Centrifugal Local Search

Chris Calabro

June 8, 2005

Abstract

We present a simple local search algorithm that starts at a solution α to a formula $\phi \in$ 3-CNF and flees from α in the hopes of reaching another solution β. We show that knowing the location of 1 solution α will be of significant help in finding a nearby solution β.

We then continue the theme of the 2CNF Code essay by using the above to upper bound the number of solutions to a $\phi \in$ 3-CNF whose solution space $\mathrm{sol}(\phi)$ is a distance-d code.

1 The Algorithm

The algorithm starts at a given assignment a and flips the given variable i_0. If it is not at a solution, then it picks an arbitrary false clause and flips a variable in it that disagrees with α, thus fleeing from α.

```
find_next(φ ∈ 3-CNF, α ∈ sol(φ), γ ∈ 2^n, i_0 ∈ [n])
  a ← γ
  flip a_{i_0}
  while φ(a) ≠ 1
    C ∈ false(φ, a)   // C is an arbitrary false clause
    S ← {j ∈ var(C) | a_j = α_j}    // since α ∈ sol(φ), |S| ≤ 2
    if S = ∅, return error
    i ←$ S
    flip a_i
  return a
```

2 Analysis

Let $\phi \in$ 3-CNF$, \beta \in \mathrm{sol}(\phi), \gamma \in 2^n, \gamma \subset \beta$. Say that β is $\subseteq$-minimal w.r.t. γ iff $\forall \delta \in \mathrm{sol}(\phi)\ (\gamma \subset \delta \subseteq \beta \rightarrow \delta = \beta)$.

Lemma 1. *Let* $\phi \in$ 3-CNF$, \alpha = 0, \beta \in \mathrm{sol}(\phi), \gamma \in 2^n, \gamma \subset \beta, i_0 \in \beta - \gamma, d = |\beta - \gamma|$*, and* β *is* $\subseteq$*-minimal w.r.t.* γ*. Then*

$$Pr(\mathrm{find\_next}(\phi, \alpha, \gamma, i_0) = \beta) \geq 2^{-(d-1)}.$$

Notice that the conclusion does not actually depend on $\alpha = 0$, but without this assumption the hypotheses would be more clumsy to state, requiring $\gamma \subset \beta$ and $\beta - \gamma$ to be replaced by $\gamma \triangle \alpha \subset \beta \triangle \alpha$ and $\beta \triangle \alpha - \gamma \triangle \alpha$.

Proof. We show the following invariant: at the beginning of step k, $|a| = |\gamma| + k$ and $Pr(\gamma \subset a \subseteq \beta) \geq 2^{-(k-1)}$. This is true for $k = 1$ since initially $a = \gamma \subset \beta$ and then we flip $i_0 \in \beta - \gamma$.

Now suppose that $|a| = k - 1$, $Pr(\gamma \subset a \subseteq \beta) \geq 2^{-(k-2)}$, and $\phi(a) \neq 1$. Let $C \in \text{false}(\phi, a), S = \{j \in \text{var}(C) \mid a_j = \alpha_j = 0\} = \text{var}(C) - a$. Then $|S| \leq 2$ since $\alpha \in \text{sol}(\phi)$.

Suppose $\gamma \subset a \subseteq \beta$. Suppose indirectly that $S \cap (\beta - a) = \emptyset$. Then

$$\begin{aligned}
&\emptyset = S \cap (\beta - a) = (\text{var}(C) - a) \cap (\beta - a) = (\text{var}(C) \cap \beta) - a \\
\Rightarrow &\text{var}(C) \cap \beta \subseteq a \\
\Rightarrow &\text{var}(C) \cap \beta \subseteq \text{var}(C) \cap a \subseteq \text{var}(C) \cap \beta \\
\Rightarrow &\text{var}(C) \cap a = \text{var}(C) \cap \beta \\
\Rightarrow &C \notin \text{false}(\phi, a),
\end{aligned}$$

which is a contradiction. So $S \cap (\beta - a) \neq \emptyset$. Note that this implies $S \neq \emptyset$ and so find_next does not return an error at this step. But also $Pr(j \in \beta - a) = \frac{|S \cap (\beta - a)|}{|S|} \geq \frac{1}{2}$. So the loop invariant is maintained. It is interesting to note that this probability is $\frac{1}{2}$ only in the case that C is a critical clause for ϕ at α.

