Maximum 3-SAT as QUBO

Michael J. Dinneen[1]

Semester 2, 2016

[1] Slides mostly based on Alex Fowler's and Rong (Richard) Wang's notes.

Boolean Formula

- A *Boolean variable* is a variable that can take only the values $\textsc{True}{=}1$ or $\textsc{False}{=}0$.
- The *negation/NOT operator* of a Boolean variable x is $\overline{x} = 1 - x$.
- Binary *Boolean operators*: AND and OR are represented by the symbols $\wedge$ and $\vee$ respectively.
- A *literal* is a Boolean variable or its negation.
- A *Boolean formula* is an expression involving only Boolean literals, Boolean operators, and parentheses.
- A *clause* is a *disjunction* of literals (literals separated by the $\vee$ operator).
- A Boolean formula is in *conjunctive normal form* (CNF) if it is the *conjunction* of several clauses.
- It is called a *3CNF-formula* if all clauses contain 3 literals.

Examples of Boolean Formula Terminology

Let a, b, x, y, z be Boolean variables.

$(x \vee y \vee z)$ and $(a \vee \overline{b})$ are clauses.

$(x \vee y) \wedge (\overline{y} \wedge z \vee x)$ is a Boolean formula.

A 3CNF formula involving the Boolean variables x_1, x_2 and x_3 is

$$(x_1 \vee x_2 \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_1 \vee x_2 \vee \overline{x_3})$$

We choose to express a 3CNF formula ϕ featuring n variables and m clauses in the form:

$$\phi = C_1 \wedge C_2 ... \wedge C_m$$

where each

$$C_i = y_{i1} \vee y_{i2} \vee y_{i3}$$

and the y_{ij} are literals of the n variables. We usually label the variables $x_1, x_2, ..., x_n$.

We say that a clause is *satisfied* by an assignment $x = [x_1, ..., x_n] \in \mathbb{Z}_2^n$ if one of its literals takes the value true for this assignment. For example

$$(x_1 \vee \overline{x_2} \vee x_3)$$

is satisfied by the assignment $[x_1, x_2, x_3] = [1, 0, 1]$.

Maximum 3SAT problem

Problem (Maximum 3SAT problem)

Instance: *A 3CNF formula ϕ, involving $n \in \mathbb{N}$ variables and $m \in \mathbb{N}$ clauses.*

Question: *Find an assignment to the x_i which satisfies the maximum number of ϕ's clauses.*

In relation to the Maximum 3SAT problem for ϕ, for an assignment to the variables $x = [x_1, ..., x_n] \in \mathbb{Z}_2^n$, we define $\phi(x)$ to be the number of clauses satisfied by the assignment x.
Thus an instance can be expressed as the pseudo-Boolean optimization problem

$$\max_{x \in \mathbb{Z}_2^n} \phi(x) \qquad (1)$$

Reduce 3SAT to Independent Set

The previous best QUBO transformation for this problem was proposed by Lucas in 2013.

It reduces I into QUBO form by a 2 step process.

1. Reduce I into a Maximum Independent Set problem, using a well-known reduction (given on next slide).
2. Use the Maximum Independent Set QUBO formulation, which uses $n = |V|$ variables.

When this QUBO transformation is applied, the resultant QUBO formulation has $3m$ variables.

3SAT $\leq_m^P$ IndSet

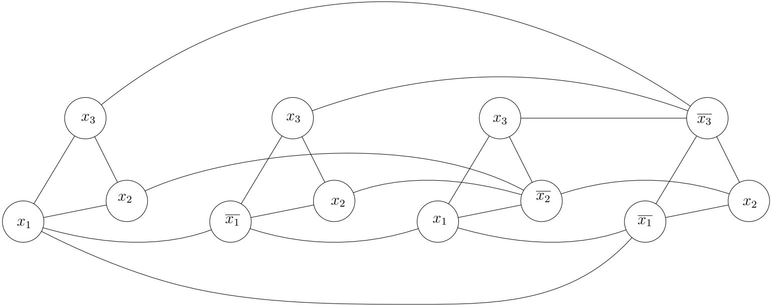
- Let ϕ be a conjunction of m clauses of 3CNF.
- Construct a graph G with $3m$ vertices that correspond to the literals in ϕ.
- For any clause in ϕ, connect the corresponding three vertices in G.
- Connect all pairs of vertices corresponding to a variable x and its negation $\overline{x}$.
- Now ϕ is satisfiable iff G has an independent set of size m.
- Furthermore, an independent set of size less than m in G corresponds to a subset of clauses of ϕ that can be satisfied.

Example Reduction

As an example, consider the Maximum 3SAT problem for

$$\phi = (x_1 \vee x_2 \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee \overline{x_3})$$

The corresponding Maximum Independent Set instance is:



Improved Direct Transformation

In our improved transformation F, $\mathrm{QUBO}(F(\phi))$ only requires $n+m$ variables. For each clause C_i, there is a variable w_i, while the original variables x_j also feature.

