arXiv:2302.00171v2 [cs.RO] 1 Nov 2023

面向安全高效交互规划的主动不确定性降低：屏蔽感知双重控制方法

期刊标题 XX(X)：1–25©作者 2022 转载与许可：
sagepub.co.uk/journalsPermissions.nav DOI:
10.1177/ToBeAssigned
www.sagepub.com/SAGE

Haimin Hu[1], David Isele[2], Sangjae Bae[2], 和 Jaime F. Fisac[1]

摘要
在交互场景中（如人机交互和多机器人协作任务），准确预测他人行为的能力对于机器人系统的安全与效率至关重要。然而，机器人往往无法获取可能影响预测的关键信息，例如其他智能体的目标、注意力和合作意愿。双重控制理论通过将预测模型的未知参数视为随机隐状态，并利用系统运行期间收集的信息在运行时推断其取值，来应对这一挑战。尽管双重控制能够以最优且自动的方式权衡探索与利用，但对于一般的交互式运动规划而言，其计算难以处理，主要原因是机器人的轨迹规划与对其他智能体意图的预测之间存在根本性的耦合。本文提出一种新颖的算法方法，基于隐式双重控制范式，实现交互式运动规划中的主动不确定性降低。我们的方法依赖于随机动态规划的基于采样的近似，从而将问题转化为模型预测控制问题，可通过实时的基于梯度的优化方法直接求解。结果表明，对于同时具有连续不确定性和类别不确定性的一大类预测模型，该策略能够保留双重控制效应。为确保交互智能体安全运行，我们采用运行时安全过滤器（也称为“安全屏蔽”方案）；当安全关键事件即将发生时，该过滤器会用安全回退策略覆盖机器人的双重控制策略。随后，我们利用近期提出的屏蔽感知鲁棒规划方案的改进变体扩展双重控制框架，主动权衡名义规划性能与低概率智能体行为所引发的高代价紧急机动风险。我们通过模拟驾驶研究以及使用 1/10 比例自动驾驶车辆进行的硬件实验，展示了该方法的有效性。

关键词
规　　划中的不确定性、人机交互、双重控制理论、随机模型预测控制、安全学习　　g.

1 引言

计算能够考虑与一个或多个其他智能体可能发生交互的机器人规划是一项颇具挑战的任务，因为机器人系统与其他智能体可能具有耦合动力学、通信能力有限且利益冲突。不确定性下交互规划的例子包括人机交互 Fisac et al. (2018a); Sadigh 等人（2018）; Bajcsy 等（2021）、多机器人协作 Leonard et al. (2007); Tokekar et al. (2016); Santos et al. (2018)、群体机器人 Swain et al. (2011); Rubenstein et al. (2014); Chung et al. (2018)、自动驾驶赛车 Liniger et al. (2015); Kabzan et al. (2019); Schwarting et al. (2021)，以及混合自动化交通 Isele (2019); Bae et al. (2020); Wu et al. (2021)。为了在这些场景中实现安全与高效，机器人必须能够准确预测其他智能体的行为，并无缝地适应其行为。意图驱动行为模型广泛用于此类预测：例如，玻尔兹曼模型 Luce (1959); Ziebart 等人 (2008) 常用于预测有限理性决策者的运动。该模型假设，其他智能体采取底层效用更高的动作的可能性呈指数级增加。如果给定的效用函数能够准确刻画其他智能体的意图，则可将交互建模为动态博弈，并通过动态规划求得参与者的反馈策略 Fisac 等人 (2019)。然而，典型的交互场景中可能存在大量先验上 看似合理的意图（例如，对应于不同均衡解的意图Peters 等人 (2020）或其他智能体的不同偏好 Sadigh 等人 (2018)，而机器人通常无法对这些意图进行完整建模，更无法观测到它们 Fisac 等人(2018b)。机器人可以尝试通过参数模型来表示其他智能体的意图，并在贝叶斯框架下推断这些参数的取值，将其视为隐状态 Fisac 等人 (2018b)； Tian 等人 (2022)； Hu 等人 (2022)，但如何在规划交互过程的同时，以可计算的方式实现这一点，仍是一个开放问题。

采用闭环贝叶斯推断的多阶段轨迹优化通常可归结为一个随机最优控制问题。隐状态随机控制的一个重要方面是，所计算的策略是否会产生所谓的双重控制效应 Feldbaum (1960);Bar-Shalom和Tse(1974); Mesbah (2018);也就是说，在

[1]普林斯顿大学电气与计算机工程系，普林斯顿，新泽西州 08544，美国[2]本田研究院，圣何塞，加利福尼亚州 95134，美国

通讯作者： Haimin Hu，普林斯顿大学电气与计算机工程系，普林斯顿，新泽西州 08544，美国 电子邮箱：haiminh@princeton.edu

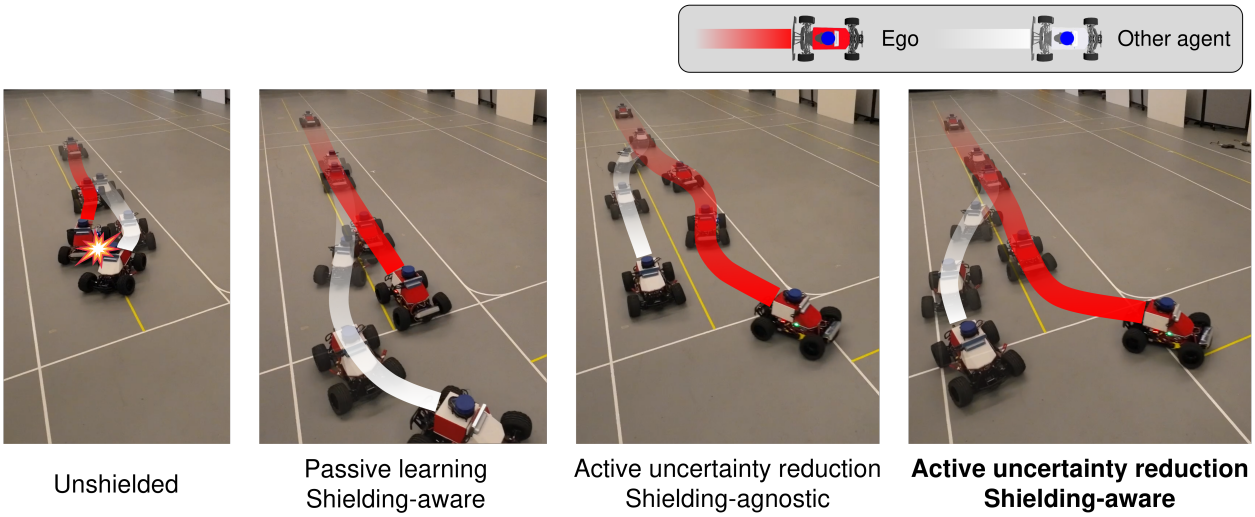


图1. 安全与不确定性降低是交互规划任务的两个核心组成部分，例如自动驾驶。在此，自车（红色车头）试图超越另一辆自动驾驶汽车（白色车头），而自车并不知道其控制策略。此例中的另一辆车被设定为合作型智能体，如果与自车足够接近，就会让行。四次试验中，自车与其他智能体之间的初始间距大致相同。左侧：缺乏形式化安全保证的无安全屏蔽规划器无法保证操作过程中不会发生碰撞。中左：安全屏蔽感知规划器会推理未来可能发生的安全屏蔽事件，并尽可能避免依赖安全覆盖。然而，缺乏主动不确定性降低可能导致自车智能体过于保守，无法完成超车任务。中右：结合主动不确定性降低的与屏蔽无关策略会采取激进的超车操作，并触发“险些碰撞”紧急覆盖，代价是性能和舒适性下降。右侧：我们提出的控制框架将主动不确定性降低与安全屏蔽感知规划相结合，兼顾安全与效率，从而显著提升闭环系统的性能。

交互规划的情境中，机器人是否会主动寻求降低对其他智能体隐状态的不确定性。双重随机最优控制问题的求解方法可分为显式方法Heirung等人(2015年); S adigh 等人(2018年); Tian 等人(2021年)，这类方法通过某种启发式探测来重新表述问题；以及隐式方法 Bar-Shalom和Tse(1974); Klenske和Hennig (2016年); Arcari 等人 (2020a)，这类方法利用随机动态规划直接求解控制问题。尽管显式双重控制问题通常比其隐式对应方法更容易表述和求解，但设计探测项并调整其权重因子并非易事，还可能导致性能不稳定。关于双重控制方法的全面综述，请参阅Mesbah (2018).

贡献：本文在随机最优控制框架下，对一大类交互式规划问题进行了建模，并提出一种采用隐式双重随机模型预测控制（SMPC）的近似求解方法。由此得到的策略能够自动权衡探索与利用的成本，使机器人在不牺牲预期规划性能的前提下，主动降低对其他智能体隐状态的不确定性。我们提出的随机模型预测控制问题同时支持连续不确定性和类别不确定性，并可使用现成的实时非线性优化求解器求解。据我们所知，这是首个无需显式信息收集策略或目标函数即可实现主动不确定性降低的交互式运动规划框架。

为了为机器人系统和其他智能体提供形式化安全保证，我们将双重控制策略与安全屏蔽结合使用Hsu 等人（2023），这是一种监督安全滤波器方案，可通过安全备用策略覆盖双重控制器。我们改进了Hu 等人（2022）提出的近期安全屏蔽感知鲁棒规划（SHARP）框架。该框架通过推理其他智能体的低概率动作可能触发的未来安全屏蔽事件来生成高效运动；我们在此基础上，明确考虑其他智能体对自车智能体探测动作的响应。我们提出了一种易于优化的凸约束，可在局部刻画由高成本安全机动支配的系统演化，并将其纳入隐式双重随机模型预测控制公式。图1展示了我们方法的核心思想。 我们的关键洞见是，安全屏蔽感知双重控制策略在主动与其他智能体交互、收集有用信息的同时，也始终警惕因代价高昂的安全屏蔽覆盖而导致效率损失的风险。

本工作的初步版本由 Hu 和 Fisac（2022）在 *2022* 年国际机器人算法基础研讨会（*WAFR*）上发表。在这篇修订和扩展后的论文中，我们补充了以下贡献：（i）将研究扩展到一般的交互规划问题（这是原论文所研究的人机交互规划问题的超类）；（ii）更深入地解释、推导和分析所提出的框架，包括更详细地讨论其他智能体行为模型的选择以及所提策略的性质；（iii）纳入屏蔽感知鲁棒规划方案，该方案提供鲁棒安全保证，并协调安全覆盖与双重控制探测动作之间的冲突；（iv）使用 Waymo 开放运动数据集补充仿真结果Sun 等人（2020）；（v）在 1/10 比例自动驾驶车辆上进行硬件演示。

表1。 对明确建模智能体之间耦合运动的部分交互规划方法进行比较。

特征	Schildbach 和 Borrelli (2015年)	Sadigh 等人. (2018年)	Peters 等人. (2020)	Tian 等人. (2021年)	Sunberg 和 Kochenderfer (2022)	**Ours**
主动不确定性降低	N	**Y**	N	**Y**	**Y**	**Y**
自动探索-利用权衡	N/A	N	N/A	**Y**	**Y**	**Y**
安全性保证	**Y**	N	N	N	N	**Y**
连续状态空间	**Y**	**Y**	**Y**	N	**Y**	**Y**
连续动作空间	**Y**	**Y**	**Y**	N	N	**Y**
可扩展至多个对手	**Y**	N	**Y**	N	**Y**	**Y**
可完全在线计算策略	**Y**	**Y**	**Y**	N	**Y**	**Y**
博弈论的	N	**Y**	**Y**	**Y**	N	**Y**

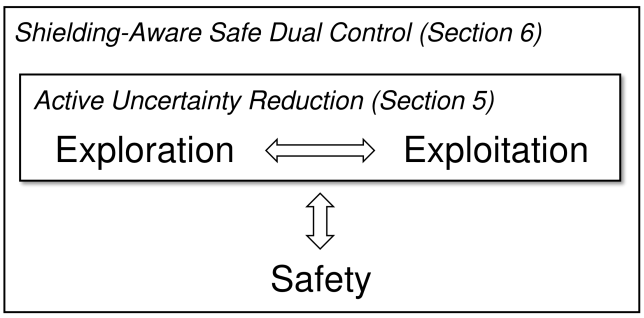


图 2. 本文研究的交互规划任务要素及其关系。在第 5 节中，我们专注于基于双重控制的主动不确定性降低，通过平衡探索与利用来提高规划效率。在第 6 节中，我们介绍屏蔽感知安全双重控制，以协调安全与效率之间可能存在的冲突。

2 相关工作

本节回顾交互规划方法的不同方面。表1比较了明确考虑自车机器人与其他智能体之间交互的规划策略。图2展示了我们方法的示意图。

2.1 双重控制与部分可观测马尔可夫决策过程

涉及识别其他智能体由未知意图所驱动的行为的机器人运动规划问题，可以建模为混合可观测性马尔可夫决策过程Bandyopadhyay 等人(2013，这是著名的部分可观测马尔可夫决策过程（POMDP）的一种变体Ong 等人(2009。历史上，控制理论和机器学习领域的研究者分别以不同术语、在不同应用领域中研究双重控制和POMDP，但后来发现二者具有相同的基本思想Dayan 和 Sejnowski(1996)；Bhambri 等人(2022)。由于双重控制与 POMDP 之间存在内在联系，通常可以将一个问题转化为另一个问题。在Sehr 和 Bitmead(2017中，一项对偶控制问题被重新表述为POMDP，并通过动态规划求解；然而，这种方法仅适用于低维问题。另一方面，研究者已开发出计算效率更高、依赖蒙特卡洛树搜索（MCTS）的算法Silver 和 Veness(2010)；Somani 等人(2013)；Sunberg 等人(2017)，用于近似求解 POMDP。

一般而言，使用基于 MCTS 的 POMDP 求解方法来解决对偶控制问题，需要对状态、控制量和观测进行采样（因而离散化）。近期的研究Sunberg 和 Kochenderfer(2022在早期算法Sunberg 和 Kochenderfer(2018的基础上，将经典的基于 MCTS 的 POMDP 求解方法扩展到连续状态与观测空间。然而，采用这种方法处理连续动作空间仍是一个开放问题，其理论上的信息获取特性也尚不明确Sunberg 和 Kochenderfer(2018)。本文则采用轨迹优化来解决对偶控制问题，从而能够直接在状态和控制量的连续空间中进行优化，并且所得控制策略具有可证明的信息获取特性。当对机器人与人类之间的交互进行显式建模时，POMDP 的表述便成为部分可观测随机博弈（POSG）（通常难以求解）Sadigh 等人(2018)。我们的方法可视为一种新的计算高效框架，用于求解表述为 POSG 的交互规划问题。

2.2 人在回路运动规划

人机交互，或称人在环规划，是交互规划中一个新兴且迅速发展的子领域。传统的人机交互规划算法通常采用两阶段流程：首先使用预测模型推断人类未来的运动，再将其输入运动规划器。在 Fisac 等人(2018b)中，利用玻尔兹曼理性模型预测人类未来的轨迹，同时通过贝叶斯推断估计模型置信度。随后，运动规划器利用这些信息获得概率无碰撞保证。之后，Tian 等人(2022)将该框架扩展为显式考虑博弈论模型，其中还会推断人类的角色，从而使规划表现不那么保守。尽管计算效率较高，上述两阶段流程将推断与规划割裂开来，可能导致规划过于保守。近期，研究者致力于通过求解联合预测规划问题，构建更集成的框架。在Sadigh 等人(2018)中，作者提出将人机交互建模为动力系统，在该系统中，机器人的动作会影响人类的动作，从而预测人类未来的状态。接下来几年中相关的

研究提出将该联合问题作为一般和动态博弈来求解 Fisac 等人(2019)；Fridovich-Keil 等人(2020)；Zanardi 等人(2021)。近期研究 Hu 等人(2023)也考虑了联合预测规划设定，作者专注于信念空间中的安全性分析，并利用哈密顿-雅可比（HJ）可达性构建了信念空间中的零和博弈。我们的方法属于联合预测与规划这一类。

2.3 双重随机模型预测控制

由于能够处理安全关键约束和通用不确定性模型，随机模型预测控制已广泛用于不确定性下的机器人运动规划。开创性工作Bernardini 和 Bemporad (2011年提出了一个利用情景优化求解随机模型预测控制问题的通用框架。在 Schildbach 和 Borrelli (2015年)，研究者提出了一种随机模型预测控制方法，用于在存在其他人类驾驶车辆时为自车提供变道辅助；该方法预测这些车辆未来的运动，并将其作为有限数量的情景纳入模型预测控制公式中。随机模型预测控制也用于交通交叉路口的自动驾驶Nair 等人(2021年)，其中利用高斯混合模型预测其他车辆的轨迹。在Chen 等人(2022年)，研究者提出了一种基于情景的随机模型预测控制算法，以捕捉非受控人类智能体的多模态反应行为。在Hu 等人(2022年)，研究者开发了一种可证明安全的随机模型预测控制规划器，用于鲁棒的人类感知机器人运动规划；该规划器会主动平衡预期性能与风险：低概率的人类行为可能触发高代价的紧急安全机动。然而，所有这些随机模型预测控制方法都无法产生双重控制效应——机器人只会被动地适应对其他智能体运动的预测，而不会主动试探对方以收集更多信息并降低其不确定性。在 Arcari 等人 (2020a)，研究者提出了一种隐式双重随机模型预测控制方法，用于同时存在参数不确定性和结构不确定性的非线性动力系统的最优控制。在Bonzanini 等人(2020)，研究者构建了一种鲁棒的基于情景的随机模型预测控制，以实现非线性动力系统的安全学习；该系统的动力学部分未知，并基于高斯过程模型从数据中学习。研究者采用了一种新颖的状态和输入相关情景树，以考虑不确定性模型对决策变量的依赖关系，从而实现主动不确定性降低并提升闭环性能。本文基于这些双重随机模型预测控制方法，旨在实现交互规划中的主动不确定性降低。

2.4 主动信息收集

迄今为止，大多数交互规划方法都遵循“被动自适应”范式。在 Peters 等人 (2020) 中，多智能体交互规划问题被建模为具有均衡不确定性的一般和微分博弈。机器人首先推断其他智能体处于哪种均衡状态，然后使自身策略与推断出的均衡解相一致。近期，主动信息收集这一概念受到机器人学领域，尤其是人机交互领域的关注；在概念上，它与双重控制非常相似。在 Sadigh 等人 (2018) 中，额外的信息收集项被添加到机器人的名义目标函数中，并在轨迹优化框架下用于在线估计人类状态-动作价值函数中的未知参数。事实上，根据 Mesbah(2018年) 提出的分类，Sadigh 等人 (2018) 的方法可归类为显式双重控制方法，该方法需要启发式地设计探测机制，并权衡优化预期性能与降低人类未知参数的不确定性两者的相对重要性。与此相反，我们提出了一种隐式双重控制方法，它能够自动平衡性能与不确定性降低，因此无需设计启发式信息收集机制。

2.5 通过安全屏蔽实现安全交互规划

在不确定性下提升运动规划安全性的一种常见方法，是采用机会约束或概率安全规划 Schildbach 和 Borrelli（2015）；Fisac 等（2018b）；Bastani 等人（2021）。然而，即便采用概率安全规划策略，当其他智能体采取低概率动作时，仍可能发生安全违规。这也被称为不太可能事件的“长尾”问题 Koopman（2018）。通过运行时安全过滤器（通常称为安全屏蔽），可以保证交互规划始终安全 Hsu 等人（2023）。在这种范式中，反应式安全回退策略被用作“最后手段”，仅在碰撞等安全关键事件迫在眉睫时才会取代名义策略。常用的屏蔽机制包括 HJ 可达性分析 Mitchell 等人（2005）；Bansal et al. (2017)；Chen 等人（2021）、控制屏障函数 Ames 等人（2016）；Robey 等人（2020）；Lindemann 等人（2021）、模型预测控制 Li 和 Bastani（2020）；Wabersich and Zeilinger (2021)，以及李雅普诺夫方法 Chow 等人（2018）。在 Li 和 Bastani（2020）中，鲁棒模型预测控制被用作备用策略，为强化学习过程提供高概率安全性保证。关于屏蔽机制的统一观点，可参见近期综述 Hsu 等人（2023）。尽管安全屏蔽能够有效保障安全，但有时会降低规划效率，因为安全控制器的设计通常未考虑任务完成时间、轨迹平滑度或功耗等性能因素。为缓解这一问题，Hu 等人（2022）提出了屏蔽感知鲁棒规划（SHARP）框架，主动权衡名义规划性能与由其他智能体的低概率行为所触发的高代价紧急机动。本文将 SHARP 融入隐式双重随机模型预测控制，以实现安全且高效的交互规划；同时，通过明确考虑其他智能体对自车智能体探测动作的响应，放宽了“其他智能体忽略自车智能体”这一过于保守的假设，从而对该方法加以改进。

