Sparse Sets: Showing $SAT \le_{btt}^{\mathrm{p}} SS \rightarrow P = NP$ Using Chains
Exposition by William Gasarch

1 Definitions and Notation

Def 1.1 A set $S \subseteq \{0,1\}^*$ is *sparse* if there exists a polynomial s such that, for all n, $|S \cap \{0,1\}^{\le n}| \le s(n)$. (So the notion of sparse only applies to sets of strings over $\{0,1\}$.)

Notation 1.2 Let B be a set. Let $k \in \mathsf{N}$. Then $FML(k,B)$ is the set of all Boolean Combinations of k atoms of the form $x \in B$. Note that an element of $FML(k,B)$ is either TRUE or FALSE.

Example 1.3 Let $B \subseteq \{0,1\}^*$. Here is an example of an elements of $FML(2,B)$:

$$(0010 \in B) \wedge (00110010101 \notin B).$$

Here is an example of an element of $FML(3,B)$:

$$((00 \in B) \wedge (1100 \notin B)) \vee (00 \notin B \wedge 1111 \in B).$$

Notation 1.4 We will write an element of $FML(k,B)$ by writing the strings and then the formula with placeholders for the strings. For example, the two in the example will be written

$$((0010, 00110010101), BLAH_1 \in B \wedge BLAH_2 \notin B$$

$$(00, 1100, 1111), (BLAH_1 \in B \wedge (BLAH_2 \notin B) \vee (BLAH_1 \notin B \wedge BLAH_3 \in B).$$

Def 1.5 If A and B are sets and $k \in \mathsf{N}$ then $A \le_{k\text{-tt}}^{\mathrm{p}} B$ means that there is a polynomial time function $f : A \rightarrow FML(k,B)$ such that $x \in A$ iff $f(x)$ is TRUE.

Def 1.6 If A and B are sets and $k \in \mathsf{N}$ then $A \leq_{btt}^{\mathrm{p}} B$ means that there is a some k such that $A \leq_{k\text{-tt}}^{\mathrm{p}} B$.

Def 1.7 LSAT (called *Left Sat*) is the set of ordered pairs (ϕ, z) such that

1. ϕ is a Boolean formula. Let n be the number of variables.
2. $z \in \{0,1\}^n$ is viewed as an assignment.
3. There exists $x \preceq z$ such that $\phi(x)$.

Exercise 1

1. Prove that LSAT is in NP.
2. Prove that LSAT is NP-complete.

We will use the following fact, which we leave as exercises for the reader, in what follows.

Exercise 2 Let ϕ be a Boolean formula on n variables. Let $m \in \mathsf{N}$. Let $z \prec z'$ and $z_1 \prec \cdots \prec z_m \in \{0,1\}^n$.

1. Prove that if $(\phi, z) \in$ LSAT then $(\phi, z') \in$ LSAT.
2. Prove that one of the following occurs
   (a) For all i, $(\phi, z_i) \in$ LSAT.
   (b) For all i, $(\phi, z_i) \notin$ LSAT.
   (c) There exists i, $1 \leq i \leq m$ such that $(\phi, z_1), \ldots, (\phi, z_i) \notin$ LSAT and $(\phi, z_{i+1}), \ldots, (\phi, z_m) \in$ LSAT
   (d) From the above we can conclude that if for some i, $(\phi, z_i) \notin LSAT$ then $(\phi, z_1) \notin LSAT$.

Exercise 3 Show that if $A \leq_{btt}^{\mathrm{p}} B$ and $B \leq_{btt}^{\mathrm{p}} C$ then $A \leq_{btt}^{\mathrm{p}} C$.

Notation 1.8 If $n \in \mathsf{N}$ then $[n]$ is the set $\{1, \ldots, n\}$.

We will need this lemma several times. We leave the proof to the reader.

Lemma 1.9 *Let $L \in \mathsf{N}$. Let W be ANY set. If $w_1, w_2, \ldots, w_L$ are elements of W then either (1) there exists at least $\sqrt{L}$ that are identical, or (2)there exists at least $\sqrt{L}$ that are different.*

2 The 2-tt case

Theorem 2.1 *If there exists a sparse set S such that* SAT $\leq^{\mathrm{p}}_{2-tt}$ S *then* P = NP.

Proof:

Let S be $s(n)$ sparse and let the reduction be via f. Let $g(n) = s(f(n))$. In all that we deal with the number of strings in S will be bounded by $g(n)$.

We use the chain method.

Def 2.2 A *chain of length m* is a sequence of the form

- $((\phi, z_1), \vec{w}_1, \psi_1))$
- $((\phi, z_2), \vec{w}_2, \psi_2))$
- $\vdots$
- $((\phi, z_m,), \vec{w}_m, \psi_m))$

such that the following hold.

1. $z_1 > z_2 > \cdots > z_m$ in lex order.
2. For all j, k
$$(\phi, z_j) \in LSAT \text{ iff } (\phi, z_k) \in LSAT.$$
(Hence either all of the (ϕ, z_j) are in $LSAT$ or all the (ϕ, z_j) are not in $LSAT$.)
3. For all j, $f(\phi, z_j) = (\vec{w}_j, \psi_j)$. Note that hence $\vec{w}_j$ is a vector of length 2 and ψ_j is a boolean formula on two atoms of the form $w \in S$. (Given the last point, either all of the $\psi_j(\vec{w}_j)$ are TRUE or all of them are FALSE.
4. All of the $(\vec{w}_i, \psi_i)$ are DIFFERENT. (NOTE- its okay to have $\vec{w}_i = \vec{w}_j$ so long as $\psi_i \neq \psi_j$.)

