arXiv:2305.01473v1 [cs.LG] 1 May 2023

参数化鲁棒马尔可夫链的高效敏感性分析 *

Thom Badings[1],Sebastian Junges[1], Ahmadreza Marandi[2],Ufuk Topcu[3],和 Nils Jansen[1]

[1] 拉德堡德大学，荷兰奈梅亨
thom.badings@ru.nl[2] 荷兰埃因霍温理工大学[3] 美国德克萨斯大学奥斯汀分校

摘要。 我们提出了一种用于参数化鲁棒马尔可夫链敏感性分析的新方法。这些模型纳入参数和概率分布集合，以缓解精确概率已知这一常常不切实际的假设。我们以不确定转移概率相对于期望奖励等指标的偏导数来衡量敏感性。作为主要贡献，我们提出了一种高效计算这些偏导数的方法。为使该方法适用于包含数千个参数的模型，我们进一步扩展该方法，以选出偏导数最大的参数子集 k 。我们的方法基于线性规划，并对参数给定取值处的这些规划进行求导。实验表明，该方法适用于包含超过一百万个状态和数千个参数的模型。此外，我们还将结果融入一种迭代学习方案，该方案可受益于专门的敏感性分析。

1 引言

离散时间马尔可夫链（MC）广泛应用于随机系统建模 [7]。一个经典假设是，马尔可夫链的所有概率都已精确确定；但在实践中，这一假设很难满足，甚至可能根本无法满足 [4]。鲁棒马尔可夫链（rMC），也称不确定马尔可夫链，通过采用概率分布集合来放宽这一假设，例如最简单的情况是采用概率区间 [11,38]。rMC 的典型验证问题是在该分布集合中这些概率的最坏情况取值 下，计算感兴趣的度量（例如期望累积奖励）的上界或下界 [51,58]。因此，对于这些集合中概率的任意取值，验证结果都具有鲁棒性。

我应该在哪里改进我的模型？作为一个贯穿示例，考虑一辆正在具有不同地形类型的环境中向目标位置导航的地面车辆。

* 本研究部分由 NWO 项目 NWA.1160.18.238 (PrimaVera)、ERC 初创基金 101077178 (DEUCE)，以及 ONR N00014-21-1-2502 和 AFOSR FA9550-22-1-0403 项目资助。

在每种地形类型上，车辆都有一定概率打滑而无法移动。假设我们获得了足够数量的 样本，以推断每种地形上打滑概率的上界和下界（即区间）。我们使用这些概率区间将网格世界建模为 rMC。然而，从 rMC 中无法看出，如果我们获得更多样本，模型（以及我们关注的度量）会如何变化。例如，如果我们针对某种地形再采样一次，rMC 的某些区间会发生变化，但我们如何预期验证结果会随之变化？如果验证结果不理想，我们应该针对哪种地形类型获得更多样本？

参数化鲁棒马尔可夫链。为了推理增加样本将如何改变我们的模型，进而影响验证结果，我们采用敏感性分析 [28]。为此，我们使用参数化鲁棒马尔可夫链（prMC），其概率分布集合由一组参数 [25],的函数定义，例如具有参数化上界和下界的区间。借助这些关于参数的函数，我们可以描述模型中各状态之间的依赖关系。为每个参数赋予数值称为实例化。将实例化应用于参数化鲁棒马尔可夫链后，可通过将参数的每次出现替换为其指定值，诱导出一个鲁棒马尔可夫链。对于这个诱导鲁棒马尔可夫链，我们计算给定度量的（鲁棒）值，并将该实例化对应的验证结果称为解。因此，我们可以将参数化鲁棒马尔可夫链与一个函数关联起来，该函数称为解函数，它将参数实例化映射为相应的值。

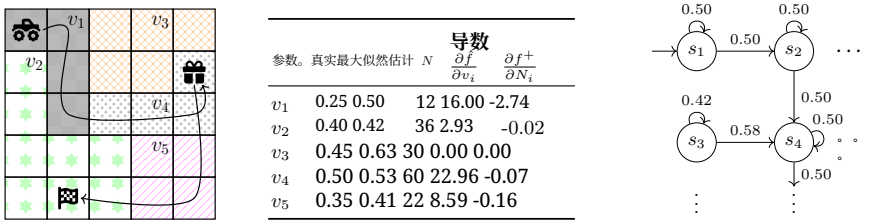
参数化鲁棒马尔可夫链的导数。在贯穿全文的示例中，我们选择用参数表示每种地形对应的样本量。自然地，该解函数的导数相对于各参数（即样本量）的变化，便对应于获取更多样本后解的预期变化。对于参数化马尔可夫链（pMC），参数实例化会得到一个确定的概率分布，关于此类链的微分已在 [33] 中得到研究。然而，对于参数化鲁棒马尔可夫链，如何计算导数以及导数在什么条件下存在，尚不明确。因此，我们考虑以下问题：

问题 1 （计算导数）。给定一个参数化鲁棒马尔可夫链和一个参数实例化，计算解函数（在该实例化处取值）相对于每个参数的偏导数。

我们的方法。我们通过求解一个参数化线性优化问题来计算参数化鲁棒马尔可夫链的导数。我们基于凸优化理论中关于如何对该优化问题的最优解求导的结果 [8,14]。我们还给出了导数存在的充分条件。

提高效率。然而，对于每个参数都显式计算其导数，无法扩展到包含数千个参数的更真实的模型。相反，我们注意到，为了确定应该针对哪个参数获取更多样本，我们并不需要显式地知道所有偏导数。相反，只需知道哪些参数具有最高的（或最低的，具体取决于应用场景）导数就可能足够了。因此，我们还解决以下（相关的）问题：

问题 2 *(k-最高阶导数)*。给定一个 prMC，确定具有最高（或最低）偏导数的 $|V|$ 参数。



参数。	真实	最大似然估计	N	导数 $\frac{\partial f}{\partial v_i}$	导数 $\frac{\partial f^+}{\partial N_i}$
v_1	0.25	0.50	12	16.00	-2.74
v_2	0.40	0.42	36	2.93	-0.02
v_3	0.45	0.63	30	0.00	0.00
v_4	0.50	0.53	60	22.96	-0.07
v_5	0.35	0.41	22	8.59	-0.16

(a) 网格世界。 (b) 最大似然估计值和导数。 (c) 马尔可夫链的部分。

图 1：网格世界环境（a）。车辆（ ）必须将包裹（ ）运送到仓库（ ）。我们在（b）中获得最大似然估计，从而得到（c）中的马尔可夫链。

我们提出了用于求解问题 2 的新颖高效方法。具体而言，我们设计了一个线性规划（LP），无需显式计算所有导数，即可找出偏导数最高（或最低）的 k 参数。该线性规划构成了一个用于求解问题 2 的多项式时间算法，实际应用中，其速度比显式计算所有导数快几个数量级，尤其是在参数数量较多时。此外，如果需要偏导数的具体值，还可以仅针对最终得到的 k 参数求解问题 1。在实验中，我们展示了可以为拥有超过一百万个状态和数千个参数的模型计算导数。

学习框架。 随机环境中的学习通常具有很强的数据密集性，可能需要数百万个样本，才能获得关于感兴趣的度量 [42,46] 的足够紧的界。已有多种基于抽样获取概率区间的方法，包括 Hoeffding 不等式等统计方法 [13]，以及通过迭代更新区间的贝叶斯方法 [56]。受降低学习算法样本复杂度这一挑战的启发，我们将所提出的方法嵌入一种迭代学习方案中，该方案充分利用了参数敏感度值的可用性。在实验中，我们展示了在学习具有数百个参数的未知马尔可夫链时，可以有效地利用导数信息来指导抽样。

贡献。 我们的贡献有三点：(1) 我们提出了首个用于计算参数化鲁棒马尔可夫链偏导数的算法。(2) 对于参数化马尔可夫链和参数化鲁棒马尔可夫链，我们都开发了一种高效的方法，用于确定导数最高的参数子集。(3) 我们将这些方法应用于迭代学习方案。我们在第 2 节概述我们的方法，并在第 3 节对问题陈述进行形式化。在第 4 节中，我们针对参数化马尔可夫链解决问题 (1) 和 (2)，并在第 5 节中针对参数化鲁棒马尔可夫链解决这两个问题。最后，第 6 节介绍学习方案和实验。

2 概述

我们将第 1 节中的示例加以扩展，以更具体地说明我们的方法。图 1a 所示的环境被划分为五个地形类型相同的区域。车辆可以沿四个基本方向移动。回想一下，

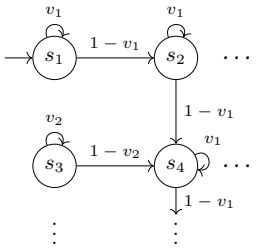


图 2：参数化马尔可夫链。

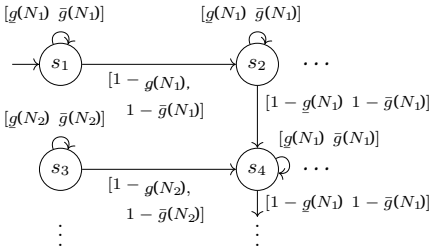


图3：参数化鲁棒马尔可夫链。

对于地形相同的所有状态，打滑概率也相同。车辆沿着指定路线收集包裹并将其送到仓库。我们的目标是估算完成任务所需的期望步数 $f^\star$ 。

概率估计。传统上，我们会通过抽样推导概率的最大似然估计（最大似然估计）。假设每种打滑概率使用 N 个样本，我们得到了图1b所示的粗略最大似然估计，由此得到图1c中的马尔可夫链。验证该马尔可夫链可知，在这些估计值下的期望行程时间（称为解）为 $f = 25.51$ 步数，与真实打滑概率下的 $f^\star = 21.62$ 步行程时间相差甚远。我们希望通过对其中一种地形类型获取更多样本来缩小这种验证与现实之间的差距。应该针对五种地形类型中的哪一种获取更多样本?

参数化模型。我们可以将网格世界建模为一个 pMC，即具有符号概率的马尔可夫链。该 pMC 的解函数是行进时间 f，它是这些符号概率的函数。我们在图 2 中示意了该 pMC 的四个状态。因此，最相关的参数自然定义为解函数的最大偏导数所对应的参数。如 图1b所示，参数 v_4 对 $\frac{\partial f}{\partial v_4} = 22.96$ 的偏导数最高，而 v_3 的导数为零，因为从未访问过与该参数相关的状态。 —

参数化鲁棒模型。上述方法未考虑每个最大似然估计中的不确定性。地形类型 v_4 的导数最高，但其样本量也最大，因此，与例如 v_1 相比，再对 v_4 抽样一次可能影响较小。那么， v_4 确实是获取额外样本的最佳选择吗? 图 3 展示了可用于回答这一问题的 prMC，其中我们使用（参数化）区间作为不确定集。对于所有地形类型，参数是样本量 $N_1, \ldots, N_5$（与 pMC 不同，在 pMC 中，参数表示打滑概率）。那么，如果我们为某种地形类型再获取一个样本，可以预期不确定集会如何变化?