3 预备知识

3.1 多智能体动力系统

我们考虑一类离散时间输入仿射动力学，用以描述自车机器人系统（e）与一个或多个其他智能体（o）之间的交互，例如附近的人类，

$$x_{t+1} = f(x_t) + B^e(x_t)u_t^e + B^o(x_t)u_t^o + d_t, \qquad (1)$$

其中，$x_t=(x_t^e,\ x_t^o)\in\mathbb{R}^n$ 是联合状态向量，$u_t^e\in\mathcal{U}^e\subseteq\mathbb{R}^{m_e}$ 和 $u_t^o\in\mathcal{U}^o\subseteq\mathbb{R}^{m_o}$ 分别是自车与其他智能体的控制向量，$f:\mathbb{R}^n\to\mathbb{R}^n$ 是描述动力学自主部分的非线性函数，$B^e:\mathbb{R}^n\to\mathbb{R}^{n\times m_e}$ 和 $B^o:\mathbb{R}^n\to\mathbb{R}^{n\times m_o}$是可依赖于状态的控制输入矩阵，而 d_t 是表示外部扰动输入（例如风）和建模误差的加性不确定项，我们假设其服从高斯分布：$d_t\sim\mathcal{N}(\mathbf{0},\ \Sigma^d)$，独立同分布。

注释 1。动力学 (1) 通过连接多个其他智能体的状态和控制向量来对其进行建模o_1，o_2，..... 即 $x^o=(x^{o_1},\ x^{o_2},\ .....)$ 以及 $u^o=(u^{o_1},\ u^{o_2},\ .....)$。

3.2 智能体行为建模

在本文中，我们将其他智能体在每个时刻的动作 t 参数化为一种随机策略：

$$u_t^o:=\textstyle\sum_{i=1}^{n_\theta}\theta_i^M u_i^{M,o},\quad \text{subject to}\quad u_t^o\in\mathcal{U}^o,\qquad (2)$$

它是 随机基策略 $u^{M\text{的线性组合},\ i\ o}$其参数为 $\theta^M:=(\theta_1^M,\ \theta_2^M,\ .....,\ \theta_{n_\theta}^M)\in\mathbb{R}^{n_\theta}$。我们进一步允许每个基策略 $u^{M,\ i\ o}$ 具有不同的 模式M，这些模式取自一个 有限集合 $\mathcal{M}$，用于表示其他智能体的不同类别行为（例如合作、非合作或无意识）。我们使用认知科学中的“带噪理性”玻尔兹曼模型 $u^{M,\ i\ o}$ 来定义每个基策略 Luce(1959)。在该模型下，其他智能体根据概率分布选择每个基策略：

$$u_i^{M,o}\sim p\left(u_i^{M,o}\mid x,u^e;M\right)=\frac{e^{Q_i^M\left(u_i^{M,o};x,u^e,\right)}}{\int_{\tilde{u}^o\in\mathcal{U}^o}e^{Q_i^M(\tilde{u}^o;x,u^e)}d\tilde{u}^o},$$

对于所有 $i=1$、2、.....、n_θ 和 $M\in\mathcal{M}$。这里，$Q_i^M($ $u^{M\text{、}\ i\ o}$; x、u^e) 是与第 i 个基策略及众数 M 相关的其他智能体的基状态-动作价值（或 Q 值）函数。该模型假设，对于一对固定的 (i、M)，其他智能体选择使状态-动作价值函数最大化的动作的可能性呈指数级增加。

注释 2。理性决策者（例如机器人）是噪声理性智能体的一种特殊情况，当它们根据 $u^{M\text{、}\ i\ o}=\arg\max Q_i^M(\cdot)$ 选择动作时亦如此；也就是说，在选择基策略 $u^{M\text{、}\ i\ o}$ 时不存在不确定性（模糊性），但模型参数 (θ^M,M) 和外部扰动输入 d 中的不确定性仍然存在。因此，我们的框架可以自然地涵盖理性智能体。

注释 3。我们的方法不依赖于确定其他智能体的状态-动作价值函数 $Q_i^M(\cdot)$、参数 θ^M 和 M的具体方法；这些通常由系统设计者根据领域知识指定，或通过先验数据学习得到。基于目标的运动预测模型在相关文献中已有成熟研究。例如，参见*Sadigh* 等人 (*2018*年）；*Ziebart* 等人(*2008*年）。下面给出两个示例。

示例 1.（采用未知策略的自主智能体）。考虑在遥控汽车测试赛道（图*1* 和*14*）上，自车需要超越一辆由基于意图且经优化的策略控制的机器人汽车，而自车并不知道该策略。根据第*3.2*节，我们将其他智能体的状态-动作价值函数作为

$$u_t^o=\theta_{\text{tr}}^M u_{\text{tr}}^M\left(x_t,u_t^e\right)+\theta_{\text{sa}}^M u_{\text{sa}}^M\left(x_t,u_t^e\right),\ \ s.t.\ u_t^o\in\mathcal{U}^o,$$

其中 $\theta^M:=(\theta_{\text{tr}}^M,\ \theta_{\text{sa}}^M)$，基策略 $u_{\text{tr}}^M(\cdot)$ 和 $u_{\text{sa}}^M(\cdot)$分别体现其他智能体的跟踪目标（例如跟随参考轨迹）和安全目标（例如避免与自车发生碰撞）。将其他智能体对安全的重视程度建模为一个连续体的做法受到近期研究*Toghi* 等人*(2021*年，*2022*）的启发。此外，其他智能体有两种不同的模式，即是否愿意通过变道向自车让行，分别记为 $M\in\{\text{Y},\ \text{NY}\}$。图*3*展示了由不同模式和基策略配置所建模的其他智能体行为。

示例2。（博弈论人机交互）。在第二个示例中，我们考虑一个涉及一辆自动驾驶车辆（自车）和一名行人（其他智能体）的交通路口场景。我们使用模型(2) 及参数 $\theta^M:=(\theta_{\text{C}}^M,\ \theta_{\text{NC}}^M)$对行人进行建模，将其视为博弈论决策者；该参数刻画了其他智能体的合作程度。合作型智能体也会优化机器人的目标函数，而非合作型智能体则不会。我们将离散模式 M 定义为其他智能体针对机器人表现出的不同（博弈论）交互行为，这与*Bandyopadhyay* 等人*(2013)*；*Tian* 等人*(2022)*的做法类似。具体而言，其他智能体的状态-动作价值函数定义如下：

$$Q_i^M=\begin{cases}Q_i^M\left(u_i^{M,o};x_t,u_t^{e,Nash}(x_t)\right), & if\ M=\text{N},\\ Q_i^M\left(u_i^{M,o};x_t,u_t^{e,worst}(x_t)\right), & if\ M=\text{p},\\ Q_i^M\left(u_i^{M,o};x_t,u_t^{e,best}(x_t)\right), & if\ M=\text{w},\\ Q_i^M\left(u_i^{M,o};x_t^o\right), & if\ M=\text{o}.\end{cases}$$

在前三种模式中，其他智能体分别假设机器人的控制 u_t^e 是一个（局部）反馈纳什均衡解 *Basar* 和 *Olsder*（*1998*）（$M=$*N*）、使 $Q_i^M(\cdot)$最小的最坏情况解（即具有$M=$*p*的受保护型智能体），以及使 $Q_i^M(\cdot)$最大的最好情况解（即具有 $M=$*w*的乐观型智能体）；最后一种模式采用与 *Fisac* 等（*2018b*）、*Bajcsy* 等（*2021*）和 *Hu* 等（*2022*）相同的假设，即其他智能体没有意识到机器人的存在，并忽略机器人（$M=o$）。

3.3 推断模型参数

一般而言，动作模型 (2) 中的参数 θ^M 和众数 M 是机器人未知的隐状态。因此，只能根据过去的观测推断它们。为此，我们定义信息向量 $\mathcal{I}_t:=\left[x_t,u_{t-1}^e,\mathcal{I}_{t-1}\right]$，将其视为机器人在时间 $t\geq 0$能够因果观测到的所有信息的集合，其中 $\mathcal{I}_0=[x_0]$。接着，我们将信念状态 $b_t:=p\left(\theta^M,\ M\mid\mathcal{I}_t\right)$ 定义为在给定 $\mathcal{I}_t$的条件下 (θ^M,M) 的联合分布，而 $b_0:=p\left(\theta^M,M\right)$ 是给定的先验分布。当自车智能体接收到新的观测 $x_{t+1}\in\mathcal{I}_{t+1}$时，当前信念状态 b_t 将

通过递归贝叶斯推断方程进行更新：

$$p(\theta_-^M \mid \mathcal{I}_{t+1}; M) = \frac{p(x_{t+1} \mid u_t^e, \mathcal{I}_t; \theta^M, M) p(\theta^M \mid \mathcal{I}_t; M)}{p(x_{t+1} \mid u_t^e, \mathcal{I}_t; M)}, \tag{3a}$$

$$p(M_- \mid \mathcal{I}_{t+1}) = \frac{p(x_{t+1} \mid u_t^e, \mathcal{I}_t; M) p(M \mid \mathcal{I}_t)}{p(x_{t+1} \mid u_t^e, \mathcal{I}_t)}, \tag{3b}$$

$$b_{t+1}^- := p(\theta_-^M \mid \mathcal{I}_{t+1}; M) p(M_- \mid \mathcal{I}_{t+1}), \tag{3c}$$

$$b_{t+1} = g^t(b_{t+1}^-) := \int p(\theta^M, M \mid \tilde{\theta}_-^M, \tilde{M}_-) \cdot p(\tilde{\theta}_-^M, \tilde{M}_- \mid \mathcal{I}_{t+1}) d\tilde{\theta}_-^M d\tilde{M}_-. \tag{3d}$$

其中 b_{t+1}^- 是利用似然$p(x_{t+1} \mid u_t^e,\ \mathcal{I}_t;\ M)$更新后的信念状态，$p(\theta^M,\ M \mid \tilde{\theta}_-^M,\ \tilde{M}_-)$是一个转移模型，而 $g^t(\cdot)$ 是信念状态转移动力学。我们可以将式(3a)-(3d)简洁地改写为一个动力学系统，

$$b_{t+1} = g(b_t, x_{t+1}, u_t^e). \tag{4}$$

遗憾的是，系统 (4) 通常无法以解析形式表示一步以上的演化。即使先验分布 b_t 是高斯分布，后验 b_{t+1} 也不再是高斯分布，因为其他智能体由 (2) 定义的动作（进而影响观测 x_{t+1}）通常是非高斯的，因此无法利用高斯分布的共轭先验性质。在第 5 节中，我们将介绍一种计算高效的方法，近似传播信念状态动力学。

4 问题陈述

4.1 规范交互规划问题

现在，我们定义本文要解决的核心问题：典型交互规划问题，其形式化为如下随机有限时域最优控制问题：

$$\min_{\Pi^e} \mathop{\mathbb{E}}_{\substack{(\theta^M, M)\sim b_{[0:N-1]},\\ u^o_{[0:N-1]}\sim(2), d_{[0:N-1]}}} \sum_{k=0}^{N-1} \ell\left(x_k, \pi_k^e(x_k, b_k)\right) + \ell_F(x_N) \tag{5a}$$

$$\text{s.t.} \quad x_0 = \hat{x}_t,\ b_0 = \hat{b}_t, \tag{5b}$$

$$\forall k = 0, \ldots, N-1:$$

$$x_{k+1} = f(x_k) + B^e \pi_k^e + B^o u_k^o + d_k, \tag{5c}$$

$$b_{k+1} = g\left(b_k, x_{k+1}, \pi_k^e(x_k, b_k)\right), \tag{5d}$$

$$x_k \notin \mathcal{F}, \qquad \forall k = 0, \ldots, N \tag{5e}$$

其中，$\Pi^e = \{\pi_0^e(\cdot) \ldots, \pi_{N-1}^e(\cdot)\}$ 是通过优化式（5）得到的策略序列^；^x_t 和 b_t 分别是现实时间t测得的状态和维护的信念状态；与决策变量相关的预测时间用 k 索引；ℓ：$\mathbb{R}^n \times \mathcal{U}^e \to \mathbb{R}_{\geq 0}$ 和 ℓ_F：$\mathbb{R}^n \to \mathbb{R}_{\geq 0}$分别是设计者指定的阶段代价函数和终端代价函数；$\pi_k(x_k, b_k)$是因果反馈策略，Bar-Shalom和Tse（1974）；Mesbah（2018），该策略利用未来状态$x_{[k,\ldots,N]}$ 和信念状态 $b_{[k,\ldots,N-1]}$（尚待获取）的知识；$\mathcal{F} \subseteq \mathbb{R}^n$ 是失效集，状态不得进入其中。

暂时先忽略安全约束（5e），并将如何高效地施加该约束的讨论留到第6节。现在，可以使用随机动态规划贝尔曼（1966）求解问题（5）。通过贝尔曼递归，可以从后向前计算得到最优机器人的价值函数（最小剩余代价）$V_k(x_k, b_k)$和控制策略 $\pi^{e,\ *}_k(x_k,\ b_k)$，

$$V_k(x_k, b_k) = \min_{\pi_k^e(x_k, b_k)} \ell(x_k, \pi_k^e) + \mathop{\mathbb{E}}_{\substack{(\theta^M, M)\sim b_k,\\ u_k^o\sim(2), d_k}} \left[V_{k+1}(x_{k+1}, b_{k+1}) \mid \mathcal{I}_k\right]$$

$$\text{s.t. } (5c) - (5e) \tag{6}$$

终端条件为 $V_N(x_N,\ b_N) = \ell_F(x_N)$。

4.2 双重控制效应

价值函数 $V_t(x_t,\ b_t)$ 通过求解 (6) 得到，它取决于未来的信念状态 $b_{t'}$ $(t' > t)$，因此最优策略能够影响其他智能体的未来不确定性，该不确定性由信念状态量化。因此，($u^{e,\ *}_t$) 的最优策略具有6所述的双重控制效应，其形式化定义见定义1。根据最优性原理，贝尔曼(1966)，该策略在优化机器人的期望性能目标 (5a) 与主动降低其对其他智能体的不确定性之间实现了最优平衡。换言之，(6) 的最优策略会自动探测其他智能体以降低其不确定性，但仅在这样做能够提升机器人的期望闭环性能时才会如此。

定义 1。双重控制效应。如果控制输入能够以非零概率影响，

(a) 隐状态变量的至少一个 r 阶（$r \geq 2$）中心矩 *Feldbaum(1960)*；*Bar-Shalom和Tse (1974)*；*Mesbah (2018年)*，或 *(b)* 类别隐状态的熵*Hijab (1984)*。

4.3 近似双重控制

遗憾的是，除了最简单的情况外，式（6）在计算上都难以处理，这主要是因为需要对机器人的动作进行嵌套优化并计算条件期望。通过对信念状态进行量化，可以将式（6）中的期望项近似到任意精度；然而，这会导致计算量呈指数增长，即贝尔曼（1966）所述的维数灾难问题。正因如此，求解双重控制问题时主要采用近似方法。近似双重控制可分为两类：显式方法，例如Heirung等人（2015）、Sadigh 等人（2018）和 Tian 等人（2021），通过人为地将探测效应或信息收集目标引入控制策略，简化原始随机最优控制问题；以及隐式方法，例如 Bar-Shalom和Tse（1974）、Klenske和Hennig（2016）和 Arcari 等人（2020a），依赖于对贝尔曼递归（6）的直接近似。本文采用的是一种基于场景的隐式双重控制方法，下一节将对此进行详细介绍。采用隐式双重控制近似的主要优势在于，策略的自动探索-利用权衡能够以最优的方式自然保留 Sehr 和 Bitmead（2017）；Mesbah（2018）。

5 利用隐式双重控制主动减少不确定性

本节将介绍一种隐式双重控制方法，用于近似求解典型交互规划问题（5）。我们首先介绍一种近似方案，以便以可处理的方式传播信念状态动力学，并计算贝尔曼递归（6）中的期望。这对于将式（6）重新表述为一个可实时求解的随机模型预测控制问题至关重要。所提出的随机模型预测控制问题的表述及所得策略的性质将在第6节中详细介绍。

5.1 参数仿射动力学

为了以可处理的方式传播信念状态动力学（4），我们提出将联合动力学（1）重新表述为以下参数仿射形式：

$$x_{t+1} = F(x_t, u_t^e)\theta^M + \bar{f}(x_t, u_t^e) + \bar{d}_t, \tag{7}$$

其中，（$B^o(x_t)\, u_t^o$）中其他智能体的控制项替换为 $F(x_t,\ u_t^e)\,\theta^M$，它关于参数θ^M呈线性；而 $\bar{d}_t$ 是一个零均值高斯随机变量，正如我们将要看到的，它用于合并（d_t）中的高斯不确定性1以及其他智能体动作的不确定性。这是推导第5.2节近似信念更新规则的关键组成部分。

为了获得具有（7）形式的高斯参数仿射动力学，我们依赖的主要技术工具是拉普拉斯近似（Bishop 2006，第4章）。具体而言，每个基策略 $u^{M,\,i}{}^{o}$ 的条件概率分布近似为：

$$p\left(u_i^{M,o} \mid x_t, u_t^e; M\right) \approx \mathcal{N}\left(\mu_i^M(\cdot), \Sigma_i^M(\cdot)\right), \tag{8}$$

其中，基策略的均值函数为

$$\mu_i^M(x_t, u_t^e) := \arg\max_{u_i^{M,o} \in \mathcal{U}^o} Q_i^M\left(u_i^{M,o}; x_t, u_t^e\right),$$

基策略的协方差函数为

$$\Sigma_i^M(x_t, u_t^e) := -\nabla^2_{u_i^{M,o}}\, Q_i^M\left(u_i^{M,o}; x_t, u_t^e\right)^{-1}\Big|_{\mu_i^M(x_t, u_t^e)}.$$

上述拉普拉斯近似方案背后的直觉是，在（8）中得到的高斯分布以原始基策略分布 $p\left(u_i^{M,o} \mid x_t, u_t^e\right.$的众数 μ_i^M（$x_t,\ u_t^e$）为中心；$\left.M\right)$这对应于与 θ_i^M相关联的其他智能体的完全理性动作。我们在附录 C.1中讨论基 Q 值函数 Q_i^M（·）的最大化。给定 θ^M 和 M时，整体的近似动作分布为：

$$p\left(u_t^o \mid x_t, u_t^e; \theta^M\right) \approx \mathcal{N}\left(\mu_t^{u^o}(\cdot), \Sigma_t^{u^o}(\cdot)\right) \tag{9}$$

其中，其他智能体策略的均值函数为

$$\mu_t^{u^o}(x_t, u_t^e; \theta^M) := \sum_{i=1}^{n_\theta} \theta_i^M \mu_i^M(x_t, u_t^e),$$

以及其他智能体策略的协方差函数为

$$\Sigma_t^{u^o}(x_t, u_t^e; \theta^M) := \sum_{i=1}^{n_\theta} \left(\theta_i^M\right)^2 \Sigma_i^M(x_t, u_t^e).$$

随后，我们放宽对以下条件的要求 $u_t^o \in \mathcal{U}^o$ 以保持 u_t^o 服从正态分布，并在信念传播过程中，采用投影后的 u_t^o 进行状态演化。我们将在附录中讨论如何处理预测 u_t^o的无界支撑集C.2。根据上文得到的 u^o协方差函数，我们可以计算扰动项 d_t 在参数仿射动力学（7）中的协方差

$$\Sigma_t^{\bar{d}}(x_t, u_t^e; \theta^M) := \Sigma^d + B^o \Sigma_t^{u^o}(x_t, u_t^e; \theta^M) B^{o\top}, \tag{10}$$

它刻画了由高斯外部扰动 d_t（见1）以及其他智能体的带噪理性动作 u_t^o（由式9）所产生的综合不确定性。

为得到参数仿射动力学，最后一步是将其他智能体各基策略的均值函数拼接起来，定义均值矩阵：

$$U^o(x_t, u_t^e) := \left[\mu_1^M(x_t, u_t^e) \quad \dots \quad \mu_{n_\theta}^M(x_t, u_t^e)\right]. \tag{11}$$

将（2）和（11）代入（1可得：

$$x_{t+1} = \underbrace{B^o(x_t)U^o(x_t, u_t^e)}_{=:F(x_t, u_t^e)}\theta^M + \underbrace{f(x_t) + B^e(x_t)u_t^e}_{=:\bar{f}(x_t, u_t^e)} + \bar{d}_t, \tag{12}$$

其形式为参数仿射动力学（ 7).

