软件学报 ISSN 1000-9825, CODEN RUXUEW
[doi: 10.13328/j.cnki.jos.007176] [CSTR: 32375.14.jos.007176]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于 PAC 学习的组合式概率障碍证书生成*

杨紫萱[1], 曾 霞[2], 任勐鑫[3], 王建林[3], 曾振柄[4], 杨争峰[1]

[1](华东师范大学 软件工程学院, 上海 200062)

[2](西南大学 计算机与信息科学学院, 重庆 400715)

[3](河南大学 计算机与信息工程学院, 河南 开封 475001)

[4](上海大学 理学院, 上海 200444)

通信作者: 曾霞, E-mail: xzeng0712@swu.edu.cn

摘 要: 连续动力系统安全验证是一个重要的研究问题, 多年来各类验证方法所能处理的问题规模非常受限. 对此, 对于给定的连续动力系统, 提出通过反例制导方法生成一组组合式概率近似正确 (PAC) 障碍证书的算法, 最终给出无限时间范畴安全验证问题在概率统计意义下的形式化描述. 通过建立和求解基于大 M 法的混合整数规划方法, 将障碍证书的求解转化为约束优化问题. 通过微分中值定理将非线性不等式进行区间线性化. 最后, 实现组合式 PAC 障碍证书生成工具 CPBC, 并在 11 个基准系统上评估其性能. 实验结果表明, CPBC 均能成功验证每个动力系统在指定不同的安全需求阈值下的安全性. 与现有方法相比, 所提方法可以更高效地为复杂系统或高维系统生成可靠的概率障碍证书, 验证的样例规模已高达百维.

关键词: 连续动力系统; 障碍证书; PAC; 区间线性化; 混合整数规划

中图法分类号: TP311

中文引用格式: 杨紫萱, 曾霞, 任勐鑫, 王建林, 曾振柄, 杨争峰. 基于PAC学习的组合式概率障碍证书生成. 软件学报. http://www.jos.org.cn/1000-9825/7176.htm

英文引用格式: Yang ZX, Zeng X, Ren MX, Wang JL, Zeng ZB, Yang ZF. Compositional Probabilistic Barrier Certificate Generation Based on PAC Learning. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7176.htm

Compositional Probabilistic Barrier Certificate Generation Based on PAC Learning

YANG Zi-Xuan[1], ZENG Xia[2], REN Meng-Xin[3], WANG Jian-Lin[3], ZENG Zhen-Bing[4], YANG Zheng-Feng[1]

[1](Software Engineering Institute, East China Normal University, Shanghai 200062, China)

[2](College of Computer and Information Science, Southwest University, Chongqing 400715, China)

[3](School of Computer and Information Engineering, Henan University, Kaifeng 475001, China)

[4](College of Sciences, Shanghai University, Shanghai 200444, China)

Abstract: Continuous dynamical systems safety verification is an important research issue, and over the years, various verification methods have been very limited in the scale of the problems they can handle. For a given continuous dynamical system, this study proposes an algorithm to generate a set of compositional probably approximately correct (PAC) barrier certificates through a counterexample-guided approach. A formal description of the infinite-time domain safety verification problem is given in terms of probability and statistics. By establishing and solving a mixed-integer programming method based on the Big-M method, the barrier certificate problem is transformed into a constrained optimization problem. Nonlinear inequalities are linearized in intervals using the mean value theorem of differentiation. Finally, this study implements the compositional PAC barrier certificate generator CPBC and evaluates its performance on 11 benchmark systems. The experimental results show that CPBC can successfully verify the safety of each dynamical system under specified different

* 基金项目: 国家重点研发计划 (2022YFA1005101); 国家自然科学基金 (12171159, 62272397); 上海市可信工业互联网软件协同创新中心; “数字丝绸之路”上海市可信智能软件国际联合实验室项目 (22510750100)
本文由“形式化方法与应用”专题特约编辑曹钦翔副教授、宋富研究员、詹乃军研究员推荐.
收稿时间: 2023-09-11; 修改时间: 2023-10-30, 2023-12-13, 2024-01-08; 采用时间: 2024-02-29; jos 在线出版时间: 2024-11-06

safety requirement thresholds. Compared with existing methods, the proposed method can more efficiently generate reliable probabilistic barrier certificates for complex or high-dimensional systems, with the verified example scale reaching up to hundreds of dimensions.

Key words: continuous dynamical system; barrier certificate; probably approximately correct (PAC); interval linearization; mixed-integer programming

安全验证问题是利用形式化方法证明系统在运行过程中满足指定安全需求的重要研究问题. 这类研究对涉及安全攸关系统的应用领域尤为重要, 例如自动驾驶汽车[1]、医疗设备[2]、航空航天[3]等多个领域, 对人类生命安全、环境保护以及重要基础设施的正常运行具有重大意义. 形式化验证是对动力系统进行验证的重要方法, 主要可分为准确验证和概率验证两种方式. 而安全攸关动力系统通常涉及多个复杂因素的相互作用, 包括物理动力学、控制算法和环境因素等. 在验证过程中, 不确定性和噪声的影响使得系统的行为难以预测. 此外, 动力系统的状态空间通常是高维的, 增加了验证的复杂性. 因此, 通过准确验证处理这些复杂系统时常常面临极大的限制和困难, 导致一系列方法在实际应用中问题规模受到一定限制, 一般在 20 维以内.

在许多实际应用中, 系统安全性的需求往往可以通过达到一定的概率精度来满足. 系统安全性的概率验证分析是一个关键的研究领域. 主要包含两种类型的问题, 其中一类问题是寻求系统的概率近似安全性, 即在高置信度下保证系统的安全性, 另一类问题通过确定系统安全概率的下界来进行安全性评估. 之前的研究中, Xue 等人采用概率近似正确 (probably approximately correct, PAC) 学习框架[4], 通过生成 PAC 障碍证书 (barrier certificate, BC) 验证系统的概率安全性[5]. 然而, 当需要达到很高的安全概率要求时, 部分样本点可能始终无法满足安全条件. 为了应对这种情况, Xue 等人采取将这些样本点进行丢弃的方法[5]. 本文提出了生成一组组合式 PAC 障碍证书的算法, 在验证复杂动力系统的安全性时, 能够更加灵活地求解. 该算法采用反例制导的方法学习一组障碍证书, 直到验证过程中没有反例点为止. 组合式障碍证书采用多个障碍证书联合的思想, 提高了整体的性能和鲁棒性, 允许在优化问题求解过程中纠正由于样本和概率不完备性导致的误差, 并且能够有效地改善验证结果. 与先前的方法相比, 本文提出的方法通过概率方式描述连续动力系统的安全性, 在准确验证方法无法验证系统安全性的情况下, 本方法能够进行定量的概率分析, 从而有效地验证高维复杂系统的安全性. 需要指出的是, 虽然概率方法提供了系统安全性的定量评估, 但这并不意味着可以确保绝对的安全.

为了学习单个 PAC 障碍证书, 本文提出一种基于大 M 法的混合整数规划方法进行求解. 在给定安全需求阈值参数 ε 和置信度水平参数 β 的情况下, 该方法的目标是学习在初始区域上至少 $1-\beta$ 的置信度下满足安全属性的概率大于等于 $1-\varepsilon$ 的概率障碍证书. 通过引入常数 M 和二进制变量, 能够将不满足约束条件的点进行调整, 以满足对障碍证书的要求. 具体地, 对于每个从初始区域中随机采样的样本点, 将其对应的约束减去 M 乘以二进制变量. 当二进制变量为 1 时, 约束条件对该点不产生影响; 当二进制变量为 0 时, 该点必须满足约束条件. 通过在优化目标中最小化所有二进制变量的和, 目标是使不满足原问题约束条件的点的数量尽可能少, 以提高验证的效率和可行性. 基于大 M 法的混合整数规划方法能够有效处理 PAC 学习中可能出现的不满足约束条件的情况. 为了处理约束条件, 本文采用场景优化方法[6]和区间不等式线性化方法[7−10], 并通过基于微分中值定理的非线性不等式区间化方法, 将非线性不等式转化为线性不等式[11], 使用数值优化器 Gurobi 进行求解. 这种方法相比直接进行区间乘法运算能够更准确地近似原始约束, 并提高优化问题的求解效率. 通过求解上述优化问题, 可以计算出具有鲁棒性和可信度的 PAC 障碍证书.

本文旨在为连续动力系统的概率安全性验证提供一种创新方法. 通过结合 PAC 学习框架、基于大 M 法的混合整数规划和反例制导方法, 可以获得一组组合式的 PAC 障碍证书, 从而提高系统概率安全性的保障. 为了验证方法的有效性和适用性, 本文将该方法应用于 11 个连续动力系统的概率安全性验证任务, 并与现有验证方法进行对比实验. 结果表明, 所提方法在保证准确性的前提下, 大大减少了计算时间和空间的开销. 进一步推动连续动力系统概率安全验证领域的发展. 综上, 本文的主要贡献包括以下内容.

- 提出了通过反例制导方法学习一组组合式 PAC 障碍证书的算法, 增强了连续动力系统的概率安全保障.
- 将问题转化为基于大 M 法的混合整数规划问题, 使不满足约束条件的点的数量最小化, 从而提高障碍证书

的鲁棒性和可信度.

● 通过场景优化方法和基于微分中值定理的区间化方法将非线性不等式转化为线性不等式, 能够更加近似原始约束, 提高优化问题的求解效率.

● 实现了组合式 PAC 障碍证书生成工具 CPBC, 并在 11 个基准系统上评估其性能. 实验结果表明该工具成功验证了所有样例的概率安全性, 并且相比现有方法对于验证复杂系统或高维系统更高效.

本文第 1 节介绍相关工作. 第 2 节介绍预备知识. 第 3 节介绍动力系统概率安全验证任务. 第 4 节介绍组合式 PAC 障碍证书的动机和计算方法. 第 5 节介绍算法流程, 并通过具体例子展示计算方法. 第 6 节对 11 个非线性连续动力系统进行实验评估. 第 7 节总结全文.

1 相关工作

在实际应用中, 许多安全攸关系统都是基于连续动力系统构建的. 对于验证连续动力系统的安全性, 不变量生成是一种成熟的近似方案, 可为系统在无限时间范围内的安全性提供可靠的证明. 不变量可以采用多种形式构建, 比如通过障碍证书[12]证明系统在无限时间范围内的安全性. 通过预先指定的模板, 可以将问题简化为数值优化或约束求解问题[13−16]. 使用障碍证书进行验证的最初想法是由 Prajna 等人[17]提出的. 障碍证书能否成功合成与 3 个因素密切相关, 即障碍证书模板的形式、障碍证书条件的编码以及生成障碍证书的计算方法. 研究一直致力于各种形式的障碍证书, 力图在表达性和合成可行性之间取得平衡. 障碍证书可以按照其形式和验证方法的不同, 分为确定形式和概率形式.