If the condition of the loop is ever falsified so that $\phi(a) = 1$, then find_next returns a. But since β is $\subseteq$-minimal, if we have $\gamma \subset a \subseteq \beta$, then we also have $a = \beta$. Also the number k of steps satisfies $|\beta| = |a| = |\gamma| + k$, and so $k = |\beta| - |\gamma| = |\beta - \gamma| = d$. □

Corollary 2. *Let $\phi \in$ 3-CNF, $0 \in \text{sol}(\phi)$ and suppose $\forall \beta \in \text{sol}(\phi) - \{0\}$ $|\beta| \geq d$. Then the number of solutions at distance exactly d from 0 is $\leq n2^{d-1}$.*

Corollary 3. *Let $\phi \in$ 3-CNF, $0 \in \text{sol}(\phi)$ and suppose $\text{sol}(\phi)$ is a distance-d code. Then the number of solutions at distance $\leq kd$ from 0 is*

$$\leq 2\left(\frac{n}{d}\right)\left(\frac{n}{d} - 1\right) \cdots \left(\frac{n}{d} - (k-1)\right) 2^{k(d-1)},$$

provided $kd \leq \frac{1}{2}n$. In particular, the number of solutions at distance $\leq 2d$ is $\leq n^2 2^{2d}$.

Proof. Consider the following algorithm

```
find(φ ∈ 3-CNF, α ∈ sol(φ), B ≤ n)
  a ← α
  do B times
    i ←$ \overline{a△α}
    flip a_i
    a ← find_next(φ, α, a, i)
  return a
```

Fix a solution β with $|\beta| \leq kd$ and let $0 = \gamma_0 \subset \cdots \subset \gamma_l = \beta$ be a longest chain of solutions starting at 0 and ending at β. Then $l \leq k$ since $kd \geq |\beta| = |\gamma_l| \geq |\gamma_{l-1}| + d \geq \cdots \geq |\gamma_0| + ld = ld$. We now lower bound the probability that $\text{find}(\phi, 0, l) = \beta$.

The algorithm (if we look inside the code of find_next rather than just treating it like a black box) will iteratively change 0's to 1's, perhaps hitting solutions along the way. Each time a solution is not hit, find will make an at-most 2-way decision, one of which will bring *a* 1 closer to β (provided all previous decisions were made correctly). But each time a solution γ_{j-1} is hit, the next decision is to flip a random variable that currently is assigned 0. We will call this a *restart*. Any variable in $\gamma_j - \gamma_{j-1}$ will bring *a* 1 closer to γ_j, and there are at least d of them. Also there are at most $n - (j-1)d$ variables currently set to 0.

The probability of finding β is then minimized when the number of 2-way decisions is minimized and the number of restarts is maximized (this depends on $kd \leq \frac{1}{2}n$). Since the number of restarts is at most l, the probability of returning β is at least $(\frac{d}{n})(\frac{d}{n-d}) \cdots (\frac{d}{n-(l-1)d})2^{-l(d-1)}$. So the number of solutions β with $|\beta| \leq ld$ with longest chain of intermediate solutions of length l is at most $(\frac{n}{d})(\frac{n}{d}-1) \cdots (\frac{n}{d}-(l-1))2^{l(d-1)}$. Summing over all $l \leq k$ gives an upperbound on the number of solutions β with $|\beta| \leq kd$ of $2(\frac{n}{d})(\frac{n}{d}-1) \cdots (\frac{n}{d}-(k-1))2^{k(d-1)}$. □