参数化鲁棒马尔可夫链的导数。 我们使用 prMC 计算真实解 $f^\star$ 的上界 f^+ 。对于地形类型 v_i ，再获取一个采样（即将 N_i 增加 1）后，区间 $[\underline{g}(N_i)\ \bar{g}(N_i)]$ 的期望会缩小，进而降低我们的上界 f^+。这里， g 和 $\bar{g}$ 是将样本量 _

映射到区间界的函数。偏导数 $\frac{\partial f^+}{\partial N_i}$ 也展示了 prMC 的

在图1b中，其结果与 pMC 的导数结果大不相同。事实上，对 v_1 进行采样会使上界 f^+ 降低最多，因此我们最终决定对地形类型 v_1 进行采样，而不是 v_4。

高效微分。 我们指出，无需明确知道所有导数，就能确定从何处获取样本。相反，只需知道哪个参数的导数最大（或最小）即可。在本文余下部分，我们将开发高效方法，用于计算参数化马尔可夫链和参数化鲁棒马尔可夫链的解函数的全部偏导数，或仅计算其中 $k \in \mathbb{N}$ 个最大的偏导数。

支持的扩展。我们的方法适用于一般的参数化马尔可夫链和参数化鲁棒马尔可夫链，其参数可以在分布之间共享（从而捕捉依赖关系，这是参数化模型普遍具有的一项优势 [39]）。除了转移概率中的参数外，我们还可以处理参数化初始状态、奖励和策略。例如，我们可以在示例中使用参数来对监视无人机的策略进行建模，并计算这些参数的导数。

3 形式化问题陈述

令 $V = \{v_1, \ldots, v_\ell\}$, $v_i \in \mathbb{R}$ 为一个有限的有序参数集合。参数实例化是一个函数 $u\colon V \to \mathbb{R}$，它将一个参数映射为实值赋值。向量函数$\mathbf{u}(v_1, \ldots, v_\ell) = [u(v_1) \ldots, u(v_\ell)]^\top \in \mathbb{R}^\ell$ 表示从 V 到 u 中所有参数的有序实例化。参数 V 上的多项式集合为 $\mathbb{Q}[V]$。多项式 f 可以解释为一个函数$f\colon \mathbb{R}^\ell \to \mathbb{R}$，其中 $f(\mathbf{u})$是通过将 v 的每次出现替换为 $u(v)$得到的。我们用 $f[\mathbf{u}]$表示这些替换。

对于任意集合 X，令 $pFun_V(X) = \{f \mid f\colon X \to \mathbb{Q}[V]\}$ 表示所有从 X 映射到参数 V上的多项式的函数所构成的集合。我们用$pDist_V(X) \subset pFun_V(X)$ 表示 X上的参数化概率分布 集合，即满足以下条件的函数 $f\colon X \to \mathbb{Q}[V]$：对于所有参数实例化 $\mathbf{u}$，均有 $f(x)[\mathbf{u}] \in [0,1]$ 且 $\sum_{x \in X} f(x)[\mathbf{u}] = 1$。

参数化马尔可夫链 我们定义 pMC 如下：

定义 1（pMC）。一个 *pMC* $\mathcal{M}$ 是一个元组 (S, s_I, V, P)，其中 S 是有限状态集合，$s_I \in Dist(S)$ 是初始状态上的分布，V 是有限参数集合，$P\colon S \to pDist_V(S)$ 是参数化转移函数。

将实例化$\mathbf{u}$ 应用于 pMC，可通过将每个转移概率 $f \in \mathbb{Q}[V]$ 替换为 $f[\mathbf{u}]$，得到一个马尔可夫链 $\mathcal{M}[\mathbf{u}]$。我们考虑基于状态奖励函数 $R\colon S \to \mathbb{R}$的期望奖励度量。对 pMC 的每个参数实例化都会得到一个马尔可夫链，我们可以据此计算期望奖励度量 [7]的解。我们将把实例化映射为解的函数称为解函数。在保持图结构的实例化集合 [40]上，解函数是光滑的。具体而言，在实例化$\mathbf{u}$ 下，期望累积奖励对应的解函数 sol 写作如下：

$$\mathsf{sol}(\mathbf{u}) = \sum_{s \in S} \Big(s_I(s) \sum_{\omega \in \Omega(s)} \mathrm{rew}(\omega) \cdot \Pr(\omega, \mathbf{u}) \Big), \tag{1}$$

其中， $\Omega(s)$ 是从 $s \in S$ 出发的路径集合，$\text{rew}(\omega) = R(s_0) + R(s_1) + \cdots$ 是 $\omega = s_0 s_1 \cdots$ 上的累积奖励，而 $\Pr(\omega, \mathbf{u})$ 是路径 $\omega \in \Omega(s)$ 的概率。如果从状态 $s \in S$ 以概率 1 到达终止（汇）状态，则式（1）中关于 $\omega \in \Omega(s)$的无穷和 [52]存在。

参数化鲁棒马尔可夫链 由矩阵 $A \in \mathbb{R}^{m \times n}$ 和向量 $b \in \mathbb{R}^m$ 定义的凸多胞体 $T_{A,\,b} \subseteq \mathbb{R}^n$ 是集合 $T_{A,\,b} = \{p \in \mathbb{R}^n \mid Ap \leq b\}$。我们用 $\mathbb{T}_n$ 表示所有维数为 n的凸多胞体的集合，即

$$\mathbb{T}_n = \{T_{A,b} \mid A \in \mathbb{R}^{m \times n},\, b \in \mathbb{R}^m,\, m \in \mathbb{N}\}. \tag{2}$$

鲁棒马尔可夫链（rMC） [53,57] 是一个元组（S， s_I， $\mathcal{P}$），其中 S和 s_I的定义与参数化马尔可夫链相同，而不确定转移函数 $\mathcal{P}\colon S \to \mathbb{T}_{|S|}$ 将状态映射到凸多胞体 $T \in \mathbb{T}_{|S|}$。直观而言，rMC 是具有可能无限多个概率分布集合的马尔可夫链。为了获得这些马尔可夫链中任意一个的验证结果的鲁棒界，对手 会通过为每个 $s \in S$固定一个概率分布 $P(s) \in \mathcal{P}(s)$，以非确定性方式选择一个精确的转移函数 ˆ。

我们将 rMCs 扩展为其半空间由关于多项式$\mathbb{Q}[V]$ 在 V上定义的多面体。为此，令 $\mathbb{T}_n[V]$ 为所有此类参数化多胞体的集合：

$$\mathbb{T}_n[V] = \{T_{A,b} \mid A \in \mathbb{Q}[V]^{m \times n},\, b \in \mathbb{Q}[V]^m,\, m \in \mathbb{N}\}. \tag{3}$$

元素 $T \in \mathbb{T}_n[V]$ 可解释为函数 $T\colon\ \mathbb{R}^\ell \to 2^{(\mathbb{R}^n)}$，该函数将一个实例化 $\mathbf{u}$ 映射为 $\mathbb{R}^n$的（可能为空的）凸多胞体子集。集合$T[\mathbf{u}]$ 通过将 T中的每个 v_i 替换为 $u(v_i)$（对于所有 $i = 1, \ldots, \ell$）得到。

示例 *1*。图 3 中 prMC 的状态 s_1 的不确定集是参数化多胞体 $T \in \mathbb{T}_2[V]$，其参数集为单元素集合 $V = \{N_1\}$，使得

$$T = \big\{[p_{1,1}, p_{1,2}]^\top \in \mathbb{R}^2 \;\big|\; \underline{g}_1(N_1) \leq p_{1,1} \leq \bar{g}_1(N_1),\\ 1 - \bar{g}_1(N_1) \leq p_{1,2} \leq 1 - \underline{g}_1(N_1),\;\; p_{1,2} + p_{1,2} = 1\big\}.$$

我们使用参数化凸多胞体来定义 prMCs：

定义 2（prMC）。 *prMC* $\mathcal{M}_R$是一个元组 $(S, s_I, V, \mathcal{P})$，其中 S、 s_I和V的定义与 *pMCs*（定义 *1*）相同，并且 $\mathcal{P}\colon S \to \mathbb{T}_{|S|}[V]$ 是一个参数化且不确定的转移函数，将状态映射到参数化凸多胞体。

对 prMC 应用一个实例化 $\mathbf{u}$，通过将每个参数化多胞体 $T \in \mathbb{T}_{|S|}[V]$ 替换为 $T[\mathbf{u}],$，即可得到一个 rMC $\mathcal{M}_R[\mathbf{u}]$，即由具体的矩阵 $A \in \mathbb{R}^{m \times n}$ 和向量 $b \in \mathbb{R}^m$ 定义的多面体。不失一般性，我们考虑使到达一组终止状态 $S_T \subseteq S$之前的期望累积奖励最小化的对手。这个最小期望累积奖励 $\text{sol}_R(\mathbf{u})$，称为实例化 prMC $\mathcal{M}_R[\mathbf{u}]$, 上的鲁棒解，其定义为

$$\text{sol}_R(\mathbf{u}) = \sum_{s \in S} \Big(s_I(s) \cdot \min_{P \in \mathcal{P}[\mathbf{u}]} \sum_{\omega \in \Omega(s)} \text{rew}(\omega) \cdot \Pr(\omega, \mathbf{u}, P) \Big). \tag{4}$$

我们将函数 $\text{sol}_R\colon\ \mathbb{R}^\ell \to \mathbb{R}$ 称为鲁棒解函数。

关于参数化马尔可夫链（*pMC*）和参数化鲁棒马尔可夫链（*prMC*）的假设。对于 pMC 和 prMC，我们均假设在任何实例化下，转移都不会消失（图保持性）。也就是说，对于每个 $s, s' \in S$，我们有 $P(s)[\mathbf{u}](s')$（对于 pMC）和 $\mathcal{P}(s)[\mathbf{u}](s')$（对于 prMC）在所有实例化 $\mathbf{u}$ 下要么为零，要么严格为正。

问题陈述 设 $f(q_1, \ldots, q_n) \in \mathbb{R}^m$ 是一个可微多元函数，满足 $m \in \mathbb{N}$。我们用偏导数 表示 f 关于 q 的偏导数，记作$\frac{\partial x}{\partial q} \in \mathbb{R}^m$。 f 的梯度 将所有偏导数组合成一个向量，记作$\nabla_q f = [\frac{\partial f}{\partial q_1}, \ldots, \frac{\partial f}{\partial q_n}] \in \mathbb{R}^{m \times n}$。我们只使用相对于参数实例化 $\mathbf{u}$的梯度$\nabla_{\mathbf{u}} f$，因此在下文中简记为 ∇f。

