基于多面体李雅普诺夫函数的区间 MDP 策略综合⋆

Negar Monir[a,*,1], SadeghSoudjani[b,c,2]

[a]纽卡斯尔大学计算机学院，英国泰恩河畔纽卡斯尔[b]马克斯·普朗克软件系统研究所，德国[c]伯明翰大学，英国伯明翰

文章信息

关键词：区间马尔可夫决策过程 多目标优化 切换仿射系统 动态规划 值迭代 多面体李雅普诺夫函数 反例引导的归纳综合（CEGIS）

摘要

不确定性下的决策是许多安全关键型应用的核心，这些应用中的决策必须以概率建模形式体系为指导。本文提出了一种新颖的方法，利用多面体李雅普诺夫函数，在多目标区间马尔可夫决策过程中进行策略综合。不同于以往主要依赖二次函数的基于李雅普诺夫函数的方法，我们的方法采用多面体函数，以提高动态规划值迭代过程中管理不确定性的准确性。我们将值迭代算法重新表述为具有区间不确定性的切换仿射系统，并运用控制理论稳定性原理，综合引导系统趋向预期目标集的策略。通过构造一个吸引域不变集，我们确保综合得到的策略能够提供收敛保证，同时最大限度地降低底层模型中转移不确定性的影响。我们的方法无需进行计算密集的帕累托曲线计算，而是直接确定一种策略，使各目标落在其目标值的指定范围内。我们通过数值案例研究验证了该方法，其中包括回收机器人和电动汽车电池，结果表明该方法能够有效地在不确定性下进行策略综合。

arXiv:2601.03445v1 [eess.SY] 6 Jan 2026

1. 引言

研究动机。不确定性下的决策在自主系统、机器人学和医疗保健等多个领域都面临挑战。马尔可夫决策过程（MDP）为此类问题提供了一种建模框架，使智能体能够根据概率状态转移选择动作，以优化目标函数（Bellman，1957年）。然而，许多应用需要处理多个相互竞争的目标。例如，自动驾驶汽车必须兼顾安全与效率，能源电网需要权衡成本与可靠性，而医学诊断往往需要在准确性和可解释性之间取舍。这种复杂性推动了多目标马尔可夫决策过程（MOMDP）框架的发展，其旨在制定能够平衡这些相互竞争准则的策略（Chatterjee、Majumdar 和 Henzinger，2006年；Etessami、Kwiatkowska、Vardi 和 Yannakakis，2008年）。当转移概率因传感器噪声、环境变化或不完备知识等因素而存在不确定性时，问题会更加复杂。区间马尔可夫决策过程（IMDP）通过将转移概率建模为区间而非固定值来应对这一问题（Givan、Leach 和 Dean，2000年；Haddad 和 Monmege，2018年）。研究人员已开发出多种技术，用于为区间马尔可夫决策过程综合最优或近似最优策略，通常还能为最坏情况性能提供保证（Nilim 和 El Ghaoui，2005年；Wu 和 Koutsoukos，2008年；Wolff、Topcu 和 Murray，2012年；Delimpaltadakis、Lahijanian、Mazo Jr 和 Laurenti，2023年）。为进一步改善决策，研究者开始关注多目标区间马尔可夫决策过程（MOIMDP）（Hahn、Hashemi、Hermanns、Lahijanian 和 Turrini，2017年、2019年），这需要专门的算法来处理不确定转移和多目标查询，同时确保策略鲁棒性。

动态规划能够有效求解马尔可夫决策过程中的优化问题。关键技术包括值迭代（VI）和策略迭代。值迭代通过迭代更新状态值，直至收敛，从而得到最优价值函数，帮助确定每个状态下的最佳策略。这一过程简化了计算，使值迭代成为概率系统中决策的有力工具（Bertsekas等人，2011年；Delgado、de Barros、Dias和

*通讯作者：s.seyedmonir2@newcastle.ac.uk（N. Monir）；sadegh@mpi-sws.org（S. Soudjani）

ORCID：[1]N. Monir 的研究得到英国工程与自然科学研究委员会（EPSRC）项目 EP/W524700/1 和纽卡斯尔大学全球奖学金的资助。[2]S. Soudjani 的研究得到以下项目资助：欧洲创新委员会项目 101070802 和欧洲研究委员会项目 101089047。

Sanner，2016年）。面对大型马尔可夫决策过程带来的挑战，研究人员正在探索近似动态规划和强化学习，以提升可扩展性（Bertsekas，2011年；Sutton和Barto，2018年；Lavaei、Perez、Kazemi、Somenzi、Soudjani、Trivedi和Zamani，2023年）。此外，相关研究还探讨了它们的收敛性质和理论局限（Tsitsiklis和Van Roy，1996年；Tsitsiklis，2002年；Ha、Wang和Liu，2021年）。

近期研究主要致力于将基于动态规划的策略综合扩展到多目标设置（Chatterjee 等人，2006年；Etessami 等人，2008年；Hahn等人，2017年，2019年），并将基于区间的不确定性纳入马尔可夫决策过程（Haddad 和 Monmege，2018年；Mathiesen、Lahijanian 和 Laurenti，2024年）。然而，先前同时研究多目标优化和基于区间的转移的工作，主要依赖计算帕累托曲线来确定能够权衡不同目标的策略（Hahn等人，2017年，2019年；Scheftelowitsch、Buchholz、Hashemi 和 Hermanns，2017年）。正如（Hahn等人，2019年）所述，通常无法在多项式时间内得到帕累托曲线的精确表示，因此需要通过 ϵ-近似来计算该曲线。这促使我们开展本文的研究：我们提出了一种新颖的基于李雅普诺夫函数的值迭代算法，将值迭代表示为离散时间切换仿射系统（dt-SAS），并利用多面体李雅普诺夫函数（PLF）分析收敛性并综合策略。

相关工作。 利用动力系统理论分析迭代算法的收敛性，已有相关文献对此进行研究。这包括用于一阶优化的梯度下降（Brockett，1988年，1991年；Helmke 和 Moore，2012年）、计算机视觉中的最速下降算法（Bloch，1990b，a），以及深度神经网络中的学习（Yeung、Kundu 和 Hodas，2019年；Rajendra 和 Brahmajirao，2020年；Xie、Tang 和 Kuang，2022年）。这一方法在策略综合中的一个重要应用，是Iervolino、Tipaldi和Forootani(2023年)的研究，该研究利用李雅普诺夫理论，通过将动态规划表述为 dt-SAS，并采用椭球吸引不变集（吸引不变集）稳定 dt-SAS，来研究马尔可夫决策过程中的值迭代的收敛性。然而，该研究仅限于单目标马尔可夫决策过程，且未考虑区间不确定性，因此其在区间马尔可夫决策过程和多目标设置中的适用性受到限制。近期，Tipaldi、Iervolino、Massenio 和 Forootani (2025年)将马尔可夫决策过程的值迭代表述为 dt-SAS，并通过将转移矩阵和奖励矩阵视为未知量来纳入模型不确定性，但该研究仅涉及单目标，且未考虑区间不确定性。

为稳定 dt-SAS，研究中采用了多种基于李雅普诺夫函数的方法，尤其是通过定义吸引不变集（ISoA）将系统轨迹限制在吸引区域内。常见方法是利用二次李雅普诺夫函数构造椭球形吸引域内集（Deaecto and Geromel，2017年）。Albea Sanchez、Ventosa-Cutillas、Seuret 和 Gordillo （2020年）的研究引入了用于策略综合的鲁棒椭球形吸引域内集，但这种设计需要进行松弛处理，可能扩大稳定区域并降低准确性。由Monir、Sadabadi 和 Soudjani （2025年）研究并由Deaecto and Geromel （2017年）针对不确定环境改进的基于场景的扩展收紧了吸引不变集，但增加了计算成本。这些权衡促使人们探索针对 dt-SAS 的保守性更低的稳定化证书。

二次李雅普诺夫函数常用于稳定性分析，但由于其对称性和固定曲率，可能导致稳定区域过于保守，从而限制策略综合。近期研究表明，多面体李雅普诺夫函数（PLF）在表示多面体吸引不变集（ISoA）方面更为灵活且准确（Sun and Ge，2011年；Blanchini、Miani 等人，2008年；Ahmadi and Jungers，2016年）。多面体李雅普诺夫函数具有分段仿射形式，可根据系统几何特征进行定制，因此在定义稳定区域时保守性更低。这一优势对于切换系统和混合系统尤为重要，因为其动力学会随不同模式而变化，需要非均匀的稳定性保证（Polański，2000年；Berger、Jungers 和 Wang，2022年）。此外，多面体李雅普诺夫函数还便于采用反例引导方法（Ahmed、Peruffo 和 Abate，2020年；Bergerand Sankaranarayanan，2022年，2023年）。

贡献。 本文提出一种新颖的、基于李雅普诺夫函数的方法，通过利用多面体李雅普诺夫函数，为多目标区间马尔可夫决策过程综合策略。该方法提高了定义稳定区域和综合策略的准确性。我们将多目标区间马尔可夫决策过程建模为具有区间不确定性的 dt-SAS，从而能够应用李雅普诺夫稳定性理论，确保系统收敛至期望的目标值。与以往依赖二次李雅普诺夫函数的基于李雅普诺夫函数的值迭代方法不同，这些方法会导致稳定区域过于保守；我们的方法采用多面体李雅普诺夫函数。这使我们能够构造更紧致的吸引不变集，从而在不确定性下保持鲁棒性的同时，提高策略综合的精确度。此外，我们采用反例引导算法，迭代优化多面体李雅普诺夫函数的参数及其对应的吸引不变集。这确保了稳定性验证过程系统且具有适应性。图1展示了所提算法的高层次表示，概述了策略综合与稳定性分析所涉及的关键步骤。

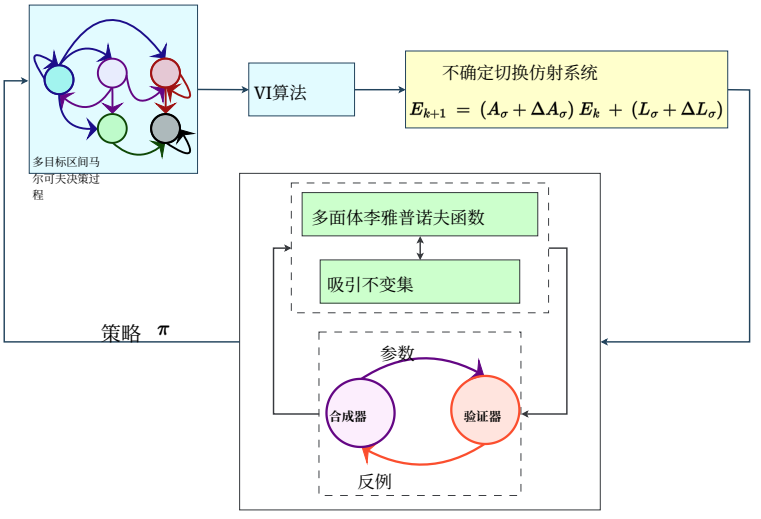


图1： 所提出方法的高层表示。

本文方法的初步版本已发表于会议论文（Monir、Schön 和 Soudjani，2024年），该版本采用二次李雅普诺夫函数和非凸优化，因此与状态方程的分段仿射结构不相符。当前稿件从以下方面扩展了会议版本：（a）我们提出了一个基于多面体李雅普诺夫函数的策略综合框架，可生成更紧致且形状更准确的不变集。由二次李雅普诺夫函数导出的椭球证书所构造的吸引域呈椭球形，但这往往会对切换系统或仿射系统的实际几何形态进行过度平滑和过度逼近。相比之下，多面体李雅普诺夫函数能够定义具有可调节面的区域，使其与切换边界和主导方向相吻合。这种额外的几何灵活性消除了椭球形状的限制，降低了保守性，并使不确定环境下的策略更加精确。（b）由于多面体李雅普诺夫函数支持反例引导计算，我们提出了两种用于设计多面体李雅普诺夫函数的反例引导算法。该方法根据约束违反情况细化稳定性条件，从而实现更准确的稳定性验证。

