\mathrm{d}y\,\mathrm{d}z \\
&\geq \int_{u-\frac{1}{2}}^{u+\frac{1}{2}}\big(1 - f_t(z)\big)\int_{u-\frac{1}{2}}^{u+\frac{1}{2}} f_t(y)\,\mathrm{d}y\,\mathrm{d}z.
\end{aligned}$$

Both inequalities come from range restrictions: in the first we use that $u \in [-1/2, 1/2]$ implies $[u - 1/2, u + 1/2] \subseteq [-1, 1]$; for the second, we have that $[u - 1/2, u + 1/2] \subseteq [z - 1, z + 1]$ for all $z \in [u - 1/2, u + 1/2]$.

To relate the expression to v_t explicitly, we choose u in a way that the window $[u - 1/2, u + 1/2]$ entails exactly half of v_t. The existence of such u may become clear by recalling the definition of v_t. To make it formal, consider the function given by

$$h(u) = \frac{1}{2}\int_{u-\frac{1}{2}}^{u+\frac{1}{2}} f_t(y)\,\mathrm{d}y,$$

and observe that

$$\min\left\{h(-1/2),\, h(1/2)\right\} \leq \frac{v_t}{2}, \qquad \text{and} \qquad \max\left\{h(-1/2),\, h(1/2)\right\} \geq \frac{v_t}{2}.$$

Thus, by the intermediate value theorem (Theorem 12), there exists $u^* \in [-1/2, 1/2]$ for which $h(u^*) = v_t/2$, that is, for which

$$\frac{1}{2}\int_{u^*-\frac{1}{2}}^{u^*+\frac{1}{2}} f_t(y)\,\mathrm{d}y = \frac{v_t}{2}.$$

Applying the arguments above, we get

$$\begin{aligned}
\mathbb{E}\left[v_{t+1} \,\middle|\, X_1, \ldots, X_t\right] &= v_t + \frac{1}{4}\int_{-1}^{1}\left(1 - f_t(z)\right)\int_{z-1}^{z+1} f_t(y)\,\mathrm{d}y\,\mathrm{d}z \\
&\geq v_t + \frac{1}{2}\int_{u^*-\frac{1}{2}}^{u^*+\frac{1}{2}}\left(1 - f_t(z)\right)\left(\frac{1}{2}\int_{u^*-\frac{1}{2}}^{u^*+\frac{1}{2}} f_t(y)\,\mathrm{d}y\right)\mathrm{d}z \\
&= v_t + \left(\frac{1}{2} - \frac{v_t}{2}\right)\frac{v_t}{2} \\
&= v_t\left[1 + \frac{1}{4}\left(1 - v_t\right)\right].
\end{aligned}$$

◀

▶ Remark 4. Our proof of Lemma 3 marks an important divergence from [20]. While their equivalent lemma provides the exact value for the expected growth, they only consider values approximated by subsets whose partial sums all lie close to the value. More precisely, they consider a value z to be approximated only if we can do so with a subset $S \in [n]$ such that, for each $j \in S$, it holds that $(z - \sum_{i \in S : i \leq j} X_i) \in [-\frac{1}{2}, \frac{1}{2}]$. They name such approximations *admissible*.

Lemma 3 tells us that, if v_t were to behave as expected, it should grow exponentially up to $1/2$, at which point $1 - v_t$ starts to decrease exponentially. The rest of the proof follows accordingly, with Section 2.2 analysing the progress of v_t up to one half, and Section 2.3 following the complementary value, $1 - v_t$, starting from one half.

2.2 Growth of the volume up to 1/2

Arguably, the main challenge in analysing the RSSP is the existence of over-time dependencies. Deciding how to overcome it sets much of the course the proof will take. Our strategy consists in constructing another process which dominates the original one while being free of dependencies.

▶ Remark 5. This approach marks yet another divergence from [20], perhaps the most important one. They proceed by constructing a suitable martingale to apply the Azuma-Hoeffding inequality [1], which is robust to dependencies. To do so, they seek a version of Lemma 3 with a tighter bound. This is difficult to compute, and we believe was an important motivation for restricting the analysis to admissible approximations. Moreover, a proper martingale is only achieved after applying a non-linear transformation to their analogous of v_t, further hindering any intuition about the involved measures.

Let τ_1 be the first time at which the volume exceeds $1/2$, that is, let

$$\tau_1 = \min\left\{t \geq 0 : v_t > \frac{1}{2}\right\}.$$

We have seen so far that up to time τ_1 the process v_t enjoys exponential growth in expectation. In the following lemma we apply a basic concentration inequality to translate this property into a constant probability of exponential growth for v_t itself.

