



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

非线性实数可满足性问题的局部搜索算法

作者姓名：王忠汉

指导教师：张立军 研究员 中国科学院软件研究所

学位类别：工学硕士

学科专业：计算机科学与技术

培养单位：中国科学院软件研究所

2025 年 6 月

Local Search Algorithm for Nonlinear Real Satisfiability

**A thesis submitted to
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Master of Engineering
in Computer Science and Technology**

By

WANG Zhonghan

Supervisor: Professor ZHANG Lijun

Institute of Software, Chinese Academy of Sciences

June, 2025

中国科学院大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘要

SMT 问题是形式化方法与软件工程领域涉及到的一类重要问题。相比于 SAT 问题（布尔可满足问题），SMT 问题可以看成是一阶逻辑的扩展，即给定特定理论下的约束，找到满足所有约束的一组解，或者证明不存在这样的解存在。SMT 问题被广泛应用于软硬件验证、程序分析、自动化推理等领域。SMT 问题视理论的不同可以分为不同理论，其中 NRA（非线性实数理论）指的是变量可以取实数值，约束可以是高次非线性多项式的一种理论。针对这类问题，目前的主流求解器 Z3、CVC5、Yices 等都提供了有效的求解方法。常见的搜索算法包括 CDCL (T)、MCSAT、增量线性化、变量替换以及基于区间算术的搜索算法等。然而，优于 NRA 问题解空间的复杂性以及高次多项式运算的复杂性，这类问题的求解仍然具有挑战性。

本文主要深入探讨 NRA 理论的解空间和求解上的难点，并提出用于求解此类复杂约束的高效的局部搜索算法。本文首先介绍 NRA 理论的解空间，并给出一些基本概念的定义，比如胞腔的划分等。接下来，本文提出目前的系统求解算法与局部搜索算法，其中的一些局部搜索算法并没有完全支持全部的 NRA 形式，并且在求解效率上也有一定的局限性。本工作基于此设计了基于边界的分数缓存机制，一方面可以减少搜索过程中的候选操作，另一方面可以有效地降低每次迭代的计算复杂度，并给出了实时更新的条件与具体实施算法。除此之外，NRA 问题具有不同于其他理论的无理数赋值特性，使得局部搜索的迭代效率极低。本文提出了等式松弛策略，可以在一定程度上延后无理数赋值，使用近似解暂时替代精确解。最后，我们也讨论了 NRA 问题独有的无单变量操作的问题，并给出多个变量移动的一种迭代策略。整体的算法实现还包括重启策略，可以在陷入局部最优的情况下跳出当前搜索空间并及时调整搜索区域。

根据上述算法，本文实现了一个名为 LS_NRA 的 SMT 求解工具，可以支持 NRA 理论的任何形式样例。我们在 SMT-LIB 上测试了工具的求解效果，包括一些来自程序验证、自动机理论以及生物网络的使用场景样例。实现结果表明，我们的求解算法和一些完备算法的求解相比具有竞争力，并在高次样例上表现非常好，打破了以往求解器求解个数零的记录。我们还探讨了不同求解器求解单个样例所需的求解时间，我们的算法在时间上可以媲美主流的求解器，可以解决 NRA、理论的实际问题。

关键词： SMT，非线性实数理论，局部搜索

Abstract

SMT problem is an important problem involved in the field of formal methods and software engineering. Compared with the SAT problem (boolean satisfiability problem), the SMT problem can be regarded as an extension of first-order logic, that is, given the constraints under a specific theory, find a set of solutions that satisfy all constraints, or prove that no such solution exists. SMT problems are widely used in software and hardware verification, program analysis, automated reasoning and other fields. SMT problems can be divided into different theories depending on the theory. Among them, NRA (nonlinear real arithmetic) refers to a theory in which variables can take real values and constraints can be high-order nonlinear polynomials. For this type of problem, current mainstream solvers including Z3, CVC5, Yices, all provide effective solutions. Common search algorithms include CDCL(T), MCSAT, incremental linearization, variable substitution, and search algorithms based on interval arithmetic. However, due to the high complexity of the solution space of the NRA problem and the computational complexity of high-order polynomial operations, the solution of this kind of problem is still challenging.

This paper mainly explores the solution space and searching difficulties of NRA theory, and proposes an efficient local search algorithm for solving such complex constraints. This paper first introduces the solution space of NRA theory and gives the definition of some basic concepts, such as the division of the sign-invariant cell. Next, this paper proposes the current system search algorithm and local search algorithm. Some of the local search algorithms do not fully support all NRA forms and have certain limitations in searching efficiency. Based on this, this work designs a boundary-based caching mechanism, which can reduce the candidate operations in the search process on the one hand, and effectively reduce the computational complexity of each iteration on the other hand, and also gives the conditions and specific implementation algorithms for real-time updates. In addition, the NRA problem has irrational number assignment characteristics that are different from other theories, which makes the iterative efficiency of local search extremely low. This paper proposes an equality relaxation strategy, which can postpone the irrational number assignment to a certain extent and use approximate solutions to temporarily replace the exact solution. Finally, we also discuss the problem of no single variable operation unique to the NRA problem, and give an iterative strategy for moving multiple variables. The overall algorithm implementation also includes a restart strategy, which can jump out of the current search space and adjust the search area in time when trapped in the local optimum.

Based on the above algorithm, this paper implements an SMT solver named LS_NRA,

which can support any form of instances of NRA theory. We tested the solving effect of the tools on SMT-LIB benchmark, including some usage scenario instances from program verification, automata theory, and even biological networks. The implementation results show that our solving algorithm is competitive with the performance of some complete algorithms, and performs very well on high-order instances, breaking the previous record of zero number of other solvers. We also explored the solution time required for different solvers to solve a single example. Our algorithm is comparable to mainstream solvers in terms of time, and can solve practical problems of NRA theory.

Keywords: SMT, Nonlinear Real Arithmetic, Local Search

目 录

第 1 章 绪论	1
1.1 研究背景及意义	1
1.1.1 可满足性模理论问题 (SMT Problem)	1
1.2 论文主要工作	3
1.3 论文组织	4
第 2 章 相关技术及研究现状	5
2.1 可满足性模理论与非线性实数问题	5
2.2 SMT(NRA) 的解空间	6
2.2.1 多项式约束的解空间	6
2.2.2 量词消去与柱形代数分解	11
2.3 非线性实数理论求解现状	13
2.3.1 CDCL(T) 算法	13
2.3.2 MCSAT/NLSAT 方法	14
2.4 约束满足问题求解现状	15
2.4.1 约束规划求解	16
2.4.2 其他求解方法	18
2.5 AllDifferent 约束求解现状	19
2.5.1 图论基础	19
2.5.2 AllDifferent 约束和二部图匹配	21
2.5.3 AllDifferent 约束的过滤算法	22
2.6 局部搜索算法现状	25
2.7 本章小结	27
第 3 章 AllDiff-LS 的基础组件和算法框架	29
3.1 问题介绍和求解思路	29
3.2 AllDifferent 约束转化为图及其优化	30
3.3 分步选择操作策略	33
3.4 禁忌策略	35
3.5 打破平局策略	36
3.6 局部搜索算法框架	37
3.7 本章小结	40

第 4 章 针对 AllDifferent 约束的加权和重启策略	41
4.1 解池技术	41
4.2 约束加权图和动态迭代策略	42
4.3 AllDiff-LS 工具设计	44
4.4 本章小结	45
第 5 章 实验设计及结果分析	47
5.1 实验设置	47
5.2 AllDiff-LS 求解 AllDifferent 约束的能力	49
5.3 AllDiff-LS 涉及的策略的有效性	53
5.4 本章小结	56
第 6 章 总结与展望	57
6.1 工作总结	57
6.2 下一步的工作	58
参考文献	59
致谢	63
作者简历及攻读学位期间发表的学术论文与研究成果	65

图形列表

1.1 工具整体框架。	3
2.1 解空间示意图。	7
2.2 延拓点集示意图。	8
2.3 柱形延拓示意图。	9
2.4 柱形代数分解步骤示意图。	12
2.5 CDCL(T) 框架示意图。	14
2.6 AllDifferent 约束和其二部图表示。	22
2.7 值图与残差有向图的转换。	24
3.1 例子中约束转化得到约束图。	31
3.2 例子中约束图的化简。	32
3.3 Local Search 的求解流程。	38
4.1 名为“AI Escargot”的数独实例。	45
5.1 四种方法在给定基准测试上的结果。当 <i>avg_time</i> 为 1000s 时，意味着 超时。	51
5.2 在四类实例上 AllDiff-LS（横坐标）和其他两个版本（纵坐标）的平 均运行时间比较。	54
5.3 在数独上 AllDiff-LS（横坐标）与其他三个版本（纵坐标）的平均运 行时间比较。	55

表格列表

2.1 柱形延拓-区间计算。	9
2.2 柱形延拓-采样点生成。	10
2.3 柱形代数分解-提升。	13
5.1 AllDiff-LS 和其他最先进的基线方法在数独实例上的结果。	50
5.2 AllDiff-LS 和其他最先进的基准方法在 N 皇后，全间隔和 2-MOLS 问题上的结果。	50
5.3 OR-Tools、LocalSolver 和 SAT 求解器 Kissat 在数独实例上的结果。 .	52
5.4 OR-Tools 和 LocalSolver 在其他实例上的结果。	52
5.5 在给定数独实例上使用不同化简规则对应的变量简化比率。	53

第 1 章 绪论

本章节主要介绍本文工作的研究背景和意义。首先，本文对一些基本概念给出定义，包括 SMT 问题、非线性实数理论 (NRA)、解空间以及符号一致胞腔。本文还会介绍目前的几种主流求解算法以及局限性，引出本文工作的研究动机。接着，本文会针对工作的几个创新点展开，详细阐述算法的设计。最后，本文将总结论文的整体结构。

1.1 研究背景及意义

随着信息技术的发展，软硬件系统的正确性和安全性日益成为人们关注的话题。现有的一些验证技术使用诸如模型检查、定理证明等手段将问题转化成为可满足性模理论 (Satisfiability Modulo Theories, SMT) 问题，并通过求解器求解。因此，SMT 求解算法的设计对工业生产、科学研究等领域具有重要意义。针对特定的约束类型，如何设计高效的求解算法，在短时间内求解更多的样例成为 SMT 研究的一个重点。基于此，研究一种高效求解 SMT 问题的算法具有很高的理论意义与实际价值。

1.1.1 可满足性模理论问题 (SMT Problem)

可满足性模理论问题 (SMT) 是一种在计算机科学中常见的问题类型，它组合了布尔可满足性问题 (SAT) 和一些理论约束。每一个约束一般可以表示为对变量集合取值范围的限制，当一组变量的赋值满足了所有约束时，当前赋值被称为 SMT 问题的一组解。SMT 求解可以理解为寻找一组解或者证明不存在解的过程。SMT 问题的一个挑战在于囊括理论的复杂性，比如算术理论、位向量理论等，不同理论需要的求解策略也不同。除此之外，SMT 问题涉及到的变量个数与约束个数十分庞大，解空间规模基本上呈指数关系，这也就造成了目前的求解器很难在短时间内处理十分庞大的 SMT 问题。

SMT 问题在很多领域有广泛应用，常用于符号执行^[1,2]，程序验证^[3,4]，程序生成^[5]，自动机学习^[6,7]以及神经网络验证等^[8-11]。非线性实数理论一般广泛应用于信息物理系统^[12-14]，程序终止条件的秩函数生成^[15,16]，非线性混成自动机的分析^[17]等。一般来说，这些问题会把需要验证的条件编码成为 SMT 问题，然后交给后端的求解器取处理。比如，在符号执行中约束包括程序的分支条件，需要对程序可能执行的每一条路径进行搜索，从而确保最终程序的状态符号认为的要求，这些条件最终被编码成为 SMT 问题中的约束，整个问题转化为全部约束是否可以同时满足，即 SMT 问题。

目前关于 SMT 求解的研究基本可以分为完备搜索算法和启发式搜索算法。完备搜索算法主要的思想是通过搜索 SMT 问题的一部分空间，然后在遇到冲突

的情况下通过推理和回溯完成对当前冲突区域的剪枝，以避免后续搜索遇到相同冲突，算法在所有变量得到赋值后结束。一般来说，完备算法因为其很强的推理能力成为主流算法，但整体推理的时间和空间复杂度较高，面对大规模的样例容易消耗过多的计算资源。启发式方法一般借助一些人工的启发式优化策略，比如引入随机等，来更好地完成整个搜索过程。启发式方法一般会针对特殊的约束类型采取不同的启发式策略，因此其求解效果可能会根据样例的形式不同而产生不同的效果。在处理大规模样例中，启发式效果对计算资源的消耗较少，有时会得到比完备算法更好的结果。本文接下来会重点概述几种不同算法的求解思路。

完备算法一般也成为系统搜索算法，可以同时处理约束可满足以及不可满足的情况。SMT 的完备算法一般包括 CDCL(T)^[18] 和 MCSAT 算法^[19,20]，其共同的思路是不断尝试新的赋值，在遇到冲突时通过冲突分析学习新的子句，通过不断试错缩小需要探索的解空间，直到最终找到满足所有约束的一组解，或者排除整个解空间从而证明原公式不可被满足。非线性实数理论一般需要通过柱形代数分解 (CAD)^[21] 进行量词消去，从而学习到特定冲突下的新子句。这方面的研究包括应用 CAD 的变种^[22]，对 CAD 投影设计启发式的变量顺序^[23] 等。除了上述两种算法之外，近年来一些其他算法也在 SMT 求解上取得了不错的效果，包括增量线性化^[24]，区间约束传播^[25,26] 和亚热带方法^[27,28]。这些方法一般作为求解器插件使用，可以在特定的样例进行快速处理。

近年来，一些基于优化方法常用来检测给定区域内是否存在符合多项式组的解，进而应用到了 SMT 问题上。其中，Cimatti 等人的工作^[29] 首次应用全局优化的方法去寻找初始解，然后通过迭代寻找附近的可行解。Ni 等人的工作也使用了优化方法去寻找可行解^[30]，然后通过解方程^[31] 等手段求出一个精确解。

Gao 等人引入 δ -完备 (δ -complete) 决策程序的概念，并基于此设计了解决非线性约束的 dReal 求解器。与一般 SMT 求解器不同的是，dReal 支持对指数函数和三角函数的求解。其中 δ -完备包括 δ -满足 (δ -sat) 和不可满足 (unsat)，通过松弛输入的公式来解决更宽泛问题的效果。本文的工作主要借鉴了这种松弛的想法来加速局部搜索的迭代。和 dReal 求解器不同的是，我们的算法最终仍然会返回一个严格满足所有约束的精确解。

局部搜索算法是本文工作的重点，也是近年来求解可满足样例的重点。局部搜索算法一般从一个完全赋值开始，针对当前尚且不可满足的约束设计操作，使用评价函数筛选合适的操作进行迭代，最终通过不断在邻域中搜索输出满足所有约束的一组赋值。其主要优点是对特定样例的求解效果很好，并且能够在很短的时间内找到足够好的一组解。主要缺点包括容易陷入局部最优、操作的设计和评价函数的设计较为困难等。局部搜索一般不可用于求解不可满足的样例。目前主流的局部搜索算法支持线性整数逻辑^[32]、非线性整数逻辑^[33]、多线性样例^[34] 和部分多项式理论^[35]。本文提出了第一个可以覆盖全部非线性实数理论的局部搜索算法。

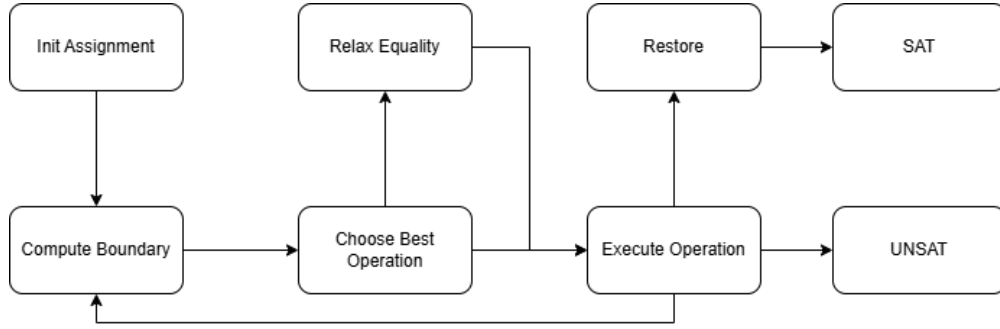


图 1.1 工具整体框架。

Figure 1.1 Overall framework of LS_NRA.

1.2 论文主要工作

本文重点关注非线性实数的 SMT 问题。本文主要讨论非线性实数问题求解的难点，并针对这些难点设计出合适的局部搜索算法，从而达到高效求解的效果。目前算法的主要难点是求解高次多项式约束需要太多求解时间，这些样例成为我们设计局部搜索算法的重点。

本工作主要在 SMT-LIB^[36] 上进行试验，在可满足样例上超越了目前的主流搜索算法。算法的创新性上，本文主要考虑以下几个方面：

- 考虑通过设计更好地数据结构和迭代策略，针对实数问题的操作采样进行优化，以期可以加速整体搜索过程；
- 针对非线性实数特有的无理数赋值问题，如何减少多项式计算上的时间消耗；
- 考虑非线性问题单变量无操作的情况，如何避免搜索陷入停滞的情况；

本文的局部搜索算法 **LS_NRA** 整体流程如图1.1所示。本工作主要包括以下几个贡献：

- 首先，本文通过分析非线性实数的解空间引入边界（boundary）数据结构，从而实现了可行域-分数对变量的缓存机制。本文还给出了邻居变量的定义以及边界的更新算法，从而可以保证算法的正确性以及数据结构的可复用性；
- 针对无理数赋值问题，本文借鉴 dReal 的做法，在强迫无理数赋值时引入等式松弛（relaxation）的概念，允许暂时的有理数赋值。在找到松弛解之后，本文给出了求解精确解的算法，保证了算法的正确性；
- 针对非线性问题独有的无单变量移动问题，本文给出了一个简单的迭代算法和前瞻策略 (look-ahead)，基本避免了算法停滞的现象；
- 本文增加了重启策略和预处理模块，相关工具在 SMT-Lib 上效果良好，可以在短时间内快速找到高次多项式的可满足赋值，打破了以往主流求解器的高次问题上的零求解。

1.3 论文组织

本文的后续章节按照以下方式组织：

第二章：介绍 SMT 问题的基本概念和解空间，并介绍目前主流的算法和研究现状。

第三章：介绍目前非线性实数理论的挑战和本文的设计思路及实现过程。

第四章：介绍预处理模块、重启策略和整体工具实现。

第五章：介绍实验设计和结果分析。

第六章：总结本文贡献，展望后续研究工作。

第2章 相关技术及研究现状

本章节主要介绍 SMT 和非线性实数理论的相关基本概念以及研究现状。首先, 本文给出 SMT 问题和非线性实数理论的语法, 然后给出解空间表达以帮助读者更好理解 SMT 问题的本质。紧接着我们给出 CDCL(T) 算法的结构、MCSAT 算法原理以及其他一些 SMT 求解器的介绍。然后我们给出近几年来局部搜索在 SMT 问题上的应用, 包括线性整数逻辑、多线性逻辑和部分多项式理论等。最后, 我们总结目前算法的缺点和非线性问题的挑战, 并引出我们的工作带来的进展和突破。

2.1 可满足性模理论与非线性实数问题

首先, 本文给出可满足性模理论 (Satisfiability Modulo Theories, SMT) 问题的一般定义, 前置概念介绍如下:

定义 2.1 (变量). SMT 问题中的变量根据赋值要求分为布尔变量和算术变量。布尔变量是只能取布尔值 (T, \perp) 的变量, 布尔变量集合一般用 B 表示。算术变量可以取整数或者实数值, 一般用 R 表示。非线性实数问题中所有算术变量必须取实数值。

定义 2.2 (多项式约束). 非线性问题主要由多项式约束构成, 多项式约束的形式为 $P(x) \sim 0$, 其中 $P(x)$ 是一个多项式, \sim 是一个关系符号, 可以是 $=, <, >$ 中的一个。

Tseitin 编码保证了任意的逻辑约束可以在多项式时间内转化为合取范式 (CNF)。为方便说明, 本文中所有 SMT 约束表达成为 CNF 形式, 一些基本概念如下:

定义 2.3. 命题文字 命题文字 (literal) 是可满足问题中的基本结构, 包括正文字 (p) 和负文字 ($\neg p$)。非线性实数理论的命题文字一般是多项式约束或布尔变量约束。正文字只有当对应的多项式约束满足时才满足, 负文字只有当对应的多项式约束不满足时才满足。一个典型的非线性实数文字比如 $x^2 + y > 0$ 或者 $\neg(x + y^3 < 0)$ 。

定义 2.4. 子句 子句 (clause) 定义为文字的析取结构, 一般记为 $C = l_1 \vee l_2 \vee \dots \vee l_n$, 其中 l_i 是一个文字。当子句包含的所有文字不满足时子句不满足, 只要有一个文字被满足子句也被满足。**空子句**一般表示不包含任何文字的子句。比如, 一个非线性实数理论的子句形如 $(x^2 + y > 0) \vee (x + y^3 < 0)$ 。

定义 2.5. 公式 公式 (formula) 定义为子句的合取结构, 一般记为 $\varphi = C_1 \wedge C_2 \wedge \dots \wedge C_m$, 其中 C_i 是一个子句。当公式包含的所有子句都满足时公式满足, 只要有一个子句不满足公式就不满足。一个非线性实数理论的公式比如 $(x^2 + y > 0) \wedge (x + y^3 < 0)$ 。

定义 2.6. 赋值 赋值 (assignment) 指的是变量到布尔值或者实数的一种映射, 布尔变量赋值为 $f : B \rightarrow \{\top, \perp\}$, 算术变量赋值为 $f : R \rightarrow \mathbb{R}$ 。一个赋值 f 满足一个公式 φ , 记作 $f \models \varphi$, 当且仅当对于公式中的每一个子句 C_i , 至少有一个文字 l_j 满足 $f \models l_j$ 。赋值可以分为部分赋值 (partial assignment) 和完全赋值 (full assignment), 部分赋值只对部分变量存在映射, 完全赋值对所有变量都存在映射。赋值一般记为一组映射, 比如 $\{x \mapsto 2, y \mapsto 3\}$ 。

定义 2.7. 评估 评估 (evaluation) 指的是给定赋值和一个文字, 判断当前文字是否满足, 记为 $eval(ass, l) \rightarrow \{\top, \perp\}$, 其中 ass 是一个赋值, l 是一个文字。评估的结果为 \top 表示文字满足, 为 \perp 表示文字不满足。比如 $eval(\{x \mapsto 2, y \mapsto 3\}, x^2 + y > 0) \rightarrow \top$ 。评估只有在文字包含的所有变量都有赋值时才有返回值。

定义 2.8. 可满足性模理论 (SMT) 可满足性模理论 (SMT) 指的是给定一个逻辑公式 φ , 如果存在一组赋值 f 使得 $f \models \varphi$, 则称 φ 是**可满足的 (satisfiable)**, 否则称 φ 是**不可满足的 (unsatisfiable)**。SMT 问题的目标是找到一个可满足的赋值, 即找到一个满足公式的赋值, 或者证明赋值不可满足。

我们给出以下例子 2.1 进一步说明 SMT 问题的可满足性。

例 2.1. 给定 SMT 公式 $\varphi = (x^2 + y > 0) \wedge (x + y^3 < 0)$, 我们可以找到一组赋值 $f = \{x \mapsto 2, y \mapsto -2\}$ 使得 $f \models \varphi$ 。因此我们称该问题是可满足的。给定 SMT 公式 $\varphi = (x^2 + y^2 < 0)$, 我们在实数空间上找不到一组赋值满足当前约束, 因此该问题是不可满足的。

逻辑公式化简: 为方便说明, 我们把所有的算术文字化简为以下几种形式 $\{p > 0, p \geq 0, p = 0\}$, 其中 p 是多项式。具体的化简规则表述如下:

- $p < 0$ 化简为 $'(p) > 0$, 其中 $'(p)$ 表示 p 的相反多项式。
- $p \leq 0$ 化简为 $'(p) \geq 0$, 其中 $'(p)$ 表示 p 的相反多项式。
- $p \leq 0 \wedge p \geq 0$ 化简为 $p = 0$ 。

2.2 SMT(NRA) 的解空间

本小节主要探讨 SMT(NRA) 公式的解空间结构, 以帮助读者更好理解 SMT 问题和搜索过程的本质。

2.2.1 多项式约束的解空间

定义 2.9. 解空间 解空间 (Solution Space) 指的是满足 SMT 公式的所有赋值的集合, 一般用 S 表示。解空间是一个高维空间, 每个维度对应一个变量的取值, 每个点对应一个赋值。解空间的维度取决于变量的个数, 解空间的大小取决于变量的取值范围。非线性实数的解空间是 R^n , 其中 n 是变量的个数。

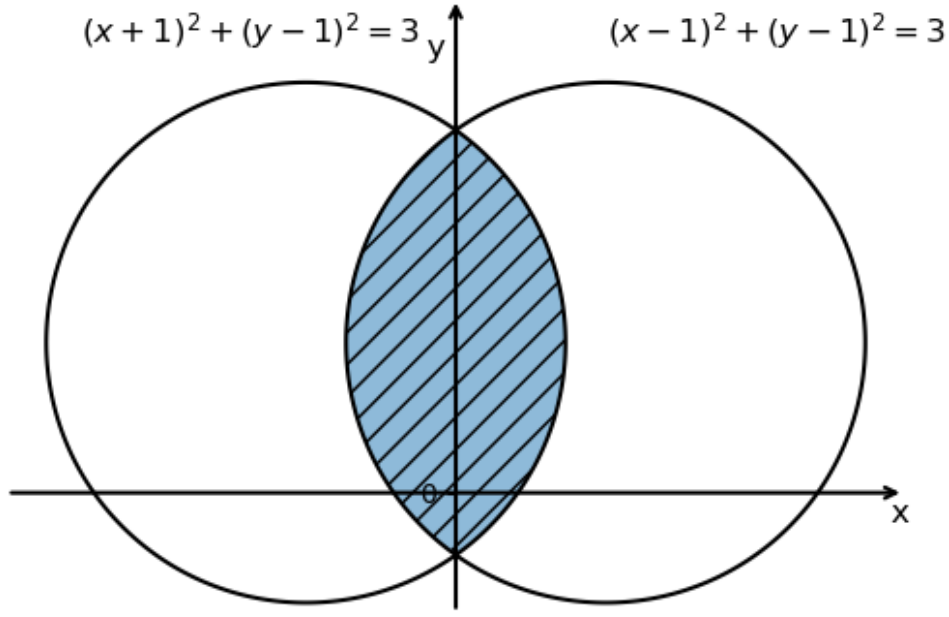


图 2.1 解空间示意图。

Figure 2.1 Solution Space Demo.

