Research Statement

of Zhonghan Wang (CS PhD applicant for 2025Fall)

My motivation to attend graduage school is a natural outcome of my research experiences gained when I roled as a master student at **Institute of Software, Chinese Academy of Sciences (ISCAS)**. During my experiences, I am mainly interested in exploring the performance and application of state-of-the-art Satisfiability Modulo Theory (SMT) solvers, whose rapid development has benefited a lot of related areas in formal verification, system safety and programming languages. The CAV 2021 award [1] is a witness of the increasing power of SMT Solvers. Generally, my research experiences can be viewed as **a systematic line for nonlinear real arithmetic (NRA) solving**, as splited into several points in the following context with my own growing experience enrolled.

Development of NRA track in Z3-Plus-Plus[2] (winner of single-query in SMT-COMP 2022 & 2023).: My first intersection with SMT solving starts with SMT-COMP. For my first year of master career, I have been working on building efficient SMT solver with **Dr. Shaowei Cai** and **Dr. Bohua Zhan** at ISCAS. During the program experience, I built heuristic methods like static or dynamic variable ordering of model based satisfiability (MCSAT) (de Moura & Jovanovic, 2013; Jovanovic, Barrett, & de Moura, 2013) and its implementation in Z3 prover called NLSAT (Jovanovic & de Moura, 2012). Besides, the implementation of interval constraint analysis (ICP) (Khanh & Ogawa, 2012; Tung, Khanh, & Ogawa, 2017) also helps prove unsatisfiable instances directly. We also design a sample-cell projection in NLSAT explain framework. Although the new operator is also a cylindrical algebraic decomposition projection, it fastens the computation speed for a single cell. Actually from the current view when I write these words, my first experience is more like an engineering work that integrates existing algorithms into our portfolio, rather than doing some research work to express my own insight. However, these ”meaningless” dirty work indeed helps me gain an overview of nonlinear arithmetic solving, by reading plenty publications and other state-of-the-art solvers, including Z3 (de Moura & Bjørner, 2008), CVC5 (Barbosa et al., 2022) and Yices2 (Dutertre, 2014).

Incomplete Method: Local Search for Nonlinear Real Arithmetic (VMCAI'2024) (Z. Wang, Zhan, Li, & Cai, 2024): After completing the submission of our solver for SMT-COMP, I started to wonder about the concrete direction that I should choose. I felt being overwhelmed in the publication ocean and helpless. Luckily, the recent development about introducing local search algorithm into SMT solving (Cai, Li, & Zhang, 2022; B. Li & Cai, 2023; H. Li, Xia, & Zhao, 2023) had triggered my interest. Local Search has shown powerful strength in boolean satisfiability (SAT) solving for a couple of years. It targets to find models near the current exploring branch, and also simplifies the searching complexity, which brings heavy burden for SMT solving. Although local search is weaker in industrial instances compared with systematic search, its advantage in random instances also help developers emerge interesting heuristics. Recently, Prof. Shaowei Cai and some students from my college have designed a ground breaking local search framework, which first introduces local search to try to solve SMT problems on linear integer arithmetic. The novel implementation beats previous competitive tools in satisfiable instances.

Based on this, my research mainly discusses the potential of local search for nonlinear real arithmetic, which has more complicated syntax than linear integer ones. NRA solving is difficult for its uncountable states. Although cylindrical algebraic decomposition helps depict a figure of cells in searching spaces, its heavy computation is actually bottleneck of some systematic algorithms. Our algorithm mainly focuses to cache previously computed score information for arithmetic variables by a novel data structure **boundary**. With this, complexity of computation on real number movements has decreased a lot. Besides, a lazy relaxation mechanism is also introduces for the consideration of number complexity during each loop step of local search. The equality constraints are slacked when necessary, temporarily expands the satisfied cell and brings assignment to an original one in the end. Our code is implemented based on Z3 solver, and beats all mainstream solvers in SMT-LIB. The research result is published in **25th International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI 2024, London)**. I also gained my first experience of expressing my idea and discussing with researchers in UK, which brought me a lot of confidence and honor[3].