Given $(\phi, \vec{z})$ we build a chain starting there. Our goal is *not* to find out if $(\phi, \vec{z}) \in LSAT$. Note that we can easily compute $(\vec{w}, \psi) \in FML(2, S)$ such that

$$(\phi, \vec{z}) \in LSAT \text{ iff } \psi(\vec{w}) = TRUE.$$

We WANT to obtain a $(\psi', \vec{w}') \in FML(1, S)$ such that

$$(\phi, \vec{z}) \in LSAT \text{ iff } \psi'(\vec{w}') = TRUE.$$

So essentially we show that if LSAT $\le^{\mathrm{p}}_{2-tt} S$ then LSAT $\le^{\mathrm{p}}_{1-tt} S$.

We will try to build a chain. One of two things must happen.

1. While trying to build it we will just find out if $(\phi, \vec{z}) \in LSAT$.
2. The chain will go all the way down to $(\phi, 0^n)$. Since it is easy to determine if $(\phi, 0^n) \in LSAT$ we will be done.
3. Let $h_1(n)$ be a function to be named later, though note that it will be polynomial. If the chain gets to be of length $h_1(n)$ then the following happens. First there are MANY that have the same ψ. Restrict to just those. So we are looking at a chain of length $h_2(n) = h_1(n)/16$. (Since there are 16 different possible ψ.) Since the ψ are all the same we have, for all $i \neq j$, $\vec{w}_i \neq \vec{w}_j$. View the sequence of elements of $FML(2, S)$ as follows: (Let $L = h_2(n)$ for ease of reading.)

   $$\begin{array}{ccccc} w_{11} & w_{12} & w_{13} & \cdots & w_{1L} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2L} \end{array}$$

   By Lemma 1.9 one of the two must occur: (1) there is a set $J \subseteq [L]$, $|J| \geq \sqrt{L}$, such that every element of $\{w_{1j} \mid j \in J\}$ is the same, or (2) there is a set $J \subseteq [L]$, $|J| \geq \sqrt{L}$, such that every element of $\{w_{1j} \mid j \in J\}$ is different.

   Restrict to this set J and renumber so that we have:

   $$\begin{array}{ccccc} w_{11} & w_{12} & w_{13} & \cdots & w_{1J} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2J} \end{array}$$

   Note that either all of the w_{1j} are the same or all are different.

By Lemma 1.9 one of the following must occur: (1) there is a set $J' \subseteq [J]$, $|J'| \geq \sqrt{J}$, such that every element of $\{w_{2j} \mid j \in J\}$ is the same, or (2) there is a set $J' \subseteq [J]$, $|J'| \geq \sqrt{J}$, such that every element of $\{w_{2j} \mid j \in J']$ is different.

Restrict to this set J' and renumber so that we have:

$$\begin{array}{ccccc} w_{11} & w_{12} & w_{13} & \cdots & w_{1J'} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2J'} \end{array}$$

We note a few things. First off, $|J'| \geq (h_1(n)/16)^{1/4}$. We use this later when we need to derive a bound for $h_1(n)$. Second it is not possible that option (1) happened both times since then you would have all of the $\vec{w}_j$ are the same when in fact you can't even have two that are the same. So either its (1) then (2), or (2) then (1), (2) then (2). The first two are symmetric so we just have two cases.

Case 1: (1) then (2). So we have

$$\begin{array}{ccccc} w & w & w & \cdots & w \\ w_1 & w_2 & w_3 & \cdots & w_{J'} \end{array}$$

Where all the w_i are different.

We take $h_1(n)$ such that $|J'| \geq h_1(n)/16)^{1/4} \geq g(n)+1$. Hence we need $h_1(n) \geq 16(g(n)+1)^4$. Since all of the w_i are different and there are $\geq g(n)$ of them, there must be one that is NOT in S. Hence for some i we have

$$(\phi, z_i) \in LSAT \text{ iff } \psi(w, F).$$

But recall that in the original chain we have that either they are ALL in $LSAT$ or NONE are in $LSAT$. So we now have

$$(\phi, 1^n) \in LSAT \text{ iff } \psi(w, F).$$

Note that $\psi(w, F) \in FML(1, S)$.

Case 2: (2) then (2). So we have

$$\begin{array}{ccccc} w_{11} & w_{12} & w_{13} & \cdots & w_{1J'} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2J'} \end{array}$$

Where every string in a row is different. Note that it is possible that there is a string in the first row that equals a string in the second row. This will not be a problem.

We would like to have that there is some column where BOTH are NOT in S. (We won't know which column that is.)

We need a simple combinatorial lemma.

Lemma 2.3 *Let w_{1j} and w_{2j} be strings. We write them as such:*

$$\begin{array}{ccccc} w_{11} & w_{12} & w_{13} & \cdots & w_{1J'} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2J'} \end{array}$$

Let S be a set of strings. If over 1/2 of the string in each row are NOT in S then there exists a column where both of the strings in it are not in S.

Proof: Assume, by way of contradiction, that every column has an element of S in it. Map column j to 1 if $w_{1j} \in S$ and to 2 if $w_{2j} \in S$. Either 1 or 2 is mapped to $\geq J'/2$ times. Assume its 1. Then there are $\geq J'/2$ elements in row 1 that are in S. This contradicts the premise. $\blacksquare$

Hence we take $|J'| \geq 3g(n)$. Unraveling it all we need

$|J'| \geq h_1(n)/16)^{1/4} \geq 3g(n)$ so we need $h_1(n) \geq 16(3g(n)^4$. This implies the prior lower bound on h_1 so we can take $h_1(n) = 16(3g(n))^4$.

Since there is a row where both strings are NOT in S we have

$$(\phi, 1^n) \in LSAT \text{ iff } \psi(F, F).$$

Here we know the status of $(\phi, \vec{z})$ in $LSAT$.

$\blacksquare$