Number Partitioning

Suppose you have a set of N integers, $A = \{a_1, \ldots, a_N\}$, with $1 \leq a_i \leq 2^M$ (i.e., each integer can be stored in M bits). Pick a subset $B \subseteq A$, and define the energy of the resulting partition of $B \cup (A \setminus B)$ as

$$E(B) = \left| \sum_{a_i \in B} a_i - \sum_{a_i \in A \setminus B} a_i \right|.$$

A basic type of algorithm is called a decision algorithm: does there exist a partition B^* such that $E(B^*) = 0$ (such a partition is called perfect)? Note that this means that we have found a partition such that the sum of the integers in each half is the same.

Let's assume that $N \gg 1$ and $M \gg 1$, and that the elements of A are i.i.d. random variables uniformly distributed on $\{1, \ldots, 2^M\}$. For computer science, this would be a reasonable assumption: data is typically well-approximated as random.[1] The idea here is to treat the problem's "energy" cost as a physical energy, and compute the partition function $Z(\beta)$ for this system. One can think of β as a tolerance level for imperfect partitions; thus, to answer the decision problem, one would set $\beta = \infty$ (i.e., no tolerance).

(a) Show that the exact expression for Z,[2] averaged over all possible realizations of the random variables a_i, is

$$\langle Z(\beta) \rangle = 2^N \int_{-\pi}^{\pi} \frac{\mathrm{d}x}{2\pi} g(x)^N \left(1 + 2\sum_{n=1}^{\infty} \mathrm{e}^{-\beta n} \cos nx\right)$$

where $g(x)$ is the function

$$g(x) = \frac{\sin\left(2^M \frac{x}{2}\right)}{2^M \sin \frac{x}{2}} \cos\left(\left(2^M + 1\right)\frac{x}{2}\right).$$

(b) The exact answer is not particularly illuminating, so let's go for an approximate one instead. Plot or sketch $g(x)$ on $[-\pi, \pi)$ for a reasonably large M.

(c) The sketch above should suggest that the integral can be very reasonably approximated by rescaling to the variable

$$y \equiv \frac{x}{2^M \sqrt{N}}.$$

Also, define

$$\kappa \equiv \frac{M}{N}.$$

Making this substitution, and any other reasonable approximations needed, show that to leading order in N and κ,

$$\langle Z(\beta) \rangle = 2^{N(1-\kappa)} \sqrt{\frac{3}{2\pi N}} \coth \frac{\beta}{2}.$$

[1] Of course, we have no idea if our assumption of a uniform distribution is reasonable!

[2] Use the integral relation that, for $z \in \mathbb{Z}$, $\int_{-\pi}^{\pi} \frac{\mathrm{d}x}{2\pi} \mathrm{e}^{\mathrm{i}xz} = \begin{cases} 1 & z = 0 \\ 0 & \text{otherwise} \end{cases}$. Clever use of this relation will allow you to take advantage of the independence of the a_i.

(d) Are there "critical" values of N, κ, or β at which we expect to see crossovers/transitions between distinctive behaviors? Relate this to the original question of solving the decision problem.

Another type of algorithm asks you to actually *find* the partition. This is unfortunately a lot harder; in fact, this problem is NP-complete, and beyond the scope of this problem.