MAJORITY-3SAT (and Related Problems) in Polynomial Time

\[ \cdots \quad \frac{1}{4} \quad \frac{1}{2} \quad 1 \]

Ryan Williams  

Shyan Akmal  

(appeared in FOCS’21)
**CNF-SAT**

Given a CNF formula $\phi$ over variables $x_1, \ldots, x_n$

**Question:** Is $\phi$ satisfiable?

**Equivalent Question:** Is $\Pr_{a \in \{0,1\}^n} [\phi(a) = 1] > 0$?

Is the fraction of satisfying assignments positive?

**#CNF-SAT**

Given a CNF formula $\phi$ over variables $x_1, \ldots, x_n$

**Question:** Compute number of SAT assigns to $\phi$

**Equivalent Question:** Compute $\Pr_{a \in \{0,1\}^n} [\phi(a) = 1]$

Determine the exact fraction of satisfying assignments

**Complexity:**
- NP-complete, even 3SAT
- #P-complete, even #2SAT
**MAJ-SAT**

Given a CNF formula $\phi$ over variables $x_1, \ldots, x_n$

**Question:** Is $\Pr_{a \in \{0,1\}^n}[\phi(a) = 1] \geq 1/2$?

**Equivalently:** Compute the most significant bit of the number of SAT assignments

---

**MAJ-kSAT**

Given a $k$-CNF formula $\phi$ over variables $x_1, \ldots, x_n$

**Question:** Compute MAJ-SAT for $\phi$

---

In some papers, the hardness of MAJ-3SAT and extensions was being *assumed* in order to show hardness for other problems...

[BDK’07] Computing if $\#\text{SAT}(\phi) \geq 2^{n/2}$ is PP-complete for 3-CNF $\phi$
It seems most people working in the area believed that MAJ-3SAT was PP-complete, and that we were just lacking a good reduction.

Unfortunately, the resulting decision problem MAJORITY-3SAT is not known to be PP-complete. In particular, the standard reduction from SAT to 3SAT [5] is not applicable here, as it requires the addition of “dummy” variables, which increase the number of possible assignments without necessarily increasing the number of satisfying ones: this can decrease the ratio of satisfying assignments over total assignments from above 1/2 to a value less than or equal to this threshold.

restrictions on k, we just write CNF). MAJSAT is PP-complete with respect to many-one reductions even if the input is restricted to be in CNF; however, it is not known whether MAJSAT is still PP-complete with respect to many-one reductions if the sentence φ is in 3CNF. Hence we will resort in proofs to a slightly different decision problem, following results by Bailey et al. [7]. The problem

Status of PP-completeness of MAJ3SAT

Asked 5 years, 8 months ago   Modified 11 months ago   Viewed 1k times

SHORT QUESTION: Is MAJ-3CNF a PP-complete problem under many-one reductions?
Theorem 1 (MAJ-3SAT is easy)

There is an algorithm which given a 3-CNF $\varphi$ decides if $\Pr[\varphi] \geq 1/2$ in linear time.

Theorem 2 (THR$_\rho$-kSAT is easy)

Fix any positive integer $k$ and rational $\rho \in (0, 1)$ with constant denominator. There is an algorithm which given a $k$-CNF $\varphi$ decides if $\Pr[\varphi] \geq \rho$ in linear time.

It turns out MAJ-kSAT is actually easy...
A Variant: Greater-Than-MAJ-SAT

| **MAJ-SAT**          | **GtMAJ-SAT**          | **GtMAJ-3SAT**          | **GtMAJ-4SAT**          |
|----------------------|------------------------|-------------------------|------------------------|
| Given a CNF formula \( \phi \), is \( \Pr[\phi] \geq 1/2 \)? | Given a CNF formula \( \phi \), is \( \Pr[\phi] > 1/2 \)? | Given a 3-CNF \( \phi \), is \( \Pr[\phi] > 1/2 \)? | Given a 4-CNF \( \phi \), is \( \Pr[\phi] > 1/2 \)? |
| PP-complete          | PP-complete            | We prove: P             | NP-complete!            |

We prove:
Greater-Than-MAJ-4SAT is NP-hard

Given a 3CNF $\phi$ on variables $x_1, \ldots, x_n$:
introduce a new variable $y$, and add $y$ to every clause of $\phi$

The new formula has strictly more than $\frac{1}{2}$ satisfying assignments if and only if $\phi$ is satisfiable!