[1] https://i-cav.org/2021/cav-award/

[2] https://z3-plus-plus.github.io/

[3] https://www.youtube.com/watch?v=yqSTGOOuHdY&t=450s

Complete Method: Clause Level MCSAT for Nonlinear Real Arithmetic (Preprint) (Z. Wang, 2024a): My third experience is about improving MCSAT algorithm for nonlinear arithmetic. Actually this project is an unexpected gift, cause I've been writing codes about NLSAT during my first experience in Z3-Plus-Plus and eventually design nothing new. However, thanks to the novel idea of introducing feasible set computation into local search (B. Li & Cai, 2023), my motivation of bringing it into complete method was born. Different from CDCL framework in SAT solving, MCSAT encounters two kinds of conflicts including semantic decision and literal decision. The former can be viewed as an extension of boolean variable decision of SAT solving, with the search space of 2^n to R^n. The latter is also required to assign literals and enable conflict analysis. However, from the search space view, one can directly assign new variables by taking the union of feasible set of literals in the current clause constraint. I designed a new mechanism called *feasible set based look-ahead* to properly decide literals in the search process. Besides, I also extend the concept of unit propagation into SMT solving, and named it *clause-level arithmetic propagation*, which could quickly detect conflict clauses. My new ideas update MCSAT a lot, and even this single algorithm can beats current portfolio solvers in satisfiable instances, and ranks second in the unsatisfiable instances. I had also thought about using complete method heursitics in MCSAT, including novel branching heuristics, lemma deletation and phase selection. This generated another ouput called *DNLSAT* (Z. Wang, 2024b), which also helped facilitate solving satisfiable instances, and brought a clear framework of dynamic variable ordering easily for future development. From my personal view, this is the first project that I start to learn about transfering different concepts into different problems.

ML4SMT: Backbone Variables for Nonlinear Real Arithmetic.: I am also thinking about boosting MCSAT using some internal structures, just like someone studies how to boost CDCL in SAT solving. The inner structure is so complex that it's really hard to find the relationship manually without the assistant of machine learning. Luckily, I found a recent work called NeuroBack W. Wang et al. (2024) which used backbone variables to boost CDCL solvers. I decided to connect with **Dr. Wenxi Wang** to further extend this idea into MCSAT. We decide to extend the concept of backbone variables to the prediction of ranged arithmetic variables. Currently this program is still ongoing.

Talking about my future research plan, the strong background on theoretical information will help my development. I am still searching for cooperation opportunity in related areas, like programming language or system safety. Traditional program analysis method like symbolic execution relies on SMT solving and heuristic for seaching space. Besides, synthesizing inductive invariant using syntax guided synthesis (Sygus) is also a related direction with formal verification.

References

Barbosa, H., Barrett, C. W., Brain, M., Kremer, G., Lachnitt, H., Mann, M., … Zohar, Y. (2022). cvc5: A versatile and industrial-strength SMT solver. In D. Fisman & G. Rosu (Eds.), *Tools and algorithms for the construction and analysis of systems - 28th international conference, TACAS 2022, held as part of the european joint conferences on theory and practice of software, ETAPS 2022, munich, germany, april 2-7, 2022, proceedings, part I* (Vol. 13243, pp. 415–442). Springer. Retrieved from https://doi.org/10.1007/978-3-030-99524-9_24 doi: 10.1007/978-3-030-99524-9_24

Cai, S., Li, B., & Zhang, X. (2022). Local search for SMT on linear integer arithmetic. In S. Shoham & Y. Vizel (Eds.), *Computer aided verification - 34th international conference, CAV 2022, haifa, israel, august 7-10, 2022, proceedings, part II* (Vol. 13372, pp. 227–248). Springer. Retrieved from https://doi.org/10.1007/978-3-031-13188-2_12 doi: 10.1007/978-3-031-13188-2_12

de Moura, L. M., & Bjørner, N. S. (2008). Z3: an efficient SMT solver. In C. R. Ramakrishnan & J. Rehof (Eds.), *Tools and algorithms for the construction and analysis of systems, 14th international conference, TACAS 2008, held as part of the joint european conferences on theory and practice of software, ETAPS 2008, budapest, hungary, march 29-april 6, 2008. proceedings* (Vol. 4963, pp. 337–340). Springer. Retrieved from https://doi.org/10.1007/978-3-540-78800-3_24 doi: 10.1007/978-3-540-78800-3_24

de Moura, L. M., & Jovanovic, D. (2013). A model-constructing satisfiability calculus. In R. Giacobazzi, J. Berdine, & I. Mastroeni (Eds.), *Verification, model checking, and abstract interpretation, 14th international conference, VMCAI 2013, rome, italy, january 20-22, 2013.*

proceedings (Vol. 7737, pp. 1–12). Springer. Retrieved from https://doi.org/10.1007/978-3-642-35873-9_1 doi: 10.1007/978-3-642-35873-9_1

