(67686) Mathematical Foundations of AI June 4, 2008

Lecture 4

Lecturer: Ariel D. Procaccia *Scribe: Zvi Vlodavsky and Bracha Hod*

1 Constraint Satisfaction Problems

Constraint satisfaction problems (CSPs) are mathematical problems where one must find states or objects that satisfy a number of constraints or criteria.

Definition 1 *A constraint satisfaction problem is defined a triple $\langle X, D, C\rangle$, where X is a set of variables, D is a domain of values, and C is a set of constraints $C_1(S_1)..C_n(S_n)$ where each S_i is a set of variables. A constraint C_i is a combination of valid values for the variables S_i. A solution to the CSP is an assignment of values to $S_1..S_n$ that satisfies all constraints.*

Example 1 The eight queens puzzle is the problem of putting eight chess queens on an 8×8 chessboard, such that none of them is able to capture any other using the standard chess queen's moves. The queens must be placed in such a way that no two queens would be able to attack each other. Thus, a solution requires that no two queens share the same row, column, or diagonal.
Variables $x_1..x_8$ represent the location of the queens.
$D = (1, a)..(8, h)$. An example constraint for queens 1 and 2 is: $C_{1,2} = \langle(1, a),(2, c)\rangle, \langle(1, a),(2, d)\rangle, \langle(1, b),(2, d)\rangle....$

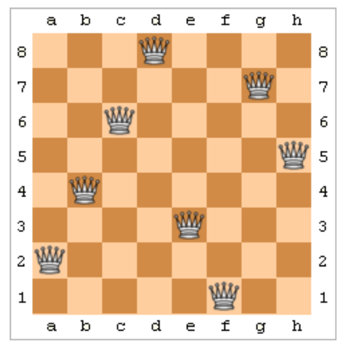

Figure 1: Example to a solution for the eight queens puzzle.

Example 2 Satisfiability (SAT) is the problem of determining if the variables of a given Boolean formula can be assigned in such a way as to make the formula evaluate to TRUE. SAT is a classic example for a CSP. Consider, for example, the formula $(x_1 \vee \bar{x}_2) \wedge (x_2 \vee x_3)$. The domain D is $\{0, 1\}$ for FALSE and TRUE, and the constraints are:
$C_1(x_1, x_2) = \langle 1, 0\rangle, \langle 1, 1\rangle, \langle 0, 0\rangle$; $C_2(x_2, x_3) = \langle 1, 1\rangle, \langle 0, 1\rangle, \langle 1, 0\rangle$.

Solving a CSP on a finite domain is an NP-complete problem in general, because 3-SAT is NP-complete.

2 Tractability via Local Properties

In some specific cases we can use local properties of CSP in order to efficiently solve the problem. Dechter [1] presents a relationship between the sizes of the variables' domains, the number of variables in each constraint and the level of local consistency sufficient to ensure global consistency. We describe here a few notable definitions and results.

Definition 2 *Let $X = (X_1, ..., X_n)$ be a set of variables in a CSP, and let $X' = (X'_1, ..., X'_i)$ be some subset of variables from X. A partial instantiation of variables $\langle X'_1 = x_1, ..., X'_i = x_i\rangle$ is* ***locally consistent*** *if it satisfies all the constraints in X'.*

Definition 3 *A **globally consistent** CSP is one in which any locally consistent partial instantiation of variables can be extended to a consistent full instantiation.*

Note that globally consistent CSPs have the property that a solution can be found without backtracking. There are CSPs that do not have the globally-consistent property, e.g. the eight queen puzzle.

Definition 4 *A CSP is said to be **i-consistent** if any locally consistent partial instantiation of $i-1$ variables, can be extended by any i-th variable to a consistent instantiation.*

Definition 5 *A CSP is **strong i-consistent** if it is k-consistent for every $k \leq i$.*

Theorem 6 *Let there be a CSP with $|D| = d$ and arity r (each constraint having at most r variables). If it is strong $(d(r-1)+1)$-consistent, then it is globally consistent [1].*

Proof For simplicity we provide the proof for the special case of $r = 2$.