So $\text{GtMAJ-4SAT}$ is NP-hard

It turns out there is also an NP verifier for this problem!

There is a huge difference between MAJ-4SAT and GtMAJ-4SAT (assuming $P \neq NP$)
Exists-MAJ-SAT

EMAJ-SAT

Given a CNF formula \( \phi(\vec{x}, \vec{y}) \) on vars \( \vec{x} \) and \( \vec{y} \), is \( \exists a \ [ \Pr_b [\phi(a, b)] \geq 1/2] \) true?

EMAJ-kSAT

Given a \( k \)-CNF formula \( \phi(\vec{x}, \vec{y}) \), is \( \exists a \ [ \Pr_b [\phi(a, b)] \geq 1/2] \) true?

NP\(^{PP} \)-complete

We prove:

\( P \) for \( k = 2 \)

NP-complete for \( k \geq 3 \)

Many other results!
Outline for the Rest

• Some Intuition
• MAJ-2SAT is Easy
• MAJ-3SAT is Easy
• Conclusion
Some Intuition...

**General CNFs**

A single clause may have a high fraction of SAT assignments

\[ \phi = (x_1 \lor \cdots \lor x_n) \]

\[ \Pr[\phi] = 1 - \frac{1}{2^n} \approx 1 \]

**2-CNFS**

A single clause already restricts the fraction considerably

\[ \phi = (x_a \lor x_b) \land \cdots \]

\[ \Pr[\phi] \leq \frac{3}{4} \]

Two “disjoint” clauses restrict the fraction further...

\[ \phi = (x_a \lor x_b) \land (x_c \lor x_d) \land \cdots \]

\[ a, b, c, d \text{ are distinct indices} \]

\[ \Pr[\phi] \leq \left(\frac{3}{4}\right)^2 < 0.57 \]
Some Intuition...

2-CNFs

Three “disjoint” clauses already restrict the fraction below $1/2$

$$\phi = (x_a \lor x_b) \land (x_c \lor x_d) \land (x_e \lor x_f) \land \cdots$$

$a, b, c, d, e, f$ are distinct

$$\Pr[\phi] \leq \left(\frac{3}{4}\right)^3 < 0.43$$

Completely analogous reasoning holds for $k$-CNFs!

If $\phi$ contains a variable-disjoint set of $t$ clauses of width $k$,

$$\Pr[\phi] \leq \left(1 - \frac{1}{2^k}\right)^t \leq e^{-\frac{t}{2^k}}$$

So let’s look for large sets of disjoint clauses! But if we can’t find them, we need to do something else...
MAJ-2SAT Algorithm

Given a 2-CNF \( \phi \), is \( \Pr_{a \in \{0,1\}^n}[\phi(a) = 1] \geq 1/2 \) ?

**Idea:** Search for Variable-Disjoint Clause Sets

\[ \phi = (x_1 \lor \neg x_2) \land (x_2 \lor \neg x_3) \land (x_4 \lor x_5) \land (x_1 \lor \neg x_6) \land \ldots \]

\( \{ \) Satisfied at most \( 3/4 \) of the time

\( \{ \) Implies that \( \Pr[\phi] \leq 3/4 \)

\( \{ \) An independent constraint

\( \{ \) Implies that \( \Pr[\phi] \leq (3/4)^2 \)
MAJ-2SAT Algorithm

Greedy Algorithm for Disjoint Sets:

Initialize $S := \emptyset$

Pass through the clauses one at a time

If clause $C$ is variable-disjoint from all of $S$, add $C$ to $S$

Produces maximal disjoint set $S$:

for all clauses $C'$ not in $S$, there is a clause $C$ in $S$ such that $C$ and $C'$ share at least one variable

\[
\varphi = (x_1 \lor \neg x_2) \land (x_2 \lor \neg x_3) \land (x_4 \lor x_5) \land (x_1 \lor \neg x_6) \land (x_4 \lor x_6) \land (x_7 \lor \neg x_2) \land (x_3 \lor x_7) \land (x_3 \lor x_4) \land \ldots
\]
MAJ-2SAT Algorithm

1. Run greedy algorithm for disjoint sets, get back a clause set $S$.
2. Suppose $|S| \geq 3$.
   Implies $\Pr[\varphi] \leq (3/4)^3 < 1/2$
   Return NO for MAJ-2SAT
3. Suppose $|S| < 2$... what to do, then?