例 2.2. 考虑逻辑公式 $F = (x - 1)^2 + (y - 1)^2 \leq 3 \wedge (x + 1)^2 + (y - 1)^2 \leq 3$ ，构成的 R^2 空间图形如图2.1所示。两个子句分别表示两个圆的内部，逻辑公式 F 表示同时存在两个圆内部的区域，即图中阴影区域。任何存在于阴影区域内的点都满足逻辑公式 F ，任何满足逻辑公式 F 的点都存在于阴影区域内。因此，解空间 S 是阴影区域的集合。

数学上，我们把这样的区域定义为胞腔。

定义 2.10. 胞腔 对于 R^n 空间上的多项式集合 Q ， Q 的一个胞腔 (cell) 是每个多项式 $P \in Q$ 保持符号不变的 R^n 最大联通集合。对于任意的点 $a \in R^n$ ，如果 a 在 Q 的胞腔内，则 a 满足 Q 中的所有多项式，我们记这个胞腔为 $cell(Q, a)$ 。显然，胞腔是 R^n 空间的一个划分。

对于逻辑公式而言，因为胞腔内的任意一点对每个多项式 $P \in Q$ 保持符号不变，因此 SMT 公式对应的所有文字仍然保持布尔值不变，不会对 SMT 公示的满足或不满足造成任何影响。当找到其中任意一个满足逻辑公式的胞腔时，可以判定原公式可满足；当所有胞腔都被证明不可满足时，可以判定原公式不可满足。

给定逻辑公式，如何快速剔除不满足的胞腔并找到可满足的胞腔非常重要，因此现有的研究工作很多聚焦在更好地胞腔划分上。一般的胞腔划分是根据多项式的根和判别式等来判断的，从一个点得到胞腔的划分成为延拓。

定义 2.11. 延拓 假定 R^n 上的多项式集合 Q 和点 $a = (a_1, \dots, a_n)$ 。给定一个变量 $x_i (i = 1, 2, \dots, n)$ ，假定多项式 $\{q(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) | q(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) \neq$

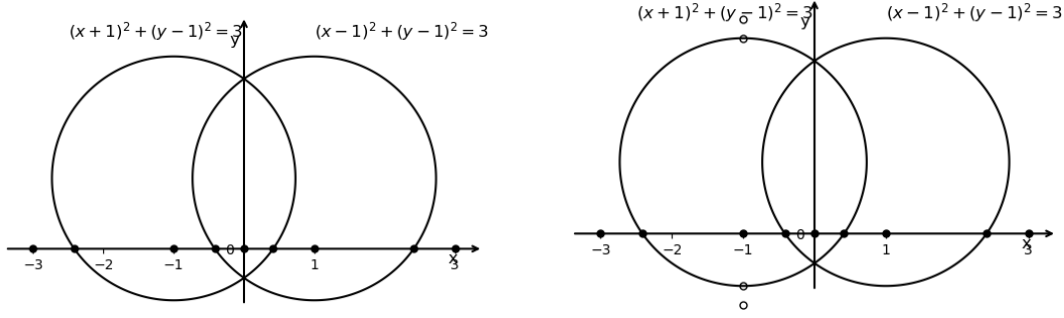


图 2.2 延拓点集示意图。

Figure 2.2 Expansion Demo.

$0, q \in Q\}$ 的实数根 $r_1 < r_2 < \dots < r_s$ 。一个点 a 关于变量 x_i 在集合 Q 上的延拓定义为一个满足如下性质的点集 $\Lambda \subseteq R^n$:

- 点 $a \in \Lambda$ 并且对于 $1 \leq j \leq s$ 均存在点 $(a_1, \dots, a_{i-1}, r_j, a_{i+1}, \dots, a_n) \in \Lambda$ 。
- 对于任意点 $b = (b_1, \dots, b_n) \in \Lambda$, 对于 $j \in 1, \dots, n$

i 有 $b_j = a_j$ 。

- 对于任意的区间 $I \in \{(-\infty, r_1), (r_1, r_2), \dots, (r_{s-1}, r_s), (r_s, +\infty)\}$, 都有唯一的 $b = (b_1, \dots, b_n) \in \Lambda$ 满足 $b_i \in I$ 。

对于点集 $\{a^{(1)}, \dots, a^{(m)}\} \subseteq R^n$, 定义集合关于变量 x_i 在 Q 上的延拓时 $\bigcup_{j=1}^m \Lambda_j$, 其中 Λ_j 是 $a^{(j)}$ 关于变量 x_i 的延拓。

例 2.3. 紧接着例子 2.2, 如图 2.2 所示, 我们可以得到点 $(0, 0)$ 关于 x 的延拓, 即点集 $\{(-3, 0), (-1 - \sqrt{2}, 0), (-1, 0), (1 - \sqrt{2}, 0), (0, 0), (\sqrt{2} - 1, 0), (1, 0), (1 + \sqrt{2}, 0), (3, 0)\}$ 。这些点在 x 变量上分割了多项式的符号区间, 从而划分了胞腔在 x 方向上的投影。右侧图的空心点构成了点 $(-1, 0)$ 关于变量 y 的延拓, 即点集 $\{(-1, -1), (-1, 1 - \sqrt{3}), (-1, 1 + \sqrt{3}), (-1, 3)\}$, 这些点在 y 方向上分开了多项式的符号区间, 从而划分了胞腔在 y 方向上的投影。

点 a 关于变量 x 的延拓点集事实上是点 a 在方向 x 上相邻胞腔的采样点, 如何能够对 R^n 空间上的所有胞腔进行采样是我们接下来的话题。为此, 我们引入柱形延拓的概念。

定义 2.12. 柱形延拓 假设 R^n 上的多项式集合 Q 和点 a 。给定一个变量顺序 $x_1 < x_2 < \dots < x_n$, 定义点 a 关于变量顺序在 Q 上的柱形延拓是 $\bigcup_{i=1}^n \Lambda_i$, 其中 Λ_1 是 a 关于变量 x_1 在 Q 上的延拓, 并且 Λ_{i+1} 是 Λ_i 关于变量 x_{i+1} 在 Q 上的延拓。我们把最终的结果 $\bigcup_{i=1}^n \Lambda_i$ 称为 Q 上的柱形延拓。

柱形延拓可以理解为给定变量顺序下多维的点集延拓, 其目的是在 R^n 空间的每一个胞腔上都有一个采样点。

例 2.4. 紧接着例子 2.3, 如图 2.3 所示, 图中所有的实心点和空心点共同构成了基于变量顺序 $x < y$ 的多项式 Q 的柱形延拓, 达到了每个胞腔上均有一个采样点

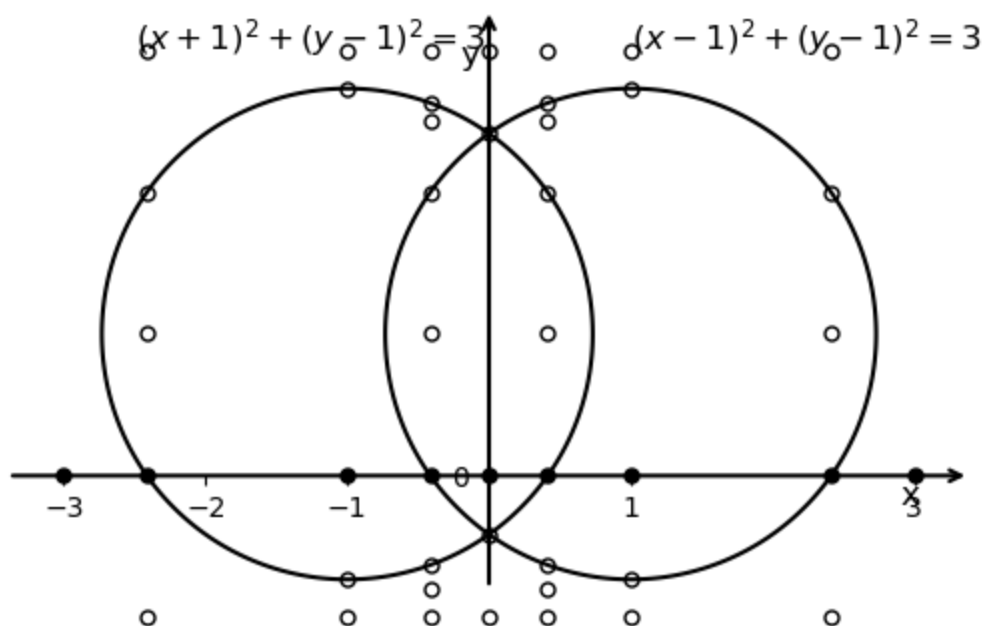


图 2.3 柱形延拓示意图。

Figure 2.3 Cylindrical Expansion Demo.

点坐标	$P_1 : (x+1)^2 + (y-1)^2 = 3$ 根	$P_2 : (x-1)^2 + (y-1)^2 = 3$ 根	根集合
$x \mapsto -3$	\emptyset	\emptyset	\emptyset
$x \mapsto -1 - \sqrt{2}$	$\{0, 2\}$	\emptyset	$\{0, 2\}$
$x \mapsto -1$	$\{1 - \sqrt{3}, 1 + \sqrt{3}\}$	\emptyset	$\{1 - \sqrt{3}, 1 + \sqrt{3}\}$
$x \mapsto 1 - \sqrt{2}$	$\{-0.63, 2.63\}$	$\{0, 2\}$	$\{-0.63, 0, 2, 2.63\}$
$x \mapsto 0$	$\{1 - \sqrt{2}, 1 + \sqrt{2}\}$	$\{1 - \sqrt{2}, 1 + \sqrt{2}\}$	$\{1 - \sqrt{2}, 1 + \sqrt{2}\}$
$x \mapsto \sqrt{2} - 1$	$\{0, 2\}$	$\{-0.63, 2.63\}$	$\{-0.63, 0, 2, 2.63\}$
$x \mapsto 1$	\emptyset	$\{1 - \sqrt{3}, 1 + \sqrt{3}\}$	$\{1 - \sqrt{3}, 1 + \sqrt{3}\}$
$x \mapsto 1 + \sqrt{2}$	\emptyset	$\{0, 2\}$	$\{0, 2\}$
$x \mapsto 3$	\emptyset	\emptyset	\emptyset

表 2.1 柱形延拓-区间计算。

Table 2.1 Demo of Computation of Cylindrical Expansion-Interval Splitting.

投影点	分割区间	延拓点集 (采样点)
$(-3, 0)$	$(-\infty, \infty)$	$(-3, 0)$
$(-1 - \sqrt{2}, 0)$	$(-\infty, 0)$	$(-1 - \sqrt{2}, -1)$
	$[0, 0]$	$(-1 - \sqrt{2}, 0)$
	$(0, 2)$	$(-1 - \sqrt{2}, 1)$
	$[2, 2]$	$(-1 - \sqrt{2}, 2)$
	$(2, \infty)$	$(-1 - \sqrt{2}, 3)$
$(-1, 0)$	$(-\infty, 1 - \sqrt{3})$	$(-1, -1)$
	$[1 - \sqrt{3}, 1 - \sqrt{3}]$	$(-1, 1 - \sqrt{3})$
	$(1 - \sqrt{3}, 1 + \sqrt{3})$	$(-1, 0)$
	$[1 + \sqrt{3}, 1 + \sqrt{3}]$	$(-1, 1 + \sqrt{3})$
	$(1 + \sqrt{3}, \infty)$	$(-1, 3)$
$(1 - \sqrt{2}, 0)$	$(-\infty, -0.63)$	$(1 - \sqrt{2}, -1)$
	$[-0.63, -0.63]$	$(1 - \sqrt{2}, -0.63)$
	$(-0.63, 0)$	$(1 - \sqrt{2}, -0.5)$
	$[0, 0]$	$(1 - \sqrt{2}, 0)$
	$(0, 2)$	$(1 - \sqrt{2}, 1)$
	$[2, 2]$	$(1 - \sqrt{2}, 2)$
	$(2, 2.63)$	$(1 - \sqrt{2}, 2.5)$
	$[2.63, 2.63]$	$(1 - \sqrt{2}, 2.63)$
	$(2.63, \infty)$	$(1 - \sqrt{2}, 3)$
$(0, 0)$	$(-\infty, 1 - \sqrt{2})$	$(0, -1)$
	$[1 - \sqrt{2}, 1 - \sqrt{2}]$	$(0, 1 - \sqrt{2})$
	$(1 - \sqrt{2}, 1 + \sqrt{2})$	$(0, 0)$
	$[1 + \sqrt{2}, 1 + \sqrt{2}]$	$(0, 1 + \sqrt{2})$
	$(1 + \sqrt{2}, \infty)$	$(0, 3)$
$(\sqrt{2} - 1, 0)$	$(-\infty, -0.63)$	$(\sqrt{2} - 1, -1)$
	$[-0.63, -0.63]$	$(\sqrt{2} - 1, -0.63)$
	$(-0.63, 0)$	$(\sqrt{2} - 1, -0.5)$
	$[0, 0]$	$(\sqrt{2} - 1, 0)$
	$(0, 2)$	$(\sqrt{2} - 1, 1)$
	$[2, 2]$	$(\sqrt{2} - 1, 2)$
	$(2, 2.63)$	$(\sqrt{2} - 1, 2.5)$
	$[2.63, 2.63]$	$(\sqrt{2} - 1, 2.63)$
	$(2.63, \infty)$	$(\sqrt{2} - 1, 3)$
$(1, 0)$	$(-\infty, 1 - \sqrt{3})$	$(1, -1)$
	$[1 - \sqrt{3}, 1 - \sqrt{3}]$	$(1, 1 - \sqrt{3})$
	$(1 - \sqrt{3}, 1 + \sqrt{3})$	$(1, 0)$
	$[1 + \sqrt{3}, 1 + \sqrt{3}]$	$(1, 1 + \sqrt{3})$
	$(1 + \sqrt{3}, \infty)$	$(1, 3)$
$(1 + \sqrt{2}, 0)$	$(-\infty, 0)$	$(1 + \sqrt{2}, -1)$
	$[0, 0]$	$(1 + \sqrt{2}, 0)$
	$(0, 2)$	$(1 + \sqrt{2}, 1)$
	$[2, 2]$	$(1 + \sqrt{2}, 2)$
	$(2, \infty)$	$(1 + \sqrt{2}, 3)$
$(3, 0)$	$(-\infty, \infty)$	$(3, 0)$

表 2.2 柱形延拓-采样点生成。

Table 2.2 Demo of Computation of Cylindrical Expansion-Sampling Points.

的效果。具体的步骤是，首先点 $(0,0)$ 关于变量 x 形成延拓（实心点），然后每一个实心点关于变量 y 各自形成延拓（空心点），最终所有点的集合共同构成了柱形延拓。表格2.1展示了在实心点处根据多项式集合的根分割得到的区间。表格2.2展示了在每一个分割区间进行采样得到的结果，即最终的柱形延拓点集（空心点）。

定义 2.13. 柱形完备 对于 R^n 上的多项式集合 Q ，给定一个变量顺序 $x_1 < x_2 < \dots < x_n$ ，称 Q 对于变量顺序是柱形完备的，当任意的点 $a \in R^n$ 和其关于 Q 的柱形延拓 Λ ， Q 的每一个胞腔都包含至少一个 Λ 的点。

具体的证明请参阅[21?]。

2.2.2 量词消去与柱形代数分解

根据前文的讨论，我们引出基于多项式解空间的一种量词消去算法-柱形代数分解 (Cylindrical Algebraic Decomposition)。

定义 2.14. 量词消去 量词消去指的是给定一个带有量词的逻辑公式 φ ，找到另一个不带有量词的逻辑公式 ψ ，使得 φ 和 ψ 的逻辑等价，记为 $\varphi \Leftrightarrow \psi$ 。量词消去的目的是将逻辑公式转化为更易处理的形式，以更好地判断问题的可满足性。

例 2.5. 考虑多项式 $P_s(x, y) = s(x^2 + y^2 - 1) + (1 - s)(xy - 1)$ ，和逻辑公式 $\varphi = \exists R \forall x, y [P_s(x, y) = 0 \Rightarrow x^2 + y^2 \leq R^2]$ 。量词消去即求解当 s 取什么值的时候逻辑公式 φ 满足。通过量词消去工具，我们可以得到 $\varphi \Leftrightarrow s \leq -1 \wedge s > \frac{1}{3}$ 。

多项式理论常用的一种量词消去是柱形代数分解 (Cylindrical Algebraic Decomposition, CAD)。其核心仍然是把 R^n 空间划分成多个符号一致的胞腔，然后通过判断每个胞腔是否满足逻辑公式进而得到给定变量的赋值区间。如图2.4所示，给定变量顺序和多项式集合，柱形代数分解可以分为三个步骤：投影、实根隔离和提升，具体的步骤如下：

- **投影 (projection):** 从多项式集合开始，每次消去一个变量，生成新的多项式集合，新多项式继续消去变量直到只剩下一个变量为止，记为 $proj(P)$ 。
- **实根隔离 (root isolation):** 当投影只剩下一个变量时，对当前多项式集合求所有根，得到顺序排列的一组根，这些根表示了胞腔在当前变量上的分割。
- **提升 (lift):** 提升是对投影阶段得到的多项式集合进行采样。投影从单变量多项式开始，从采样点继续对上一个投影多项式进行采样，直到最终对 R^n 多项式采样，最终保证了每个胞腔得到了采样点。

投影算子 (projection operator) 指的是投影过程中对多项式集合的计算，比如 Collins 算子[?] 和 McCallum 算子[?]。后者因为其计算相对便捷，被广泛应用于量词消去工具和 SMT 求解器中。

McCallum 算子用到的计算包括结式、判别式和求根等操作。

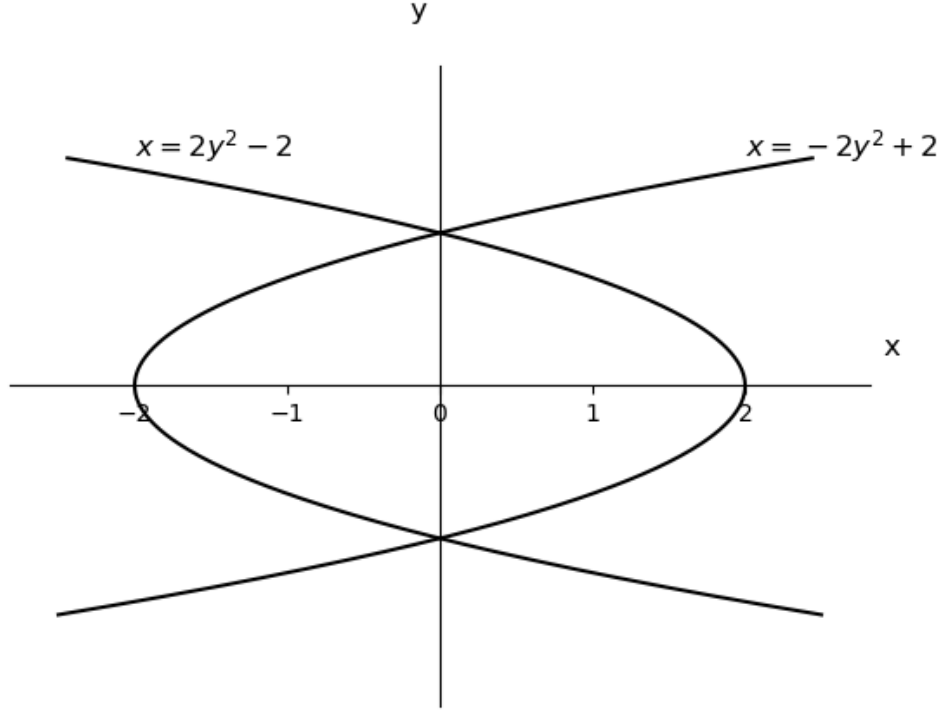


图 2.4 柱形代数分解步骤示意图。

Figure 2.4 Demo of Cylindrical Algebraic Decomposition Steps.

定义 2.15. 结式 (resultant) 对于两个多项式 $P_1, P_2 \in R[x_1, \dots, x_n]$, 假定

$$P_1 = a_m x_n^{d_m} + a_{m-1} x_n^{d_{m-1}} + \dots + a_0,$$

$$P_2 = b_n x_n^{e_n} + b_{n-1} x_n^{e_{n-1}} + \dots + b_0.$$

定义 P_1 和 P_2 的结式 $Res(P_1, P_2, x_n)$ 为:

$$Res(P_1, P_2, x_n) = \begin{vmatrix} a_m & a_{m-1} & \dots & a_0 & & & \\ & a_m & a_{m-1} & \dots & a_0 & & \\ & & \ddots & \ddots & \ddots & \ddots & \\ & & & a_m & a_{m-1} & \dots & a_0 \\ b_n & b_{n-1} & \dots & b_0 & & & \\ & b_n & b_{n-1} & \dots & b_0 & & \\ & & \ddots & \ddots & \ddots & \ddots & \\ & & & b_n & b_{n-1} & \dots & b_0 \end{vmatrix}$$

定义 2.16. 判别式 (discriminant)

对于多项式 $P \in R[x_1, \dots, x_n]$, 假定

$$P = a_m x_n^{d_m} + a_{m-1} x_n^{d_{m-1}} + \dots + a_0.$$

分割区间	x 采样点	y 根集合	(x, y) 采样点	符号 (P_1, P_2)
$(-\infty, -2)$	$x \rightarrow -3$	$\{-\frac{\sqrt{10}}{2}, \frac{\sqrt{10}}{2}\}$	$(-3, 2)$	$(-, +)$
			$(-3, \frac{\sqrt{10}}{2})$	$(-, 0)$
			$(-3, 0)$	$(-, -)$
			$(-3, -\frac{\sqrt{10}}{2})$	$(-, 0)$
			$(-3, -2)$	$(-, +)$
$[-2, -2]$	$x \rightarrow -2$	$\{0, -\sqrt{2}, \sqrt{2}\}$	$(-2, 2)$	$(-, +)$
			$(-2, \sqrt{2})$	$(-, 0)$
			$(-2, 1)$	$(-, -)$
			$(-2, 0)$	$(0, -)$
			$(-2, -1)$	$(-, -)$
			$(-2, -\sqrt{2})$	$(-, 0)$
			$(-2, -2)$	$(-, +)$

表 2.3 柱形代数分解-提升。

Table 2.3 Demo of Cylindrical Algebraic Decomposition - Lift.