注 4. 即使动力学（12）关于参数呈线性 θ^M，协方差矩阵 $\Sigma_t^{\bar{d}}$对 θ^M的依赖性，如（10）所示，仍然使得无法以闭式更新信念状态。为此，我们通过固定 θ^M，用估计值 $\bar{\theta}^M$对 $\Sigma_t^{\bar{d}}$进行近似。本文中，我们使用基于滚动模拟的方法估计 $\bar{\theta}^M$：将其设为在附录 θ^M 所述初始化流程的第 *3* 步中计算得到的 *C.3*的条件分布的均值。实际应用中，用点估计 $\bar{\theta}^M$替换随机变量 θ^M不可避免地会在协方差 $\Sigma_t^{\bar{d}}$中引入误差。

注 5. 替代动作模型。本文主要使用模型（2）（基策略的线性组合）在仿真研究和实验中预测其他智能体的动作。尽管如此，我们强调，我们的方法适用于任何能够产生参数仿射动力学 的动作模型（7）。这里，我们提供另一种属于此类的动作模型。与（*Bobu* 等人，*2020*，第 *IV.A* 节）类似，我们将其他智能体的状态-动作价值函数参数化为基函数的线性组合：*

$$Q_\theta^M\left(u^o; x, u^e, \theta^M\right) = \sum_{i=1}^{n_\theta} \theta_i^M Q_i^M\left(u^o; x, u^e\right),$$

以及与之相关的带噪声理性玻尔兹曼模型：

$$u_t^o \sim p\left(u^o \mid x, u^e; \theta^M\right) = \frac{e^{Q_\theta^M\left(u^o; x, u^e, \theta^M\right)}}{\int_{\tilde{u}^o \in \mathcal{U}^o} e^{Q_\theta^M\left(\tilde{u}^o; x, u^e, \theta^M\right)} d\tilde{u}^o}.$$

我们再次利用拉普拉斯近似，在局部将上述模型近似为 $p\left(u_t^o \mid x_t, u_t^e;\ \theta^M\right) \approx \mathcal{N}\left(\mu_t^{u^o}(x_t, u_t^e;\ \theta^M)\ \Sigma_t^{u^o}(x_t, u_t^e;\ \theta^M)\right)$.接下来我们进行一

*在Bobu 等人（2020）中，作者使用了仅依赖于状态的基代价（即价值）函数的线性组合，而非状态-动作价值函数，来对其他智能体的行为建模。

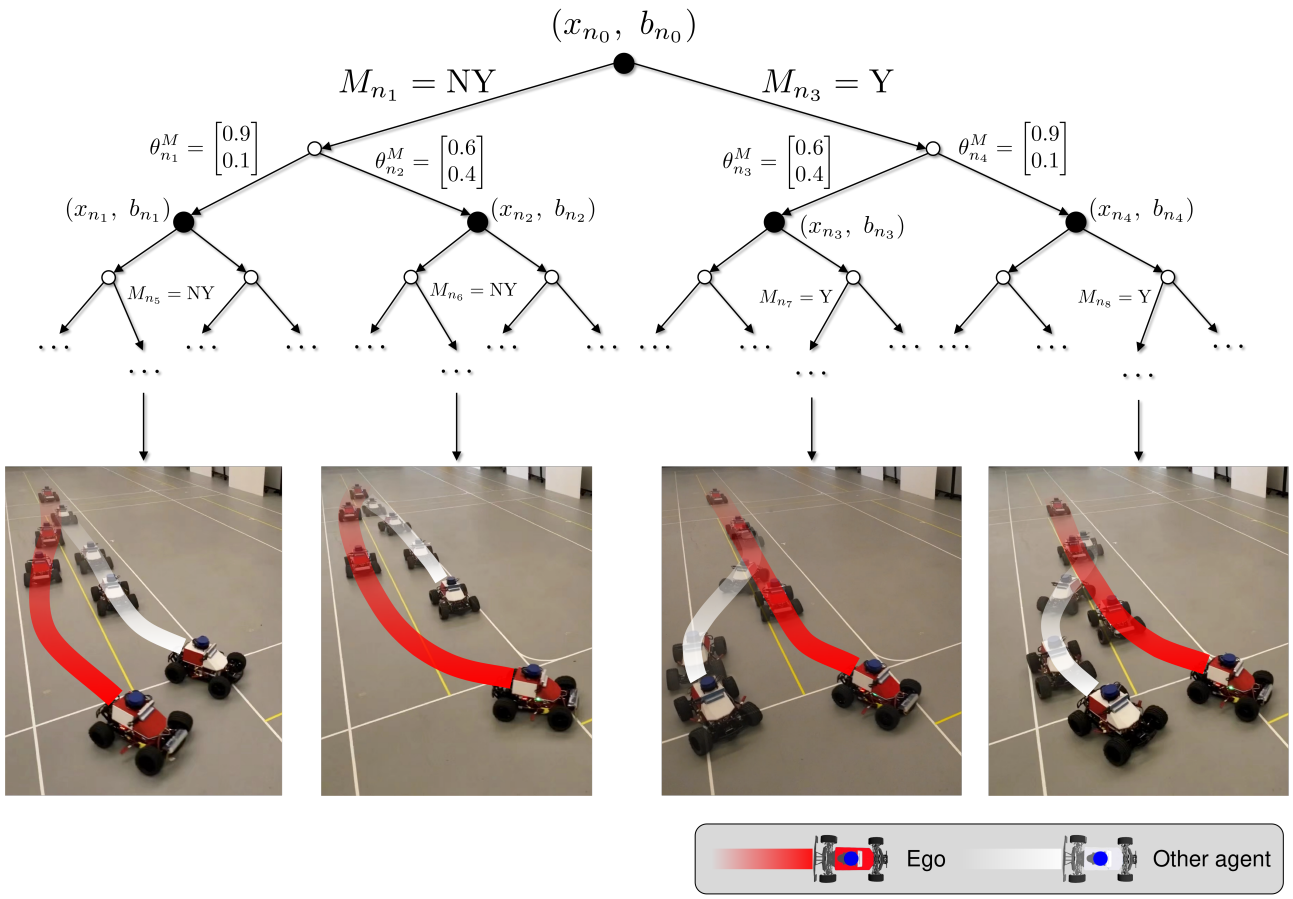


图3。 示例 1 的优化情景树示意图。自车和其他车辆分别以红色车头和白色车头表示。与节点 $\tilde{n}$ 相关联的不确定性样本 $M_{\tilde{n}}$ 和 $\theta^M_{\tilde{n}}$（见第 5.3 节中的情景树符号说明）用于近似计算随机目标函数 (20a)，并沿每个场景分支传播信念状态。白色圆圈 ○表示由一个 $M_{\tilde{n}}$ 样本部分确定的中间节点，较大的黑色圆圈 ● 表示由 $M_{\tilde{n}}$ 和 $\theta^M_{\tilde{n}}$样本共同完全确定的节点。左侧场景： 其他智能体未让行，且不太倾向于避免与自车发生碰撞（采样的连续隐状态 $\theta^M_{\text{tr}} \gg \theta^M_{\text{sa}}$）。左中场景： 其他智能体未让行，但减速以避免与自车发生碰撞。右中场景： 其他智能体让行，并采取更大的转弯幅度以避免撞上自车。右侧场景： 其他智能体让行，但不太倾向于避免与自车发生碰撞。

围绕给定的 $\bar{\theta}^M$ 进行一阶泰勒展开：

$$\begin{aligned}\mu_t^{u^o}(\cdot) &\approx \nabla_{\theta^M}\mu_t^{u^o}\left(x,u^e;\theta^M\right)\Big|_{\bar{\theta}^M}\delta\theta^M + \bar{\mu}_t^{u^o}(x_t,u_t^e;\bar{\theta}^M)\\ &=: U^o(x,u^e)\delta\theta^M + \bar{\mu}_t^{u^o}(x_t,u_t^e;\bar{\theta}^M),\end{aligned}$$

其中 $\delta\theta^M := \theta^M - \bar{\theta}^M$ 是误差项。这将得到式（7）形式的参数仿射动力学，其中 $F(x_t,u_t^e) = B^o(x_t)U^o(x_t,u_t^e)$ 和 $f(x_t,u_t^e) = f(x_t) + B^e(x_t)u_t^e + B^o(x_t)\bar{\mu}_t^{u^o}(x_t,u_t^e;\bar{\theta}^M) - F(x_t,u_t^e)\bar{\theta}^M$。

5.2 信念更新的易处理重构

根据第 5.1 节中的参数仿射动力学，我们现在可以推导出信念状态动力学（4）的可处理递归更新规则。由于 θ^M 的支撑集是连续的，我们将其条件信念建模为高斯分布，即 $p(\theta^M \mid \mathcal{I}_t;\ M) \sim \mathcal{N}(\mu_t^{\theta^M},\ \Sigma_t^{\theta^M})$。同样，$p(M \mid \mathcal{I}_t)$ 由于支撑集是离散的，因此将其建模为分类分布 $\mathcal{M}$。

我们的核心思路是利用高斯分布的自共轭性质，高效地更新 $p(\theta^M \mid \cdot)$（Bishop 2006，附录B）；也就是说，给定高斯先验 $p(\theta^M \mid \mathcal{I}_t;\ M)$，如果似然 $p(x_{t+1} \mid u_t^e, \mathcal{I}_t;\ \theta^M, M)$ 为高斯分布，那么后验 $p(\theta^M_- \mid \mathcal{I}_{t+1};\ M)$ 也为高斯分布，其均值和协方差是状态 x_t 及自车动作 u_t^e的解析函数。

以下引理概述了 $p(\theta^M \mid \cdot)$ 的近似信念更新过程，其证明见附录B.1。

引理 *1.* 对于高斯噪声动力学 *(7)* 以及式 *(8)* 中其他智能体的策略分布 $p\left(u_t^o \mid x_t, u_t^e;\ \theta^M\right)$，如果 θ^M的先验分布为 $p(\theta^M \mid \mathcal{I}_t;\ M) \sim \mathcal{N}(\mu_t^{\theta^M}, \Sigma_t^{\theta^M})$，那么后验分布为 $p(\theta^M_- \mid \mathcal{I}_{t+1};\ M) \sim \mathcal{N}(\mu_{t+1}^{\theta^M_-}, \Sigma_{t+1}^{\theta^M_-})$，其均值和协方差为

$$\begin{aligned}\mu_{t+1}^{\theta^M_-} &= \Sigma_{t+1}^{\theta^M_-}\left[F\left(x_t,u_t^e\right)^\top\left(\Sigma_{t+1}^x\right)^{-1}\left(x_{t+1} - \bar{f}(x_t,u_t^e)\right)\right.\\ &\qquad\left. + \left(\Sigma_t^{\theta^M}\right)^{-1}\mu_t^{\theta^M}\right]\\ \Sigma_{t+1}^{\theta^M_-} &= \left[\left(\Sigma_t^{\theta^M}\right)^{-1} + F\left(x_t,u_t^e\right)^\top\left(\Sigma_{t+1}^x\right)^{-1}F\left(x_t,u_t^e\right)\right]^{-1}\end{aligned}$$

where $\Sigma_{t+1}^x := \Sigma^d + B^o(x_t)\Sigma_t^{u^o}(x_t,u_t^e;\theta^M)B^o(x_t)^\top$.

重要的是，引理1中的近似信念状态动力学，是在近似求解随机动态规划（6）时保留双重控制效应的必要步骤。直观来看，机器人控制量 u_t^e 会进入协方差矩阵的更新

方程 $\Sigma_{t+1}^{\theta^M}$。因此，u_t^e 会影响$\Sigma_{t+1}^{\theta^M}$，并通过影响未来状态 $x_{t'}\,(t'>t+1)$，隐式影响所有未来的协方差矩阵，从而根据定义1，为 θ^M 产生双重控制效应。使用引理1 中的近似信念状态动力学，由 SMPC 策略产生双重控制效应的正式证明见第6 节（定理1）。对于 M，通过对似然关于 θ^M 边缘化，再应用贝叶斯规则，即可轻松计算测量更新（3b），与Arcari 等人(2020a）类似。我们将近似信念状态动力学简写为

$$b_{t+1}=\tilde{g}\left(b_t,x_{t+1},u_t^e\right), \tag{13}$$

这将在后续的随机模型预测控制中用作动力学约束，以实现计算上可行的信念传播。

5.3 动态情景树

本节中，我们提出一种近似求解典型交互规划问题（5）的方法，该方法采用基于情景树的随机模型预测控制（ST-SMPC）Bernardini 和 Bemporad (2011)，从而得到具有双重控制效应的控制策略。ST-SMPC 的核心思想是基于不确定性样本（即量化信念状态）来近似贝尔曼递归（6）中的期望。由此得到的情景树使我们能够将（6）展开为确定性有限时域最优控制问题，并可通过基于梯度的算法轻松求解。由于我们的方法直接依赖于对贝尔曼递归（6）的近似，因此可将其理解为一种隐式双重控制方法 Mesbah (2018)。与传统的 ST-SMPC 方法不同，后者在优化过程中固定不确定性样本，例如 Lucia 等人(2013); Schildbach 和 Borrelli(2015); Hu 等人(2022)，我们的情景树具有状态和输入依赖的不确定性实现值，因此形成了动态情景树。这样，不确定性样本便能在在线优化过程中根据预测的状态和输入调整其取值。从根本上说，正是这一特性使自车智能体能够以促进双重控制效应的方式与其他智能体交互；定理 1 对此作了详细解释。

我们将情景树中的一个节点记为 n，其时间、状态和信念状态分别记为 t_n、x_n 和 b_n。类似地，该节点的不确定性样本记为 θ_n^M、$\bar{d}_n^M$ 和 M_n。这里请回忆，$\bar{d}_n^M$是式 (10) 中定义的组合扰动，它隐式地表征了其他智能体动作的一个样本。所有节点的集合定义为 $\mathcal{N}$。我们将从父节点 $\mathfrak{p}\,(n)$ 到其子节点 n 的转移概率定义为 $\bar{P}_n:=P(\theta_n^M\mid\mathcal{I}_{\mathfrak{p}(n)};M_n)P\left(\bar{d}_n^M\mid\mathcal{I}_{\mathfrak{p}(n)};M_n\right)P\left(M_n\mid\mathcal{I}_{\mathfrak{p}(n)}\right)$. 随后，节点 n的路径转移概率，即从根节点 n_0 到节点 n 的转移概率，可递归计算为$P_n:=P_n\cdot P_{\mathfrak{p}(n)}\cdots P_{n_0}$，其中 $P_{n_0}=1$。

为了快速计算$(\theta_n^M,\bar{d}_n^M)$ 的条件概率并避免在线采样（即在优化过程中采样），我们利用它们是（条件）高斯随机变量这一事实，采用离线采样过程，方法与Arcari 等人 (2020b);Bonzanini 等人 (2020)的做法类似。我们首先从标准高斯分布中离线生成样本。然后，在在线优化过程中，使用以下解析均值和协方差表达式对这些样本进行变换：

算法 1 构建情景树

输入： 当前状态 $\hat{}x_t$ 和信念状态 b_t，时域 $N>0$，双重控制时域 $1\le N^d\le N$，模式集 $\mathcal{M}$，分支数 $K>0$

输出： 由节点集 $\mathcal{N}_t$定义的情景树

// 初始化：

设定 $x_{n_0}\leftarrow\hat{}x_t, b_{n_0}\leftarrow\hat{b}_t, t_{n_0}\leftarrow 0, \mathcal{N}_t\leftarrow\{n_0\}$

// 对偶控制步骤：

对所有 $t'=0,1,\ldots,N^d$ **执行**

  对所有 $\tilde{n}\leftarrow\mathcal{N}_t$ **执行**

    如果 $\tilde{n}.t=t'$ **则**

      分支： $\mathcal{N}_t\leftarrow\mathcal{N}_t\cup\text{BRANCH}(\tilde{n},\mathcal{M},K)$

    结束条件判断

  结束循环

结束循环

// 利用步骤：

对所有 $t'=N^d,N^d+1,\ldots,N$**执行**

  对所有 $\tilde{n}\leftarrow\mathcal{N}_t$ **执行**

    如果 $\tilde{n}.t=t'$，**则**

      正在扩展： $\mathcal{N}_t\leftarrow\mathcal{N}_t\cup$扩展$(\tilde{n})$

    结束条件判断

  结束循环

结束循环

算法 2 为对偶控制步骤生成子节点

函数 BRANCH$(n,\mathcal{M},K)$

初始化子节点集合： $\mathcal{C}\leftarrow\emptyset$

对于所有 $M\leftarrow\mathcal{M}$，**执行**

  随机采样一个集合 $\{\theta_1^{M,o},\theta_2^{M,o},\ldots,\theta_K^{M,o}\}$ 从标准高斯分布 $\mathcal{N}(0,I)$ 中

  随机采样一组 $\{\bar{d}^{M,1,o},\bar{d}^{M,2,o},\ldots\ldots,\bar{d}^{M,K,o}\}$ 从标准高斯分布 $\mathcal{N}(0,\ I)$

  对于所有 $k\leftarrow 1,\ 2,\ \ldots\ldots,\ K$**执行**

    创建一个节点 $\tilde{n}$，$\theta^{M,\tilde{n},o}=\theta^{M,k,o}$，$\bar{d}^{M,\tilde{n},o}=\bar{d}^{M,k,o}$，$M_{\tilde{n}}=M$

    $t_{\tilde{n}}=t_n+1$

    $\mathcal{C}\leftarrow\mathcal{C}\cup\{\tilde{n}\}$

  结束循环

结束循环

返回 $\mathcal{C}$

算法3 为利用步骤生成一个子节点

函数 EXTEND(n)

创建节点 $\tilde{n}$，$t_{\tilde{n}}=t_n+1$，$\theta^{M,\tilde{n},o}=0,\bar{d}^{M,\tilde{n},o}=0,M_{\tilde{n}}=M^n$

返回 $\{\tilde{n}\}$

解析均值和协方差表达式：

$$\theta_n^M=\mu^{\theta_n^M}\left(x_n,u_n^e\right)+\left(\Sigma^{\theta_n^M}\left(x_n,u_n^e\right)\right)^{1/2}\theta_n^{M,o}, \tag{14a}$$

$$\bar{d}_n^M=\left(\tilde{\Sigma}^{\bar{d}_n^M}\left(x_n,u_n^e\right)\right)^{1/2}\bar{d}_n^{M,o}. \tag{14b}$$

这揭示了我们所提出的情景树的动态特性：不确定性样本可在优化过程中通过变换（14a）和（14b）进行调整；此外，由于对 M（3b）进行贝叶斯更新，路径转移概率也取决于状态和输入。情景树构建过程总结于算法 1中。

备注 6。 为缓解与情景树相关的复杂度呈指数增长问题，仅对少量 θ^M、$_n$ ^{o}、 $\bar{d}^M$、$_n$ ^{o}和 M 进行采样。我们在先前的工作 *Hu* 等人 （*2022*）中针对静态情景树提出了一种情景剪枝机制；在此处，可将其作为启发式方法，根据最近优化得到的情景树剪除分支。对于涉及状态和输入相关不确定性实现的动态情景树，具有理论依据的剪枝方法仍是一个悬而未决的问题。