We will prove the theorem by showing that strong $(d+1)$-consistent binary CSPs are $(d+i+1)$-consistent for any $i \geq 1$.

According to the definitions, we need to show that if $\bar{x} = (x_1, ...x_{d+i})$ is any locally consistent subtuple of the subset of variables $\{X_1, ..., X_{d+i}\}$, and if X_{d+i+1} is any additional variable, then there is an assignment x_{d+i+1} to X_{d+i+1} that is consistent with $\bar{x}$.

We call an assignment to a single variable a unary assignment and we view $\bar{x}$ as a set of such unary assignments. With each value $j \in D$ we associate a subset A_j that contains all unary assignments in $\bar{x}$ that are consistent with the assignment $X_{d+i+1} = j$. Since variable X_{d+i+1} may take on d possible values $1, 2, ...d$ this results in d such subsets, $A_1, ..., A_d$.

We claim that there must be at least one set, say A_1, that contains the set $\bar{x}$. If this were not the case, each subset A_j would be missing some member, say x'_j, which means that the tuple generated by taking a missing unary assignment from each of the A_j's , i.e. $\bar{x}' = (x'_1, x'_2, ..., x'_d)$ whose length is d or less (there might be repetitions), could not possibly be consistent with any of X_{d+i+1}'s values.

This leads to a contradiction because as a subset of $\bar{x}$, $\bar{x}'$ is locally consistent, and from the assumption of strong $(d+1)$-consistency, this tuple should be extensible by any additional variable including X_{d+i+1}.

Note that we need not assume that the x'_i's are distinct unary assignments because strong $(d+1)$-consistency renders the argument applicable to subtuples $\bar{x}'$ of length less than k.

We found a subset, without loss of generality A_1, that contains the set $\bar{x}$. From the definition of A_1, it is consistent with $X_{d+i+1} = 1$. Hence, we found a value consistent with $\bar{x}$. ■

Theorem 7 *Let there be a CSP with arity r. Let t be an upper bound on the number of constraints each variable appears in. Let q be a lower bound on the probability of choosing a satisfying assignment for a constraint.*
If $q \geq 1 - \frac{1}{e(r(t-1)+1)}$ then there is a solution to the CSP.

Proof Let there be a random assignment of variables. $\mathcal{E}_i$ is the event of C_i not being satisfied. Since a constraint has at least a q probability of being satisfied, $Pr[\mathcal{E}_i] \leq 1-q$. Since a constraint

has at most r variables, each appearing in at most $(t-1)$ constraints, $\mathcal{E}_i$ is independent of all other events except for at most $r(t-1)$ events.
According to The Lovász Local Lemma, that is described below, by assigning $p = 1-q$, $m = r(t-1)$, if $e(1-q)(r(t-1)+1) \leq 1$ then $Pr[\bigcap_{i=1}^{n} \bar{\mathcal{E}}_i] > 0$. Hence, there is a solution to the CSP. ■

3 The Lovász Local Lemma

The Lovász Local Lemma [2] is mostly used in probabilistic methods to give existence proofs. The main idea of this lemma is that as long as the events are "mostly" independent from one another and aren't individually too likely, then there will still be a positive probability that none of them occurs.

Lemma 8 *We denote by $\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n$ the series of events such that each event occurs with probability at most p and such that each event is independent of all the other events except for at most m of them. If $ep(m+1) \leq 1$ (where e = 2.718...), then there is a nonzero probability that none of the events occur, $Pr[\bigcap_{i=1}^{n} \bar{\mathcal{E}}_i] > 0$.*

Proof Let S denote a subset of the indices from $1..n$. It is enough to show that, for any S and for any i such that $i \notin S$

$$Pr[\mathcal{E}_i | \bigcap_{j \in S} \bar{\mathcal{E}}_j] \leq \frac{1}{1+m}. \tag{1}$$