定义 P 关于变量 x_n 的判别式 $Disc(P, x_n)$ 为:

$$Disc(P, x_n) = \frac{(-1)^{\frac{m(m-1)}{2}}}{a_m} Res(f, \frac{\partial f}{\partial x_n}, x_n).$$

定义 2.17. McCallum 投影算子 (McCallum Projection Operator)

假定 $F = \{f_1, f_2, \dots, f_m\}$ 是一组 R^n 上的多项式。McCallum 投影算子 $proj(F, x_i)$ 是从 F 到 R^{n-1} 上的多项式集合 $proj_m(F)$ 的一组映射, 包含以下元素:

- F 中每个多项式的系数
- F 中每个多项式关于变量 x_n 的判别式
- F 中每两个不同多项式 f_i, f_j 关于变量 x_n 的结式

实际计算中, 多项式集合的元素通过因式分解化为最简系数的形式, 对结果不产生影响。

例 2.6. 我们给出 CAD 算法的一个例子。假设 $F = \{P_1 : x-2y^2+2, P_2 : x+2y^2-2\}$, 我们规定变量顺序为 $y < x$, 按步骤计算如下:

1. **投影:** 计算对于变量 y 的投影 $proj(F, y)$
 - 化简后系数多项式为 $\{x+2, x-2\}$
 - 化简后判别式为 $\{x+2, x-2\}$
 - 化简后结式为 $\{x\}$ 。

得到投影多项式 $proj(F, y) = \{x, x-2, x+2\}$

2. **实根隔离:** 我们对 $proj(F, y)$ 的每个多项式求根, 得到根集合 $\{-2, 0, 2\}$ 。

3. **提升:** 我们对根集合分得的区间进行提升, 进一步采样所有胞腔, 结果如表格 2.3 所示。

2.3 非线性实数理论求解现状

2.3.1 CDCL(T) 算法

CDCL(T) (Conflict Driven Clause Learning for Theories) 算法主要是在 SAT 问题的 CDCL 搜索框架上增加了理论求解器 (Theory Solver), 主要的搜索思路包括

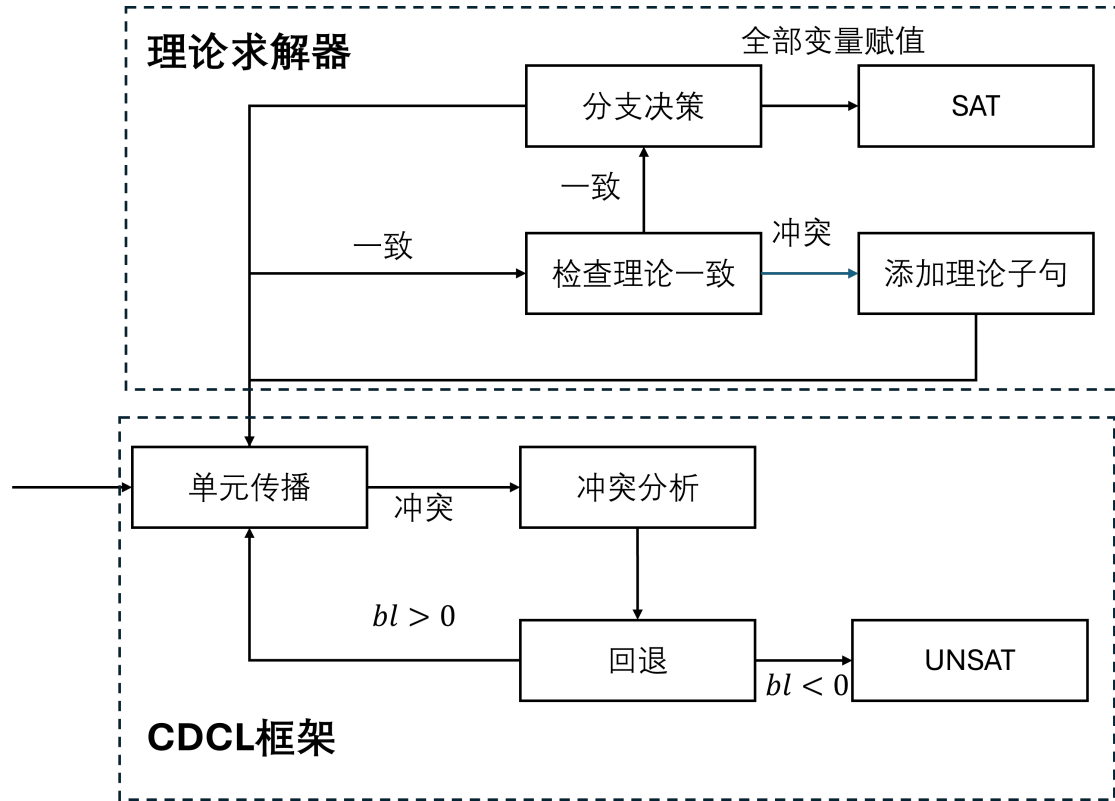


图 2.5 CDCL(T) 框架示意图。

Figure 2.5 Demo of CDCL(T) Framework.

单元传播、分支决策、冲突分析及回退等 CDCL 常见技术。理论求解器的作用是对给定的布尔骨架赋值进行推理，当全部布尔骨架得到赋值时通过理论分析返回问题可满足或者报告冲突。具体框架请参见图2.5。

非线性实数的理论求解器主要目的是判断多项式不等式集合的一致性，主要方法可以分为以下几种：

- **区间约束传播 (interval constraint propagation, ICP):** 通过变量或者单项式的赋值区间推导出多项式的赋值区间，根据这些区间快速判断一些常见的不一致性。
- **虚拟替代 (virtual substitution, VS):** 通过替换多项式中的变量，将多项式转化为更简单的形式，进而判断问题的可满足性。
- **柱形代数分解 (cylindrical algebraic decomposition, CAD):** 如前文所述，通过把 R^n 空间划分为多个符号一致的胞腔，逐一排查的方式来推出问题的可满足性。
- **增量线性化 (incremental linearization):** 把一些多项式线性化，然后使用线性求解器检测问题可满足性，一般是不完备的。

2.3.2 MCSAT/NLSAT 方法

MCSAT(Model-Constructing Satisfiability) 和 NLSAT(NonLinear Satisfiability) 是另一种具有 CDCL 结构的算法，核心的思想仍然是决策赋值、冲突分析和回

退。与 CDCL(T) 不同的是, MCSAT/NLSAT 算法加入了基于可行域的算术变量推理, 允许在脱离布尔骨架的情况下直接对算术变量进行赋值, 一些基本概念介绍如下:

- **决策层数 (level):** 指文字层面的决策层数, 包括布尔文字和算术文字。
- **决策阶段 (stage):** 指算术变量赋值次数。
- **日志 (trace):** 一种记录算法更新的线性结构, 包括可行域的更新、决策变量的赋值等。
- **算术传播 (arithmetic propagation):** 基于可行域的文字层面的传播, 可以直接推断某些文字的可满足性。

其中, MCSAT/NLSAT 最关键的步骤就是根据当前变量和文字的可行域关系进行算术文字赋值, 假设算术变量的可行域是 $curr_set$, 文字的可行域是 lit_set , 算术传播分为以下几种情况:

- $lit_set = \emptyset$: 任何赋值都不能使文字满足, 判断文字为 \perp 。比如当 $x \mapsto -1$ 时, 文字 $y^2 \leq x$ 不可满足。
- $lit_set = R$: 任何赋值都可以使文字满足, 直接赋值文字为 \top 。比如文字 $x^2 \geq -1$ 一定满足。
- $curr_set \subseteq lit_set$: 当前文字的可行域包括了算术变量的可行域, 因此可以直接判断文字满足。比如当 $x \in [-2, 2]$ 时, 文字 $x^2 \leq 10$ 一定满足。
- $curr_set \cap lit_set = \emptyset$: 任何可行域内的赋值都不会让文字满足, 因此直接赋值文字为 \perp 。比如当 $x \in [-2, 2]$ 时, 文字 $x^2 \geq 10$ 不可满足。

例2.7给出了一个 MCSAT/NLSAT 搜索的具体例子。

例 2.7. 考虑公式 $F = (P_1 : (x-2)^2 + y^2 \leq 4) \wedge (P_2 : x - y \geq 0)$ 。假设我们的搜索顺序是 $x < y$, 算法的搜索过程如下:

1. **决策算术变量:** 赋值 $x \mapsto 5$ 。
2. **可行域计算:** 计算 P_1 关于变量 y 的可行域是 \emptyset , 发现冲突。
3. **冲突分析:** 对公式 $\exists y. (x-2)^2 + y^2 \leq 4$ 进行量词消去, 通过柱形代数分解得到实根 $\{0, 4\}$, 生成新的学习子句 $x \leq 4$ 。
4. **回退:** 回退到 x 未赋值的状态。
5. **决策算术变量:** 重新赋值 $x \mapsto 2$ 。
6. **可行域计算:** 计算 P_1 关于变量 y 的可行域是 $[-2, 2]$, P_2 关于变量 y 的可行域是 $(-\infty, 2]$, 取交集得到可行域是 $[-2, 2]$ 。
7. **决策算术变量:** 赋值 $y \mapsto 0$ 。所有变量得到赋值, 返回 SAT。

2.3.3 局部搜索算法

2.3.4 其他方法

2.4 约束满足问题求解现状

首先, 本文先介绍 CSP 的定义, 它的特点是要求变量的值域有限且离散。前置概念的定义如下。

对于一个变量 x ，本文用 $D(x)$ 来表示 x 的域，即可以赋予 x 的一组可能的值。通常，使用简写 $x \in \{d_1, \dots, d_m\}$ 来定义 $D(x) = \{d_1, \dots, d_m\}$ 。特别地，在本文的讨论中仅考虑那些具有有限域的变量。

设 $Y = y_1, y_2, \dots, y_k$ 是一组有限的变量序列，其中 $k > 0$ 。一个关于 Y 的约束 $c \in C$ 是变量序列 Y 中变量的域的笛卡尔积的子集，即 $c \subseteq D(y_1) \times D(y_2) \times \dots \times D(y_k)$ 。约束可以写作 $c(Y)$ 或 $c(y_1, y_2, \dots, y_k)$ 。如果约束定义在两个变量上，称之为二元约束。如果约束定义在两个以上的变量上，称之为全局约束。

定义 2.18 (约束满足问题)。一个 CSP，由一组有限的变量序列 $X = x_1, x_2, \dots, x_n$ 及其各自的域 $D = D(x_1), D(x_2), \dots, D(x_n)$ 组成，同时伴有一组约束 C ，每个约束都在 X 的一个子序列上。为了简化表示，在表示特定的约束集合时经常省略大括号 “{}”，从而一个 CSP 被表示为 $P = (X, D, C)$ 。

之后，本文给出 CSP 解的定义。设 $P = (X, D, C)$ 为一个 CSP，其中 $X = x_1, x_2, \dots, x_n$ 且 $D = D(x_1), D(x_2), \dots, D(x_n)$ 。如果元组 $(d_1, \dots, d_n) \in D(x_1) \times \dots \times D(x_n)$ ，并且 $(d_{i_1}, d_{i_2}, \dots, d_{i_m}) \in c$ ，其中 $c \in C$ ，那么，称关于 X 的这组赋值满足在变量 $x_{i_1}, x_{i_2}, \dots, x_{i_m}$ 上的约束 C 。

对于变量序列 K ，定义 $D(K) = \bigcup_{x \in K} D(x)$ 。当变量 x 的域 $D(x)$ 是单元素集，即 $D(x) = \{d\}$ 时，将其简写为 $x = d$ ，记作单域。

一个 CSP 是一致的，如果这个 CSP 存在解。相反，不存在解的 CSP 是不一致的。一个失败的 CSP 指具有空域（也即存在变量的值域为空）的 CSP，或者只有单域的 CSP 且这些值并不能共同构成 CSP 的解。一个 CSP 是已解决的，若该 CSP 只有单域，并且其中的赋值可以共同构成 CSP 的解。需要注意的是，一个失败的 CSP 也是不一致的，但并非所有不一致的 CSP 都是失败的。

设 $P = (X, D, C)$ 和 $P' = (X, D', C')$ 都是 CSP。如果 P 和 P' 有相同的解集，那么它们被称为等价的。如果 P 和 P' 等价，且对所有的 $x \in X$ ，都有 $D(x) \subseteq D'(x)$ ，那么 P 小于 P' 。这种关系记作 $P \leq P'$ 。如果 $P \leq P'$ 并且至少存在一个 $x \in X$ 使得 $D(x) \subset D'(x)$ ，那么 P 严格小于 P' 。这种关系记作 $P < P'$ 。当 $P \leq P'$ 和 $P' \leq P$ 都成立时，记作 $P \equiv P'$ 。

2.4.1 约束规划求解

约束规划的目标是寻找给定的 CSP 的一个解（或所有解）。求解过程中，值域过滤、约束传播和搜索是交织进行的。

对 CSP 的解（或所有解）的寻找是通过迭代地将一个 CSP 分解为更小的 CSP 来进行的。这个分解过程被称为构建搜索树。树的一个节点代表一个 CSP。最初，在根节点，得到一个待解决的 CSP，记作 P_0 。如果 P_0 既没有被解决也没有失败，就将 P_0 分解为两个或更多的 CSP，记作 P_1, P_2, \dots, P_k ($k > 1$)。同时，必须确保 P_0 的所有解都被保留，而且没有多余的解被添加到 P_0 中。从而， P_1, P_2, \dots, P_k 的解集的并集等于 P_0 的解集。此外，每个 CSP P_i ($i > 0$) 应该

严格小于 P_0 ，以确保分解过程能够终止。接下来，会根据上述相同的标准分解每个 CSP P_1, P_2, \dots, P_k 。分解会一直进行，直到所有 CSP 被分解为失败的或已解决的 CSP。失败和已解决的 CSP 是搜索树的叶子节点。

搜索树的大小与 P_0 中变量数量呈指数关系。为了减小搜索树的大小，约束规划使用了一个叫做约束传播的过程。给定一个约束 C ，一个值域过滤算法会从 C 中的变量的值域中移除与 C 不一致的值。算法必须保留所有的解并且不向 C 中添加任何解。每当一个不一致的值域值被移除，其效果会通过所有其他共享相同对应变量的约束进行传播。这个迭代过程持续进行，直到在 CSP 中的所有约束都未检测到更多的不一致值。此时，可以称 CSP 是局部一致的。局部一致性反映了算法并没有得到一个全局一致的 CSP，而是得到了一个所有约束都是局部的，也即单独的、一致的 CSP。这里本文将介绍四个常见的局部一致性：弧一致性、超弧一致性、边界一致性和区间一致性。

定义 2.19 (弧一致性). 一个二元约束 $c(x_1, x_2)$ 是弧一致的，如果对于 x_1 的所有值 $d_1 \in D(x_1)$ ，存在一个值 $d_2 \in D(x_2)$ 使得 $(d_1, d_2) \in c$ ，并且对于 x_2 的所有值 $d_2 \in D(x_2)$ ，存在一个值 $d_1 \in D(x_1)$ 使得 $(d_1, d_2) \in c$ 。

定义 2.20 (超弧一致性). 一个约束 $c(x_1, \dots, x_m)$ ($m > 1$) 是超弧一致的，如果对于所有 $i \in \{1, \dots, m\}$ 和所有 $d_i \in D(x_i)$ ，存在值 $d_j \in D(x_j)$ (其中 $j \in \{1, \dots, m\} - i$)，使得 $(d_1, \dots, d_m) \in c$ 。

超弧一致性又称全局弧一致性 (Generalized Arc Consistency, GAC)，相较于弧一致性限制性更强，应用也更广。需要注意的是，弧一致性相当于应用于二元约束的超弧一致性。弧一致性和超弧一致性都确保每个域中的所有值都属于满足约束的元组，与当前变量域相关。另外两种局部一致性概念主要关注变量域的边界。因此，当应用这些定义时，假设涉及的变量域是一个固定的、线性有序的、有限域的元素集的子集。设 D 是这样的一个域，定义 $\min D$ 和 $\max D$ 分别为其最小值和最大值。此外，使用大括号 “{ }” 和方括号 “[]” 来分别表示一组域值和一个域值区间。

定义 2.21 (边界一致性). 一个约束 $c(x_1, \dots, x_m)$ ($m > 1$) 是边界一致的，如果对于所有 $i \in \{1, \dots, m\}$ 和每个值 $d_i \in \{\min D(x_i), \max D(x_i)\}$ ，存在值 $d_j \in [\min D(x_j), \max D(x_j)]$ (其中 $j \in \{1, \dots, m\} - i$)，使得 $(d_1, \dots, d_m) \in c$ 。

定义 2.22 (区间一致性). 一个约束 $c(x_1, \dots, x_m)$ ($m > 1$) 是区间一致的，如果对于所有 $i \in \{1, \dots, m\}$ 和所有 $d_i \in D(x_i)$ ，存在值 $d_j \in [\min D(x_j), \max D(x_j)]$ (其中 $j \in \{1, \dots, m\} - i$)，使得 $(d_1, \dots, d_m) \in c$ 。

在分解一个 CSP 之后，值域过滤和约束传播被应用到更小的 CSP 上。值的移除导致搜索树变小，从而加速了解决过程。同时，花费在约束传播上的时间应该小于它引发的加速，以便改善效果显著。关于约束传播过程的详细描述可以

在^[2]和^[3]中找到, 这里不再展开。总之, 本文希望应用高效的过滤算法, 其效率通常由应用于约束的局部一致性的概念来决定, 而在解决过程中何时应用哪种局部一致性的概念则取决于问题本身。

当前主流的求解器分为完备的和启发式的求解器, 一些求解器是商用的, 而一些则是开源的, 他们具有不同的特点, 本文介绍三个主流的 CP 求解器。

- **Choco**。Choco¹是一个基于 Java 的约束满足问题 (CSP) 求解库。它提供了一种声明式语言, 用于描述和解决各种复杂的约束满足问题。Choco 常被用作研究和教学工具, 并且支持各种类型的约束, 包括数值约束、逻辑约束等。

- **CPLEX Optimizer**。CPLEX Optimizer²是 IBM ILOG 的一款产品, 它包含一个称为 CP Optimizer 的模块, 这是一个约束规划求解器。其主要优点是拥有丰富的 API, 支持多种编程语言, 如 C++、Java、Python 等, 并且它的求解速度非常快, 可以处理非常大规模的问题。

- **Yuck**。Yuck 是一个基于 Scala 的运筹学和约束规划库, 它的主要特点是它采用了一种基于邻域搜索的算法, 可以在解空间中快速找到优质的解。作为一款启发式求解器, 他在近几年的 CP 求解启发式赛道中都是第一名。

这些求解器都支持集成在 Minizinc^[2]中。MiniZinc 是一种中间级别的约束建模语言, 设计用于描述各种组合优化问题。它的主要目标是提供一种简单、可移植且高效的方式来描述这些问题, 以便可以使用各种不同的求解器来解决它们。并且, 它支持数组、集合和函数等高级数据结构, 以及各种数学和逻辑运算符, 并且可以使用 CPLEX、Choco、Gecode 等求解器作为后端求解器, 这使得用户可以在不改变模型的情况下尝试不同的求解器和算法。

2.4.2 其他求解方法

除了 CP 求解器外, 有许多其他的求解技术也可以用来处理 CSP, 它们来自于不同的研究领域。其中, 布尔可满足性 (SAT) 求解器、可满足性模理论 (SMT) 求解器和整数线性规划 (ILP) 求解器是最常用的三种。接下来, 本文将介绍这三种求解技术, 以及它们在处理 CSP 时的优势和应用情景。

逻辑公式可满足性问题要求判定一个逻辑公式是否存在使得公式为真的赋值, SAT 和 SMT 问题是其中最主要的两个可满足性问题。SAT 问题是第一个被证明为 NP 完全的问题, 是计算机科学的核心问题, 也是数理逻辑的基础问题。然而, SAT 问题的表达能力和应用范围有其局限性。许多现实问题牵涉到各种领域知识, 这就需要用表达能力更强的逻辑公式, 比如一阶逻辑公式。尽管一阶逻辑语言是不可判定的, 但许多应用只需处理一些背景理论来解释特定的谓词和函数记号的一阶逻辑公式。这类公式被称为 SMT 公式, 它对 SAT 问题进行了扩展, 将其中的布尔变量用背景理论谓词取代。常见的 SAT 求解器有 MiniSAT、Kissat 等, 而主流的 SMT 求解器则包含 Z3、CVC5 等。

¹<https://choco-solver.org/>

²<https://www.ibm.com/products/ilog-cplex-optimization-studio>

定义 2.23 (布尔可满足性). 一个 SAT 问题由一系列的布尔变量和逻辑运算符（如与、或、非）构成，对该问题的求解涉及到查找一个给定的布尔公式的满足赋值，使得整个公式的结果为真。

定义 2.24 (可满足性模理论). 一个 SMT 问题可以表示为一个三元组，包含变量集合（可以是整数、实数或者字符串）、函数符号集合（比如常见的运算符、量词等操作）、以及由变量和函数组成的约束集合。对该问题的求解涉及到一组变量的赋值，使得约束集中的约束全部可满足。

SAT 问题是 NP 完全问题，但现代的 SAT 求解器已经能够处理数百万个变量和数千万个约束的问题。将 CSP 转化为 SAT 问题的关键步骤是将 CSP 的变量和约束转化为布尔变量和布尔公式。具体来说，对于 CSP 中的每一个变量和它的每一个可能的值，都定义一个对应的布尔变量。如果 CSP 变量被赋予该值，则对应的布尔变量为真；否则为假。然后，可以使用“at most one”和“at least one”的编码方式来保证 CSP 中的每个变量只能取一个值。同时，CSP 的其他约束也可以通过定义相应的布尔公式来实现。

将 CSP 编码为 SMT 问题的过程相对直接。由于 SMT 支持整数变量和约束，以及各种数学函数，因此可以直接将 CSP 中的变量、约束和函数映射到 SMT 问题中。这样，CSP 就可以被转化为一个等价的 SAT/SMT 问题，进而利用相应求解器求解。一般来言，在编码 CSP 时，SAT 的编码复杂度要高很多，但比涉及高阶逻辑的 SMT 求解器拥有更好的性能。

最后是 ILP，它是线性规划（LP）的一种扩展。线性规划问题包含一组线性约束条件和一个线性目标函数，问题的目标是在满足这些约束条件的前提下，找到使目标函数取得最大（或最小）值的变量的取值，而 ILP 则要求所有的决策变量都是整数。由于 CSP 中的变量通常也是整数，因此 CSP 可以很自然地映射为 ILP 问题。常见的 ILP 求解器有 Gurobi、SCIP 等。

这些算法为解决 CSP 提供了新的视角，尽管某些转换可能相对复杂，可能会影响性能或效率，但它们也弥补了 CP 求解器在特定问题时的不足。

2.5 AllDifferent 约束求解现状

由于目前 AllDifferent 约束的过滤算法及后续算法都要用到图论的相关知识，所以本文先介绍图论以及相关的最大匹配算法，之后介绍 AllDifferent 约束的过滤算法以及基于该算法的一些改进策略。

2.5.1 图论基础

本文用 $G = (V, E)$ 表示一个（无向）图，其中 V 是一个有限的顶点集，而 E 是来自 V 的无序对的多重集，称为边。顶点 $u \in V$ 和 $v \in V$ 之间的边记作 uv 。如果存在一个分区 S, T ，使得 $E \subseteq \{st | s \in S, t \in T\}$ ，那么图 $G = (V, E)$ 称为二部图，写作 $G = (S, T, E)$ 。

在图 $G = (V, E)$ 中, 游走指一个序列 $P = v_0, e_1, v_1, \dots, e_k, v_k$, 其中 $k \geq 0$, $v_0, v_1, \dots, v_k \in V$, $e_1, e_2, \dots, e_k \in E$, 并且对于 $i = 1, \dots, k$, 有 $e_i = v_{i-1}v_i$ 。如果 v_0, \dots, v_k 是不同的, 那么这个游走就被称为一条路径。一个闭合的路径, 即 $v_0 = v_k$, 称为回路。

图 $G = (V, E)$ 的子图是一个图 $G' = (V', E')$, 它满足 $V' \subseteq V$ 和 $E' \subseteq \{uv | u \in V', v \in V', uv \in E\}$ 。称子图 $G' = (V', E')$ 为图 $G = (V, E)$ 的最大连通子图, 若对于 V' 中的每一对 u, v , 在 G' 中都存在一条 u 到 v 的路径。

有向图是一个对 $G = (V, A)$, 其中 V 是一个有限的顶点集, A 是来自 V 的有序对的多重集, 称为弧, 从 $u \in V$ 到 $v \in V$ 的弧记作 (u, v) 。类似于无向二部图, 如果存在一个分区 S, T , 使得 $A \subseteq \{(s, t) | s \in S, t \in T\}$, 那么有向图 $G = (V, A)$ 就是二部图。也写作 $G = (S, T, A)$ 。此外, 关于有向图的路径、回路、子图、最大连通子图等概念和无向图类似, 这里不再展开。

在有向图中, 如果存在两个顶点 s 和 t , 从 s 可以通过一条有向路径到达 t , 同时也可以从 t 通过一条有向路径到达 s , 那么就说两个顶点相互到达。如果在一个有向图中, 存在一个顶点集 V , 对于集合中的任意两个顶点 u 和 v 都可以相互到达, 称 V 为强连通分量 (Strongly Connected Component, SCC)。

图 $G = (V, E)$ 的一个匹配指一组不相交的边集 $M \subseteq E$, 即 M 中的任意两条边不共享顶点。如果 M 包含了所有属于 $S \subseteq V$ 的顶点, 称匹配 M 覆盖了 S 。如果 M 没有覆盖顶点 $v \in V$, 那么 v 就被称为 M -free, 又称自由节点; 如果 M 覆盖了顶点 $v \in V$, 称 v 为匹配节点。匹配 M 的大小是 $|M|$, 若最大匹配包含了图中所有的顶点, 那么这个匹配被称为完全匹配。

定义 2.25 (最大图匹配). 给定一个无向图 $G = (V, E)$, 其中 V 是顶点集, E 是边集。最大图匹配问题要求找出图中一个最大的匹配。

设 M 是图 $G = (V, E)$ 中的一个匹配。如果一条路径 P 的长度是奇数, 它的两端没有被 M 覆盖, 并且它的边交替出现在 M 的外部 and 内部, 那么称 P 为 M -增广路径。如果回路 P 的边交替出现在 M 的外部 and 内部, 那么称 P 为 M -交替路径。在 M -增广路径上, 可以交换在 M 中和不在 M 中的边, 得到的 M' 仍然是一个匹配, 并且其中的 $|M'| = |M| + 1$ 。因此可以得到如下的定理, 证明如下。