5.4 利用步骤

为缓解情景树中节点数量呈指数增长所带来的计算难题，我们可以在阶段 $N^d < N$停止扩展树，我们将该阶段称为双重控制时域。此后，剩余的 N^e : $=N - N^d$ 个阶段构成利用时域，其中每种情景均在不扩展分支的情况下延伸，且信念状态仅根据转移动力学 $g^t(\cdot)$（3d）中定义的动力学进行传播，对应于一个非双重控制随机模型预测控制问题。得益于情景树结构，利用步骤的控制输入仍保留因果反馈性质，使机器人能够谨慎行事，并对其他智能体未来的不确定性实现“被动适应”。

6 安全屏蔽感知型安全双重控制

在本节中，我们重新审视安全约束 (5e)，并为采用所提出的隐式双重随机模型预测控制策略的闭环系统提供安全性保证。以下概述本节的主要思路。首先，为确保对所有 $t \geq 0$ 均满足 $x_t \notin \mathcal{F}$，即使其他智能体采取最坏情况动作且存在外部扰动[†]，我们采用一类最小限制性监督控制方案，通常称为安全屏蔽。安全控制中安全屏蔽方法及其应用的示例包括但不限于：Bansal et al. (2017)；Fisac et al. (2018a)；Bonzanini et al. (2020)；Bastani (2021)；Wabersich and Zeilinger (2021)；Hu 等人（2022）。这种方法依赖于安全回退策略，通常将其作为“最后手段”，在安全关键事件（例如碰撞）迫在眉睫时覆盖名义策略。屏蔽策略的响应式特性使其不考虑自车的规划效率，并且往往与双重控制策略生成的探测动作相冲突，导致规划性能下降。为缓解安全覆盖与主动不确定性降低之间的冲突，我们将第5节提出的隐式双重随机模型预测控制方法与近期提出的安全屏蔽感知鲁棒规划（SHARP）框架 Hu 等人（2022）相结合。SHARP 的关键特点是利用传播后的信念状态，因果地预判未来可能发生的安全屏蔽事件，从而主动权衡名义规划性能与由其他智能体不太可能出现的行为触发的高成本紧急机动。最终得到的策略旨在降低需要采用高成本屏蔽策略的可能性，而非单纯降低碰撞概率。请注意，Hu 等人（2022）提出的原始 SHARP 框架假设其他智能体不会关注自车智能体。接下来，我们将放宽这一过于保守的假设，并明确考虑其他智能体的响应。

6.1 面向安全交互规划的安全屏蔽

本文专注于为交互规划提供鲁棒（最坏情况）安全性保证。因此，我们作出运行设计域（ODD）Lee 等人 (2020) 假设：外部扰动 d_t 有界。

假设 1。 外部扰动 d_t 逐元素有界，即 $d_t \in \mathcal{D} \subseteq \mathbb{R}^{n_x}$ 且 $\bar{d}^i_{\min} \leq d^i_t \leq \bar{d}^i_{\max}$ 对所有 $i = 1, \ldots, n_x$均成立。

一种屏蔽机制定义为一个元组(Ω π^s)，其中集合Ω $\subseteq \mathbb{R}^{n_x}$ 是一个安全集，并且是鲁棒控制不变的Blanchini (1999)，且满足$\Omega \cap \mathcal{F} = \emptyset$，而 π^s : $\mathbb{R}^{n_x} \to \mathcal{U}^e$ 是一个安全控制策略，即使面对其他智能体和外部扰动的最坏情况动作，也能使联合系统（1）的状态保持在Ω 内。

定义2。 鲁棒控制不变集。 给定联合动力学(1)，其中其他智能体的动作 $u^o_t \in \mathcal{U}^o$ 有界，且输入 $d_t \in \mathcal{D}$不确定，若存在控制策略$\Omega \subseteq \mathbb{R}^{n_x}$，使得控制策略$\pi^s$: $\mathbb{R}^{n_x} \to \mathcal{U}^e$ 能够使 x_t 始终不会离开Ω，则集合

$$x_0 \in \Omega \\ \Rightarrow x_t \in \Omega,\ \forall t > 0,\ \forall u^o_t \in \mathcal{U}^o,\ \forall d_t \in \mathcal{D},\ u^e_t = \pi^s(x_t).$$

令 x^+ 表示在给定1的情况下，根据（ x， u^e， $\tilde{u}^o$， $\tilde{}$和 d，通过式（1）演化得到的下一状态。随后，我们定义屏蔽集：

$$\mathcal{S}^e = \{(x, u^e) \in \Omega \times \mathcal{U}^e \mid \exists \tilde{u}^o \in \mathcal{U}^o,\ \exists \tilde{d} \in \mathcal{U}^o : \\ x^+ \left(x, u^e, \tilde{u}^o, \tilde{d}\right) \notin \Omega\},$$

其中包含所有可能导致下一状态脱离安全集的状态-动作对。基于屏蔽机制(Ω π^s)，我们定义一种限制最少的监督安全过滤器Hsu 等人 (2023)，将其定义为切换策略：

$$u^e_t = \pi^f(x_t; \tilde{u}^e_t) := \begin{cases} \tilde{u}^e_t, & \text{if } (x_t, \tilde{u}^e_t) \notin \mathcal{S}^e \\ \pi^s(x_t), & \text{if } (x_t, \tilde{u}^e_t) \in \mathcal{S}^e. \end{cases} \tag{15}$$

安全过滤器（15）允许自车智能体应用任意名义控制器 π_t : $\mathbb{R}^{n_x} \to \mathcal{U}^e$ 只要 $(x_t,\ \pi_t\ (x_t)\ )$ 不在屏蔽集 $\mathcal{S}^e$中；否则，它会用屏蔽策略 $\pi_t\ (x_t)$ 替代 $\pi^s\ (x_t)$ 。由此可得以下结果。

命题 1。安全过滤器（命题 1，Hu 等人（2022））。若集合Ω在屏蔽策略下是鲁棒控制不变的$\pi^s\ (\cdot)$，则它在安全过滤策略下也是鲁棒控制不变的 $\pi^f(\cdot;\ \pi_t\ (\cdot)\ )$对于任意名义策略 $\pi_t\ (\cdot)$ 。

注 7。我们可以用（5e）替换为机会约束，以处理具有无界支撑集的不确定输入，即$P\left[x \notin \mathcal{F}\right] \geq 1 - \delta$，其中 $\delta \in (0, 1]$ 是容差水平。随后，可通过概率屏蔽方法（例如*Bastani* 等人（*2021*年）获得概率安全保证。

6.2 交互感知 *SHARP*

在 SHARP 框架中，安全过滤器（15）接收沿情景树各分支预测的未来状态和其他智能体的动作

[†]这也可以解释为，在使用（2）预测另一智能体的动作和未来状态演变时，所产生的最大模型失配。

，以因果地预见未来可能发生的安全屏蔽事件，从而使最终策略能够感知（通常代价高昂的）屏蔽机动。然而，（15）涉及基于集合包含关系的条件约束，因而难以在随机模型预测控制问题中进行优化。为此，我们对 Hu 等人 (2022) 提出的近似局部安全滤波方案进行修改和改进，将（15）重新表述为一个凸约束，同时额外考虑其他智能体的响应和外部扰动。

我们从给定的标称状态处将联合系统 (1) 线性化，该状态 $\bar{x}_{\tilde{n}}$ 与节点 $\tilde{n}$ 相关：

$$\delta x_{\tilde{n}}^{+} = A_{\tilde{n}}\delta x_{\tilde{n}} + B_{\tilde{n}}^{e}u_{\tilde{n}}^{e} + B_{\tilde{n}}^{o}u_{\tilde{n}}^{o} + d_{\tilde{n}}, \tag{16}$$

其中 $A_{\tilde{n}} := \nabla_x f(x)|_{\bar{x}_{\tilde{n}}}$，$B_{\tilde{n}}^{e} := B^{e}(\bar{x}_{\tilde{n}})$，$B_{\tilde{n}}^{o} := B^{o}(\bar{x}_{\tilde{n}})$，且 $\delta x_{\tilde{n}} = x_{\tilde{n}} - \bar{x}_{\tilde{n}\circ}$ 然后，我们将安全集 Ω 在 $x_{\tilde{n}}$处局部近似为一个半空间：

$$\Omega_{\tilde{n}} = \{\delta x \mid H_{\tilde{n}}{}^{\top}\delta x \geq 0\}, \tag{17}$$

其中 $H_{\tilde{n}} := \bar{x}_{\tilde{n}}^{+} - \bar{x}_{\tilde{n}}$ 近似表示 Ω 在 $x_{\tilde{n}}$处切空间的法向量，如图4所示。现在，给定不确定线性系统 (16) 和半空间安全集 $\Omega_{\tilde{n}}$，我们将 Agrawal 和 Sreenath(2017年) 的结果加以扩展，构造一个仿射鲁棒控制障碍函数（RCBF）约束，以近似表示 (15)。该 RCBF 通过施加一个用于认证近似安全集$\Omega_{\tilde{n}}$受控不变性的约束，在局部捕捉安全屏蔽操作；这一点在引理 2中得到了形式化表述。

定义 3。 离散时间指数型鲁棒控制障碍函数。 映射$h(x)$是系统 $x_{t+1} = f(x_t, u_t, d_t)$ 的离散时间指数型鲁棒控制障碍函数，当且仅当：

(a) $h(x_0) \geq 0$ *and,*

(b) $\exists u_t \in \mathcal{U}$ *such that* $h(x_{t+1}) - h(x_t) + \gamma h(x_t) \geq 0$ *for all* $t \geq 0$, $d_t \in \mathcal{D}$, *and some* $\gamma \in (0, 1]$.

引理 2。 设假设 *1*成立。对于不确定线性系统 (16)，给定其他智能体的控制 $u^{o} \in \mathcal{U}^{o}$，令$\Omega_{\tilde{n}}$ 由 (17) 定义的集合在自车控制 $u_{\tilde{n}}^{e} \in \mathcal{U}^{e}$ 下是鲁棒控制不变的，当且仅当其满足：

$$\begin{aligned} &H_{\tilde{n}}^{\top}\left[\left(A_{\tilde{n}} + (\gamma - 1)I\right)\delta x_{\tilde{n}} + B_{\tilde{n}}^{e}u_{\tilde{n}}^{e} + B_{\tilde{n}}^{o}u^{o} + d_{\tilde{n}}^{*}\right] \\ &=: K_{\tilde{n}}(\delta x_{\tilde{n}}, u_{\tilde{n}}^{e}, u^{o}) \geq 0, \end{aligned} \tag{18}$$

对于所有 $\delta_x \in \Omega_{\tilde{n}}$ 以及某个 $\gamma \in (0, 1]$。这里，最优（最坏情况）扰动 $d_{\tilde{n}}^{*}$ 按元素定义如下：

$$d_{\tilde{n}}^{*,i} := \begin{cases} \bar{d}_{\min}^{i}, & \text{if } H_{\tilde{n}}^{i} \geq 0 \\ \bar{d}_{\max}^{i}, & \text{otherwise.} \end{cases} \tag{19}$$

证明见附录 B.2。不等式（18）关于 $\delta x_{\tilde{n}}$ 和 $u_{\tilde{n}}^{e}$以及 u^{o} 是仿射的（因此也是凸的），优化效率高，并且可作为不等式约束纳入 ST-SMPC 问题，以因果地预测未来的安全屏蔽事件。我们不采用式（18）中的最坏情况动作，而是允许该动作取决于其他智能体的预测动作$u_{\tilde{n}}^{o}$，它是受自车动作和系统状态影响的决策变量。这样，SHARP 在预测安全屏蔽事件时便能考虑交互。在时刻 t，基于上一次优化得到的情景树 $\mathcal{N}_{t-1}^{*}$，我们可按照式（15）执行一次单步仿真展开，以确定需要使用屏蔽策略的节点。我们将时刻 t 的所有屏蔽节点集合记为 $\mathcal{N}_t^{s}$。对所有 $\tilde{n}^{s} \in \mathcal{N}^{s}$均施加凸屏蔽感知约束（18）。寻找 $\mathcal{N}_t^{s}$的过程总结于算法 4。

算法4 识别安全屏蔽节点

输入： 由节点集定义的上一个优化场景树 $\mathcal{N}_{t-1}^{*}$，由节点集 $\mathcal{N}_t$定义的情景树

输出： 安全屏蔽节点集 $\mathcal{N}_t^{s}$

初始化安全屏蔽节点集：$\mathcal{N}_t^{s} \leftarrow \emptyset$

对于所有 $(\tilde{n}^{*}, \tilde{n}) \leftarrow \mathtt{zip}(\mathcal{N}_{t-1}^{*}, \mathcal{N}_t)$ **执行**

  $\tilde{x}_{\tilde{n}^{*}}^{+} \leftarrow$ 使用 $x_{\tilde{n}^{*}}$, $u_{\tilde{n}^{*}}^{e}$, 进行单步前向仿真 $u_{\tilde{n}^{*}}^{o}$, $d_{\tilde{n}^{*}}^{M}$ 以及动力学 (20c)

  若 $\tilde{x}_{\tilde{n}^{*}}^{+} \notin \Omega$，**则**

    根据以下内容计算 $H_{\tilde{n}}$、$A_{\tilde{n}}$、$B_{\tilde{n}}^{e}$、$B_{\tilde{n}}^{o}$ 和 $d_{\tilde{n}}^{*}$ 第 6.2

    $\mathcal{N}_t^{s} \leftarrow \mathcal{N}_t^{s} \cup \{\tilde{n}\}$

  结束 if

结束 for

6.3 总体算法方法

给定当前状态测量 $\hat{}x_t$、最近更新的信念状态$\hat{}$ b_t以及由节点集定义的情景树 $\mathcal{N}_t$，我们可以将（5）重写为与Bernardini 和 Bemporad（2011）类似形式的 ST-SMPC 问题：

$$\min_{\mathbf{U}_t^{e}} \sum_{\tilde{n}\in\mathcal{N}_t\setminus\mathcal{L}_t} P_{\tilde{n}}\ell(x_{\tilde{n}}, u_{\tilde{n}}^{e}) + \sum_{\tilde{n}\in\mathcal{L}_t} P_{\tilde{n}}\ell_F(x_{\tilde{n}}) \tag{20a}$$

$$\text{s.t. } x_{n_0} = \hat{x}_t,\ b_{n_0} = \hat{b}_t, \tag{20b}$$

$$x_{\tilde{n}} = f(x_{\mathfrak{p}(\tilde{n})}) + B^{o}(x_{\mathfrak{p}(\tilde{n})})u_{\tilde{n}}^{o}(x_{\mathfrak{p}(\tilde{n})},\ u_{\mathfrak{p}(\tilde{n})}^{e}) + B^{e}(x_{\mathfrak{p}(\tilde{n})})u_{\mathfrak{p}(\tilde{n})}^{e} + \bar{d}_{\tilde{n}}^{M}, \quad \forall\tilde{n}\in\mathcal{N}_t\setminus\{n_0\} \tag{20c}$$

$$b_{\tilde{n}} = \tilde{g}\left(b_{\mathfrak{p}(\tilde{n})}, x_{\tilde{n}}, u_{\mathfrak{p}(\tilde{n})}^{e}\right), \quad \forall\tilde{n}\in\mathcal{N}_t^{d} \tag{20d}$$

$$b_{\tilde{n}} = g^{t}\left(b_{\mathfrak{p}(\tilde{n})}\right), \quad \forall\tilde{n}\in\mathcal{N}_t^{e} \tag{20e}$$

$$u_{\tilde{n}}^{e}\in\mathcal{U}^{e},\ u_{\tilde{n}}^{o}\in\mathcal{U}^{o}, \quad \forall\tilde{n}\in\mathcal{N}_t\setminus\mathcal{L}_t \tag{20f}$$

$$K_{\tilde{n}}(\delta x_{\tilde{n}}, u_{\tilde{n}}^{e}, u_{\tilde{n}}^{o}) \geq 0, \quad \forall\tilde{n}\in\mathcal{N}_t^{s} \tag{20g}$$

其中，$\mathcal{L}_t$ 是所有叶节点的集合（即没有后代的节点），$\mathcal{N}_t^{d}$和 $\mathcal{N}_t^{e}$分别是双重控制节点和利用节点的集合，$\mathbf{U}_t^{e}$:$= \{u_{\tilde{n}}^{e} \in \mathbb{R}^{m_e} : \tilde{n} \in \mathcal{N}_t \setminus \mathcal{L}_t\}$ 是与所有非叶节点相关联的自车控制输入集合，$\mathcal{N}_t^{s}$是算法4返回的屏蔽节点集，其中包含施加凸屏蔽感知约束（18）的节点。目标函数（20a）基于情景树的不确定性样本对（5a）进行近似。约束（20b）-（20g）分别表示初始（信念）状态、物理状态动力学、探索和利用步骤的信念动态、控制限制以及（18）中的凸屏蔽感知约束。问题（20）是一个非凸轨迹优化问题，可使用通用非凸求解器求解，例如 SNOPT Gill 等人（2005）和 IPOPT W¨achter 和 Biegler（2006年）。问题（20）的最优解$\mathbf{U}_t^{e}{}^{*}$，以滚动时域方式实施，即 $\pi_{\text{IDSMPC-SHARP}}^{e}(\hat{}x_t,\ b_t) := u^{e,}{}_{n_0}^{*}$，我们将其称为屏蔽感知隐式双重随机模型预测控制（IDSMPC-SHARP）策略。我们的总体算法方法围绕

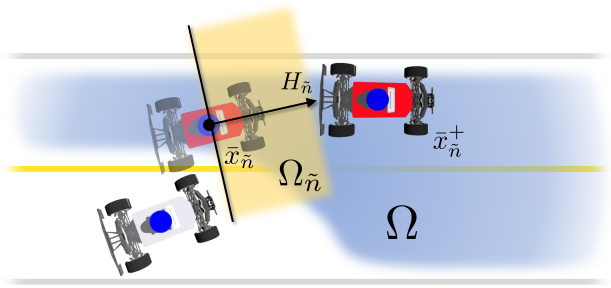


图 4. 基于RCBF的半空间安全集$\Omega_{\tilde{n}}$（黄色）的示意图，该安全集在$x_{\tilde{n}}$处局部近似安全集 Ω（蓝色）。

IDSMPC-SHARP策略展开，详见算法5。

注释8. 将*RCBF*约束 (18)（其近似安全过滤策略 (15)）纳入 *ST-SMPC* (20g) 的唯一目的，是 预测 未来的安全屏蔽覆盖，从而主动提升长期规划性能。这种近似不会影响系统的递归安全性；递归安全性由屏蔽步骤（算法 *5*第*6*行，已在定理*2*中证明）保证。

注释*9.* 当 *(20a)* 中的代价函数为二次型，即 $\ell(x,u) = \|x\|_Q^2 + \|u\|_R^2$ 且 $\ell_F^e(x) = \|x\|_{Q_F}^2$ 时，可以采用一种近似技术，解析地计算关于状态的期望代价。令 $A_{\mathfrak{p}(\tilde{n})}$ 表示在状态 $\mu^x_{\mathfrak{p}(\tilde{n})}$ 的均值和自车控制 $u^e_{\mathfrak{p}(\tilde{n})}$ 处评估的动力学 *(20c)* 的雅可比矩阵。随后，沿情景树的每个分支，状态都将保持高斯分布，其协方差由递归公式给出：$\Sigma^x_{\tilde{n}} = \Sigma^{\bar{d}}_{\tilde{n}}(\mu^x_{\mathfrak{p}(\tilde{n})}, u^e_{\mathfrak{p}(\tilde{n})}) + A_{\mathfrak{p}(\tilde{n})}\Sigma^x_{\mathfrak{p}(\tilde{n})}A^\top_{\mathfrak{p}(\tilde{n})}$。我们回顾一下，$\mathbb{E}_{\xi\sim\mathcal{N}(0,I)}\left[(W\xi)^\top Q(W\xi)\right] = tr(W^\top QW)$。因此，目标函数 *(20a)* 可近似为

$$\sum_{\tilde{n}\in\mathcal{N}_t\setminus\mathcal{L}_t} P_{\tilde{n}}\ell(x_{\tilde{n}}, u^e_{\tilde{n}}) + \sum_{\tilde{n}\in\mathcal{L}_t} P_{\tilde{n}}\ell_F(x_{\tilde{n}}) + \sum_{\tilde{n}\in\mathcal{N}_t} \mathrm{tr}(Q\Sigma^x_{\tilde{n}})$$