概述。 本文的结构如下。第 2 节介绍预备知识和问题陈述。在第 3节中，我们将 MOIMDP 建模为带不确定性的 dt-SAS。第 4 节提出两种反例引导算法，用于为不确定的 dt-SAS 设计多面体 ISoA，并提出一种鲁棒 VI 算法，用于为 MOIMDP 设计策略。在第 5节中，我们介绍案例研究，以展示我们方法的有效性，并将结果与基线方法进行比较。最后，第 6 节总结全文，并提出未来研究方向。

2. 预备知识和问题陈述

本节介绍关键定义、预备知识和问题陈述。

记号。 $\mathbb{N}$、$\mathbb{R}$、$\mathbb{R}_{>0}$、$\mathbb{R}_{\geq 0}$、$\mathbb{R}^n$ 和 $\mathbb{R}^{n\times m}$ 分别表示包含零的自然数集、实数集、正实数集、非负实数集、n 维欧几里得空间以及 $n\times m$ 实矩阵集。$\mathbb{N}_m$ 表示集合 $\{1,2,\ldots,m\}$，$0_{n\times m}$ 表示元素均为零的 $n\times m$ 矩阵。任意集合 A的基数记为 $|A|$。对于任意 $M\in\mathbb{R}^{n\times n}$，$\det(M)$ 表示 M的行列式。对于任意维数适当的矩阵 $A = A^\top$、B 和$C = C^\top$，我们简写为

$$\begin{bmatrix} A & B \\ * & C \end{bmatrix} := \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}.$$

$\mathrm{Co}(A,B)$ 是由矩阵 A 和 B生成的凸包。集合的内部用 $\mathrm{int}(\cdot)$表示。将两个维度适当的矩阵 A 和 B 按行拼接记为 $[A,B]$，按列拼接记为 $[A;B]$。最后，维度为 $m\in\mathbb{N}$ 的单位单纯形定义为 $\Lambda_m := \{[\lambda_1,\ldots,\lambda_m] \mid \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1\}$。

2.1. 多目标区间马尔可夫决策过程

我们研究在多目标区间马尔可夫决策过程（MOIMDP）上优化多目标的动态规划框架。

定义 1. 一个多目标区间马尔可夫决策过程 是一个元组 $\underline{\bar{\Sigma}} = (\bar{X}, x_0, U, P, \bar{P}, R, \bar{R})$，由有限状态集 X、初始状态 $x_0 \in X$和有限动作集 U组成。状态之间转移概率的下界和上界分别由 $P, \bar{P} : X \times U \times X \to [0,1]$给出。一组 q 奖励函数的下界和上界由 $R, \bar{R} : X \times U \to \mathbb{R}^q$给出，其中 $R := [r_1; r_2, \ldots; r_q]$ 以及 $\bar{R} := [\bar{r}_1; \bar{r}_2, \ldots; \bar{r}_q]$。

对于任意状态-动作对 $(x,u) \in X \times U$，下一个状态 $x' \in X$ 按照某个概率分布 $x' \sim \mathcal{I}(\cdot|x,u)$进行选择，其中对于所有 x，$\mathrm{P}(x'|x,u) \le \mathcal{I}(x'|x,u) \le \bar{\mathrm{P}}(x'|x,u)$ $u\,x'$。对于任意 $(x,u) \in X \times U$，奖励 $r_m(x,u)$、$m \in \mathbb{N}_q$属于区间 $[\mathrm{r}_m, \bar{\mathrm{r}}_m]$。马尔可夫 策略 $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \ldots)$是在任意时刻将状态映射到动作的一系列函数 $\pi_t : X \to U$，$t \in \mathbb{N}$。我们用 Π 表示所有此类策略的集合。注意，对于有限的 X和 U，可能的函数 $\pi : X \to U$的集合也是有限的，记为 $\bar{\Pi} = \{\pi^1, \pi^2, \ldots, \pi^M\}$，其中$M = |U|^{|X|}$。如果存在 $\pi : X \to U$使得 $\boldsymbol{\pi} = (\pi, \pi, \pi, \ldots)$，则策略 $\boldsymbol{\pi}$称为平稳策略。在这种情况下，略去记号滥用，我们简单地用 π 表示该平稳策略。

任意策略 $\boldsymbol{\pi}$与每个时间步选择的概率分布 $\mathcal{I}(x'|x,u)$ 共同决定状态序列 $x(0), x(1), x(2), \ldots$上的一个概率测度。策略 $\boldsymbol{\pi}$的性能根据其诱导的期望总折扣奖励进行评估，而该奖励也取决于从区间 $[\mathrm{P}(x'|x,u), \bar{\mathrm{P}}(x'|x,u)]$ 中选取的转移概率 $\mathcal{I}(x'|x,u)$，以及从区间 $[\mathrm{R}(x,u), \bar{\mathrm{R}}(x,u)]$中选取的奖励。因此，研究转移概率和奖励中的不确定性所对应的最优和最坏情况性能至关重要。

给定策略 $\boldsymbol{\pi}$ 和折扣因子 $0 < \gamma_i < 1, i \in \mathbb{N}_q$，将期望总折扣奖励的最优和最坏情况性能定义为任意初始状态的函数，如下所示

$$w^m_{\boldsymbol{\pi},\mathrm{opt}}(x_0) := \mathrm{opt}_{\boldsymbol{\mu}} \mathbb{E}^{\boldsymbol{\pi},\boldsymbol{\mu}} \left[\sum_{t=0}^{\infty} \gamma_m^t r_m(x(t), \pi_t(x(t))) | x(0) = x_0 \right],$$

其中，$\mathrm{opt} \in \{\max, \min\}$ 是针对对抗策略 $\boldsymbol{\mu}$ 而言的，该策略选择转移概率 $\mathcal{I}$，其取值来自区间 $[\mathrm{P}, \bar{\mathrm{P}}]$，以及奖励 r_m，其取值来自区间 $[\mathrm{r}_m, \bar{\mathrm{r}}_m]$。期望算子 $\mathbb{E}^{\boldsymbol{\pi},\boldsymbol{\mu}}$是相对于策略 $(\boldsymbol{\pi}, \boldsymbol{\mu})$下状态序列所诱导的概率分布而言的。当策略$w^m_{\pi,\mathrm{opt}}(\cdot)$为平稳策略时，我们将这些价值函数记为 $\boldsymbol{\pi}$，它们满足以下贝尔曼方程（Bertsekas 等人, 2011年）

$$w^m_{\pi,\mathrm{opt}}(x) = \mathrm{opt}_{\mathcal{I}, r_m} \left[r_m(x, \pi(x)) + \sum_{x' \in X} \gamma_m \mathcal{I}(x'|x, \pi(x)) w^m_{\pi,\mathrm{opt}}(x') \right],$$

对于所有 $x \in X$ 和 $m \in \mathbb{N}_q$。

假设 X 具有 n 个状态，并且状态具有某个固定顺序，则价值函数可以表示为一个向量。稍微滥用一下记号，该向量也记作$w^m_{\pi,\ \mathrm{opt}} \in \mathbb{R}^n$。贝尔曼算子也可以用向量形式定义为 $\mathcal{T}_\pi : \mathbb{R}^n \to \mathbb{R}^n$，其中

$$\mathcal{T}_\pi w = \mathrm{opt}_{\mathcal{I}, R_m} \left[R_m(\pi) + \gamma_m \mathcal{I}(\pi) w \right], \quad \forall w \in \mathbb{R}^n,$$

其中，$R_m(\pi) \in \mathbb{R}^n$ 是包含 $r_m(x, \pi(x))$的奖励区间的向量，$\mathcal{I}(\pi)$ 是在 π下包含转移概率区间的矩阵；向量 R_m 和矩阵 $\underline{\mathcal{I}}$ 的每个元素都可以分别写成区间$R_{mi} \in [\mathrm{R}_{mi}, \bar{\mathrm{R}}_{mi}]$和 $\mathcal{I}_{ij} \in [\mathrm{P}_{ij}, \bar{\mathrm{P}}_{ij}]$。用于计算$w^m_{\pi,\ \mathrm{opt}}$的值迭代为

$$w^m_{k+1} = \mathcal{T}_\pi w^m_k, \quad \forall k \in \mathbb{N} \text{ with } w^m_0 = 0, \tag{2.1}$$

其中$w^m_{\pi,\ \mathrm{opt}} = \lim_{k \to \infty} w^m_k$。值迭代（2.1）需要在$\mathrm{opt} \in \{\max, \min\}$ 下求解两次，以获得针对转移概率和奖励中的不确定性的最优和最坏情况性能。

注 1（多目标马尔可夫决策过程）。若定义 $\underline{\bar{\Sigma}}$ 中的 $\underline{P}$和 $\bar{P}$ 分别为 I，则可得到一个转移概率为 P、奖励为 R 的多目标马尔可夫决策过程，记为 $\Sigma = (X, x_0, U, P, R)$。将 P和 R代入多目标区间马尔可夫决策过程的方程中，

去掉关于不确定性的优化算子 opt，并考虑价值函数 v_π^m，可得多目标马尔可夫决策过程的仿射贝尔曼算子：

$$\mathcal{T}_\pi v = R_m(\pi) + \gamma_m P(\pi) v, \quad \forall v \in \mathbb{R}^n.$$

用于计算v_π^m的相应 VI 算法为

$$v_{k+1}^m = \mathcal{T}_\pi v_k^m, \quad \forall k \in \mathbb{N} \; with \; v_0^m = 0, \tag{2.2}$$

其中 $v_\pi^m = \lim_{k\to\infty} v_k^m$。

2.2. 问题陈述

给定一个多目标区间马尔可夫决策过程 $\underline{\bar{\Sigma}} = (X, x_0, U, \mathrm{P}, \mathrm{P}, \mathrm{R}, \mathrm{R})$，我们的目标是设计一种策略 $\boldsymbol{\pi}$，使式（2.1）中的价值函数收敛到目标值 W_{tar}的一个邻域。

问题1。给定一个多目标区间马尔可夫决策过程 Σ和目标值 $W_{tar} := [w_{tar}^1; w_{tar}^2; \ldots; w_{tar}^q]$，求一个集合 $\mathcal{G}$ 和一种策略 $\boldsymbol{\pi}$，使得 $W_{tar} \in \mathcal{G}$ 且 $W_{\boldsymbol{\pi},\ opt} = [w_{\boldsymbol{\pi},opt}^1; w_{\boldsymbol{\pi},opt}^2; \ldots; w_{\boldsymbol{\pi},\text{最优}}^q] \in \mathcal{G}$。

对于任意 W_{tar} 和任意平稳策略 π，只需取足够大的 $\mathcal{G}$，上述问题便显然可行。本文旨在通过下一节所述的新型基于李雅普诺夫函数的值迭代算法，找到一个尽可能紧的集合 $\mathcal{G}$。

3. 具有不确定性的 dt-SAS 的 VI 算法

为解决问题 1，我们将多目标区间马尔可夫决策过程 $\Sigma = (X, x_0, U, \mathrm{P}, \mathrm{P}, \mathrm{R}, \mathrm{R})$ 的 VI 算法（2.1）重新表述为一个带不确定性的 dt-SAS（dt-USAS）稳定性问题，用以描述其底层误差动力学。随后，我们通过反例引导计算设计一个多面体 ISoA，从而合成一种鲁棒切换策略，使价值函数趋向期望的目标值。

令 $W_k = [w_k^1; w_k^2; \ldots; w_k^q]$ 为所有 q 个目标的增广价值函数。因此，VI 算法（2.1）的解包含于以下解集之中

$$W_{k+1} = A_{\pi_k} W_k + B_{\pi_k},$$

带有区间矩阵

$$A_\pi = \mathrm{diag}(\gamma_1 \mathcal{I}(\pi), \ldots, \gamma_q \mathcal{I}(\pi)), \text{ and } B_\pi = \left[R_1(\pi); R_2(\pi); \ldots; R_q(\pi)\right], \tag{3.1}$$

