P ≠ NP proof

Mario Stöckli
stoeckli.mario@gmail.com

July 2025

We prove $P \neq NP$ by contradiction by showing that there is no polynomial-time algorithm to solve the set partition problem.

1 Proof

1.1 The set partition problem

Given integers $a_1, \ldots, a_n$ and target c, determine whether there exist $s_i \in \{-1, 1\}$ such that

$$\sum_{i=1}^{n} a_i s_i = c.$$

This problem is in NP since, given witness s, the verifier $\sum_{i=1}^{n} a_i s_i - c$ is computable in $O(n)$.

1.2 Indistinguishable instance generation

Define generators G_1, G_2:
1. G_1: sample $a'_i \sim U(-2^n, 2^n]$, set $a_i = |a'_i|$, $c = \sum_i a'_i$.
2. G_2: sample $a'_i \sim U(-2^n, 2^n]$, set $a_i = |a'_i|$, $c = \sum_i a'_i + 2$.

Then every G_1 instance is satisfiable with exactly two solutions (w.h.p.) and G_2 instances are unsatisfiable (w.h.p). Without solving the NP problem, no polynomial-time algorithm can distinguish G_1 from G_2 with more than negligible advantage, since the two distributions are statistically indistinguishable.

Suppose $P = NP$. Then a polynomial-time solver S exists. Feeding instances from G_1, G_2 to S yields a distinguisher:

$$S(I) = \begin{cases} G_1, & \text{if a solution found,} \\ G_2, & \text{otherwise.} \end{cases}$$

This distinguisher succeeds, contradicting indistinguishability.

Thus,

$$P \neq NP. \quad \square$$

A Support and Total Variation Analysis

A.1 Distributions of X and Y (sum-only version)

For the purpose of analyzing the generators G_1 and G_2, define

$$X = \sum_{i=1}^{n} a_i \quad \text{under } G_1, \qquad Y = \sum_{i=1}^{n} a_i + 2 \quad \text{under } G_2,$$

where a_i are sampled independently from $U[0, 2^n]$.
The distributions are then

$$\mathcal{L}(X) : \Pr[X = t] = \Pr\Big[\sum_{i=1}^{n} a_i = t\Big], \qquad \mathcal{L}(Y) : \Pr[Y = t] = \Pr[X = t{-}2].$$

A.2 Support Preservation of G_2

For fixed $a_1, \ldots, a_n$, consider the lattice span

$$h(a) = \gcd(2a_1, 2a_2, \ldots, 2a_n).$$

The support of X lies in a single residue class modulo $h(a)$. Since $h(a) = 2{\cdot}\gcd(a_1, \ldots, a_n)$, a shift of $+2$ preserves the support iff $\gcd(a_1, \ldots, a_n) = 1$.
For a_i sampled uniformly at random from $\{0, \ldots, 2^n\}$,

$$\Pr[\gcd(a_1, \ldots, a_n) = 1] = \frac{1}{\zeta(n)} \to 1 \quad \text{as } n \to \infty.$$

Hence with overwhelming probability, $h(a) = 2$, and the shift $+2$ remains within the same support.

A.3 Total Variation Distance

By the central local limit theorem,

$$\mathrm{TV}\big(\mathcal{L}(X), \mathcal{L}(Y)\big) = \frac{1}{2} \sum_{k \in \mathbb{Z}} \big|\Pr[X = k] - \Pr[X = k - 2]\big| = O\Big(\frac{1}{\sigma\sqrt{n}}\Big).$$

As $\sigma = \Theta(2^n\sqrt{n})$, the total variation distance between the sums of generator G_1 and G_2 is negligible in n.

A.4 Conclusion: Statistical Indistinguishability

The distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ satisfy:

1. They are supported on the same lattice (span 2).

2. Their total variation distance obeys

$$\mathrm{TV}(\mathcal{L}(X), \mathcal{L}(Y)) \leq \mathrm{negl}(n).$$

Hence $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ are *statistically indistinguishable*. That is, for any distinguisher D (even unbounded),

$$\left| \Pr[D(X) = 1] - \Pr[D(Y) = 1] \right| \leq \mathrm{negl}(n).$$

Therefore, no algorithm (efficient or not) can distinguish G_1 from G_2 with more than negligible advantage.

B Solutions of the Generators

In the previous sections we have already sufficiently analyzed the magnitude of $\Pr[X = k]$ and similar statements.