

Automating the Axiomatic Analysis of Matching Mechanisms with SAT Solving Technology

Ulle Endriss

Institute for Logic, Language and Computation

University of Amsterdam

Talk Outline

I will try to demonstrate how the AI technique of *SAT solving* can be used for the *axiomatic analysis* of *matching mechanisms*.

- Model: one-to-one matching
- Preservation Theorem for axioms expressed in a formal language
- Approach to proving impossibility theorems via SAT solving
- Application: two impossibility theorems for matching

The Model: One-to-One Matching

Two groups of *agents*: $A_n = \{a_1, \dots, a_n\}$ and $B_n = \{b_1, \dots, b_n\}$.

Each agent *rank*s all the agents on the opposite side of the market.

Need *mechanism* to return one-to-one *matching* given such a *profile*.

Examples: job markets, marriage markets, ...

Would like a mechanism with good normative properties (*axioms*):

- *Stability*: no a_i and b_j prefer one another over assigned partners
- *Strategyproofness*: best strategy is to truthfully report preferences
- *Fairness*: (for example) no advantage for one side of the market

Gale-Shapley (1962): stable (✓); strategyproof for left side (✓) but not right side (✗) of the market; unfair advantage for left side (✗).

D. Gale and L.S. Shapley. College Admissions and the Stability of Marriage. *American Mathematical Monthly*, 69:9–15, 1962.

Formal Language for Axioms

Would like to have formal language with clear semantics (i.e., a logic) to express axioms, to be able to get results for entire families of axioms.

First-order logic with *sorts*, one for *profiles* and one for agent *indices*, with these basic ingredients:

- $p \triangleright (i, j)$ — in profile p , agents a_i and b_j will get matched
- $j \succ_{p,i}^A j'$ — in profile p , agent a_i prefers b_j to $b_{j'}$ (also for B)
- $\text{top}_{p,i}^A = j$ — in profile p , agent a_i most prefers b_j (also for B)
- $p \sim_i^A p'$ — profiles p and p' are a_i -variants (also for B)
- $p \rightleftarrows p'$ — swapping sides in profile p yields profile p'
- \forall_P / \exists_P and \forall_N / \exists_N — quantifiers for variables of two sorts

Example

$$\forall_P p. \forall_P p'. \forall_N i. \forall_N j. \forall_N j'. [(j \succ_{p,i}^A j' \wedge p \sim_i^A p') \rightarrow \neg(p \triangleright (i, j') \wedge p' \triangleright (i, j))]$$

The Preservation Theorem

Call a mechanism *top-stable* if it always matches all mutual favourites.

Call an axiom *universal* if it can be written in the form $\forall \vec{x}.\varphi(\vec{x})$.

Theorem 1 *Let μ^+ be a top-stable mechanism of dimension n that satisfies a given set Φ of universal axioms. If $n > 1$, then there also exists a top-stable mechanism μ of dimension $n - 1$ that satisfies Φ .*

Proof idea: Construct larger profile in which extra agents most prefer each other and are least liked by everybody else.

Corollary: enough to prove impossibility theorems for smallest n !

Counterexample

Preservation Theorem might look trivial. *Doesn't this always hold?*

No: some axioms we can satisfy for large but not for small domains.

Suppose we want to design a mechanism under which at least one agent in each group gets assigned to their most preferred partner:

$$\forall_P p. \exists_N i. \forall_N j. [(top_{p,i}^A = j) \rightarrow (p \triangleright (i, j))] \wedge$$

$$\forall_P p. \exists_N j. \forall_N i. [(top_{p,j}^B = i) \rightarrow (p \triangleright (i, j))]$$

This is *not* universal! Mechanism exists for $n = 3$ but not for $n = 2$.

Proving Impossibility Theorems

Suppose we want to prove an impossibility theorem of this form:

“for $n \geq k$, no matching mechanism satisfies all the axioms in Φ ”

Our Preservation Theorem permits us to proceed as follows:

- Check all axioms in Φ indeed are universal axioms.
- Check Φ includes (or implies) top-stability.
- Express all axioms for special case of $n = k$ in *propositional CNF*.
- Using a *SAT solver*, confirm that this CNF is unsatisfiable.
- Using an *MUS extractor*, find a short proof of unsatisfiability.

For example, *stability* for $n = 3$ can be expressed in CNF like this:

$$\bigwedge_{p \in B_3!^3 \times A_3!^3} \bigwedge_{i \in \{1,2,3\}} \bigwedge_{j \in \{1,2,3\}} \bigwedge_{\substack{i' \text{ s.t. } p \text{ has} \\ a_i \succ b_j a_{i'}}} \bigwedge_{\substack{j' \text{ s.t. } p \text{ has} \\ b_j \succ a_i b_{j'}}} (\neg x_{p \triangleright (i,j')} \vee \neg x_{p \triangleright (i',j)})$$

Remark: This is a conjunction of *419,904 clauses* (big, yet manageable).

Application: A Variant of Roth's Theorem

Recall this classic result:

Theorem 2 (Roth, 1982) *For $n \geq 2$, no matching mechanism for incomplete preferences is both stable and two-way strategyproof.*

Remark: In our model (with complete preferences) true only for $n \geq 3$.