$\bar{}$对所有 $\pi \in \Pi$均成立。此外，我们将相对于目标值 w_{tar}^m 的误差定义为 $e_k^m := w_k^m - w_{tar}^m$ 对所有目标 $m \in \mathbb{N}_q$均成立，其动力学为

$$e_{k+1}^m = \gamma_m \mathcal{I}(\pi_k) e_k^m + l_{\pi_k}^m,$$

其中 $l_\pi^m := R_m(\pi) - (I - \gamma_m \mathcal{I}(\pi)) w_{tar}^m$。与增广价值函数 W_k类似，增广误差$E_k = [e_k^1; e_k^2; \ldots; e_k^q]$ 由受不确定性约束的切换仿射系统描述

$$E_{k+1} = A_{\pi_k} E_k + L_{\pi_k}, \tag{3.2}$$

其中 $L_{\pi_k} = [l_{\pi_k}^1; l_{\pi_k}^2; \ldots; l_{\pi_k}^q] = B_{\pi_k} - (I - A_{\pi_k}) W_{tar}$。请注意，矩阵 A_{π_k}和 L_{π_k}均具有不确定性，可表示为凸包

$$[A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa\in\mathbb{D}}, \quad \forall \pi \in \bar{\Pi}, \tag{3.3}$$

由有限个 $\mathbb{D}$矩阵 A_π^j 和 L_π^j生成。

3.1. 构造可行值集合

在本小节中，我们通过分析平稳策略评估的渐近稳定性，来求解问题1 中的可行目标值集合。我们确定名义模型下价值函数能够达到的稳态值。

我们将 $\mathcal{I}$ 改写为 $\mathcal{I} = \underline{\mathcal{I}} + \bar{\Delta}\mathcal{I}$，其中 $\mathcal{I}$ 是该区间马尔可夫决策过程的名义转移概率矩阵。一种选择是令元素 $\hat{\mathcal{I}}_{ij}$ 尽可能接近 $\frac{1}{2}(\mathrm{P}^{ij} + \underline{\mathrm{P}}_{ij})$，并满足 $[\hat{\mathcal{I}}_{ij} - \Delta_{ij}, \hat{\mathcal{I}}_{ij} + \Delta_{ij}] \supset [\mathrm{P}^{ij}, \mathrm{P}_{ij}]$。类似地，我们将 R 改写为 $R = \hat{R} + \Delta R$，其中 $\hat{R}$ 是该区间马尔可夫决策过程的名义奖励。一种选择是选择 $\hat{R}_{ij} = \frac{1}{2}(\mathrm{R}^{ij} + \underline{\mathrm{R}}_{ij})$。注意，$(X, x_0, U, , \hat{\mathcal{I}}, \hat{R})$ 是作为名义模型的多目标马尔可夫决策过程。此外，定义名义矩阵

$$\hat{A}_\pi := \mathrm{diag}(\gamma_1 \hat{\mathcal{I}}(\pi), \ldots, \gamma_q \hat{\mathcal{I}}(\pi)), \text{ and } \hat{B}_\pi := \left[\hat{R}_1(\pi); \hat{R}_2(\pi); \ldots; \hat{R}_q(\pi)\right]. \tag{3.4}$$

名义模型的值迭代为

$$\hat{W}_{k+1} = \hat{A}_{\pi_k} \hat{W}_k + \hat{B}_{\pi_k}, \tag{3.5}$$

其稳态值为 $W_{\pi,\ ss}$ 并满足

$$\hat{W}_{\pi,ss} = \hat{B}_\pi + \hat{A}_\pi \hat{W}_{\pi,ss}. \tag{3.6}$$

注意，A_π 始终可逆，因为 $\gamma_m \in (0,1)$ 对所有 $m \in \mathbb{N}_q$ 成立，且 $\mathcal{I}(\pi)$ 的谱半径对于任意π均等于 1。这里，我们可以将（3.6）改写如下

$$\hat{W}_{\pi,ss} = (I - \hat{A}_\pi)^{-1} \hat{B}_\pi.$$

将 $W_{\pi,\ ss}$ 作为目标值 W_{tar}，可得 $L_\pi := B_\pi - (I - A_\pi) W_{tar} = 0$。因此，名义模型的误差动力学变为

$$\hat{E}_{k+1} = \hat{A}_\pi \hat{E}_k, \tag{3.7}$$

这是一个经典的线性时不变离散时间模型。无论初始误差如何，名义误差都会收敛到零（即模型是全局渐近稳定的，$\hat{A}_\pi$ 的所有特征值均位于复平面的单位圆内）。

引理1. 对于任意平稳策略 π、任意满足 $m \in \mathbb{N}_q$ 的 $\gamma_m \in (0,1)$，以及任意初始值 W_0，（3.5）中的 W_k 渐近收敛到相应的稳态值 $\hat{W}_{\pi,\ ss} = (I - \hat{A}_\pi)^{-1} \hat{B}_\pi$。

我们研究问题 1，以引导价值函数 W_k 趋向给定的目标值 $W_{tar} = [w^1_{tar}; w^2_{tar}; \ldots; w^q_{tar}]$。考虑以下相对于名义矩阵指定的广义平衡点可行集：

$$\mathcal{F} := \{W \in \mathbb{R}^{qn} : W = (I - \hat{A}_\lambda)^{-1} \hat{B}_\lambda, \lambda \in \Lambda_M\}, \tag{3.8}$$

其中，Λ_M 是维度为 M 的单位单纯形，并且

$$\hat{A}_\lambda := \sum_{\pi \in \bar{\Pi}} \lambda_\pi \hat{A}_\pi, \text{ and } \hat{B}_\lambda = \sum_{\pi \in \bar{\Pi}} \lambda_\pi \hat{B}_\pi. \tag{3.9}$$

这些广义平衡点被计算为与策略 $\bar{}$ 相关的动力学（3.5）的凸组合所产生的平衡点 $\pi \in \Pi$。对于任意 $W_{tar} \in \mathcal{F}$，误差动力学与（3.2）中的相同。

3.2. 带不确定性的离散时间随机切换系统的稳定性分析

在本小节中，我们利用李雅普诺夫理论分析（3.2）所描述的带不确定性的离散时间随机切换系统的稳定性。我们的目标是构造一个由多面体李雅普诺夫函数导出的吸引不变集，从而保证误差动力学收敛到原点周围的指定区域。目标是构造切换律 $\pi(E)$，使得理想情况下，$E_k \to 0$ 当 $k \to \infty$ 时，对所有初始条件均成立。然而，在系统存在不确定性的情况下，通常无法实现这一条件。相反，我们设计 $\pi(E)$ 在 Π 的元素之间进行切换，以引导系统轨迹趋向包含原点的吸引不变集。为形式化分析稳定性并构造该集合，我们给出吸引不变集的定义。

定义 2（吸引不变集）。集合 $\Omega \subset \mathbb{R}^{qn}$ 是系统（吸引不变集）（3.2）在切换策略 π_k控制下的吸引不变集，当存在李雅普诺夫函数 $\mathbb{V} : \mathbb{R}^{qn} \to \mathbb{R}_{\geq 0}$ 使得以下条件同时满足时：

（C1） $0 \in \Omega$，（C2）若 $E_k \notin \Omega$，则
$\Delta\mathbb{V}(E_k) := \mathbb{V}(E_{k+1}) - \mathbb{V}(E_k) < 0$，（C3）若 $E_k \in \Omega$，
则 $E_{k+1} \in \Omega$。

若存在这样的李雅普诺夫函数，则可以保证在该切换策略下，对于任意初始误差，误差始终会收敛到集合 Ω，并且 E_0 此后始终保持在 Ω 内。此时称该系统为实际稳定（Deaecto 和 Egidio，2016年）。

根据定义 2，我们采用候选多面体李雅普诺夫函数

$$\mathbb{V}(E) := \max_{c \in \mathcal{V}} \left(c^\top E - d_c \right) \tag{3.10}$$

对于某个有限向量集 $\mathcal{V}$ 和实数值 d_c，这些值使我们能够移动水平集的中心。根据条件 (C2)，选取以下最小类型状态反馈切换控制律：

$$\pi(E) = \arg\min_{\pi \in \bar{\Pi}} \mathbb{V}(A_\pi E + L_\pi). \tag{3.11}$$

我们还将 $\mathbb{V}$ 的水平集

$$\Omega := \{E \in \mathbb{R}^{qn} : \mathbb{V}(E) \leq \rho\}, \tag{3.12}$$

作为候选多面体 ISoA，其中某个 ρ 待设计。

问题 2。 给定 (3.2) 中带不确定性的离散时间随机切换系统的误差动力学，设计 (3.10) 中的多面体李雅普诺夫函数参数，以及 $\rho > 0$，使得 Ω 在 (3.12) 中，根据定义 (3.11) 所定义的切换律，成为带不确定性的离散时间随机切换系统的吸引不变集 2。

为了解决这个问题，我们将在下一节给出算法解。

4. 多目标区间马尔可夫决策过程的策略综合

本节针对问题1-2，提出两种利用反例引导的归纳综合（CEGIS）设计多面体 ISoA 的算法（Abate、David、Kesseli、Kroening 和 Polgreen，2018年），并提供一种鲁棒 VI 算法来综合策略。

从高层次来看，CEGIS方法通过交替执行以下操作来运行：提出满足当前要求集的候选证书，并向验证器发起查询。验证器要么返回反例，要么确认候选证书的有效性；这一过程持续进行，直到不再发现反例。第一种算法在小节 4.1 中介绍，利用模理论可满足性（SMT）求解器（Barrett 和 Tinelli，2018年）。第二种算法在小节 4.2 中介绍，该算法基于优化，通过求解优化问题而非使用 SMT 求解器来确定证书和反例。此外，小节 4.3提出了一种鲁棒 VI 算法。该算法利用计算得到的任一多面体李雅普诺夫函数（PLF）及其多面体 ISoA 来求解问题 1。

4.1. 通过反例引导的归纳综合和 SMT 综合多面体吸引不变集

在本节中，我们不失一般性地设定 $\rho = 1$ 为式（3.12）中的吸引不变集（ISoA），并建立多面体李雅普诺夫函数（PLF）的参数条件，以求解问题2。

定理 1。 函数 $\mathbb{V}(E)$ 在 （3.10）、切换策略 π 在 （3.11）以及吸引不变集 Ω 在 （3.12）与 $\rho = 1$ 共同为问题 *2* 提供一个解，只要 $\mathbb{V}(E)$ 的参数满足以下陈述：

$$\begin{aligned}\Psi : \forall E, \exists \pi, \{&(\forall c_i \in \mathcal{V},\, d_{c_i} \geq -1) \wedge (\exists c_i \in \mathcal{V}, {c_i}^\top E - d_{c_i} \geq 0) \wedge ([\forall c_i \in \mathcal{V}, (c_i^\top E - d_{c_i}) \leq 1] \vee \\ &[\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, \exists c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} < c_\ell^\top E - d_{c_\ell})]) \wedge ([\exists c_i \in \mathcal{V},\, (c_i^\top E - d_{c_i}) > 1] \vee \\ &[\forall \kappa \in \mathbb{D}, \forall c_j \in \mathcal{V}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} \leq 1])\}. \end{aligned} \tag{4.1}$$