▶ **Lemma 6.** *For all* $0 \leq t < \tau_1$ *and* $\beta \in (0, 1/8)$ *it holds that*

$$\Pr\left[v_{t+1} \geq v_t(1+\beta) \,\middle|\, X_1, \ldots, X_t,\, t < \tau_1\right] \geq p_\beta,$$

where $p_\beta = 1 - \frac{7}{8(1-\beta)}$.

Proof. The result shall follow easily from reverse Markov's inequality (Lemma 14) and the bound from Lemma 3. However, doing so requires a suitable upper bound on v_{t+1} and, while $2v_t$ would serve the purpose, such bound does not hold in general.

We overcome this limitation by fixing t and considering how much v_t would grow in the next step if we were to consider only values ε-approximated at time t that happen to lie in $[-1, 1]$ after being translated by X_{t+1}. Making it precise by the means of the recurrence in Eq. (1), we define

$$\tilde{v} = \frac{1}{2}\int_{-1}^{1} f_t(z) + \big(1 - f_t(z)\big)\, f_t(z - X_{t+1}) \cdot \mathbb{1}_{[-1,1]}(z - X_{t+1})\,\mathrm{d}z.$$

This expression differs from the one for v_{t+1} only by the inclusion of the characteristic function of $[-1, 1]$. This not only implies that $\tilde{v} \leq v_{t+1}$, but also that $\tilde{v}$ can replace v_{t+1} in the bound from Lemma 3, since the argument provided there eventually restricts itself to integrals within $[-1, 1]$, trivialising $\mathbb{1}_{[-1,1]}$. Moreover, as we obtain $\tilde{v}$ without the influence of values from outside $[-1, 1]$, we must have $\tilde{v} \leq 2v_t$. Finally, using that $t < \tau_1$ implies $v_t < 1/2$ and chaining the previous conclusions in respective order, we conclude that

$$\begin{aligned}
\Pr\left[v_{t+1} \geq v_t(1+\beta) \,\middle|\, X_1, \ldots, X_t, t < \tau_1\right] &\geq \Pr\left[\tilde{v} \geq v_t(1+\beta) \,\middle|\, X_1, \ldots, X_t, t < \tau_1\right] \\
&\geq \frac{\mathbb{E}\left[\tilde{v} \,\middle|\, X_1, \ldots, X_t, t < \tau_1\right] - v_t(1+\beta)}{2v_t - v_t(1+\beta)} \\
&\geq \frac{\frac{9}{8}v_t - v_t(1+\beta)}{2v_t - v_t(1+\beta)} \\
&= 1 - \frac{7}{8(1-\beta)},
\end{aligned}$$

where we applied the reverse Markov's inequality in the second step. ◀

The previous lemma naturally leads us to look for bounds on τ_1, that is, to estimate the time needed for the process to reach volume $1/2$. As expected, the exponential nature of the process yields a logarithmic bound.

▶ **Lemma 7.** *For all* $\beta \in (0, 1/8)$ *and all integers* t *with*

$$t \geq \frac{1}{p_\beta}\left\lceil \frac{\log \frac{1}{2\varepsilon}}{\log(1+\beta)} \right\rceil,$$

we have that

$$\Pr\left[\tau_1 \leq t\right] \geq 1 - \exp\left[-\frac{2p_\beta^2}{t}\left(t - \frac{1}{p_\beta}\left\lceil \frac{\log \frac{1}{2\varepsilon}}{\log(1+\beta)} \right\rceil\right)^2\right].$$

Proof. The main idea behind the proof is to define a new random variable which stochastically τ_1 while being simpler to analyse. We begin by discretizing the domain $(0, 1/2]$ of the volume

into sub-intervals $\{I_i\}_{1\leq i\leq i^*}$ as follows:

$$\begin{cases} I_1 = (0, \varepsilon], \\ I_i = \left(\varepsilon\,(1+\beta)^{i-1}\,,\, \varepsilon\,(1+\beta)^{i}\right] \text{ for } 2 \leq i < i^*, \\ I_{i^*} = \left(\varepsilon\,(1+\beta)^{i^*-1}\,,\, \frac{1}{2}\right], \end{cases}$$

where i^* is the smallest integer for which $\varepsilon\,(1+\beta)^{i^*} \geq \frac{1}{2}$, i.e.

$$i^* = \left\lceil \frac{\log \frac{1}{2\varepsilon}}{\log(1+\beta)} \right\rceil.$$