鲁棒解函数在实例化$\mathbf{u}$处的梯度为 $\nabla \mathsf{sol}_R[\mathbf{u}] = \left[\left(\frac{\partial \mathsf{sol}}{\partial u}^{R}_{(v_1)}\right)[\mathbf{u}], \ldots, \left(\frac{\partial \mathsf{sol}}{\partial u}^{R}_{(v_\ell)}\right)[\mathbf{u}]\right]$。我们求解以下问题。

> 问题 *1.* 给定一个 *prMC* $\mathcal{M}_R$ 和一个参数实例化 $\mathbf{u}$，计算在 $\mathbf{u}$ 处求值的鲁棒解函数的梯度 $\nabla \mathsf{sol}_R[\mathbf{u}]$ 。

求解问题 1 的计算量随参数数量线性增长；如果参数数量较多，可能会带来显著开销。通常，只需获得具有最高阶导数的参数：

> 问题 *2.* 给定一个 *prMC* $\mathcal{M}_R$、一个实例化 $\mathbf{u}$ 和一个 $k \leq |V|$，计算 k 参数的一个子集 $V^\star$，使其偏导数最大。

对于这两个问题，我们分别为 pMCs（第 4 节）和 prMCs（第 5 节）提出了多项式时间算法。第 6 节定义了我们通过实证研究进行考察的问题变体。

4 pMC 解函数的求导

我们可以基于一个由 $|S|$ 线性方程构成的方程组，计算具有实例化 $\mathbf{u}$ 的 MC $\mathcal{M}[\mathbf{u}]$ 的解；此处针对的是期望奖励度量 [7]。令$x = [x_{s_1}, \ldots, x_{s_{|S|}}]^\top$和 $r = [r_{s_1}, \ldots, r_{s_{|S|}}]^\top$分别表示每个状态 $s \in S$ 中的期望累积奖励和瞬时奖励的变量。于是，对于一组终端（汇）状态 $S_T \subset S$，我们得到方程组

$$x_s = 0, \qquad \forall s \in S_T \tag{5a}$$

$$x_s = r_s + P(s)[\mathbf{u}]x, \quad \forall s \in S \backslash S_T. \tag{5b}$$

对所有 $s \in S_T$，令 $P(s)[\mathbf{u}] = 0$，并通过将所有 $s \in S$ 对应的行 $P(s)[\mathbf{u}]$ 依次叠放，定义矩阵 $P[\mathbf{u}] \in \mathbb{R}^{|S| \times |S|}$。那么，式 (5) 可写成矩阵形式 $(I_{|S|} - P[\mathbf{u}])x = r$。例如，可使用高斯消元或更高级的迭代方程求解器，高效求解式 (5) 中的方程组。

4.1 显式计算导数

我们对方程组 (5) 关于参数 $v_i \in V$ 的实例化$u(v_i)$ 求导，类似于例如 [33]。对于所有 $s \in S_T$，导数

$\frac{\partial x_s}{\partial u}(v_i)$ 显然为零。对于所有 $s \in S \setminus S_T$，根据乘积法则可得

$$\frac{\partial x_s}{\partial u(v_i)} = \frac{\partial P(s)x}{\partial u(v_i)}[\mathbf{u}] = (x^\star)^\top \frac{\partial P(s)^\top}{\partial u(v_i)}[\mathbf{u}] + P(s)[\mathbf{u}]\frac{\partial x}{\partial u(v_i)}, \tag{6}$$

其中，$x^\star \in \mathbb{R}^{|S|}$ 是方程（5）的解。对所有 $s \in S$，其矩阵形式为

$$\left(I_{|S|} - P[\mathbf{u}]\right)\frac{\partial x}{\partial u(v_i)} = \frac{\partial P x^\star}{\partial u(v_i)}[\mathbf{u}]. \tag{7}$$

式（1）中定义的解计算为 $\mathsf{sol}[\mathbf{u}] = s_I^\top x^\star$。因此，解函数关于 $u(v_i)$ 的偏导数的闭式为

$$\left(\frac{\partial \mathsf{sol}}{\partial u(v_i)}\right)[\mathbf{u}] = s_I^\top \frac{\partial x}{\partial u(v_i)} = s_I^\top \left(I_{|S|} - P[\mathbf{u}]\right)^{-1}\frac{\partial P x^\star}{\partial u(v_i)}[\mathbf{u}]. \tag{8}$$

问题 *1* 的算法。 下面给出一种求解参数化马尔可夫链的 问题 1 的算法。式 (8) 给出了解函数的偏导数的闭式表达式，该偏导数是式 (5) 中向量 $x^\star$ 的函数。然而，由于需要求 $(I_{|S|} - P[\mathbf{u}])$ 的逆，通常直接求解式 (7) 中的方程组更有效。具体而言，通过以下步骤可得到解关于 $u(v_i)$ 的偏导数：(1) 使用 $\mathbf{u}$ 求解式 (5)，得到 $x^\star \in \mathbb{R}^{|S|}$；(2) 使用 $|S|$ 个未知数求解式 (7) 中的方程组，以求得该向量 $x^\star$。对于所有 $|V|$ 个参数，我们重复步骤 2。因此，通过求解 $|V|+1$ 个线性方程组（每个方程组含有 $|S|$ 个未知数），即可求解问题 1。

4.2 计算 k 个最高阶导数

为了求解参数化马尔可夫链的 问题 2，我们提出一种方法，无需显式计算所有导数，即可仅计算偏导数最高（或最低）的 $k \leq \ell = |V|$ 参数。不失一般性，我们聚焦于最高阶导数。我们可以通过求解一个包含二元变量的组合优化问题来确定这些参数，其中 $z_i \in \{0,1\}$ 适用于 $i = 1, \ldots, \ell$。我们的目标是构造该优化问题，使得$z_i^\star = 1$ 的最优值意味着参数 $v_i \in V$ 属于 k 高阶导数的集合。具体而言，我们构造如下混合整数线性规划问题（MILP） [59]:

$$\underset{y \in \mathbb{R}^{|S|},\, z \in \{0,1\}^\ell}{\text{maximize}} \quad s_I^\top y \tag{9a}$$

$$\text{subject to } \left(I_{|S|} - P[\mathbf{u}]\right) y = \sum_{i=1}^{\ell} z_i \frac{\partial P x^\star}{\partial u(v_i)}[\mathbf{u}] \tag{9b}$$

$$z_1 + \cdots + z_\ell = k. \tag{9c}$$

约束（9c）确保式（9）的任意可行解恰有 k 个非零项。由于矩阵（$I_{|S|} - P[\mathbf{u}]$）按构造可逆（例如，参见 [52]），

式（9）对于 y 中的每个 $z \in \{0,1\}^\ell$ 取值都存在唯一解。因此，目标值 $s_I^\top y$ 是满足 $z_i = 1$ 的参数 $v_i \in V$ 的导数之和。由于我们最大化该目标，式（9）的最优解 $y^\star$，$z^\star$ 保证对应于使式（8）中解的导数最大的 k 参数。我们针对该混合整数线性规划陈述这一正确性主张：

命题 1. 设 $y^\star$, $z^\star$ 是式 (9) 的一个最优解。那么，集合 $V^\star = \{v_i \in V \mid z_i^\star = 1\}$ 是具有最大导数的 $k \leq \ell$ 个参数的子集。

集合 $V^\star$ 可能不是唯一的。然而，要解决问题 2，只需获得一组 k 参数，使其偏导数达到最大。因此，集合 $V^\star$ 为问题 2 提供了一个解。我们指出，要解决问题 2 中求 k 个最小导数的问题，只需将式 (9a) 中的目标改为最小化 $s_I^\top y$。

线性松弛。当 ℓ 和 k 的值较大时，式 (9) 中的混合整数线性规划在计算上难以处理。因此，我们通过对混合整数线性规划进行线性松弛 来计算集合 $v^\star$ 。具体而言，我们将二元变量 $z \in \{0,1\}^\ell$ 松弛为连续变量 $z \in [0,1]^\ell$。由此得到式 (9) 的以下线性规划松弛：

$$\underset{y \in \mathbb{R}^{|S|},\, z \in \mathbb{R}^\ell}{\text{maximize}} \; s_I^\top y \tag{10a}$$

$$\text{subject to } \left(I_{|S|} - P[\mathbf{u}]\right) y = \sum_{i=1}^{\ell} z_i \frac{\partial P x^\star}{\partial u(v_i)}[\mathbf{u}] \tag{10b}$$

$$0 \leq z_i \leq 1, \quad \forall i = 1, \ldots, \ell \tag{10c}$$

$$z_1 + \cdots + z_\ell = k. \tag{10d}$$

将式 (10) 中线性规划松弛的解记为 y^+，z^+ 。关于此类整数问题的线性松弛，详见 [35,45]。在我们的情形下，具有二进制值 $z_i^+ \in \{0,1\}$的线性规划松弛的每个最优解 y^+，z^+ 也都是混合整数线性规划的最优解，从而得到以下定理。

定理 1。 式 (10) 中的线性规划松弛具有一个最优解 y^+, z^+，其中 $z^+ \in \{0,1\}^\ell$ （即每个最优变量 z_i^+ 都是二进制值），且每个这样的解也是式 (9)中混合整数线性规划的最优解。

证明。 由 $\left(I_{|S|} - P[\mathbf{u}]\right)$的可逆性可知，式 (9) 等价于

$$\underset{z \in \{0,1\}^\ell}{\text{maximize}} \; \sum_{i=1}^{\ell} z_i \left(s_I^\top \left(I_{|S|} - P[\mathbf{u}]\right)^{-1} \frac{\partial P x^\star}{\partial u(v_i)}[\mathbf{u}] \right) \tag{11a}$$

$$\text{subject to } z_1 + \cdots + z_\ell = k. \tag{11b}$$

式（11）的线性松弛是一个线性规划，其可行域的顶点均为整数（例如参见 [36]）。因此，式（11）及其松弛式（10）都有一个整数最优解 z^+，该解构造了式（9）中的 $z^\star$。□

二进制解 $z^+ \in \{0,1\}^\ell$ 是式（10）中线性规划可行集的顶点。可以将单纯形法线性规划求解器设置为返回此类解。[1]问题 *2* 的算法。我们给出一种算法，通过两个步骤求解由两步组成的 pMCs 上的问题 2。首先，对于 pMC $\mathcal{M}$ 和参数实例化$\mathbf{u}$，我们求解式（7）中的线性方程组，得到 $x^\star$ 的解 $\mathsf{sol}[\mathbf{u}] = s_I^\top x^\star$。其次，我们固定若干参数 $k \leq \ell$ 并求解线性规划松弛（10）。然后，按照命题 1 中的定义，得到具有最大导数的参数集合 $V^\star$。参数集合 $V^\star$ 是问题 2 的一个解。