该定理的证明见附录。注意，条件 Ψ 的否定可以明确写成

$$\begin{aligned}\neg\Psi : \exists E, \forall \pi, \{&(\exists c_i \in \mathcal{V},\, d_{c_i} < -1) \vee (\forall c_i \in \mathcal{V}, {c_i}^\top E - d_{c_i} < 0) \vee ([\exists c_i \in \mathcal{V}, (c_i^\top E - d_{c_i}) > 1] \wedge \\ &[\exists c_j \in \mathcal{V}, \exists \kappa \in \mathbb{D}, \forall c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} \geq c_\ell^\top E - d_{c_\ell})]) \vee ([\forall c_i \in \mathcal{V},\, (c_i^\top E - d_{c_i}) \leq 1] \wedge \\ &[\exists \kappa \in \mathbb{D}, \exists c_j \in \mathcal{V}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} > 1])\}. \end{aligned} \tag{4.2}$$

尽管定理 1 给出了所需条件，但由于系统中的不确定性以及参数空间的高维性，寻找满足这些要求的有效多面体李雅普诺夫函数颇具挑战性。直接尝试解析求解可能会过于保守，或者在计算上不可行。为了逐步微调参数，我们采用反例引导的归纳综合框架，通过综合与验证的过程迭代改进这些条件（Solar-Lezama、Rabbah、Bodík 和 Ebcioğlu，2005年；Jha、Gulwani、Seshia 和 Tiwari，2010年；Abate 等，2018年）。反例引导的归纳综合框架通过以下两个主要步骤运行：

综合步骤： 此步骤基于初始参数化生成候选多面体李雅普诺夫函数和多面体 ISoA。通过 SMT 求解器选择满足定理 Ψ 中的条件1 的参数。

验证步骤： 在此步骤中，SMT 求解器通过检查 $\neg\Psi$ 的可满足性并寻找反例，验证候选多面体李雅普诺夫函数是否满足条件 Ψ。如果没有找到反例，则该过程成功结束。如果找到了反例，则该反例是一个违反条件 Ψ的状态 E。生成反例后，会利用它来优化多面体李雅普诺夫函数的参数。此过程将迭代重复，以确保最终综合出的函数符合所有必要条件。

CEGIS方法的结构如图 2所示，图中展示了综合步骤与验证步骤之间的交互。基于 CEGIS 的完整方法用于设计多面体李雅普诺夫函数及其多面体 ISoA，详见算法 1。

4.2. 基于优化的反例引导归纳综合来综合多面体 ISoA

在本小节中，我们不再像上一小节那样假设 $\rho = 1$，而是通过在方程（$d_c = 0$ 中设定3.12），简化多面体吸引不变集。我们为多面体李雅普诺夫函数建立了一个简化条件，并使用优化而非 SMT 求解器。

定理 2。 函数 $\mathbb{V}(E)$ 在 （3.10）中，满足 $d_c = 0$，切换策略 π 在 （3.11）中，以及吸引不变集 Ω 在 （3.12）中，满足$\rho > 0$ 时，构成问题 *2* 的一个解，前提是 $\mathbb{V}(E)$ 的参数满足以下条件：

$$\begin{aligned}\Phi = \forall E, \exists \pi, (\exists c_i \in \mathcal{V}, {c_i}^\top E \geq 0) \wedge \{&([\forall c_i \in \mathcal{V}, (c_i^\top E) \leq \rho] \wedge [\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) \leq \rho]) \vee \\ &([\exists c_i \in \mathcal{V},\, (c_i^\top E) > \rho] \wedge [\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, \exists c_\ell \in \mathcal{V},\, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) < c_\ell^\top E)])\}. \end{aligned} \tag{4.3}$$

定理的证明放在附录中。注意，条件 Φ 的否定为

$$\begin{aligned}\neg\Phi = \exists E, \forall \pi, (\forall c_i \in \mathcal{V},\, {c_i}^\top E < 0) \vee \{&[\exists c_i \in \mathcal{V}, (c_i^\top E) > \rho] \wedge [\exists c_j \in \mathcal{V}, \exists \kappa \in \mathbb{D}, \forall c_\ell \in \mathcal{V}, \\ &(c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) \geq c_\ell^\top E)]\} \vee \{[\forall c_i \in \mathcal{V},\, (c_i^\top E) \leq \rho] \wedge [\exists c_j \in \mathcal{V},\, \exists \kappa \in \mathbb{D},\, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) > \rho]\}. \end{aligned} \tag{4.4}$$

我们提出一种基于优化的反例引导归纳综合方法，包含两个主要步骤，用于设计多面体李雅普诺夫函数及其多面体 ISoA，依据条件 Φ （4.3）。

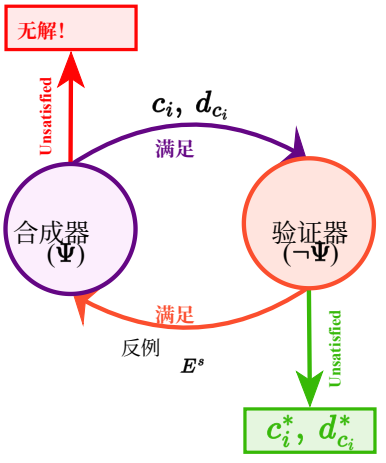


图 2：用于寻找有效多面体李雅普诺夫函数的 CEGIS 方法示意图。

算法 1 通过基于 SMT 的反例引导归纳综合合成多面体李雅普诺夫函数及其多面体 ISoA

输入：用于 $\pi \in \Pi$ 和 $\kappa \in \mathbb{D}$ 的矩阵 A_π^κ 和 L_π^κ

1 初始化向量集合 $\mathcal{V}^0 = \emptyset$2 选择有限状态集作为反例 $\mathcal{E}^0$3 对于 $s \in \mathbb{N}$，执行以下操作：4 使用 SMT 求解器生成多面体李雅普诺夫函数候选 $\mathbb{V}(E)$ 的 $(c_i^s, d_{c_i}^s)$，使得式 (4.1) 中的 Ψ 对所有 $E \in \mathcal{E}^s$ 均成立 5 如果步骤 4 不可行，则 6 返回 FAIL 并退出 7 结束 8 $\mathcal{V}^{s+1} = \mathcal{V}^s \cup \{(c_i^s, d_{c_i}^s)\}$9 使用式 (4.2) 中的条件 $\neg\Psi$ 和 SMT 求解器查找反例 E^s 10 如果未找到反例，则 11 返回 $\mathcal{V}^s$ 和多面体李雅普诺夫函数 $\mathbb{V}(E)$ 并退出 12 结束 13 $\mathcal{E}^{s+1} = \mathcal{E}^s \cup \{E^s\}$14 结束 输出：带有支持向量 $\mathcal{V}$ 的多面体李雅普诺夫函数 $\mathbb{V}(E)$

综合步骤：在此步骤中，目标是根据条件 $c_i \in \mathcal{V}$，基于 Φ （4.3）中的条件，求得向量。因此，定义如下优化问题：

$$\min_{c_i \in \mathcal{V}, \rho \geq 0} \quad \rho$$
$$\text{s.t. } (c_i, \rho) \vDash \Phi, \tag{4.5}$$

其中，优化受到以下约束： $\mathcal{V}$ 和 $\rho \geq 0$ 中的向量以及为 E选定的元素必须满足 Φ 。

验证步骤：求解优化问题后，将对所得的多面体李雅普诺夫函数和多面体 ISoA 进行验证。我们不会一次性检查 $\neg\Phi$ 式（4.4）中的条件，而是将验证过程分为以下三个不同阶段，每个阶段针对该条件中的特定约束 $\neg\Phi$。这样最多可找到三个反例。完整过程详见算法 2。

下面，我们定义三个优化问题，分别对应于 $\neg\Phi$ 中的每个约束，具体如下： $E_1 = [e_1^1, e_1^2, \dots, e_1^{n\times q}]^\top$、$E_2 = [e_2^2, e_2^2, \dots, e_2^{n\times q}]^\top$ 和 $E_3 = [e_3^1, e_3^2, \dots, e_3^{n\times q}]^\top$

$$\max_{E_1} \left\{ \Sigma_{k=1}^{n\times q} e_1^k \text{ s.t. } \forall c_i \in \mathcal{V},\ {c_i}^\top E < 0 \right\}, \tag{4.6}$$

$$\max_{E_2} \{ \Sigma_{k=1}^{n\times q} e_2^k \text{ s.t. } \forall \pi,\ [\exists c_i \in \mathcal{V}, (c_i^\top E) > \rho] \ \wedge [\exists c_j \in \mathcal{V}, \exists \kappa \in \mathbb{D}, \forall c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) \geq c_\ell^\top E)] \}, \tag{4.7}$$

$$\max_{E_3} \{ \Sigma_{k=1}^{n\times q} e_3^k \text{ s.t. } \forall \pi, [\forall c_i \in \mathcal{V},\ (c_i^\top E) \leq \rho] \wedge\ [\exists c_j \in \mathcal{V},\ \exists \kappa \in \mathbb{D}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) > \rho] \}. \tag{4.8}$$

由于集合 $\mathcal{V}$、$\mathbb{D}$ 和 $\mathcal{E}$ 是有限的，综合和验证步骤中涉及的量词与析取可以转化为混合整数线性规划（MILP）优化。全称量词通过针对各自索引集重复添加约束来表示，而存在量词和析取则通过大 M 法或指示约束，使用二元选择变量或激活变量来表示（Williams, 2013; Conforti, Cornuéjols and Zambelli, 2014; Vielma, 2015）。这种方法使问题能够直接使用 MOSEK 或 Gurobi 等混合整数线性规划求解器求解（MOSEK ApS, 2025a; Gurobi Optimization, LLC, 2024）。 MOSEK Fusion 和 Pyomo 等建模工具（通过 pyomo.gdp）直接支持析取（MOSEK ApS, 2025b; Hart, Watson and Woodruff, 2011; Bynum, Hackebeil, Hart, Laird, Nicholson, Siirola, Watson, Woodruff et al., 2021）。

算法 2 通过基于优化的反例引导归纳综合多面体李雅普诺夫函数及其多面体吸引不变集 S

输入： 矩阵 A_π^κ 和 L_π^κ，用于 $\pi \in \Pi$和 $\kappa \in \mathbb{D}$

初始化向量集 $\mathcal{V}^0 = \emptyset$
选择有限状态集作为反例 $\mathcal{E}^0$
对于 $s \in \mathbb{N}$ **执行**
  生成 c_i^s多面体李雅普诺夫函数候选 $\mathbb{V}(E)$ 满足 $d_c = 0$，通过求解优化问题（4.5），其中约束在 on $E \in \mathcal{E}^s$
  如果步骤*4* 不可行**则**
    返回失败
    中断
  end
  $\mathcal{V}^{s+1} = \mathcal{V}^s \cup \{c_i^s\}$
  设置标志位 $\delta_1 = 0,\ \delta_2 = 0,\ \delta_3 = 0$
  通过求解优化问题 E^s（4.6）
  如果步骤*10* 不可行**则**
    $\delta_1 = 1$
  end
  求解 E_2^s优化问题（4.7）
  如果 步骤 *14* 不可行 **则**
    $\delta_2 = 1$
  end
  求解 E_3^s优化问题（4.8）
  如果 步骤 *18* 不可行 **则**
    $\delta_3 = 1$
  end
  如果 $\delta_1\delta_2\delta_3 == 1$**则**
    返回 $\mathcal{V}^s$ 和多面体李雅普诺夫函数 $\mathbb{V}(E)$
    中断
  end
  $\mathcal{E}^{s+1} = \mathcal{E}^s \cup \{E_1^s\ E_2^s\ E_3^s\}$
结束
输出： 多面体李雅普诺夫函数 $\mathbb{V}(E)$ 及其支持向量 $\mathcal{V}$

4.3. 基于李雅普诺夫函数的鲁棒 VI 算法

算法 1 和 2 为计算多面体李雅普诺夫函数及其多面体 ISoA，分别给出了问题 2 的两种不同解。利用这些算法，可以求得误差动力学 ($\mathbb{V}$ 及多面体 ISoA Ω 。基于这些结果，我们提出了一种系统化算法，将基于李雅普诺夫函数的 VI 程序整合其中，如算法 3.2所述。对于一般的 3，若其不属于 W_{tar}，该算法首先寻找与 $\mathcal{F}$最接近的 $W'_{tar} \in \mathcal{F}$，然后应用定理 W_{tar}来计算策略。计算得到的不变集 1 和 2 随后与 Ω 平移，并在必要时扩展以包含 W'_{tar}，从而得到 W_{tar}作为问题 $\mathcal{G}$的解。1。