If Equation (1) is true, and by using the chain rule:

$$\begin{aligned} Pr[\bigcap_{i=1}^{n} \bar{\mathcal{E}}_i] &= Pr[\bar{\mathcal{E}}_1] \cdot Pr[\bar{\mathcal{E}}_2 | \bar{\mathcal{E}}_1] \cdots Pr[\bar{\mathcal{E}}_n | \bigcap_{i=1}^{n-1} \bar{\mathcal{E}}_i] \\ &= (1 - Pr[\mathcal{E}_1]) \cdot (1 - Pr[\mathcal{E}_2 | \bar{\mathcal{E}}_1]) \cdots (1 - Pr[\mathcal{E}_n | \bigcap_{i=1}^{n-1} \bar{\mathcal{E}}_i]) \\ &\geq (1 - \frac{1}{1+m})^n \\ &> 0. \end{aligned} \tag{2}$$

Equation (1) can be verified by Induction on $|S| = n$.
The base case is when $S = \emptyset$. Since $Pr[\mathcal{E}_i] \leq p$ and $ep(m+1) \leq 1$ we get that:

$$Pr[\mathcal{E}_i] \leq \frac{1}{e(1+m)} \leq \frac{1}{1+m}. \tag{3}$$

Inductive step: Given a set S and $i \notin S$, let $S_2 \subseteq S$ be a set of all the events that $\mathcal{E}_i$ is independent of. $S_1 = S \setminus S_2$ is denoted the set that contains the remaining events in S.
By definition of conditional probability:

$$Pr[\mathcal{E}_i | \bigcap_{j \in S} \bar{\mathcal{E}}_j] = \frac{Pr[\mathcal{E}_i \cap (\bigcap_{j \in S_1} \bar{\mathcal{E}}_j) | \bigcap_{k \in S_2} \bar{\mathcal{E}}_k]}{Pr[\bigcap_{j \in S_1} \bar{\mathcal{E}}_j | \bigcap_{k \in S_2} \bar{\mathcal{E}}_k]} \tag{4}$$

Since $\mathcal{E}_i$ is mutually independent of $\{\mathcal{E}_k | k \in S_2\}$, the numerator can be expressed by:

$$Pr[\mathcal{E}_i \cap (\bigcap_{j \in S_1} \bar{\mathcal{E}}_j) | \bigcap_{k \in S_2} \bar{\mathcal{E}}_k] \leq Pr[\mathcal{E}_i | \bigcap_{k \in S_2} \bar{\mathcal{E}}_k] = Pr[\mathcal{E}_i] \leq p. \quad (5)$$

Suppose that $S_1 = \{j_1..j_r\}$ where $r \leq m$. By the induction hypothesis, and using the chain rule, the denominator is:

$$\begin{aligned} Pr[\bigcap_{j \in S_1} \bar{\mathcal{E}}_j | \bigcap_{k \in S_2} \bar{\mathcal{E}}_k] &= (1 - Pr[\mathcal{E}_{j_1} | \bigcap_{k \in S_2} \bar{\mathcal{E}}_k]) \cdots (1 - Pr[\mathcal{E}_{j_r} | \bar{\mathcal{E}}_{j_1} \cap \cdots \cap \bar{\mathcal{E}}_{j_{r-1}} \bigcap_{k \in S_2} \bar{\mathcal{E}}_k]) \\ &\geq (1 - \frac{1}{1+m}) \cdots (1 - \frac{1}{1+m}) \\ &\geq (\frac{1}{1+m})^m. \end{aligned} \quad (6)$$

The last transition derived from the fact that $r \leq m$.

Using previous assumption of $ep(m+1) \leq 1$, as well as the results from Equations (5) and (6), it follows that:

$$\begin{aligned} Pr[\mathcal{E}_i | \bigcap_{j \in S} \bar{\mathcal{E}}_j] &\leq \frac{p}{(1 - \frac{1}{1+m})^m} \\ &\leq \frac{1}{(1+m)e(1 - \frac{1}{1+m})^m} \\ &\leq \frac{1}{1+m}. \end{aligned} \quad (7)$$

In the last transition, we used the fact that $(1 - \frac{1}{1+m})^m \geq \frac{1}{e}$. ■

4 References

[1] R. Dechter, From local to global consistency, Artificial Intelligence 55 87-107, 1992.

[2] N. Alon and J. H. Spencer, The Probabilistic Method, Wiley, 2000.