5 参数化鲁棒马尔可夫链解函数的微分

我们将重点转向参数化鲁棒马尔可夫链。回顾一下，解 $\mathsf{sol}_R[\mathbf{u}]$ 是针对不确定性的最坏情况实现计算得出的，称为鲁棒解。我们推导出以下方程组，其中，与 pMC 一样，$x \in \mathbb{R}^{|S|}$ 表示每个状态的期望累积奖励。

$$x_s = 0, \qquad \forall s \in S_T \tag{12a}$$

$$x_s = r_s + \inf_{p \in \mathcal{P}(s)[\mathbf{u}]} \left(p^\top x\right), \quad \forall s \in S \setminus S_T. \tag{12b}$$

由于式 (12b) 中包含内部下确界，直接求解式 (12) 等价于求解一个非线性方程组。鲁棒优化 [11] 中的标准方法是利用每个内部下确界对应的对偶问题，例如 [19,51]中采用的方法。对于每个 $s \in S$，$\mathcal{P}(s)$ 是一个参数化凸多胞体 $T_{A,\,b}$ 其定义如式 (3) 所示。该多胞体的维数取决于后继状态的数量，而后继状态的数量通常远小于状态总数。为了明确表示后继状态的数量，我们用 $\mathsf{post}(s) \subseteq S$ 表示 $s \in S$ 的后继状态，并定义 $T_{A,\,b} \in \mathbb{T}_{|\mathsf{post}(s)|}[V]$ 其中 $A_s \in \mathbb{Q}^{m_s \times |\mathsf{post}(s)|}$和 $b_s[\mathbf{u}] \in \mathbb{Q}^{m_s}$（回顾一下，$m_s$ 是该多面体的半空间数量）。那么，对于每个 $s \in S \setminus S_T$，式 (12b) 中的下确界为

$$\text{minimize } p^\top x \tag{13a}$$

$$\text{subject to } A_s[\mathbf{u}]p \leq b_s[\mathbf{u}] \tag{13b}$$

$$\mathbb{1}^\top p = 1, \tag{13c}$$

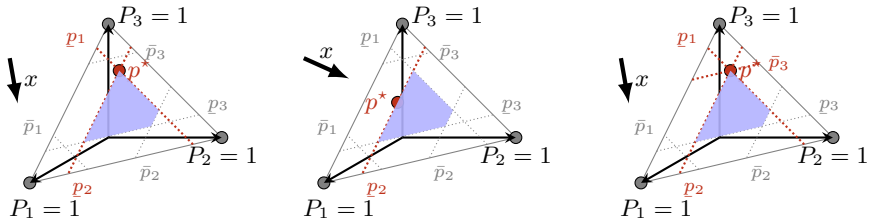
其中，$\mathbb{1}$ 表示适当维度的全 1 列向量。设 $x_{\mathsf{post}(s)} = [x_s]_{s \in \mathsf{post}(s)}$ 为与 $\mathsf{post}(s)$ 中（按顺序排列的）后继状态相对应的决策变量向量。式（13）的对偶问题，其对偶变量为 $\alpha \in \mathbb{R}^{m_s}$ 和 $\beta \in \mathbb{R}$（详情参见 [10]），可写为：

$$\text{maximize } -b_s[\mathbf{u}]^\top \alpha - \beta \tag{14a}$$

$$\text{subject to } A_s[\mathbf{u}]^\top \alpha + x_{\mathsf{post}(s)} + \beta \mathbb{1} = 0 \tag{14b}$$

$$\alpha \geq 0. \tag{14c}$$

[1] 即使得到一个非顶点解 y^+，z^+ 我们也可以对 z^+采用任意平局裁决规则，从而强制每个 z_i^+ 取二进制值，并保持式（10d）中的和不变。



(a) 最优值定义良好。 (b) 最优值不唯一。 (c) 活跃约束过多。

图4：三个多面体不确定集（蓝色阴影），其中向量 x、最坏情况点 $p^\star$ 和活跃约束以红色标出。

将该对偶问题代入式（12b），可得到以下线性规划，其决策变量为 $x \in \mathbb{R}^{|S|}$，并且对每个 $s \in S$都有 $\alpha_s \in \mathbb{R}^{m_s}$ 和 $\beta_s \in \mathbb{R}$：

$$\begin{aligned}
&\text{最大化} \quad s_I^\top x && && \text{(15a)}\\
&\text{满足} \quad x_s = 0, && \forall s \in S_T && \text{(15b)}\\
&\qquad x_s = r_s - \left(b_s[\mathbf{u}]^\top \alpha_s + \beta_s\right), && \forall s \in S \setminus S_T && \text{(15c)}\\
&\qquad A_s[\mathbf{u}]^\top \alpha_s + x_{\mathsf{post}(s)} + \beta_s \mathbb{1} = 0, \quad \alpha_s \geq 0, && \forall s \in S \setminus S_T. && \text{(15d)}
\end{aligned}$$

将式（12）重写为式（15）要求 $s_I \geq 0$，这显然成立，因为 s_I 是一个概率分布。设 $x^\star, \alpha^\star, \beta^\star$ 为式（15）的一个最优点。该最优值的 $x^\star$ 元素也是式（12）[11]的一个最优解。因此，式（4）中定义的鲁棒解为 $\mathsf{sol}_R[\mathbf{u}] = s_I^\top x^\star$。

5.1 使用参数化马尔可夫链计算导数（以及何时不起作用）

为求解问题 1，我们对如何计算参数化鲁棒马尔可夫链的鲁棒解提供一些直观解释。式 (12b) 中的下确界在每个集合 $\mathcal{P}(s)$[内找到使 $(p^\star)^\top x$ 最小化的最坏情况 点 $p^\star$] 。对于一个包含三个概率区间 $p_i \leq p_i \leq \bar{p}_i$, $i = 1,2,3$ 的不确定集，图 4a 对这一最小化过程进行了可视化。给定优化方向 x（图 4a 中的箭头），点 $p^\star$（红点）位于约束 $\underline{p}_1 \leq p_1$ 和 $\underline{p}_2 \leq p_2$ 均处于活跃状态的顶点处，因此可以达到。[2] 由此，我们得到多面体中使 $(p^\star)^\top x$ 最小化的点为 $p^\star = [p_1, p_2, 1 - p_1 - p_2]^\top$。通过这一过程，我们可以为每个状态 $s \in S$ 获得一个最坏情况 点 $p_s^\star$。我们可以利用这些点，将参数化鲁棒马尔可夫链转换为诱导参数化马尔可夫链，其转移函数为 $P(s) = p_s^\star$，适用于每个状态 $s \in S$。

当参数发生微小变化时，图4a中的点 $p^\star$ 会平滑变化，其闭式表达式（即函数形式）保持不变。因此，直观上我们可以应用第 4 节中的方法，计算诱导参数化马尔可夫链的偏导数。然而，如以下两个边界情况所示，这种方法并非总是有效。

[2] 不等式约束 $gx \leq h$ 在最优解下处于活跃状态 $x^\star$，当且仅当 $gx^\star = h$ [14]。

1. 考虑图4b，其中 x 所定义的优化方向与不确定集的一个面平行。在这种情况下，最坏情况点 $p^\star$ 并不唯一，但优化方向 x 发生无穷小变化就会使该点再次落在某个顶点上。我们应选择哪个点来获得诱导参数化马尔可夫链（这种选择会影响导数吗）？
2. 考虑图4c，其中点 $p^\star$ 处有超过 $|S|-1$ 个活跃约束。注意，减小 $\bar{p}_3$ 会改变点 $p^\star$，而增大 $\bar{p}_3$ 则不会。事实上，最优点 $p^\star$ 会随着多面体的半空间非光滑地变化。因此，解也会非光滑地变化，导数因而未定义。我们该如何处理这种情况?

这些例子表明，通过获取每个点 $p_s^\star$ 来计算诱导 pMC 的导数可能很棘手，在某些情况下甚至完全不可行。接下来，我们将介绍一种方法，它直接推导出一组线性方程，以根据式 (15) 中线性规划的解，求出参数化鲁棒马尔可夫链的导数（全部导数或仅 k 个最大导数）；该方法能够从本质上识别出上述导数未定义的边界情况。

5.2 显式计算导数

现在，我们提出一种专门的方法，用于判断 prMC 的解函数的导数是否存在；如果存在，则计算该导数。从图4可以看出，点 $p^\star$ 只有在图4a中由两个活跃约束确定时才是唯一的，且其导数光滑。只有一个活跃约束时（图4b），该点是欠定的；而有三个活跃约束时（图4c），导数可能不光滑。在一般情况下，恰好有 $n-1$ 个活跃约束（其对应的面不平行），是概率单纯形中 n维点 $p^\star$ 唯一且平滑变化的充分条件。

最优对偶变量。对于式（15）中的每个 $s \in S \setminus S_T$，最优对偶变量 $\alpha_s^\star \geq 0$ 表明多面体 $A_s[\mathbf{u}]p \leq b_s[\mathbf{u}]$ 的哪些约束处于激活状态，即矩阵 $A_s[\mathbf{u}]$ 中哪些行 $a_{s,i}[\mathbf{u}]$ 满足 $a_{s,i}[\mathbf{u}]p^\star = b_s[\mathbf{u}]$。具体而言，$\alpha_{s,i} > 0$ 的取值意味着 i^{th} 约束处于激活状态，而 $\alpha_{s,i} = 0$ 则表示约束 [14] 未处于激活状态。我们将 $E_s = [e_1, \ldots, e_{m_s}] \in \{0,1\}^{m_s}$ 定义为一个向量，其二进制值 $e_i \, \forall i \in \{1, \ldots, m_s\}$ 由 $e_i = [\![\alpha_{s,i}^\star > 0]\!]$ 给出。[3] 此外，将 $\mathbf{D}(E_s)$ 表示为一个对角线上为 E_s、其余位置均为零的矩阵。我们将式（15）中的线性规划简化为一个线性方程组，该方程组仅对每个 $s \in S \setminus S_T$ 对应的最坏情况点 $p_s^\star$ 下处于激活状态的约束进行编码：

$$x_s = 0, \qquad \forall s \in S_T \tag{16a}$$

$$x_s = r_s - \left(b_s[\mathbf{u}]^\top \mathbf{D}(E_s)\alpha_s + \beta_s\right), \qquad \forall s \in S \setminus S_T \tag{16b}$$

$$A_s[\mathbf{u}]^\top \mathbf{D}(E_s)\alpha_s + x_{\mathsf{post}(s)} + \beta_s \mathbb{1} = 0, \quad \alpha_s \geq 0, \quad \forall s \in S \setminus S_T. \tag{16c}$$