算法3：面向多目标区间马尔可夫决策过程的鲁棒李雅普诺夫函数值迭代算

法 输入：多目标区间马尔可夫决策过程 $\underline{\Sigma} \Xi (X, x_0, U, \mathrm{P}, \mathrm{P}, \mathrm{R}, \mathrm{R})$ 和目标值 W_{tar}– 1 确定平稳策略集合 Π2 根据 (3.3) 确定所有 $(\pi, j) \in \bar{\Pi} \times \mathbb{D}$ 对应的 A_π^j 和 L_π^j ^ ^ –3 根据 (3.4) 计算所有 $\pi \in \Pi$ 对应的 A_π, B_π 4 将 $\mathcal{F}$ 定义为 (3.8) 中的 $\mathcal{F}$ 5 计算 $W'_{tar} := \arg\min_{W'}\{\|W' - W\|, W \in \mathcal{F}\}$6 选择满足 $W'_{tar} = (I - \hat{A}_\lambda)^{-1}\hat{B}_\lambda$ 的 $\lambda \in \Lambda_M$，并根据 (3.9) 计算 $\hat{A}_\lambda$ 和 $\hat{B}_\lambda$ 7 通过算法 1 或算法 2 计算多面体李雅普诺夫函数 $\mathbb{V}(E)$ 和多面体 ISoA Ω 8 计算满足 $\Omega + W'_{tar} \subset \mathcal{G}$ 和 $W_{tar} \in \mathcal{G}$ 的最小集合 $\mathcal{G}$ 9 设置 $W_0 = 0$ 和 $E_0 = W_0 - W'_{tar}$10 对于 $k \in \mathbb{N}$，执行 11 使用 (3.11) 计算 $\pi_k(E_k)$12 使用 (3.2) 更新 E_{k+1}，并对不确定性应用 opt 13 结束 输出：集合 $\mathcal{G}$ 和切换策略 $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$，使得 $W_{tar} \in \mathcal{G}, W_{\boldsymbol{\pi},\mathsf{opt}} \in \mathcal{G}$

4.4. 计算复杂度

算法的计算复杂度1. SMT 求解器 Z3/Z3py 无法为带量词的 SMT 提供封闭形式的最坏情况复杂度保证。该工具采用 DPLL(T) 算法，将针对布尔结构的 SAT 搜索与线性实数算术（LRA）的多项式时间检查相结合。因此， Z3/Z3py 的性能取决于所求解的具体实例。此外，即使没有量词，一旦引入析取，涉及线性不等式的 SMT 也可能变为 NP 完全。线性规划检查的时间复杂度是多项式的，但布尔搜索可能需要指数时间。考虑到合成器在步骤 $|\mathcal{E}| = |\mathcal{V}| = s$ 时的 s，我们将进行 $4|\bar{\Pi}|s^3$次布尔搜索，并执行 $(nq+d)s^2+s$次 LRA。因此，策略的数量会使复杂度呈指数级增长，而状态、目标和不确定性顶点的数量则会使复杂度呈多项式级增长。同样，在验证器的步骤 s 中，我们将进行 $s^4|\mathbb{D}|(1+|\mathbb{D}|)$次布尔¯搜索，并执行 $|\Pi| \times \max(s+1, nq)$ 次 LRA。因此，不确定性顶点的数量会使复杂度呈指数级增长，而状态、目标和策略的数量则会使验证器的复杂度呈多项式级增长。

算法的计算复杂度2。 析取约束（DJC）通常会被重构为混合整数问题进行优化。诸如 MOSEK 之类的软件会尝试用大M约束替代DJC，并在此过程中简化问题。求解混合整数问题所需的计算量会随问题规模呈指数级增长，因此这类问题是NP难的。例如，一个含有 n 个二进制变量的问题可能需要 2^n 次松弛（见（MOSEKApS，2025a，b））。在¯我们的设定中，合成器采用带析取约束的优化。在步骤 s 中将这些约束转换为MILP会产生 $(|\Pi|+3s+1)$个整数变量，导致复杂度随策略数量呈指数级增长。验证器分为三个阶段：（i）阶段 1 中的（4.6）可转换为线性规划，并可在多项式时间内求解。（ii）阶段 2 中的（4.7）包含析取约束。将其转换为MILP后，会产生 $s^2|\mathbb{D}|$ 个整数变量，由于不确定性集中的顶点，复杂度呈指数级增长。（iii）阶段 3 中的（4.8）同样包含析取约束，会产生 $s|\mathbb{D}|$ 个整数变量，并由于不确定性集中的顶点而导致复杂度呈指数级增长。因此，算法2 的复杂度会随着策略数量和不确定性集顶点数量的增加而显著上升。

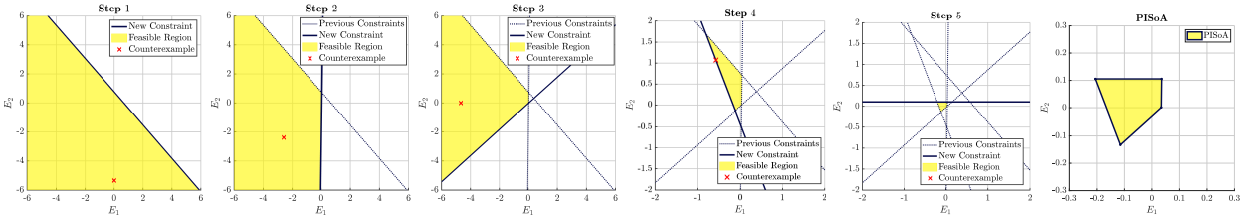


图3：回收机器人。 此图展示了算法 1 计算多面体吸引不变集的反例引导的归纳综合过程。(a) 学习器提出一个初始的多面体李雅普诺夫函数候选，由此生成一个可行区域（黄色），但验证器发现了一个反例（红色叉号）。(b)–(d) 每个反例都会生成一个新约束，用于细化可行区域；更新后的候选由验证器重新评估。(e) 经过五次迭代后，不再发现反例，这表明候选多面体李雅普诺夫函数满足所有必要条件。(f) 最终的多面体吸引不变集由所得可行区域定义。

5. 案例研究

我们将算法3 应用于三个案例研究。第一个是回收机器人（Sutton和Barto，2018年；Iervolino等人，2023年）的马尔可夫决策过程模型。第二个是改编自（Hahn等人，2019年；Monir等人，2025年）的区间马尔可夫决策过程模型。第三个案例研究是改编自（Thein和Chang，2014年）的电动汽车（EV）电池生命周期区间马尔可夫决策过程模型。

5.1. 基于马尔可夫决策过程模型的回收机器人

回收机器人模型中的转移概率矩阵 $P(\pi)$ 对于 $\pi \in \{1,\dots,6\}$ 为

$$P(1)=\begin{bmatrix}\beta & 1-\beta\\ 1-\alpha & \alpha\end{bmatrix},\quad P(2)=\begin{bmatrix}\beta & 1-\beta\\ 1 & 0\end{bmatrix},\quad P(3)=\begin{bmatrix}1 & 0\\ 1-\alpha & \alpha\end{bmatrix},\quad P(4)=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$
$$P(5)=\begin{bmatrix}0 & 1\\ 1-\alpha & \alpha\end{bmatrix},\quad P(6)=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix},$$

其中 $\alpha=0.7$ 和 $\beta=0.4$。奖励向量为 $R(1)=[\beta r_{search}-3(1-\beta);r_{search}]$、$R(2)=[\beta r_{search}-3(1-\beta);r_{wait}]$、$R(3)=[r_{wait};r_{search}]$、$R(4)=[r_{wait};r_{wait}]$、$R(5)=[0;r_{search}]$和 $R(6)=[0;r_{wait}]$，其中 $r_{search}=8$ 和$r_{wait}=2$。折扣因子 γ设为 0.5。我们使用与 $\lambda=[0;0;1;0;0;0]$ 对应的 W_{tar}，以及初始值 $W_0=[0;0]$。

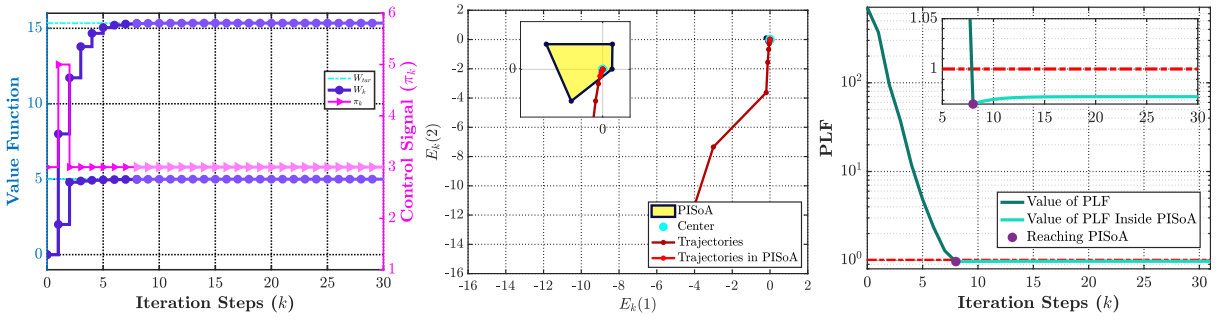
我们将算法 3 与算法 1作为子程序应用，以计算该模型的策略。CEGIS综合过程的中间步骤如图3所示。计算出多面体 ISoA 后，我们继续执行算法 3的其余步骤来计算策略。图 4 展示了价值函数的演变、得到的策略、误差轨迹向所计算出的多面体 ISoA 收敛的过程，以及多面体李雅普诺夫函数的下降情况。这些结果表明，值迭代能够收敛，轨迹保持在多面体 ISoA 内，且多面体李雅普诺夫函数下降至 $\rho=1$以下，随后保持在该值以下，从而证实了不变性。

5.2. 区间马尔可夫决策过程示例

图 5 展示了一个取自（Hahn等人，2019年）的区间马尔可夫决策过程模型，其状态集合为 $X=\{s,t,u\}$，初始状态为 s，动作集合为 $U=\{a,b\}$。图中展示了非零转移概率区间。策略集合为 $\Pi=\{1,2\}$。区间转移概率矩阵定义如下：

$$\underline{\mathrm{P}}(1)=\begin{bmatrix}0 & \frac{1}{3} & \frac{1}{10}\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix},\quad \overline{\mathrm{P}}(1)=\begin{bmatrix}0 & \frac{2}{3} & 1\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix},\quad \underline{\mathrm{P}}(2)=\begin{bmatrix}0 & \frac{2}{5} & \frac{1}{10}\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix},\quad \overline{\mathrm{P}}(2)=\begin{bmatrix}0 & \frac{3}{5} & 1\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

区间奖励向量定义为 $\underline{\mathrm{R}}(1)=\left[\frac{13}{10};\frac{3}{10};\frac{1}{10}\right]$、$\overline{\mathrm{R}}(1)=\left[5;\frac{3}{10};\frac{1}{10}\right]$、$\underline{\mathrm{R}}(2)=\left[\frac{1}{2};\frac{3}{10};\frac{1}{10}\right]$ 和 $\overline{\mathrm{R}}(2)=\left[\frac{8}{5};\frac{3}{10};\frac{1}{10}\right]$。折扣因子 $\gamma=0.7$ 和目标 W_{tar} 对应于 $\lambda=\left[\frac{9}{10};\frac{1}{10}\right]$，其初始值为 $W_0=[0;0;0]$。



(a) (b) (c)

