Scribe: Tim Kunisky
Lecturer: Tim Kunisky

Lecture 1: Phase transition in random integer partitioning

1 Problem statement

The first average-case problem we will study, to start getting used to using moment methods to extract information about optimization landscapes, is the *integer partitioning* problem. In general, the problem statement is as follows.

- **Input:** $a_1, \dots, a_n \in \mathbb{N}$.[1]
- **Problem:** Find a set $A \subseteq [n] := \{1, \dots, n\}$ so that $\sum_{j \in A} a_j = \sum_{j \notin A}$. Or, a little more generally, solve the optimization problem

$$\begin{array}{ll} \text{minimize} & |\sum_{j\in A} a_j - \sum_{j \notin A} a_j| \\ \text{subject to} & A \subseteq [n]. \end{array} \tag{1}$$

This is a classical computational problem (similar to things like *knapsack problems* that you may be more familiar with) that is NP-hard in the worst case. Here, we will ask about the *average-case* computational complexity of this problem. That means we will take the a_j to be *random* and try to determine whether there are efficient algorithms that typically solve integer partitioning effectively over a particular distribution of inputs. We will fix some $B \in \mathbb{N}$, and consider the distribution

$$a_1, \dots, a_n \overset{\text{iid}}{\sim} \mathsf{Unif}(\{0, 1, \dots, B-1\}). \tag{2}$$

The question we are most interested in is the *algorithmic* question of whether the optimizer $A^\star$ of (1) is easy to find or approximate. But, we will start out thinking about the simpler *combinatorial* (or, thinking of this problem as an optimization landscape over the n-dimensional hypercube identified with subsets $A \subseteq [n]$, more on which below, *geometric*) question of when there exist *perfect* partitions with $\sum_{j\in A} a_j = \sum_{j \notin A} a_j$, and more generally of what the value (just some random integer) of the optimization problem (1) typically is.

2 First moment method

Let us first focus on the most specific problem of determining for which sequences of the parameters (n, B) perfect partitions exist or do not exist with high probability.[2] Intuitively, it should be clear that it becomes less likely for perfect partitions to exist as B gets larger,

[1] For us, $\mathbb{N} = \mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$.
[2] A sequence of events happens *with high probability* if their probability tends to 1.

since this requires an increasingly "special" cancellation to happen among the a_j (think about, for example, comparing $B = 2$ and $B = 100$). So, we will view $B = B(n)$, and will try to identify the *critical scaling* of B as a function of n around which we switch from typically having perfect partitions to typically not having them.

As a side note, to prepare for our calculations, let us identify $A \subseteq [n]$ with $\boldsymbol{x} \in \{\pm 1\}^n$ so that $x_j = 1$ if $j \in A$ and $x_j = -1$ if $j \notin A$. Then, our question is the same as whether there exists $\boldsymbol{x} \in \{\pm 1\}^n$ such that

$$\langle \boldsymbol{x}, \boldsymbol{a} \rangle = \sum_{j=1}^{n} x_j a_j = 0. \tag{3}$$

A basic idea for starting to do this is to *count* the number of perfect partitions:

$$Z := \#\left\{ \boldsymbol{x} \in \{\pm 1\}^n : \langle \boldsymbol{x}, \boldsymbol{a} \rangle = 0 \right\}. \tag{4}$$

This is a random variable, inheriting its randomness from the randomness in the a_j. We are interested in whether $Z = 0$ with high probability or $Z \geq 1$ with high probability. It turns out that we can sometimes establish the former just by calculating the *expectation* of Z, which is usually much easier than establishing the entire distribution of Z.

Proposition 2.1 (Basic first moment method)**.** *Suppose $Z_n \in \mathbb{N}$ is a sequence of random variables, and $\mathbb{E}Z_n \to 0$. Then, $Z_n = 0$ with high probability (i.e., $\mathbb{P}[Z_n = 0] \to 1$).*

Proof. Using Markov's inequality,

$$\mathbb{P}[Z_n \neq 0] = \mathbb{P}[Z_n \geq 1] \leq \frac{\mathbb{E}Z_n}{1} = \mathbb{E}Z_n \to 0 \tag{5}$$

by the assumption. □

We will see that it is *not* always reasonable to hope for the converse, that if $\mathbb{E}Z$ is large enough then $Z \geq 1$ with high probability. But, let us proceed optimistically and see what a heuristic calculation of our $\mathbb{E}Z$ leads us to predict.