微分。然而，方程（16）何时有（唯一的）最优解？为了提供一些直观理解，我们将方程组写成矩阵形式，即

[3] 我们使用艾弗森括号：当 x 为真时为 $[\![x]\!] = 1$，否则为 $[\![x]\!] = 0$。

$C\begin{bmatrix} x & \alpha & \beta \end{bmatrix}^\top = d$，为简洁起见，我们省略了对矩阵 C 和向量 d 的明确定义。显然，如果矩阵 C 是非奇异的，那么方程（16）具有唯一解。这要求矩阵 C 为方阵；当对于每个 $s \in S \setminus S_T$，都有 $|\text{post}(s)| = \sum E_s + 1$ 时即可满足这一条件。换句话说，s 的后继状态数量等于多面体的活跃约束数量加一。这证实了我们在第 5.1 节中关于具有 $|\text{post}(s)| = 3$ 个后继状态的多面体的先前直觉，该多面体需要 $\sum_{i=1}^{m_s} E_i = 2$ 个活跃约束。

让我们将这一关于如何计算参数化鲁棒马尔可夫链导数的直觉形式化。我们可以通过乘积法则对方程（16）中的方程组进行求导，从而计算解 $x^\star$ 的导数，其方法与第 4 节中的方法非常相似。我们将在下面的定理中陈述这一关键结果。

定理2。 给定一个 *prMC* $\mathcal{M}_R$ 和一个实例化 $\mathbf{u}$，计算 $x^\star, \alpha^\star, \beta^\star$ 以用于方程 (15)，并选择一个参数 $v_i \in V$。偏导数 $\frac{\partial x}{\partial u}(v_i)$、$\frac{\partial \alpha}{\partial u}(v_i)$ 和 $\frac{\partial \beta}{\partial u}(v_i)$ 可通过求解以下线性方程组得到。

$$\frac{\partial x_s}{\partial u(v_i)} = 0, \qquad \forall s \in S_T \tag{17a}$$

$$\frac{\partial x_s}{\partial u(v_i)} + b_s[\mathbf{u}]^\top \mathbf{D}(E_s)\frac{\partial \alpha_s}{\partial u(v_i)} + \frac{\partial \beta_s}{\partial u(v_i)} = -(\alpha_s^\star)^\top \mathbf{D}(E_s)\frac{\partial b_s[\mathbf{u}]}{\partial u(v_i)}, \quad \forall s \in S \setminus S_T \tag{17b}$$

$$A_s[\mathbf{u}]^\top \mathbf{D}(E_s)\frac{\partial \alpha_s}{\partial u(v_i)} + \frac{\partial x_{\text{post}(s)}}{\partial u(v_i)} + \frac{\partial \beta_s}{\partial u(v_i)}\mathbb{1} = -(\alpha_s^\star)^\top \mathbf{D}(E_s)\frac{\partial A_s[\mathbf{u}]}{\partial u(v_i)}, \quad \forall s \in S \setminus S_T. \tag{17c}$$

该证明可通过对方程（16）应用乘积法则得到，具体证明见附录 A.1。因此，为了计算参数 $v_i \in V$ 对应的导数，我们需要求解一个规模为 $|S| + \sum_{s \in S \setminus S_T} |\text{post}(s)|$ 的线性方程组。利用定理2，我们可以得到解函数可微的充分条件。

引理 1. 将方程（17）中的线性方程组写成矩阵形式，即

$$C\left[\frac{\partial x}{\partial u(v_i)}, \frac{\partial \alpha}{\partial u(v_i)}, \frac{\partial \beta}{\partial u(v_i)}\right]^\top = d, \tag{18}$$

对于 $C \in \mathbb{R}^{q \times q}$ 和 $d \in \mathbb{R}^q$，$q = |S| + \sum_{s \in S \setminus S_T} |\text{后验}(s)|$，它们由方程（17）隐式给出。解函数 $\text{解}_R[\mathbf{u}]$ 在实例化 $\mathbf{u}$ 处可微，当矩阵 C 非奇异时，此时可得 $\left(\frac{\partial \text{解}}{\partial u}^R(v_i)\right)[\mathbf{u}] = s_I^\top \frac{\partial x}{\partial u}(v_i)$。

证明. 解函数的偏导数为 $\frac{\partial \text{解}}{\partial u}^R(v_i)[\mathbf{u}] = s_I^\top \frac{\partial x^\star}{\partial u}(v_i)$，其中 $\frac{\partial x^\star}{\partial u}(v_i)$ 是方程（16）的解（的一部分）。因此，只要方程（16）存在（唯一）解，解函数就可微；而矩阵 C 非奇异即可保证这一点。因此，引理 1 中的结论得证。□

问题*1*的算法。我们使用定理2求解参数化鲁棒马尔可夫链的 问题 1，方法与参数化马尔可夫链类似。给定一个参数化鲁棒马尔可夫链 $\mathcal{M}_R$ 和一个实例化 $\mathbf{u}$，我们首先求解

利用公式 (15) 获得 $x^\star$，$\alpha^\star$，$\beta^\star$。其次，我们使用 $\alpha_s^\star$ 来计算每个 $s \in S \setminus S_T$的活跃约束向量 E_s 。第三，对于每个参数 $v \in V$，我们求解公式 (17) 中的方程组。因此，为了计算解函数的梯度，我们求解一个线性规划和 $|V|$ 个线性方程组。

5.3 计算 k-最高阶导数

我们直接采用第 4.2 节中的相同过程，计算具有 $k \leq \ell$ 最高阶导数的参数。与 pMC 类似，我们可以通过求解一个 MILP 来计算 k 最高阶导数；该 MILP 对每个参数 $v \in V$都对式（17）中的方程组进行编码。为简洁起见，我们在附录 A.2 中给出该 MILP。它与式（9）具有相同的结构，因此我们可以采用相同的线性松弛，得到一个具有定理 1 所述保证的线性规划。换言之，求解线性规划松弛可得具有最大导数的参数集合 $V^\star$，如命题 1 所述。对于参数化鲁棒马尔可夫链，该集合 $V^\star$ 是问题 2 的一个解。

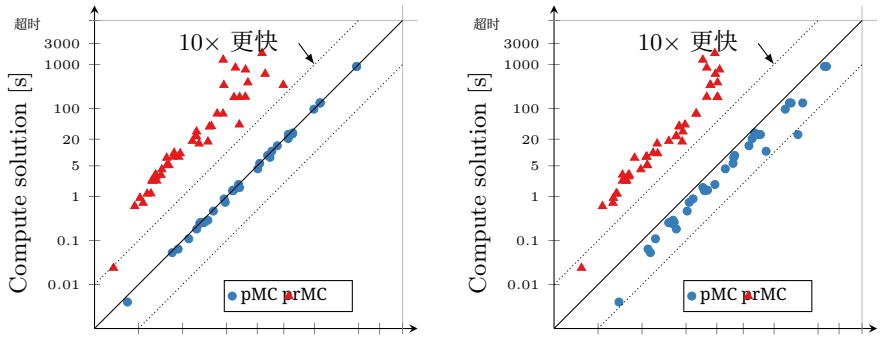
6 数值实验

我们通过实验来回答以下有关我们方法的问题：

1. 从计算复杂度和运行时间来看，计算所有导数是否可行，尤其是与计算（鲁棒）解相比？2. 仅计算 k 最高阶导数与计算所有导数相比如何？3. 我们能否应用该方法，有效确定在学习框架中应对哪些参数进行采样？

让我们简要总结一下回答这些问题所涉及的计算。首先，计算 pMC 的解 $\mathsf{sol}(\mathbf{u})$（定义见式 (1)），意味着求解式 (5) 中的线性方程组。同样，计算 prMC 的鲁棒解 $\mathsf{sol}_R(\mathbf{u})$，意味着求解式 (15) 中的线性规划。接下来，解决问题 1 （即计算所有 $|V|$ 偏导数），相当于针对每个参数 $v \in V$ 求解一个线性方程组（即 prMC 对应式 (5)，prMC 对应式 (17)）。相较之下，解决问题2（即计算导数最大（或最小）的参数子集$V^\star$ ），对于 pMC 而言，意味着求解式 (10) 中的线性规划（或 prMC 对应的等价线性规划），然后根据命题 1 提取参数子集 $V^\star$ 。

问题 *3*：计算 k阶导数。 问题 2 的一个解是一组 $V^\star$ k 参数，但不包括导数的计算。不过，也可以直接求得每个参数 $v \in V^\star$对应的实际导数 $\left(\frac{\partial}{\partial u}\mathsf{sol}\,(v)\right)[\mathbf{u}]$。具体而言，我们针对 $V^\star$ 中的 k 个参数求解问题 1，从而得到所有 $v \in \overline{V^\star}$对应的偏导数。值得注意的是，对于 $k = 1$，导数可直接由式 (10) 中线性规划的最优值 $s_I^\top y^+$ 得出，因此无需执行这一步。我们将计算 k 阶导数的实际值称为问题 *3*。



一个导数 [s] 最高阶导数 [s]

图 5：计算单个导数（左侧，问题 1）或最高阶导数（右侧，问题 3）的运行时间（对数刻度），与计算解 $\mathrm{sol}[\mathbf{u}]$/$\mathrm{sol}_R[\mathbf{u}]$ 的运行时间进行比较。

设置。 我们使用 Python 3.10 实现了该方法，使用 Storm [34] 解析参数化马尔可夫链，使用 Gurobi [30] 求解线性规划问题，并使用 SciPy 稀疏求解器求解方程组。所有实验均在配备 4GHz 英特尔酷睿 i9 CPU 和 64 GB RAM 的计算机上运行，超时时间为一小时。我们的实现可在 https://doi.org/10.5281/zenodo.7864260 获取。

网格世界基准测试。 我们使用第 2 节示例中的网格世界的扩展版本，其状态数量超过一百万，地形类型最多可达 10 000 种。车辆仅向右或向下移动，两种移动方向的概率均为 50%（离开网格时会绕回另一侧）。打滑仅发生在向下移动时，与第 2 节中的示例略有不同，打滑意味着车辆会移动两个单元格而非一个。我们针对每种打滑概率获取 $N = 500$ 至 1 000 个样本。对于 pMCs，我们使用基于这些样本进行最大似然估计（$\frac{\bar{p}}{N}$，其中 $\bar{p}$ 为样本均值）得到的概率；而对于 prMCs，我们则利用霍夫丁不等式推断概率区间（详情请参见 Q3）。

文献中的基准测试。我们还使用了标准基准测试套件 [32,43] 中马尔可夫链和马尔可夫决策过程（MDP）的多个参数化扩展实例。我们也使用了 [5,22] 中的 pMC 基准测试，因为这些模型的参数比传统基准测试更多。我们通过围绕 pMC 的概率构造概率区间，将这些基准测试扩展为参数化鲁棒马尔可夫链。