确定形式的障碍证书通常用于验证确定性连续动力系统的安全性. 其中, 障碍证书的形式是一个确定的函数或多项式, 能够精确地描述系统的安全区域和边界. 确定形式的障碍证书可以通过数值计算或数学推导方法来判断其是否存在, 并可用于保证系统在给定约束条件下保持在一个安全区域内. Prajna 等人[17,18]采用正立方体定理推导出障碍证书的平方和程序, 这导致了一个属于 NP 难问题的双线性矩阵不等式求解问题. Sloth 等人[19]建立了一种组合形式的条件, 用于验证特定类别的动力系统的安全性. Kapinski 等人[20]提出了一种基于 Lyapunov 函数的障碍证书, 比 Prajna 提出的障碍证书更保守但更易于处理. Platzer 等人[21]探讨了微分不变量来扩展障碍证书的思想, 考虑多项式等式和不等式的布尔组合, 扩展了障碍证书的范围, 从而可以描述更广泛的系统安全性问题. Sogokon 等人[22]采用一种称为类 Lyapunov 函数方法的技术, 来放宽障碍证书的导数条件, 以保持搜索空间的凸性, 从而能够搜索具有更一般特征的更广泛的障碍证书. 对于障碍证书条件的编码, 一种常用方法是使用平方和多项式和半定规划理论将验证条件转化为凸优化问题[18,23,24]. 另一种方法是使用 Hamilton-Jacobi 可达性分析的概念, 将验证条件编码为偏微分方程[25]. 在确定形式的障碍证书生成的计算方面, 最广泛使用的方法是平方和编程和 BMI 求解[23], 已成功应用于各种类型的系统. Zeng 等人[26]提出使用 Darboux 形式的障碍证书来保证系统的安全性, 其基于代数曲线定义, 并可以对系统的轨迹施加约束.

概率形式的障碍证书适用于具有确定性或随机性的系统, 可以描述系统在一定概率下保持在安全区域内的可能性, 更全面地考虑系统的不确定性. 现有研究大多数采用统计模型检验[27,28]或概率可达集计算[29,30]来验证随机系统. 但统计模型检验只能提供统计保证[31]. 概率形式的障碍证书可以通过随机模型或随机过程来定义. 随机模型通常基于概率分布函数或随机变量, 用于描述系统的随机性质. 随机过程是一种描述随机行为随时间演变的数学工具. Xue 等人[5]使用 PAC 框架定义概率障碍证书, 旨在研究在给定样本和一定置信度的情况下, 如何从训练数据中学习出一个在未知数据上表现良好的模型, 来验证受扰动输入的动力系统的概率安全性. Huang 等人[32]提出一种基于障碍证书的随机初始状态半代数混成系统的概率安全验证方法, 将概率安全验证问题转化为在初始状态空间中寻找可找到的障碍证书的最大内切区域的问题. Kumar 等人[33]使用高斯过程对残差动态的投影建模为控制障碍函数, 结果表明相比神经网络的确定性方法, 概率方法能够显著减少安全违规次数. Luo 等人[34]提出了概率安全障碍证书 (PrSBC), 使用控制障碍证书来定义概率安全的可能控制动作的空间, 通过约束机会约束安全集的二次规划来确定安全控制器. Jing 等人[35]提出了一种安全性和性能规范的障碍证书, 该证书通过在长期概率上

定义的前向不变性的新概念, 集成了基于可达性和基于不变性的方法, 能保证多全局目标的随机安全控制技术. 随着概率形式障碍证书技术的发展, 其将在越来越多的实际应用中发挥重要作用, 为复杂系统的概率安全性提供更全面的保障.

2 预备知识

本节将介绍动力系统的概率安全性问题、场景优化方法和基于微分中值定理的区间线性化方法.

2.1 动力系统安全性

考虑一类由自治型常微分方程建模的连续动力系统:

$$\dot{x} = f(x) \tag{1}$$

其中, $x \in \mathbb{R}^n$ 是状态向量, $\dot{x}$ 表示其时间导数 $\mathrm{d}x/\mathrm{d}t$, $t \in \mathbb{R}_0^+$ 表示时间, $f: \mathbb{R}^n \to \mathbb{R}^n$ 是在状态空间 $D \subseteq \mathbb{R}^n$ 上定义的向量场. 假设 f 满足局部利普希茨条件, 使得在状态空间 D 中对于每个初始状态 x_0, 存在一个有限的时间区间 $[0,T]$, 在该区间内动力系统有唯一解 $x(x_0,t)$.

动力系统 (1) 通常配有系统状态空间 $X \subseteq D$ 和初始区域 $X_0 \subseteq X$, X 为一个有界紧集. Ω 为 X_0 服从的分布, 用于选择初始连续状态. 将动力系统表示为四元组 $C = (f, X_0, \Omega, X)$. 然而, 动力系统在实际应用中面临着安全性的挑战. 安全性指系统在运行过程中具有始终不会进入非安全区域的性质. 对于连续动力系统而言, 安全性的保证至关重要, 因为系统的异常状态可能导致严重的后果. 给定一个预先指定的非安全区域 $X_u \subseteq X$, 如果从初始区域 X_0 出发的系统轨迹在任意时间内都不会进入非安全区域 X_u, 则称动力系统 C 是安全的.

定义 1 (安全性). 对于一个连续动力系统 $C = (f, X_0, \Omega, X)$ 和给定的非安全区域 $X_u \subseteq X$, 其中 $X_u \cap X_0 = \varnothing$, 若系统轨迹始终在 X 中并且不会进入非安全区域 X_u, 即对任意 $x_0 \in X_0$, 任意 $T \geqslant 0$:

$$\forall t \in [0,T],\ (x(x_0,t) \in X) \wedge (x(x_0,T) \notin X_u),$$

则称该系统是安全的.

定义 1 是对于动力系统安全性的语义定义, 旨在确保系统在运行过程中始终不会进入非安全区域. 动力系统通常由非线性常微分方程表示, 由于难以获得精确解, 对于动力系统的安全性验证一直以来是一个具有挑战性的问题. 目前能够验证的问题规模受到一定的限制, 一般局限在 20 维以内. 然而, 如果允许系统在一定概率情况下不满足安全性质, 即不要求所有从初始区域出发的状态轨迹都完全避开非安全区域, 这样的系统被称为概率安全系统. 动力系统的概率安全性对系统的安全性进行定量分析, 扩展了问题规模的可行性, 使其适应更为复杂的实际工业问题, 有效地处理了不确定性因素. 定义 2 为动力系统的概率安全性定义.

定义 2 (概率安全性). 连续动力系统 $C = (f, X_0, \Omega, X)$ 关于 $\varepsilon \in (0,1)$ 和 $\beta \in (0,1)$ 是概率近似安全的, 如果在至少 $1-\beta$ 的置信度下, 动力系统满足安全性质的概率大于等于 $1-\varepsilon$.

定义 2 表明若动力系统是概率近似安全的, 则在至少 $1-\beta$ 的置信度下, 从满足 Ω 的随机初始状态 x_0 出发的状态点在后续演化中不会进入非安全区域的概率至少为 $1-\varepsilon$. 系统的概率安全性是通过 PAC 学习框架定义的. PAC 学习框架广泛应用于机器学习领域, 是一种基于统计学习理论的方法. 该方法提供了在有限数据集上通过随机采样评估系统安全性的方法. 目的是在给定安全需求阈值参数和置信度水平参数的情况下保证学习部分的训练效果, 使得该框架适用于形式验证的目的. 本文仅考虑 Ω 为均匀分布的情况, 然而, 本文的方法也适用于其他分布情况.

2.2 场景优化方法

场景优化方法 (scenario optimization approach)[6]是一种用于解决具有无限个约束的凸优化问题的有效技术. 在这种方法中, 假设对不确定性进行概率描述, 通过一组 Δ 以及 Δ 上的概率分布 P 来表示不确定性, 将不确定性引入优化问题的建模中, 具有统计形式的保证. 由于在计算中无法处理大量的约束 $h_\delta(\gamma) \leqslant 0, \forall \delta \in \Delta$, 通过概率分布 P 在 Δ 上随机采样不确定性参数 δ 的 K 个独立同分布的样本 $\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(K)}$, 并将原问题转化为求解与这 K 个样本

相关的约束优化问题, 从而降低计算复杂度. 假设原凸优化问题为:

$$\begin{cases}\min\limits_{c} \ c^{\mathrm{T}}\gamma \\ \text{s.t. } h_{\delta}(\gamma)\leqslant 0, \forall \delta\in \Delta\end{cases} \tag{2}$$

将求解无限个约束的凸优化问题转化为求解有限个约束的凸优化问题:

$$\begin{cases}\min\limits_{c} \ c^{\mathrm{T}}\gamma \\ \text{s.t. } h_{\delta^{(i)}}(\gamma)\leqslant 0,\ i=1,2,\ldots,K\end{cases} \tag{3}$$

公式 (3) 放松了公式 (2), 因为它只考虑了与采样的样本 $\delta^{(i)}$ 相对应的 K 个有限约束, 若公式 (3) 无解, 则公式 (2) 的解集为空集. 由于公式 (3) 的约束有限, 并且是一个凸优化问题, 因此只要 K 不是一个非常大的数, 可以通过数值优化器进行求解.

在场景优化方法中, 最关键的是如何选择适当的样本数 K, 以保证得到足够准确的优化结果. 通常, K 的取值要足够大, 以确保样本的覆盖性和可靠性. 然而 K 又不能过大, 应尽量减少计算成本. 因此, 对于不同的问题和应用场景, 需要综合考虑样本数 K 的选择, 以平衡准确性和计算效率之间的权衡. 对于样本数 K, 存在以下定理.

定理 1. 若公式 (3) 可行并且存在唯一的最优解 γ^{*}, 给定安全需求阈值参数 $\varepsilon\in(0,1)$ 和置信度水平参数 $\beta\in(0,1)$, 如果满足

$$\varepsilon\geqslant\frac{2}{K}\left(\ln\frac{1}{\beta}+m\right) \tag{4}$$

其中, K 为从 Δ 中随机采样的样本数, m 为约束中优化变量的个数, 那么至少在 $1-\beta$ 的置信度下最优解 γ^{*} 满足 Δ 中所有约束的概率大于等于 $1-\varepsilon$, 即 $P(\{\delta\in\Delta|h_{\delta}(\gamma^{*})\leqslant 0\})\geqslant 1-\varepsilon$.

