Methods for Analysing Infinite-domain Constraint Satisfaction Problems

Peter Jonsson (peter.jonsson@liu.se)
Department of Computer and Information Science (IDA)
Linköping University
Sweden

Constraint Satisfaction Problems (CSP) are a well-known and important class of computational problems. An instance of the constraint satisfaction problem consists of a set of variables, a set of possible variable values (known as the *domain*), and a set of constraints which impose restrictions on value assignments to the variables. To solve such a problem, a value has to be found for each variable so that all the constraints are satisfied. Many problems encountered in computer science, artificial intelligence, and mathematics can be viewed as instances of the CSP: examples can be found in, for instance, spatiotemporal reasoning, computer vision, machine learning, scheduling, bioinformatics, and graph theory. CSPs are additionally the basis for many optimisation problems, counting problems, etc. One of the most prominent complexity-theoretic questions concerning CSPs is the following: given a set of relations Γ (referred to as a *constraint language*), what is the complexity of CSP(Γ), i.e. the CSP where only the relations in Γ are allowed in constraints. CSP(Γ) and mathematical methods for studying this problem is the focal point of this project.

CSPs over finite domains are extremely well-studied in the literature. A major breakthrough took place in 2017 when Bulatov and Zhuk proved (independently) the *dichotomy conjecture*: a problem CSP(Γ) is either polynomial-time solvable or NP-complete. This result was based on a thorough understanding of the mathematics behind finite-domain CSPs. In particular, various results from algebra were the key for obtaining this result, and it has been argued that the recent vitalisation of universal algebra is an effect of interplay with CSP research.

If we turn our attention to infinite-domain CSPs, then there is still an abundance of interesting applications e.g. in spatiotemporal reasoning, computer vision, computational linguistics, and bioinformatics. However, there cannot exist a dichotomy result since an infinite number of complexity classes is needed for capturing the complexity of CSP(Γ). This is one of several reasons for infinite-domain CSPs behaving very differently compared to finite-domain CSPs, implying that we need an extended mathematical toolbox for analysing infinite-domains CSPs. During the last few years, there has been a rapid development of new tools for studying infinite-domain CSPs. These tools have been based on ideas from various mathematical fields such as universal algebra, mathematical logic/model theory, graph theory, and Ramsey theory. This has led to a number of full complexity classification results for restricted classes of CSPs, such as certain temporal constraints, graph-based constraints, phylogeny constraints, and others.

The main goal when analysing the complexity of CSPs has historically been to separate NP-hard cases from polynomial-time cases. From both a theoretical and practical point of view, it is additionally highly interesting to analyse the time complexity of algorithms for the NP-hard cases. There is a substantial body of such results for finite-domain CSPs while our understanding of infinite-domain CSPs is comparatively weak. Hence, this project is directed towards mathematical methods for studying computational aspects of infinite-domain CSPs in a broad sense: we want to both separate polynomial-time solvable cases and NP-hard cases, and analyse the time complexity of the NP-hard cases. We expect strong cross-fertilisation effects from such an approach, and that it will lead to innovative methods for studying CSPs.

It is convenient to separate the project description into two parts based on the mathematical properties of the constraint languages at hand. The balance between work on Part 1 and Part 2 below will be up to discussion with the PhD student.

Part 1. Here, we consider ω-categorical constraint languages, i.e. when all countable models of the first-order theory of Γ are isomorphic. From the AI perspctive, it is very common that *qualitative* CSPs are ω-categorical, and there is a large number of CSPs for qualitative reasoning described in the literature. Understanding and exploiting the ω-categoricity of qualitative CSPs has been one of the keys for obtaining powerful complexity results. This part has a large degree of built-in flexibility: one may approach ω-categorical language from many different angles, ranging from purely mathematical questions (concerning, for instance, clone theory) via questions concerning computational complexity to (fairly practical) questions concerning concrete algorithms for particular subproblems. The common denominator for all these approaches is that interesting mathematics enter the stage in a natural way. A concrete starting point would be to analyse computational aspects of the RCC-5 formalism (by Cohn et al.) for spatial reasoning: not much is known concerning time complexity of algorithms nor the computational complexity of first-order definable fragments. Despite its apparent simplicity, RCC-5 has resisted deeper analyses and this can be attributed to an insufficient mathematical understanding.

Part 2. This part concerns *numeric constraints*. The term numeric constraints is not well-defined in the literature—the basic feature is first-order definability over some structure $(\mathbb{A}; R, R_1, R_2, \ldots)$ where $R(x_1, \ldots, x_n) \equiv x_1 = f(x_2, \ldots, x_n)$ for some "mathematically relevant" function $f : \mathbb{A}^{n-1} \to \mathbb{A}$. It is common to choose $\mathbb{A}$ and R with $\mathbb{A} \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ and $R(x, y, z) \equiv x = y + z$. Important types of relations in this context are *semilinear* relations that are first-order definable in $\{x = y + z, \leq, \{1\}\}$, *semialgebraic* relations that are first-order definable in $\{x = y + z, x = y \cdot z, \leq, \{1\}\}$, and *Presburger arithmetic*, i.e. semilinear relations over the integers. Numeric constraints have important applications and, from an AI perspective, we particularly want to point out the importance of numeric constraints when extending spatiotemporal formalisms with metric information. Numeric constraints are typically not ω-categorical so the algebraic approach is not applicable in general; nevertheless, certain complexity classifications have been obtained by using "non-systematic" methods. Recently, there has been quite some work aimed at constructing general methods for analysing numeric constraints. Examples include the exploitation of saturated models and their Galois connections, connections with algorithmic game theory, and the invention of new algorithmic methods for numeric constraints. The area is in a state of flux and this may provide opportunities for very rapid progress. A concrete starting point would be to analyse the computational complexity of various extensions of the influential *simple temporal problem* by Dechter, and to develop suitable analysis methods.