Now, for each $i \geq 0$, we direct our interest to the number of steps required for v_t to exit the sub-interval I_i after first entering it. By Lemma 6, this number is majorized by a geometric random variable $Y_i \sim \mathit{Geom}(p_\beta)$. Therefore, we can conclude that τ_1 is stochastically dominated by the sum of such variables, that is, for $t \in \mathbb{N}$, we have

$$\Pr\left[\tau_1 \geq t\right] \leq \Pr\left[\sum_{i=1}^{i^*} Y_i \geq t\right]. \tag{2}$$

Let $B_t \sim \mathrm{Bin}(t, p_\beta)$ be a binomial random variable. For the sum of geometric random variables, it holds that

$$\Pr\left[\sum_{i=1}^{i^*} Y_i \leq t\right] = \Pr\left[B_t \geq i^*\right].$$

Since $\mathbb{E}\left[B_t\right] = tp_\beta$, the Hoeffding bound for binomial random variables (Lemma 15) implies that, for all $\lambda \geq 0$, we have

$$\Pr\left[B_t \leq tp_\beta - \lambda\right] \leq \exp\left[-\frac{2\lambda^2}{t}\right].$$

Setting t such that $tp_\beta - \lambda = i^*$, we get

$$\begin{aligned} \Pr\left[\sum_{i=1}^{i^*} Y_i \geq t\right] &\leq \Pr\left[B_t \leq i^*\right] \\ &\leq \exp\left[-\frac{2\left(tp_\beta - i^*\right)^2}{t}\right] \\ &= \exp\left[-\frac{2p_\beta^2\left(t - \frac{i^*}{p_\beta}\right)^2}{t}\right], \end{aligned}$$

as long as $\lambda = tp_\beta - i^* \geq 0$, that is, for all

$$t \geq \frac{1}{p_\beta}\left\lceil \frac{\log \frac{1}{2\varepsilon}}{\log(1+\beta)} \right\rceil.$$

Finally, applying this to Eq. (2) and passing to complementary events, we obtain that

$$\Pr[\tau_1 \leq t] \geq 1 - \exp\left(-\frac{2p_\beta^2\left(t - \frac{i^*}{p_\beta}\right)^2}{t}\right).$$

◀

2.3 Growth of the volume from 1/2

From now we study the second half of the process: from the moment the volume reaches $1/2$ up to the time it gets to $1 - \varepsilon/2$. We do so by analysing the complementary stochastic process, i.e., by tracking, from time τ_1 onwards, the proportion of the interval $[-1, 1]$ that does not admit an ε-approximation. More precisely, we consider the process $\{w_t\}_{t\geq 0}$, defined by $w_t = 1 - v_{\tau_1+t}$.

We shall obtain results for w_t similar to those we have proved for v_t. Fortunately, as we will see, those proofs offer even less resistance. We start with an analogous of Lemma 3.

▶ **Lemma 8.** *For all $t \geq 0$, it holds that*

$$\mathbb{E}\left[w_{t+1} \,\middle|\, X_1, \ldots, X_{\tau_1+t}\right] \leq w_t\left[1 - \frac{1}{4}(1 - w_t)\right].$$

Proof. From the definition of w_{t+1} and Lemma 3, it follows that

$$\begin{aligned}
\mathbb{E}\left[w_{t+1} \,\middle|\, X_1, \ldots, X_{\tau_1+t}\right] &= 1 - \mathbb{E}\left[v_{\tau_1+t+1} \,\middle|\, X_1, \ldots, X_{\tau_1+t}\right] \\
&\leq 1 - v_{\tau_1+t}\left[1 + \frac{1}{4}\left(1 - v_{\tau_1+t}\right)\right] \\
&= w_t - \frac{1}{4}w_t(1 - w_t) \\
&= w_t\left[1 - \frac{1}{4}(1 - w_t)\right].
\end{aligned}$$

◀

Let τ_2 the first time that w_t gets smaller than or equal to ε, that is, let

$$\tau_2 = \min\left\{t \geq 0 : w_t \leq \frac{\varepsilon}{2}\right\}.$$

The following lemma bounds this quantity, in analogy to Lemma 7.