3 Heuristic threshold derivation

We start with a standard manipulation of such expectations:

$$\begin{aligned}
\mathbb{E}Z &= \mathbb{E} \sum_{\boldsymbol{x} \in \{\pm 1\}^n} \mathbb{1}\{\langle \boldsymbol{x}, \boldsymbol{a} \rangle = 0\} \\
&= \sum_{\boldsymbol{x} \in \{\pm 1\}^n} \mathbb{E}[\mathbb{1}\{\langle \boldsymbol{x}, \boldsymbol{a} \rangle = 0\}] \\
&= \sum_{\boldsymbol{x} \in \{\pm 1\}^n} \mathbb{P}[\langle \boldsymbol{x}, \boldsymbol{a} \rangle = 0]. \qquad (6)
\end{aligned}$$

$$
\begin{array}{ccccccccccccccc}
 & & & & & 1 & 0 & 0 & 1 & \cdots & 0 & 1 & \cdot & x_1 \\
 & & & & & 0 & 0 & 1 & 0 & \cdots & 1 & 1 & \cdot & x_2 \\
 & & & & & \vdots & & & & \vdots & & & & \vdots \\
 & & & & & 0 & 0 & 1 & 0 & \cdots & 1 & 1 & \cdot & x_n \\
\hline
1 & 0 & 0 & \cdots & & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & = & \sum_{j=1}^n x_j a_j
\end{array}
$$

Figure 1: A schematic illustration of the calculation of $\sum_{j=1}^n x_j a_j$ in binary that is discussed in Section 3.

We will see more precise ways to calculate these probabilities later, but for now let us give a heuristic derivation (taken from Section 14.5 of [MM11], along with lots of our discussion here). Let

$$b := \log_2 B, \tag{7}$$

the number of bits needed to specify a number in $\{0, 1, \dots, B-1\}$. Now, consider writing out each a_j in binary. Then, our random model of the a_j is realized by taking each of b bits in each a_j to be drawn from $\mathsf{Unif}(\{0,1\})$. Consider now calculating $\langle \boldsymbol{x}, \boldsymbol{a} \rangle = \sum_{j=1}^n x_j a_j$ in binary. The number of bits in the output will be roughly

$$\log_2\left(|\langle \boldsymbol{x}, \boldsymbol{a} \rangle|\right) \approx \log_2\left(B \cdot \left|\sum_{j=1}^n x_j\right|\right) = b + \log_2\left(\left|\sum_{j=1}^n x_j\right|\right). \tag{8}$$

Note that the least significant b of these bits are distributed i.i.d. as $\mathsf{Unif}(\{0,1\})$, as they are just XOR's of n bits distributed i.i.d. as $\mathsf{Unif}(\{0,1\})$, possibly further XOR'd with bits carried from less significant "columns" of the summation. Once we consider bits beyond the bth place, this will no longer hold, since now we are just adding carried bits from the less significant places of the summation. (It may help to consult Figure 3.)

But, let us make the heuristic leap that *all* of the bits in $\sum_{j=1}^n x_j a_j$ may be viewed as i.i.d. random variables drawn from $\mathsf{Unif}(\{0,1\})$. Then, we would predict:

$$\begin{aligned} \mathbb{P}[\langle \boldsymbol{x}, \boldsymbol{a} \rangle = 0] &\approx \left(\frac{1}{2}\right)^{b + \log_2(|\sum_{j=1}^n x_j|)} \\ &= \frac{1}{B|\sum_{j=1}^n x_j|}. \end{aligned}$$

Finally, recall that for most $\boldsymbol{x} \in \{\pm 1\}^n$, say, all but any small fraction, we have $|\sum_{j=1}^n x_j| = \Theta(\sqrt{n})$. This may be viewed as a standard observation about the magnitude of simple random walks on the integers after n steps, or as the special case of the central limit theorem sometimes called the de Moivre–Laplace theorem. You should ideally be familiar with these ideas already, but in case not, Section 3.1 of [Dur19], which you can find online, gives a good treatment.

Combining our calculations, we are led to the following prediction.

Conjecture 3.1. $\mathbb{E}Z = \Theta(2^n / B\sqrt{n}) = \Theta(2^{n - \log_2 n/2 - b})$.

Using the first moment method and trusting that its converse might hold as well, we can also conjecture the following.

Conjecture 3.2. *Suppose $B = B(n)$. Let us write*

$$B^* = B^*(n) = \frac{2^n}{\sqrt{n}}. \tag{9}$$

Then:

- *If $B/B^* \to \infty$, then with high probability there do* not *exist perfect partitions of $a_1, \ldots, a_n$.*
- *If $B/B^* \to 0$, then with high probability there exist perfect partitions of $a_1, \ldots, a_n$.*

Note that the first claim would follow from Conjecture 3.1 together with Proposition 2.1, while we do not yet have tools from which the second claim would follow.

Perhaps surprisingly, one can do numerical experiments and confirm that this conjecture appears to be true, despite our quite sloppy reasoning above! In particular, even the $\sqrt{n}$ term which came from maybe the most questionable one of our sequence of approximations appears to be the correct scaling. In the next lecture, we will perform a first moment calculation more carefully to verify Conjecture 3.1, and will introduce the *second moment method* for proving the second claim of Conjecture 3.2.

References

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge University Press, 2019.

[MM11] Cristopher Moore and Stephan Mertens. *The nature of computation.* OUP Oxford, 2011.