2.3 基于微分中值定理的区间线性化方法

对于线性区间不等式组 $A^{I}x\leqslant b^{I}$, 其中 $A^{I}=\{A\mid \underline{A}\leqslant A\leqslant \overline{A}\}$ 是一个 $m\times n$ 维区间矩阵, $b^{I}=\{b\mid \underline{b}\leqslant b\leqslant \overline{b}\}$ 是一个 m 维区间向量. 向量 x_0 是 $A^{I}x\leqslant b^{I}$ 的一个强解, 若满足 $Ax_0\leqslant b$ 对于每个 $A\in A^{I}$ 和 $b\in b^{I}$ 都成立. 用 X_S 表示所有强解的集合, 有以下命题成立.

命题 1. 若线性区间不等式组 $A^{I}x\leqslant b^{I}$ 有解, 则所有强解的集合 $X_S=\{x_1-x_2\mid \overline{A}x_1-\underline{A}x_2\leqslant \underline{b}, x_1\geqslant 0, x_2\geqslant 0\}$. 若线性区间不等式组 $A^{I}x\geqslant b^{I}$ 有解, 则所有强解的集合 $X_S=\{x_1-x_2\mid \underline{A}x_1-\overline{A}x_2\geqslant \overline{b}, x_1\geqslant 0, x_2\geqslant 0\}$.

若想得到具有区间数值的线性不等式系统的强解, 则可以通过命题 1 将区间不等式转化为线性规划问题计算. 定义 3 定义了区间运算的基本规则.

定义 3. 已知区间 $X=[\underline{a},\overline{a}]$ 和区间 $Y=[\underline{b},\overline{b}]$, 则区间 $X\cdot Y=[\min\{\underline{ab},\underline{a}\overline{b},\overline{a}\underline{b},\overline{ab}\},\max\{\underline{ab},\underline{a}\overline{b},\overline{a}\underline{b},\overline{ab}\}]$, 区间 $X-Y=[\underline{a}-\overline{b},\overline{a}-\underline{b}]$. 给定一个常数 c, $X\pm c=[\underline{a}\pm c,\overline{a}\pm c]$, $cX=[c\underline{a},c\overline{a}]$.

每个区间表示为一对下界和上界, 当线性化一个区间不等式时, 更小的区间长度通常近似效果更好, 因为这意味着下界和上界的差异更小, 可以减少线性化过程中的误差和近似带来的不确定性. 对于已知的连续函数 $g(x)$, $x\in[\underline{x},\overline{x}]$, 由微分中值定理可知:

$$g(\overline{x})-g(\underline{x})=\frac{\partial g}{\partial x}(\zeta)(\overline{x}-\underline{x}),\ \zeta\in[\underline{x},\overline{x}].$$

定理 2 引入基于中值定理的连续函数区间化方法.

定理 2. 给定一个连续函数 $g(x):\mathbb{R}^{n}\to\mathbb{R}, x\in[\underline{x},\overline{x}]$, 中点 $\check{x}=\dfrac{\underline{x}+\overline{x}}{2}$, 则

$$g(x)\in g(\check{x})+\frac{\partial g}{\partial x}([\underline{x},\overline{x}])[\underline{x}-\check{x},\overline{x}-\check{x}],$$

其中, $\dfrac{\partial g}{\partial x}([\underline{x},\overline{x}])$ 表示对 $g(x)$ 的偏导直接进行区间乘法运算得到的结果.

基于微分中值定理的区间线性化方法首先通过定理 2 将连续函数转化为区间形式, 再通过命题 1 进行线性化. 与直接通过区间乘法运算得到的结果相比, 该方法得到的区间长度更短, 更加近似精确值. 接下来通过一个例

子说明直接进行区间乘法运算与基于微分中值定理区间化方法的不同之处.

例 1: 假设连续函数 $g(x)=x^3-2x^2+4x, x\in[4,5]$, 通过单调性可知 $g(x)\in[48,95]$. 如果对于函数的每一项直接通过区间乘法运算后可以得到 $g(x)\in[30,113]$. 基于微分中值定理的区间化方法首先得到中点 $\check{x}=4.5$, 再将 $g(x)$ 进行求导得 $g'(x)=3x^2-4x+4, x\in[4,5]$. 将 $g'(x)$ 通过定义 3 进行区间运算得 $g'(x)\in[32,63]$, 所以 $g(x)\in g(4.5)+g'(x)(x-4.5)=[37.125,100.125]$, 区间长度比 $[30,113]$ 小.

通过例 1 的演示, 可以发现, 使用定理 2 进行区间化运算相比于直接对连续函数的每一项进行区间乘法运算, 更加接近于原问题的准确值.

3 动力系统的概率安全验证

在动力系统的安全性验证方面, 一种流行的方法是使用障碍证书. 由于动力系统的安全性验证问题通常具有不确定性, 即使对于具有简单动力学的系统也是如此. 近年来, 对于系统的概率安全性分析问题受到广泛关注, 其中一种方法是通过定量安全性取代定性安全性, 即不要求系统轨迹一定不能到达某个非安全区域. 本节首先介绍通过障碍证书 (barrier certificates, BC) 来验证系统的安全性概念, 属于定性安全性. 然后介绍概率障碍证书验证系统的安全性概念, 属于定量安全性.

3.1 障碍证书

障碍证书是一种提供形式化框架的方法, 通常用一个函数来定义, 具体形式取决于系统的特定要求和约束, 可以是多项式函数、线性不等式或差分不等式. 障碍证书将系统的状态空间划分为两个区域, 分别包含初始状态的前向可达状态集和非安全区域的反向可达状态集. 障碍证书的设计满足一定的条件, 即系统的状态始终保持在定义的安全边界内, 并满足所需的安全属性. 通过验证障碍证书的存在性, 可以确定系统是否能完全避免进入非安全区域, 正式证明系统的安全性. 在没有精确解的情况下, 也能够确保系统满足非线性方程的要求. 障碍证书有多种变体, 对应多种不同的计算方法. 本文将障碍证书定义为多项式形式, 但本文的方法并不局限于多项式形式.

定理 3 (障碍证书)[22]. 给定一个连续动力系统 $C=(f,X_0,\Omega,X)$ 和相应的非安全区域 $X_u\subseteq X$, 若障碍证书 $B:\mathbb{R}^n\to\mathbb{R}$ 存在, 则需要满足以下条件:

BC1. $\forall x\in X_0,\ B(x)\leqslant 0$;

BC2. $\forall x\in X_u,\ B(x)>0$;

BC3. $\forall x\in X,\ \dfrac{\partial B}{\partial x}(x)f(x)+\lambda B(x)\leqslant 0$, 其中, $\lambda\in\mathbb{R}$ 为一个固定的常数.

障碍证书 $B(x)$ 将可达区域中的所有状态映射为非正实数, 将非安全区域中的所有状态映射为正实数. $B(x)$ 的零级集合将初始区域与非安全区域分开, 要求系统状态在连续演化过程中, 当 $B(x)=0$ 时不会将满足 $B(x)\leqslant 0$ 的状态演变为满足 $B(x)>0$ 的状态[36]. 定理 3 将障碍证书视为与时间无关的量来刻画系统的安全性. 若能找到一个满足 BC1–BC3 的 $B(x)$, 则可以说明该系统是安全的.

3.2 概率障碍证书

概率障碍证书用于验证动力系统的概率安全性. Xue 等人通过限制系统出现不安全行为的概率在一定置信度下保持在定量安全的目标范围内, 将满足这一性质的系统称为概率近似安全的. 通过统计推断和有限样本建模, 利用 PAC 障碍证书验证动力系统的概率安全性, 可以更全面地评估复杂动力系统或高维动力系统的安全性. 基于 PAC 学习的保证和场景优化方法将训练数据的大小与安全需求阈值参数 ε 和置信度水平参数 β 相关联. 基于定理 3, 提出了通过学习连续动力系统的 PAC 障碍证书 (probably approximately correct barrier certificates, PBC) 来证明系统的概率近似安全性.

定义 4 (PAC 障碍证书). 给定一个连续动力系统 $C=(f,X_0,\Omega,X)$ 和相应的非安全区域 $X_u\subseteq X$, 若 $B:\mathbb{R}^n\to\mathbb{R}$ 是关于安全需求阈值参数 $\varepsilon\in(0,1)$ 和置信度水平参数 $\beta\in(0,1)$ 的 PAC 障碍证书, 则需满足下述条件:

PBC1. $\forall x\in X_u,\ B(x)>0$;

PBC2. $\forall x \in X,\ \dfrac{\partial B}{\partial x}(x)f(x) + \lambda B(x) \leqslant 0$, 其中 $\lambda \in \mathbb{R}$ 为一个固定的常数;

PBC3. 在置信度至少为 $1-\beta$ 的条件下, $P(\{x \in X_0 \mid B(x) \leqslant 0\}) \geqslant 1-\varepsilon$.

若函数 $B(x)$ 满足 PBC1–PBC3, 则在至少 $1-\beta$ 的置信度下, 可以说明从初始区域 X_0 出发的状态在连续演化过程中不会进入非安全区域的概率至少为 $1-\varepsilon$. PAC 学习框架定义了一个学习算法可以产生 PAC 障碍证书的概率, 即其安全需求阈值参数小于某个给定的阈值 ε, 并且安全需求阈值参数可以通过训练数据的规模来控制. 相较于传统的穷尽搜索或确定性方法, 该方法能够更高效地计算出具有一定置信度的概率障碍证书, 从而减少计算的复杂性. PAC 障碍证书满足以下定理.

定理 4. 若 $B: \mathbb{R}^n \to \mathbb{R}$ 是关于 $\varepsilon \in (0,1)$ 和 $\beta \in (0,1)$ 的 PAC 障碍证书, 则连续动力系统 $C = (f, X_0, \Omega, X)$ 在至少 $1-\beta$ 的置信度下, 满足安全性质的概率大于等于 $1-\varepsilon$.

对于 PAC 障碍证书的求解, 可以将问题转化为在初始区域上通过场景优化方法求解的优化问题. 给定一个连续动力系统 $C = (f, X_0, \Omega, X)$ 和相应的非安全区域 $X_u \subseteq X$, 存在一个函数 $B(x,c): \mathbb{R}^n \to \mathbb{R}$ 和一个常数 λ, c 为障碍证书中每一项的系数, $B(x,c)$ 是关于安全需求阈值参数 $\varepsilon \in (0,1)$ 和置信度水平参数 $\beta \in (0,1)$ 的 PAC 障碍证书, 如果满足以下约束条件:

$$\begin{cases} \min\limits_{c} \displaystyle\int_{X_0} B(x,c)\mathrm{d}x \\ \text{s.t.} \begin{cases} \bigwedge\limits_{i=1,2,\ldots,K} B(x^{(i)},c) \leqslant 0,\ x^{(i)} \in X_0 \\ B(x,c) > 0,\ \forall x \in X_u \\ \dfrac{\partial B}{\partial x}(x,c)f(x) + \lambda B(x,c) \leqslant 0,\ \forall x \in X \end{cases} \end{cases} \tag{5}$$

通过最小化 $B(x,c)$ 在初始区域上的积分使得从初始区域出发的状态尽可能多地满足 $B(x,c) \leqslant 0$, 即一组状态使得轨迹永远不会进入非安全区域 X_u.