证明. 如果 M' 是一个比 M 大的匹配, 考虑图 $G' = (V, M \cup M')$ 。在 G' 中, 每个顶点最多连接两条边。因此, G' 的每个组件要么是一个回路, 要么是一条路径 (可能长度为零)。由于 $|M'| > |M|$, 因此至少有一个组件包含的 M' 的边比 M 的边多。因为所有的回路都包含偶数条边, 所以这个组件必须是一个 M 增广路径。如果 M' 是一个比 M 大的匹配, 考虑图 $G' = (V, M \cup M')$ 。在 G' 中, 每个顶点最多连接两条边。因此, G' 的每个组件要么是一个回路, 要么是一条路径 (可能长度为零)。由于 $|M'| > |M|$, 因此至少有一个组件包含的 M' 的边比 M 的边多。因为所有的回路都包含偶数条边, 所以这个组件必须是一个 M 增广路径。□

定理 2.1. 设 $G = (V, E)$ 是一个图, M 是 G 中的一个匹配。那么 M 要么是最大匹配, 要么存在一个 M -增广路径。

因此, 可以通过在 G 中迭代计算 M -增广路径并扩展 M 的方式来求一个图最大匹配, 它是求解最大图匹配的一类经典方法, 步骤如下。

设 $G = (U, W, E)$ 是一个二部图, M 是 G 的一个匹配。通过将 M 中的所有边从 W 指向 U , 并将所有其他边从 U 指向 W , 构造有向二部图 $G_M = (U, W, A)$, 即:

$$A = \{(w, u) | uw \in M, u \in U, w \in W\} \cup \{(u, w) | uw \in E \setminus M, u \in U, w \in W\}$$

然后, 在 G_M 中从 U 中的一个自由顶点开始并在 W 中的一个自由顶点结束的每一条有向路径对应于 G 中的一个 M -增广路径。通过选择 $|U| \leq |W|$, 最多需要找到 $|U|$ 条这样的路径。由于每条路径可以通过广度优先搜索在最多 $O(|A|)$ 时间内被识别, 因此这个算法的时间复杂度是 $O(|U||A|)$ 。

在^[2]中提出了其改进算法, 可以在 $O(|V|^{1/2}|A|)$ 时间内运行, 其中 $V = U \cup W$ 。其思想为, 与其反复沿着单个 M -增广路径增强 M , 不如反复同时沿着一组不相交的 M -增广路径增强 M 。这样的路径集合可以再次在 $O(|A|)$ 时间内找到。可以证明, 在 $|V|^{1/2}$ 次迭代之后, 通过对路径长度的推理, 最多可能还有 $O(|V|^{1/2})$ 次迭代, 这导致总的时间复杂度为 $O(|V|^{1/2}|A|)$ 。而在一个普通的图 $G = (V, E)$ (不一定是二部图) 中, 可以在 $O(|V||E|)$ 时间内^[2] 或者 $O(|V|^{1/2}|E|)$ 时间内^[2] 计算出最大的匹配。

2.5.2 AllDifferent 约束和二部图匹配

本文首先展示 AllDifferent 约束的解和二部图中的匹配的等价性。为此, 本文将 AllDifferent 约束涉及的变量使用二部图进行表示, 称为值图。它的定义如下, 注意, 定理中的匹配 M 覆盖了 X , 因此是最大匹配。此外, 本文给出了一个例子来描述这一转换过程。

定义 2.26 (值图). 设 X 是一组变量集合, 则二部图 $G = (X, D(X), E)$, 其中 $E = \{xd | d \in D(x), x \in X\}$, 被称为 X 的值图。

定理 2.2. 设 $X = x_1, x_2, \dots, x_n$ 是一系列变量, G 是 X 的值图。那么 $(d_1, \dots, d_n) \in \text{AllDifferent}(x_1, \dots, x_n)$ 当且仅当 $M = \{x_1d_1, \dots, x_nd_n\}$ 是 G 中的一个匹配。

证明. 在 M 中的边 x_id_i (对于某个 $i \in \{1, \dots, n\}$) 对应于赋值 $x_i = d_i$ 。由于 M 中的边不共享顶点, 所以对于所有 $i \neq j$, 都有 $x_i \neq x_j$ 。□

例 2.8. 假设有四个任务 (1, 2, 3 和 4) 要被分配给五台机器 (A, B, C, D 和 E)。每台机器最多只能分配一个任务, 但并非每个任务都能分配给每台机器。图2.6左中展示了可能的组合, 例如, 任务 2 可以分配给机器 B 和 C。由于任务必须分配

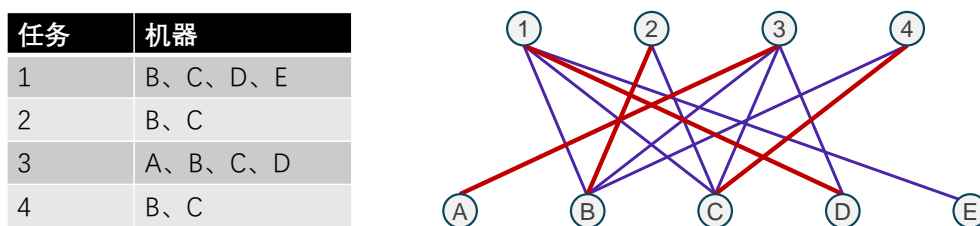


图 2.6 AllDifferent 约束和其二部图表示。

Figure 2.6 AllDifferent constraint and its bipartite graph representation.

给不同的机器，可以使用 AllDifferent 约束表述该问题，从而将问题建模为 CSP。具体地，引入一个变量 x_i ，用于表示任务 $i = 1, \dots, 4$ ，其赋值代表任务 i 被分配到的机器。变量的初始值域为图 2.6 左表示的任务和机器之间可能的组合，下式是构造的 CSP 约束：

$$\text{AllDifferent}(x_1, x_2, x_3, x_4).$$

其中， $x_1 \in \{B, C, D, E\}$, $x_2 \in \{B, C\}$, $x_3 \in \{A, B, C, D\}$, $x_4 \in \{B, C\}$ 。设 $X = x_1, \dots, x_n$ ，其值图如图 2.6 右所示。值图中的红粗线表示覆盖 X 的一个匹配，它对应于 CSP 的解，即 $x_1 = D$, $x_2 = B$, $x_3 = A$ 和 $x_4 = C$ 。

最后，本文要介绍 Hall 的婚配定理^[2]，它是用于推导 AllDifferent 约束过滤算法的图论中的一个经典定理，旨在描述一个二部图存在完全匹配的必要和充分条件。这个定理的直观解释是，如果每个女孩（A 集合）都有足够多的男孩（B 集合）可以选择，那么就有可能为每个女孩找到一个男孩作为配对，形成一个完全匹配，反之亦然。接下来，本文给出 AllDifferent 约束语境下，该定理的形式化描述，证明略。

定理 2.3. 约束条件 $\text{AllDifferent}(x_1, \dots, x_n)$ 有解，当且仅当对于所有的 $K \subseteq \{x_1, \dots, x_n\}$ ，下式成立：

$$|K| \leq |D(K)|$$

例 2.9. 考虑如下 CSP：

$$\text{AllDifferent}(x_1, x_2, x_3, x_4).$$

其中， $x_1 \in \{2, 3\}$, $x_2 \in \{2, 3\}$, $x_3 \in \{1, 2, 3\}$, $x_4 \in \{1, 2, 3\}$ 。可见，对于任何 $K \subseteq \{x_1, x_2, x_3, x_4\}$ 且 $|K| \leq 3$ ，都有 $|K| \leq |D(K)|$ 。然而，对于 $K = \{x_1, x_2, x_3, x_4\}$ ，有 $|K| > |D(K)|$ 。由定理 2.3 可知，这个 CSP 没有解。

2.5.3 AllDifferent 约束的过滤算法

前面，本文介绍了局部一致性的概念，本文以弧一致性为例介绍 AllDifferent 约束如何应用局部一致性，本文首先利用二元分解将 AllDifferent 约束分解成一

组二元约束，其定义如下。

定义 2.27 (二元分解). 设 C 是变量 x_1, \dots, x_n 上的一个约束。 C 的二元分解指在 x_1, \dots, x_n 的变量对上的一组最小的二元约束 $C_{dec} = \{C_1, \dots, C_k\}$ (其中 k 为整数且 $k > 0$)，使得 C 的解集等于 $\bigcap_{i=1}^k C_i$ 的解集。

关于 $\text{AllDifferent}(x_1, x_2, \dots, x_n)$ 的二元分解是

$$\bigcup_{1 \leq i < j \leq n} \{x_i \neq x_j\}. \quad (2.1)$$

建立二元分解上的弧一致性的过滤算法很简单：每当一个变量的域只包含一个值时，这个值就从在 AllDifferent 约束中出现的其他变量的值域中被移除。这个过程一直重复，直到没有更多的变化发生或者一个域变为空。通过这个算法，可以建立 AllDifferent 约束上的局部一致性，或者证明它是不一致的。换句话说，当一个值域包含多个元素时，就不能推理出更多的信息了。

这个算法的一个缺点是，需要 $\frac{1}{2}(n^2 - n)$ 个不等约束来表示一个 n 元的 AllDifferent 约束，从而这种方法的最坏情况时间复杂度是 $O(n^2)$ 。另一个更重要的缺点是信息的丢失。当二元约束集合被弧一致化时，一次只比较两个变量。而当 AllDifferent 约束被超弧一致化时，所有的变量都被同时考虑，这允许更强的局部一致性，下面本文给出一个简单的例子。

例 2.10. 设一个 CSP 包含三个变量 x_1, x_2, x_3 ，其中 $x_1 \in \{2, 3\}, x_2 \in \{2, 3\}, x_3 \in \{1, 2, 3\}$ 。对于约束 $\text{AllDifferent}(x_1, x_2, \dots, x_n)$ ，在不采用二元分解时，通过超弧一致性，可以删减 x_3 值域中的 2 和 3；而对该 AllDifferent 约束使用二元分解时，由于约束被分解为多个二元约束，此时无法推导出任何信息。

前面，提到了 AllDifferent 约束和二部图的等价性，以及求解二部图最大匹配的方法。接下来，本文介绍一种利用二部图实现的 GAC 算法，它基于最大匹配和强连通分量，称为 *Régin* 算法^[2]。算法思路是构造残差图表示 AllDifferent 约束，并通过在图上作图匹配删减冗余边，实现在约束上的全局一致性。

后文需要用到前面定义的 CSP 的一些概念，为了方便后文叙述，本文额外引入几个新定义。约束 c 的变量集是它所限制的有序变量集，记作 $X(c) = \{x_1, x_2, \dots, x_r\}$ 。 c 的值域是 $X(c)$ 的值域的并集，记作 $D(c) = \bigcup_{x \in X(c)} D(x)$ 。此外，本文使用 $B_c(a)$ 来表示 $X(c)$ 中值域包含 a 的变量集 (即， $B_c(a) = \{x | x \in X(c), a \in D(x)\}$)。在 AllDifferent 约束表示的二部图中，冗余边指在任何最大匹配中都不出现的边。允许边则表示属于一些但不一定所有最大匹配的边，冗余边和允许边是互补的。

一个节点是允许的，当且仅当对于一个任意的最大匹配 M ，它可以通过一个从自由节点开始的偶数交替路径到达。交替路径中的变量节点集被记为 Γ ，其值节点集被记为 A 。关于允许边有如下定理，证明略。

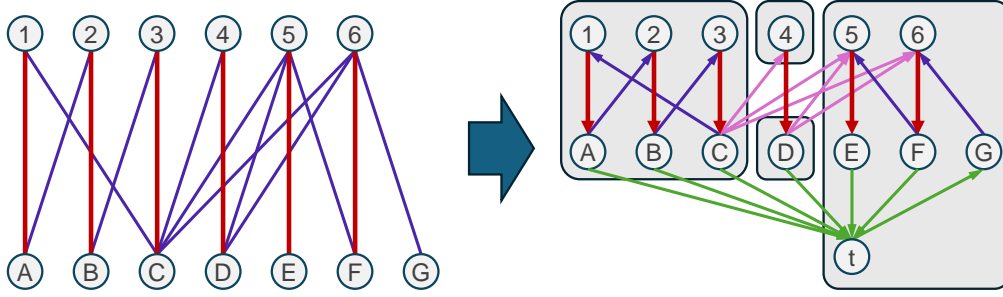


图 2.7 值图与残差有向图的转换。

Figure 2.7 Value graph to residual directed graph transformation.

定理 2.4. 一个 *AllDifferent* 约束是 *GAC* 的，当且仅当它的值图的每个边都是允许边，也即属于在 $B(c)$ 中覆盖 $X(c)$ 的一些匹配。

定理 2.5 (Berge 定理). 一个边是允许的，当且仅当，对于一个任意的最大匹配 M ，它属于从自由节点开始的偶数增广路径或者偶数交替路径。

基于上述定理，Régin 提出了执行 *GAC* 的第一个 *AllDifferent* 约束算法，通过从值图中移除冗余边来实现 *GAC*。该算法首先计算一组 *AllDifferent* 约束的值图的最大匹配，通过引入新的节点和将边改为有向边，构造残差有向图如下。

定义 2.28 (残差有向图). 残余有向图被定义为 $R = \langle V_R, E_R \rangle$ ，其中 $V_R = X(c) \cup D(c) \cup \{t\}$ ， t 是与 $D(c)$ 连接的一个新引入的汇节点， $E_R = E_M \cup E_U \cup E_{t_1} \cup E_{t_2}$ 。具体来说，每个边 $e \in E_M \cup E_U$ 表示约束 c 中的一个变量-值对。匹配边 E_M 将变量连接到它们各自的匹配值： $E_M = \{x \rightarrow a | x \in X(c), a = M(x)\}$ 。未匹配的边 E_U 将值连接到它们各自的未匹配变量： $E_U = \{a \rightarrow x | a \in D(c), x \in B_c(a) \setminus \{M(a)\}\}$ 。 E_{t_1} 中的边将匹配值节点连接到 t ： $E_{t_1} = \{a \rightarrow t | M(a) \in X(c)\}$ 。 E_{t_2} 中的边将 t 连接到未匹配的值节点（自由节点）： $E_{t_2} = \{t \rightarrow a | M(a) \notin X(c)\}$ 。

通过残差有向图，可以发现，通过引入汇节点 t ，从自由节点开始的偶数增广路径被扩展为偶数的交替路径，从而允许边表现为残差图上的交替路径。此外，很容易发现，交替路径上的顶点之间的其他边也可以扩展成一个交替路径，因此也属于允许边。从而得到如下定理：

定理 2.6. 一个非匹配边是允许的，当且仅当其顶点处于同一个强连通分量中。

因此，算法会计算残差有向图的强连通分量（SCC），这是算法最耗时的部分，通常通过 Tarjan 算法来解决。之后会删除两个端点处于不同强连通分量的非匹配边。图2.7展示了一个简单的值图转换为残差图的例子，在右图中每个框表示一个 SCC，而紫红色的线段就是应当被删除的冗余边。

近几年,有许多基于该算法的改进工作,主要思路是减少或避免 SCC 的计算。例如:SCC-分割^[21],它只计算在回溯搜索过程中其变量的域已经改变的 SCC,从而减少了计算 SCCs 所需的变量范围;匹配优化^[21],它证明了冗余边可以分为两类,其中一类可以直接删除,无需计算相应的 SCCs;早期检测^[21],它避免了计算那些删除后不会分割当前 SCCs 的无关紧要的边;在文献^[21]中,证明了二部图中的所有冗余边都指向某些交错环,从而对值进行筛选;在文献^[21]中,发现 GAC 算法只关心图模型的一个节点是否在 SCC 中,而不是它属于哪个 SCC,从而使用比特位提高算法的性能。

2.6 局部搜索算法现状

局部搜索是相对于完备搜索来说的,它是一种用于解决优化问题的常用启发式策略,而几乎所有问题都可以看作是一类优化问题,例如组合优化问题。局部搜索的基本思想是从一个初始状态开始,然后在解的邻域中寻找更好的状态,通过一系列的移动来改进当前解。一些基本的概念的定义如下。

定义 2.29 (状态). 状态是解空间中的一个元素,表示问题的一个可能解。

定义 2.30 (移动). 移动是从当前状态转移到邻域中的另一个状态的操作。

定义 2.31 (打分函数). 打分函数(评价函数)是一个将解空间中的每个状态映射到实数的函数,用于评估每个状态的优劣,即解的质量。

定义 2.32 (邻域). 邻域是一个状态执行一次移动后的所有可达状态的集合。

为了实施局部搜索算法,首先需要明确候选解的邻域结构,即确定哪些候选解可以被视为邻居。一旦邻域结构被定义,问题的解空间就可以被视为一个图形结构。因此,局部搜索实际是在图论问题上求解。以下是其求解流程:

1. 设 $S = \{s_1, s_2, \dots, s_n\}$ 为所有候选解构成的解空间,也即状态空间。
2. 定义初始状态 $s_0 \in S$,一般由构造的初始化函数得到。
3. 定义邻域函数 $N : S \rightarrow \mathcal{P}(S)$,将每个状态映射到其邻域的状态集合。
4. 定义目标函数 $f : S \rightarrow \mathbb{R}$,用于评估每个状态的好坏,即解的质量。
5. 设置停止条件,当满足某个特定条件时终止算法。常见的停止条件如达到最大迭代次数、最大时间限制等。
6. 由打分函数(可能多个)构成移动规则,指导从当前状态的邻域中选择下一个状态。一般选择使得目标函数变好最多的状态。
7. 从初始状态 s_0 开始,按照移动规则在邻域中选择下一个状态,更新当前状态。重复这个过程,直到达到最优解或满足其他停止条件。

若将被满足的约束数量作为最小化目标,则 CSP 可以被看作组合优化问题,通过局部搜索,寻找使得 CSP 中一致约束最多的赋值。本文以一个简单的 CSP 为例,介绍上面提到的概念和算法流程。

例 2.11. 假设有一个简单的约束满足问题(CSP),其中有三个变量 $X = \{x_1, x_2, x_3\}$, 每个变量的取值范围是 $\{1, 2, 3\}$ 。目标是找到一组赋值 $A = \{a_1, a_2, a_3\}$, 使得所有的约束 $C = \{c_1, c_2, c_3\}$ 都被满足。在这个问题中, 状态就是变量的一组赋值, 比如 $s = \{2, 1, 3\}$ 。移动就是改变一个变量的赋值, 比如从状态 $\{2, 1, 3\}$ 移动到状态 $\{2, 2, 3\}$ 。目标函数就是被满足的约束数量, 本文希望最小化未被满足的约束数量, 也就是最大化被满足的约束数量。邻域就是通过一次移动可以达到的所有状态, 比如从状态 $\{2, 1, 3\}$ 可以移动到的状态有 $\{1, 1, 3\}$ 、 $\{3, 1, 3\}$ 、 $\{2, 2, 3\}$ 、 $\{2, 3, 3\}$ 、 $\{2, 1, 1\}$ 和 $\{2, 1, 2\}$ 。此时从一个初始赋值开始, 通过移动来最小化未满足约束的数量, 直到得到可满足解。

近几年, 局部搜索算法发展出了很多策略, 这些策略可以根据问题的特性和需求进行组合, 以提高局部搜索算法的效率和效果。下面是一些例子:

- **禁忌搜索**^{[2]1}: 这种策略旨在避免在已经访问过的区域中陷入循环。在禁忌搜索中, 会维护一个禁忌表, 记录一些在近期内已经访问过的状态。当从当前状态的邻域中选择下一个状态时, 禁忌表中的状态将被排除在外。

- **重启策略**: 重启策略能够有效避免搜索过程陷入局部最优解。若如果搜索过程在一段时间内没有找到更好的解, 算法会随机选择一个初始状态重新开始搜索。这样可以使搜索过程有机会探索解空间的其他区域。

- **迭代局部搜索**^{[2]1}: 这种策略在没有达到局部最优时, 持续执行迭代改进算法。当达到局部最优时, 执行随机游走算法, 即从当前的局部最优解出发, 进行一次或多次随机的移动, 以跳出当前的局部最优区域。策略的目标是搜索一系列的局部最优解, 然后返回最好的一个作为结果。

- **解池技术**: 解池技术在搜索过程中不仅保存当前最优解, 还保存一些具有代表性的非最优解, 这些解构成了一个解池。通常他和重启策略一起使用, 在重启时, 算法从解池中选择一个解执行随机游走得到初始状态, 重新开始搜索。

在近十年来, 已经有不少工作采用局部搜索来对 CSP 问题进行求解, 并帮助解决了许多此前未解决的难题, 下面本文列举一些经典及最新的工作。

在文献^{[2]1}中, 作者设计了一种新的启发式算法, 它利用了问题在约束和变量方面的结构, 可以比全局代价函数更精确地指导搜索进行优化 (例如, 违反约束的数量)。该工作在拉丁方、N 皇后、全间隔等问题上进行了实验, 并且取得了较好的实验结果。在文献^{[2]1}中, 作者设计了将拉丁方问题转化为图的方法, 从而将问题归约到图着色问题上, 并使用遗传算法, 对给定补全拉丁方实例进行求解, 取得了较好的求解效率。在文献^{[2]1}中, 作者设计了针对数独问题的蚁群算法, 并尝试了多种邻域的定义, 算法可以在 25x25 阶的困难数独实例上取得高达 95% 的求解成功率。在文献^{[2]1}中, 作者设计了针对性的解池技术和子代价函数, 并通过高效的化简规则对 CSP 进行了化简, 其在拉丁方问题上, 可以求解绝大部分 70 阶的拉丁方补全问题。

2.7 本章小结

本章节介绍了 CSP、AllDifferent 约束的形式化定义以及主流的求解它们的算法。在第一节中，本文介绍了 CSP 的求解现状，主要介绍了 CP 求解器，以及其他可用于求解的技术，如 SAT、SMT、ILP 求解器等。在第二节中，本文介绍了 AllDifferent 约束的求解现状，主要介绍了它和图匹配的关系，以及两种主流的针对该约束的过滤算法。在第三节中，本文介绍了局部搜索算法的概念和求解流程，以及相关的各类技术。

第3章 AllDiff-LS 的基础组件和算法框架

本章首先对本文的研究问题进行详细介绍，即如何设计一个高效的求解算法来求解一个包含 AllDifferent 约束的约束满足问题。基于以上研究问题，本章介绍了本文提出的适用于 AllDifferent 约束求解的局部搜索算法。首先，本章节讨论了使用什么数据结构对约束进行表示，以及如何对约束进行建模。具体地，算法采用了一个包含两种类型的顶点和边的异构图来表示 AllDifferent 约束，不是单独处理每个 allDifferent 约束，而是将它们视为一个整体。其次，算法采用两个低复杂度的顶点简化规则对图进行化简，以减少不必要的搜索空间。接下来，算法针对构造的图设计求解算法，提出了分两步选择操作的方法，以及相应的打破平局和禁忌策略。最后，本文给出算法的整体流程图，并展示了包含 AllDifferent 约束的 CSP 是如何被求解的。

3.1 问题介绍和求解思路

本文主要研究以 AllDifferent 约束为主的 CSP 问题，这是一类特殊的 CSP，其中所有的全局约束由 AllDifferent 约束组成，但允许出现其他类型的二元约束，使得 CSP 约束的全部或大部由 AllDifferent 约束组成。在这类问题中，一个变量或包含它的变量表达式将受到多个 AllDifferent 约束的共同限制，问题的目标是寻找使得所有约束满足的变量赋值。这类问题的一个主要特点是，由于 AllDifferent 约束的全局性，解空间的大小会随着问题规模的增大而指数级增长。因此，寻找有效的求解策略和算法以提高求解效率，是研究这类问题的一个重要方向。

使用 AllDifferent 约束可以编码很多经典的 CSP，例如数独、n 皇后、All-Interval 和各类拉丁方问题。以数独问题为例，问题中每一行、每一列以及每一个宫（3x3 的小格子）的数字都需要满足 AllDifferent 约束，即这些位置上的数字都需要是 1 到 9 且互不相同。而除此之外，数独问题没有其他的约束条件。

设 $P = (X, D, C)$ 是一个约束满足问题，其中 C 是约束集。定义 $AD(c)$ 成立，当且仅当 c 是一个 AllDifferent 约束。取 $C_A = \{c | c \in C, AD(c)\}$ 为问题中所有 AllDifferent 约束组成的集合，本章节后续部分将主要研究如何获得在该约束集上的可满足赋值。不失一般性的，在本章节的最后会介绍如何将其他二元约束整合到求解的框架中，从而得到整个 CSP 的可满足赋值。

考虑到 AllDifferent 约束强大的约束力，现有的求解器在解决约束数量和长度不断增加的各种 CSP 问题时，表现出了性能下降的趋势。本文注意到在 AllDifferent 约束中，出现的大部分变量都是用来表示变量表达式的，同一个变量不可避免地会出现在多个约束中，这导致约束之间存在强烈的关联性。此前，关于 AllDifferent 约束求解的算法分为两类，一类是借助过滤算法对约束化简后再对问题进行回溯求解，第二类则是通过提前满足 AllDifferent 约束的方式将约