图4：回收机器人。算法3的结果。（a）价值函数的迭代演化 W_k，其收敛于W_{tar}，以及综合策略 π_k。（b）误差轨迹 E_k 在计算所得策略下，均收敛至多面体 ISoA，此后始终保持在其中。（c）PLF 的演化过程单调递减，降至1以下，并保持低于1，从而验证了不变性。

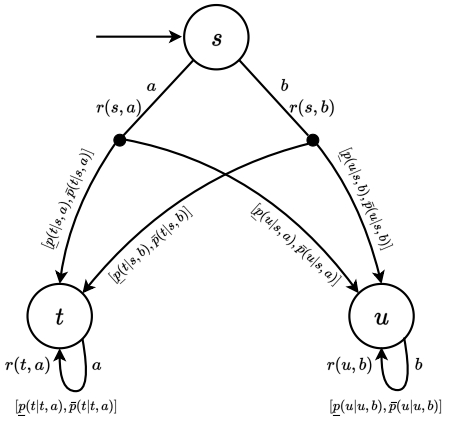


图5：IMDP示例，改编自（Hahn等人，2019年）。

我们使用算法3 ，并以算法 2 作为子程序，计算该模型的策略。所得的多面体 ISoA 如图 6所示。目标函数下界和上界的价值函数以及切换策略如图 7所示。图中还显示，计算得到的多面体李雅普诺夫函数降至 ρ以下，随后保持在 ρ以下，从而证实了不变性

5.3. 电动汽车电池生命周期

我们考虑了改编自（Thein和Chang，2014年）的锂离子电动汽车电池生命周期模型。我们将电动汽车电池的寿命终止阶段路径规划问题建模为一个如图8所示的折扣区间马尔可夫决策过程。该模型有六个状态：$X = \{S_0, S_I, S_R, S_M, S_C, S_D\}$。这些状态分别对应以下阶段：使用老化 ($S_0$)、检测 ($S_I$)、再利用 ($S_R$)、再制造 ($S_M$)、回收 ($S_C$)和处置 ($S_D$)。只有在决策状态$\{S_0, S_I\}$才能作出决策，其中状态 S_0 中的可用动作是 $\mathcal{A}_0 = \{$检测、再利用、再制造、回收$\}$，状态 S_I中的可用动作是 $\mathcal{A}_I = \{$再利用、再制造、回收、处置$\}$。在其余状态 $\{S_R, S_M, S_C, S_D\}$中不作出决策。为考虑建模误差和运行变化，转移概率矩阵中的非零元素最多存在 3% 变化。

奖励按每个动作和状态分配，并分为三类：(i) 经济成本：每个动作都有名义成本，反映所消耗的精力和资源（例如，再制造比再利用成本更高；回收和处置会产生处理及合规费用）。这些成本作为负奖励，对于每个特定动作保持不变。(ii) 健康惩罚：用于衡量电池预期的磨损；使电池继续或重新投入使用的动作（老化/再利用）所产生的惩罚，高于设施处理步骤（再制造、回收、处置）。我们允许根据不同状态进行小幅调整，以准确反映动作的应用场景。(iii)

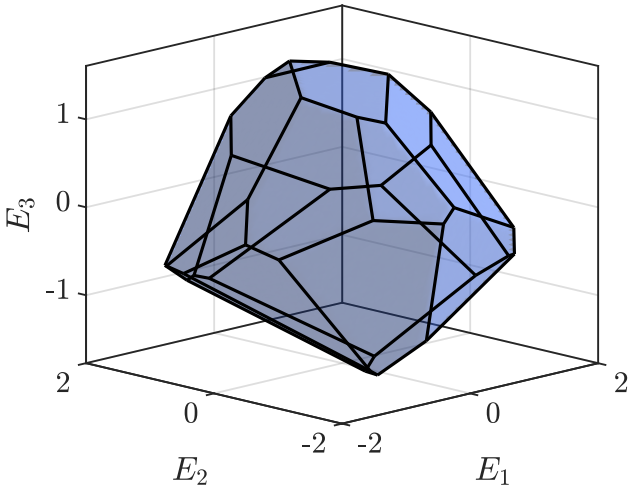


图 6：IMDP示例。使用算法2 在算法3中求得的 IMDP 示例误差动力学的多面体 ISoA。

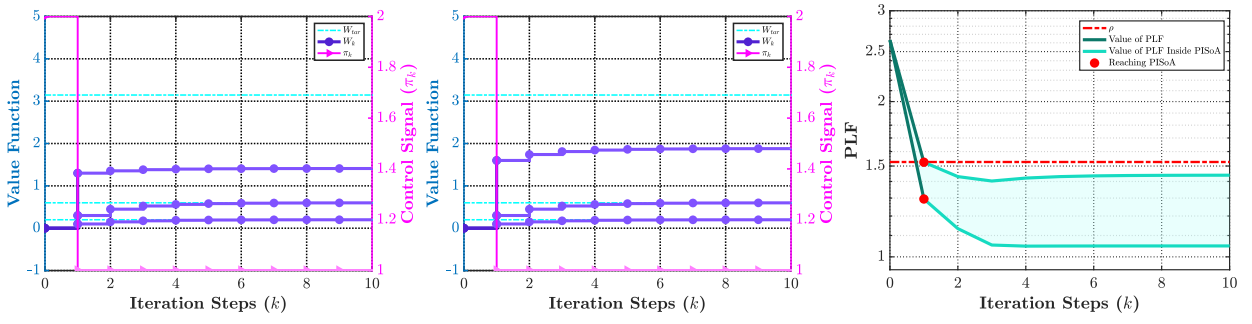


图 7：IMDP示例。算法3的结果。(a) 和 (b) 目标函数下界和上界对应的价值函数及所选策略的演化。(c) 这些界对应的多面体李雅普诺夫函数下降至 ρ以下，并保持在不变集内。

环境效益：循环利用类动作可获得正奖励（再利用 >再制造 >回收），而处置则为负奖励。每项奖励均采用基于动作的名义值，并设定对称的相对区间（0.01λ0.03% 以反映估计误差）。

我们采用三种实现：(i) SMT-PLF，使用基于 SMT 的反例引导归纳综合算法1来实例化算法3；(ii) Opt-PLF，使用基于优化的反例引导归纳综合算法2来实例化算法3；以及 (iii) Quadratic，该方法改编自基线论文中的二次李雅普诺夫函数（Monir等人，2024年）。我们考虑单目标、双目标和三目标三种场景，以评估目标维度增加带来的影响。定量结果汇总于表1，其中包括收敛行为、不确定性下的下界和上界，以及计算时间。

精度比较。在表1的所有目标上，基于多面体李雅普诺夫函数的方法 SMT-PLF 和 Opt-PLF 在选择相近策略的同时，达到或提升了基线 Quadratic 方法的准确性。对于 W_1–W_3，三种方法选出的策略 π 完全相同，报告的误差也相同。对于 W_4，在 SMT-PLF 和 Opt-PLF 下，$\|E\|$均有所降低，而 $\|E\|$ 略有增加；不过，总和 $\|E\| + \|E\|$ 有所降低，因此整体误差带收窄。在这些行中，Opt-PLF 达到的总误差小于 SMT-PLF。在 W_5 处，总误差大于 Quadratic（即相较基线，准确性有所下降），但在多面体类方法中，Opt-PLF 优于 SMT-PLF：Opt-PLF 的 $\|E\|$ 较小，$\|E\|$ 略大，而总和较小。

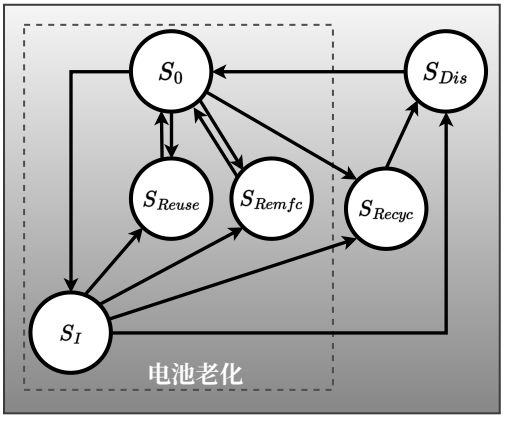


图8：电动汽车电池生命周期的马尔可夫决策过程模型。

表1 电动汽车电池生命周期。状态数为 $n = 6$，目标数量为 q。计算得到的策略为 π。对于每个目标数量 $q \in \{1, 2, 3\}$，考虑三个目标向量 W_{tar}（第 W_1 行至第 W_9 行）。误差下界和上界的范数列于 $\|\underline{\mathsf{E}}\|$、$\|\bar{\mathsf{E}}\|$列。PLF综合的计算时间分别为 $\mathcal{T}_1$（算法1） $\mathcal{T}_2$（算法2），值迭代（算法 $\mathcal{T}_3$）的计算时间为3。基线算法的计算时间为 $\mathcal{T}$，该算法给出一个二次函数。在第 W_6行中，Quadratic 块中的“–”表示基线优化问题不可行，因此未报告结果。

			SMT-PLF					Opt-PLF					Quadratic			
n	q	W_{tar}	π	$\|\underline{\mathsf{E}}\|$	$\|\bar{\mathsf{E}}\|$	$\mathcal{T}_1$ [s]	$\mathcal{T}_3$ [s]	π	$\|\underline{\mathsf{E}}\|$	$\|\bar{\mathsf{E}}\|$	$\mathcal{T}_2$ [s]	$\mathcal{T}_3$ [s]	π	$\|\underline{\mathsf{E}}\|$	$\|\bar{\mathsf{E}}\|$	$\mathcal{T}$ [s]
6	1	W_1	1	4.6817	4.6421	206.08	2.76	1	4.6817	4.6421	281.26	2.08	1	4.6817	4.6421	8.07
6	2	W_2	1	4.6968	4.7083	1091.07	3.08	1	4.6968	4.7083	1656.20	2.81	1	4.6968	4.7083	20.96
6	3	W_3	1	4.7017	4.7241	2387.70	3.84	1	4.7017	4.7241	4059.09	3.18	1	4.7017	4.7241	51.68
6	1	W_4	15	3.4865	4.6175	328.41	3.08	16	2.8859	4.669	449.54	2.83	11	4.7259	4.4079	15.40
6	2	W_5	15	3.5436	4.6541	1207.42	3.08	16	2.9546	4.7085	1922.10	3.19	1	14.8044	9.1324	103.11
6	3	W_6	15	3.5896	4.6852	2532.60	4.16	16	3.0453	4.7616	4478.60	4.14	–	–	–	–
6	1	W_7	2	2.8353	4.6752	234.73	1.87	4	2.3887	4.8574	316.89	1.12	3	2.9637	4.7976	7.42
6	2	W_8	2	2.9483	4.8541	1018.42	2.12	4	2.4144	4.8831	1727.63	2.12	3	2.9845	4.8236	18.66
6	3	W_9	3	2.9913	4.8291	2029.80	3.09	4	2.4247	4.8875	3485.20	3.09	3	2.9913	4.8291	69.37

对于 W_6，基线方法中的优化问题不可行，而 SMT-PLF 和 Opt-PLF 均返回了可行证书和策略；同样，尽管 Opt-PLF 的 $\|E\|$ 略大、$\|E\|$ 小于 SMT-PLF，但其总和更小。类似地，对于 W_7 和 W_8，Opt-PLF 的总误差小于 SMT-PLF，而 SMT-PLF 的总误差小于基线。最后，在 W_9 处，SMT-PLF 和 Quadratic 选择相同的策略并取得相同的误差，而 Opt-PLF 选择了不同的策略，且总误差更小。总之，两种多面体方法都利用几何灵活性，相较基线收窄了误差带；在多面体方法族中，Opt-PLF 的总误差始终最小，其次是 SMT-PLF。