定理 5[5]. 若上述优化问题可行并且存在唯一的最优解 c^*, 给定安全需求阈值参数 $\varepsilon \in (0,1)$ 和置信度水平参数 $\beta \in (0,1)$ 的条件下, 随机采样数 K 满足:

$$K \geqslant \frac{2}{\varepsilon}\left(\ln\frac{1}{\beta} + m\right),$$

其中, m 为优化变量的个数. 则在至少 $1-\beta$ 的置信度下 $P(\{x \in X_0 \mid B(x) \leqslant 0\}) \geqslant 1-\varepsilon$, 即动力系统 $C = (f, X_0, \Omega, X)$ 满足安全性质的概率至少为 $1-\varepsilon$.

4 组合式概率障碍证书生成方法

本节通过使用场景优化方法和基于微分中值定理的区间线性化方法将 PAC 障碍证书的求解问题转化为基于大 M 法的混合整数规划问题, 并且提出通过反例制导方法生成一组组合式的 PAC 障碍证书来验证动力系统的概率安全性.

4.1 基于大 M 法的混合整数规划方法

在约束条件 (5) 中, 通过场景优化方法从初始区域 X_0 中随机采样 K 个样本点, 然而, 这些样本点并不一定能够全部满足约束条件. 为了确保能够得到一个 PAC 障碍证书, 引入常量 M 作为罚因子, 将优化问题 (5) 转化为一个混合整数规划问题. 希望不满足约束条件的样本点的数量尽可能少, 即使在某些样本点为反例点的情况下, 依然能够得到优化问题的解. 本文选择障碍证书模板 $B(x,c) = \sum_{\alpha \in \mathcal{M}} c_\alpha x^\alpha$. 具体地, 通过引入一个很大的正常数 M 和 K 个二进制变量 $y^{(i)} \in \{0,1\}, i = 1,2,\ldots,K$ 来调整原始约束条件 (5). 给定一个连续动力系统 $C = (f, X_0, \Omega, X)$ 和相应的非安全区域 $X_u \subseteq X$, 存在一个函数 $B(x,c): \mathbb{R}^n \to \mathbb{R}$ 和一个常数 λ, 在安全需求阈值参数 ε 和置信度水平参数 β 给定的情况下, 通过求解以下优化问题来计算 $B(x,c)$:

$$\begin{cases} \min\limits_{c,y^{(i)},i=1,\ldots,K} \displaystyle\int_{X_0} B(x,c)\mathrm{d}x + \sum_{i=1}^{K} y^{(i)} \\ \text{s.t.} \begin{cases} \bigwedge\limits_{i=1,2,\ldots,K} B(x^{(i)},c) - My^{(i)} \leqslant 0,\ x^{(i)} \in X_0 \\ B(x,c) - \varsigma \geqslant 0,\ \forall x \in X_u \\ \dfrac{\partial B}{\partial x}(x,c)f(x) + \lambda B(x,c) \leqslant 0,\ \forall x \in X \end{cases} \end{cases} \tag{6}$$

(7)

其中, ς 为一个很小的常数, λ 为一个固定的常数, $x^{(1)},x^{(2)},\ldots,x^{(K)}$ 是从初始区域 X_0 中随机采样的 K 个样本点. M 为一个很大的正数, 通过预先定义来确保该约束对目标函数的影响足够大, 从而实现对 $y^{(i)}$ 的取值的限制. 当 $y^{(i)}=0$ 时, 对应的 $x^{(i)}$ 满足约束条件. 当 $y^{(i)}=1$ 时, 约束条件对 $x^{(i)}$ 不产生影响, $x^{(i)}$ 为一个反例点. 通过最小化反例点的数量 $\sum_{i=i}^{K} y^{(i)}$ 来获得具有鲁棒性和可信度的障碍证书, 以提高系统的概率安全性保障.

定理 6. 若 c^* 是上述优化问题的最优解, 在给定安全需求阈值参数 $\varepsilon \in (0,1)$ 和置信度水平参数 $\beta \in (0,1)$ 的条件下, 随机采样数 K 满足:

$$K \geqslant \frac{2}{\varepsilon}\left(\ln\frac{1}{\beta} + m + 1\right),$$

其中, m 为 $B(x,c)$ 中优化变量的个数, 则在至少 $1-\beta$ 的置信度下 $P(\{x \in X_0 \mid B(x,c^*) \leqslant 0\}) \geqslant 1-\varepsilon$, 即动力系统 $C=(f,X_0,\Omega,X)$ 满足安全性质的概率至少为 $1-\varepsilon$.

直接求解上述优化问题是很困难的, 因为求解公式 (7) 是一个 NP 难问题. 通过基于微分中值定理的区间线性化方法, 将公式 (7) 转化为线性不等式, 从而将障碍证书的求解转化为线性优化问题, 便于直接通过数值优化器 Gurobi 进行求解, 为 PAC 障碍证书的计算提供了一个新方法. 在将公式 (7) 通过区间不等式线性化之前, 首先要将所给非安全区域 X_u 和系统状态空间 X 分别用区间 I_{X_u} 和 I_X 表示, 其中 $X_u \subseteq I_{X_u}, X \subseteq I_X$. 先通过定理 2 将非线性不等式转化为区间不等式, 再使用命题 1 将每个优化变量用两个非负变量的差表示, 从而转化为线性不等式. 本文通过实验证明了该区间线性化方法在验证复杂动力系统时, 能在更高的安全概率下成功验证动力系统的概率安全性. 这表明使用定理 2 能有效减小区间长度, 更接近准确解, 从而大大提高区间运算的效率.

4.2 组合式障碍证书

由于将原始优化问题 (5) 转化为基于大 M 法的混合整数规划方法存在一个潜在的问题, 即从初始区域中随机采样的样本点可能存在部分样本点不满足原始优化问题 (5) 的约束条件的情况, 也就是计算出来的障碍证书在某些样本点处的值并不满足在初始区域上非正的条件. 基于此, 本文提出学习一组组合式 PAC 障碍证书 (compositional PAC barrier certificates, CPBC) 来解决动力系统概率安全性验证的问题. 通过场景优化方法和基于大 M 法的混合整数规划方法学习 PAC 障碍证书. 将样本点代入原问题中进行验证, 不满足初始区域上的约束条件的点作为反例点. 本文通过使用反例制导方法学习一组组合式的 PAC 障碍证书, 来提高动力系统概率安全验证的保障. 为了能更具体地说明本文所提方法的应用和优势, 通过例 2 进行阐述.

例 2: 给定一个连续动力系统

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -x_1 + x_1 x_2 \\ -x_2 \end{bmatrix},$$

系统状态空间 $X=\{x \in \mathbb{R}^2 : -2.5 \leqslant x_1, x_2 \leqslant 2.5\}$, 初始区域 $X_0=\{x \in \mathbb{R}^2 : -0.5 \leqslant x_1 \leqslant 0, -1 \leqslant x_2 \leqslant 1\}$, X_0 服从均匀分布, 非安全区域 $X_u=\{x \in \mathbb{R}^2 : 1.2 \leqslant x_1 \leqslant 2, -2 \leqslant x_2 \leqslant -0.17\}$.

若安全需求阈值参数 $\varepsilon=0.1$, 置信度水平参数 $\beta=10^{-12}$, $M=10^4$. 由定理 6 可知在初始区域中随机采样的样本数 $K=813$. 通过基于大 M 法的混合整数规划方法, 在初始区域中随机采样 813 个样本点学习 PAC 障碍证书 $B_0=-12.858+8.958x_1-15.0x_2$.

如图 1 (a) 所示, 绿框和红框分别代表初始区域和非安全区域, B_0 由黑色直线表示, 目标是希望 B_0 将两个区域分隔开. 接下来通过将 K 个样本点代入 B_0, 验证是否满足初始区域上的原始约束条件 $B_0(x) \leqslant 0$. 计算可知存在 5

个反例点 $A(-0.112,-0.993)$、$B(-0.104,-0.928)$、$C(-0.072,-0.922)$、$D(-0.009,-0.963)$、$E(-0.063,-0.974)$. 将初始区域划分为上半部分 X_1 和下半部分 X_2 两个区域, 其中 X_2 包含所有反例点. 从 X_1 中随机采样 K 个点, 学习 PAC 障碍证书 $B_1 = -8.907 + 7.287x_1 - 10.998x_2$, 如图 1 (b) 中的黑色直线所示. 从 X_2 中随机采样 K 个点, 学习 PAC 障碍证书 $B_2 = -0.099 + 3.666x_1 + 1.573x_2$, 如图 1 (b) 中的蓝色直线所示.

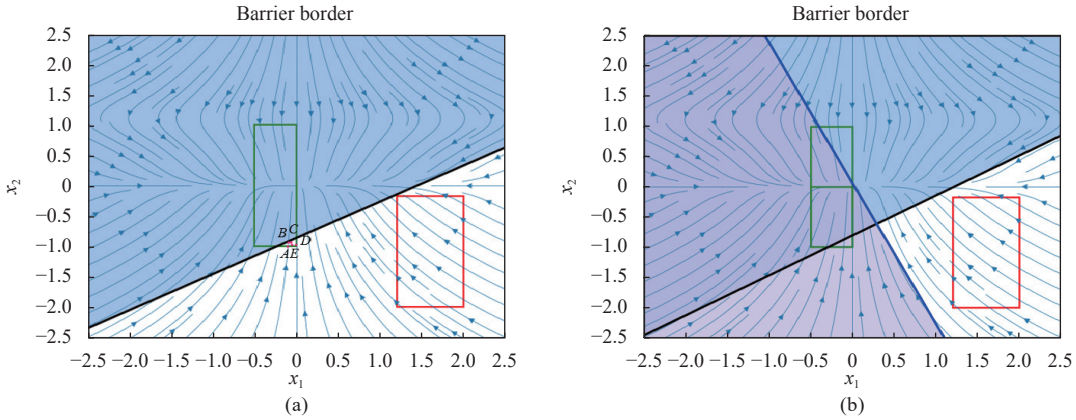

图 1 组合式障碍证书效果图

将 X_1 和 X_2 中的采样点分别代入 B_1 和 B_2 验证可知满足原始约束条件 $B_1(x) \leqslant 0$ 和 $B_2(x) \leqslant 0$, 分别可以说明在至少 $1-10^{-12}$ 的置信度下, 从 X_1 出发的状态在连续演化过程中满足安全性质的概率至少为 0.9, 从 X_2 出发的状态在连续演化过程中满足安全性质的概率至少为 0.9. B_1 和 B_2 一起将初始区域 X_0 与非安全区域 X_u 分隔开. 因此可以表明在至少 $1-10^{-12}$ 的置信度条件下, 该动力系统满足安全性的概率至少为 0.9. B_1 和 B_2 称为该系统的一组 PAC 障碍证书, 共同验证系统的概率安全性.