束作为隐含约束处理。这两种策略都是将问题中的 AllDifferent 约束分别进行处理，而没有考虑约束之间也存在约束关系。此外，随着约束数量的增多，第二类策略无法处理绝大多数 AllDifferent 约束，使得策略的有效性会大的折扣。

基于以上观察，本文并没有单独对待每一个 AllDifferent 约束，而是将它们作为一个整体来处理。本文采用了包含两种类型顶点和边的异构图来表示 AllDifferent 约束，称为 AllDifferent 约束图 (ACG)。构造这个图的动机来源于针对 AllDifferent 约束的二元分解，本文在此基础上将变量和变量表达式的概念进行了区分，从而使得此框架可以应用于更为广泛的弧一致性算法。

3.2 AllDifferent 约束转化为图及其优化

设 $P = (X, D, C)$ 是一个约束满足问题， C_A 是其 AllDifferent 约束组成的集合。约为了表示该问题，本文引入一个异构图 \mathcal{G} ，称之为 AllDifferent 约束图 (AllDifferent Constraint Graph, ACG)。为了方便后续描述，约束中涉及的基本单位是变量表达式，其中，变量表达式可以是包含加减乘除等操作的复杂表达式，也可以是单位表达式。例如，在约束条件 $AllDifferent(x_1, x_2 + x_3)$ 中， x_1 和 $x_2 + x_3$ 都是表达式。异构图的形式化表示如下。

定义 3.1 (AllDifferent 约束图). 异构图被定义为 $\mathcal{G} = (V, E)$ ，其中 $V = X \cup NV$ ，表示顶点集合，其中 NV 指出现在 AllDifferent 约束中的变量表达式集合。 $E = E_p \cup E_x$ 表示边集合，它由两种类型的边表示。具体来说，一种类型是变量表达式之间的边，称为表达边，用 E_p 表示，如果两个变量表达式 p 和 q 出现在同一个 AllDifferent 约束中，它表示为 $(p, q) \in E_p$ 。另一种类型是变量和变量表达式之间的边，用 E_x 表示，如果一个变量表达式 p 包含一个变量 x ，则有 $(p, x) \in E_x$ 。

AllDifferent 约束图的思想是，将全局约束用二元分解分解为多组二元约束的并集，再针对二元约束构成的二元关系以及变量表达式和变量之间的二元关系，构造图结构。和使用残差图对 AllDifferent 约束进行表示不同，本文没有引入变量-值这一关系。通过隐去值域这一信息，ACG 得以表示多组 AllDifferent 约束，从而将所有约束使用一张图进行表示，以施展更强的弧一致性算法。下面，本文通过一个简单的例子来介绍 ACG。

例 3.1. 给定四个变量 x_1, x_2, x_3, x_4 ，其有限值域分别为 $\{1, 2, 3\}$ ， $\{1, 2\}$ ， $\{2, 3\}$ 和 $\{1, 2, 3\}$ 。如果要求 x_1, x_2 和 $x_3 - 1$ 的赋值不能相同，并且要求 $x_1 + x_2$ 和 $x_3 + x_4$ 的赋值不能相同，那么可以如下建模此问题：

$$AllDifferent(x_1, x_2, x_3 - 1), \quad (3.1)$$

$$AllDifferent(x_1 + x_2, x_3 + x_4), \quad (3.2)$$

容易看出，该问题的一个解是 $\mathcal{A} = (3, 1, 3, 2)$ 。它的 ACG 如图3.1所示。

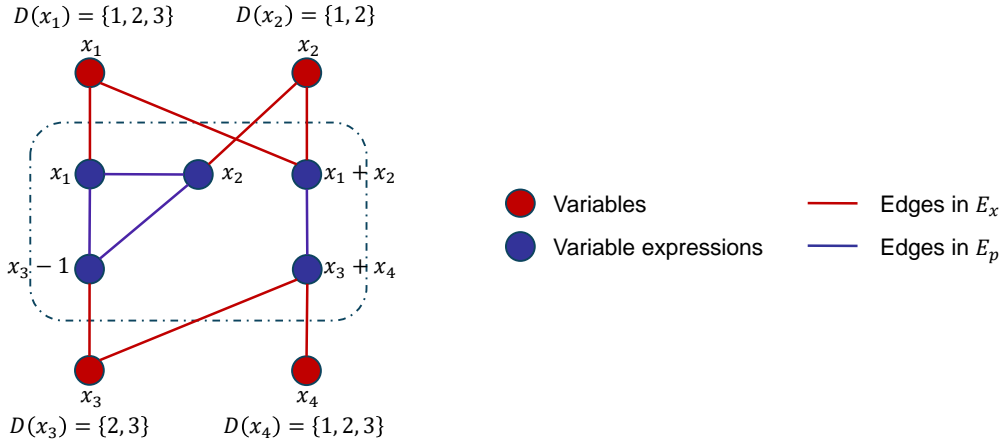


图 3.1 例子中约束转化得到约束图。

Figure 3.1 The constraints in the example are transformed into a constraint graph.

对于得到的 ACG，本文需要根据 ACG 中二元分解得到的二元约束，对变量的值域进行化简。由于二元约束涉及的变量表达式可能包含多个变量或其他常数，直接套用二元约束的弧一致性算法的效率可能不太理想。因此，本文根据变量的值域设计了两个化简规则。其核心思想是，对于一个域大小为 1 的变量 x （例如， $\{i\}$ ），可以直接将值 i 分配给 x ，并实施基于二元约束的全局弧一致性检测。同时，在图中，将 x 的变量表达式邻居中相应的变量更改为值 i ，并删除变量顶点 x 及其相关边。需要注意的是，虽然这些规则以减少 ACG 中的顶点为主，并不保证整组或单个 AllDifferent 约束的全局弧一致性。

本文先给出一些基础定义。在 ACG 中，用 $N(x)$ 表示变量 x 的邻居，用 $XN(p)$ 和 $PN(p)$ 分别表示变量表达式 p 的变量邻居和表达式邻居。在 ACG 中，用 $N(x)$ 表示变量 x 的邻居，用 $XN(p)$ 和 $PN(p)$ 分别表示变量表达式 p 的变量邻居和表达式邻居。在一个 CSP 中，设 $Y = x_{i_1}, x_{i_2}, \dots, x_{i_m}$ 是一串有限的变量，其中 $m > 0$ 。称一个元组 $\mathcal{A}(Y) = (d_{i_1}, d_{i_2}, \dots, d_{i_m}) \in D(x_{i_1}) \times D(x_{i_2}) \times \dots \times D(x_{i_m})$ 为 Y 的赋值，并且使用 $\mathcal{A}(x)$ 来表示在赋值 \mathcal{A} 下赋给变量 x 的值。通常，如果一个关于 X 的赋值给出了 CSP 中的每个变量的赋值，那么它被称为完整赋值；否则，它就是部分赋值。定义赋值操作如下：

定义 3.2 (赋值操作). 给定一个变量 x 和一个值 i ，如果将值 i 赋给 x ，那么就意味着在每个 $p \in N(x)$ 中将对应的变量改为值 i ，并在 ACG 中移除顶点 x 及其相关的边。

显然，对于值域大小为 1 的变量 x （例如， $\{i\}$ ），可以直接将值 i 赋给 x 。而对于线性单变量表达式 p （也即， $XN(p) = \{x\}$ ）， p 的值域与 x 的值域一一对应。在 AllDifferent 约束中，这种变量表达式很常见。基于此，本文给出如下两条化简规则，这两条规则可以移除变量的值域中不满足一致性的赋值。

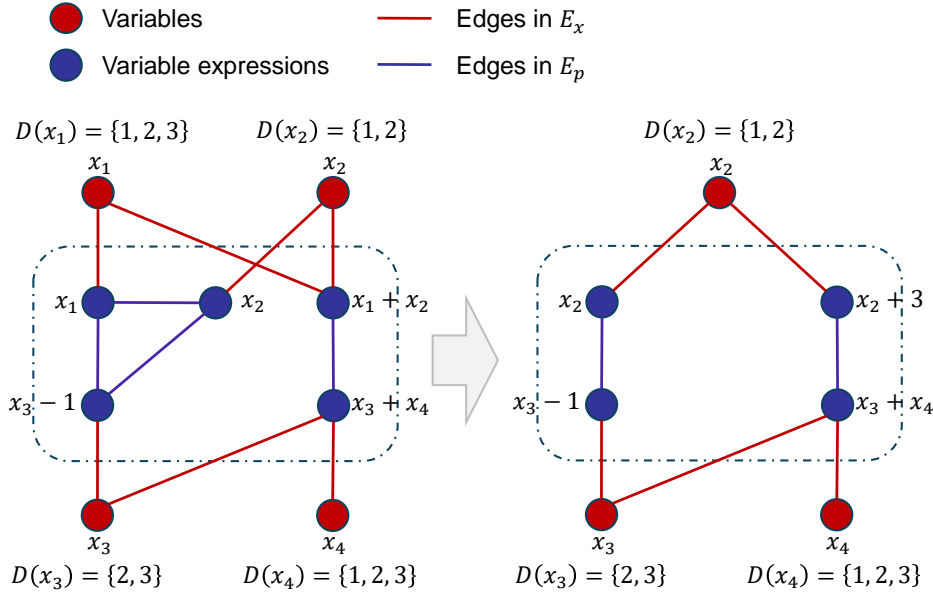


图 3.2 例子中约束图的化简。

Figure 3.2 Simplification of the constraint graph in the example.

- **规则 1.** 给定一个约束 c , 设 $N(c)$ 为 c 中包含的表达式, $D(c)$ 为 $\forall p \in N(c)$ 的可能赋值集的并集。对于一个单变量表达式 $p \in N(c)$ 和一个值 $i \in D(c)$, 如果 $|N(c)| = |D(c)|$, 并且 p 是 c 中唯一可以取值 i 的表达式, 那么将值 i 赋给 x 。
- **规则 2.** 给定一个表达式 p , 若 $|XN(p)|$ 变为 0, 意味着 p 变成一个常数 (例如, i)。对于它的一个单变量表达式邻居 $q \in PN(x)$ (例如, $XN(q) = \{x\}$), 如果 i 是 q 的可能赋值, 那么从 x 的域中移除使得 $q = i$ 的值。如果 p 的所有邻居都是单变量表达式, 那么移除顶点 p 及其相关的边。

本文对 ACG 反复应用两个规则, 从而实施赋值操作以减少 ACG 中的顶点和边。直到 ACG 中不能再移除任何顶点时, 得到最终的 ACG。同样的, 本文以一个简单的例子介绍化简的过程。

例 3.2. 对于例3.1中构造的 ACG, 考虑约束 c_1 , 其中 $N(c_1) = \{x_1, x_2, x_3 - 1\}$ 和 $D(c_1) = \{1, 2, 3\}$ 。可以通过应用规则 1 将值 3 赋给 x_1 , 赋值操作会将变量表达式 $x_1 + x_2$ 修正为 $x_2 + 3$, 变成单变量表达式, 同时让表达式 x_1 变为一个常数 3, 此外变量顶点 x_1 和以它为端点的边也会被删除掉。然后, 由于表示常数 3 的结点的所有邻居都是单变量结点, 可以应用规则 2 删除这个常数顶点和相关的边。ACG 的化简过程如图 3.2所示。

对于剩余的变量, 他们的值域在后续算法中将不再发生改变。后面要用到的局部搜索算法需要从一个初始的完整赋值开始, 本文选择了随机构造完整赋值的方法。具体来说, 对于剩余的变量, 从每个变量的域中随机选择一个值作为其赋值。Initialization 函数在算法 1中给出, 第 9 行是关键步骤。

算法 1 *Initialization function*输入: A CSP $\mathcal{P} = (X, D, C)$ 输出: An ACG \mathcal{G} and an complete assignment \mathcal{A}

```

1: Build an ACG  $\mathcal{G} = (V, E)$  for  $\mathcal{P}$ , where  $V = X \cup NV$  and  $E = E_p \cup E_x$ ;
2:  $SimpSet \leftarrow \{x \mid x \in X, |D(x)| = 1\}$ ;
3: repeat
4:   Select a variable  $x$  from  $SimpSet$  and let  $i$  be the unique value in  $D(x)$ ;
5:   for all  $v$  in  $N(x)$  do
6:     Assign  $x$  contained in  $v$  to  $i$ ;
7:   end for
8:   Remove  $x$  from  $X$ ;
9:   Simplify  $D$  and  $\mathcal{G}$  according to Rule 1 and Rule 2;
10:  // Both  $V$  and  $E$  in  $\mathcal{G}$  may potentially be simplified.
11:   $SimpSet \leftarrow \{x \mid x \in X, |D(x)| = 1\}$ ;
12: until  $SimpSet$  is empty
13: // Initialize  $\mathcal{A}$  of  $X$  randomly;
14: for all  $x$  in  $X$  do
15:   Select a random value  $i$  from  $D(x)$ ;
16:    $\mathcal{A}(x) \leftarrow i$ ;
17: end for
18: return  $\mathcal{G}, \mathcal{A}$ ;

```

除此之外, 还可以使用针对 AllDifferent 约束的 GAC 算法对约束进行过滤, 在过滤后再将约束转化为 ACG。这种做法对值域的限制较强, 保证了每个 AllDifferent 约束的超弧一致性, 但也存在计算复杂度较高等问题。

3.3 分步选择操作策略

在局部搜索中, 一个移动 (又称操作) 意味着对赋值进行轻微的修改, 它定义了候选解的邻域。一般来说, 移动的复杂性应尽可能地保持低, 同时确保它能够充分探索当前解的邻域。在启发式的局部搜索框架 CBLS (Constraint-Based Local Search, 基于约束的局部搜索) 中, 会先满足部分 AllDifferent 约束, 之后通过交换约束中两个变量的赋值的操作实现对当前状态的修改, 这里提到的交换操作称为 *swap*, 就是一种移动。

相对的, 本文采用的是一个简单但有效的操作符 *mov*, 它通过将一个变量顶点的值修改为其值域中的其他值的方式来修改当前的完全赋值, 上面提到的交换操作相当于对交换的两个变量分别作了一次修改赋值的操作 *mov*。这样做的动机是 *mov* 相较于 *swap* 对约束修改的粒度更细, 使得其邻域范围更广, 能够实现

更快的收敛速度，实际上，实验也表明了 *swap* 移动在求解包含复杂 AllDifferent 约束的 CSP 时存在困难。下面是 *mov* 的形式化定义。

定义 3.3. 给定一个变量顶点 x 和一个值 i ，将 x 的值从 $\mathcal{A}(x)$ 改变为 i 的操作符定义为 $mov(x, i)$ 。

对于每个 $(p, q) \in E_p$ ，如果 $\mathcal{A}(p) = \mathcal{A}(q)$ ，称 (p, q) 为冲突边。对于一个 ACG \mathcal{G} ，一个完整赋值 \mathcal{A} 的代价，表示为 $cost(\mathcal{G}, \mathcal{A})$ ，是在 \mathcal{A} 下的冲突边的总数。本文采用 ACG 中冲突边的改进数量，作为 *mov* 的度量，因此一个 *mov* 的打分函数可以表示为

$$score(mov) = cost(\mathcal{G}, \mathcal{A}) - cost(\mathcal{G}, \mathcal{A}'), \quad (3.3)$$

其中 \mathcal{A}' 是通过在 \mathcal{A} 上应用 *mov* 获得的。本文使用 $cost(x, i)$ 来表示将值 i 赋给 x 的代价，即当 $\mathcal{A}(x) = i$ 时，以 $\forall y \in N(x)$ 为终点的冲突边的总数。由于 $mov(x, i)$ 只影响那些以 x 的邻居为终点的边，所以 $mov(x, i)$ 的冲突分数可以由 x 的改变代价来表示，定义如下：

定义 3.4. 给定一个 ACG \mathcal{G} ，操作符 $mov(x, i)$ 的冲突分数定义为

$$score(mov) = cost(x, \mathcal{A}(x)) - cost(x, i). \quad (3.4)$$

通常人们倾向于通过反复选择最合适的 *mov*，不断降低 \mathcal{G} 在 \mathcal{A} 下的冲突边数，也即 $cost(\mathcal{G}, \mathcal{A})$ ，直到找到一个解。一个最常见的贪心选择准则是选择使得 $score(mov)$ 最大的 *mov*，这个方法叫做爬山法，它是很多启发式算法的基础。

在选择操作符时，需要确保它不会陷入循环，并且能在考虑约束条件的满足情况下逃离局部最优。给定一个变量 x ，本文用 $mc(x) = \min(\{cost(x, i) \mid i \in D(x) \setminus \mathcal{A}(x)\})$ 表示 x 在除当前赋值外的所有可能赋值下的最小 $cost$ 。基于此，本文定义一个变量集 $X' = \{x \mid cost(x, \mathcal{A}(x)) \geq mc(x), x \in X\}$ ，称为候选变量集，它拥有一个重要的性质：对于 X' 中的每个变量，都有至少一个 *mov* 保证冲突边的数量不会增加。考虑到 *mov* 包含两部分：变量和值，本文提出以下两步选择策略：

- **步骤 1.** 从 X' 中选择在 $\mathcal{A}(x)$ 下具有最高 $cost$ 的变量 x ；
- **步骤 2.** 从 $D(x)$ 中选择使所选 x 的 $cost$ 最小的值 i 。

在依次选择变量 x 和值 i 后，已经确定了要执行的 $mov(x, i)$ 。虽然两步选择可能会错过一些更贪婪的 *mov*（在第一步未被选择的变量可能有更好的赋值），但它可以大大降低选择 *mov* 的时间复杂度。设 $|D|$ 表示每个变量的平均值域的大小，那么直接的 *mov* 选择方法需要 $|X| \cdot |D|$ 的复杂度，而两步 *mov* 选择只需要 $|X| + |D|$ 的复杂度，从而降低了最坏情况下的时间成本。此外，候选变量集 X' 确保了所选变量的下限质量。

然而，局部搜索算法经常会陷入局部最优，这意味着 X' 是空的。在这种情况下，本文定义了一个新的候选变量集 X'' ，它表示当前存在冲突的变量集，称

为冲突变量集。这个变量集的动机来自于改变没有冲突的变量的赋值不会减少冲突的总数。 X'' 不能保证变量的质量，所以在这种情况下，本文直接选择 *score* 最大的 *mov*，*mov* 中的 x 只能从 X'' 中选择。为了和两步选择策略有所区分，本文称这种选择方式为直接选择策略。

3.4 禁忌策略

局部搜索算法使用禁忌策略^{[?]1}来避免陷入循环。鉴于需要通过两步来选择操作，本文提出了一个禁忌策略组合，包括以下策略。

- **策略 1.** 在一个变量表达式改变其值后，它将被禁忌，直到出现一个以它为端点的新的冲突边。此外，如果一个变量的所有邻居都被禁忌，那么该变量也将被禁忌；

- **策略 2.** 在一个变量 x 被赋予值 i 后，接下来的 tt 次迭代中，操作 $mov(x, i)$ 将被禁忌，其中 tt 是一个被称为禁忌期限的参数；

- **策略 3.** 如果选择的 *mov* 被禁忌的同时其变量也被禁忌，那么算法在接下来的 β 步中切换到直接选择策略。

在算法的求解流程中，策略 1 在变量选择阶段实施。策略 3 确保算法在两步选择策略的性能下降时，能够快速切换到直接选择策略。在直接选择中，只使用策略 2。接下来，本文将具体介绍这三个策略的动机和细节。

策略 1 可以看作是一种格局检测策略^{[?]1}的变体，它最初是为了解决最小顶点覆盖问题中局部搜索算法的循环问题而被设计出来的，并在此后应用在许多图论问题上。格局检测策略的核心目标是在向当前候选解添加点时考虑该点的邻域。在本文中，本文用 S_x 表示关于变量的禁忌表，它初始为空。在选择一个操作 $mov(x, i)$ 执行时，会同步地将变量 x 放入 S_x 中，在禁忌表中的变量在第一阶段选择变量时不会被选中，除非 $S_x = X'$ ，也即所有变量都在禁忌表中。基于此，本文定义一个变量的格局如下：

定义 3.5 (格局). 对于 ACG 中的一个变量顶点 x ，其格局定义为其所有邻居的状态向量 F_v 。其中，一个变量表达式顶点的状态表示为 $f_p \in \{0, 1\}$ ，具体地：

$$f_p = \begin{cases} 1, & \text{若出现新的以 } p \text{ 为端点的冲突边;} \\ 0, & \text{其他情况.} \end{cases}$$

如果把变量表达式的格局定义为其与其变量表达式邻居构成的边，那么变量的格局可以视为变量表达式格局构成的格局，形成嵌套关系。一个变量的格局发生改变，当且仅当它所关联的变量表达式的格局发生改变。在 S_x 中的变量的格局改变时，会从 S_x 中移除该变量。这种策略是合理且直观的，因为它避免了重复选取同一个变量导致的循环情况。

然而，在实际求解过程中发现，仅采用策略一时仍然会发生循环。考虑如下场景：两个相关联的变量对应的 *mov* 会互相改变对方的格局，在不限制 *mov* 的

情况下，一个变量在被禁忌后会被另一个变量对应的 *mov* 激活，导致两个 *mov* 被无限循环地选择。基于此，本文构造了关于变量值域的禁忌表 S_m ，采用传统的 *tabu* 策略维护 S_m 。该策略基于以下思路：通过禁止对最近的变化进行反转，避免局部搜索马上返回刚刚访问过的状态进而导致循环。在算法求解时，本文在选择一个 *mov* 时，会优先选择没有被禁忌的变量，并优先从变量的值域中选择没有被禁忌的赋值。

在应用策略 1 和策略 2 时，需要注意到，在执行一个操作后，该操作在被禁忌的同时可能无法从禁忌表中激活任何变量。因此，禁忌搜索可能会出现如下情况：在两步选择策略中，会出现所有变量都被禁忌的情况。这种情况往往发生在局部搜索趋于稳定的后期。

考虑一种极端情况，在两步选择策略中，若所有 X' 中的变量都被禁忌，在步骤一中将不得不从禁忌表里选择一个变量 x ，而恰好变量 x 的所有赋值也都被禁忌，导致在步骤二中还需要从禁忌表中选择一个被禁忌的赋值 i 。在这种情况下，会得到一个被双重禁忌的操作 $mov(x, i)$ 。一个更合理的策略是，直接切换到直接选择策略，通过更贪心的选择策略寻找当前状态的邻域下最合适的操作。因此，本文设计了策略 3，它实际上是一种快速转换机制，作为前面提到的根据冲突变量集 X'' 切换的补充。

3.5 打破平局策略

另一方面，由于 *AllDifferent* 约束的性质，在直接选择 *mov* 时，经常有许多最大 *score* 的 *mov* 不唯一的情况。常用的打破平局策略有随机选择策略、子打分函数策略等。其中，随机选择策略会在所有具有相同优化目标值的解决方案中随机选择一个；而子打分函数策略除了主要的打分函数外，还会设计子打分函数。当且仅当通过主打分函数得到的最优操作有多个时，我们会使用子打分函数进行评分和排序，以此来打破平局。

本文提出了一个新的 *nscore* 作为 *score* 的补充，它利用了变量的邻域信息。如果两个变量的邻居之间存在冲突边，就说这两个变量处于冲突中。本文使用 $n(x, i)$ 来表示当值 i 被赋予 x 时，与 x 冲突的变量的数量。

定义 3.6. 给定一个 ACG \mathcal{G} ，一个 $mov(x, i)$ 的 *nscore* 定义为

$$nscore(mov) = n(x, \mathcal{A}(x)) - n(x, i). \quad (3.5)$$

与 *score* 相比，*nscore* 关注的是变量之间的冲突，由于在 ACG 中的侧重点不同，一个 *mov* 在两个打分机制下不会出现高度相似的现象，从而保证打破平局的有效性。下面通过一个例子介绍该打分函数如何起作用。

例 3.3. 考虑图 3.2 右图所示的 ACG 示例，并假设它当前在赋值 $\mathcal{A} = (1, 2, 2)$ 下。那么 $score(mov(x_2, 2)) = score(mov(x_3, 3)) = 2$ ，而 $nscore(mov(x_2, 2)) = 2$ 和 $nscore(mov(x_3, 3)) = 1$ 。因此，选择执行 $mov(x_2, 2)$ 。

算法 2 *Select Move* function

输入: A complete assignment \mathcal{A} , a control variable $mode$ and candidate variable sets X', X''

输出: A variable x and a value i

```

1: if  $mode = 0$  then
2:   Select a variable  $x$  from  $X'$  based on  $cost(x, \mathcal{A}(x))$  and Strategy 1;
3:   Select a value  $i$  from  $D(x)$  based on  $cost(x, i)$  and Strategy 2;
4:   if  $x$  is Tabu and  $mov(x, i)$  is Tabu then
5:      $mode \leftarrow \beta$ ; // Strategy 3
6:     Initialize  $X''$  according to  $\mathcal{A}$ ;
7:   end if
8: else
9:   Select  $x$  from  $X''$  and  $i$  based on  $score(mov)$  and  $nscore(mov)$ ;
10:   $mode \leftarrow mode - 1$ ;
11: end if
12: return  $x, i$ 

```

最终的评价准则如下: 首先通过打分函数 $score$ 寻找评分最高的 mov , 在 mov 不唯一时, 使用更细粒度的子打分函数 $nscore$ 对 mov 进行评价, 优先选择 $nscore$ 最大的 mov 打破平局, 若 mov 仍然不唯一, 通过随机选择的方式打破进一步的平局。在本节的最后, 本文给出最终的 *Select Move* 函数, 如算法 2 所示。它最初采用两步选择策略, 当 X' 为空时, 接下来的 β 步将使用直接选择策略, 其中 β 是用于控制步长的常数。

3.6 局部搜索算法框架

在构造了状态、打分函数、邻域以及评价准则后, 可以得到构造一个局部搜索算法所需的基本框架。基于前几节提到的定义和策略, 本文设计了一个用于解决包含 AllDifferent 约束问题的 CSP 的局部搜索算法, 称为 AllDiff-LS。算法的主体部分如图 3.3 所示, 算法采用了迭代局部搜索的框架, 主要由三个部分组成, 即两步选择策略、直接选择策略和重启策略, 本章中已经介绍了前两个部分, 重启策略部分将在下一章进行介绍。

本文在算法 3 中给出了方法的伪代码描述。AllDiff-LS 算法的输入是一个仅包含 AllDifferent 约束的 CSP $\mathcal{P} = (X, D, C)$ 以及一个时间限制 T , 输出则是一个关于输入的 CSP 的完整赋值 \mathcal{A}_b 。首先, 本文使用前述的初始化算法 1 随机生成了一个初始的完整赋值 \mathcal{A} 。接下来, 算法以迭代的方式运行, 包含一个外循环和一个内循环。注意, 内循环由最大迭代次数 $maxIter$ (第 24 行) 控制, 而外循环由截止时间 T (第 29 行) 控制。

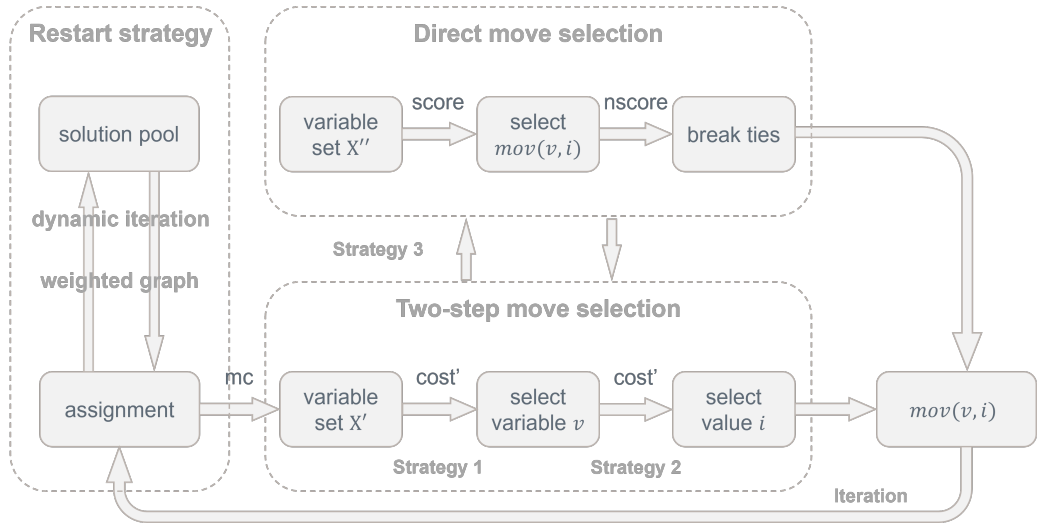


图 3.3 Local Search 的求解流程。

Figure 3.3 The solving process of Local Search.

