DRAFT: Spectral and Modular Analysis of #P Problems

Ohad Asor

January 7, 2016

Abstract

We present various analytic and number theoretic results concerning the #SAT problem as reflected when reduced into a #PART problem. As an application we present a heuristic to probabilistically estimate the solution of #SAT problems.

1 Overview

#SAT is the problem of counting the number of satisfying assignments to a given 3CNF formula, while #PART is the problem of counting the number of zero partitions in a given set of integers. Precise definitions will be given later on. Those problems lie on the complexity class #P, as whether merely deciding if the count is zero or not is an NP-Complete problem. We present various results concerning #PART and analyze their connection with #SAT. On section 2 we skim some preliminaries. Section 3 presents the core of the analytic results. Section 4 presents a modular-arithmetic formula for computing #PART, and section 5 presents implications to complexity theory. On section 6 we present miscellaneus results and a conjecture, where section 7 is analyzes how multiple reductions may give probabilistic answer to #SAT as a consequence our analysis.

2 Preliminaries

Our setting is counting the number of solutions given an instance of the Partition problem. We sometimes use custom terminology as there is no unified one.

Definition 2.1. Given $n \in \mathbb{N}, \mathbf{x} \in \mathbb{N}^n$, a *Partition* σ of $\mathbf{x}$ is some $\sigma \in \{-1,1\}^n$. The *size* of the partition σ is $\langle \mathbf{x}, \sigma \rangle = \sum_{k=1}^{n} \sigma_k x_k$. A partition is called a *zero partition* if its size is zero. The problem #PART is to determine the number of zero partitions given $\mathbf{x}$. The problem PART is deciding whether a zero partition exists or not for $\mathbf{x}$. The *Weak* setting of the problem is when $\mathbf{x}$ is supplied in unary radix, and the *Strong* setting is when it is supplied in binary radix (or another format with same efficiency), therefore the input size is logarithmically smaller on the strong setting.

#PART is in #P complexity class. The setting of the #PART after being reduced from the counting Boolean Satisfiability problem (SAT) is n integers to partition each having up to $\mathcal{O}(n)$ binary digits, demonstrating why the rather strong setting is of interest. In fact, there exist polynomial time algorithms solving the weak setting of PART, notably Dynamic Programming algorithms, as well as the formula derived on Theorem 4.1 below. However, solving PART on the strong setting is not possible in polynomial time (as a function of the input length), unless P=NP.

#SAT can be reduced to #SUBSET-SUM using an algorithm described in [1], while various slight variations appear on the literature. We summarize the reductions on the Appendix.

3 Analytic Setting

Theorem 3.1. *Given $n \in \mathbb{N}, \mathbf{x} \in \mathbb{N}^n$ then the (probability-theoretic) characteristic function of the random variable $\langle \mathbf{x}, \sigma \rangle = \sum_{k=1}^{n} x_k \sigma_k$ over uniform $\sigma \in \{-1, 1\}^n$ is $\prod_{k=1}^{n} \cos(x_k t)$.*

Proof. Consider the formula $2 \cos a \cos b = \cos(a+b) + \cos(a-b)$ and the cosine being even function to see that:

$$\psi(t) \equiv \psi(x_1, \ldots, x_n, t) \equiv \prod_{k=1}^{n} \cos(x_k t) = 2^{-n} \sum_{\sigma \in \{-1,1\}^n} \cos(t \langle \mathbf{x}, \sigma \rangle) = \mathbb{E}\left[e^{it\langle \mathbf{x}, \sigma \rangle}\right] \tag{1}$$

□

Corollary 3.2. *Given $n \in \mathbb{N}, \mathbf{x} \in \mathbb{N}^n$ then*

$$\int_0^1 \prod_{k=1}^{n} \cos(2\pi x_k t)\, dt \tag{2}$$

is the number of zero partitions of $\mathbf{x}$.