▶ **Lemma 9.** *For every $t > 0$, it holds that*

$$\Pr[\tau_2 \leq t] \geq 1 - \frac{1}{\varepsilon} \cdot \left(\frac{7}{8}\right)^t.$$

Proof. Applying that $1 - w_t = v_{\tau_1+t} > \frac{1}{2}$ to Lemma 8 gives the bound

$$\mathbb{E}\left[w_{t+1} \,\middle|\, X_1, \ldots, X_{\tau_1+t}\right] \leq \frac{7}{8}w_t. \tag{3}$$

Moreover, from the conditional expectation theory, for any two random variables X and Y, we have $\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}\left[X\right]$. From this and Eq. (3), we can conclude that

$$\begin{aligned}\mathbb{E}\left[w_t\right] &= \mathbb{E}\left[\mathbb{E}\left[w_t \mid X_1, \ldots, X_{\tau_1+t-1}\right]\right] \\ &\leq \frac{7}{8}\mathbb{E}\left[w_{t-1}\right],\end{aligned}$$

which, by recursion, yields

$$\begin{aligned}\mathbb{E}\left[w_t\right] &\leq \left(\frac{7}{8}\right)^t \mathbb{E}\left[w_0\right] \\ &\leq \frac{1}{2}\left(\frac{7}{8}\right)^t.\end{aligned}$$

Finally, by Markov's inequality (Lemma 13),

$$\begin{aligned}\Pr\left[\tau_2 \geq t\right] &\leq \Pr\left[w_t \geq \frac{\varepsilon}{2}\right] \\ &\leq \frac{2\mathbb{E}\left[w_t\right]}{\varepsilon} \\ &\leq \frac{1}{\varepsilon}\left(\frac{7}{8}\right)^t,\end{aligned}$$

and the thesis follows from considering the complementary event. ◀

2.4 Putting everything together

In this section, we conclude our argument, finally proving Theorem 2. Let $\tau = \tau_1 + \tau_2$, the first time at which the sequence of v_t's reaches at least $1 - \varepsilon/2$.

▶ **Lemma 10.** *Let $\varepsilon \in (0, 1/3)$. There exist constants $C > 0$ and $\kappa > 0$ such that for every $t \geq C\log(1/2\varepsilon)$, it holds that*

$$\Pr\left[\tau \leq t\right] \geq 1 - \exp\left[-\frac{\left(t - C\log\frac{1}{2\varepsilon}\right)^2}{\kappa t}\right].$$

Proof. The definition of τ allows us to apply Lemmas 7 and 9 quite directly. Indeed if, for the sake of Lemma 7, we assume $t \geq \frac{2}{p_\beta}\left\lceil\frac{\log\frac{1}{2\varepsilon}}{\log(1+\beta)}\right\rceil$ for some $\beta \in (0, 1/8)$, we have

$$\begin{aligned}\Pr\left[\tau \leq t\right] &= \Pr\left[\tau_1 + \tau_2 \leq t\right] \\ &\geq \Pr\left[\tau_1 \leq t/2, \tau_2 \leq t/2\right] \\ &\geq \Pr\left[\tau_1 \leq t/2\right] + \Pr\left[\tau_2 \leq t/2\right] - 1 \\ &\geq 1 - \exp\left[-\frac{p_\beta^2}{t}\left(t - \frac{2}{p_\beta}\left\lceil\frac{\log\frac{1}{2\varepsilon}}{\log(1+\beta)}\right\rceil\right)^2\right] - \frac{1}{\varepsilon}\cdot\left(\frac{7}{8}\right)^{t/2}.\end{aligned}$$

where the second inequality holds by the union bound.

The remaining of the proof consists in computations to connect this expression to the one in the statement.

To this end, we start by choosing $\beta = 1/16$ and by observing that for every $\lambda \geq 3/2$, we have

$$\exp\left[-\frac{p_\beta^2}{t}\left(t - \frac{2}{p_\beta}\left\lceil\frac{\log\frac{1}{2\varepsilon}}{\log(1+\beta)}\right\rceil\right)^2\right] + \frac{1}{\varepsilon}\left(\frac{7}{8}\right)^{t/2} \leq \exp\left[-\frac{p_\beta^2}{t}\left(t - \frac{2\lambda\log\frac{1}{2\varepsilon}}{p_\beta\log(1+\beta)}\right)^2\right], \quad (4)$$

for all $\varepsilon < 1/3$ and

$$t \geq \frac{2\lambda\log\frac{1}{2\varepsilon}}{p_\beta\log(1+\beta)}.$$

More precisely, let $\alpha_\varepsilon = C_\beta \log\frac{1}{2\varepsilon}$ where $C_\beta = \frac{2}{p_\beta\log(1+\beta)}$. By dividing Eq. (4) by its right-hand side, we obtain

$$\exp\left[(\lambda^2-1)p_\beta^2\tfrac{\alpha_\varepsilon^2}{t} + 2(1-\lambda)p_\beta^2\alpha_\varepsilon\right] + \exp\left[\lambda^2 p_\beta^2\tfrac{\alpha_\varepsilon^2}{t} + \left(\tfrac{p_\beta\log(1+\beta)}{2} - 2\lambda p_\beta^2\right)\alpha_\varepsilon + \left(p_\beta^2 - \tfrac{\log\frac{8}{7}}{2}\right)t + \log 2\right] \leq 1. \quad (5)$$