结果。 所有基准测试的结果均列于附录B中的表 2 和表 3。

Q1. 计算解与导数

我们研究在 p(r)MC 上计算导数是否可行。具体而言，我们比较在 p(r)MC 上计算导数（问题1和问题3）所需的时间与求解这些模型所需的时间。

表 1：所选网格世界模型的模型规模、运行时间和导数。

模型统计信息				验证		问题 1	问题 3		导数	
Type	$\|S\|$	$\|V\|$	转移数	解$_{(R)}$ [u]	时间 [s]	全部导数 [s]	$k=1$ [s]	$k=10$[s]	最高误差	百分比
pMC	5000	50	14995	5.07	1.39	3.32	2.64	2.69	1.54e+00	0.0
pMC	5000	100	14995	5.05	1.36	4.17	2.63	2.66	1.28e+00	0.0
pMC	5000	921	14995	4.93	1.87	19.92	4.52	2.87	1.20e+00	0.0
pMC	80000	100	239995	8.01	25.54	98.47	45.18	46.87	1.95e+00	0.0
pMC	80000	1000	239995	8.01	25.64	612.97	48.92	58.20	2.08e+00	0.0
pMC	80000	9831	239995	7.93	25.52	5,650.25	347.76	1,343.59	2.10e+00	0.0
pMC	1280000	100	3839995	12.90	902.52	4,747.43	1,396.51	1,507.77	3.32e+00	0.0
pMC	1280000	1000	3839995	12.79	902.67	37,078.12	1,550.45	1,617.27	3.18e+00	0.0
pMC	1280000	10000	3839995	超时[b]		—	—	—	—	—
prMC	5000	100	14995	136.07	23.46	3.55	0.60	1.58	-1.26e-02	-0.0
prMC	5000	921	14995	138.74	29.82	25.23	0.85	1.09	-4.44e-03	-0.0
prMC	20000	100	59995	2,789.77	1,276.43	15.68	2.40	2.70	-4.96e-01	-0.1
prMC	20000	1000	59995	2,258.41	339.96	159.70	3.53	4.09	-9.51e-02	-0.0
prMC	80000	100	239995	超时[b]		—	—	—	—	—

[a] 根据 10 个参数的运行时间外推至全部 $|V|$ 个参数。[b] 验证 p(r)MC 时发生超时（1 小时），计算导数时未发生超时。

在图 5 中，我们展示了所有基准测试中计算解（定义于公式 (1) 和 (4)）的时间，并将其与计算问题 1 的单个导数（左图）或问题 3 所得所有参数的最高阶导数（右图）的时间进行比较。左图中的点 (x, y) 表示，计算单个导数耗时 x 秒，而计算解耗时 y 秒。位于（中心）对角线之上的线表示相较于计算解的时间获得了加速；位于上方对角线上的点表示获得了 $10\times$ 加速或更高。

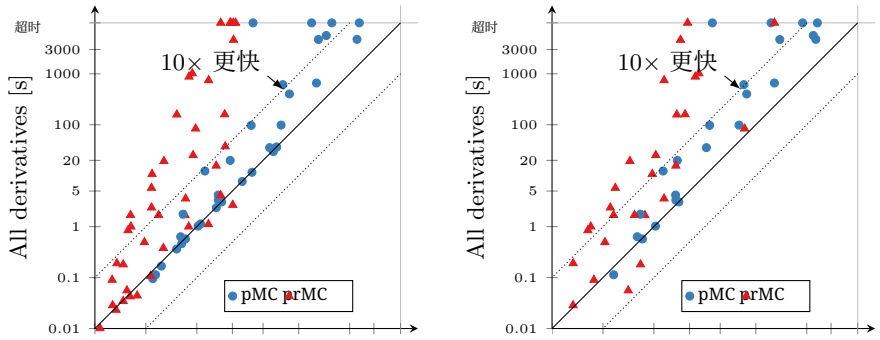
一个导数。图 5 左侧的图显示，对于参数化马尔可夫链，计算解和计算单个导数所需的时间大致相同。这是预期之中的，因为这两个问题本质上都是求解一个包含 $|S|$ 个未知数的方程组。回顾一下，对于参数化鲁棒马尔可夫链，计算解意味着求解公式（15）中的线性规划，而计算导数则需要求解一个方程组。因此，与计算解相比，计算参数化鲁棒马尔可夫链的导数所需的开销相对较小，这一点也得到了图 5 中实验结果的证实。

最高阶导数。图 5 右图显示，对于参数化马尔可夫链，计算最高阶导数比计算解稍慢（计算最高阶导数的线性规划比计算解的方程组耗时更长）。另一方面，对于 prMC，计算最高阶导数仍比计算解便宜得多。因此，如果我们无论如何都要使用 prMC，那么计算导数的成本相对较低。

问题 2. 仅计算 k 个导数带来的运行时间提升

我们希望了解求解问题 3 相较于求解问题 1 的计算优势。对于问题 2，我们考虑所有具有 $|V| \geq 10$ 个参数的模型。

网格世界基准测试的部分结果摘录见表 1。回顾一下，在获得（鲁棒）解之后，求解问题 1 相当于



$k = 1$ 最高阶导数 [s] $k = 10$ 高阶导数 [s]

图 6：计算最高阶导数（左）或 10 个最高阶导数（右）（问题 3）的运行时间（对数刻度），与计算所有导数（问题 1）的运行时间对比。

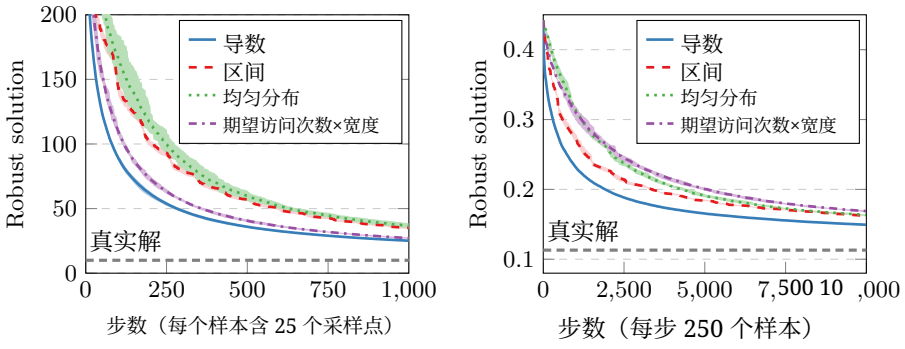
用于求解 $|V|$ 线性方程组，而问题 3 涉及求解一个线性规划和 k 方程组。从表 1 中可以清楚看出，计算k 导数比计算全部 $|V|$ 导数快几个数量级，尤其是当参数总数较多时。

我们比较了计算所有导数（问题 1）与仅计算 $k = 1$ 个或 10 个最高阶导数（问题 3）的运行时间。图 6 左图显示了计算 $k = 1$的运行时间，右图则显示了计算 $k = 10$ 个最高阶导数的运行时间。图 6 的解读方式与图 5 相同。从图 6 可以看出，仅计算 k 个最高阶导数通常能显著加快计算速度，往往超过 10 倍（模型非常小时除外）。此外，$k = 1$与 $k = 10$之间的差异很小，这表明求解问题 2 后提取实际导数的成本相对较低。

数值稳定性。 虽然我们的算法是精确的，但为了提高效率，实现中使用了浮点运算。为评估数值稳定性，我们将最高阶导数（针对 $k = 1$ 求解问题 3）与通过将参数扰动 1×10^{-3} 得到的导数经验近似进行比较。两者之间的差异（表 1 和表 2 中的 ‘误差百分比’ 列）微乎其微，表明我们的实现具有足够的数值稳定性，能够准确地计算导数。

Q3. 学习框架中的应用

降低样本复杂度是应对不确定性进行学习的一项关键挑战 [42,46]。尤其是在随机环境中学习非常依赖数据，现实应用往往需要数百万个样本，才能为感兴趣的度量提供紧界 [15]。受这一挑战的启发，我们将该方法应用于学习框架，以探究与替代探索策略相比，能否利用导数有效地引导探索。



(a) 湿滑网格世界。 （b）无人机运动规划。

图 7：网格世界（a）和无人机（b）基准测试中，学习框架下每种抽样策略对应的鲁棒解。图中展示了 10 次（网格世界）或 5 次（无人机）重复实验的平均值，阴影区域表示最小值/最大值。

模型。 我们考虑以下场景中应在哪里进行采样的问题：1）包含 $|S| = 800$ 种和 $|V| = 100$ 种地形类型的湿滑网格世界；2）来自[22] 的无人机基准，其中包含 $|S| = 4\,179$ 个参数和 $|V| = 1\,053$ 个参数。与第 2 节中的启发性示例一样，我们以 prMC 的形式学习未知马尔可夫链的模型，其中参数是每个参数对应的样本量。我们假设可以使用一个模型，对每个参数任意采样（即网格世界中的打滑概率）。每个参数 $i \in \{1, \ldots, |V|\}$的初始样本量为 $N_i = 100$，据此我们利用霍夫丁不等式推导出一个 $\beta = 0.9$（90%）置信区间。参数 i的区间为 $[\hat{}p_i - \epsilon_i,\ \hat{}p_i + \epsilon_i]$，其中 $\hat{}p_i$ 为样本均值，$\epsilon_i = \sqrt{\log 2 - \log(1-\beta)\, 2N}$（详情见例如 [13]）。

学习方案。 我们迭代地选择要为哪个参数 $v_i \in V$ 获取 25 个（对于网格世界）或 250 个（对于无人机）额外样本。我们比较了四种选择要为其 v_i 进行抽样的参数的策略：1）选择导数最高的参数，即针对 $k = 1$求解问题 3；2）选择区间宽度 ϵ_i最大的参数；3）均匀抽样；4）根据期望访问次数乘以区间宽度进行抽样（详情请参见附录B.1）。每一步之后，我们都会利用额外样本更新 prMC 解的鲁棒上界。

结果。 图 7 展示了每种抽样策略对应的解的上界，以及采用真实参数值的马尔可夫链（MC）的解。对于这两个基准测试，我们的导数引导的采样策略比其他策略更快地收敛到真实解。值得注意的是，我们的导数引导策略同时考虑了每个参数的不确定性和重要性，从而降低了逼近真实解所需的样本复杂度。

7 相关工作

我们将从三个方面讨论相关研究：参数化马尔可夫链（pMCs）、其扩展形式——参数化区间马尔可夫链（piMCs），以及通用敏感性分析方法。

参数化马尔可夫链。pMC [23,44]传统上主要通过计算解函数 [12,24,27,28,31]来研究。近期许多文献关注综合（寻找满足某个规约的参数赋值）或验证（证明所有赋值都满足某个规约）。近期概述请参见 [37] 。与本文尤其相关的是 [54]„ 它用于检查导数是否为正（对所有参数赋值均成立）；以及 [33], ，它通过梯度下降求解参数综合。我们注意到，所有这些问题都是（余）ETR 完全的 [40] ，且解函数的规模随参数数量呈指数级增长 [6], ，而我们考虑的是多项式时间算法。此外，针对不可控参数的实用验证程序（如本文所做）所能处理的参数数量不足 10 个。文献 [50] 利用参数化验证，通过检测哪些参数取值满足规约来指导模型细化。相比之下，我们考虑略微更保守的 rMC，并旨在逐步优化目标。解函数也提供了一种计算和细化置信区间的方法 [16]；然而，解函数的规模妨碍了可扩展性。