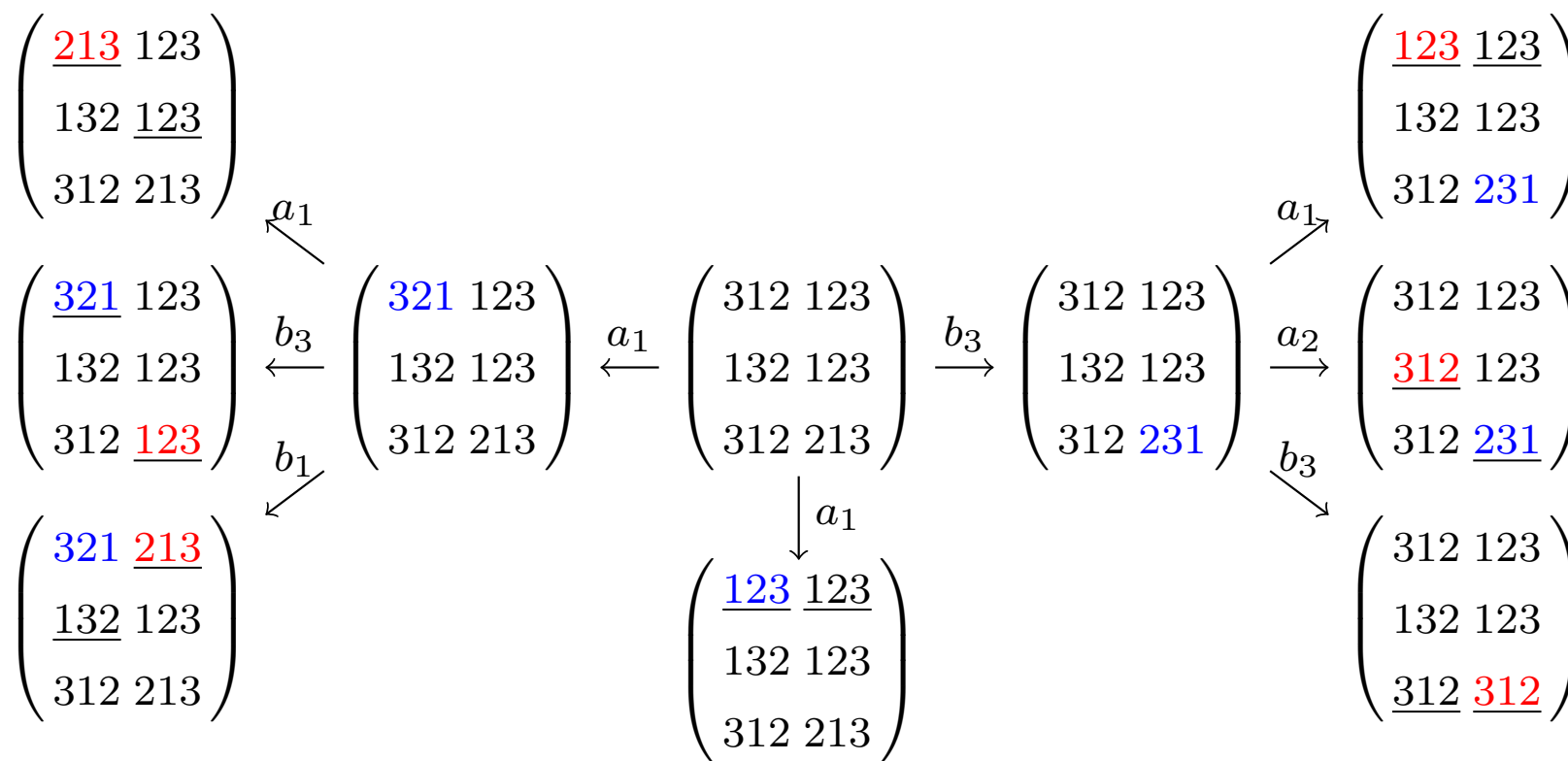
We can use our approach to prove this stronger variant:

Theorem 3 *For $n \geq 3$, no matching mechanism is both top-stable and two-way strategyproof (even in our model).*

By the Preservation Theorem, we are done if the claim holds for $n = 3$. SAT solver says it does, and MUS provides human-readable proof (\hookrightarrow).

A.E. Roth. The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7:617–628, 1982.

Proof of Base Case



Application: Stability vs. Gender-Indifference

Call a mechanism *gender-indifferent* if swapping the two sides of the market (“genders”) yields the corresponding swap in the outcome:

$$\forall_P p. \forall_P p'. \forall_N i. \forall_N j. [(p \rightleftharpoons p') \rightarrow (p \triangleright (i, j) \rightarrow p' \triangleright (j, i))]$$

Bad news:

Theorem 4 *For $n \geq 3$, there exists no matching mechanism that is both *stable* and *gender-indifferent*.*

Here the MUS extractor finds a particularly simple proof: it identifies a “swap-symmetric” profile for which there exists no admissible outcome (two matchings are ruled out by G-I and the other four by stability).

F. Masarani and S.S. Gokturk. On the Existence of Fair Matching Algorithms. *Theory and Decision*, 26(3):305–322, 1989.

Last Slide

By the *Preservation Theorem*, for top-stable mechanisms and universal axioms, proving impossibilities can be automated. Specific results:

- Impossible to get *top-stability* and *two-way strategyproofness*.
- Impossible to get *stability* and *gender-indifference*.

Future potential of SAT for economic theory *beyond impossibilities*: independence of axioms, designing mechanisms, justification, . . .

Paper and full code available here: tinyurl.com/satmatching

U. Endriss. Analysis of One-to-One Matching Mechanisms via SAT Solving: Impossibilities for Universal Axioms. Proc. AAI-2020.