Proof. Following and integrating both sides. Stronger statements are possible (e.g. for $\mathbf{x} \in \mathbb{R}^n$ or $\mathbf{x} \in \mathbb{C}^n$) using characteristic function inversion theorems. □

Theorem 3.3. *Given $\{n, N\} \subset \mathbb{N}, j \in \mathbb{Z}, \mathbf{x} \in \mathbb{N}^n$ then*

$$\frac{1}{N} \sum_{m=1}^{N} e^{2\pi i j \frac{m}{N}} \prod_{k=1}^{n} \cos\left(2\pi x_k \frac{m}{N}\right) \tag{3}$$

is the number of partitions of $\mathbf{x}$ having size that is divisible by N with remainder j.

Proof. Following (1)

$$\frac{1}{N} \sum_{m=1}^{N} \prod_{k=1}^{n} \cos\left(2\pi x_k \frac{m}{N}\right) = \sum_{\sigma \in \{-1,1\}^n} \frac{1}{N} \sum_{m=1}^{N} e^{2\pi i \frac{m}{N} \langle \mathbf{x}, \sigma \rangle} \tag{4}$$

the sum of the roots of unity on the rhs is zero if N does not divide $\langle \mathbf{x}, \sigma \rangle$, and is one if N does divide it, therefore (3) is equal to

$$\sum_{u=-\infty}^{\infty} c_{uN} \tag{5}$$

where c_u denoting the number of partitions that sum to u:

$$c_u = \left| \{\sigma \in \{-1, 1\}^n \mid \langle \mathbf{x}, \sigma \rangle = u\} \right| \tag{6}$$

As for the remainder, observe that

$$\sum_{\sigma \in \{-1,1\}^n} e^{2\pi i t(\langle \mathbf{x}, \sigma \rangle + j)} = e^{2\pi i t j} 2^n \prod_{k=1}^{n} \cos(2\pi x_k t) \tag{7}$$

□

4 Modular Arithmetic Formula

Theorem 4.1. *Given $n \in \mathbb{N}, \mathbf{x} \in \mathbb{N}^n$, the number of $\mathbf{x}$'s zero partition out of all possible 2^n partitions is encoded as a binary number in the binary digits of*

$$\prod_{k=1}^{n} \left[1 + 4^{n x_k}\right] \tag{8}$$

from the s'th dight to the $s + n$ digit, where $s = n \langle \mathbf{x}, 1 \rangle$.

Proof. We write down the following sum and perform substitution according to (1):

$$S = \frac{1}{n} \sum_{m=1}^{n} \psi\left(\frac{2\pi m}{n} + i \ln 2\right) = \sum_{\sigma \in \{-1,1\}^n} \frac{2^{-\langle \mathbf{x}, \sigma \rangle}}{n} \sum_{m=1}^{n} e^{\frac{2\pi i m}{n} \langle \mathbf{x}, \sigma \rangle} \tag{9}$$

multiplying all x_k by n (while preserving partitions, since we can always multiply all x_k by the same factor and keep the exact number of zero partitions) puts $e^{\frac{2\pi i m}{n} \langle n\mathbf{x}, \sigma \rangle} = 1$ and we get:

$$S = \sum_{\sigma \in \{-1,1\}^n} 2^{-n \langle \mathbf{x}, \sigma \rangle} = \sum_{u=-\infty}^{\infty} c_u 2^{-u} \tag{10}$$

where c_u is defined in. Recalling that $\sum_{u=-\infty}^{\infty} c_u = 2^n$ and c_u are all positive, while on (9) being multiplied by distinct powers $2^{\pm n}$, therefore the summands' binary digits never interfere with each other. Recalling that c_0 is our quantity of interest, we have proved that the number of zero partitions in $\mathbf{x}$

$$\left\lfloor \frac{2^n}{n} \sum_{m=1}^{n} \prod_{k=1}^{n} \cos\left[n x_k \left(\frac{2\pi m}{n} + i \ln 2\right)\right] \right\rfloor \mod 2^n \tag{11}$$

$$= \left\lfloor \prod_{k=1}^{n} \left[2^{n x_k} + 2^{-n x_k}\right] \right\rfloor \mod 2^n \tag{12}$$

$$= \left\lfloor 2^{-n \sum_{k=1}^{n} x_k} \prod_{k=1}^{n} \left[1 + 2^{2n x_k}\right] \right\rfloor \mod 2^n \tag{13}$$