在内循环（第 13-23 行）中，算法搜索具有最小代价的完整赋值。算法首先通过 *SelectMove* 函数获取一个操作 *mov*，然后实施 *mov*，同时更新两个禁忌列表 S_x 和 S_m ，以及有冲突的顶点。每个内循环结束后，下一轮的初始完整赋值和最大迭代次数将通过 *SelectSolution*（第 27-28 行，下一章介绍）获得。如果在给定的时间限制 T 内找到了一个没有冲突的完整赋值，那么返回该赋值（第 25 行）；否则，返回一个空集（第 30 行）。

算法 3 AllDiff-LS algorithm

输入: A CSP $\mathcal{P} = (X, D, C)$ and time limit T

输出: A complete assignment \mathcal{A}_b

```

1: Initialize  $\mathcal{G}, \mathcal{A}$ ;
2:  $\mathcal{AC} \leftarrow []$ ;
3:  $maxIter \leftarrow \alpha$ ;
4: repeat
5:   Initialize  $mc(x)$  for  $\forall x \in X$ ;
6:   Initialize  $X'$  according to  $mc(x)$  and  $\mathcal{A}$ ;
7:    $mode \leftarrow 0$ ;
8:    $\mathcal{A}_b \leftarrow \mathcal{A}$ ;
9:   if  $X' = []$  then
10:    Initialize  $X''$  according to  $\mathcal{A}$ ,  $mode \leftarrow \beta$ ;
11:   end if
12:   repeat
13:      $x, i \leftarrow SelectMove(\mathcal{A}, mode, X', X'')$ ;
14:      $\mathcal{A}(x) \leftarrow i$ ;
15:     Update  $mc(x)$ , update  $X'$  or  $X''$ ;
16:     if  $mode > 0$  and  $X' = \emptyset$  then
17:       Initialize  $X''$  according to  $\mathcal{A}$ ;
18:        $mode \leftarrow \beta$ ;
19:     end if
20:     if  $cost(\mathcal{G}, \mathcal{A}) \leq cost(\mathcal{G}, \mathcal{A}_b)$  then
21:        $\mathcal{A}_b \leftarrow \mathcal{A}$ ;
22:     end if
23:      $iter \leftarrow iter + 1$ ;
24:   until  $iter \geq maxIter$ 
25:   if  $cost(\mathcal{G}, \mathcal{A}_b) = 0$  then return  $\mathcal{A}_b$ 
26:   end if
27:    $\mathcal{A} \leftarrow SelectSolution(\mathcal{G}, \mathcal{A}_b, \mathcal{AC})$ ;
28:    $maxIter \leftarrow \mathcal{A}.step$ ;
29: until  $T$  is reached
30: return  $\emptyset$ 

```

3.7 本章小结

本章节将 CSP 转化为了图结构，并设计了一个用于求解此类问题的局部搜索算法，具体如下。在第一节中，本文给出了要研究的问题，以 AllDifferent 约束为主的 CSP 的形式化定义。在第二节中，本文将 CSP 中包含的 AllDifferent 约束转化为约束图，并设计了两个化简规则对约束进行化简。在第三节中，本文介绍了分两步选择操作的策略，作为直接选择操作策略的补充。在第四节中，本文设计了局部搜索的禁忌搜索框架。在第五节中，本文设计了包含打破平局策略的操作评价准则。在第六节中，本文将前面提到的策略整合在了一起，实现了局部搜索算法：AllDiff-LS，并介绍了算法的流程细节。

第4章 针对 AllDifferent 约束的加权和重启策略

本章节主要介绍用于局部搜索算法 AllDiff-LS 的重启策略，它主要介绍了基于解池构造的两个策略：图加权和动态迭代。首先，本文介绍了拟采用的解池技术，用来存储每轮局部搜索过程中得到的最优解。之后，本文介绍了如何对 AllDifferent 约束图加权，以及权值如何起作用。此外，本文将解池中的候选解绑定了新的信息，用来指导后续局部搜索的迭代次数。最后，本文还介绍了结合重启策略后的局部搜索算法的实现。

4.1 解池技术

在上一章，本文介绍了 AllDiff-LS 的算法框架，它基于迭代局部搜索的框架，其中迭代局部搜索是一个将局部搜索作为组件的元启发式方法。在迭代局部搜索流程中，会迭代执行局部搜索改进当前最优解，并在迭代一定次数还没有找到改进时重启，以生成新的解来扩展搜索空间。生成解的过程一般是通过施加扰动实现的，并且该算法会重复执行，直到达到算法设置的终止条件。引入扰动，有助于算法跳出当前的局部最优解。

在 AllDiff-LS 算法流程中，重启时会从局部最优解中抽取新的解，用于后续搜索，这是通过从一组选定的解生成重启点来实现的。这组解，称为解池 (solution pool)^[2]，它维护了搜索过程中得到的当前的最优解，也就是说，ACG 在解池 \mathcal{A} 中的每个赋值下的冲突边的数量是相同的，并且当前已经是最小的。在搜索的初始阶段，解池被初始化为空并且拥有一个最大池尺寸限制。搜索过程中解池的维护机制如下：

- 当搜索在数次迭代后**没有找到改进**时，如果此时的局部最优解还没有在解池中，就将其添加到解池中，此时解池中的每个解都拥有相同的代价。
- 当搜索在数次迭代后**找到改进**时，为了保持最好的解，算法会直接更新解池，清空当前解池的同时将最新的局部最优解加入解池中。

本文会额外记录解池中每个候选解被访问的次数，如果此解池的大小超过了池尺寸限制，就会从解池中删除被访问最多的候选解，这样做的动机源于需要探索那些较少被访问的候选解，以探索更广的搜索空间。接下来，本文会介绍局部最优解的进入准则和候选解的生成准则，以通过解池存储和生成用于后续搜索的优良解。

解的准入策略。一个解 \mathcal{A}_1 在被加入到解池之前，需要先判断这个解在解池中是否已经存在。也就是说，需要保证对于解池中的每个候选解 \mathcal{A}_2 ，都有 $\mathcal{A}_1 \neq \mathcal{A}_2$ ，这个过程的复杂度是较高的。为了简化和候选解比较的复杂度，在保存局部最优解 \mathcal{A} 的同时，保存该解的冲突边集合。本文采用了文献^[2]的思路，引入相似解的概念如下。

定义 4.1. 两个解相似，当且仅当这两个解具有相同的冲突边。

从而，本文在判断一个解是否需要加入到解池中时，仅需要保证它和池中的任何解不相似即可。仅在出现相似解的情况下，本文才会比较两者是否相同，在不同时将池中的候选解更新为新的解，否则则抛弃这个解。在搜索的后期，冲突边的数量会非常少，由于判断相似仅需要比较解的冲突边，所以这极大地减小了比较的时间开销。

解的生成策略。一个解 \mathcal{A}_1 在被选中作为下一轮搜索的初始解之后，需要对它进行扰动以逃离局部最优。过去的研究表明，用于扰动的方法对性能有显著影响。扰动的方法有很多，例如，交换几对随机选定的相邻变量的赋值；删除几组随机选定的变量的赋值，使用随机或启发式的方法填充它们的赋值；以及随机游走策略等。而修改的范围也需要进行考量，Tasgetiren 等人^[2]发现，通过将一个或两个随机选择的移动插入到随机选定的位置，性能相当好。本文通过随机修改变量赋值的方法来扰动解。

关于扰动的程度，本文将选定的扰动次数与当前解的冲突边个数，以及当前解被选中的频率挂钩。定义扰动次数为对选定解使用指定扰动技术扰动的次数。具体地，本文令基础的扰动次数为冲突边个数乘以一个系数，并在这个解被多次选中时，相应地增加系数的权重以得到最终的扰动次数。本文发现搜索的前期可能存在搜索不充分的情况，导致冲突边数量较多，进而导致扰动程度过大的问题。因此，在实际实验中，本文仅在冲突边的数量低于一定阈值时才对初始解施加扰动。

4.2 约束加权图和动态迭代策略

在局部搜索的过程中，加权策略是常用的一种调整搜索方向和搜索步长的方法。常见的加权策略有动态权重策略、自适应权重策略等。比如模拟退火算法中的温度参数，就是一种动态权重。在搜索初期，为了增加搜索的随机性和全局搜索能力，权重设置得较大；随着搜索的进行，为了逐渐收敛到全局最优解，权重逐渐减小。

为了避免找到重复的赋值，本文在重启部分提出了一种基于解池技术的加权策略，对 ACG 中的边集进行加权。具体地，对于 ACG 中的每个边 $e \in E_p$ ，本文使用 $e.w$ 作为 e 的权重，初始设置为 1，由此得到的 ACG，称为加权 *AllDifferent* 约束图，简称 WACG (Weighted ACG)。关于如何加权，本文采用了 PAWS 策略的带概率版本，具体地，边集的权值按照如下方式更新：当一个新的赋值被添加到 \mathcal{AC} 时，每个在赋值下的冲突边的权重以概率 θ 增加一。

由于解池中仅保存当前最优解，因此本文将 WACG 的权值与解池进行绑定。当解池 \mathcal{AC} 被重置时（即，找到了一个更小的 *cost* 赋值），加权图中所有的边 $e \in E$ 的 $e.w$ 也被重置为 1。这样做的动机是找到新解意味着跳出了一个局部最优解，因此，之前关于边集的权值更新对于新的搜索空间可能起不到好的效果，

算法 4 *SelectSolution* function

输入: An ACG $\mathcal{G}(V, E)$, a complete assignment C and a solution pool \mathcal{AC}

输出: A complete assignment \mathcal{A}'

```

1:  $\mathcal{A}.step \leftarrow \alpha$ ;
2: if  $\mathcal{AC} = \emptyset$  or  $cost(\mathcal{G}, \mathcal{A}) < \mathcal{AC}.cost$  then
3:   Reset  $e.w$  for  $\forall e \in E_p$ ;
4:    $\mathcal{AC} \leftarrow \{\mathcal{A}\}$ ;
5:    $\mathcal{AC}.cost \leftarrow cost(\mathcal{G}, \mathcal{A})$ ;
6:   Update  $e.w$  ( $e \in E_p$ ) based on weight strategy;
7: else if  $cost(\mathcal{G}, \mathcal{A}) = \mathcal{AC}.cost$  then
8:   if  $\mathcal{A} \neq \mathcal{A}'$  for  $\forall \mathcal{A}' \in \mathcal{AC}$  then
9:      $\mathcal{AC} \leftarrow \mathcal{AC} \cup \{\mathcal{A}\}$ ;
10:    Update  $e.w$  ( $e \in E_p$ ) based on weight strategy;
11:   else
12:      $\mathcal{A}'.step \leftarrow \mathcal{A}'.step + \gamma * \alpha$ ;
13:   end if
14: end if
15: Update  $\mathcal{AC}$  based on its size and weight strategy;
16: Select a  $\mathcal{A}'$  from  $\mathcal{AC}$  randomly;
17: Disturb  $\mathcal{A}'$  based on disturbance strategy;
18: return  $\mathcal{A}'$ 

```

甚至影响搜索的改进方向。

对边进行加权后, 得到加权代价, 即当 $\mathcal{A}(x) = i$ 时, 以 $\forall y \in N(x)$ 为终点的冲突边的总权重, 其计算如下。在得到加权代价后, 本文在两步选择策略的变量和值选择过程中使用 $cost'$ 代替 $cost$, 而算法的其他部分仍然使用 $cost$ 。

定义 4.2. 给定一个 ACG \mathcal{G} , 变量 x 赋值 i 的加权代价定义为

$$cost'(x, i) = \sum_{p \in N(x)} \sum_{q \in CN(p)} (p, q).w, \quad (4.1)$$

其中 $CN(p) = \{q \mid q \in PN(p)\}$ 。

本文最后介绍的一个策略是动态迭代策略。考虑前一节中提到的扰动策略中, 本文提到了前期搜索不充分的问题, 除了调整扰动的程度外, 这种情况可以通过将静态的迭代次数设为动态次数来改善。

此外, 考虑一个候选解被选中的次数较多的情况, 除了表明此候选解在解池中存在时间较长 (通过更新相似解解决), 还说明以该候选解为初始解的搜索过程始终无法得到一个有效的解。其原因是搜索空间探索不充分, 除了通过施加更

加强烈的扰动外，还可以通过增加迭代次数的方式，对该解处于的搜索空间进行充分探索。基于以上两点，本文设计了用于充分探索的动态迭代策略如下，具体实现在介绍伪代码时介绍。

- 在选中候选解后，下一轮迭代的次数随着候选解的冲突边数量的增多而适当增加。
- 在一轮搜索没有提升（即当前最优解即初始解）时，下一轮以该解为初始解的内循环的迭代次数会适当增加。

本文在算法 4 中显示了选择初始完整赋值的伪代码，AllDiff-LS 算法会通过 *SelectSolution* 获得一个赋值 \mathcal{A} 和其相应的最大迭代次数 $\mathcal{A}.step$ 。在有新的拥有更少冲突边的赋值 \mathcal{A} 出现时，重置 WACG 的边权，并将 \mathcal{AC} 重置为 $\{\mathcal{A}\}$ 。本文用 $\mathcal{AC}.cost$ 表示解池中的赋值的代价。如果上一轮得到的赋值与 \mathcal{AC} 中的赋值代价相同且是新的赋值，本文将其放入 \mathcal{AC} 。当 \mathcal{A} 被插入到 \mathcal{AC} 时，要做两件事：首先， $\mathcal{A}.step$ 被初始化为 α ，其中 α 是一个手动设置的超参数，并在大循环迭代的过程中逐步减少；其次，更新 WACG，令 \mathcal{A} 的冲突边的权重以概率 θ 增加一。在更新 \mathcal{AC} 后，本文从中随机选择一个赋值 \mathcal{A} 进行扰动后作为下一轮内循环的初始解，并将 $\mathcal{A}.step$ 作为其最大迭代次数。如果一个以 \mathcal{A} 为初始赋值的内循环没有找到更好的解， $\mathcal{A}.step$ 就会增加 $\gamma * \alpha$ 并在其过大时将该赋值从 \mathcal{AC} 中删掉，其中 γ 和 α 是常数。

4.3 AllDiff-LS 工具设计


通过前面的叙述，本文能够得到一个求解仅由 AllDifferent 约束构成的 CSP 的算法。这一小节，本文先介绍如何将其他的二元约束引入 AllDiff-LS 求解算法的框架，再简单介绍基于该算法实现的求解工具。

ACG 是通过对 AllDifferent 约束进行二元分解构造的。因此，在这个框架下，可以引入其他类型的二元约束，只需构造类似的边类型即可。具体来说，可以考虑等于和不等偏序这两种二元约束。等式约束要求两个变量的值必须相等，而不等偏序约束则要求一个变量的值必须大于或小于另一个变量的值。

对于等式关系，将等式关系涉及的两个变量表达式顶点合并，并合并两个变量表达式涉及的边，得到新的 ACG。此外，计算打分函数时，需要额外计算合并的变量表达式顶点内部的冲突。通过调整顶点内部冲突和冲突边之间的权重，可以将等式约束引入到 AllDiff-LS 的框架之中。实际上，ACG 的构造过程中已经存在边集的合并。例如，对于约束 $AllDifferent(x_1, x_2, x_3)$ 和 $AllDifferent(x_1, x_2, x_4)$ ，通过二元分解得到的二元约束中， x_1 和 x_2 之间的不相等关系出现了两次，但由于该关系重复，所以它们在 ACG 中表现为一条边。

对于不等偏序关系的引入则比较自然，仅需新引入两个边类型，分别表示大于号和小于号，并修改针对这两类符号的冲突边定义。例如，对于约束 $x_1 < x_2$ ，本文引入边 (x_1, x_2) ，并在 $x_1 \geq x_2$ 时将该边定义为冲突边。

1					7		9	
	3			2				8
		9	6			5		
		5	3			9		
	1			8				2
6					4			
3							1	
	4							7
		7				3		



1	6	2	8	5	7	4	9	3
5	3	4	1	2	9	6	7	8
7	8	9	6	4	3	5	2	1
4	7	5	3	1	2	9	8	6
9	1	3	5	8	6	7	4	2
6	2	8	7	9	4	1	3	5
3	5	6	4	7	8	2	1	9
2	4	1	9	3	5	8	6	7
8	9	7	2	6	1	3	5	4

图 4.1 名为 “AI Escargot” 的数独实例。

Figure 4.1 A Sudoku instance called "AI Escargot".

接下来介绍工具的实现环境和实验参数。求解工具由 C++ 实现，并用 g++ 编译，选项为 ‘-O3’。运行环境在一台配备有 Intel Xeon Platinum 8153 和 2048G 内存的服务器上进行，操作系统为 Centos 7.7.1908。AllDiff-LS 是用 C++ 实现的，AllDiff-LS 有五个参数： α 用于初始最大迭代次数， β 用于模式切换阈值， tt 用于禁忌方案， γ 用于增加赋值步骤，以及 θ 用于权重策略。参数根据^[2]和初步实验进行调整，设置如下： $\alpha = 100,000$ ， $\beta = 100$ ， $tt = rand(10) + 0.6 \times cost(\mathcal{G}, \mathcal{A})$ ， $\gamma = 5$ 和 $\theta = 0.25$ 。参数 β 被固定为 100，因为经验表明这个值可以产生良好的性能：在算法优化过程的后期，可以观察到，算法会频繁切换到直接移动选择策略，因此将参数 β 配置为一个相对较大的常数是较为合理的。

工具的输入是由两部分组成，一个 CSP，由整数变量、值域和 AllDifferent 约束组成，以及一个求解时间。工具可以对大数独、全间隔、N 皇后问题和互正交拉丁方问题进行求解。对于符合输入规格的输入文件，在给定时间内求解器会输出一个符合约束的可行解，或超时求解返回 unknown。以数独为例，本文给出一个著名的 9 阶数独 “AI Escargot”，如图 4.1 所示，算法可以在小于 0.01 秒内找到该问题的解。这是一个简单的小问题实例，在实验部分本文主要针对大规模实例进行比较。

4.4 本章小结

在这一章中，介绍了算法的补充策略，解池技术及其维护。具体地，在第一节中，本文介绍了解池技术的基本概念，以及分别通过解近似和扰动策略构造的解的准入和生成策略。在第二节中，本文介绍了基于解池技术的加权和动态迭代策略，具体地，算法构造了新的 WACG，并根据求解情况动态地调整后续求解的迭代次数。在第三节中，本文通过构造新的冲突类型，将二元约束引入到求解的框架中，并介绍了实现的求解工具的具体细节。

第5章 实验设计及结果分析

本章节针对基于 AllDiff-LS 算法实现的求解工具设计了实验，并对实验结果进行了分析。章节主要包含三部分：第一部分介绍实验安排，比如采用的数据集、比较的方法等；第二部分介绍算法在多个实验基准上同其他方法的性能比较；第三部分对实验中提到的各种策略的有效性进行消融实验。

5.1 实验设置

实验样例：在本章节，本文选择了四个经典问题作为基准：数独，N 皇后，全间隔和二正交拉丁方 (2-MOLS) 问题^[?]]。它们都可以使用 AllDifferent 约束进行建模，并且这些问题的约束生成的 ACG 具有独特的特征，使这些问题极具代表性。下面，本文给出这些问题的编码的具体细节。

定义 5.1 (数独). 一个数独 (Sudoku) 问题实例 S^n 可以用一个 $n^2 \times n^2$ 的网格表示，其中填充着 1 到 n^2 范围内的数字，这里 n 被称为问题的阶数。网格被划分为 n^2 个大小为 $n \times n$ 的子网格。通常一些单元格已经被固定为特定的值。问题要求每单元格中的数字满足每行、每列和每个子网格中每个数字恰好出现一次。

在数独问题中，所有变量表达式都是单变量表达式，并出现在多个 AllDifferent 约束中。以下为其编码：

$$\begin{aligned} & \text{AllDifferent}(x_{i,1}, x_{i,2}, \dots, x_{i,n^2}) \text{ for } 1 \leq i \leq n^2, \\ & \text{AllDifferent}(x_{1,j}, x_{2,j}, \dots, x_{n^2,j}) \text{ for } 1 \leq j \leq n^2, \\ & \text{AllDifferent}(x_{i*n+u}, x_{j*n+v} \mid \text{ for } 1 \leq u, v \leq n) \text{ for } 0 \leq i, j \leq n-1, \\ & x_{i,j} \in \{1, 2, \dots, n^2\} \text{ for } 1 \leq i, j \leq n^2. \end{aligned}$$

定义 5.2 (N 皇后). 一个 N 皇后 (N-queens) 问题实例 Q^n 是将 n 个皇后放置在一个 $n \times n$ 的棋盘上，使得没有皇后互相攻击（任何两个皇后都不能处于同一行、同一列或同一对角线上），这里 n 被称为问题的阶数。

在 N 皇后问题中，一个变量可以有多个变量表达式，但是每个变量表达式只存在于一个 AllDifferent 约束中。建模这个问题的一种方法是为每一行引入一个整数变量 x_i ，其中 $i = 1, 2, \dots, n$ ，它的取值范围是从列 1 到 n 。这意味着在每一行 i 中，皇后被放置在第 x_i 列上。以下为其编码：

$$\begin{aligned} & \text{AllDifferent}(x_1, x_2, \dots, x_n), \\ & \text{AllDifferent}(x_1 - 1, x_2 - 2, \dots, x_n - n), \\ & \text{AllDifferent}(x_1 + 1, x_2 + 2, \dots, x_n + n), \\ & x_i \in \{1, 2, \dots, n\} \text{ for } 1 \leq i \leq n. \end{aligned}$$

定义 5.3 (全间隔). 一个全间隔 (*All-Interval*) 问题实例, 表示为 \mathcal{A}^n , 涉及将 n 个不同的元素排列成一个序列, 使得任意两个元素之间的绝对差形成一个包含所有可能整数从 1 到 $n-1$ 的集合。 n 的值被称为问题的阶数。