Given a 2-CNF $\varphi$ is $\Pr[\varphi] \geq 1/2$?

\[ \varphi = (x_1 \lor \neg x_2) \land (x_2 \lor \neg x_3) \land (x_4 \lor x_5) \land (x_1 \lor \neg x_6) \land (x_4 \lor x_6) \land (x_7 \lor \neg x_2) \land (x_3 \lor x_7) \land (x_3 \lor x_4) \land \ldots \]
Fact: If $S$ is a maximal disjoint set, then the union of all variables in all clauses of $S$ forms a hitting set for all clauses in $\phi$.

Hitting set $H = \{x_1, x_2, x_4, x_5\}$

Consider any assignment $\alpha : H \rightarrow \{0, 1\}$.

This sets at least one variable in every clause, so the formula simplifies to a $1$-CNF.

Example: If $x_1, x_2, x_4 \mapsto 0$ and $x_5 \mapsto 1$ ...
MAJ-2SAT Algorithm: case of small $|S|$ 

It is easy to solve #SAT on 1-CNF!

If constraints are inconsistent: $\Pr[\varphi_\alpha] = 0$

If constraints are consistent, and $v$ distinct variables appear:

$$
\Pr[\varphi] = \sum_{\alpha:H \rightarrow \{0,1\}} \Pr[\varphi|_{\alpha}]
$$

$$
\Pr[\varphi_\alpha] = 1/2^v
$$

Idea: Enumerate all assignments to $H$ that satisfy the clauses they appear in, solve #1SAT on each subformula obtained. We’ll compute #SAT exactly in this case!

$\varphi_\alpha = \neg x_3 \land \neg x_6 \land x_8 \land x_7 \land x_9 \land \cdots$
MAJ-2SAT Algorithm

Given a 2-CNF $\varphi$ is $\Pr[\varphi] \geq 1/2$?

1. Run greedy algorithm for disjoint sets, get back a clause set $S$.
2. Suppose $|S| \geq 3$.
   Implies $\Pr[\varphi] \leq (3/4)^3 < 1/2$
   Return NO for MAJ-2SAT
3. Suppose $|S| \leq 2$.
   Try all SAT assignments to $S$, obtaining 1-CNFs. Solve #SAT on each of them to determine #SAT for the entire formula.

$$\varphi = (x_1 \lor \neg x_2) \land (x_2 \lor \neg x_3) \land (x_4 \lor x_5) \land (x_1 \lor \neg x_6) \land (x_4 \lor x_6) \land (x_7 \lor \neg x_2) \land (x_3 \lor x_7)$$

The same strategy works for all thresholds, not just $1/2$. 
Alternative Perspective: MAJ-2SAT

Every 2-CNF has one of two possible forms:

Random-Like

Has “Bad” Subformula

\[ \varphi = (x_1 \lor \neg x_2) \land (x_2 \lor \neg x_3) \land (x_4 \lor x_5) \land (x_1 \lor \neg x_6) \land (x_4 \lor x_6) \land (x_7 \lor \neg x_2) \land (x_3 \lor x_7) \land (x_3 \lor x_4) \land \cdots \]

Pr[\varphi] is small

Structured

Small Sum of 1-CNFs

\[ \Pr[\varphi] \text{ is easy to compute} \]
MAJ-3SAT Algorithm

Given a 3-CNF \( \varphi \) is \( \Pr[\varphi] \geq 1/2 \)?

\[ \varphi = (x_1 \lor \neg x_2 \lor x_7) \land (x_2 \lor \neg x_3 \lor x_6) \land (x_4 \lor x_5 \lor \neg x_8) \land (x_1 \lor \neg x_6 \lor x_5) \land \quad \{ \text{Satisfied at most } 7/8 \text{ of the time} \} \]

Implies that \( \Pr[\varphi] \leq 7/8 \)

If we find at least \( d \) disjoint clauses...

Implies that \( \Pr[\varphi] \leq (7/8)^d \)

For \( d \geq 6 \) we have \( \Pr[\varphi] \leq (7/8)^6 < 1/2 \) and we can report NO

What can we do when \( d < 6 \)?
As before, we get a “small” hitting set

Hitting set

\[ H = \{ x_1, x_2, x_3, x_4, x_6, x_7 \} \]

Any assignment \( \alpha : H \rightarrow \{0, 1\} \)

to \( \varphi \) induces a 2-CNF \( \varphi_\alpha \)

\[
\Pr[\varphi] = \sum_{\alpha:H\rightarrow\{0,1\}} \Pr[\varphi_\alpha]
\]

But now \( \Pr[\varphi_\alpha] \) is \#P-hard to compute...
Search for Disjoint Sets... Again!