Corollary 4. *Let $\phi \in$ 3-CNF, $\alpha \in 2^n$ and suppose that* $\text{sol}(\phi)$ *is a distance-d code. Then the number of solutions at distance $\leq d$ from α is $\leq n^2 2^d$.*

Proof. If there is no solution *a* within distance d of α, then we are done. Otherwise, we choose a solution *a* within distance d of α and apply corollary 3. □

We claim that this bound is tighter than if we simply counted the number of distance-d codewords that we can pack into the volume of a radius-d sphere or even onto the surface. These are both sphere-packing problems and yield bounds of $n^{O(d)}$. We justify this in the next section, but it will take some time.

3 An inferior bound

Lemma 5. *Suppose $\alpha, \beta \in 2^n, |\alpha| = |\beta| = d, |\alpha \triangle \beta| = d'$. Then $2 \mid d$ and $|\alpha - \beta| = |\beta - \alpha| = \frac{d}{2}$.*

Proof. Let $\gamma = \alpha \cap \beta$. Then

$$
\begin{aligned}
&|\alpha - \beta| + |\gamma| = |\alpha| = |\beta| = |\beta - \alpha| + |\gamma| \\
\Rightarrow &|\alpha - \beta| = |\beta - \alpha| \\
\Rightarrow &d = |\alpha \triangle \beta| = |\alpha - \beta| + |\beta - \alpha| = 2|\alpha - \beta|.
\end{aligned}
$$

□

Fix a point α on the surface of the sphere centered at 0 of radius d. The lemma tells us that the number of points β also on the sphere at distance d'

from α is 0 if d' is odd, and $\binom{d}{\frac{d'}{2}}\binom{n-d}{\frac{d'}{2}}$ if d' is even. So the number of points on the sphere at distance $< d$ from α is

$$\sum_{\substack{d'=0 \\ \text{step } 2}}^{d-1} \binom{d}{\frac{d'}{2}}\binom{n-d}{\frac{d'}{2}} = \sum_{d'=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{d}{d'}\binom{n-d}{d'}.$$

So a randomly chosen point β on the surface of the sphere has probability

$$\frac{\sum_{d'=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{d}{d'}\binom{n-d}{d'}}{\binom{n}{d}}$$

of being distance $< d$ from α. This is also the probability that 2 points chosen uniformly and independently on the surface of the sphere have distance $< d$. We will use this calculation in a moment.

Let $C = \{X_1, \dots, X_p\}$ be points chosen independently and uniformly on the surface of the sphere. Then

$$\begin{aligned}
&Pr(C \text{ is a distance-}d \text{ code}) \\
=&1 - Pr(\exists i, j \in [p], i \neq j \ |X_i \triangle X_j| < d) \\
\geq&1 - p^2 Pr(|X_i \triangle X_j| < d) \\
=&1 - p^2 \frac{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{d}{i}\binom{n-d}{i}}{\binom{n}{d}} \\
\geq&1 - 2p^2 \frac{\binom{d}{\frac{d}{2}}\binom{n-d}{\frac{d}{2}}}{\binom{n}{d}} \\
\geq&1 - 2p^2 \frac{2^d}{n^{\frac{d}{2}}} \quad \text{though this is a bit tedious to show,}
\end{aligned}$$

and this is > 0 iff $p < \frac{1}{\sqrt{2}}(\frac{\sqrt{n}}{2})^{\frac{d}{2}}$.

So there is a distance-d code of size $p \in n^{\Omega(d)}$ on the surface of the sphere. So any *upper* bound on the number of codewords that can fit on the surface of the sphere must be $n^{\Omega(d)}$.

Our previous bound for the number of solutions of ϕ at distance $\leq 2d$ from 0 when $\text{sol}(\phi)$ is a distance-d code was $n^2 2^{2d}$, which is superior. Oddly, this improves to $n2^d$ when we know that $0 \in \text{sol}(\phi)$.