其中，节点 $\tilde{n}$ 的路径转移概率计算结果为 $P_{\tilde{n}} = P(\theta^M_n \mid \mathcal{I}_{\mathfrak{p}(n)};\ M_n)\ P\left(M_n \mid \mathcal{I}_{\mathfrak{p}(n)}\right) P_{\mathfrak{p}(\tilde{n})}$。这样，只需对隐状态进行采样以计算期望成本，无需进行 $\bar{d}^M$ 采样。

6.4 规划框架的性质

本节将考察所提出的 IDSMPC-SHARP 控制策略的性质，以及联合系统（1）在算法5作用下的闭环行为。

定理1。双重控制效应。通过求解 $\pi^e_{IDSMPC\text{-}SHARP}(\cdot,\ \cdot)$（20）得到的反馈控制策略 $\pi^e_{IDSMPC\text{-}SHARP}(\cdot,\ \cdot)$ 会产生双重控制效应。

证明。根据引理 1，对于任意给定的时刻 $t \geq 0$，机器人的控制 u^e_t 可以影响 $\Sigma^{\theta^M}_{t+1}$，即关于 θ^M的信念的协方差（二阶矩）。因此，策略$\pi^e_{\text{IDSMPC-SHARP}}$ 对隐状态θ^M产生双重控制效应，符合定义 1(a)。回顾式（3b）中关于 M 的分类分布信念测量更新方程：

$$p(M_- \mid \mathcal{I}_{t+1}) = \frac{p(x_{t+1} \mid u^e_t, \mathcal{I}_t; M)p(M \mid \mathcal{I}_t)}{p(x_{t+1} \mid u^e_t, \mathcal{I}_t)},$$

算法5 屏蔽感知安全双重控制

输入： 初始状态 x_0，初始信念状态 b_0，Q 值函数 $Q^M_i(\cdot)$ 位于 (2) 中，安全屏蔽机制 (Ω π^s)

初始化：$t \leftarrow 0, \mathcal{N}^*_0 \leftarrow \emptyset$

当 尚未达到规划目标**时**

// 规划：

$\mathcal{N}_t \leftarrow$使用算法构建情景树1

$\mathcal{N}^s_t \leftarrow$ 识别屏蔽节点并构建使用算法构建屏蔽感知约束4

$\pi^e_{\text{IDSMPC-SHARP}}(x_t, b_t)$ $\mathcal{N}^*_t \leftarrow$ 求解 ST-SMPC (20)

// 安全屏蔽：

将安全过滤策略 (15) 应用于自车智能体：$u^e_t = \pi^f(x_t;\ \pi^e_{\text{IDSMPC-SHARP}}(x_t,\ b_t))$

// 信念更新：

测量新状态 x_{t+1}

$b_{t+1} \leftarrow$使用（3a）-（3d）更新信念状态

更新时间：$t \leftarrow t+1$

结束循环

这表明自车的控制 u^e_t 可以影响关于 M的分类分布的所有分量，进而影响其熵：

$$H(M_- \mid \mathcal{I}_{t+1}) \propto \sum_{\tilde{M}\in\mathcal{M}} P(\tilde{M} \mid \mathcal{I}_{t+1}) \log P(\tilde{M} \mid \mathcal{I}_{t+1}),$$

意味着对 M 根据定义1(b)的双重控制效应。 □

注释 10。最优性。*ST-SMPC* (20) 的解通常不是贝尔曼递归 (6)的最优解，这是由于近似信念状态动力学 $\tilde{g}$、基于不确定性样本近似的期望代价、截断的探索步数，以及非凸程序 (20) 的解往往只是局部最优。因此，通常无法实现最优的探索-利用权衡。尽管如此，由于自车的控制根据定理*1*对隐状态产生双重控制效应，因此，在达到 (20) 的局部最优解这一程度上，它会自动平衡名义规划性能与不确定性降低。

注释 *11*。可行性保证。只要满足凸的安全屏蔽约束 *(20g)*，*ST-SMPC (20)* 就是可行的。正如 *Agrawal* 和 *Sreenath*（*2017*）所指出的，当自车的控制输入有界时，并不能保证约束 *(20g)* 存在可行解。为确保 *(20)* 递归可行，我们可以将 *(20g)* 作为带松弛变量的软约束纳入，即 $K_{\tilde{n}}(\delta x_{\tilde{n}}, u^e_{\tilde{n}}, u^o_{\tilde{n}}) \geq s_{\tilde{n}},\ s_{\tilde{n}} \leq 0$。请注意，放宽约束 *(20g)* 不会影响递归安全性；递归安全性由屏蔽步骤（算法*5*第*6*行）强制保证，并已在定理*2*中证明。

定理2。递归安全性。假设初始状态位于安全集内，即 $x_0 \in \Omega$，则联合系统 (1) 与算法 *5* 构成闭环后仍保持安全，即 $x_t \in \Omega\ \ \forall t > 0$.

证明。该结果可直接由命题 1和定义 2得出。 □

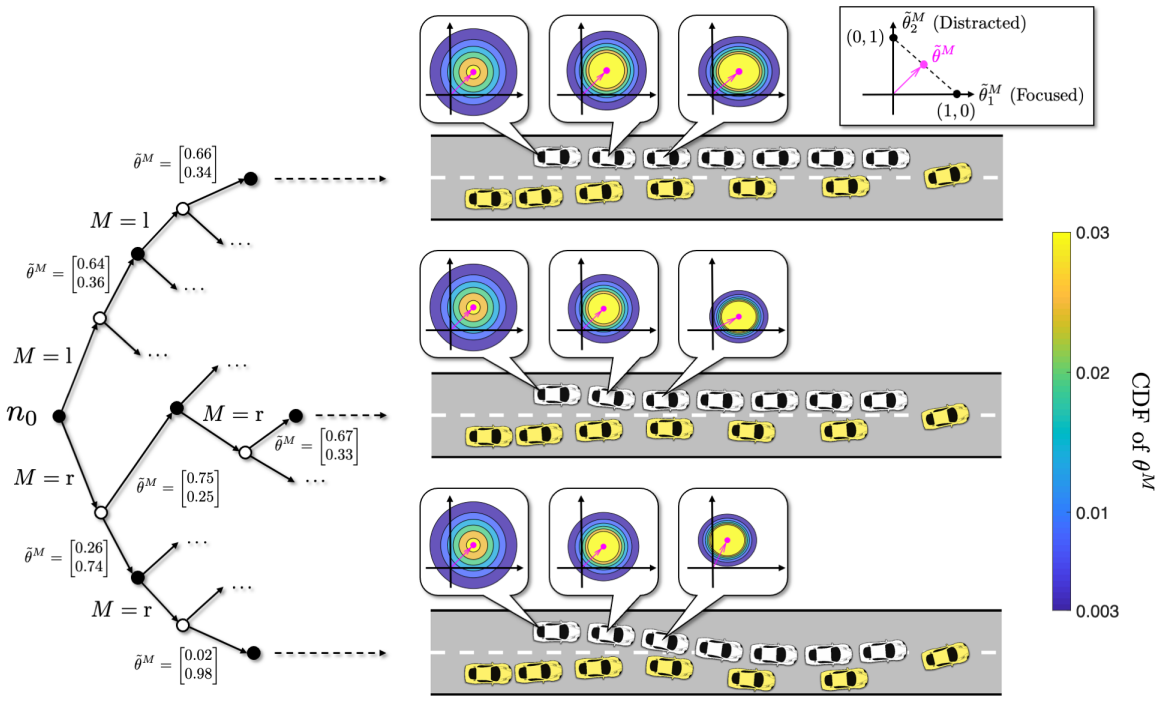


图 5. 高速公路超车示例（见第 7.2.2 节）中的情景树，以及由此得到的经过优化的情景轨迹，其中包含 $N^d = 2$ 个双重控制时间步和 $N^e = 4$ 个利用步骤。人类驾驶车辆和自动驾驶汽车分别以白色和黄色绘制。隐状态 θ^M 建模为二维高斯随机变量，并归一化到 1-单纯形以便可视化。白色圆圈 ∘ 表示一个由一个 M 样本部分确定的中间节点，黑色圆圈 ● 则表示一个由 $M_{\bar{n}}$ 和 $\tilde{\theta}^M$ 样本共同完全确定的节点。洋红色箭头表示归一化隐状态 $\tilde{\theta}^M$的最大后验均值。等高线图显示了 θ^M的累积分布函数（CDF）等值集。与人类偏好左车道（上方分支）的情况相比，不确定性在这种情况下降低得明显较少，因为人类的行为受机器人（探测）动作的影响小于其偏好右车道时的情况。

7 仿真研究

7.1 仿真设置

我们在仿真的驾驶场景中评估所提出的隐式对偶情景树式随机模型预测控制（IDSMPC，即不包含屏蔽感知约束（20g）的情况下求解（20g）得到的规划器）和 IDSMPC-SHARP（通过求解（20）得到的规划器）。在规划和仿真中，车辆和行人动力学分别由 Zhang 等人 (2020) 提出的四维运动学自行车模型和 Fridovich-Keil 等人 (2020) 提出的四维独轮车模型描述，两者均以 $\Delta t = 0.2$s 的时间步长进行离散化。所有仿真均使用 MATLAB 和 YALMIP L¨ofberg (2004) 在配备英特尔酷睿 i7-10700K 处理器的台式机上完成。所有非线性模型预测控制问题均使用 SNOPT Gill 等人 (2005) 求解。仿真所用的参数值见表 2。开源代码可在线获取。[‡]

7.1.1 基线方法。 我们将所提出的 IDSMPC 和 IDSMPC-SHARP 规划器与四种基线方法进行比较：

- 显式对偶随机模型预测控制（EDSMPC）在阶段代价 ℓ 中加入了信息增益项（5a） $\lambda(H(b_k) - H(b_{k+1}))$，由 Sadigh 等人(2018年)提出，从而形成一个显式对偶随机最优控制问题。这里， $\lambda > 0$是经过微调的权重因子。与 IDSMPC 类似，期望代价也通过不确定性样本进行近似计算。

- 非对偶情景随机模型预测控制（NDSMPC）基于求解（5），并采用情景树不通过测量更新（20d）传播信念状态（因此，得到的策略不具有双重控制效应）。类似的情景规划也见于 Bernardini 和 Bemporad (2011年)；Schildbach 和 Borrelli (2015年)；Hu 等人(2022)。

- 确定性等价模型预测控制（CEMPC）基于确定性等价原理求解（5）Mesbah (2018年)；Arcari 等人 (2020a)。最大后验（MAP）策略对齐方法由 Peters 等人 (2020)用于自车智能体修改规划问题。例如，在示例1中，如果自车和其他车辆都在内侧车道行驶，且 M 的最大后验估计为 NY（不让行），则会将自车的参考车道设为外侧车道（通过修改代价函数 $\ell(\cdot)$ 和 $\ell_F(\cdot)$），以便安全地超越其他智能体。

- 采用基于推断的策略对齐的迭代线性二次博弈（ISA-iLQ），由Peters 等人（2020）提出，最初由 Fridovich- Keil 等人（2020）开发。输入投影器用于强制执行$u^o \in \mathcal{U}^o$ 和 $u^e \in \mathcal{U}^e$。

EDSMPC 是一种双重控制规划器，而其他三种规划器不会产生双重控制效应。所有规划器都使用相同的二次代价函数 ℓ 和 ℓ_F 来惩罚参考跟踪误差和控制量幅值，并且都配备了

[‡]https://github.com/SafeRoboticsLab/Dual_控制_HRI

相同的基于贝叶斯推断的方法来推断其他智能体的意图。为逐步解读结果，IDSMPC 和四种基线方法均不采用安全屏蔽；为考虑安全性，我们使用 Zeilinger 等人提出的软约束模型预测控制方法（2014年），该方法通过为每个节点设置松弛变量 $x_{\tilde{n}} \notin \mathcal{F}$ 来放宽原有的硬约束$\tilde{n} \in \mathcal{N}_t$，使这些规划器具备安全感知能力，但不依赖安全屏蔽。IDSMPC-SHARP 使用由应急规划器设计提供的屏蔽机制，参见Hardy 和 Campbell（2013）；Bajcsy 等人（2021年）提出的方法可防范其他智能体的行为不确定性和外部扰动（假设1）。应急规划问题通过迭代线性二次调节器（iLQR）方法在线求解，参见Todorov 和 Li（2005）。我们采用 Bastani提出的模型预测屏蔽（MPS）算法（2021），以确保屏蔽策略的递归可行性。

7.1.2 指标。为衡量规划性能，我们考虑以下两个指标：

- 闭环代价，定义为 $J^{e}_{\text{cl}} := \sum_{t=0}^{T_{\text{sim}}} \ell(x_t, u^{e}_t)$，其中 T_{sim} 是仿真时域，$x_{[0:T_{\text{sim}}]}, u_{[0:T_{\text{sim}}]}$ 是执行的轨迹（通过重新规划）。
- 碰撞率，定义为 $N_{\text{coll}}/N_{\text{trial}} \times 100\%$，其中$N_{\text{coll}}$ 是发生碰撞的试验次数，即 $x_t \in \mathcal{F}$，而 N_{trial} 是试验总次数。

7.1.3 假设。我们提出了三个假设，并通过仿真结果加以验证。

•H1（性能与安全权衡）。双重控制规划器比非双重控制基线方法实现更好的性能-安全权衡。

• H2（隐式双重控制与显式双重控制）。即使经过微调，显式双重控制的效率仍低于其隐式对应方法。

• H3（SHARP 提升效率）。与 *IDSMPC* 相比，*IDSMPC-SHARP* 提升了规划效率，即使经过针对安全性的微调也是如此。

7.2 仿真智能体

7.2.1 智能体策略。为了展示我们的方法在一般交互规划场景中的有效性，我们使用一个类似于Baseggio 等人（2011）Guo 等人（2013）提出的基于优化的仿真器来生成其他智能体的运动。自车智能体无法获取仿真器的设计参数。尽管我们已尽最大努力生成合理的仿真人类智能体行为，使其不落入 ST-SMPC 所用运动预测模型（20）所涵盖的假设范围，但我们承认，仿真人类行为仍可能与真实世界的人类数据有所不同，而这类数据通常获取成本高且难以获得，尤其是在交互场景中。尽管如此，我们在第7.3节展示了一个示例，其中人类的轨迹来自 Waymo 开放运动数据集Sun 等人（2020）。

7.2.2 目标与意识不确定性。考虑图中所示的高速公路驾驶场景 7，其中一辆自动驾驶车辆（自车，以黄色标示）需要超越一辆人类驾驶车辆（其他智能体，以白色标示）。连续隐状态定义为 $\theta^M := (\theta^M_D, \theta^M_F)$ 其中 θ^M_D和 θ^M_F分别表示人类的分心水平和专注程度。专注的人类会考虑联合系统的安全（例如，在机器人尝试并入其前方时避让并为其让出空间），而分心的人类则不会。离散隐状态 M 用于描述人类是否偏好在左车道或右车道行驶，即 $M \in \{$左，右$\}$。通过求解 IDSMPC 得到的该示例的优化情景树如图5所示。

图中绘制的性能-安全权衡曲线 6验证了 H1。在这里，我们为每个 未经过安全屏蔽的 规划器设计了一组经过精细调优的安全关键参数，例如软约束避碰代价的权重、失效集的鲁棒裕度以及加速度限制。对于给定的规划器设计，我们使用不同的随机种子对该场景进行 50 次仿真；随机种子会影响包括初始条件、加性扰动、人类的车道偏好和安全意识在内的不确定性来源。这些随机变量与人类和机器人之间的（闭环）交互相互独立。请注意，即使是保守性最低的 IDSMPC 策略，其碰撞率仍低于基线方法，且闭环代价与非对偶策略相近。尽管 EDSMPC 策略也凭借主动降低人类不确定性的能力实现了较低的碰撞率，但其整体闭环性能始终不如 IDSMPC，这验证了 H2。最后，我们在由相同随机种子控制的 50 个场景下测试 IDSMPC-SHARP。得益于安全屏蔽，所有 50 次试验均能保证安全；同时，由于具备屏蔽感知，IDSMPC-SHARP 达到的（归一化）平均闭环成本（1.30）降低了约19%（$\pm 5.6\%$ 标准差），相比之下，IDSMPC 的平均闭环成本为（1.61）；IDSMPC 经过精细调优，实现了零碰撞率。这验证了 H3。

图 7 展示了一次仿真试验中 $P(M = \text{l})$ 的轨迹快照及演变过程。如图 7 底部所示，真实值中人类的车道偏好在前 3 s[§] 为右车道，随后在仿真剩余时间内变为左车道。先验设为 $P(M = \text{l}) = 1$ 和 $\theta^{\text{l}}, \theta^{\text{r}} \sim \mathcal{N}((0.5, 0.5), 5I)$。与非对偶控制规划器不同，IDSMPC 控制机器人沿道路中央接近人类驾驶车辆，从而使机器人能够对人类进行信息性探测——这使 M （底部）的预测更加准确——并引导机器人进入一个更容易避免碰撞的区域。事实上，当人类的隐状态 M 在 $t = 6$ s 时从 r 切换为 l，使用 IDSMPC 的机器人执行了急剧右转，并成功避免与人类发生碰撞。EDSMPC规划器虽然在初期有效降低了不确定性，却未能意识到从右侧超越人类会产生更高效的轨迹。所有非对偶控制规划器即使进行重新规划，也因对 M 的了解不足而导致与人类发生碰撞。还值得注意的是，即使 IDSMPC 使用 NDSMPC解进行

[§]第7节中描述智能体行为的所有时间数据均为模拟数据。

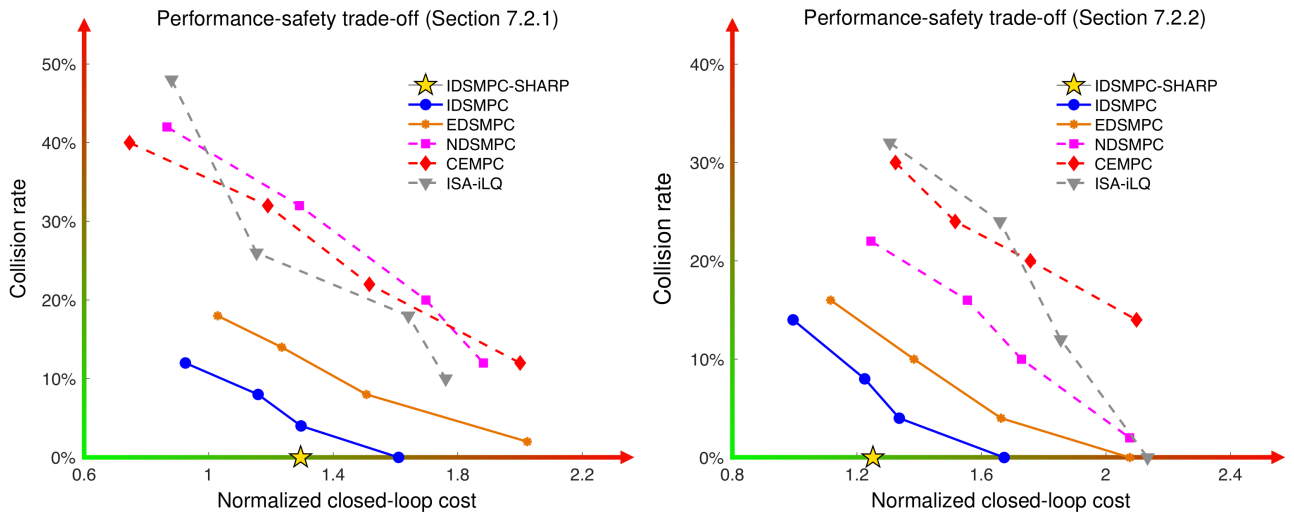


图 6。 第 7.2.2 节和7.2.3节仿真实例的性能-安全权衡。闭环代价 J_{cl}^{e} 通过 1×10^{4}进行归一化。对于未加屏蔽的规划器（IDSMPC、EDSMPC、NDSMPC、CEMPC 和 ISA-iLQ），每个数据点均基于一组不同的安全关键调参参数获得，每组参数进行 50 次具有不同随机种子的试验。使用 IDSMPC-SHARP 进行 50 次仿真试验所对应的数据点以黄色星形标出。对偶控制策略的曲线以实线表示，而非对偶控制策略的曲线以虚线表示。在整体性能-安全权衡方面，IDSMPC 优于所有基线方法。得益于安全屏蔽和 SHARP 框架，IDSMPC-SHARP 的碰撞率为零，且相比最安全的 IDSMPC 设计，代价降低约20% 。