基于以上例子所带来的启发, 本文提出学习一组组合式 PAC 障碍证书的方法, 以解决样本点中可能存在反例点的问题.

定义 5 (组合式 PAC 障碍证书). 给定一个连续动力系统 $C = (f, X_0, \Omega, X)$ 和相应的非安全区域 $X_u \subseteq X$, 若 $B_1, B_2, \ldots, B_n$ 是关于安全需求阈值参数 $\varepsilon \in (0,1)$ 和置信度水平参数 $\beta \in (0,1)$ 的一组组合式 PAC 障碍证书, 则需满足下述条件:

CPBC1. $\forall x \in X_u, B_i(x) > 0,\ i = 1,2,\ldots,n$;

CPBC2. $\forall x \in X, \dfrac{\partial B_i}{\partial x}(x) f(x) + \lambda B_i(x) \leqslant 0,\ i = 1,2,\ldots,n$, 其中 $\lambda \in \mathbb{R}$ 为一个固定的常数;

CPBC3. $\forall x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(K)} \in X_i$, $\bigwedge\limits_{j=1,2,\ldots K} B_i(x_i^{(j)}) \leqslant 0$, $i = 1,\ldots,n$ 为真, 其中 $\{X_1,\ldots,X_n\}$ 是 X_0 的划分, K 是从子块 X_i 中随机采样的样本数且满足公式 (8).

对于所得到的一组组合式 PAC 障碍证书 $B_1, B_2, \ldots, B_n$, 只需每个 X_i 中 K 个样本点对于 $B_i, i = 1,2,\ldots,n$ 都有能满足的安全性质. 由于样本数由安全需求阈值参数 $\varepsilon \in (0,1)$ 和置信度水平参数 $\beta \in (0,1)$ 决定, 并且对于每个 PAC 障碍证书所学习的样本点不同, 因此系统的安全概率为 $1-\varepsilon$. 组合式 PAC 障碍证书的安全概率由以下定理可得.

定理 7 (组合式 PAC 障碍证书概率安全性). 假设 $B_1, B_2, \ldots, B_n$ 为一组组合式 PAC 障碍证书, 满足 CPBC1–3, 则在至少 $1-\beta$ 的置信度下, 可以说明系统 $C = (f, X_0, \Omega, X)$ 在至少 $1-\varepsilon$ 的概率下满足安全性质.

如图 2 所示为组合式 PAC 障碍证书的构造框架, 具体流程如下:

(1) 令 $\pi^{(1)} = \{X_0\}$ 是初始区域 X_0 的一个划分; 组合式 PAC 障碍证书集合 CPBC=$\{\}$; $i = 1$; $L = 100$;

(2) 对于 $\pi^{(i)} = \{X_1^{(i)}, \ldots, X_{n_i}^{(i)}\}$ 中每个子块, n_i 为当前 $\pi^{(i)}$ 中包含的子块个数, 随机采样 K 个点, 并利用大 M 法分别计算相应子块的 PAC 障碍证书 $B_1^{(i)}, \ldots, B_{n_i}^{(i)}$;

(3) 令 $T = \{\}$; 分别对每个子块 $X_j^{(i)}$ 中的 K 个样本点检验是否满足条件 $B_j^{(i)}(x) \leqslant 0$, $j = 1,2,\ldots,n_i$, 若均满足则将

$B_j^{(i)}(x)$ 存入集合 CPBC; 否则 $X_j^{(i)}$ 中存在反例点, 将 $X_j^{(i)}$ 存入到 T 中;

(4) 若 $T=\{\}$, 则返回组合式 PAC 障碍证书集合 CPBC; 否则, 令 $i=i+1$, 构造 $\pi^{(i)}$ 为 T 的加细划分, 即将 T 中每一个元素 X_j 划分成两个子块, 使得其中一个子块包含 X_j 的 K 个样本点中所有的反例点, 另一子块包含非反例点;

(5) 若 $i=L$ 则结束, 否则重复步骤 (2)–(4).

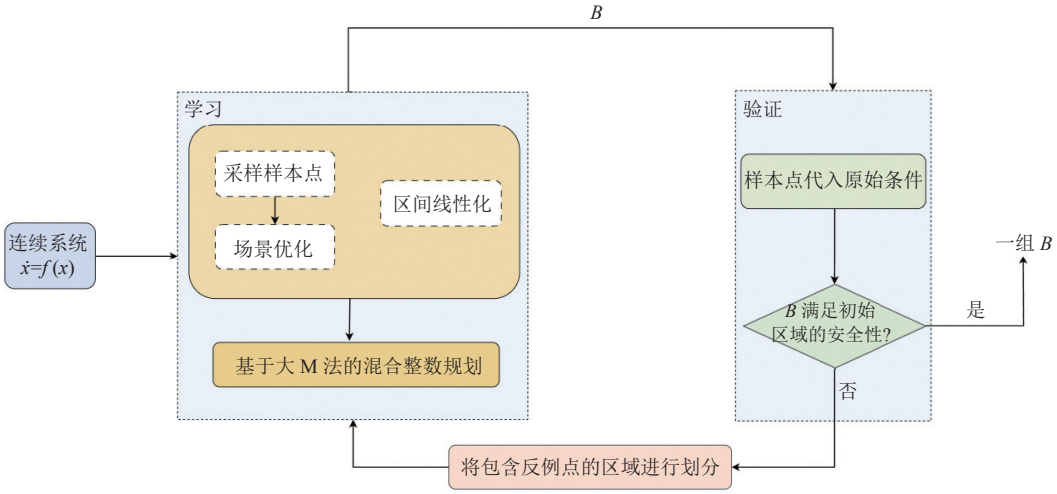

图 2 组合式 PAC 障碍证书构造框架

由于优化问题采用基于大 M 法的混合整数规划方法, 在约束中引入了新的变量, 因此可能存在反例点. 通过将包含反例点的区域进行划分后再分别对每个区域计算 PAC 障碍证书, 能够更好地逼近原问题的约束, 从而生成一组组合式 PAC 障碍证书来共同验证动力系统的概率安全性, 提高验证的准确性. 这种反例制导方法允许在优化过程中纠正由于样本和概率不完备性导致的误差, 并且能够改善验证结果.

在应用场景方面, 生成一组组合式 PAC 障碍证书的方法适用于各种安全攸关系统, 如机器人控制系统、自动驾驶等, 从而验证系统的概率安全性. 例如, 在给定安全需求阈值参数和置信度水平参数的条件下, 对机器人控制系统进行验证, 以确保机器人在复杂环境中执行任务时不会出现意外行为. 对自动驾驶系统验证其在不同交通条件和道路情况下的概率安全性, 确保驾驶的可靠性和安全性.

因此, 本文提出的求解组合式障碍证书的方法为验证动力系统的概率安全性提供了一种有效的、全面的、可靠的方式, 在许多复杂系统的概率安全性验证中具有重要作用, 并在自动化、机器学习等领域中有广泛的应用潜力.

5 算法与用例展示

本节首先介绍使用混合整数规划方法生成一组组合式 PAC 障碍证书的算法, 再通过一个三维实例进行详细介绍.

5.1 算 法

算法 1 展示了本文所提出的组合式 PAC 障碍证书的生成算法. 对于一个给定的连续动力系统和非安全区域, 首先定义安全验证的安全需求阈值参数 ε 和置信度水平参数 β, 给定一个很大的正常数 M, 将划分 $\pi^{(i)}$ 初始化为 X_0, 将临时存储集合 τ 初始化为空集, 将组合式障碍证书集合 CPBC 初始化为空集, 将变量 i 初始化为 1, L 初始化为 100, K 为满足定理 6 的随机采样数. 当 i 小于 L 时, 开始循环 (第 1 行). 对于 $\pi^{(i)}$ 中每一个划分块 $\bar{U}$ (第 2 行), 随机采样 K 个点得到样本点集 $\bar{u}$ (第 3 行), 并将 (7) 通过定理 2 使非线性不等式区间化, 再由命题 1 将区间不等式转化为线性不等式, 将问题转化为第 4.1 节所提出的基于大 M 法的混合整数规划方法, 可以直接带入数值优化器 Gurobi 进行求解, 得到 PAC 障碍证书 B (第 4 行). 将 $\bar{u}$ 中样本点带回初始区域上的原始约束条件 (5), 验证是否满足 B 的值在样本点上非正的安全性条件, 若满足则将 B 加入 CPBC 集合中 (第 5、6 行), 若不满足则将该划分块 $\bar{U}$

分割成 $\bar{U}_1$ 和 $\bar{U}_2$ 两块, 其中一块包含 $\bar{u}$ 中所有反例点, 将 $\bar{U}_1$ 和 $\bar{U}_2$ 加入临时集合 τ 中 (第 7–9 行). 若 τ 是空集, 表示现在没有待求解的初始区域的划分块, 返回的 CPBC 即为一组 PAC 障碍证书 (第 10、11 行). 否则, 将 i 自增, τ 赋值给 $\pi^{(i)}$ (第 12–14 行), 然后对 $\pi^{(i)}$ 中每个划分块进行新一轮的遍历. 若 i 的值为 100, 则停止循环, 表示无法证明在至少 $1-\beta$ 的置信度下, 该系统在至少 $1-\varepsilon$ 的概率下满足安全性质.

算法 1. 组合式 PAC 障碍证书生成算法 CPBC.

输入: 连续动力系统 $C=(f,X_0,\Omega,X)$, 非安全集 X_u, 安全需求阈值参数 ε, 置信度水平参数 β, 大数 M, $\pi^{(i)}=\{X_0\}$, $\tau=\{\}$, CPBC = {}, $i=1$, $L=100$, $K\geqslant\dfrac{2}{\varepsilon}\left(\ln\dfrac{1}{\beta}+m+1\right)$;

输出: 一组 PAC 障碍证书 CPBC.