在全间隔问题中, 仅存在两个 AllDifferent 约束, 变量和变量表达式之间是多对多的关系, 并且一个变量可能在一个约束中的两个变量表达式中出现。问题建模的形式化表示如下:

$$\begin{aligned} & \text{AllDifferent}(x_1, x_2, \dots, x_n), \\ & \text{AllDifferent}(|x_1 - x_2|, |x_2 - x_3|, \dots, |x_{n-1} - x_n|), \\ & x_i \in \{1, 2, \dots, n\} \text{ for } 1 \leq i \leq n. \end{aligned}$$

定义 5.4 (互正交拉丁方). 互正交拉丁方 (*MOLS*) 是由数学家欧拉提出的一个极具挑战性的组合问题。拉丁方阵可以看作没有子网格约束的数独问题。给定两个相同阶数 n 的拉丁方阵, 如果它们在每个对应位置的元素组合是唯一的, 那么它们是互相正交的, 表示为 $2\text{-MOLS}(n)$ 。

在互正交拉丁方问题中, 在采用 AllDifferent 约束进行建模的情况下, 一个变量存在于多个变量表达式中, 而且一个单一的变量表达式中包含多个变量, 这些变量也出现在多个 AllDifferent 约束中。因此, MOLS 的 AllDifferent 约束图比前三种情况更为复杂。以下为其一种编码方式:

$$\begin{aligned} & \text{AllDifferent}(x_{i,1}, x_{i,2}, \dots, x_{i,n}) \text{ for } 1 \leq i \leq n, \\ & \text{AllDifferent}(x_{1,j}, x_{2,j}, \dots, x_{n,j}) \text{ for } 1 \leq j \leq n, \\ & \text{AllDifferent}(y_{i,1}, y_{i,2}, \dots, y_{i,n}) \text{ for } 1 \leq i \leq n, \\ & \text{AllDifferent}(y_{1,j}, y_{2,j}, \dots, y_{n,j}) \text{ for } 1 \leq j \leq n, \\ & \text{AllDifferent}(x_{i,j} * n + y_{i,j} \mid \text{ for } 1 \leq i, j \leq n), \\ & x_{i,j}, y_{i,j} \in \{1, 2, \dots, n\} \text{ for } 1 \leq i, j \leq n. \end{aligned}$$

虽然存在一些人工制作的实例^[2, 3], 但这些实例的规模较小, 相对容易解决。因此, 本文主要生成大规模随机实例进行实验。使用^[2, 3]和^[2, 3]中提到的方法, 本文生成了从 4 到 9 阶的数独实例 (即, 从 16×16 到 81×81 的大小)。本文为每个固定单元格比例生成了 100 个实例, 步长为 10%, 从 0% 到 90%, 总共生成了 6000 个单独的实例, 并使用 ' $\mathcal{S}\text{-}n\text{-}r$ ' 来表示阶数为 n 、固定单元格比例为 r 的数独实例族。对于 N 皇后问题, 本文比较了从 500 到 6000 阶的十二个实例 (步长为 500), 记为 ' $\mathcal{Q}\text{-}n$ '。对于全间隔问题, 本文比较了从 12 到 26 阶的八个实例 (步长为 2), 记为 ' $\mathcal{A}\text{-}n$ '。对于 2-MOLS 问题, 由于其难度级别, 本文只比较了阶数为 3、4、5、7、8 和 9 的六个实例 (已经证明 2-MOLS(6), 也称为 36 官问题, 没有解), 记为 ' $\mathcal{L}\text{-}n$ '。

比较方法: 本文将 AllDiff-LS 与一个启发式求解器 Yuck¹, 以及两个完备求解器 ILOG CPLEX Optimizer (版本 20.10, 简称 CPLEX) 和 Choco (版本 4.10.13)^[2] 进行了比较, 所有这些都支持 ‘AllDifferent’ 约束。Yuck 采用了^[2]的想法, 并在 2022 MiniZinc 挑战赛的局部搜索赛道中获得了冠军, 而 CPLEX 和 Choco 分别是知名的商业和开源求解器, 它们都支持各种化简策略。此外, 对于数独, 本文还比较了两个专用的启发式算法, 来自^[2]的蚁群算法 (ACS) 和来自^[2]的迭代局部搜索算法 (ILS), 其中 ACS 被调整为接受大于 5 的阶数。在 ACS 算法中, 作者将约束传播和蚁群算法相结合, 通过 Minimum Remaining Values Heuristic 搜索求解空间。而在 ILS 算法中, 算法会在初始化时满足部分约束, 并通过交换变量赋值的方式进行局部搜索。

实验安排及指标: 后续所有实验都在一台配备有 Intel Xeon Platinum 8153 (2.00GHz) 和 2048G RAM 的服务器上进行, 系统为 Centos 7.7.1908。每个算法在一个实例上的执行时间限制 T 为 1000 秒。启发式算法在每个实例上运行十次, 其中支持随机种子 (AllDiff-LS, ACS) 的方法将使用不同的随机种子 (从 1 到 10)。其他算法在每个实例上只运行一次。对于每个算法, 本文使用 R 来表示一个实例族的成功运行百分比, 并使用 $time$ (以秒为单位) 来表示其平均成功时间。当 R 为 0 时, 其 $time$ 被记录为 ‘-’。

实验结果展示: 根据上文提到的求解方法和求解样例, 后续本文将比较这些求解器的求解性能, 以及通过一系列消融实验来证明设计的算法框架中各个策略的有效性, 具体指标为求解的成功率、比例和开销。

5.2 AllDiff-LS 求解 AllDifferent 约束的能力

根据每个问题的特点, 本文选择了具有代表性的基线方法进行比较, 具体地, 本文将问题实例分为了两类, 第一类是数独, 它是一类包含固定点的 CSP, 因此在搜索之前需要进行化简, 而其他三类无固定点的 CSP 属于第二类。

关于数独, 由于已经有专门的启发式算法性能优于一般的启发式求解器, 本文没有展示 Yuck 的实验结果, 其整体实验结果如表 5.1 所示。可以观察到, 在填充率超过 70% 后, 数独的难度急剧下降, 所有算法都可以通过推理规则轻易找到解决方案。其中, ACS 算法和 ACP-LS 算法的效率优于其他算法, 这证明了 ACP-LS 减少规则的有效性。

当填充比例在 40% 和 60% 之间时, 其求解难度最大。随着阶数的增加, 本文选择的算法中没有一个算法能保证 100% 的成功率。尤其是当阶数高于 5 时, 这些算法的求解性能开始恶化, 出现无法解决的情况。而 AllDiff-LS 的性能下降较慢, 直到最困难的 9 阶, 它仍然可以在 10 秒内解决除填充比例在 40%-60% 范围外的每一个实例。

¹<https://github.com/informarte/yuck>

表 5.1 AllDiff-LS 和其他最先进的基线方法在数独实例上的结果。

Table 5.1 Results of AllDiff-LS and other state-of-the-art baseline methods on Sudoku.

Instance	ACS		ILS		CPLEX		Choco		AllDiff-LS		Instance	ACS		ILS		CPLEX		Choco		AllDiff-LS	
family	R(%)	time	R(%)	time	R(%)	time	R(%)	time	R(%)	time	family	R(%)	time	R(%)	time	R(%)	time	R(%)	time	R(%)	time
S-4-0	100	0.04	100	0.08	100	18.50	100	1.11	100	<0.01	S-7-0	94.1	550.37	100	51.60	0	—	100	2.85	100	0.15
S-4-10	100	0.03	100	0.10	100	18.18	100	1.09	100	<0.01	S-7-10	92.2	559.82	29.2	544.69	1	814.81	90	58.82	100	0.15
S-4-20	100	0.03	100	0.15	100	17.15	100	1.09	100	<0.01	S-7-20	24.9	665.38	0	—	0	—	68	163.23	100	0.18
S-4-30	100	0.03	100	0.71	100	16.48	100	1.08	100	<0.01	S-7-30	0.3	798.53	0	—	0	—	4	226.72	100	0.24
S-4-40	100	0.02	100	1.52	100	16.79	100	1.08	100	<0.01	S-7-40	0	—	0	—	0	—	0	—	100	0.63
S-4-50	100	<0.01	100	0.04	100	14.01	100	1.03	100	<0.01	S-7-50	0	—	0	—	0	—	0	—	98.4	100.53
S-4-60	100	<0.01	100	<0.01	100	9.59	100	1.01	100	<0.01	S-7-60	99.2	8.92	68.3	271.46	49	619.98	100	1.48	100	0.06
S-4-70	100	<0.01	100	<0.01	100	7.40	100	1.02	100	<0.01	S-7-70	100	<0.01	100	3.14	100	30.86	100	1.32	100	<0.01
S-4-80	100	<0.01	100	<0.01	100	5.96	100	0.99	100	<0.01	S-7-80	100	<0.01	100	0.52	100	17.94	100	1.29	100	<0.01
S-4-90	100	<0.01	100	<0.01	100	5.39	100	1.05	100	<0.01	S-7-90	100	<0.01	100	0.20	100	9.60	100	1.27	100	<0.01
S-5-0	100	0.75	100	0.47	100	34.93	100	1.24	100	0.01	S-8-0	0	—	100	238.08	0	—	0	—	100	0.45
S-5-10	100	1.18	100	0.72	100	31.87	100	1.55	100	0.01	S-8-10	0	—	0	—	0	—	6	106.90	100	0.56
S-5-20	100	2.25	100	2.27	100	29.28	100	1.26	100	0.01	S-8-20	0	—	0	—	0	—	0	—	100	0.66
S-5-30	100	3.93	61.2	17.20	100	35.53	100	1.60	100	0.01	S-8-30	0	—	0	—	0	—	0	—	100	0.95
S-5-40	98.7	9.31	68.9	47.15	100	130.53	100	2.58	100	0.01	S-8-40	0	—	0	—	0	—	0	—	100	3.19
S-5-50	96.4	2.99	41.2	13.67	100	68.06	100	1.80	100	0.25	S-8-50	0	—	0	—	0	—	0	—	79.4	168.75
S-5-60	100	<0.01	100	0.08	100	15.73	100	1.01	100	<0.01	S-8-60	0	—	0	—	0	—	0	—	62	258.15
S-5-70	100	<0.01	100	0.02	100	10.23	100	1.04	100	<0.01	S-8-70	100	<0.01	100	18.24	100	90.17	100	1.28	100	<0.01
S-5-80	100	<0.01	100	<0.01	100	6.63	100	0.98	100	<0.01	S-8-80	100	<0.01	100	1.62	100	33.06	100	1.24	100	<0.01
S-5-90	100	<0.01	100	<0.01	100	5.58	100	1.03	100	<0.01	S-8-90	100	<0.01	100	0.47	100	5.90	100	1.22	100	<0.01
S-6-0	100	42.61	100	238.08	100	260.11	100	2.09	100	0.03	S-9-0	0	—	5	784.62	0	—	0	—	100	1.74
S-6-10	100	45.56	100	28.75	100	159.55	99	1.89	100	0.04	S-9-10	0	—	0	—	0	—	0	—	100	2.03
S-6-20	100	67.08	97.1	240.98	100	264.67	99	5.16	100	0.04	S-9-20	0	—	0	—	0	—	0	—	100	3.50
S-6-30	100	175.69	20.7	574.28	37	610.38	95	78.67	100	0.06	S-9-30	0	—	0	—	0	—	0	—	100	7.60
S-6-40	5.4	320.93	3.9	687.09	0	—	21	207.36	100	0.14	S-9-40	0	—	0	—	0	—	0	—	100	23.47
S-6-50	0	—	0	—	0	—	0	—	99.2	45.76	S-9-50	0	—	0	—	0	—	0	—	40.7	629.96
S-6-60	100	<0.01	100	4.49	100	40.45	100	1.24	100	<0.01	S-9-60	0	—	0	—	0	—	0	—	0	—
S-6-70	100	<0.01	100	0.54	100	16.78	100	1.20	100	<0.01	S-9-70	100	<0.01	100	109.72	100	114.35	100	1.64	100	<0.01
S-6-80	100	<0.01	100	0.16	100	11.56	100	1.19	100	<0.01	S-9-80	100	<0.01	100	5.19	100	39.13	100	1.59	100	<0.01
S-6-90	100	<0.01	100	0.08	100	8.80	100	1.17	100	<0.01	S-9-90	100	<0.01	100	1.18	100	10.92	100	1.57	100	<0.01

表 5.2 AllDiff-LS 和其他最先进的基准方法在 N 皇后, 全间隔和 2-MOLS 问题上的结果。

Table 5.2 Results of AllDiff-LS and other state-of-the-art baseline methods on N-queens, All-interval and 2-MOLS.

Instance CPLEX Choco							Yuck							AllDiff-LS							Instance CPLEX Choco							Yuck							AllDiff-LS						
family	time	time	time	R(%)	time	R(%)	family	time	time	time	R(%)	time	R(%)	family	time	time	time	R(%)	time	R(%)	family	time	time	time	R(%)	time	R(%)	family	time	time	time	R(%)	time	R(%)							
Q-500	8.33	—	896.07	100	0.09	100	Q-2500	436.40	—	—	0	14.98	100	Q-4500	—	—	—	0	80.13	100	Q-5000	—	—	—	0	119.18	100	Q-5500	—	—	—	0	135.37	100							
Q-1000	69.93	—	—	0	1.08	100	Q-3000	624.28	—	—	0	28.03	100	Q-6000	—	—	—	0	167.59	100																					
Q-1500	138.22	—	—	0	2.94	100	Q-3500	837.99	—	—	0	39.70	100																												
Q-2000	272.83	—	—	0	8.48	100	Q-4000	—	—	—	0	60.98	100																												
A-10	0.52	0.54	1.16	100	< 0.01	100	A-16	2.86	174.36	6.84	100	0.01	100	A-22	217.25	—	565.35	100	6.01	100																					
A-12	0.56	1.51	1.58	100	< 0.01	100	A-18	3.24	276.35	52.44	100	0.23	100	A-24	164.58	—	—	0	42.31	100																					
A-14	0.62	19.24	2.56	100	< 0.01	100	A-20	76.51	—	64.10	100	1.13	100	A-26	—	—	—	0	214.75	90																					
L-3	1.80	2.62	3.31	100	< 0.01	100	L-5	3.20	1.22	29.27	100	< 0.01	100	L-8	—	—	—	0	345.53	90																					
L-4	2.16	0.95	2.71	100	< 0.01	100	L-7	14.38	—	—	0	3.86	100	L-9	—	—	—	0	—	0																					

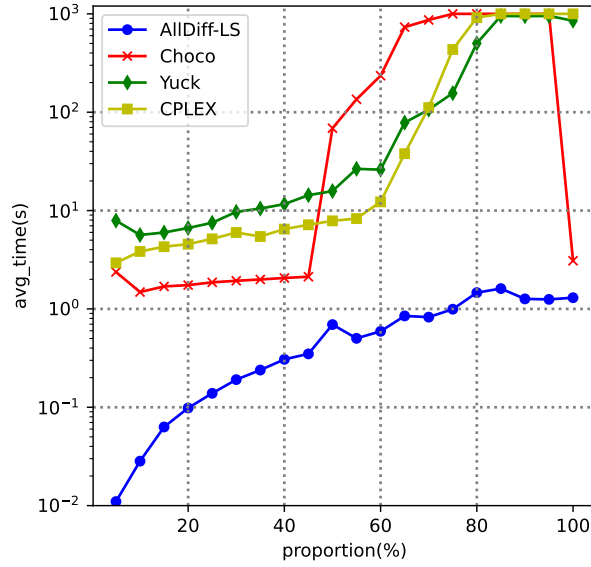


图 5.1 四种方法在给定基准测试上的结果。当 avg_time 为 1000s 时，意味着超时。

Figure 5.1 Results of methods on given benchmarks. when avg_time is 1000s, it means timeout.

此外，AllDiff-LS 算法可以保证除了阶数为 9 的实例外，其它的成功率超过 60%。特别是，在 11 个实例族上，ACS、ILS 和 CPLEX 无法解决任何实例，而 AllDiff-LS 在其中 8 个上可以达到 100% 的成功率，其余三个的成功率则分别是 98.4%、79.4% 和 40.7%。

对于 N 皇后、全间隔和 2-MOLS 问题，本文将 AllDiff-LS 的性能与 Yuck、CPLEX 和 Choco 进行了比较。由于每个实例族只包含一个实例，本文只显示启发式算法（AllDiff-LS，Yuck）的解决成功率。代表这三种问题的 ACGs 不需要被简化，但它们的结构更复杂，使得局部搜索有更高的求解压力。根据表 5.2，可以看出，除 \mathcal{L} -9 外，AllDiff-LS 可以解决几乎所有给定的实例。对于一些困难的实例，如 \mathcal{A} -26 和 \mathcal{L} -8，AllDiff-LS 也可以达到 90% 的解决成功率。而对于其他算法，CPLEX 的表现最好，但仍然无法求解大部分 N 皇后问题，以及一些高阶的全间隔问题和 2-MOLS 问题。因此在这三类实例上，相较于其他三个求解器，AllDiff-LS 的效果仍然是最好的。

本文还比较了 AllDifferent 约束数量对解决难度的影响。具体来说，本文选择了难度适中的实例族 \mathcal{S} -7-0，并通过随机删除实例族的约束，按顺序得到了 20 个新的约束逐渐减少的实例族（累计 2000 个实例）（约束从原始的 5% 到 100% 不等）。结果如图 5.1 所示。由于一些实例不再是数独，本文使用 Yuck 代替 ACS 和 ILS。横坐标表示约束逐渐增加的实例族，纵坐标表示每个实例族的平均时间消耗。结果表明，随着约束数量的增加，问题的解决难度在一定比例内会急剧增加，与其他方法相比，AllDiff-LS 受影响最小。

最后，本文给出额外的三个求解器：OR-Tools、LocalSolver 和 Kissat，出于类型原因，本文把它们的求解结果放在最后介绍。其中，Google 的 OR-Tools 是一套开源的软件库，它包含了丰富的解决方案，可以用于求解 CSP、LP、MIP 等各

表 5.3 OR-Tools、LocalSolver 和 SAT 求解器 Kissat 在数独实例上的结果。

Table 5.3 Results of OR-Tools, LocalSolver and SAT Solver Kissat on Sudoku.

Instance	OR-Tools	Kissat	LocalSolver	Instance	OR-Tools	Kissat	LocalSolver	Instance	OR-Tools	Kissat	LocalSolver						
family	R(%)	time	R(%)	time	time	family	R(%)	time	R(%)	time	time						
S-4-0	100	1.06	100	0.03	0.79	S-6-0	100	101.77	100	3.51	12.71	S-8-0	100	230.82	100	5.05	—
S-4-10	100	0.81	100	0.04	0.84	S-6-10	100	70.36	100	12.90	19.32	S-8-10	98	396.82	0	—	—
S-4-20	100	0.67	100	0.04	0.55	S-6-20	100	15.78	100	12.38	95.83	S-8-20	99	295.43	0	—	—
S-4-30	100	0.58	100	0.03	0.51	S-6-30	100	7.45	100	11.45	307.31	S-8-30	100	229.50	0	—	—
S-4-40	100	0.51	100	0.03	0.47	S-6-40	100	15.71	100	54.29	—	S-8-40	10	—	0	—	—
S-4-50	100	0.03	100	0.03	0.22	S-6-50	80	255.22	100	475.99	—	S-8-50	0	—	0	—	—
S-5-0	100	8.89	100	0.33	3.69	S-7-0	100	386.07	100	88.60	49.62	S-9-0	0	—	0	—	—
S-5-10	100	2.64	100	0.36	4.56	S-7-10	100	213.12	95	264.54	83.90	S-9-10	0	—	0	—	—
S-5-20	100	1.53	100	0.65	9.36	S-7-20	100	216.22	83	307.86	605.231	S-9-20	0	—	0	—	—
S-5-30	100	1.26	100	0.72	35.66	S-7-30	100	143.73	42	289.48	—	S-9-30	0	—	0	—	—
S-5-40	100	1.08	100	1.40	65.53	S-7-40	76	—	0	—	—	S-9-40	0	—	0	—	—
S-5-50	100	0.53	100	0.52	5.03	S-7-50	0	—	0	—	—	S-9-50	0	—	0	—	—

表 5.4 OR-Tools 和 LocalSolver 在其他实例上的结果。

Table 5.4 Results of OR-Tools and LocalSolver on other benchmarks.

Instance	OR-Tools	LocalSolver	Instance	OR-Tools	LocalSolver	Instance	OR-Tools	LocalSolver	Instance	OR-Tools	LocalSolver
family	time	time	family	time	time	family	time	time	family	time	time
Q-500	105.29	212.35	Q-2000	—	—	Q-3500	—	—	Q-5000	—	—
Q-1000	559.75	—	Q-2500	—	—	Q-4000	—	—	Q-5500	—	—
Q-1500	—	—	Q-3000	—	—	Q-4500	—	—	Q-6000	—	—
A-12	0.94	0.14	A-16	0.60	0.20	A-20	0.73	0.22	A-24	0.93	0.25
A-14	0.56	0.14	A-18	0.66	0.22	A-22	0.88	0.23	A-26	1.05	0.24
L-3	0.89	0.15	L-5	0.51	0.20	L-8	13.41	—			
L-4	0.37	0.16	L-7	6.37	17.55	L-9	63.69	—			

类优化问题，是比较新的开源求解工具，并且支持多线程求解；LocalSolver 是一款基于局部搜索的优化求解器，专门设计用来解决各种复杂的组合优化问题，它是一款商业求解器，同样支持多线程求解；Kissat 是一款高效的 SAT 求解器，它的一个显著特点是它的求解策略旨在利用现代多核处理器的并行计算能力，以提高求解效率。

由于 LocalSolver 不支持并行运行多个实例，对于数独的每个实例族，本文仅运行其中随机的一个实例，在表中本文不给出求解比例，仅提供单个实例的求解时间作为参考。同时，由于除数独外的例子都包含变量表达式，使用 SAT 对其编码比较困难（需要借助加法器），所以本文仅给出数独的 Kissat 求解结果。这些求解器的多线程求解结果如表 5.3 和表 5.4 所示，其中数独实例仍然舍弃了容易求解的部分样例族。可以观察到，OR-Tools 的求解能力最好，并且能够求解 AllDiff-LS 无法求解的实例 L-9，但在除此之外的其他实例上仍然不如 AllDiff-LS。而将 CSP 编码成 SAT 问题再使用 Kissat 求解的方法，相较于 OR-Tools 求解器在低阶数独上效果较好，但在高阶例子上表现较差。

5.3 AllDiff-LS 涉及的策略的有效性

本文首先研究两个简化规则的有效性。由于 N-queens 和 2-MOLS 的 ACG 无法被简化, 本文在数独实例上进行了实验, 其中推理规则的效果是非琐碎的 (去除了填充比例小于 40% 的实例)。结果如表 5.5 所示, 其中 *avg* 表示每个实例族变量数量的平均简化比例。可以看出, 使用两个简化规则比只使用 rule1 或 rule2 更好, 并且在填充率高的情况下, 大部分变量都可以被简化。

表 5.5 在给定数独实例上使用不同化简规则对应的变量简化比率。

Table 5.5 The variable simplification ratio of different simplification rules on given Sudoku.