计算时间对比。所需时间主要由两部分构成：通过算法进行 PLF 综合1或2，以及算法3中的值计算。表1 所报告的计算时间表明，每个实例的值计算仅需几秒，而 PLF 综合阶段占据了总运行时间的大部分。Quadratic 基线的执行时间最短，其次是Opt-PLF，而SMT-PLF耗时最长：$\mathcal{T} < \mathcal{T}_2 + \mathcal{T}_3 < \mathcal{T}_1 + \mathcal{T}_3$。较高的计算时间 $\mathcal{T}_1$ 源于SMT-PLF 算法，该算法迭代使用 SMT 求解器来综合策略并验证其是否满足要求。相比之下，Opt-PLF ($\mathcal{T}_2$) 通过交替求解混合整数线性规划来实现，而其规模通常以适中的速度增长。总之，采用多面体 ISoA 可利用几何灵活性提高准确性，但代价是综合时间更长。在两种多面体方法中，Opt-PLF 比 SMT-PLF准确性更高且耗时更少。

6. 结论

本文提出了一种利用多面体李雅普诺夫函数，为多目标马尔可夫决策过程综合策略的新方法。我们将相关的值迭代算法重新表述为具有区间不确定性的切换仿射系统，并应用控制理论定理综合策略，使系统轨迹趋向一个包含各目标目标值的不变集。该多面体函数适用于动态规划方程的仿射结构，能够更准确地处理区间不确定性。通过在切换仿射系统模型中增添不同目标，我们的方法无需进行耗费大量计算的帕累托前沿计算或其近似计算。涉及回收机器人和电动汽车电池应用的数值研究，证明了该策略综合方法在不确定性下的有效性。未来工作将考虑需要求解约束动态规划的优化问题，并结合带有量化误差界的抽象方法，通过缩减模型规模来提升可扩展性。

参考文献

Abate, A., David, C., Kesseli, P., Kroening, D., Polgreen, E., 2018. 模理论的反例引导归纳综合，载于：国际计算机辅助验证会议，Springer。第270–288页。Ahmadi, A.A., Jungers, R.M., 2016. 切换线性系统李雅普诺夫函数复杂度的下界。《非线性分析：混合系统》21，118–129。Ahmed, D., Peruffo, A., Abate, A., 2020. 使用SMT 求解器自动且可靠地综合李雅普诺夫函数，载于：系统构造与分析工具和算法：第26届国际会议，TACAS 2020，作为欧洲软件理论与实践联合会议ETAPS 2020的一部分举行，爱尔兰都柏林，2020年4月25日至30日，论文集，第I部分26，Springer。第97–114页。Albea Sanchez, C., Ventosa-Cutillas, A., Seuret, A., Gordillo, F., 2020. 不确定离散时间切换仿射系统的鲁棒切换控制设计。《国际鲁棒与非线性控制期刊》30，7089–7102。Barrett, C., Tinelli, C., 2018. 模理论可满足性，载于：《模型检验手册》。Springer，第305–343页。Bellman, R., 1957. 一种马尔可夫决策过程。《数学与力学杂志》，679–684。Berger, G.O., Jungers, R.M., Wang, Z., 2022. 黑箱切换线性系统的数据驱动不变子空间辨识，载于：2022年IEEE第61届决策与控制会议（CDC），IEEE。第32–37页。Berger, G.O., Sankaranarayanan, S., 2022. 从反例中学习固定复杂度多面体李雅普诺夫函数，载于：2022年IEEE第61届决策与控制会议（CDC），IEEE。第3250–3255页。Berger, G.O., Sankaranarayanan, S., 2023. 通过反例引导计算分段线性系统的多面体李雅普诺夫函数。《自动化》155，111165。Bertsekas, D.P., 2011. 近似策略迭代：综述与若干新方法。《控制理论与应用杂志》9，310–335。Bertsekas, D.P., et al., 2011. 《动态规划与最优控制》第3版，第II卷。马萨诸塞州贝尔蒙特：Athena Scientific 1。Blanchini, F., Miani, S., et al., 2008. 《控制中的集合论方法》。第78卷。Springer。Bloch, A.M., 1990a. 总体最小二乘问题的Kähler结构、Brockett 最速下降方程与约束流，载于：《系统理论中的实现与建模：MTNS-89 国际研讨会论文集》第I卷，Springer。第83–88页。Bloch, A.M., 1990b. 最速下降、线性规划与哈密顿流。《当代数学》AMS 114，77–88。Brockett, R., 1988. 对列表排序、矩阵对角化并求解线性规划问题的动力系统，载于：第27届IEEE决策与控制会议论文集，第799–803页，第1卷。Brockett, R.W., 1991. 对列表排序、矩阵对角化并求解线性规划问题的动力系统。《线性代数及其应用》146，79–91。Bynum, M.L., Hackebeil, G.A., Hart, W.E., Laird, C.D., Nicholson, B.L., Siirola, J.D., Watson, J.P., Woodruff, D.L., et al., 2021. Pyomo——Python中的优化建模。第67卷。Springer。Chatterjee, K., Majumdar, R., Henzinger, T.A., 2006. 多目标马尔可夫决策过程，载于：Durand, B., Thomas, W.（编），STACS 2006，Springer Berlin Heidelberg，柏林，海德堡。第325–336页。Conforti, M., Cornuéjols, G., Zambelli, G., 2014. 整数规划模型，载于：《整数规划》。Springer，第45–84页。Deaecto, G.S., Egidio, L.N., 2016. 离散时间切换仿射系统的实用稳定性，载于：2016年欧洲控制会议（ECC），IEEE。第2048–2053页。Deaecto, G.S., Geromel, J.C., 2017. 离散时间切换仿射系统的稳定性分析与控制设计。《IEEE 自动控制汇刊》62，4058–4065。Delgado, K.V., de Barros, L.N., Dias, D.B., Sanner, S., 2016. 不精确概率下马尔可夫决策过程的实时动态规划。《人工智能》230，192–223。Delimpaltadakis, G., Lahijanian, M., Mazo Jr, M., Laurenti, L., 2023. 连续动作空间下的区间马尔可夫决策过程，载于：第26届ACM混合系统：计算与控制国际会议论文集，第1–10页。Etessami, K., Kwiatkowska, M., Vardi, M.Y., Yannakakis, M., 2008. 马尔可夫决策过程的多目标模型检验。《计算机科学逻辑方法》4。Givan, R., Leach, S., Dean, T., 2000. 有界参数马尔可夫决策过程。《人工智能》122，71–109。Gurobi Optimization, LLC, 2024. 《Gurobi 优化器参考手册》。Ha, M., Wang, D., Liu, D., 2021. 用于折扣最优控制的广义值迭代及其稳定性分析。《系统与控制快报》147，104847。

Haddad, S., Monmege, B., 2018. 马尔可夫决策过程和区间马尔可夫决策过程的区间迭代算法。《理论计算机科学》735，111–131。《可达性问题2014：特刊》。Hahn, E.M., Hashemi, V., Hermanns, H., Lahijanian, M., Turrini, A., 2017. 区间马尔可夫决策过程的多目标鲁棒策略综合，载于：国际系统定量评估会议，Springer。第207–223页。Hahn, E.M., Hashemi, V., Hermanns, H., Lahijanian, M., Turrini, A., 2019. 多目标区间马尔可夫决策过程：从鲁棒策略到帕累托曲线。《ACM建模与计算机仿真汇刊》（TOMACS）29，1–31。Hart, W.E., Watson, J.P., Woodruff, D.L., 2011. Pyomo：使用Python对数学规划进行建模与求解。《数学规划计算》3，219–260。Helmke, U., Moore, J.B., 2012.《优化与动力系统》。施普林格科学与商业媒体。Iervolino, R., Tipaldi, M., Forootani, A., 2023. 值迭代算法的基于李雅普诺夫的版本及其离散时间切换仿射系统表述。《国际控制杂志》96，577–592。Jha, S., Gulwani, S., Seshia, S.A., Tiwari, A., 2010. 预言机引导的基于组件的程序综合，载于：第32届ACM/IEEE国际软件工程会议论文集，第1卷，第215–224页。Lavaei, A., Perez, M., Kazemi, M., Somenzi, F., Soudjani, S., Trivedi, A., Zamani, M., 2023. 离散时间随机控制系统的组合式强化学习。《IEEE开放控制系统期刊》。Mathiesen, F.B., Lahijanian, M., Laurenti, L., 2024. IntervalMDP.jl：用于区间马尔可夫决策过程的加速值迭代。arXiv预印本 arXiv:2401.04068。Monir, N., Sadabadi, M.S., Soudjani, S., 2025. 通过情景优化对不确定切换仿射系统进行鲁棒控制，载于：第7届年度动力学与控制学习会议，PMLR。第1460–1471页。Monir, N., Schön, O., Soudjani, S., 2024. 多目标区间马尔可夫决策过程的基于李雅普诺夫的策略综合。IFAC在线论文58，99–106。MOSEK ApS, 2025a.《MOSEK MATLAB优化工具箱》，11.0版。MOSEK ApS, 2025b.《MOSEK Python Fusion API手册》。11.0版。Nilim, A., El Ghaoui, L., 2005. 转移矩阵不确定的马尔可夫决策过程鲁棒控制。《运筹学》53，780–798。Polański, A., 2000. 基于多面体李雅普诺夫函数的绝对稳定性分析。《自动化》36，573–578。Rajendra, P., Brahmajirao, V., 2020. 通过深度学习对动力系统建模。《生物物理评论》12，1311–1320。Scheftelowitsch, D., Buchholz, P., Hashemi, V., Hermanns, H., 2017. 转移参数不确定的马尔可夫决策过程的多目标方法，载于：第11届EAI国际性能评估方法与工具会议论文集，第44–51页。Solar-Lezama, A., Rabbah, R., Bodík, R., Ebcioğlu, K., 2005. 面向位流程序的基于草图的编程，载于：2005年ACM SIGPLAN编程语言设计与实现会议论文集，第281–294页。Sun, Z., Ge, S.S., 2011.《切换动力系统稳定性理论》。Sutton, R.S., Barto, A.G., 2018.《强化学习：导论》。麻省理工学院出版社。Thein, S., Chang, Y.S., 2014. 电动汽车锂离子电池生命周期评估决策模型——以韩国智能电动巴士项目为例。《电源杂志》249，142–147。Tipaldi, M., Iervolino, R., Massenio, P.R., Forootani, A., 2025. 一种用于求解随机动态规划问题的数据驱动实用稳定化方法。《自动化》178，112372。Tsitsiklis, J., Van Roy, B., 1996. 函数逼近下时序差分学习的分析，载于：Mozer, M., Jordan, M., Petsche, T.（编），《神经信息处理系统进展》，麻省理工学院出版社。Tsitsiklis, J.N., 2002. 乐观策略迭代的收敛性。《机器学习研究杂志》3，59–72。Vielma, J.P., 2015. 混合整数线性规划建模技术。《SIAM评论》57，3–57。Williams, H.P., 2013.《数学规划中的模型构建》。约翰·威立父子出版公司。Wolff, E.M., Topcu, U., Murray, R.M., 2012. 具有时序逻辑规范的不确定马尔可夫决策过程鲁棒控制，载于：2012年第51届IEEE决策与控制会议（CDC），第3372–3379页。Wu, D., Koutsoukos, X., 2008. 使用有界参数马尔可夫决策过程对不确定系统进行可达性分析。《人工智能》172，945–954。Xie, W., Tang, W., Kuang, Y., 2022. 一种用于加速深度神经网络随机优化的新型混合优化器：动力系统视角。《神经计算》514，341–350。Yeung, E., Kundu, S., Hodas, N., 2019. 学习非线性动力系统Koopman算子的深度神经网络表示，载于：2019年美国控制会议（ACC），IEEE。第4832–4839页。