Try all assignments $\alpha : H \rightarrow \{0, 1\}$

For each 2-CNF $\varphi_\alpha$, look for maximal disjoint set in $\varphi_\alpha$

Either (1) all these disjoint sets are “small”,
or (2) a disjoint set is “large”

If all are small, obtain 1-CNFs

In case (1), compute $\Pr[\varphi]$ exactly!
Picking out a Sunflower

Suppose $\varphi_\alpha$ has a disjoint set of size at least $d$ ...

$$\varphi_\alpha = (x_5 \lor \neg x_8) \land (x_9 \lor \neg x_{10}) \land (x_{11} \lor x_{12}) \land (x_{13} \lor \neg x_{14}) \land \ldots$$

$$H = \{x_1, x_2, x_3, x_4, x_6, x_7\}$$

$$\varphi = (l_1 \lor x_5 \lor \neg x_8) \land (l_2 \lor x_9 \lor \neg x_{10}) \land (l_3 \lor x_{11} \lor x_{12}) \land (l_4 \lor x_{13} \lor \neg x_{14}) \land \ldots$$
Picking out a Sunflower

Suppose $\varphi_\mathcal{A}$ has a disjoint set of size at least $d$ ...

Some literal $\ell$ from $H$ must appear many times...

$$\geq d/(2|H|)$$

...by the pigeonhole principle

We obtain a sunflower with core $\{\ell\}$
How to use the sunflower

If \( \ell \) appears in every clause, then \( \Pr[\varphi] \geq 1/2 \)

Otherwise, for \( s \geq 8 \),

\[
\Pr[\varphi] \leq \frac{1}{2} \cdot \left( \frac{7}{8} \right) + \frac{1}{2} \cdot \left( \frac{3}{4} \right)^s < \frac{1}{2}
\]

Different from \( \ell \)

\( \ell = 1 \) \quad \ell = 0 \)

MAJ-3SAT resolved in either case!
MAJ-3SAT Algorithm

1. Find a maximal disjoint set of clauses in $\varphi$

2. If disjoint set has size $\geq 6$, return NO (same as MAJ-2SAT)

3. Otherwise, find a hitting set $H$ of $\leq 18$ variables for $\varphi$

4. Try all SAT assignments to $H$, obtaining 2-CNFS

5. For each 2-CNFS, find a maximal disjoint set

6. If all disjoint sets are $\leq 7$, obtain 1-CNFS and compute $Pr[\varphi]$

7. Otherwise some disjoint set is $\geq 8$, yielding a “large” sunflower. If the sunflower core hits all clauses return YES, otherwise return NO
High-Level Intuition for MAJ-3SAT Algorithm

"Bad" Subformula

Big Disjoint Set

Sunflower + Extra

Sum of 1-CNFs

Just Sunflower

\[ \varphi = (x_1 \lor x_2 \lor x_7) \land (x_3 \lor x_4 \lor x_6) \land (x_5 \lor x_8 \lor x_9) \land (x_{10} \lor x_{13} \lor x_{14}) \land (x_{11} \lor x_{12} \lor x_{15}) \land (x_{16} \lor x_{17} \lor x_{18}) \land \ldots \]

\[ 
\ell \quad \ldots \\
\quad \ldots \\
\quad (\ell_1 \lor \ell_2 \lor \ell_3) \quad s \\
\quad \ldots \\
\]

\[ \varphi = \ell \quad \ldots \]

\( \ell \) appears in all clauses
Going Beyond MAJ-3SAT

MAJ-3SAT
3-CNF
\[ \Pr[\varphi] \geq 1/2 \]

MAJ-\(k\)SAT
\(k\)-CNF
\[ \Pr[\varphi] \geq 1/2 \]

THR\(\rho\)-\(k\)SAT
3-CNF
\[ \Pr[\varphi] \geq \rho \]

Extract More Disjoint Sets!
Extract More Sunflowers!
Conclusion

For 2-CNFs, either $\Pr[\varphi]$ is either “easy” to compute, or small.

For 3-CNFs, similar, but single literal may hit all clauses.

In general: testing $k$-CNFs at any constant threshold is “easy”

Some Future Directions:

What other problems have easy threshold versions?
Generalization to $k$-CSPs of domain $d \geq 3$?
Better parameterized algorithms?
(Terrible running time dependence on $k$)

Thank You!