Instance family	Rule1+2 <i>avg</i> (%)	Rule1 <i>avg</i> (%)	Rule2 <i>avg</i> (%)	Instance family	Rule1+2 <i>avg</i> (%)	Rule1 <i>avg</i> (%)	Rule2 <i>avg</i> (%)	Instance family	Rule1+2 <i>avg</i> (%)	Rule1 <i>avg</i> (%)	Rule2 <i>avg</i> (%)
S-4-40	9.44	8.22	1.18	S-6-40	0.43	0.43	0.01	S-8-40	0.02	0.02	0.00
S-4-50	79.34	70.41	27.20	S-6-50	4.28	3.97	0.21	S-8-50	0.48	0.47	0.01
S-4-60	94.91	94.90	93.95	S-6-60	96.54	96.54	10.63	S-8-60	7.10	6.42	0.41
S-4-70	95.29	95.29	95.28	S-6-70	98.93	98.93	98.93	S-8-70	99.12	99.12	99.09
S-4-80	91.06	91.06	91.06	S-6-80	98.10	98.10	98.09	S-8-80	99.40	99.40	99.40
S-4-90	82.96	82.96	82.96	S-6-90	93.78	93.78	93.78	S-8-90	97.22	97.22	97.22
S-5-40	1.53	1.42	0.09	S-7-40	0.09	0.09	0.00	S-9-40	0.01	0.01	0.00
S-5-50	20.92	14.69	2.25	S-7-50	1.35	1.31	0.04	S-9-50	0.16	0.16	0.00
S-5-60	95.98	95.92	83.16	S-7-60	43.80	18.48	1.63	S-9-60	3.05	2.92	0.10
S-5-70	97.65	97.65	97.64	S-7-70	99.11	99.11	99.10	S-9-70	99.31	99.30	15.45
S-5-80	95.54	95.54	95.54	S-7-80	98.90	98.90	98.90	S-9-80	99.56	99.56	99.56
S-5-90	89.42	89.42	89.42	S-7-90	95.95	95.95	95.95	S-9-90	98.28	98.28	98.28

此外, 需要对文章中提到的主要策略进行消融实验, 本文将实验分为了两大组, 一组用来研究移动选择方法和重启策略的影响, 而另一组用于展示了禁忌策略 1、禁忌策略 3 和打破平局技术对算法解决效率的影响。

第一组实验结果如图 5.2 所示, 每个点代表一个随机种子实验。为了使结果更清晰, 对于数独实验, 本文选择了五个困难的实例族 (S-6-50、S-7-50、S-8-50、S-8-60 和 S-9-50), 一个点代表十个随机种子实验的结果, 其中运行时间是它们运行时间的平均值。其中, AllDiff-LS-DM 用直接移动选择方法替换了本文的移动选择策略; 在 AllDiff-LS-FL 中, 使用固定数量的内循环迭代替代动态迭代, 并从算法 4 中移除权重策略。可见, 在数独问题上, AllDiff-LS-DM 的性能在几乎所有实例中都比 AllDiff-LS 弱, 从运行时间的比较中可以观察到移动选择策略的有效性。此外, AllDiff-LS-FL 在耗时的数独实例中表现出劣势。在其他三种类型的问题上, AllDiff-LS 对 AllDiff-LS-DM 的优势仍然存在, 但由于一些实例的规模较小, AllDiff-LS 对 AllDiff-LS-FL 的效率提升不明显。

在图 5.3 中, 本文展示了其他策略对数独实例的提升效果。具体的, 相较于原 AllDiff-LS 算法, 在 AllDiff-LS-TA 中本文将策略 1 替换为了传统的 tabu 策略, 在 AllDiff-LS-SW 中本文删除了策略 3, 而在 AllDiff-LS-BT 中本文删除了打破平局的策略。AllDiff-LS-TA 算法在一些较难的例子上会出现循环, 因此求解会超

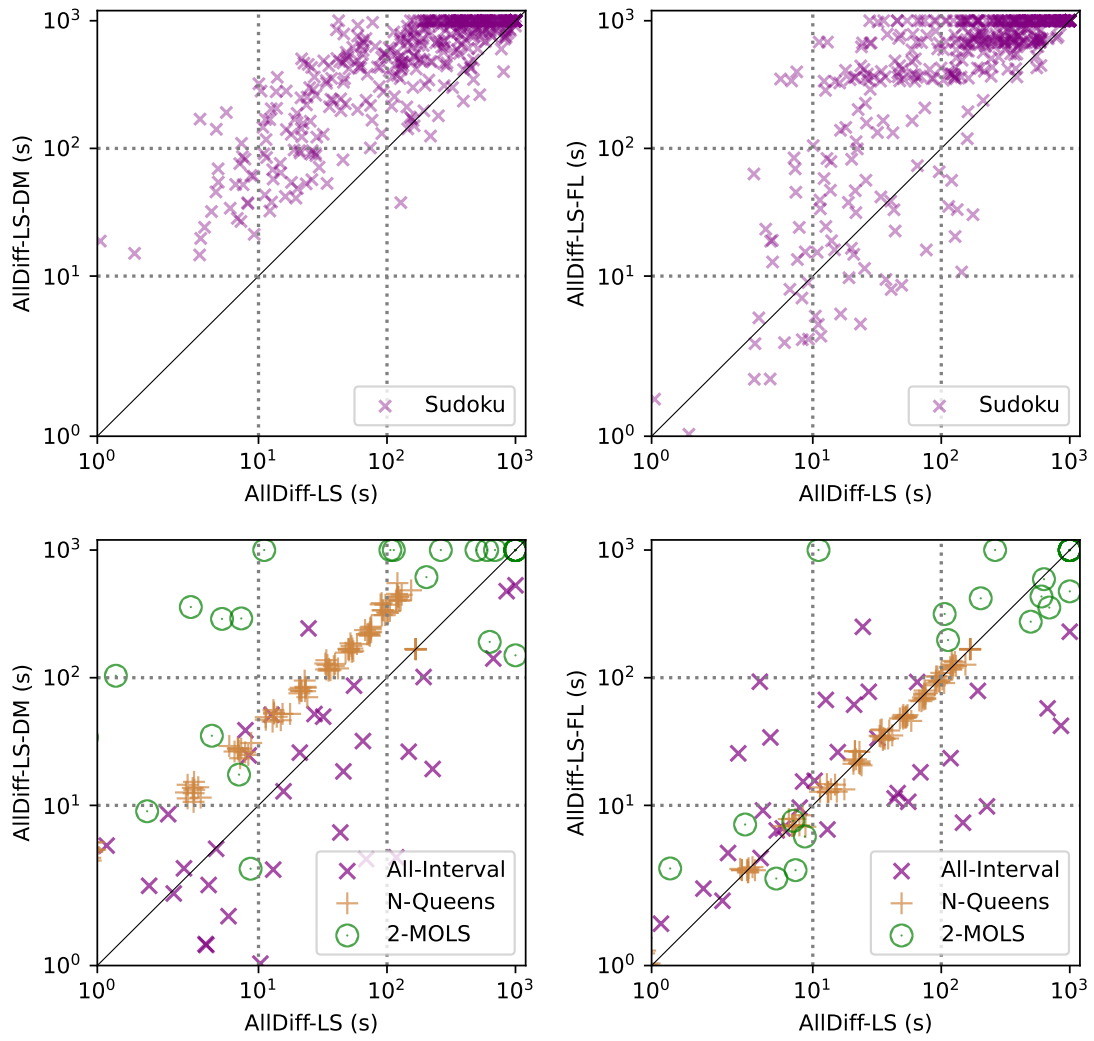


图 5.2 在四类实例上 AllDiff-LS（横坐标）和其他两个版本（纵坐标）的平均运行时间比较。
 Figure 5.2 Comparison of the average running time of AllDiff-LS (abscissa) and the other two versions (ordinate) on four kinds of benchmarks.

时。

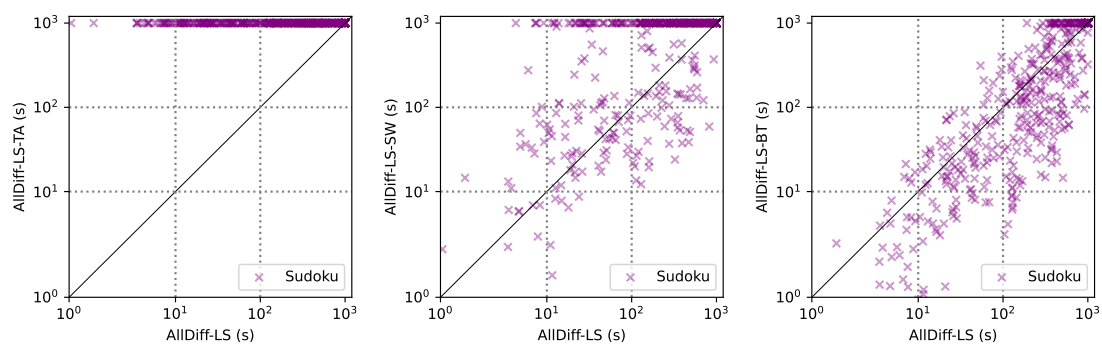


图 5.3 在数独上 AllDiff-LS（横坐标）与其他三个版本（纵坐标）的平均运行时间比较。

Figure 5.3 Comparison of the average running time of AllDiff-LS (abscissa) and the other three versions (ordinate) on Sudoku benchmarks.

5.4 本章小结

本章节通过设计实验，对基于 AllDiff-LS 算法实现的求解工具进行了深入的性能评估和策略分析。在第一节中，本文详细介绍了实验的安排，包括所使用的数据集和比较的方法。在第二节中，本文展示了算法在多个实验基准上与其他方法的性能比较结果，证明了其优越的性能。在第三节中，本文通过消融实验，对实验中提到的各种策略的有效性进行了验证。实验结果表明，基于 AllDiff-LS 算法实现的求解工具在处理各类问题时都展现出了强大的性能，证明了本文选择和实施的策略的有效性。

第6章 总结与展望

本章节主要对前面介绍的工作进行系统性的总结概括，归纳其中的主要贡献点，并对下一步的工作进行展望。

6.1 工作总结

包含 AllDifferent 约束的约束满足问题既重要又具有挑战性。虽然已有学者提出了一些算法来解决这些问题，但它们在大规模问题实例上的扩展性不佳。本工作主要提出了一种新的用于求解 AllDifferent 约束的局部搜索算法——AllDiff-LS，具体地，它由三部分组成：将 CSP 转化为局部搜索算法易处理的图结构，构造适用于给定图结构的局部搜索算法，以及设计用于该算法的重启策略。在三、四章中，本文对这三部分进行了详细的介绍。通过一系列实验，表明本文的算法在解决 AllDifferent 约束中的有效性，它可以在几分钟内解决大规模和复杂的问题实例。

对于 CSP 中的 AllDifferent 约束，本文构造了一个异构图称为 AllDifferent 约束图 (ACG)，它的思路是使用图结构对这些 AllDifferent 约束二元分解后得到的二元约束的并集进行表示。由于本文隐去了值域这个信息，一个 ACG 得以描述整组 AllDifferent 约束，从而可以施展更强的弧一致性算法。之后，本文定义了图上的赋值操作，以及基于它的化简规则，其核心思想是，根据变量和变量表达式顶点的度对图中的顶点和边进行删减。在对约束化简之后，本文通过随机赋值的方式获得最初的候选解。

本文将初始化后的 ACG 和候选解作为局部搜索的输入，通过定义状态、移动、评价准则等要素，设计了局部搜索算法。鉴于一个移动可以被分为两部分，本文设计了两步选择的移动策略作为直接选择策略的轮换策略。在此基础上，本文设计了一整套禁忌策略，它由三个小策略组成，前两个策略用于保证算法可以跳出循环，后一个策略则提供了选择操作的快速轮换机制，保证了算法的性能。同时，本文设计了打破对称的策略，借助变量表达式的邻域关系打破僵局。

此外，本文设计了基于解池技术的一整套重启策略。首先本文介绍了解池的维护机制，并提出了解的准入和生成策略，这两个策略分别用到了解近似的概念，和对解的扰动策略。此外，本文介绍了通过解池对 ACG 进行加权，从而减少对已探索区域的重复探索。最后，本文介绍了动态迭代的思想，在解加入解池和从解池中选择时，通过一定的策略增加下一轮迭代的最大迭代次数，这样做的动机是对复杂区域进行充分的探索，以寻找潜在的更优解。

基于上述思想，本文得到了求解 AllDifferent 约束的 AllDiff-LS 算法，通过将等式约束、偏序不等式约束引入到 ACG 中，本文可以处理比 AllDifferent 约束更广泛的 CSP 类型。比较实验和消融实验表明 AllDiff-LS 算法拥有更好的求解

能力，并且证明了算法中涉及各个策略的有效性。

6.2 下一步的工作

在未来，希望将局部搜索算法更加完善，让其适用于更广泛的约束满足问题上，实现一个更加通用的求解器，用来提升 CSP 求解的能力和效率。此外，将启发式策略和完备策略结合在一起对 CSP 进行求解，是一个有希望的改进方向。完备求解器一般依赖于传播和搜索，并因此具有较强的推理能力。一方面，在局部搜索之前或中途都可以使用约束传播对解空间进行化简；另一方面，完备搜索算法中关于变量序、分支等的启发式，同样可以应用于局部搜索中，做为指引。而其中一个思路是先在一个小的搜索窗口中使用完备的求解算法，如果无法求解则转为调用启发式求解方法。

参考文献

- [1] Cadar C, Dunbar D, Engler D. Klee: unassisted and automatic generation of high-coverage tests for complex systems programs [C]//OSDI'08: Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation. USA: USENIX Association, 2008: 209–224.
- [2] Godefroid P, Klarlund N, Sen K. Dart: directed automated random testing [C/OL]//PLDI '05: Proceedings of the 2005 ACM SIGPLAN Conference on Programming Language Design and Implementation. New York, NY, USA: Association for Computing Machinery, 2005: 213 – 223. <https://doi.org/10.1145/1065010.1065036>.
- [3] Yao P, Shi Q, Huang H, et al. Program analysis via efficient symbolic abstraction [J/OL]. Proc. ACM Program. Lang., 2021, 5(OOPSLA). <https://doi.org/10.1145/3485495>.
- [4] Beyer D, Dangl M, Wendler P. A unifying view on smt-based software verification [J/OL]. J. Autom. Reason., 2018, 60(3): 299–335. <https://doi.org/10.1007/s10817-017-9432-6>.
- [5] Wang J, Wang C. Learning to synthesize relational invariants [C/OL]//ASE '22: Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering. New York, NY, USA: Association for Computing Machinery, 2023. <https://doi.org/10.1145/3551349.3556942>.
- [6] Tappler M, Aichernig B K, Lorber F. Timed automata learning via smt solving [C/OL]//NASA Formal Methods: 14th International Symposium, NFM 2022, Pasadena, CA, USA, May 24–27, 2022, Proceedings. Berlin, Heidelberg: Springer-Verlag, 2022: 489–507. https://doi.org/10.1007/978-3-031-06773-0_26.
- [7] Xu R, An J, Zhan B. Active learning of one-clock timed automata using constraint solving [C/OL]//Automated Technology for Verification and Analysis: 20th International Symposium, ATVA 2022, Virtual Event, October 25–28, 2022, Proceedings. Berlin, Heidelberg: Springer-Verlag, 2022: 249–265. https://doi.org/10.1007/978-3-031-19992-9_16.
- [8] Amir G, Wu H, Barrett C, et al. An smt-based approach for verifying binarized neural networks [C/OL]//Tools and Algorithms for the Construction and Analysis of Systems: 27th International Conference, TACAS 2021, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2021, Luxembourg City, Luxembourg, March 27 – April 1, 2021, Proceedings, Part II. Berlin, Heidelberg: Springer-Verlag, 2021: 203 – 222. https://doi.org/10.1007/978-3-030-72013-1_11.
- [9] Katz G, Barrett C, Dill D L, et al. Reluplex: An efficient smt solver for verifying deep neural networks [C]//Majumdar R, Kunčák V. Computer Aided Verification. Cham: Springer International Publishing, 2017: 97–117.
- [10] Paulsen B, Wang C. Linsyn: Synthesizing tight linear bounds for arbitrary neural network activation functions [C/OL]//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. Berlin, Heidelberg: Springer-Verlag, 2022: 357 – 376. https://doi.org/10.1007/978-3-030-99524-9_19.
- [11] Paulsen B, Wang C. Example guided synthesis of linear approximations for neural network verification [C/OL]//Computer Aided Verification: 34th International Conference, CAV 2022,

- Haifa, Israel, August 7–10, 2022, Proceedings, Part I. Berlin, Heidelberg: Springer-Verlag, 2022: 149–170. https://doi.org/10.1007/978-3-031-13185-1_8.
- [12] Bae K, Gao S. Modular smt-based analysis of nonlinear hybrid systems [C/OL]//2017 Formal Methods in Computer Aided Design (FMCAD). 2017: 180-187. DOI: [10.23919/FMCAD.2017.8102258](https://doi.org/10.23919/FMCAD.2017.8102258).
- [13] Shoukry Y, Chong M, Wakaiki M, et al. Smt-based observer design for cyber-physical systems under sensor attacks [J/OL]. ACM Trans. Cyber-Phys. Syst., 2018, 2(1). <https://doi.org/10.1145/3078621>.
- [14] Cimatti A. Application of smt solvers to hybrid system verification [C]//2012 Formal Methods in Computer-Aided Design (FMCAD). 2012: 4-4.
- [15] Leike J, Heizmann M. Ranking templates for linear loops [J/OL]. Log. Methods Comput. Sci., 2015, 11(1). [https://doi.org/10.2168/LMCS-11\(1:16\)2015](https://doi.org/10.2168/LMCS-11(1:16)2015).
- [16] Heizmann M, Hoenicke J, Leike J, et al. Linear ranking for linear lasso programs [C/OL]// Hung D V, Ogawa M. Lecture Notes in Computer Science: volume 8172 Automated Technology for Verification and Analysis - 11th International Symposium, ATVA 2013, Hanoi, Vietnam, October 15-18, 2013. Proceedings. Springer, 2013: 365-380. https://doi.org/10.1007/978-3-319-02444-8_26.
- [17] Cimatti A, Mover S, Tonetta S. A quantifier-free SMT encoding of non-linear hybrid automata [C/OL]//Cabodi G, Singh S. Formal Methods in Computer-Aided Design, FMCAD 2012, Cambridge, UK, October 22-25, 2012. IEEE, 2012: 187-195. <https://ieeexplore.ieee.org/document/6462573/>.
- [18] Nieuwenhuis R, Oliveras A, Tinelli C. Solving SAT and SAT modulo theories: From an abstract Davis–Putnam–Logemann–Loveland procedure to DPLL(T) [J/OL]. J. ACM, 2006, 53(6): 937-977. <https://doi.org/10.1145/1217856.1217859>.
- [19] Jovanovic D, de Moura L M. Solving non-linear arithmetic [C/OL]//Gramlich B, Miller D, Sattler U. Lecture Notes in Computer Science: volume 7364 Automated Reasoning - 6th International Joint Conference, IJCAR 2012, Manchester, UK, June 26-29, 2012. Proceedings. Springer, 2012: 339-354. https://doi.org/10.1007/978-3-642-31365-3_27.
- [20] de Moura L M, Jovanovic D. A model-constructing satisfiability calculus [C/OL]// Giacobazzi R, Berdine J, Mastroeni I. Lecture Notes in Computer Science: volume 7737 Verification, Model Checking, and Abstract Interpretation, 14th International Conference, VMCAI 2013, Rome, Italy, January 20-22, 2013. Proceedings. Springer, 2013: 1-12. https://doi.org/10.1007/978-3-642-35873-9_1.
- [21] Caviness B F, Johnson J R. Quantifier elimination and cylindrical algebraic decomposition [C]//Texts and Monographs in Symbolic Computation. 2004.
- [22] Ábrahám E, Davenport J H, England M, et al. Deciding the consistency of non-linear real arithmetic constraints with a conflict driven search using cylindrical algebraic coverings [J/OL]. J. Log. Algebraic Methods Program., 2021, 119: 100633. <https://doi.org/10.1016/j.jlamp.2020.100633>.
- [23] Li H, Xia B, Zhang H, et al. Choosing better variable orderings for cylindrical algebraic decomposition via exploiting chordal structure [J/OL]. J. Symb. Comput., 2023, 116: 324-344. <https://doi.org/10.1016/j.jsc.2022.10.009>.
- [24] Cimatti A, Griggio A, Irfan A, et al. Incremental linearization for satisfiability and verification modulo nonlinear arithmetic and transcendental functions [J/OL]. ACM Trans. Comput. Log., 2018, 19(3): 19:1-19:52. <https://doi.org/10.1145/3230639>.

- [25] Khanh T V, Ogawa M. SMT for polynomial constraints on real numbers [C/OL]//Jeannet B. *Electronic Notes in Theoretical Computer Science*: volume 289 Third Workshop on Tools for Automatic Program Analysis, TAPAS 2012, Deauville, France, September 14, 2012. Elsevier, 2012: 27-40. <https://doi.org/10.1016/j.entcs.2012.11.004>.
- [26] Tung V X, Khanh T V, Ogawa M. raSAT: an SMT solver for polynomial constraints [J/OL]. *Formal Methods Syst. Des.*, 2017, 51(3): 462-499. <https://doi.org/10.1007/s10703-017-0284-9>.
- [27] Fontaine P, Ogawa M, Sturm T, et al. Subtropical satisfiability [C/OL]//Dixon C, Finger M. *Lecture Notes in Computer Science*: volume 10483 Frontiers of Combining Systems - 11th International Symposium, FroCoS 2017, Brasília, Brazil, September 27-29, 2017, Proceedings. Springer, 2017: 189-206. https://doi.org/10.1007/978-3-319-66167-4_11.
- [28] Nalbach J, Ábrahám E. Subtropical satisfiability for SMT solving [C/OL]//Rozier K Y, Chaudhuri S. *Lecture Notes in Computer Science*: volume 13903 NASA Formal Methods - 15th International Symposium, NFM 2023, Houston, TX, USA, May 16-18, 2023, Proceedings. Springer, 2023: 430-446. https://doi.org/10.1007/978-3-031-33170-1_26.
- [29] Cimatti A, Griggio A, Lipparini E, et al. Handling polynomial and transcendental functions in SMT via unconstrained optimisation and topological degree test [C/OL]//Bouajjani A, Holík L, Wu Z. *Lecture Notes in Computer Science*: volume 13505 Automated Technology for Verification and Analysis - 20th International Symposium, ATVA 2022, Virtual Event, October 25-28, 2022, Proceedings. Springer, 2022: 137-153. https://doi.org/10.1007/978-3-031-19992-9_9.
- [30] Ni X, Wu Y, Xia B. Solving smt over non-linear real arithmetic via numerical sampling and symbolic verification [C]//SETTA 2023. 2023.
- [31] Li H, Xia B, Zhao T. Square-free pure triangular decomposition of zero-dimensional polynomial systems [J]. *J. Syst. Sci. Compl.*, 2023.
- [32] Cai S, Li B, Zhang X. Local search for SMT on linear integer arithmetic [C/OL]//Shoham S, Vizel Y. *Lecture Notes in Computer Science*: volume 13372 Computer Aided Verification - 34th International Conference, CAV 2022, Haifa, Israel, August 7-10, 2022, Proceedings, Part II. Springer, 2022: 227-248. https://doi.org/10.1007/978-3-031-13188-2_12.
- [33] Cai S, Li B, Zhang X. Local search for satisfiability modulo integer arithmetic theories [J/OL]. *ACM Trans. Comput. Logic*, 2023, 24(4). <https://doi.org/10.1145/3597495>.
- [34] Li B, Cai S. Local search for smt on linear and multi-linear real arithmetic [C/OL]//2023 Formal Methods in Computer-Aided Design (FMCAD). 2023: 1-10. DOI: [10.34727/2023/isbn.978-3-85448-060-0_25](https://doi.org/10.34727/2023/isbn.978-3-85448-060-0_25).
- [35] Li H, Xia B, Zhao T. Local search for solving satisfiability of polynomial formulas [C/OL]//Enea C, Lal A. *Lecture Notes in Computer Science*: volume 13965 Computer Aided Verification - 35th International Conference, CAV 2023, Paris, France, July 17-22, 2023, Proceedings, Part II. Springer, 2023: 87-109. https://doi.org/10.1007/978-3-031-37703-7_5.
- [36] Barrett C, Fontaine P, Tinelli C. The Satisfiability Modulo Theories Library (SMT-LIB) [Z]. 2016.

致 谢

我要对我的导师张健研究员，以及马菲菲研究员表达深深的感谢。他们的精心指导和无私帮助，使我从一个对科研知之甚少的本科毕业生，逐步深入到科学研究的广阔天地。他们像明灯一样照亮了我探索未知的道路。在科研过程和论文投稿的各个阶段，他们都以严谨负责的态度给予我指导和帮助，这种敬业精神深深地感染和激励着我。

我也要向我的师兄们表示感谢。特别是贾富琦师兄，他从我研究生一年级开始就引领我进行实验和阅读论文，使我在科研知识的海洋中打下了坚实的基础。刘明昊师兄在我研究生二年级时带领我熟悉独立进行科研和论文写作，使我逐渐对科研生涯充满了热爱。此外，我还要感谢董宇航师弟，我们在研究生三年级时互相鼓励，共同完成了许多科研任务。在我完成毕业论文的过程中，师弟柴文健、董宇航和师妹师朗辰的帮助也是我无法忘记的。

在未来，我将继续在张老师的指导下进行博士研究。我会一直以两位老师为我学术生涯的灯塔，他们的专业精神和严谨态度将永远是我学术追求的榜样。

作者简历及攻读学位期间发表的学术论文与研究成果

作者简历：

韩瑞，男，山东泰安人，2000 年生，中国科学院软件研究所硕士研究生。

2017 年 9 月——2021 年 6 月，在吉林大学软件学院获得学士学位。

2021 年 9 月——2024 年 6 月，在中国科学院软件研究所攻读硕士学位。

已发表（或正式接受）的学术论文（加星号的表示共一作者）：

1. Fuqi Jia*, **Rui Han***, Xutong Ma, Baoquan Cui, Minghao Liu, Pei Huang, Feifei Ma, Jian Zhang. PSMT: Satisfiability Modulo Theories Meets Probability Distribution. (ASE 2023, New Ideas Track)
2. Fuqi Jia, **Rui Han**, Pei Huang, Minghao Liu, Feifei Ma, Jian Zhang. Improving Bit-Blasting for Nonlinear Integer Constraints. (ISSTA 2023)
3. Minghao Liu, **Rui Han**, Fuqi Jia, Pei Huang, Feifei Ma, Hantao Zhang, Jian Zhang. Investigating the Existence of Holey Latin Squares via Satisfiability Testing. (PRICAI 2023)
4. Minghao Liu*, Kunhang Lv*, Pei Huang, **Rui Han**, Fuqi Jia, Yu Zhang, Feifei Ma, Jian Zhang. NRAgo: Solving SMT(NRA) Formulas with Gradient-Based Optimization. (ASE 2023, Tool Track)

投稿经历

1. AllDiff-LS: Solving Alldifferent Constraints with Efficient Local Search, AAAI 2023 过第一阶段，未中。

参加的研究项目及获奖情况：

1. 参与了课题组优化模理论求解工具的工具开发和测试。
2. 参与了课题组非线性求解工具的开发和测试。
3. 获得 2023 年国科大三好学生。