Observe that, since $t \geq \lambda\alpha_\varepsilon$ and $p_\beta^2 - \frac{\log\frac{8}{7}}{2} < 0$, the left-hand side of Eq. (5) is at most

$$\exp\left[-\left(\frac{1-\lambda^2}{\lambda} + 2(\lambda-1)\right)p_\beta^2\alpha_\varepsilon\right] + \exp\left[\left(\frac{p_\beta\log(1+\beta)}{2} - \frac{\lambda}{2}\log\frac{8}{7}\right)\alpha_\varepsilon + \log 2\right].$$

We rewrite the above expression as

$$e^{-a_0\alpha_\varepsilon} + 2e^{-a_1\alpha_\varepsilon} \overset{(a)}{\leq} 1,$$

by setting $a_0 = \left(\frac{1-\lambda^2}{\lambda} + 2(\lambda-1)\right)p_\beta^2 > 0$ and $a_1 = -\left(\frac{p_\beta\log(1+\beta)}{2} - \frac{\lambda}{2}\log\frac{8}{7}\right) > 0$. As for (a), it holds for all values of $\varepsilon \in (0, 1/3)$. ◀

Choosing $\beta = 1/16$, $\kappa = 1/p_\beta^2$ and $C = \frac{3}{p_\beta\log(1+\beta)}$, for all $\varepsilon \in (0, 1/3)$ and $t \geq C\log\frac{1}{2\varepsilon}$, Lemma 10 can be reformulated as

$$\Pr\left[v_t \geq 1-\varepsilon\right] \geq 1 - \exp\left[-\frac{\left(t - C\log\frac{1}{2\varepsilon}\right)^2}{\kappa t}\right].$$

Theorem 2 follows by observing that once we can approximate all but an $\varepsilon/2$ proportion of the interval $[-1, 1]$, any $z \in [-1, 1]$ either is ε-approximated, or it is distant at most ε from a value that it is ε-approximated, which implies that z is 2ε-approximated.

▶ Remark 11. Our proof worsens the minimum number of variables t for which the theorem holds. In particular, from [20] we get that t must be at least $2(1+\log_2 e)\log(1/\varepsilon) \sim 4.89\log(1/\varepsilon)$. In our case, t must be at least $3 \cdot 15/\log(17/16) \cdot (\log(1/\varepsilon) - \log 2) \sim 742.27(\log(1/\varepsilon) - \log 2)$. We highlight that this work provides an existence proof, proposing no algorithms. Thus, we focused on the proof simplicity rather than on constant optimisation.

References

1 Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.

2 Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

3 Daniel J Bernstein, Stacey Jeffery, Tanja Lange, and Alexander Meurer. Quantum algorithms for the subset-sum problem. In *International Workshop on Post-Quantum Cryptography*, pages 16–33. Springer, 2013.

4 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. doi:10.1002/rsa.10004.

5 Karl Bringmann and Philip Wellnitz. On near-linear-time algorithms for dense subset sum. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1777–1796. SIAM, 2021.

6 Rebekka Burkholz, Nilanjana Laha, Rajarshi Mukherjee, and Alkis Gotovos. On the existence of universal lottery tickets. In *International Conference on Learning Representations*, 2022. URL: https://openreview.net/forum?id=SYB4WrJql1n.

7 Arthur da Cunha, Emanuele Natale, and Laurent Viennot. Proving the strong lottery ticket hypothesis for convolutional neural networks. In *International Conference on Learning Representations*, 2022. URL: https://openreview.net/forum?id=Vjki79-619-.

8 Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009. URL: http://www.cambridge.org/gb/knowledge/isbn/item2327542/.

9 Andre Esser and Alexander May. Low weight discrete logarithms and subset sum in $2^{0.65n}$ with polynomial memory. *Cryptology ePrint Archive*, 2019.