1. **while** $i<L$ **do**
2. **for** $\pi^{(i)}$ 中每一个划分块 $\bar{U}$
3. 采样 $\bar{U}$ 中 K 个点得到样本点集 $\bar{u}$
4. 通过基于大 M 法的混合整数规划方法计算 PAC 障碍证书 B
5. **if** $\bar{u}$ 中每个样本点 x 都满足 $B(x)\leqslant 0$
6. 将 B 加入 CPBC
7. **else**
8. 将 $\bar{U}$ 划分为 $\bar{U}_1$ 和 $\bar{U}_2$, 分别包含 $\bar{u}$ 中所有反例点和非反例点
9. 将 $\bar{U}_1$, $\bar{U}_2$ 加入 τ
10. **if** τ 是空集
11. **return** CPBC
12. **else**
13. $i=i+1$
14. $\pi^{(i)}=\tau$

5.2 用例展示

本节通过一个具体的三维例子来展示在给定一个连续动力系统 $C=(f,X_0,\Omega,X)$ 和相应的非安全区域 $X_u\subseteq X$ 的情况下, 如何得到该系统的组合式 PAC 障碍证书.

例 3[37]: 给定一个连续动力系统

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}=\begin{bmatrix}-x_2\\ -x_3\\ -x_1-2x_2-x_3+x_1^3\end{bmatrix},$$

已知系统的状态空间为 $X=\{x\in\mathbb{R}^3:-2\leqslant x_1,x_2,x_3\leqslant 2\}$, 初始区域为 $X_0=\{x\in\mathbb{R}^3:0\leqslant x_1,x_2,x_3\leqslant 1\}$, 非安全区域为 $X_u=\{x\in\mathbb{R}^3:1.5\leqslant x_1\leqslant 2,0\leqslant x_2,x_3\leqslant 1\}$, Ω 为均匀分布.

设置障碍证书 $B(x,c)=c_0+c_1x_1+c_2x_2+c_3x_3$, 由命题 1 可知 $c_i=c_{i1}-c_{i2}$, $i=0,1,2,3$. 设置安全需求阈值参数 $\varepsilon=0.1$, 置信度水平参数 $\beta=10^{-12}$, 大数 $M=10^4$. 由定理 6 可知从初始区域中随机采样的样本数 $K\geqslant 732.624$, 取 $K=733$, $y^{(i)}\in\{0,1\}$ 为二进制变量, $i=1,2,\ldots,K$.

在 X_0 上有线性不等式约束:

$$\bigwedge_{i=1,2,\ldots,K}B(x^{(i)},c)-My^{(i)}\leqslant 0,\ x^{(i)}\in X_0.$$

在 X_u 上通过定理 2 将不等式 $B(x)-\varsigma\geqslant 0$ 区间化得:

$$c_0+1.75c_1+0.5c_2+0.5c_3+c_1[-0.25,0.25]+c_2[-0.5,0.5]+c_3[-0.5,0.5]-\varsigma\geqslant 0.$$

再线性化得:

$$c_{01}-c_{02}+1.5c_{11}-2c_{12}-c_{22}-c_{32}-\varsigma\geqslant 0.$$

在 X 上通过定理 2 将不等式 $\dfrac{\partial B}{\partial x}(x)f(x)+\lambda B(x)\leqslant 0$ 区间化得:

$$\lambda c_0+(\lambda+1)c_1[-2,2]+(\lambda+1)c_2[-2,2]+(\lambda+16)c_3[-2,2]\leqslant 0.$$

再线性化得:

$$\lambda c_{01}-\lambda c_{02}+2(\lambda+1)c_{11}+2(\lambda+1)c_{12}+2(\lambda+1)c_{21}+2(\lambda+1)c_{22}+2(\lambda+16)c_{31}+2(\lambda+16)c_{32}\leqslant 0.$$

目标是最小化 $\int_{X_0}B(x,c)\mathrm{d}x$ 和不满足约束条件的点的个数, 令 $\lambda=0.132$, $\varsigma=0.05$, 令第 1 轮划分 $\pi^{(1)}=\{X_1^{(1)}\}$, 其中 $X_1^{(1)}=X_0$. 从 $X_1^{(1)}$ 中随机采样 K 个样本点, 通过大 M 法学习 PAC 障碍证书:

$$B_1^{(1)}=-13.709+11.603x_1+2.989x_2-3.446x_3,$$

验证样本点是否满足 $B_1^{(1)}(x)\leqslant 0$, 得到 $(0.999,0.841,0.013)$ 、 $(0.983,0.936,0.142)$ 、 $(0.963,0.984,0.024)$ 这 3 个反例点. 将 $X_1^{(1)}$ 存入到临时存储空集 T 中, 并构造 T 的加细划分, 赋给 $\pi^{(2)}$, 此时 $\pi^{(2)}=\{X_1^{(2)},X_2^{(2)}\}$, 其中 $X_1^{(2)}=\{x\in\mathbb{R}^3:0\leqslant x_1\leqslant 0.5,0\leqslant x_2,x_3\leqslant 1\}$, $X_2^{(2)}=\{x\in\mathbb{R}^3:0.5\leqslant x_1\leqslant 1,0\leqslant x_2,x_3\leqslant 1\}$.

从 $X_1^{(2)}$ 中随机采样 K 个样本点, 通过大 M 法学习 PAC 障碍证书:

$$B_1^{(2)}=-9.749+12.389x_1+3.717x_2-4.284x_3,$$

验证样本点是否满足约束条件 $B_1^{(2)}(x)\leqslant 0$, 可知不存在反例点, 此时组合式 PAC 障碍证书集合 $\mathrm{CPBC}=\{B_1^{(2)}\}$.

从 $X_2^{(2)}$ 中随机采样 K 个样本点, 通过大 M 法学习 PAC 障碍证书:

$$B_2^{(2)}=-6.103+6.835x_1+8.882\times 10^{-16}x_2,$$

验证样本点是否满足 $B_2^{(2)}(x)\leqslant 0$, 得到 $(0.946,0.957,0.980)$ 、 $(0.903,0.782,0.762)$ 、 $(0.992,0.267,0.109)$ 这 3 个反例点. 将 $X_2^{(2)}$ 存入到临时存储空集 T 中, 并构造 T 的加细划分, 赋给 $\pi^{(3)}$, 此时 $\pi^{(3)}=\{X_1^{(3)},X_2^{(3)}\}$, 其中 $X_1^{(3)}=\{x\in\mathbb{R}^3:0.5\leqslant x_1\leqslant 1,0\leqslant x_2\leqslant 0.5,0\leqslant x_3\leqslant 1\}$, $X_2^{(3)}=\{x\in\mathbb{R}^3:0.5\leqslant x_1,x_2\leqslant 1,0\leqslant x_3\leqslant 1\}$.

从 $X_1^{(3)}$ 中随机采样 K 个样本点, 通过大 M 法学习 PAC 障碍证书:

$$B_1^{(3)}=-3.216+2.564x_1+0.747x_2-0.860x_3,$$

验证样本点是否满足约束条件 $B_1^{(3)}(x)\leqslant 0$, 可知不存在反例点, 此时组合式 PAC 障碍证书集合 $\mathrm{CPBC}=\{B_1^{(2)},B_1^{(3)}\}$.

从 $X_2^{(3)}$ 中随机采样 K 个样本点, 通过大 M 法学习 PAC 障碍证书:

$$B_2^{(3)}=-3.223+2.415x_1+0.264x_2-0.429x_3,$$

验证样本点是否满足 $B_2^{(3)}(x)\leqslant 0$, 可知不存在反例点, 此时组合式 PAC 障碍证书集合 $\mathrm{CPBC}=\{B_1^{(2)},B_1^{(3)},B_2^{(3)}\}$. 由于此时没有将 T 赋值, 所以 T 为空集, 因此组合式 PAC 障碍证书为 $B_1^{(2)},B_1^{(3)},B_2^{(3)}$. 可以验证在置信度至少为 $1-10^{-12}$ 的条件下, 系统满足安全性质的概率至少为 0.9.

6 实验结果及分析

在本节中, 我们对提出的 CPBC 算法进行了评估, 以实现对 11 个非线性连续动力系统进行概率安全验证. 同时比较了 CPBC 算法和 Xue 等人提出的 PACBC 方法[5], 并采用基于微分中值定理的区间化方法 PACBC-DM 进行了实验. 所有实验都是在 Ubuntu 系统下使用 3.30 GHz Intel(R) Xeon(R) Gold 6246 处理器、NVIDIA GeForce RTX 2080 Ti CPU 和 64 GB RAM 的机器上运行的. 本文使用数值优化器 Gurobi 求解混合整数规划问题, 实验结果见表 1.

针对 11 个基准样例进行了对比实验, 其中包括高达 200 维的样例, 以评估基于微分中值定理区间化方法的优势以及 CPBC 与 PACBC 方法之间的效果差异. 对于 15 维以上的高维样例, 从 Ratschan 的研究中选取了一个可扩展性样例[38], 并针对几个有代表性的样例进行了深入分析. PACBC 方法通过在初始区域的约束中引入松弛变量, 并在目标函数中最小化这一变量来求解障碍证书. 在将非线性不等式线性化时, PACBC 直接采用区间乘法运算.

相比之下, PACBC-DM 在将非线性不等式线性化时采用基于微分中值定理的区间线性化方法, 其余方面与 PACBC 相同. 此外, CPBC 采用基于大 M 法的混合整数规划方法来求解障碍证书, 并使用基于微分中值定理的区间线性化方法处理非线性约束条件. 对于一些复杂的动力系统, 本文通过求解 1 组组合式的 PAC 障碍证书, 以验证其概率安全性.

如表 1 所示, n_x 为动力系统中状态变量的个数, d_f 表示向量场中多项式的最高次数, ε 表示安全需求阈值参数, 对于所有样例, 置信度水平参数 $\beta = 10^{-12}$. T 表示计算时间 (以 s 为单位), d_B 表示障碍证书的次数. 允许最大运行时间为 20 000 s, 画横杠表示该方法在用例上失败, 即不能在最大运行时间内生成 PAC 障碍证书来验证动力系统的概率安全性.