Set

$$M = \prod_{k=1}^{n} \left[1 + 2^{2n x_k}\right] = \sum_{\sigma \in \{0,1\}^n} 2^{2n \langle \mathbf{x}, \sigma \rangle} \tag{14}$$

then (12) tells us that the number of zero partitions is encoded as a binary number in the binary digits of M, from the s'th dight to the $s + n$ digit. □

5 Hardness of Integration

Note that the expression in (10) is nothing but the trapezoid rule of order N applied to the following integral:

Corollary 5.1. $\mathbf{x} \in \mathbb{Q}^n$ *has a zero partition if and only if*

$$\int_0^\infty \prod_{k=1}^n \cos(2\pi x_k t)\, dt = \infty \tag{15}$$

and does not have a zero partition if and only if

$$\int_0^\infty \prod_{k=1}^n \cos(2\pi x_k t)\, dt = 0 \tag{16}$$

Proof. Follows from Theorem 4.1, the integrand being periodic, and change of variable to support rationals. □

Corollary 5.2. *There is no algorithm that takes any function that can be evaluated in polynomial time, and decides in polynomial time whether its integral over the real line is zero (conversley, infinity) unless P=NP.*

Theorem 5.3. *[Theorem 2.2 on [3]] If u is an analytic function satisfying $|u(z)| \leq M$ in $\frac{1}{r} \leq |z| \leq r$ for some $r > 1$, then for any $N \geq 1$ the trapezoid rule with N points will be far from the exact integral by no more than $\frac{4\pi M}{r^N - 1}$.*

Corollary 5.4. *If there exists a function w such that given ψ as in (1) that corresponts to a #PART problem, and*

$$u(z) = \psi(w(z))\, w'(z) \tag{17}$$

is computable in polynomial time (wrt the input length and the desired output accuracy) and satisfies the conditions of Theorem 4.4 with $r = 2$ and $M = \mathcal{O}\left(poly\left(\sum_{k=1}^n x_k\right)\right)$, then P=NP.

Proof. Observe that ψ behaves like $e^{x_k t}$ for imaginary input. It therefore satisfies $M = e^{r\sum_{k=1}^n x_k}$ at the setting of Theorem 4.4. For exponential convergence wrt PART's input length we need $\frac{4\pi M}{r^N - 1}$ diminish exponentially. Therefore if we can change the variable of integration in (14) using some w and result with $M = \mathcal{O}\left(\text{poly}\left(\sum_{k=1}^n x_k\right)\right)$, we could estimate the integral in (14) to our desired accuracy (2^{-n}) in subexponential time. □

6 Additional Results and Conjectures

Theorem 6.1. *Given $n \in \mathbb{N}, N \in \mathbb{N}, \mathbf{x} \in \mathbb{N}^n$, the variance of the sizes of all partitions is the sum of the squares of the input. Formally:*

$$\sum_{k=1}^n x_k^2 = 2^{-n} \sum_{\sigma \in \{-1,1\}^n} \langle \mathbf{x}, \sigma \rangle^2 \tag{18}$$

while

$$\frac{2^n}{N^3} \sum_{m=1}^N \frac{\partial^2}{\partial t^2} \prod_{k=1}^n \cos(2\pi x_k t) \bigg|_{t=\frac{m}{N}} \tag{19}$$

is the variance of the sizes of all partitions that their size is divisible by N without remainder.

Proof. Following (1) and differentiating:

$$\prod_{k=1}^{n} \cos(\pi x_k t) = 2^{-n} \sum_{\sigma\in\{-1,1\}^n} \cos(\pi t \langle \mathbf{x}, \sigma\rangle) \tag{20}$$

$$\implies \sum_{\ell=1}^{n} x_\ell \sin(\pi x_\ell t) \prod_{k\neq\ell}^{n} \cos(\pi x_k t) = 2^{-n} \sum_{\sigma\in\{-1,1\}^n} \langle \mathbf{x}, \sigma\rangle \sin(\pi t \langle \mathbf{x}, \sigma\rangle) \tag{21}$$