Dutertre, B. (2014). Yices 2.2. In A. Biere & R. Bloem (Eds.), *Computer aided verification - 26th international conference, CAV 2014, held as part of the vienna summer of logic, VSL 2014, vienna, austria, july 18-22, 2014. proceedings* (Vol. 8559, pp. 737–744). Springer. Retrieved from https://doi.org/10.1007/978-3-319-08867-9_49 doi: 10.1007/978-3-319-08867-9_49

Jovanovic, D., Barrett, C., & de Moura, L. (2013). The design and implementation of the model constructing satisfiability calculus. In *2013 formal methods in computer-aided design* (p. 173-180). doi: 10.1109/FMCAD.2013.7027033

Jovanovic, D., & de Moura, L. M. (2012). Solving non-linear arithmetic. In B. Gramlich, D. Miller, & U. Sattler (Eds.), *Automated reasoning - 6th international joint conference, IJCAR 2012, manchester, uk, june 26-29, 2012. proceedings* (Vol. 7364, pp. 339–354). Springer. Retrieved from https://doi.org/10.1007/978-3-642-31365-3_27 doi: 10.1007/978-3-642-31365-3_27

Khanh, T. V., & Ogawa, M. (2012). SMT for polynomial constraints on real numbers. In B. Jeannet (Ed.), *Third workshop on tools for automatic program analysis, TAPAS 2012, deauville, france, september 14, 2012* (Vol. 289, pp. 27–40). Elsevier. Retrieved from https://doi.org/10.1016/j.entcs.2012.11.004 doi: 10.1016/j.entcs.2012.11.004

Li, B., & Cai, S. (2023). Local search for smt on linear and multi-linear real arithmetic. In *2023 formal methods in computer-aided design (fmcad)* (p. 1-10). doi: 10.34727/2023/isbn.978-3-85448-060-0_25

Li, H., Xia, B., & Zhao, T. (2023). Local search for solving satisfiability of polynomial formulas. In C. Enea & A. Lal (Eds.), *Computer aided verification - 35th international conference, CAV 2023, paris, france, july 17-22, 2023, proceedings, part II* (Vol. 13965, pp. 87–109). Springer. Retrieved from https://doi.org/10.1007/978-3-031-37703-7_5 doi: 10.1007/978-3-031-37703-7_5

Tung, V. X., Khanh, T. V., & Ogawa, M. (2017). raSAT: an SMT solver for polynomial constraints. *Formal Methods Syst. Des.*, *51*(3), 462–499. Retrieved from https://doi.org/10.1007/s10703-017-0284-9 doi: 10.1007/s10703-017-0284-9

Wang, W., Hu, Y., Tiwari, M., Khurshid, S., McMillan, K., & Miikkulainen, R. (2024). *Neuroback: Improving cdcl sat solving using graph neural networks.* Retrieved from https://arxiv.org/abs/2110.14053

Wang, Z. (2024a). *clausesmt: A nlsat-based clause-level framework for satisfiability modulo nonlinear real arithmetic theory.* Retrieved from https://arxiv.org/abs/2406.02122

Wang, Z. (2024b). *Dnlsat: A dynamic variable ordering mcsat framework for nonlinear real arithmetic.* Retrieved from https://arxiv.org/abs/2406.18964

Wang, Z., Zhan, B., Li, B., & Cai, S. (2024). Efficient local search for nonlinear real arithmetic. In *Verification, model checking, and abstract interpretation: 25th international conference, vmcai 2024, london, united kingdom, january 15–16, 2024, proceedings, part i* (p. 326–349). Berlin, Heidelberg: Springer-Verlag. Retrieved from https://doi.org/10.1007/978-3-031-50524-9_15 doi: 10.1007/978-3-031-50524-9_15