表 1　实验评估以及相关工作对比

Ex	n_x	d_f	ε	PACBC-DM		PACBC		CPBC		
				T (s)	d_B	T (s)	d_B	T (s)	d_B	N
C_1 [39]	2	2	0.2	1.50	1	3.71	2	1.17	1	1
			0.1	2.89	1	7.13	2	2.14	1	1
			0.05	5.51	1	14.21	2	4.17	1	1
C_2 [38]	3	3	0.2	2.16	1	2.19	1	1.79	1	1
			0.1	4.02	1	16.35	2	3.33	1	1
			0.05	7.78	1	32.63	2	6.54	1	1
C_3 [40]	7	2	0.2	5.54	1	75.54	2	5.05	1	3
			0.1	14.30	1	166.14	2	9.58	1	4
			0.05	120.74	1	315.03	2	20.33	1	6
C_4 [40]	9	2	0.2	23.00	1	186.58	1	6.74	1	1
			0.1	30.27	1	190.07	1	13.31	1	1
			0.05	116.40	1	207.56	1	26.27	1	1
C_5 [41]	12	1	0.2	118.93	1	668.79	2	10.30	1	1
			0.1	129.20	1	1 333.60	2	19.85	1	1
			0.05	150.67	1	2 639.76	2	40.05	1	1
C_6 [38]	15	5	0.2	15.74	1	—	—	14.51	1	1
			0.1	550.07	1	—	—	28.17	1	1
			0.05	588.67	1	—	—	56.47	1	3
C_7 [38]	17	5	0.2	21.77	1	258.19	1	18.40	1	2
			0.1	39.68	1	—	—	35.64	1	2
			0.05	418.40	1	—	—	71.21	1	1
C_8 [38]	19	5	0.2	23.25	1	312.65	1	21.53	1	1
			0.1	45.96	1	—	—	41.93	1	1
			0.05	93.95	1	—	—	83.94	1	3
C_9 [38]	101	5	0.2	—	—	—	—	789.11	1	1
			0.1	—	—	—	—	1 574.74	1	1
			0.05	—	—	—	—	3 136.34	1	1
C_{10} [5]	101	5	0.2	884.38	1	864.17	1	824.72	1	1
			0.1	1 735.06	1	1 704.87	1	1 616.99	1	1
			0.05	3 373.56	1	3 362.61	1	3 185.07	1	1
C_{11} [38]	201	5	0.2	—	—	—	—	4 012.86	1	1
			0.1	—	—	—	—	7 894.78	1	1
			0.05	—	—	—	—	15 714.11	1	1

为了充分展示使用微分中值定理对非线性不等式进行区间化的优势, 我们开展了一系列实验, 并在 PACBC-DM 中呈现这些实验结果. 在对 11 个样例进行比较后, 我们发现在多数情况下 PACBC-DM 表现优于 PACBC. 这主要是因为 PACBC 在将非线性不等式线性化时直接采用区间乘法运算, 导致较大的区间扩张, 这在某些情况下使得

优化问题难以求解. 相反, 基于微分中值定理的区间化方法提供了更精确的区间估计. 在 PACBC 无法找到解的情况下, 通常需要划分区间或增加障碍证书模板的次数来寻找可行解, 但这可能导致求解复杂度呈指数型增加. 相比之下, PACBC-DM 无需进行额外区域划分就能找到可行解, 显著提高了计算效率.

在样例 C_1-C_5 和 C_{10} 中, CPBC 和 PACBC 都能成功验证动力系统的概率安全性, 适应不同的安全需求阈值要求. 从表 1 中的 $T(s)$ 列数据可以看出, 就时间效率而言, CPBC 在这些样例中的表现均优于 PACBC. 例如, 对于样例 C_5, 在安全阈值 $1-\varepsilon$ 为 0.95 的情况下, PACBC 耗时 2 639.76 s 以验证系统的概率安全性, 而 CPBC 仅需 40.05 s, 效率是 PACBC 的 65 倍. 对于样例 C_3, 在安全阈值为 0.95 的情况下, PACBC 需耗时 315.03 s 在 2 次障碍证书模板下验证系统的概率安全性, 而 CPBC 只需 20.33 s 即可生成组合形式的 1 次 PAC 障碍证书, 效率是 PACBC 的 15 倍. 样例 C_3 对于不同要求的安全阈值均生成多个 PAC 障碍证书来验证系统的概率安全性, 实验效率明显高于 PACBC, 表明多个 1 次的 PAC 障碍证书具有表达力的同时生成效率更高. 与 PACBC 不同的是, CPBC 在将非线性不等式线性化时采用基于微分中值定理的方法, 有效改善了区间扩张现象, 提高了计算结果的精确性. 在 PACBC 未能获得可行解的情况下, 不得不对非安全区域和系统状态域进行区域划分以解决由于区间扩张导致的问题, 但这也增加了求解时间的开销. 在实验中, 对于 PACBC, 首先尝试在不划分区域的情况下在 1 次障碍证书模板下求解, 若无解则对区域进行划分. 考虑到求解时间随划分数的指数型增长, 每个维度的最大划分数设为 4, 通过逐步提高区域划分的细度来迭代求解. 若在 1 次障碍证书条件下仍然无解, 则采用 2 次障碍证书进行计算.

在样例 C_6-C_9 中, 观察到 CPBC 能成功验证所有样例在不同安全阈值下的概率安全性. 与此相比, PACBC 只能在样例 C_7 和 C_8 中, 在安全阈值为 0.8 时成功验证系统的概率安全性, 而 CPBC 能在安全阈值为 0.95 的情况下成功验证, 且时间效率至少快 14 倍. CPBC 通过求得组合式 PAC 障碍证书, 迅速验证了系统的概率安全性. 对于 PACBC 验证失败的情况, 我们进行了一系列探索和尝试: 首先, 在 1 次障碍证书模板下, 对非安全区域和系统状态域的各个维度进行最多 4 次划分并求解, 以缓解由区间扩张导致的无可行解问题. 如果仍无解, 则采用 2 次障碍证书进行验证, 在必要时对非安全区域和系统状态域的各个维度进行划分后求解. 如果在最大运行时间范围内仍无解, 则用横杠表示验证失败. 这些实验结果表明, PACBC 的验证过程高度依赖于障碍证书的复杂表达力, 主要受非线性不等式在区间化时直接通过区间乘法运算带来的误差影响. 相比之下, CPBC 通过使用基于微分中值定理的区间化方法, 无需对区域进行划分即可成功验证样例的概率安全性. 这不仅体现了其验证的有效性, 还表现出其在复杂动力系统概率安全验证方面的广泛适用性.

本文还针对两个 101 维的样例进行了对比实验, 样例 C_9 源自 Ratschan 等人[38]的工作, 样例 C_{10} [5]则是 Xue 等人对 Ratschan 样例的修改, 其中系统状态空间每个维度被限制在 $[-0.3, 0.3]$ 的范围内. 对于样例 C_9, 其系统状态空间的范围较广 ($[-10, 10]$), 这在区间线性化的过程中凸显了 PACBC 方法直接采用区间乘法运算引入的误差问题. 如果对每个维度进行区域划分处理, 我们发现即使是在最大允许的运行时间内, 也无法求解. 我们尝试将区域随机划分为 1 024、2 048 和 4 096 个区域进行运算, 但在所有这些尝试中, 都未能在规定的最大运行时间内求解. 如果不进行区域划分, 由于区间扩张的问题, 同样无法找到可行解. 相比之下, CPBC 通过成功求解单个 1 次障碍证书, 有效地进行了动力系统的概率安全验证, 这进一步证明了, 基于微分中值定理进行的区间线性化方法, 相比直接区间相乘方法, 在优化问题求解上能更接近原始问题的解. 对于样例 C_{10}, 由于其较小状态空间范围, PACBC 在区间相乘操作中的劣势不太明显, 从而在区间线性化时能够有效地求解. 然而, 实验结果显示, CPBC 在效率方面仍然表现得更优越. 此外, 本文还进一步探索了一个 201 维的样例, 我们发现, 在 PACBC 无法验证动力系统的概率安全性的情况下, CPBC 在不进行区域划分的前提下, 成功地验证了动力系统的概率安全性.

综上所述, 基于 CPBC 算法框架的 PAC 障碍证书生成方法在实验中展现了出色的性能. 对于 PACBC 无法验证的情况, CPBC 通过生成组合式 PAC 障碍证书, 更高效地实现了动力系统的概率安全验证. 而基于 PACBC 算法框架的方法在一些复杂的动力系统下表现出较差的效果, 尤其是在高维动力系统中. 因此, 本文提出的 CPBC 算法在提升动力系统的概率安全保障方面具有显著潜力和优势, 为高维动力系统的概率安全验证提供了一种有效的解决方案.

7 总 结

本文提出基于反例制导方法学习一组组合式 PAC 障碍证书的算法, 以解决非线性连续动力系统概率安全验证中的不确定性和高维状态空间等挑战. 将求解 PAC 障碍证书的优化问题转化为基于大 M 法的混合整数规划问题. 对于在 3 个区域上的不等式约束, 通过使用场景优化方法和基于微分中值定理的区间线性化方法, 将其转化为线性优化问题, 并通过数值优化器 Gurobi 进行求解. 最后对 11 个非线性连续动力系统进行对比实验. 实验结果表明, 本文提出的 CPBC 算法在不需要对区域进行划分的情况下, 成功验证了所有用例的概率安全性, 同时证明了基于微分中值定理的区间线性化方法能更接近原始非线性约束, 从而得到最优解. 综合实验结果, 本文为实际系统的安全性评估提供了实用且可靠的方法, 有望推动非线性连续动力系统概率安全性验证领域的发展.

在区间线性化过程中, 可以将区间分割成多个部分, 并使用中值定理来提高精度, 从而有望减少组合式障碍证书的数量, 但会显著增加求解时间. 本文所提方法不使用区域划分, 而是通过构造组合式障碍证书来平衡精度问题和前期计算的要求. 未来, 我们将探索如何平衡区域划分带来的 PAC 障碍证书数量与求解时间, 以进一步提高算法的效率和实用性.

References:

[1] Chen B, Li TF. Formal modeling and verification of autonomous driving scenario. In: Proc. of the 2021 IEEE Int'l Conf. on Information Communication and Software Engineering (ICICSE). Chengdu: IEEE, 2021. 313–321. [doi: 10.1109/ICICSE52190.2021.9404128]

[2] Jiang Y, Song HB, Wang R, Gu M, Sun JG, Sha L. Data-centered runtime verification of wireless medical cyber-physical system. IEEE Trans. on Industrial Informatics, 2017, 13(4): 1900–1909. [doi: 10.1109/TII.2016.2573762]

[3] Wang Q, Turriate V, Burgos R, Boroyevich D, Sagona J, Kheraluwala M. Towards a high performance motor drive for aerospace applications: Topology evaluation, converter optimization and hardware verification. In: Proc. of the 43rd Annual Conf. of the IEEE Industrial Electronics Society (IECON 2017). Beijing: IEEE, 2017. 1622–1628. [doi: 10.1109/IECON.2017.8216275]

[4] Haussler D. Probably approximately correct learning. In: Proc. of the 8th National Conf. on Artificial Intelligence. Boston: ACM, 1990. 1101–1108. [doi: 10.5555/1865609.1865663]

[5] Xue B, Fränzle M, Zhao HJ, Zhan NJ, Easwaran A. Probably approximate safety verification of hybrid dynamical systems. In: Proc. of the 21st Int'l Conf. on Formal Engineering Methods. Shenzhen: Springer, 2019. 236–252. [doi: 10.1007/978-3-030-32409-4_15]

[6] Campi MC, Garatti S, Prandini M. The scenario approach for systems and control design. Annual Reviews in Control, 2009, 33(2): 149–157. [doi: 10.1016/j.arcontrol.2009.07.001]

[7] Rohn J, Kreslová J. Linear interval inequalities. Linear and Multilinear Algebra, 1994, 38(1–2): 79–82. [doi: 10.1080/03081089508818341]

[8] Hladík M. Interval linear programming: A survey. In: Mann ZA, ed. Linear Programming—New Frontiers in Theory and Applications. New York: Nova Science Publishers, 2012. 85–120.