$$\implies \sum_{\ell=1}^{n}\sum_{\ell\prime=1}^{n} -x_\ell \sin(\pi x_\ell t)\, x_{\ell\prime} \sin(\pi x_{\ell\prime} t) \prod_{k\neq\ell,\ell\prime}^{n} \cos(\pi x_k t) + x_\ell^2 \prod_{k=1}^{n} \cos(\pi x_k t) \tag{22}$$

$$= 2^{-n} \sum_{\sigma\in\{-1,1\}^n} \langle \mathbf{x}, \sigma\rangle^2 \cos(\pi t \langle \mathbf{x}, \sigma\rangle) \tag{23}$$

and (19) follows by substituting $t = 0$. (19) can be proved using Parseval identity as well. Turning to (20):

$$\frac{2^n}{N} \sum_{m=1}^{N} \frac{\partial^2}{\partial t^2} \prod_{k=1}^{n} \cos(2\pi x_k t)\bigg|_{t=\frac{m}{N}} = \sum_{\sigma\in\{-1,1\}^n} \langle \mathbf{x}, \sigma\rangle^2 \cos\left(2\pi \frac{m}{N} \langle \mathbf{x}, \sigma\rangle\right) = \sum_{u=-\infty}^{\infty} u^2 N^2 c_{Nu} \tag{24}$$

due to aliasing of roots of unity, and c_{Nu} the number of partitions whose size is divisible by Nu as in (4). ☐

Remark 6.2. It is easy to derive all moments and cumulants of our random variable since we're given its characteristic function.

Theorem 6.3. *Let $Z^{\mathbf{x}}$ be the number of zero partitions of a vector of naturals X. Let $D_x^{\mathbf{x}}$ be the number of zero partitions of X after multiplying one if its elements by two, where this element is denoted by x. Let $A_x^{\mathbf{x}}$ be the number of zero partitions of X after appending it x (so now x appears at least twice). Then*

$$Z^{\mathbf{x}} = D_x^{\mathbf{x}} + A_x^{\mathbf{x}} \tag{25}$$

Proof. Denote

$$\psi(x_1, \ldots, x_n) = 2^n \int_0^\pi \prod_{k=1}^{n} \cos(x_k t)\, dt \tag{26}$$

then, using the identity $\cos 2x = 2\cos^2 x - 1$:

$$\psi(x_1, \ldots, 2x_m, \ldots, x_n) = 2^n \int_0^\pi \cos(2x_m t) \prod_{k\neq m}^{n} \cos(x_k t)\, dt \tag{27}$$

$$= 2^n \int_0^\pi \left[2\cos^2(x_m t) - 1\right] \prod_{k\neq m}^{n} \cos(x_k t)\, dt$$

$$\implies \psi(x_1, \ldots, x_m, \ldots, x_n) - \psi(x_1, \ldots, 2x_m, \ldots, x_n) = \tag{28}$$

$$2^{n+1} \int_0^\pi \cos^2(x_m t) \prod_{k\neq m}^{n} \cos(x_k t)\, dt = \psi(x_1, \ldots, x_m, \ldots, x_n, x_m)$$

and the result follows by derivation similar to Theorem 3.2. ☐

Conjecture 6.4. *For all even n, for all $\mathbf{x} \in \mathbb{N}^n$ the number of $\mathbf{x}$'s zero partitions is no more than the number of zero partitions of vector of size n with all its elements equal 1. Namely, never more than $\binom{n}{\frac{1}{2}n}$ zero partitions.*

Furthermore, for all odd n, for all $\mathbf{x} \in \mathbb{N}^n$ the number of $\mathbf{x}$'s zero partitions is no more than the number of zero partitions of vector of size n with all its elements equal 1 except one element that equals 2.

7 Estimating #SAT

The numbers produced by the reduction from #SAT to #PART have digits that does not exceed 4, and if using radix 6, they never even carry. Therefore the very same digits produced by the reduction can be interpreted in any radix larger than 5, being reduced to a different #PART problem. Still, it is guaranteed that the number of solution to those #PART problems are independent of the radix, as they're all reduced from the same #SAT problem. This property might be used to approximate #SAT using results as Theorem 3.3. We can obtain the number of partitions that their (shifted) size divides a given number N in polynomial time wrt n (the number of numbers to partition) and the number of digits of x_k. Nevertheless, it takes exponential time in the number of digits of N.