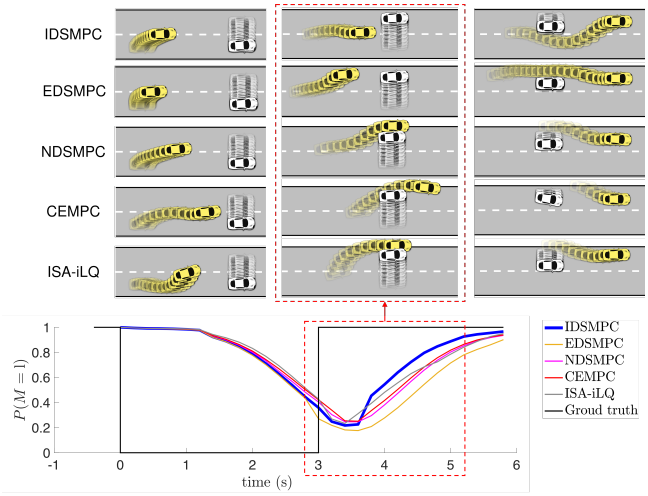


图 7。 第7.2.2节所述高速公路驾驶示例的仿真快照，其中黄色的自车试图超越白色的人类驾驶车辆。纵向位置以相对坐标表示，$p_x^H = 0$。左、中、右三列分别显示 $t = [0, 3]$ s、 $t = [3, 5]$ s 及剩余时段的轨迹。底部的图显示了四种规划器随时间变化的 $P(M = 1)$。我们提出的 IDSMPC 规划器使机器人顺利、安全地完成了超车操作，而非双重控制规划器则导致了不安全轨迹。机器人使用 EDSMPC 时，被困在紧邻人类驾驶车辆的一处狭窄空间中，导致轨迹效率较低。

初始化，其闭环行为也大不相同，这本质上源于双重控制效应。

在图 8中，我们考察了另一次仿真试验。使用 IDSMPC，机器人在 6 秒内安全地超越了人类驾驶的汽车。然而，使用 ISA-iLQ 规划器时，机器人未能在 10 秒内超越人类。由于缺乏对偶控制作用，机器人被困在人类后方，且没有意识到人类愿意为机器人让出空间。在图 8中，我们还展示了使用 CEMPC 规划器生成的一条不安全轨迹。这些结果表明，在与其他智能体交互时，凭借双重控制作用，机器人能够获得更好的安全性和活跃性。

7.2.3 行为与合作不确定性。接下来，我们考虑示例2中引入人类不确定性的无信号灯交通路口场景。图 9展示了使用 IDSMPC、具有不同隐状态的四次仿真试验的轨迹快照。我们为所有 $M \in \mathcal{M}$ 选择了无信息先验分布 $p(M \mid \mathcal{I}_0) = [0.25 \quad 0.25 \quad 0.25 \quad 0.25]$ 和 $\theta^M \sim \mathcal{N}((0.5, 0.5)\ 5I)$。可以看到，四次试验均安全完成，且自动驾驶车辆和行人都到达了目标位置。根据 50 次仿真试验绘制的性能-安全权衡曲线如图 6 所示，与第 7.2.2 节类似。再次可以看到，在所有未加屏蔽的规划器中，IDSMPC策略实现了最佳的性能-安全权衡，且碰撞率始终低于 15%。采用经安全屏蔽的 IDSMPC-SHARP 策略后，所有试验均实现了安全，其（归一化）平均闭环代价为 1.25，比最安全的 IDSMPC 设计（1.67）低 25%（±3% 标准差）。

7.2.4 四智能体交互示例。最后，我们将 IDSMPC 应用于示例2中相同的交通路口场景，其中涉及三个人类智能体：两辆人类驾驶车辆和一名行人。人类驾驶车辆采用目标与意识不确定性模型（第7.2.2节），行人则采用行为与合作不确定性模型（示例2和第7.2.3节）。此外，我们设置仿真，使行人在前 2 秒内忽略其他智能体（$M = \text{o}$），之后在剩余仿真时间内变为安全感知（$M = \text{p}$）。图10展示了一次具有代表性的试验中的轨迹快照。自动驾驶车辆能够

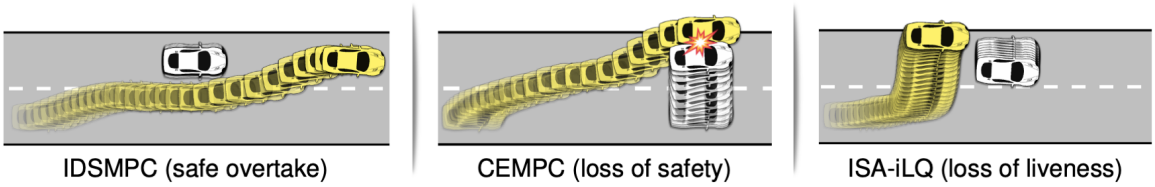


图 8. 第 7.2.2 节中的高速公路驾驶示例的一次试验。左图显示，机器人（黄色）使用 IDSMPC 在 6 秒内成功超越了人类驾驶车辆（白色）。中图和右图分别展示了使用 CEMPC 生成的不安全轨迹 $t = [0, 4.6]$ 秒，以及使用 ISA-iLQ 生成的轨迹，其中机器人未能在 10 秒内超越人类。

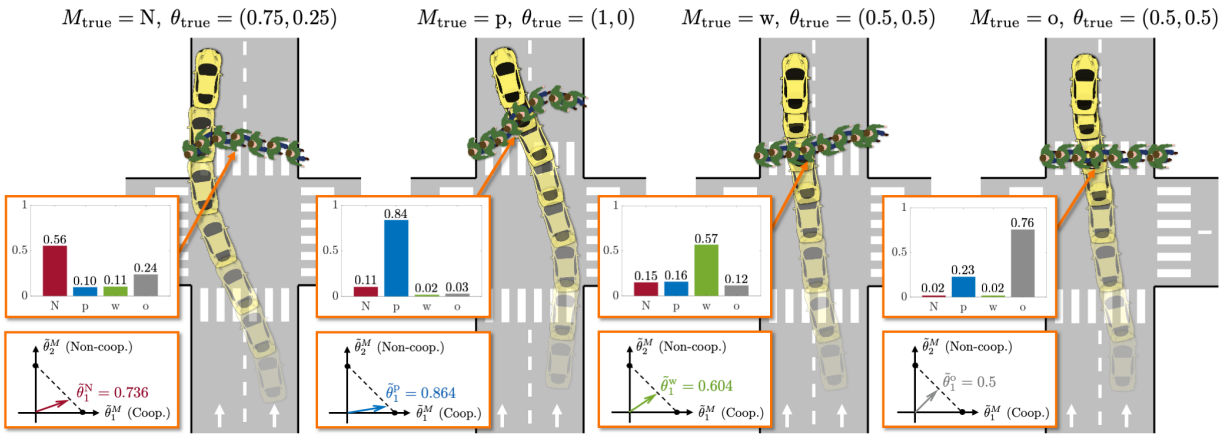


图9. 使用所提出的 IDSMPC 规划器进行示例2的仿真快照。整个仿真过程中，人类的众数保持不变。在橙色方块中，我们展示了 $t = 2.6$ 秒时机器人的运行信念 $p(M \mid \mathcal{I}_t)$ 以及 $\tilde{\theta}^M$，后者是将 θ^M 归一化为 1-单纯形后的最大后验均值。机器人能够快速识别人类的隐状态，并在全部四次试验中规划出无碰撞轨迹，同时考虑了预期的不确定性降低以及与人类的交互。

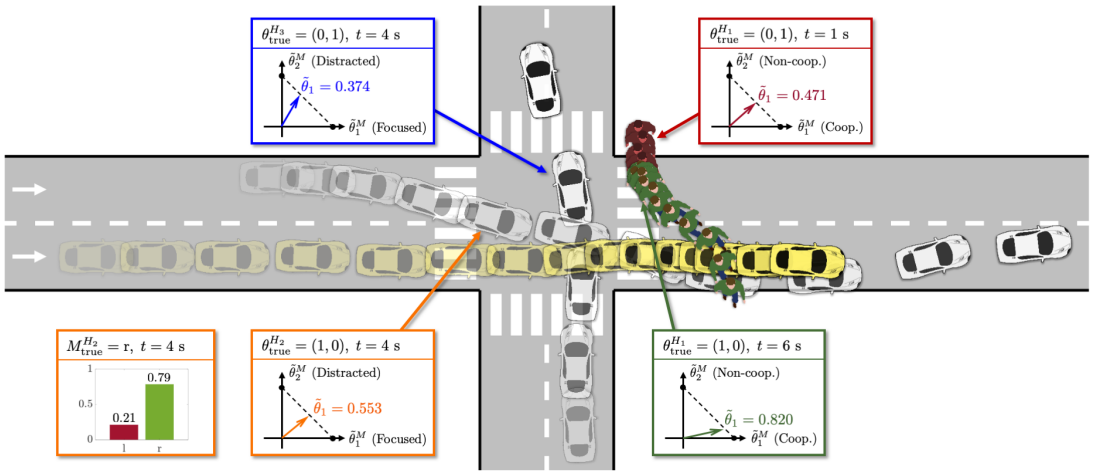


图10. 使用 IDSMPC 规划器时，包含一名行人（H_1）和两辆人类驾驶车辆（H_2 和 H_3）的多智能体交互场景的仿真快照及隐状态估计。方框中显示的是在 $t = 4$ 秒时其他智能体的连续隐藏状态估计 $\tilde{\theta}^M$（归一化至 1-单纯形）。对于 H_2，我们还展示了机器人的运行信念 $p(M^{H_2} \mid \mathcal{I}_t)$。

迅速降低对其他智能体的不确定性，并安全通过交通路口。

7.3 在*Waymo*运动数据集上的评估

本节中，我们为高速公路驾驶场景（第 7.2.2 节）提供更多仿真结果。该场景中的人类驾驶员轨迹取自 Waymo 开放运动数据集Sun 等人 (2020)。我们从原始数据集中筛选出 50 条具有不同人类动作和目标车道的轨迹数据。由于人类驾驶车辆的运动是通过重放轨迹数据生成的，因此可以认为人类完全没有安全意识，而机器人对此并不知情。图 1×10^4展示了 50 次试验所得的闭环代价统计数据（经归一化），其中所有规划器均经过安全屏蔽，因此所有试验都安全。11中绘出了相关数据。IDSMPC-SHARP、IDSMPC、EDSMPC、NDSMPC、CEMPC 和 ISA-iLQ 的（归一化）平均闭环代价分别为1。09、1。35、1。57、1。81、1。89和1。86，如图11中箱线图的中位标记所示。即使人类没有响应，双重控制规划器仍因主动不确定性降低而比非双重控制规划器更高效。IDSMPC-SHARP 的表现优于所有其他规划器，表明其适用于真实的交互场景。

7.4 统计显著性检验

为验证结果在统计上显著，从而确认第 7.1.3 节提出的三个假设，我们针对第 7.2.2 节、第 7.2.3 节和第 7.3 节报告的结果进行了方差分析（ANOVA）检验，将闭环代价作为响应变量，并将规划方法两两分组作为自变量。首先，对于第 7.2.2 节（$F(1, 92) = 38.18$，$p < 0.001$）、第 7.2.3 节（$F(1, 98) = 97.74$，$p < 0.001$）和第 7.3 节（$F(1, 98) = 9.01$，$p = 0.003$）报告的数据，我们发现非对偶随机模型预测控制（NDSMPC）与隐式双重随机模型预测控制（IDSMPC）之间存在显著的主效应，这验证了 H1。接着，通过比较显式双重随机模型预测控制（EDSMPC）与隐式双重随机模型预测控制（IDSMPC），我们再次发现，对于第 7.2.2 节（$F(1, 98) = 75.49$，$p < 0.001$）和第 7.2.3 节（$F(1, 98) = 74.34$，$p < 0.001$）报告的数据，二者之间存在不可忽略的效应，这支持了 H2。最后，对于第 7.2.2 节（$F(1, 98) = 39.08$，$p < 0.001$）和第 7.2.3 节（$F(1, 98) = 110.63$，$p < 0.001$）报告的数据，我们还发现，具有屏蔽感知与不具有屏蔽感知的隐式双重随机模型预测控制（IDSMPC）之间存在显著的主效应，这支持了 H3。在对第 7.3 节所得结果进行方差分析检验时，我们注意到，在尝试验证 H2（$F(1, 98) = 2.43$，$p = 0.123$）和 H3（$F(1, 98) = 5.24$，$p = 0.024$）时，F 值降低而 p 值升高；诚然，这未达到统计显著性。我们指出，这一结果符合预期，因为其他智能体的行为是根据 Waymo 开放运动数据集中的真实轨迹数据重放的，因此其他智能体不会对自车作出响应。因此，双重随机模型预测控制和屏蔽感知都高度依赖于对智能体间交互的推理，因而其效果不可避免地较弱。

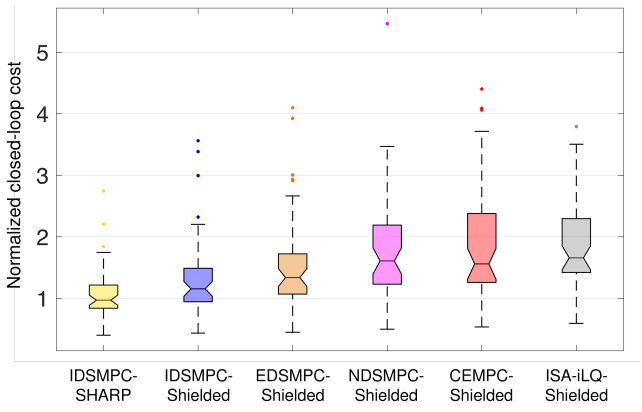


图 11. 高速公路驾驶场景（第 1×10^4节）的闭环代价（归一化处理）7.2.2，使用从 Waymo 开放运动数据集中选取的 50 条真实的人类驾驶车辆轨迹 Sun 等人（2020）。箱体中的中线、底边和顶边分别表示中位数、第 25 百分位数和第 75 百分位数。最大须线长度设为 1.5；若数据服从正态分布，则覆盖率为 99.3%。离群值以点表示。

8 硬件演示

8.1 实验设置

本节中，我们在示例 1（超车）中，使用两辆定制的 1/10 比例非完整约束赛车多智能体系统（MuSHR）自动驾驶车辆（一辆自车和一辆对等车辆），在位于加利福尼亚州圣何塞的本田研究院美国公司测试赛道上，演示所提出的 IDSMPC-SHARP 规划框架（第 6 节）。赛道的二维地图见图14。我们使用机器人操作系统（ROS）建立传感器、执行器与计算单元之间的通信。每个 MuSHR 机器人（图 12）都使用激光雷达，根据给定的赛道栅格地图和周边地标确定自身状态（位置、速度和朝向），并与另一辆车共享其当前状态。IDSMPC-SHARP 和所有对比规划器以 10Hz[¶] 的频率运行于一台配备 E5-2640 处理器的英特尔至强台式机上，并向 MuSHR 车辆发送规划轨迹；车载英特尔 NUC 迷你电脑上运行的基于 PID 的低层控制器负责跟踪该轨迹。台式机和 NUC 均运行 Ubuntu 20.04 LTS 操作系统。我们采用 Zhang 等人（2020）中的四维运动学自行车模型作为自车和其他智能体模型预测控制问题中的车辆动力学。本节中的所有模型预测控制问题均使用 Python 中的 CasADi Andersson 等人（2019）建模为非线性规划（NLP），并使用 IPOPT W¨achter 和 Biegler（2006）及其线性系统求解子程序 MA57 Duff（2004）实时求解。我们使用 Bansal 等人（2017）；Leung 等人（2020）的 HJ 可达性来构建屏蔽机制。基于 HJ 的屏蔽策略使用 OptimizedDP Bui 等人（2022）预先计算，并部署在车载英特尔 NUC 上。硬件实验所用的参数值见附录 A.2 中的表 3。

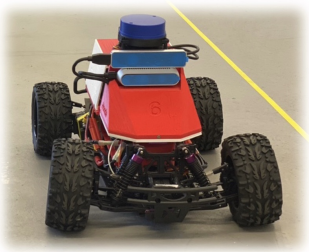


图12. 配备用于定位的激光雷达的 MuSHR 自动驾驶车辆。

8.1.1 对比方法。 我们在硬件实验中将提出的 IDSMPC-SHARP 规划器与另外四种规划器进行比较：

- **Oracle：** 一种模型预测控制规划器，根据当前时刻计算出的其他智能体的共享的规划方案来计算策略。
- **消融实验 I：** IDSMPC 移除了随机模型预测控制问题（18）中的安全屏蔽感知约束（20），因此与屏蔽无关。

[¶]我们的代码实现使所有控制策略的每个规划周期计算时间均低于 100 毫秒。

- **消融实验 II：** NDSMPC-SHARP 移除了信念动态（测量更新）（20d），仅使用转移模型（20e）更新信念状态。因此，该策略只能被动地降低不确定性。

- **基线：** CEMPC-SHARP 将第 7.1 节介绍的 CEMPC 规划器与屏蔽感知约束（18）相结合。

在所有四种对比方法中，只有 IDSMPC 策略会产生双重控制效应。所有规划器均使用相同的二次代价函数 ℓ 和 ℓ_F，并配备相同的基于 HJ可达性的安全屏蔽机制。

8.1.2 安全法规与性能指标。 由于车辆尺寸相对于车道宽度较大（车辆宽约 40 cm，车道宽约 60 cm），因此我们将满足以下两项规定的试验定义为安全：

- **SR1：** 自车与其他车辆不会相互碰撞（即发生物理接触），并且
- **SR2：** 自车至少有一个车轮接触或位于赛道边界内（如图 14）所示的外侧灰线处。

这项安全法规的制定受到典型汽车比赛的启发。在这类比赛中，与我们的情况类似，车辆可以在相对狭窄的车道上采取激进的超车操作。

受 Leung 等人 (2020) 的启发，我们考虑以下两个指标来评估安全性与效率之间的权衡：

- 安全指数，定义为

$$SI := \sum_{k=0}^{T_{\text{exp}}} \mathbf{1}\left[\ell_{\text{HJ}}\left(x_k\right) \geq 0\right] \ell_{\text{HJ}}\left(x_k\right),$$

其中 $T_{\text{exp}} > 0$ 是一次实验试验的总时间步数，$\ell_{\text{HJ}}(\cdot)$ 是 HJ 可达性计算中使用的运行代价（带符号距离函数 Bansal 等人 (2017年)），用于捕捉违反安全法规 SR1 和 SR2 的惩罚。该指标反映了安全违规的时间累积严重程度（相对于安全屏蔽安全集 Ω）。

- 效率指数，定义为

$$EI := \frac{1}{T_{\text{exp}}} \sum_{k=0}^{T_{\text{exp}}} \|x_k^e - x_{k,\text{ref}}^e\|_{Q_{\text{exp}}}^2 + \|u_k^e\|_{R_{\text{exp}}}^2 + \left(p_{y,k}^o - p_{y,k}^e\right),$$

其中，Q_{exp} 和 R_{exp} 是五种MPC运行代价 $\ell(\cdot)$共用的代价函数矩阵，其取值见表3；当自车尚未超越其他智能体时，$(p_{y,k}^o - p_{y,k}^e)$这一项会产生惩罚代价，否则则给予奖励。该指标通过结合跟踪精度、控制量幅值和超车进度，反映平均时间规划效率。

8.1.3 智能体策略。其他智能体采用配备屏蔽感知约束（18）的模型预测控制策略，在跟踪参考车道的同时避免与自车发生碰撞。约束（18）的软约束代价权重在不同试验中随机设置，以使其他智能体对安全的重视程度有所不同。当自车进入半径为 r_{detect}^o的检测圆时，经过一段时间延迟后，其他智能体会以固定概率变道至另一车道，为自车让行。让行概率和时间延迟均在不同试验中随机设置。请注意，其他智能体的（随机化）屏蔽感知模型预测控制策略所产生的交互行为，与示例1 中定义并用于ST-SMPC公式（20）的运动预测模型不同，也不使用该模型计算出的任何量。因此，我们的实验并非自我验证。