Letting $x = [x_1, \ldots, x_n]$ (respectively $w = [w_1, \ldots, w_m]$) represent assignments to the x_i (respectively w_j) variables, and

$$[x, w] = [x_1, \ldots, x_n, w_1, \ldots, w_m] \in \mathbb{Z}_2^{n+m}$$

$\mathrm{QUBO}(F(\phi))$ is the problem

$$\min_{[x,w]\in\mathbb{Z}_2^{n+m}} [x, w]^T F(\phi)[x, w] = \min_{[x,w]\in\mathbb{Z}_2^{n+m}} -g(x, w) - K_\phi \qquad (2)$$

where K_ϕ is a constant dependent on ϕ, and

$$g(x, w) = \sum_{i=1}^{m} C_i \text{ is satisfied by } x = \text{“number of satisfied clauses”} \qquad (3)$$

Firstly, each of ϕ's clauses $C_i = (y_{i1} \vee y_{i2} \vee y_{i3})$ is formulated as

$$C_i = y_{i1} + y_{i2} + y_{i3} - y_{i1}y_{i2} - y_{i1}y_{i3} - y_{i2}y_{i3} + y_{i1}y_{i2}y_{i3}$$

Thus $\phi(x)$ (the number of clauses satisfied in ϕ by an assignment x) can be expressed as the cubic pseudo-Boolean function

$$\phi(x) = \sum_{i=1}^{m} (y_{i1} + y_{i2} + y_{i3} - y_{i1}y_{i2} - y_{i1}y_{i3} - y_{i2}y_{i3} + y_{i1}y_{i2}y_{i3}) \tag{4}$$

Now by adding in an extra variable w_i, each $y_{i1}y_{i2}y_{i3}$ can be represented quadratically as

$$y_{i1}y_{i2}y_{i3} = \max_{w_i \in \mathbb{Z}_2} w_i(y_{i1} + y_{i2} + y_{i3} - 2)$$

Hence by substituting this representation for $y_{i1}y_{i2}y_{i3}$ into (4), we conclude that for **every** $x \in \mathbb{Z}_2^n$, (4) equals

$$\max_{w \in \mathbb{Z}_2^m} \sum_{i=1}^{m} ((1 + w_i)(y_{i1} + y_{i2} + y_{i3}) - y_{i1}y_{i2} - y_{i1}y_{i3} - y_{i2}y_{i3} - 2w_i) \tag{5}$$

Example Transformation (1/3)

As an example, take the Maximum 3SAT problem for

$$\phi = (x_1 \vee x_2 \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee \overline{x_3})$$

Since the variables of ϕ are x_1, x_2 and x_3, and ϕ has 4 clauses, the variables of $\mathrm{QUBO}(F(\phi))$ are x_1, x_2, x_3 and w_1, w_2, w_3, w_4.
From formula (3), $g(x, w) =$

$$\begin{aligned}
&= \sum_{i=1}^{m} ((1 + w_i)(y_{i1} + y_{i2} + y_{i3}) - y_{i1}y_{i2} - y_{i1}y_{i3} - y_{i2}y_{i3} - 2w_i) \\
&= (1 + w_1)(x_1 + x_2 + x_3) - x_1x_2 - x_1x_3 - x_2x_3 - 2w_1 \\
&+ (1 + w_2)((1 - x_1) + x_2 + x_3) - (1 - x_1)x_2 - (1 - x_1)x_3 - x_2x_3 - 2w_2 \\
&+ (1 + w_3)(x_1 + (1 - x_2) + x_3) - x_1(1 - x_2) - x_1x_3 - (1 - x_2)x_3 - 2w_3 \\
&+ (1 + w_4)((1 - x_1) + x_2 + (1 - x_3)) - (1 - x_1)x_2 - (1 - x_1)(1 - x_3) \\
&- x_2(1 - x_3) - 2w_4
\end{aligned}$$

Example Transformation (2/3)

Summing this out into its separate components we conclude

$$\begin{aligned}-g(x,w) = {} & -4 + 0x_1 - 2x_1x_2 + 2x_1x_3 - x_1w_1 + x_1w_2 - x_1w_3 + x_1w_4 \\ & + x_2 - 0x_2x_3 - x_2w_1 - x_2w_2 + x_2w_3 - x_2w_4 - x_3 - x_3w_1 \\ & - x_3w_2 - x_3w_3 + x_3w_4 + 2w_1 + 0(w_1w_2 + w_1w_3 + w_1w_4) \\ & + w_2 + 0(w_2w_3 + w_2w_4) + w_3 + 0w_3w_4 + 0w_4\end{aligned}$$

Letting $z = [x, w]$ (for readability), we present

$$-g(x,w) = K + z^T F(\phi) z = -4 + \sum_{1 \le i \le j \le 7} F(\phi)_{i,j} z_i z_j$$

Example Transformation (3/3)

The entries of $F(\phi)$ are

$$F(\phi) = \begin{bmatrix} 0 & -2 & 2 & -1 & 1 & -1 & 1 \\ 0 & 1 & 0 & -1 & -1 & 1 & -1 \\ 0 & 0 & -1 & -1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and QUBO$(F(\phi))$ is the problem

$$\min_{[x,w]\in\mathbb{Z}_2^7} [x, w]^T F(\phi)[x, w]$$

Conclusion

For an instance ϕ for the MAX 3SAT problem, we usually have the number of variables n being less than the number of clauses m.

Observation

Thus, our second direct approach will generally use at least 33% less variables than the number of variables for the reduction to the Maximum Independent Set approach.

To see this compare $n + m$ with $3m$.

Some Final Facts about MAX 3SAT

Theorem

The expected number of clauses satisfied by a random assignment to a 3SAT instance (with all clauses different) is within an approximation factor 7/8 of optimal.

Proof.

The probability of a clause not being satisfied is $\left(\frac{1}{2}\right)^3 = 1/8$. Using linearity of expectation we expect $\left(\frac{7}{8}\right)m$ to be true. □

Corollary

For every instance of 3SAT, there is a truth assignment that satisfies at least $\frac{7}{8}m$ clauses.

Application: Every instance of 3SAT with at most 7 clauses is satisfiable.