[9] Hladík M. Optimal value range in interval linear programming. Fuzzy Optimization and Decision Making, 2009, 8(3): 283–294. [doi: 10.1007/s10700-009-9060-7]

[10] Hladík M. Weak and strong solvability of interval linear systems of equations and inequalities. Linear Algebra and its Applications, 2013, 438(11): 4156–4165. [doi: 10.1016/j.laa.2013.02.012]

[11] Alefeld G, Mayer G. Interval analysis: Theory and applications. Journal of Computational and Applied Mathematics, 2000, 121(1–2): 421–464. [doi: 10.1016/S0377-0427(00)00342-3]

[12] Gan T, Xia BC. Barrier certificate generation for safety verification of continuous systems for a bounded time. Ruan Jian Xue Bao/Journal of Software, 2016, 27(3): 645–654 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4986.htm [doi: 10.13328/j.cnki.jos.004986]

[13] Gulwani S, Tiwari A. Constraint-based approach for analysis of hybrid systems. In: Proc. of the 20th Int'l Conf. on Computer Aided Verification. Princeton: Springer, 2008. 190–203. [doi: 10.1007/978-3-540-70545-1_18]

[14] Kapinski J, Deshmukh JV, Sankaranarayanan S, Aréchiga N. Simulation-guided Lyapunov analysis for hybrid dynamical systems. In: Proc. of the 17th Int'l Conf. on Hybrid Systems: Computation and Control. Berlin: ACM, 2014. 133–142. [doi: 10.1145/2562059.2562139]

[15] Sankaranarayanan S, Sipma HB, Manna Z. Constructing invariants for hybrid systems. In: Proc. of the 7th Int'l Workshop on Hybrid Systems: Computation and Control. Philadelphia: Springer, 2004. 539–554. [doi: 10.1007/978-3-540-24743-2_36]

[16] Tiwari A. Approximate reachability for linear systems. In: Proc. of the 6th Int'l Workshop on Hybrid Systems: Computation and Control. Prague: Springer, 2003. 514–525. [doi: 10.1007/3-540-36580-X_37]

[17] Prajna S, Jadbabaie A, Pappas GJ. A framework for worst-case and stochastic safety verification using barrier certificates. IEEE Trans. on Automatic Control, 2007, 52(8): 1415–1428. [doi: 10.1109/TAC.2007.902736]

[18] Prajna S, Jadbabaie A. Safety verification of hybrid systems using barrier certificates. In: Proc. of the 7th Int'l Workshop on Hybrid Systems: Computation and Control. Philadelphia: Springer, 2004. 477–492. [doi: 10.1007/978-3-540-24743-2_32]

[19] Sloth C, Pappas GJ, Wisniewski R. Compositional safety analysis using barrier certificates. In: Proc. of the 15th ACM Int'l Conf. on Hybrid Systems: Computation and Control. Beijing: ACM, 2012. 15–24. [doi: 10.1145/2185632.2185639]

[20] Kapinski J, Deshmukh J. Discovering forward invariant sets for nonlinear dynamical systems. In: Proc. of the 2015 Interdisciplinary Topics in Applied Mathematics, Modeling and Computational Science. Cham: Springer, 2015. 259–264. [doi: 10.1007/978-3-319-12307-3_37]

[21] Platzer A, Clarke EM. Computing differential invariants of hybrid systems as fixedpoints. In: Proc. of the 20th Int'l Conf. on Computer Aided Verification. Princeton: Springer, 2008. 176–189. [doi: 10.1007/978-3-540-70545-1_17]

[22] Sogokon A, Ghorbal K, Tan YK, Platzer A. Vector barrier certificates and comparison systems. In: Proc. of the 22nd Int'l Symp. on Formal Methods. Oxford: Springer, 2018. 418–437. [doi: 10.1007/978-3-319-95582-7_25]

[23] Kong H, Song XY, Han D, Gu M, Sun JG. A new barrier certificate for safety verification of hybrid systems. The Computer Journal, 2014, 57(7): 1033–1045. [doi: 10.1093/comjnl/bxt059]

[24] Liu J, Zhan NJ, Zhao HJ. Computing semi-algebraic invariants for polynomial dynamical systems. In: Proc. of the 9th ACM Int'l Conf. on Embedded Software. Taipei: ACM, 2011. 97–106. [doi: 10.1145/2038642.2038659]

[25] Mitchell IM, Bayen AM, Tomlin CJ. A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. IEEE Trans. on Automatic Control, 2005, 50(7): 947–957. [doi: 10.1109/TAC.2005.851439]

[26] Zeng X, Lin W, Yang ZF, Chen X, Wang LL. Darboux-type barrier certificates for safety verification of nonlinear hybrid systems. In: Proc. of the 13th Int'l Conf. on Embedded Software. Pittsburgh: ACM, 2016. 11. [doi: 10.1145/2968478.2968484]

[27] Larsen KG, Legay A. Statistical model checking: Past, present, and future. In: Proc. of the 7th Int'l Symp. on Leveraging Applications of Formal Methods, Verification and Validation: Foundational Techniques. Imperial: Springer, 2016. 3–15. [doi: 10.1007/978-3-319-47166-2_1]

[28] Legay A, Viswanathan M. Statistical model checking: Challenges and perspectives. Int'l Journal on Software Tools for Technology Transfer, 2015, 17(4): 369–376. [doi: 10.1007/s10009-015-0384-z]

[29] Shmarov F, Zuliani P. ProbReach: Verified probabilistic delta-reachability for stochastic hybrid systems. In: Proc. of the 18th Int'l Conf. on Hybrid Systems: Computation and Control. Seattle: ACM, 2015. 134–139. [doi: 10.1145/2728606.2728625]

[30] Shmarov F, Zuliani P. Probabilistic hybrid systems verification via SMT and Monte Carlo techniques. In: Proc. of the 12th Int'l Haifa Verification Conf. Haifa: Springer, 2016. 152–168. [doi: 10.1007/978-3-319-49052-6_10]

[31] Ellen C, Gerwinn S, Fränzle M. Statistical model checking for stochastic hybrid systems involving nondeterminism over continuous domains. Int'l Journal on Software Tools for Technology Transfer, 2015, 17(4): 485–504. [doi: 10.1007/s10009-014-0329-y]

[32] Huang C, Chen X, Lin W, Yang ZF, Li XD. Probabilistic safety verification of stochastic hybrid systems using barrier certificates. ACM Trans. on Embedded Computing Systems, 2017, 16(5s): 186. [doi: 10.1145/3126508]

[33] Kumar AR, Liu SL, Fisac JF, Adams RP, Ramadge PJ. ProBF: Learning probabilistic safety certificates with barrier functions. arXiv: 2112.12210, 2021.

[34] Luo WH, Kapoor A. Airborne collision avoidance systems with probabilistic safety barrier certificates. In: Proc. of the 33rd Conf. on Neural Information Processing Systems. Vancouver: NeurIPS, 2019. 1–11.

[35] Jing HM, Nakahira Y. Probabilistic safety certificate for multi-agent systems. In: Proc. of the 61st IEEE Conf. on Decision and Control (CDC). Cancun: IEEE, 2022. 5343–5350. [doi: 10.1109/CDC51059.2022.9992692]

[36] Zhao QY, Wang Y, Li XD. Safe neural network controller synthesis and verification for hybrid systems. Ruan Jian Xue Bao/Journal of Software, 2023, 34(7): 2981–3001 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6857.htm [doi: 10.13328/j.cnki.jos.006857]

[37] Shields DN, Storey C. The behaviour of optimal Lyapunov functions. Int'l Journal of Control, 1975, 21(4): 561–573. [doi: 10.1080/00207177508922012]

[38] Ratschan S, She ZK. Safety verification of hybrid systems by constraint propagation-based abstraction refinement. ACM Trans. on Embedded Computing Systems, 2007, 6(1): 8–37. [doi: 10.1145/1210268.1210276]

[39] Ratschan S. Simulation based computation of certificates for safety of dynamical systems. In: Proc. of the 15th Int'l Conf. on Formal

Modeling and Analysis of Timed Systems. Berlin: Springer, 2017. 303–317. [doi: 10.1007/978-3-319-65765-3_17]

[40] Abate A, Ahmed D, Edwards A, Giacobbe M, Peruffo A. FOSSIL: A software tool for the formal synthesis of lyapunov functions and barrier certificates using neural networks. In: Proc. of the 24th Int'l Conf. on Hybrid Systems: Computation and Control. Nashville: ACM, 2021. 24. [doi: 10.1145/3447928.3456646]

[41] Llibre J, Valls C. On the integrability of the Einstein–Yang–Mills equations. Journal of Mathematical Analysis and Applications, 2007, 336(2): 1203–1230. [doi: 10.1016/j.jmaa.2007.03.049]

附中文参考文献:

[12] 甘庭, 夏壁灿. 运用栅栏函数验证连续系统的有界时间安全性. 软件学报, 2016, 27(3): 645–654. http://www.jos.org.cn/1000-9825/4986.htm [doi: 10.13328/j.cnki.jos.004986]

[36] 赵庆晔, 王豫, 李宣东. 安全的混成系统神经网络控制器生成与验证. 软件学报, 2023, 34(7): 2981–3001. http://www.jos.org.cn/1000-9825/6857.htm [doi: 10.13328/j.cnki.jos.006857]

杨紫萱(2000－), 女, 硕士生, 主要研究领域为智能系统分析与验证.

王建林(1978－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为人工智能, 符号计算, 工业软件开发.

曾霞(1987－), 女, 博士, 副教授, CCF 专业会员, 主要研究领域为智能系统分析与验证, 优化理论, 符号-数值混合计算.

曾振柄(1963－), 男, 博士, 教授, 博士生导师, 主要研究领域为计算机数学, 人工智能程序设计.

任勐鑫(2002－), 男, 硕士生, CCF 学生会员, 主要研究领域为智能系统分析与验证, 强化学习.

杨争峰(1980－), 男, 博士, 教授, 博士生导师, 主要研究领域为计算机数学, 形式化方法.