8.2 实验结果

我们首先在图 14中展示一组具有代表性的试验（每个规划器各一个）。为确保公平比较，我们在全部五次试验中为其他智能体选用了相同的策略参数。所有试验中，自车的参考车道均设为内侧车道。凭借主动不确定性降低和屏蔽感知鲁棒规划，我们提出的 I DSMPC-SHARP 规划器生成了安全且高效的轨迹，其质量与 Oracle MPC 策略生成的轨迹相当。相比之下，由于缺乏屏蔽感知，IDSMPC（消融实验 I）在自车近距离急左转以避让其他智能体时，触发了紧急安全屏蔽机动，从而产生了不平滑且摇摆不定的轨迹。由于 NDSMPC-SHARP（消融实验 II）移除了信念状态动力学（测量更新），它只能被动地学习其他智能体隐状态的取值。因此，自车变得过于保守，无法超越其他智能体。CEMPC-SHARP（基线）仅根据隐状态的最大后验估计做出决策，而该估计往往滞后于其他智能体的实际运动。实际上，当最大后验估计的隐状态为 NY（不让行）时，自车向右转，试图从外侧车道超越其他智能体；但其他智能体已明显表现出为自车让行的意图。因此，这次右转并无必要，也被认为降低了整体规划性能。

接下来，我们使用硬件实验开展了一项性能-安全权衡研究，结果绘于图13中。我们为每个规划器进行了 5 次试验，每次使用不同的随机种子，这会影响初始条件和其他智能体的策略参数。根据 *SR1* 和 *SR2* 安全法规，所有试验均满足安全要求。请注意，为了考虑模型失配和通信延迟，我们设计了比 SR1 和 SR2 更保守的 HJ 可达性安全屏蔽机制，因此几乎所有试验的安全指数 $SI > 0$ 均为如此。我们发现，IDSMPC-SHARP 在安全性和效率之间保持了良好的平衡，与 Oracle MPC 类似。值得注意的是，即使是最保守的 NDSMPC-SHARP 策略，也可能导致较差的安全性能。这是因为，在缺乏主动不确定性降低的情况下，面对不确定性较高的信念状态，SHARP 无法有效预测未来的安全屏蔽事件，最终导致系统

进入近乎不安全的区域，导致安全指数较大SI。由于策略对齐，CEMPC-SHARP 对实验随机性引起的信念波动较为敏感（鲁棒性较低）。因此，在所有五种规划器中，它的数据点方差最大。

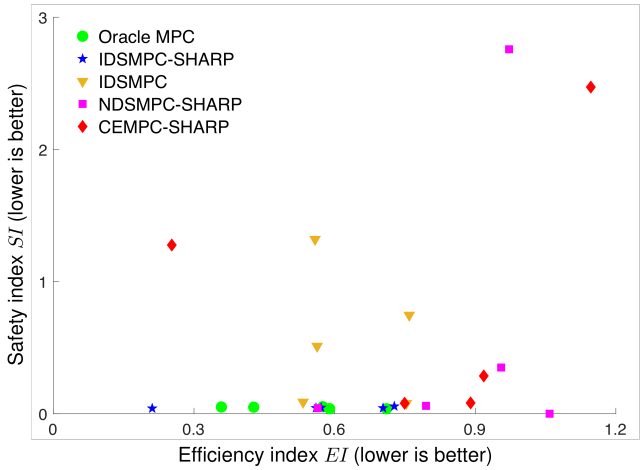


图 13. 硬件实验的性能与安全权衡（示例 1）。每个数据点均基于一次试验的闭环轨迹获得。左下角是理想区域，其中 EI 和 SI 都较低。

9 结论

我们提出了一种隐式双重控制方法，用于通过主动降低不确定性来进行交互规划。该方法得到的策略通过对双重控制问题的贝尔曼递归进行易于处理的近似，提高了规划效率，从而形成一种隐式双重、基于场景的随机模型预测控制策略，可自动在优化预期性能与获取有关其他智能体未来行为的信息之间实现高效平衡。

通过对双重控制策略施加安全屏蔽（即监督安全滤波器），可获得鲁棒安全保证。随机模型预测控制问题中加入了一项具有凸性的安全屏蔽感知约束，该约束基于近期提出的 SHARP 框架的改进版本推导而来。由此得到的 IDSMPC-SHARP策略使自车机器人智能体能够与其他智能体高效交互，同时考虑到其他智能体采取罕见动作所触发的高代价屏蔽机动所带来的风险。我们通过模拟驾驶示例以及使用 1/10 比例自动驾驶车辆开展的基于 ROS 的硬件实验，展示了所提出的框架。

9.1 局限性

尽管我们已在一个包含三个同级智能体的交互规划示例中展示了该方法，但如何将其推广到更多智能体仍是一项尚待解决的挑战。在最坏情况下，节点（因而决策变量）的数量会随着交互智能体数量和探索（双重控制）时间步数的增加而呈指数增长。尽管如此，基于场景的模型预测控制方法适用于中等规模的交互规划问题。此外，当前框架假设自车智能体能够完美观测其他智能体的状态和过去的动作，这在实际中往往并不现实。观测不确定性条件下交互规划的最新进展Isele 等人(2018b); Sunberg 和 Kochenderfer(2022); Hu 等人(2023) 为在此类情形下改进并推广我们的方法提供了一个有前景的方向。

9.2 未来方向

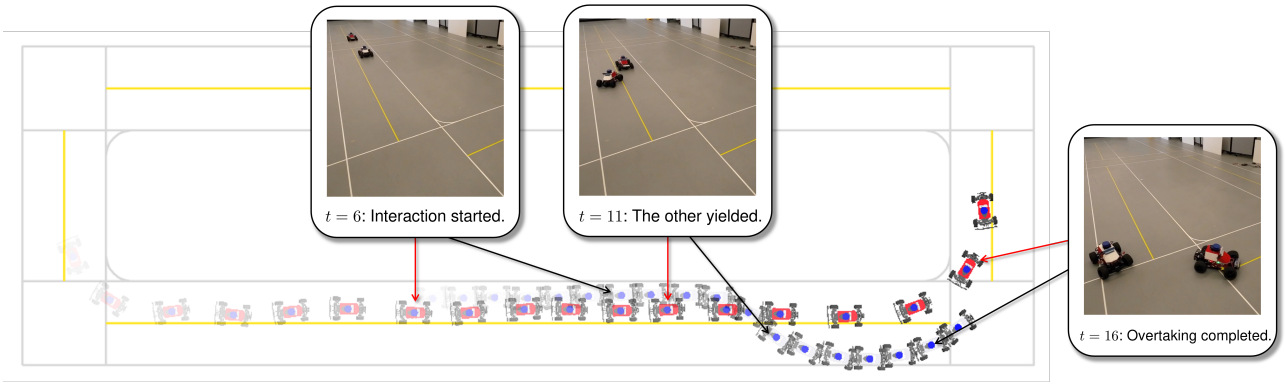
我们认为，我们的工作是迈向更广泛方法的一项重要进展，这类方法能够处理其他智能体行为的不同参数化形式，包括分位数层级k 模型Stahl II 和 Wilson(1994)；Tian 等人(2021)、基于学习的预测Isele 等人(2018a)、非线性意见动力学Bizyaeva 等人(2022)，以及依赖于状态和输入的信念状态转移动力学 $g^t(b^-,\ x,\ u^e)$，这些动力学能够捕捉自车的决策对其他智能体隐状态的影响。尽管本文关注的是机器人的自身表现，我们的方法也可以调整，以在人机协作场景中考虑社会协调和利他主义Toghi 等人(2022)。我们也期待在其他交互规划任务中测试我们的框架，例如在人机交互场景中开展人机无人机交互研究Fisac 等人 (2018b)，并邀请真实人类参与者参加。

致谢

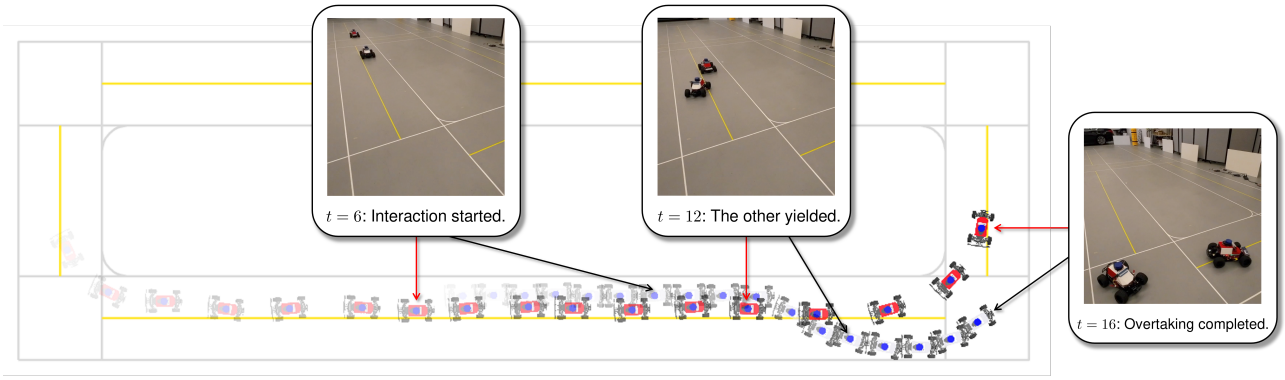
本研究由普林斯顿 SEAS Project X 创新基金和本田研究院（HRI）美国公司资助。本文仅反映作者的观点和结论，并不代表 HRI 或本田旗下任何其他实体的立场。作者感谢 Thang Lian、Huan D. Nguyen 和 Zhaobo K. Zheng 在硬件实验方面提供的帮助。作者还感谢 Faizan M. Tariq、Piyush Gupta、Aolin Xu、Yichen Song、Zixu Zhang 和 Kai- Chieh Hsu，就不确定性下的决策、模型预测控制和安全屏蔽展开了非常有益的讨论。

参考文献

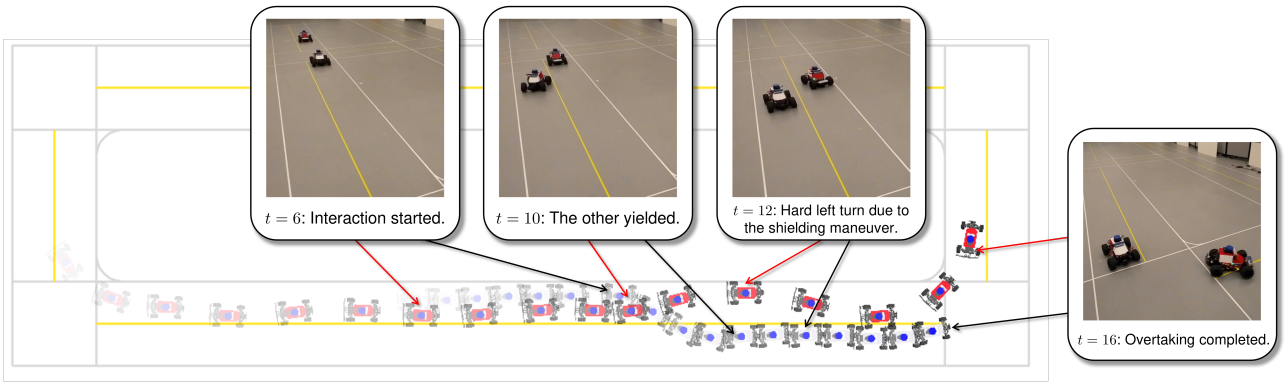
Agrawal A 和 Sreenath K（2017年）离散系统安全关键控制中的离散控制屏障函数及其在双足机器人导航中的应用。载于：机器人学：科学与系统会议论文集，第 13 卷。美国马萨诸塞州剑桥。DOI：10.15607/RSS.2017.XIII.073。Ames AD、Xu X、Grizzle JW 和 Tabuada P（2016年）用于安全关键系统的基于控制屏障函数的二次规划。*IEEE*自动控制汇刊 62(8)：3861–3876。Andersson JA、Gillis J、Horn G、Rawlings JB 和 Diehl M（2019）Casadi：用于非线性优化和最优控制的软件框架。数学规划计算11(1)：1–36。Arcari E、Hewing L、Schlichting M 和 Zeilinger M（2020a）针对具有参数和结构不确定性的系统的双重随机模型预测控制。载于：动力学与控制学习。第 894–903 页。Arcari E、Hewing L 和 Zeilinger MN（2020b）一种用于双重随机模型预测控制的近似动态规划方法。*IFAC-PapersOnLine*53(2)：8105–8111。Bae S、Saxena D、Nakhaei A、Choi C、Fujimura K 和 Moura S（2020）基于采用循环神经网络的模型预测控制，在密集交通中实现协作感知换道动作。载于：*2020*年美国控制会议（*ACC*）。IEEE，第 1209–1216 页。



(a) 硬件 实验试验示例 1 使用 Oracle MPC。绘制的轨迹存在轻微的激光雷达定位误差 r.



(b)硬件实验试验 示例 1 使用 IDSMPC-SHARP（提出的方法）。绘制的轨迹存在轻微的激光雷达定位误 or.

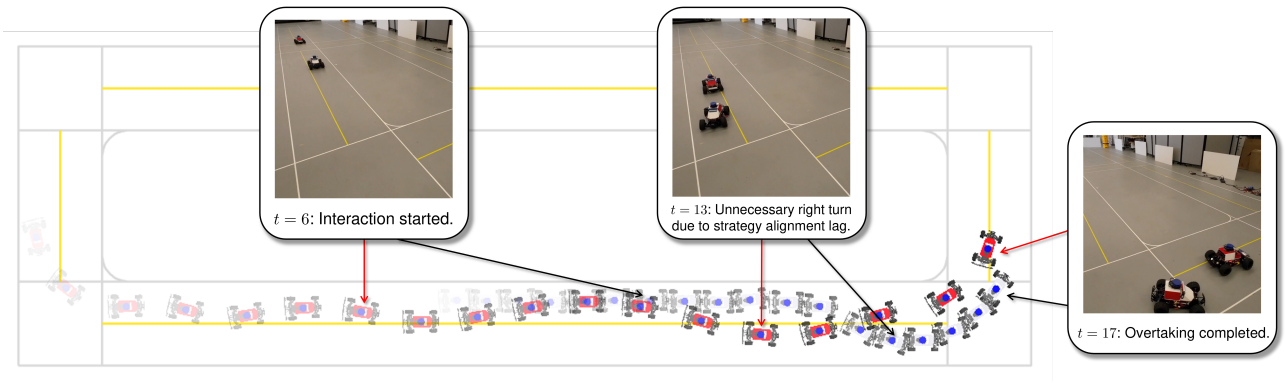


(c)硬件实验试验示例 1 使用 IDSMPC。绘制的轨迹存在轻微的激光雷达定位误差。



(d) 硬件 示例的实验试验 1 使用 NDSMPC-SHARP。绘制的轨迹受到轻微激光雷达定位误差的影响 ror.

图14。 所提出的 IDSMPC-SHARP 规划器与其他四种规划策略的比较。自车和其他车辆的车头分别涂为红色和白色。每隔 1 秒绘制一次车辆快照。所提出的 IDSMPC-SHARP 规划器准确预测了其他智能体的让行意愿，并完成了干净利落的超车操作，其表现与 Oracle MPC 产生的运动相当；Oracle MPC 根据其他智能体传达的未来计划计算策略。IDSMPC 规划器（不具备屏蔽感知）因遭遇意料之外的紧急安全屏蔽机动而导致性能下降。由于缺少主动不确定性降低，NDSMPC-SHARP 产生了过于保守的运动，未能完成超车任务。CEMPC-SHARP 对其他智能体隐状态最大后验估计中的误差十分敏感，并因未能及时识别其他智能体的让行意愿而执行了不必要的变道机动。视频请见 https://youtu.be/sJawMtO5QgY。 使用以下文件制作：*sagej.cls*



（**e**）硬件 示例1的硬件实验试验1 使用 CEMPC-SHARP。所绘制的轨迹存在轻微的激光雷达定位误差 r.

图14。所提出的 IDSMPC-SHARP 规划器与其他四种规划策略的比较（续）。

Bajcsy A、Siththaranjan A、Tomlin CJ 和 Dragan AD（2021）分析在线自适应的人类模型。载于：*2021*年 *IEEE* 机器人与自动化国际会议（*ICRA*）。IEEE，第2754–2760页。Bandyopadhyay T、Won KS、Frazzoli E、Hsu D、Lee WS 和 Rus D（2013）意图感知运动规划。载于：机器人学算法基础 *X*。施普林格，第475–491页。Bansal S、Chen M、Herbert S 和 Tomlin CJ（2017）Hamilton-Jacobi 可达性：简要综述与最新进展。载于：*IEEE*决策与控制会议（*CDC*）。第2242–2253页。Bar-Shalom Y 和 Tse E（1974）随机控制中的双重效应、确定性等价与分离。*IEEE*自动控制汇刊19(5)：494–500。Bas¸ar T 和 Olsder GJ（1998）动态非合作博弈论。工业与应用数学学会。Baseggio M、Beghi A、Bruschetta M、Maran F 和 Minen D（2011）用于驾驶模拟器运动提示算法设计的模型预测控制方法。载于：*2011*年第*14*届*IEEE*智能交通系统国际会议（*ITSC*）。IEEE，第692–697页。Bastani O（2021）通过模型预测屏蔽实现具有非线性动力学的安全强化学习。载于：*2021*年美国控制会议（*ACC*）。IEEE，第3488–3494页。Bastani O、Li S 和 Xu A（2021）通过统计模型预测屏蔽实现安全强化学习。载于：机器人学：科学与系统。Bellman R（1966）动态规划。《科学》153(3731)：34–37。Bernardini D 和 Bemporad A（2011）随机约束线性系统的稳定化模型预测控制。*IEEE*自动控制汇刊57(6)：1468–1480。Bhambri S、Bhattacharjee A 和 Bertsekas D（2022）Wordle 的强化学习方法：一种部分可观测马尔可夫决策过程/自适应控制方法。*arXiv* 预印本 *arXiv:2211.10298*。Bishop CM（2006）《模式识别与机器学习》。施普林格。Bizyaeva A、Franci A 和 Leonard NE（2022）具有可调灵敏度的非线性意见动力学。*IEEE*自动控制汇刊。Blanchini F（1999）控制中的集合不变性。*Automatica* 35(11)：1747–1767。

Bobu A, Bajcsy A, Fisac JF, Deglurkar S and Dragan AD（2020）量化人机演示与物理修正学习中的假设空间设定错误。*IEEE*机器人学汇刊36(3)：835–854。Bonzanini AD, Paulson JA and Mesbah A（2020）使用情景树，在状态与输入相关的不确定性下进行基于安全学习的模型预测控制。载于：*IEEE*决策与控制会议（*CDC*）。第2448–2454页。Bui M, Giovanis G, Chen M and Shriraman A（2022）Optimizeddp：用于最优控制与动态规划的高效、易用库。*arXiv*预印本 *arXiv:2204.05520*。Chen M, Herbert SL, Hu H, Pu Y, Fisac JF, Bansal S, Han S and Tomlin CJ（2021）Fastrack：用于实时运动规划和有保证安全跟踪的模块化框架。*IEEE*自动控制汇刊 66(12)：5861–5876。Chen Y, Rosolia U, Ubellacker W, Csomay-Shanklin N and Ames AD（2022）基于分支模型预测控制的交互式多模态运动规划。*IEEE*机器人与自动化快报7(2)：5365–5372。Chow Y, Nachum O, Duenez-Guzman E and Ghavamzadeh M（2018）一种基于李雅普诺夫的方法，用于安全强化学习。载于：神经信息处理系统进展。网址 https://dl.acm.org/doi/10.5555/3327757.3327904。Chung SJ, Paranjape AA, Dames P, Shen S and Kumar V（2018）空中群体机器人学综述。*IEEE*机器人学汇刊34(4)：837–855。Dayan P and Sejnowski TJ（1996）探索奖励与双重控制。机器学习25：5–22。Duff IS（2004）MA57——用于求解稀疏对称正定与不定系统的代码。*ACM*数学软件汇刊（*TOMS*）30(2)：118–144。Feldbaum AA（1960）双重控制理论 I。*Avtomatika iTelemekhanika*：1240–1249。Fisac JF, Akametalu AK, Zeilinger MN, Kaynama S, Gillula J and Tomlin CJ（2018a）面向不确定机器人系统中基于学习的控制的通用安全框架。*IEEE*自动控制汇刊 64(7)：2737–2752。Fisac JF, Bajcsy A, Herbert SL, Fridovich-Keil D, Wang S, Tomlin CJ and Dragan AD（2018b）基于置信度的人类预测下的概率安全机器人规划。载于：机器人学：科学与系统。Fisac JF, Bronstein E, Stefansson E, Sadigh D, Sastry SS and Dragan AD（2019）面向