10 Jonas Fischer and Rebekka Burkholz. Towards strong pruning for lottery tickets with non-zero biases. *CoRR*, abs/2110.11150, 2021. URL: https://arxiv.org/abs/2110.11150, arXiv:2110.11150.

11 M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

12 Peter Gemmell and Anna M. Johnston. Analysis of a subset sum randomizer. *IACR Cryptol. ePrint Arch.*, page 18, 2001. URL: http://eprint.iacr.org/2001/018.

13 Brian Hayes. The easiest hard problem. *American Scientist*, 90:113–117, 2002.

14 Alexander Helm and Alexander May. Subset sum quantumly in 1.17ˆ n. In *13th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

15 Ce Jin, Nikhil Vyas, and Ryan Williams. Fast low-space algorithms for subset sum. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1757–1776. SIAM, 2021.

16 Ce Jin and Hongxun Wu. A simple near-linear pseudopolynomial time randomized algorithm for subset sum. *arXiv preprint arXiv:1807.11597*, 2018.

17 David S Johnson, Cecilia R Aragon, Lyle A McGeoch, and Catherine Schevon. Optimization by simulated annealing: an experimental evaluation; part ii, graph coloring and number partitioning. *Operations research*, 39(3):378–406, 1991.

18 Aniket Kate and Ian Goldberg. Generalizing cryptosystems based on the subset sum problem. *Int. J. Inf. Sec.*, 10(3):189–199, 2011. doi:10.1007/s10207-011-0129-2.

19 Yang Li and Hongbo Li. Improved quantum algorithm for the random subset sum problem. *arXiv preprint arXiv:1912.09264*, 2019.

20 G. S. Lueker. Exponentially small bounds on the expected optimum of the partition and subset sum problems. *Random Structures and Algorithms*, 12:51–62, 1998.

21 Stephan Mertens. A physicist's approach to number partitioning. *Theor. Comput. Sci.*, 265(1-2):79–108, 2001. doi:10.1016/S0304-3975(01)00153-0.

22 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. URL: https://proceedings.neurips.cc/paper/2020/hash/1b742ae215adf18b75449c6e272fd92d-Abstract.html.

23 Wheeler Ruml, J Thomas Ngo, Joe Marks, and Stuart M Shieber. Easily searched encodings for number partitioning. *Journal of Optimization Theory and Applications*, 89(2):251–291, 1996.

24 Zhi-Wei Sun. Unification of zero-sum problems, subset sums and covers of z. *Electronic Research Announcements of The American Mathematical Society*, 9:51–60, 2003.

25 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. doi:10.18653/v1/2021.emnlp-main.606.

A Tools

▶ **Theorem 12** (Intermediate Value Theorem). *Let $g\colon [a,b] \to \mathbb{R}$ be a continuous real-valued function such that $\lambda = \min\{g(a), g(b)\} < \max\{g(a), g(b)\} = \Lambda$. Then, for any value $\delta \in [\lambda, \Lambda]$, there exists a point $a < c_\delta < b$ such that $g(c_\delta) = \delta$.*

▶ **Lemma 13** (Markov's inequality). *Let X be a non-negative random variable. Then for all $c > 0$, we have*

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}.$$

▶ **Lemma 14** (Reverse Markov's inequality). *Let X be a random variable such that $X \leq u$ for some constant $u \in \mathbb{R}$. Then for all $c < u$, we have*

$$\Pr[X > c] \geq \frac{\mathbb{E}[X] - c}{u - c}.$$

Proof. We apply Markov's inequality (Lemma 13) to the random variable $Y = u - X$. Note that Y is non-negative, since $X \leq u$. We get

$$\Pr[X \leq c] = \Pr[Y \geq u - c] \leq \frac{\mathbb{E}[Y]}{u - c} = \frac{u - \mathbb{E}[X]}{u - c},$$

and the thesis follows. ◀

▶ **Lemma 15** (Hoeffding bounds [8]). *Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $\Pr[0 \leq X_i \leq 1] = 1$ for all $i \in [n]$. Let $X = \sum_{i=1}^{n} X_i$ and $\mathbb{E}[X] = \mu$. Then*
(i) for any $\lambda \geq 0$ and $\mu \leq \mu_+$, it holds that

$$\Pr\left[X \geq \mu_+ + \lambda\right] \leq \exp\left(-\frac{2\lambda^2}{n}\right);$$

(ii) for any $\lambda \geq 0$ and $0 \leq \mu_- \leq \mu$, it holds that

$$\Pr\left[X \leq \mu_- - \lambda\right] \leq \exp\left(-\frac{2\lambda^2}{n}\right).$$