The probability that a there exists a partition with size divisible by a given prime p is roughly

$$\mathcal{P}\left[p|\left\langle \mathbf{x},\sigma\right\rangle\right] \approx 1-\left(1-\frac{1}{p}\right)^{2^n} \tag{29}$$

taking K reductions of #SAT and a set P of primes, the probability that in all of the there exists a partition with size divisible by a given prime p is therefore roughly

$$\begin{aligned}\prod_{p\in P}\prod_{k=1}^{K}\mathcal{P}\left[p|\left\langle \mathbf{x}_k,\sigma\right\rangle\right] &\approx \prod_{p\in P}\left[1-\left(1-\frac{1}{p}\right)^{2^n}\right]^K \\ &\leq \exp\left(-K\sum_{p\in P}\left(1-\frac{1}{p}\right)^{2^n}\right)\end{aligned} \tag{30}$$

recalling that for $x\in[0,1]$ we have $e^{-x}\geq 1-x$. This doesn't seem to be helpful since it seem to require many or large primes. However, if conjecture 6.4 is true, then we can bound our hueristic approximation with rather

$$\exp\left(-K\sum_{p\in P}\left(1-\frac{1}{p}\right)^{\binom{n}{\frac{1}{2}n}}\right) \tag{31}$$

8 Further Research

1. By using (11) we can get successive estimates to (12) by selecting e.g. primes $N = 2,3,5,\ldots$. We then could accelerate this sequence using Shanks, Romberg, Pade or similar sequence-acceleration method.

2. It is also interesting to consider a paper-and-pencil algorithm for calculating a single digit of the result of numbers that are given in the following form: all numbers have the form $1000\ldots0001$ so they're fully characterized by the number of zeros in the middle. The numbers are then given as naturals expressing the numbers of zeros, and we'd like to calculate the k'th digit of the result of the multiplication of all those numbers, as (9) and Theorem 4.1 suggest.

3. Detecting whether x_1, x_2 appear with different (resp. equal) signs in some zero partition can be done by examining the integral correlations such as

$$-\sin(x_1 t)\sin(x_2 t)\prod_{k=3}^{n}\cos(x_k t) = 2^{-n}\sum_{\sigma\in\{-1,1\}^n}\sigma_1\sigma_2\cos(t\langle\mathbf{x},\sigma\rangle) \tag{32}$$

Acknowledgments

Thanks to HunterMinerCrafter for many valuable discussions.

References

[1] Sipser, "Introduction to the Theory of Computation". International Thomson Publishing (1996).

[2] Kac, "Statistical Independence in Probability, Analysis and Number Theory". Carus Mathematical Monographs, No. 12, Wiley, New York (1959)

[3] Trefethen, Weideman, "The Exponentially Convergent Trapezoidal Rule" SIAM Review 08/2014; 56(3):385-458. DOI: 10.1137/130932132

A Appendix

Reduction of #SAT to #SUBSET-SUM Given variables $x_1, \ldots, x_l$ and clauses $c_1, \ldots, c_k$ and let natural $b \geq 6$. we construct a set S and a target t such that the resulted subset-sum problem requires finding a subset of S that sums to t. The number t is l ones followed by k 3s (i.e. of the form $1111\ldots3333$). S contains four groups of numbers $y_1, \ldots, y_l$, $z_1, \ldots, z_l$, $g_1, \ldots, g_k$, $h_1, \ldots, h_k$ where $g_i = h_i = b^{k-i}$, and y_i, z_i are b^{k+l-i} plus b^m for y_i if variable i appears positively in clause m, or for z_i if variable i appears negated in clause m. Then, every subset that sum to t matches to a satisfying assignment in the input CNF formula and vice versa, as proved in [1].

Reduction of #SUBSET-SUM to #PART Given S, t as before and denote by $s = \sum_{x\in S} x$ the sum of S members, the matching PART problem is $S \cup \{2s - t, s + t\}$. Here too all solutions to both problems are preserved by the reduction and can be translated in both directions.