A. 定理1的证明

证明。PLF $\mathbb{V}(E)$ 为非负函数，这意味着

$$\forall E, \mathbb{V}(E) \geq 0 \;\Leftrightarrow\; \forall E, \max_{c_i \in \mathcal{V}} \left({c_i}^\top E - d_{c_i} \right) \geq 0 \;\Leftrightarrow\; \forall E,\, \exists c_i \in \mathcal{V},\, {c_i}^\top E - d_{c_i} \geq 0.$$

因此，如果参数满足条件 ψ_0，则

$$\psi_0 : \forall E,\, \exists c_i \in \mathcal{V},\, {c_i}^\top E - d_{c_i} \geq 0, \tag{A.1}$$

对于所有 E，PLF $\mathbb{V}(E)$ 都将是非负的。条件 ψ_0 的否定为

$$\neg\psi_0 : \exists E,\, \forall c_i \in \mathcal{V},\, {c_i}^\top E - d_{c_i} < 0. \tag{A.2}$$

对于定义2中的条件（C1），我们有

$$0 \in \Omega \;\Leftrightarrow\; \mathbb{V}(0) \leq 1 \;\Leftrightarrow\; \max_{c_i \in \mathcal{V}} (0 - d_{c_i}) \leq 1 \;\Leftrightarrow\; \forall c_i \in \mathcal{V},\, d_{c_i} \geq -1.$$

因此，当参数满足条件 ψ_1 as 时，条件 (C1) 成立

$$\psi_1 : \forall c_i \in \mathcal{V},\, d_{c_i} \geq -1. \tag{A.3}$$

ψ_1 的否定为

$$\neg\psi_1 : \exists c_i \in \mathcal{V},\, d_{c_i} < -1. \tag{A.4}$$

对于定义2中的条件 (C2)，我们有

$$\mathbb{V}(E) > 1 \to \Delta\mathbb{V}(E) < 0 \;\Leftrightarrow\; \max_{c_i \in \mathcal{V}} (c_i^\top E - d_{c_i}) > 1 \to$$
$$\min_{\pi} \max_{c_j \in \mathcal{V}} (c_j^\top (A_\pi E + L_\pi) - d_{c_j}) - \max_{c_\ell \in \mathcal{V}} (c_\ell^\top E - d_{c_\ell}) < 0,$$

对于所有不确定矩阵 $[A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa \in \mathbb{D}^\circ}$。等价地，我们有

$$\forall [A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa \in \mathbb{D}},\; \max_{c_i \in \mathcal{V}} (c_i^\top E - d_{c_i}) \leq 1 \;\vee$$
$$\min_{\pi} \max_{c_j \in \mathcal{V}} (c_j^\top (A_\pi E + L_\pi) < d_{c_j}) < \max_{c_\ell \in \mathcal{V}} (c_\ell^\top E - d_{c_\ell}). \tag{A.5}$$

基于 (A.5) 及凸性的性质，以下条件 ψ_2 满足定义 2中的条件 (C2)：

$$\psi_2 : \forall E,\, \exists \pi,\, [\forall c_i \in \mathcal{V}, (c_i^\top E - d_{c_i}) \leq 1] \vee$$
$$[\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, \exists c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} < c_\ell^\top E - d_{c_\ell})]. \tag{A.6}$$

条件 ψ_2 的否定为

$$\neg\psi_2 : \exists E,\, \forall \pi,\, [\exists c_i \in \mathcal{V}, (c_i^\top E - d_{c_i}) > 1] \;\wedge\; [\exists c_j \in \mathcal{V}, \exists \kappa \in \mathbb{D}, \forall c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} \geq c_\ell^\top E - d_{c_\ell})]. \tag{A.7}$$

最后，对于定义2中的条件（C3），我们有

$$\mathbb{V}(E) \leq 1 \to \mathbb{V}(A_\pi E + L_\pi) \leq 1 \;\Leftrightarrow\; \max_{c_i \in \mathcal{V}} (c_i^\top E - d_{c_i}) \leq 1$$
$$\to \min_{\pi} \max_{c_j \in \mathcal{V}} (c_j^\top (A_\pi E + L_\pi) - d_{c_j}) \leq 1,$$

对于所有不确定矩阵 $[A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa\in\mathbb{D}}$，这等价于

$$\forall [A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa\in\mathbb{D}}, \max_{c_i\in\mathcal{V}}(c_i^\top E - d_{c_i}) > 1 \ \vee \ \min_{\pi}\max_{c_j\in\mathcal{V}}(c_j^\top(A_\pi E + L_\pi) - d_{c_j}) \le 1. \tag{A.8}$$

利用（A.8）以及凸性性质，以下条件 ψ_3 满足定义2中的条件（C3）：

$$\psi_3 : \forall E, \exists \pi, \ [\exists c_i \in \mathcal{V}, (c_i^\top E - d_{c_i}) > 1] \ \vee \ [\forall \kappa \in \mathbb{D}, \forall c_j \in \mathcal{V}, c_j^\top(A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} \le 1]. \tag{A.9}$$

条件 ψ_3 的否定为

$$\neg\psi_3 : \exists E, \forall \pi, \ [\forall c_i \in \mathcal{V}, (c_i^\top E - d_{c_i}) \le 1] \ \wedge \ [\exists \kappa \in \mathbb{D}, \exists c_j \in \mathcal{V}, c_j^\top(A_\pi^\kappa E + L_\pi^\kappa) - d_{c_j} > 1]. \tag{A.10}$$

将定义2中的所有要求综合起来，可得

$$\Psi = \psi_0 \ \wedge \ \psi_1 \ \wedge \ \psi_2 \ \wedge \ \psi_3,$$

由此得到（4.1）中的表达式，其否定为

$$\neg\Psi = \neg\psi_0 \vee \neg\psi_1 \vee \neg\psi_2 \vee \neg\psi_3,$$

从而得到表达式（4.2），证明结束。 ■

B. 定理2的证明

证明。PLF $\mathbb{V}(E)$ 为非负，这意味着

$$\mathbb{V}(E) \ge 0 \ \Leftrightarrow \ \max_{c_i\in\mathcal{V}} {c_i}^\top E \ge 0 \ \Leftrightarrow \ \forall E, \exists c_i \in \mathcal{V}, {c_i}^\top E \ge 0.$$

因此，如果参数满足条件 ϕ_0，则

$$\phi_0 : \forall E, \exists c_i \in \mathcal{V}, {c_i}^\top E \ge 0, \tag{B.1}$$

对于所有 E，PLF $\mathbb{V}(E)$ 都将是非负的。条件 ϕ_0 的否定为

$$\neg\phi_0 : \exists E, \forall c_i \in \mathcal{V}, {c_i}^\top E < 0. \tag{B.2}$$

对于定义2中的条件（C1），有

$$0 \in \Omega \Rightarrow \mathbb{V}(0) \le \rho \Rightarrow \max_{c_i\in\mathcal{V}} 0 \le \rho \Rightarrow 0 \le \rho,$$

根据假设，该条件已经 $\rho \ge 0$得到满足。对于定义 2中的条件（C2），有

$$\mathbb{V}(E) > \rho \to \Delta\mathbb{V}(E) < 0 \ \Leftrightarrow \ \max_{c_i\in\mathcal{V}}(c_i^\top E) > \rho \to \min_{\pi}\max_{c_j\in\mathcal{V}}(c_j^\top(A_\pi E + L_\pi)) - \max_{c_\ell\in\mathcal{V}}(c_\ell^\top E) < 0,$$

对于所有不确定矩阵 $[A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa\in\mathbb{D}}$。等价地，我们有

$$\forall [A_\pi, L_\pi] \in \mathrm{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa\in\mathbb{D}}, \max_{c_i\in\mathcal{V}}(c_i^\top E) \le \rho \ \vee \ \min_{\pi}\max_{c_j\in\mathcal{V}}(c_j^\top(A_\pi E + L_\pi)) < \max_{c_\ell\in\mathcal{V}}(c_\ell^\top E). \tag{B.3}$$

利用（B.3）以及余 凸性性质，以下条件 ϕ_2 满足定义中的条件（C2） 2:

$$\phi_2 : \forall E, \exists \pi, [\forall c_i \in \mathcal{V}, (c_i^\top E) \le \rho] \ \vee \ [\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, \exists c_\ell \in \mathcal{V}, (c_j^\top(A_\pi^\kappa E + L_\pi^\kappa) < c_\ell^\top E)]. \tag{B.4}$$

条件 ϕ_2 的否定是

$$\neg\phi_2 : \exists E, \forall \pi, [\exists c_i \in \mathcal{V}, (c_i^\top E) > \rho] \ \wedge [\exists c_j \in \mathcal{V}, \exists \kappa \in \mathbb{D}, \forall c_\ell \in \mathcal{V}, (c_j^\top(A_\pi^\kappa E + L_\pi^\kappa) \ge c_\ell^\top E)]. \tag{B.5}$$

最后，对于定义中的条件（C3）2，我们有

$$\mathbb{V}(E) \leq \rho \rightarrow \mathbb{V}(A_\pi E + L_\pi) \leq \rho \;\; \Leftrightarrow \;\; \max_{c_i \in \mathcal{V}} c_i^\top E \leq \rho \rightarrow$$
$$\min_{\pi} \max_{c_j \in \mathcal{V}} c_j^\top (A_\pi E + L_\pi) \leq \rho,$$

对于所有不确定矩阵 $[A_\pi, L_\pi] \in \text{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa \in \mathbb{D}}$。等价地，我们可以将其表示为

$$\forall [A_\pi, L_\pi] \in \text{Co}([A_\pi^\kappa, L_\pi^\kappa])_{\kappa \in \mathbb{D}}, \max_{c_i} c_i^\top E > \rho \;\vee\; \min_{\pi} \max_{c_j \in \mathcal{V}} c_j^\top (A_\pi E + L_\pi) \leq \rho. \tag{B.6}$$

应用（B.6）以及凸性的性质后，条件 ϕ_3 满足定义2中规定的条件（C3）：

$$\phi_3 : \forall E, \exists \pi, [\exists c_i \in \mathcal{V}, (c_i^\top E) > \rho] \vee [\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) \leq \rho]. \tag{B.7}$$

条件 ϕ_3 的否定为

$$\neg\phi_3 : \exists E, \forall \pi, [\forall c_i \in \mathcal{V}, (c_i^\top E) \leq \rho] \wedge [\exists c_j \in \mathcal{V}, \exists \kappa \in \mathbb{D}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) > \rho]. \tag{B.8}$$

将定义2中的所有要求合并起来，我们得到

$$\Phi = \phi_0 \wedge \phi_2 \wedge \phi_3 = \forall E, \exists \pi, \{(\exists c_i \in \mathcal{V}, {c_i}^\top E \geq 0) \wedge ([\forall c_i \in \mathcal{V}, (c_i^\top E) \leq \rho] \vee$$
$$[\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, \exists c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) < c_\ell^\top E)]) \wedge ([\exists c_i \in \mathcal{V}, (c_i^\top E) > \rho] \vee$$
$$[\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) \leq \rho])\},$$

利用布尔恒等式 $(A \vee B) \wedge (\neg A \vee C) = (A \wedge C) \vee (\neg A \wedge B)$，我们得到

$$\Phi = \forall E, \exists \pi, (\exists c_i \in \mathcal{V}, {c_i}^\top E \geq 0) \wedge \{([\forall c_i \in \mathcal{V}, (c_i^\top E) \leq \rho] \wedge [\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) \leq \rho]) \vee$$
$$([\exists c_i \in \mathcal{V}, (c_i^\top E) > \rho] \wedge [\forall c_j \in \mathcal{V}, \forall \kappa \in \mathbb{D}, \exists c_\ell \in \mathcal{V}, (c_j^\top (A_\pi^\kappa E + L_\pi^\kappa) < c_\ell^\top E)])\},$$

这就是式（4.3）中的表达式，其否定为

$$\neg\Phi = \neg\phi_0 \vee \neg\phi_2 \vee \neg\phi_3,$$

利用布尔恒等式可得式（4.4），从而完成证明。 ■