参数化区间马尔可夫链（参数化区间马尔可夫链）。据我们所知，尽管参数化鲁棒马尔可夫链尚未得到研究，但其限制稍多的版本——参数化区间马尔可夫链——已有相关研究。具体而言，参数化区间马尔可夫链具有带参数边界的区间值转移。关于参数化区间马尔可夫链的研究分为两类。第一，一致性 [26,49]：是否存在一种参数实例化，使得诱导区间马尔可夫链（可达部分）包含有效概率分布？第二，针对最多含有 12 个参数的参数化区间马尔可夫链，研究定量和定性可达性的参数综合 [9]。

扰动分析。扰动分析研究参数实例化受到任意扰动向量 X 影响时解的变化，该向量的范数上界为 δ，即 $||X|| \leq \delta$ （或者反过来说，确定哪种 δ 能确保解的扰动低于给定的最大值）。类似地， [20] 利用 pMC 的两种实例化（称为增广区间马尔可夫链）之间的距离，来界定可达概率的变化。针对稳态分布 [1]也存在类似分析。这些问题与 pMC 中的验证问题密切相关；若存在多参数依赖关系，它们同样难以处理。为提高可处理性，后续研究 [55] 基于一阶或二阶泰勒展开推导了渐近界。其他扰动分析方法则针对系统的单条路径进行分析 [17,18,29]。关于（无参数）不精确马尔可夫链（一种鲁棒马尔可夫链的变体）中的敏感性分析，文献 [21]进行了全面研究。

学习中的探索。与第 6 节中的 Q3 类似，在许多学习场景中，确定采样位置都很重要。可能近似正确（PAC）统计模型检测 [2,3]和基于模型的强化学习 [46] 等方法通常采用乐观探索策略 [47]。相比之下，我们根据解函数相对于参数化模型的敏感性分析来指导探索。

8 结语

我们提出了高效方法，用于计算参数化马尔可夫链和参数化鲁棒马尔可夫链的解函数偏导数。对于这两种模型，我们展示了如何显式计算所有参数对应的这些导数，以及如何仅计算 k 最高阶导数。实验表明，我们能够为拥有超过一百万个状态和数千个参数的模型计算导数。具体而言，计算 k 最高阶导数相较于显式计算所有导数可显著加速；对于可验证的参数化鲁棒马尔可夫链，这种方法也是可行的。未来，我们希望支持模型中的非确定性，并将我们的方法应用于（在线）学习框架，尤其是降低不确定性在计算上代价高昂的场景 [41,48]。

参考文献

1. Abbas, K., Berkhout, J., Heidergott, B.：马尔可夫链扰动分析述评。arXiv预印本 arXiv:1609.04138（2016） 2. Agarwal, C., Guha, S., Kret´ınsk´y, J., Muruganandham, P.：离散时间和连续时间马尔可夫决策过程平均收益的PAC统计模型检测。载于：CAV (2)。计算机科学讲义，第13372卷，第3–25页。施普林格（2022） 3. Ashok, P., Kret´ınsk´y, J., Weininger, M.：马尔可夫决策过程和随机博弈的PAC统计模型检测。载于：CAV (1)。计算机科学讲义，第11561卷，第497–519页。施普林格（2019） 4. Badings, T., Sim˜ao, T.D., Suilen, M., Jansen, N.：不确定条件下的决策：超越概率。《国际软件工具与技术转移期刊》（2023） 5. Badings, T.S., Cubuktepe, M., Jansen, N., Junges, S., Katoen, J., Topcu, U.：不确定参数化马尔可夫决策过程的基于场景的验证。《国际软件工具与技术转移期刊》**24**(5), 803–819（2022） 6. Baier, C., Hensel, C., Hutschenreiter, L., Junges, S., Katoen, J., Klein, J.：参数化马尔可夫链：概率计算树逻辑复杂度与无分数高斯消元。Inf. Comput. **272**, 104504（2020） 7. Baier, C., Katoen, J.：《模型检测原理》。麻省理工学院出版社（2008） 8. Barratt, S.：凸优化问题的解的可微性。arXiv预印本 arXiv:1804.05098（2018） 9. Bart, A., Delahaye, B., Fournier, P., Lime, D., Monfroy, ´E., Truchet, C.：使用约束求解参数化区间马尔可夫链的可达性。Theor. Comput. Sci. **747**, 48–74（2018） 10. Bazaraa, M.S., Jarvis, J.J., Sherali, H.D.：《线性规划与网络流》。John Wiley & Sons（2011） 11. Ben-Tal, A., Ghaoui, L.E., Nemirovski, A.：《鲁棒优化》，普林斯顿应用数学丛书，第28卷。普林斯顿大学出版社（2009） 12. Bortolussi, L., Milios, D., Sanguinetti, G.：不确定连续时间马尔可夫链的平滑模型检测。Inf. Comput. **247**, 235–253（2016） 13. Boucheron, S., Lugosi, G., Massart, P.：《集中不等式：独立性的非渐近理论》。牛津大学出版社（2013） 14. Boyd, S.P., Vandenberghe, L.：《凸优化》。剑桥大学出版社（2014）

15. Buckman, J., Hafner, D., Tucker, G., Brevdo, E., Lee, H.：基于随机集成价值扩展的高效采样强化学习。载于：NeurIPS，第 8234–8244 页（2018） 16. Calinescu, R., Ghezzi, C., Johnson, K., Pezz`e, M., Rafiq, Y., Tamburrelli, G.：利用置信区间进行形式化验证，以确立软件系统的服务质量属性。IEEE可靠性汇刊。**65**(1), 107–125（2016） 17. Cao, X., Chen, H.：马尔可夫过程的扰动实现、势函数与敏感性分析。IEEE自动控制汇刊。**42**(10), 1382–1393（1997） 18. Cao, X., Wan, Y.：通过势函数和扰动实现对马尔可夫系统进行敏感性分析的算法。IEEE控制系统技术汇刊。**6**(4), 482–494（1998） 19. Chen, T., Feng, Y., Rosenblum, D.S., Su, G.：离散时间马尔可夫链验证中的扰动分析。载于：CONCUR。《计算机科学讲义》，第 8704 卷，第 218–233 页。施普林格（2014） 20. Chonev, V.：增广区间马尔可夫链中的可达性。载于：RP。《计算机科学讲义》，第 11674 卷，第 79–92 页。施普林格（ 2019 ） 21. Cooman, G.D., Hermans, F., Quaeghebeur, E.：离散时间有限马尔可夫链的敏感性分析。载于：UAI，第 129–136 页。AUAI出版社（2008） 22. Cubuktepe, M., Jansen, N., Junges, S., Katoen, J., Topcu, U.：马尔可夫决策过程参数综合的凸优化。IEEE自动控制汇刊。**67**(12), 6333–6348（2022） 23. Daws, C.：离散时间马尔可夫链的符号化与参数化模型检验。载于：ICTAC。《计算机科学讲义》，第 3407 卷，第 280–294 页。施普林格（2004） 24. Dehnert, C., Junges, S., Jansen, N., Corzilius, F., Volk, M., Bruintjes, H., Katoen, J., ´Abrah´am, E.：Prophesy：一种概率参数综合工具。载于：CAV (1)。《计算机科学讲义》，第 9206 卷，第 214–231 页。施普林格（ 2015 ） 25. Delahaye, B.：参数化区间马尔可夫链的一致性。载于：SynCoP。OASIcs，第 44 卷，第 17–32 页。达格施图尔城堡——莱布尼茨信息学中心（2015） 26. Delahaye, B., Lime, D., Petrucci, L.：参数化区间马尔可夫链的参数综合。载于：VMCAI。《计算机科学讲义》，第 9583 卷，第 372–390 页。施普林格（2016） 27. Fang, X., Calinescu, R., Gerasimou, S., Alhwikem, F.：通过模型分片实现快速参数化模型检验。载于：ICSE，第 835–846 页。IEEE（2021） 28. Filieri, A., Tamburrelli, G., Ghezzi, C.：通过运行时定量验证与敏感性分析支持自适应。IEEE软件工程汇刊。**42**(1), 75–99（2016） 29. Fu, M.C., Hu, J.：马尔可夫链的平滑扰动分析导数估计。运筹学快报。**15**(5), 241–251（1994） 30. Gurobi Optimization, LLC：Gurobi优化器参考手册（2023），https: //www.gurobi.com 31. Hahn, E.M., Hermanns, H., Zhang, L.：参数化马尔可夫模型的概率可达性。《软件工具与技术转移国际期刊》。**13**(1), 3–19（2011） 32. Hartmanns, A., Klauck, M., Parker, D., Quatmann, T., Ruijters, E.：定量验证基准集。载于：TACAS。LNCS，第 11427 卷，第 344–350 页。施普林格（2019） 33. Heck, L., Spel, J., Junges, S., Moerman, J., Katoen, J.：部分可观测条件下随机化控制器的梯度下降。载于：VMCAI。《计算机科学讲义》，第 13182 卷，第 127–150 页。施普林格（2022）