自动驾驶车辆的分层博弈论规划。载于：IEEE机器人与自动化国际会议（ICRA）。第9590–9596页。Fridovich-Keil D, Ratner E, Peters L, Dragan AD and Tomlin CJ（2020）用于非线性多参与者一般和微分博弈的高效迭代线性二次近似。载于：IEEE国际机器人与自动化会议（ICRA）。第1475–1481页。Gill PE, Murray W and Saunders MA（2005）SNOPT：一种用于大规模约束优化的序列二次规划算法。《SIAM评论》47(1)：99–131。Guo H, Ji Y, Qu T and Chen H（2013）基于模型预测控制理解并建模人类驾驶员行为。IFAC会议论文集46(21)：133–138。Hardy J and Campbell M（2013）针对自动驾驶道路车辆的概率障碍物预测进行应急规划。IEEE机器人学汇刊29(4)：913–929。Heirung TAN, Foss B and Ydstie BE（2015）基于模型预测控制的双重控制与在线实验设计。《过程控制杂志》32：64–76。Hijab O（1984）熵与双重控制。载于：IEEE决策与控制会议（CDC）。第45–50页。Hsu KC, Hu H and Fisac JF（2023）安全过滤器：自主系统中安全关键控制的统一视角。《控制、机器人与自主系统年度评论》（即将出版）。Hu H and Fisac JF（2022）面向人机交互的主动不确定性降低：一种隐式双重控制方法。载于：《机器人学算法基础》第十五卷。施普林格，第385–401页。Hu H, Nakamura K and Fisac JF（2022）SHARP：面向安全高效人机交互的屏蔽感知鲁棒规划。IEEE机器人与自动化快报7(2)。Hu H, Zhang Z, Nakamura K, Bajcsy AV and Fisac JF（2023）欺骗博弈：闭合交互式机器人自主性的安全学习闭环。载于：机器人学习会议。Isele D（2019）密集交通中自动驾驶车辆的交互式决策。载于：2019年IEEE智能交通系统会议（ITSC）。IEEE，第3981–3986页。Isele D, Nakhaei A and Fujimura K（2018a）自动驾驶车辆上的安全强化学习。载于：2018年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，第1–6页。Isele D, Rahimi R, Cosgun A, Subramanian K and Fujimura K（2018b）利用深度强化学习驾驶自动驾驶车辆通过视线受阻的交叉路口。载于：2018年IEEE机器人与自动化国际会议（ICRA）。IEEE，第2034–2039页。Kabzan J, Hewing L, Liniger A and Zeilinger MN（2019）面向自动驾驶赛车的基于学习的模型预测控制。IEEE机器人与自动化快报4(4)：3363–3370。Klenske ED and Hennig P（2016）用于近似贝叶斯强化学习的双重控制。《机器学习研究杂志》17(1)：4354–4383。Koopman P（2018）重尾安全上限。载于：自动化与网联车辆系统测试研讨会，第1145卷。网址 https://users.ece.cmu.edu/~koopman/ pubs/koopman18_heavy_tail_ceiling.pdf。Lee CW, Nayeer N, Garcia DE, Agrawal A and Liu B（2020）确定自动化系统的运行设计域

通过评估风险的驾驶系统。载于：*2020*年*IEEE*智能车辆研讨会（*IV*）。IEEE，第1317–1322页。Leonard NE、Paley DA、Lekien F、Sepulchre R、Fratantoni DM和Davis RE（2007）集体运动、传感器网络与海洋采样。*IEEE*会刊95(1)：48–74。Leung K、Schmerling E、Zhang M、Chen M、Talbot J、Gerdes JC和Pavone M（2020）在人–机器人车辆交互规划框架中融入基于可达性的安全保障。《国际机器人研究期刊》39(10-11)：1326–1345。Li S和Bastani O（2020）用于随机动力学下安全强化学习的鲁棒模型预测屏蔽。载于：*IEEE*机器人与自动化国际会议（*ICRA*）。第7166–7172页。DOI:10.1109/ICRA40945.2020.9196867。Lindemann L、Hu H、Robey A、Zhang H、Dimarogonas D、Tu S和Matni N（2021）从数据中学习混合控制障碍函数。载于：机器人学习会议。PMLR，第1351–1370页。Liniger A、Domahidi A和Morari M（2015）基于优化的1:43比例遥控汽车自动驾驶赛车。《最优控制应用与方法》36(5)：628–647。L¨ofberg J（2004）YALMIP：用于在MATLAB中建模和优化的工具箱。载于：*CACSD*会议论文集。中国台北。Luce RD（1959）《个体选择行为》。英国牛津：John Wiley。Lucia S、Finkler T和Engell S（2013）应用于不确定性下半间歇聚合反应器的多阶段非线性模型预测控制。《过程控制杂志》23(9)：1306–1319。Mesbah A（2018）结合主动不确定性学习的随机模型预测控制：双重控制综述。《控制年度评论》45：107–117。Mitchell IM、Bayen AM和Tomlin CJ（2005）连续动态博弈中可达集的时变Hamilton–Jacobi表述。*IEEE*自动控制汇刊50(7)：947–957。Nair SH、Govindarajan V、Lin T、Meissen C、Tseng HE和Borrelli F（2021）面向交通交叉路口的多模态预测随机模型预测控制。*arXiv*预印本 *arXiv:2109.09792*。Ong SC、Png SW、Hsu D和Lee WS（2009）用于混合可观测性机器人任务的部分可观测马尔可夫决策过程。载于：机器人学：科学与系统，第5卷。第4页。Peters L、Fridovich-Keil D、Tomlin CJ和Sunberg ZN（2020）一般和微分博弈中基于推断的策略对齐。载于：第*19*届自主智能体与多智能体系统国际会议。第1037–1045页。Robey A等（2020）从专家示范中学习控制屏障函数。载于：*IEEE*决策与控制会议（*CDC*）。第3717–3724页。Rubenstein M、Cornejo A和Nagpal R（2014）千机器人群的可编程自组装。《科学》345(6198)：795–799。Sadigh D、Landolfi N、Sastry SS、Seshia SA和Dragan AD（2018）规划与人类协同的汽车行为：利用对人类行为的影响，针对人类内部状态进行规划和主动信息收集。《自主机器人》42(7)：1405–1426。

Santos M、Diaz-Mercado Y和Egerstedt M（2018）面向具有异构感知能力的多机器人团队的覆盖控制。*IEEE*机器人与自动化快报3(2)：919–925。Schildbach G和Borrelli F（2015）用于高速公路变道辅助的情景模型预测控制。载于：*IEEE*智能车辆研讨会（*IV*）。第611–616页。Schwarting W、Pierson A、Karaman S和Rus D（2021）信念空间中的随机动态博弈。*IEEE*机器人学汇刊37(6)：2157–2172。Sehr MA和Bitmead RR（2017）易于求解的对偶最优随机模型预测控制：医疗保健领域的一个示例。载于：*IEEE*控制技术与应用会议（*CCTA*）。第1223–1228页。Silver D和Veness J（2010）大型部分可观测马尔可夫决策过程中的蒙特卡罗规划。载于：神经信息处理系统会议。Somani A、Ye N、Hsu D和Lee WS（2013）DESPOT：带正则化的在线部分可观测马尔可夫决策过程规划。神经信息处理系统进展26。Srinivasa SS、Lancaster P、Michalove J、Schmittle M、Summers C、Rockett M、Smith JR、Chouhury S、Mavrogiannis C和Sadeghi F（2019）MuSHR：一种用于教育和研究的低成本开源机器人赛车。*CoRR*abs/1908.08031。Stahl II DO和Wilson PW（1994）关于参与者对其他参与者建模的实验性证据。《经济行为与组织杂志》25(3)：309–327。Sun P、Kretzschmar H、Dotiwalla X、Chouard A、Patnaik V、Tsui P、Guo J、Zhou Y、Chai Y、Caine B等（2020）自动驾驶感知的可扩展性：Waymo开放数据集。载于：*IEEE/CVF*计算机视觉与模式识别会议。第2446–2454页。Sunberg ZN、Ho CJ和Kochenderfer MJ（2017）推断交通参与者内部状态对高速公路自动驾驶的价值。载于：*2017*年美国控制会议（*ACC*）。IEEE，第3004–3010页。Sunberg ZN和Kochenderfer MJ（2018）适用于连续状态、动作和观测空间中部分可观测马尔可夫决策过程的在线算法。载于：第二十八届国际自动规划与调度会议。Sunberg ZN和Kochenderfer MJ（2022）通过考虑人类内部状态的部分可观测马尔可夫决策过程规划改进自动驾驶。*IEEE*《智能交通系统汇刊》。Swain DT、Couzin ID和Leonard NE（2011）用于鱼群行为实验的实时反馈控制机器鱼。*IEEE*会刊100(1)：150–163。Tian R、Sun L、Bajcsy A、Tomizuka M和Dragan AD（2022）通过具备置信度感知能力的博弈论人类模型，为人机交互提供安全保障。载于：*2022*年国际机器人与自动化会议（*ICRA*）。IEEE，第11229–11235页。Tian R、Sun L、Tomizuka M和Isele D（2021）针对以人为中心的机器人，基于对人类潜在状态的主动推理开展任意时刻博弈论规划。载于：*IEEE*机器人与自动化国际会议（*ICRA*）。第4509–4515页。Todorov E和Li W（2005）用于受约束非线性随机系统局部最优反馈控制的广义迭代LQG方法。载于：*2005*年美国控制会议论文集。IEEE，第300–306页。

Toghi B、Valiente R、Sadigh D、Pedarsani R 和 Fallah YP（2021）与人类驾驶员共情的协作式自动驾驶车辆。载于：*2021*年*IEEE/RSJ*国际智能机器人与系统会议（*IROS*）。IEEE，第 4517–4524 页。Toghi B、Valiente R、Sadigh D、Pedarsani R 和 Fallah YP（2022）自动驾驶中的社会协调与利他主义。*IEEE*《智能交通系统汇刊》。Tokekar P、Vander Hook J、Mulla D 和 Isler V（2016）用于精准农业的无人机与无人地面车辆协同系统的传感器规划。*IEEE*机器人学汇刊 32(6)：1498–1511。Wabersich KP 和 Zeilinger MN（2021）用于受约束非线性动力系统基于学习的控制的预测安全滤波器。*Automatica*129：109597。W¨achter A 和 Biegler LT（2006年）关于大型非线性规划内点滤波线搜索算法的实现。*Mathematical Programming*106(1)：25–57。Wu C、Kreidieh AR、Parvate K、Vinitsky E 和 Bayen AM（2021年）Flow：混合自主交通的模块化学习框架。*IEEE*机器人学汇刊。Zanardi A、Mion E、Bruschetta M、Bolognani S、Censi A 和 Frazzoli E（2021年）具有字典序偏好和社会效率纳什均衡的城市驾驶博弈。*IEEE*机器人与自动化快报6(3)：4978–4985。Zeilinger MN、Morari M 和 Jones CN（2014年）具有鲁棒稳定性保证的软约束模型预测控制。*IEEETrans. Autom.Control*：1190–1202。Zhang X、Liniger A 和 Borrelli F（2020）基于优化的避碰。*IEEE*控制系统技术汇刊29(3)：972–983。Ziebart BD、Maas AL、Bagnell JA 和 Dey AK（2008年）最大熵逆强化学习。载于：*AAAI*，第 8 卷。第 1433–1438 页。

符号列表

Th下列清单总结了本文中的重要符号 。

交互规划

b 信念状态g 信念状态动力学g^t 信念转移动力学$\tilde{g}$ 近似信念状态动力学u 控制输入d 外部扰动输入$_d$ 参数仿射动力学中的组合扰动f 联合动力学的自主部分e 自车ℓ 自车阶段代价函数ℓ_F 自车终端代价函数$\mathcal{I}$ 信息向量o 其他智能体

Q 状态-动作价值（Q 值）函数x 物理状态
θ 连续隐状态F 参数仿射动力学中的系统矩阵M 离散隐状态（众数）

情景树与模型预测控制

n 节点n_0 根节点$\mathcal{N}$ 节点集$\mathcal{N}^d$ 双重控制节点集$\mathcal{N}^e$ 利用节点集$\mathcal{N}^s$ 屏蔽节点集$\mathcal{L}$ 叶节点集$\mathfrak{p}(n)$ 节点的父节点 nP_n 节点的路径转移概率 $n_{-}P_n$ 节点的转移概率 n 自其父节点N 规划时域 N^d 双重控制时域N^e 利用时域

安全屏蔽与 SHARP

$\mathcal{U}$ 控制界限$\mathcal{D}$ 扰动界限$\mathcal{F}$ 失效集 Ω 安全集（控制不变集）Ω_n 节点处的近似局部安全集 $n\pi^f$ 安全过滤策略π^s 屏蔽策略h SHARP 中使用的鲁棒控制屏障函数

A 参数值

A.1 第7节中使用的参数值

表 2. 第7节中使用的规划器和仿真参数。

符号	值	定义
Δt	0.2 秒	采样时间
l_w	3.7 m	车道宽度
Q_{sim}	diag(1, 2, 1, 1)	自车状态代价矩阵
R_{sim}	diag(0.1, 1)	自车控制代价矩阵
Σ^d	$0.1I$	扰动协方差
N^d	2	双重控制时间步
N^e	4	利用时间步
K	2	θ^M的分支数
γ	0.5	RCBF约束参数
		SHARP中使用的

A.2 第 *8* 节中使用的参数值

表 3. 第 8 节中使用的规划器和实验参数。

符号	值	定义
Δt	0.2 秒	采样时间
l_w	0.6 米	车道宽度
Q_{exp}	diag(0.5,2, 1, 1)	自车状态代价矩阵
R_{exp}	diag(0.1, 0.5)	自车控制代价矩阵
Σ^d	$0.01I_8$	扰动协方差
N^d	3	双重控制时间步
N^e	15	利用时间步
K	2	θ^M的分支数
γ	0.5	RCBF约束参数-
		SHARP 中使用的参数
$r^o_{\text{检测}}$	2.5 m	检测圆半径
		其他智能体
$\bar{t}^o_d$	2.5 s	平均反应时间
		其他智能体

B 证明

B.1 引理证明*1*

回顾一下，式（d_t 7）中的组合扰动项是一个均值为零的高斯分布随机变量，其协方差在式（10）中定义。在给定 $x_t \in \mathcal{I}_t$、u^e_t 和 θ^M 的条件下，x_{t+1}（似然）的状态分布服从高斯分布，即 $p(x_{t+1} \mid u^e_t, \mathcal{I}_t; \theta^M, M) \sim \mathcal{N}(\mu^x_{t+1}, \Sigma^x_{t+1})$ 其均值和协方差为

$$\mu^x_{t+1} = F(x_t, u^e_t)\theta^M + \bar{f}(x_t, u^e_t),$$
$$\Sigma^x_{t+1} = \Sigma^d + B^o \Sigma^{u^o}_t(x_t, u^e_t; \theta^M) B^{o\top} = \Sigma^{\bar{d}}_t.$$

由于动力学（7）关于参数 θ^M 呈仿射形式，利用高斯分布的自共轭性质可得 $\mu^{\theta^M}_{t+1}$ 和 $\Sigma^{\theta^M}_{t+1}$。

□

B.2 引理证明*2*

令 $h(\delta x) = H^\top_{\tilde{n}} \delta x$。不失一般性，令 $\delta x_0 = x_0 - \bar{x}_0 = 0$，这意味着 $h(\delta x_0) = 0$。将（16）代入定义3中的条件2，得到：

$$\forall \tilde{d} \in \mathcal{D}:$$
$$H^\top_{\tilde{n}} \left[(A_{\tilde{n}} + (\gamma - 1)I)\, \delta x + B^e_{\tilde{n}} u^e + B^o_{\tilde{n}} u^o + \tilde{d} \right] \geq 0,$$

若存在 $u^e \in \mathcal{U}^e$ 使得以下条件成立，则该条件得到满足：

$$\min_{\tilde{d} \in \mathcal{D}} H^\top_{\tilde{n}} \left[(A_{\tilde{n}} + (\gamma - 1)I)\, \delta x + B^e_{\tilde{n}} u^e + B^o_{\tilde{n}} u^o + \tilde{d} \right] \geq 0,$$

其中，最优扰动 d^* 由（19）给出。因此，映射 $h(\delta x)$ 是系统（16）和安全集Ω_n的有效离散时间指数型鲁棒控制障碍函数。利用Agrawal 和 Sreenath (2017年)中的命题4，我们得出$\Omega_{\tilde{n}}$ 是鲁棒控制不变的。

□

C 实践方面

C.1 智能体的理性动作

（8）中使用的拉普拉斯近似要求均值函数（人类的理性动作）$\mu^M_i(x_t, u^e_t)$ 作为

基 Q 值函数 $Q_i^M\left(u_i^{M,o}\right.$ 的最大化器；$\left.x_t、\ u_t^e\right)$。在本文中，我们采用 Fridovich-Keil 等人（2020）的博弈论方法作为后端，在线计算 $Q_i^M(\cdot)$，该方法使用解析最大化器。当无法预先计算 $\mu_i^M(\cdot)$ 的表达式时，我们使用一种类似 Sadigh 等人（2018）的方法，通过在线优化计算局部最大化器 $\mu_i^M(\cdot)$。在假设 $Q_i^M(\cdot)$ 是一个可达到最大值的光滑函数这一温和条件下，我们可以将 $Q_i^M(\cdot)$ 对 u_i^M 的梯度设为 0。此条件可以通过微分代数方程（DAE）约束（Andersson 等人（2019））或情景树随机模型预测控制问题（20）中的惩罚成本来施加。

C.2 投影预测智能体的动作

采用近似智能体动作模型（9）的约束$u_{\tilde{n}}^o \in \mathcal{U}^o$ （见（20））可能不可行，因为预测的人类动作 $u_{\tilde{n}}^o$ 是由（无界）基函数与（无界）正态分布权重的加权和给出的。为解决这一问题，我们针对每个节点 $\tilde{n}$定义两个独立的决策变量：$\tilde{u}_{\tilde{n}}^o$，其必须等于基函数的采样线性组合；以及 $u_{\tilde{n}}^o$，其必须满足 $u_{\tilde{n}}^o \in \mathcal{U}^o$。在（20）中添加代价项 $C\|\tilde{u}_{\tilde{n}}^o - u_{\tilde{n}}^o\|_2$（20），并取较大的 $C > 0$ （我们使用$C = 10^8$）后，求解器会将 $u_{\tilde{n}}^o$ 设为集合 $\mathcal{U}^o$ 中距离符合采样条件的 $\tilde{u}_{\tilde{n}}^o$最近的点。这个可行的“投影”控制 $u_{\tilde{n}}^o$会代入（20）中的动力学。

C.3 初始化流程

由于问题（20）通常是一个大规模非凸优化问题，因此初始化对于实时快速、可靠地求解该问题至关重要。本文采用以下流程为（20）生成初始猜测：

1.（可选）将 θ^M 和 M 设为基于当前信念状态 b_t 得出的最大后验估计值，求解确定性等价模型预测控制。2. 求解一个非对偶随机模型预测控制，其情景树结构与对偶 SMPC 相同；对于所有对偶控制步骤，将信念状态动力学 $\tilde{g}(\cdot)$ 替换为 $g^t(\cdot)$，并将确定性等价模型预测控制的解用作初始猜测。3. 对情景树中的所有对偶控制节点，使用非对偶随机模型预测控制的解，通过 $\tilde{g}(\cdot)$ 前向传播信念状态。

步骤 1 是可选的，仅当步骤 2 中的非对偶随机模型预测控制难以直接求解时才需要。在第 7 节和8节中，我们将说明，即使（20）以非对偶随机模型预测控制的结果作为初始化，所得闭环轨迹仍存在显著差异。