34. Hensel, C., Junges, S., Katoen, J.P., Quatmann, T., Volk, M.：概率模型检测器 Storm。Softw Tools Technol Transfer（2021） 35. Hoffman, A.J., Kruskal, J.B.：凸多面体的整边界点。载于：整数规划五十年，第49–76页。施普林格（2010） 36. Hoffman, A.J., Kruskal, J.B.：凸多面体的整边界点。《整数规划五十年（1958–2008）》，第49页（2010） 37. Jansen, N., Junges, S., Katoen, J.：马尔可夫模型中的参数综合：简明综述。载于：系统设计原理。计算机科学讲义，第13660卷，第407–437页。施普林格（2022） 38. Jonsson, B., Larsen, K.G.：概率过程的规约与细化。载于：LICS。第266–277页。IEEE计算机学会（1991） 39. Junges, S., ´Abrah´am, E., Hensel, C., Jansen, N., Katoen, J., Quatmann, T., Volk, M.：马尔可夫模型的参数综合。CoRR **abs/1903.07993**（2019） 40. Junges, S., Katoen, J., P´erez, G.A., Winkler, T.：参数化马尔可夫决策过程中的可达性复杂度。计算机与系统科学杂志 **119**，183–210（2021） 41. Junges, S., Spaan, M.T.J.：分层概率模型的抽象 - 细化。载于：CAV （ 1 ）。计算机科学讲义，第13371卷，第102–123页。施普林格（2022） 42. Kakade, S.M.：论强化学习的样本复杂度。博士论文，伦敦大学、伦敦大学学院（英国）（2003） 43. Kwiatkowska, M.Z., Norman, G., Parker, D.：PRISM 4.0：概率实时系统的验证。载于：CAV。计算机科学讲义，第6806卷，第585–591页。施普林格（2011） 44. Lanotte, R., Maggiolo-Schettini, A., Troina, A.：用于系统设计与分析的参数化概率转移系统。Formal Aspects Comput. **19**（1），93–109（2007） 45. Matousek, J., G¨artner, B.：整数规划与线性规划松弛，第29–40页。施普林格柏林海德堡（2007） 46. Moerland, T.M., Broekens, J., Jonker, C.M.：基于模型的强化学习：综述。CoRR **abs/2006.16712**（2020） 47. Munos, R.：从多臂老虎机到蒙特卡洛树搜索：将乐观原则应用于优化与规划。Found. Trends Mach. Learn. **7**（1），1–129（2014） 48. Neary, C., Verginis, C.K., Cubuktepe, M., Topcu, U.：可验证且可组合的强化学习系统。载于：ICAPS。第615–623页。AAAI出版社（2022） 49. Petrucci, L., van de Pol, J.：参数化区间马尔可夫链的参数综合算法。载于：FORTE。计算机科学讲义，第10854卷，第121–140页。施普林格（2018） 50. Polgreen, E., Wijesuriya, V.B., Haesaert, S., Abate, A.：用于参数化马尔可夫决策过程数据高效验证的自动化实验设计。载于：QEST。计算机科学讲义，第10503卷，第259–274页。施普林格（2017） 51. Puggelli, A., Li, W., Sangiovanni-Vincentelli, A.L., Seshia, S.A.：具有凸不确定性的MDP之PCTL性质的多项式时间验证。载于：CAV。计算机科学讲义，第8044卷，第527–542页。施普林格（2013） 52. Puterman, M.L.：马尔可夫决策过程：离散随机动态规划。威利概率与统计丛书，威利（1994） 53. Sen, K., Viswanathan, M., Agha, G.：存在不确定性时的马尔可夫链模型检测。载于：TACAS。计算机科学讲义，第3920卷，第394–410页。施普林格（2006） 54. Spel, J., Junges, S., Katoen, J.：寻找可证明最优的马尔可夫链。载于：TACAS（1）。计算机科学讲义，第12651卷，第173–190页。施普林格（2021）

55. Su, G., Feng, Y., Chen, T., Rosenblum, D.S.：基于经验确定的概率参数进行概率模型检测的渐近扰动界。IEEE Trans. Software Eng. **42**（7），623–639（2016） 56. Suilen, M., Sim˜ao, T.D., Parker, D., Jansen, N.：马尔可夫决策过程的鲁棒随时学习。载于：NeurIPS。第35卷，第28790–28802页。Curran Associates, Inc.（2022） 57. Wiesemann, W., Kuhn, D., Sim, M.：分布鲁棒凸优化。Oper. Res. **62**（6），1358–1376（2014） 58. Wolff, E.M., Topcu, U., Murray, R.M.：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：控制与决策会议。第3372–3379页。IEEE（2012） 59. Wolsey, L.A.：整数规划。John Wiley & Sons（2020）

A 证明与数学细节

A.1 定理2的证明

该证明可通过对方程（16）中的方程组应用乘积法则得出。右侧方程（16a）的导数（即对于所有终止状态 $s \in S_T$）显然为零。方程（16b）的导数为

$$\begin{aligned}\frac{\partial x_s}{\partial u(v_i)} &= -\frac{\partial \left(b_s[\mathbf{u}]^\top \mathbf{D}(E_s)\alpha_s\right)}{\partial u(v_i)} - \frac{\partial \beta_s}{\partial u(v_i)} \\ \frac{\partial x_s}{\partial u(v_i)} &= -(\alpha_s^\star)^\top \mathbf{D}(E_s)\frac{\partial b_s[\mathbf{u}]}{\partial u(v_i)} - b_s[\mathbf{u}]^\top \mathbf{D}(E_s)\frac{\partial \alpha}{\partial u(v_i)} - \frac{\partial \beta_s}{\partial u(v_i)},\end{aligned} \tag{19}$$

重新整理后即可得到方程（17b）。类似地，方程（16c）的导数为

$$(\alpha_s^\star)^\top \mathbf{D}(E_s)\frac{\partial A_s[\mathbf{u}]}{\partial u(v_i)} + A_s[\mathbf{u}]^\top \mathbf{D}(E_s)\frac{\partial \alpha_s}{\partial u(v_i)} + \frac{\partial x_{\mathsf{post}(s)}}{\partial u(v_i)} + \frac{\partial \beta_s}{\partial u(v_i)}\mathbb{1}, \tag{20}$$

重新整理后即可得到方程（17c），由此完成证明。

A.2 计算参数化鲁棒马尔可夫链的 k 个最高阶导数

与式 (9) 类似，我们可以根据混合整数线性规划的解计算 prMC 的 $k \leq \ell = |V|$ 高阶导数。为简洁起见，我们定义以下记号$x'_s = \frac{\partial x_s}{\partial u}(v_i) \in \mathbb{R}$，$\alpha'_s = \frac{\partial \alpha_s}{\partial u}(v_i) \in \mathbb{R}^{m_s}$ 和 $\beta'_s = \frac{\partial \beta_s}{\partial u}(v_i) \in \mathbb{R}$。使用这些记号，

我们得到以下混合整数线性规划：

$$\underset{x',\alpha',\beta',z\in\{0,1\}^\ell}{\text{maximize}} \quad s_I^\top x' \tag{21a}$$

$$\text{满足} \quad x'_s = 0, \qquad \forall s \in S_T \tag{21b}$$

$$x'_s + b_s[\mathbf{u}]^\top \mathbf{D}(E_s)\alpha'_s + \beta'_s \qquad = -(\alpha^\star_s)^\top \mathbf{D}(E_s) \sum_{i=1}^{\ell} z_i \frac{\partial b_s[\mathbf{u}]}{\partial u(v_i)}, \quad \forall s \in S \setminus S_T \tag{21c}$$

$$A_s[\mathbf{u}]^\top \mathbf{D}(E_s)\alpha'_s + x'_{\mathsf{post}(s)} + \beta'_s \mathbb{1} = -(\alpha^\star_s)^\top \mathbf{D}(E_s) \sum_{i=1}^{\ell} z_i \frac{\partial A_s[\mathbf{u}]}{\partial u(v_i)}, \quad \forall s \in S \setminus S_T. \tag{21d}$$

$$z_1 + \cdots + z_\ell = k. \tag{21e}$$

请注意，式 (21c) 和 (21d) 中的约束与式 (17) 中的方程组之间的差异在于对 $i = 1, \ldots, \ell$ 求和。我们推导出与式 (10) 相同的线性规划松弛，即将二元变量 $z \in \{0,1\}^\ell$ 松弛为连续变量 $z \in [0,1]^\ell$。由于式 (21) 与式 (9) 具有完全相同的特性，定理 1 同样适用于 prMCs 的情形。换言之，线性规划松弛是精确的，我们可以使用所得解，利用命题 1 找出具有最大导数的参数集合 $V^\star$。

B 详细基准测试结果

我们对第 6 节中所进行数值实验的设置和结果作更详细的概述。网格世界基准测试统计数据的完整概述见表 2。此外，所有其他基准测试的完整表格见表 3。

B.1 学习框架的详细信息

如第 6 节所述，我们使用这些方法来确定在两个基准测试中的采样位置。首先，我们考虑一个包含 $|S| = 800$ 个状态和 $|V| = 100$ 种地形类型的湿滑网格世界（为使重要性偏向特定地形类型，我们将 10 种地形类型分布在 50% 的状态中，其余地形类型分布在另外 50% 的状态中）。其次，我们考虑来自 [22] 的无人机运动规划问题，其中包含 $|S| = 4\,179$ 个状态和 $|V| = 1\,053$ 个参数。对于这两个基准测试，我们比较四种采样策略：

1. 针对最高阶导数进行贪心采样，即对 $k = 1$ 求解问题 3；2. 针对最大区间宽度进行贪心采样 ϵ_i（由于我们在每个区间上使用相同的置信水平 β，请注意，该策略等同于选择样本量最低的参数）；3. 对所有参数进行均匀采样；

4. 根据每个参数的重要性因子进行加权采样。具体而言，受 [41]，启发，我们将每个参数$i = \{1, \ldots, |V|\}$的重要性定义为包含该参数的状态的预期访问次数 i 乘以 ϵ_i 出边概率区间的宽度。

如第 6 节所述，对于每个参数 $v_i \in V$，我们使用霍夫丁不等式计算一个概率区间，从而得到以样本均值 $\hat{}p_i$ 为中心的对称概率区间。为了应用我们的方法，接下来需要计算每个参数 $v_i \in V$ 的区间界关于样本量 N_i 的导数。计算 $\epsilon_i = \sqrt{\text{对数 } 2 - \text{对数}\,(1-\beta)\,2N}$ 关于 N_i 的导数很简单，但样本均值 $\hat{}p_i$ 的变化取决于额外样本的结果。为避免这一问题，我们假设样本均值 $\hat{}p_i$ 保持不变；对于较大的样本量，这一假设尤其合理。尽管采用了这一启发式方法，图 7 显示，我们的导数引导的采样策略比其他策略更快地收敛到真实解。

表 2：所有网格世界基准测试的模型规模、运行时间和导数。

模型统计信息				验证		问题 1	问题 3		导数	
Type	$\|S\|$	$\|V\|$	转移数	解$_{(R)}[\mathbf{u}]$	时间 [s]	所有导数 [s]	$k=1$ [s]	$k=10[s]$	最高误差百分比	
pMC	200	10	595	1.87	0.05	0.11	0.15	0.16	5.53e-01	0.0
pMC	800	100	2395	7.91	0.24	0.63	0.48	0.47	1.92e+00	0.0
pMC	5000	50	14995	5.07	1.39	3.32	2.64	2.69	1.54e+00	0.0
pMC	5000	100	14995	5.05	1.36	4.17	2.63	2.66	1.28e+00	0.0
pMC	5000	921	14995	4.93	1.87	19.92	4.52	2.87	1.20e+00	0.0
pMC	20000	100	59995	1.99	14.22	35.51	26.88	10.72	4.92e-01	0.0
pMC	20000	1000	59995	2.04	5.65	97.23	11.68	12.41	5.18e-01	0.0
pMC	80000	100	239995	8.01	25.54	98.47	45.18	46.87	1.95e+00	0.0
pMC	80000	1000	239995	8.01	25.64	612.97	48.92	58.20	2.08e+00	0.0
pMC	80000	9831	239995	7.93	25.52	5,650.25[a]	347.76	1,343.59	2.10e+00	0.0
pMC	320000	100	959995	3.22	134.39	659.06	223.71	231.73	8.69e-01	0.0
pMC	320000	1000	959995	3.20	133.25	4,712.32[a]	244.07	295.71	8.02e-01	0.0
pMC	320000	10000	959995	3.28	133.66	45,655.90[a]	447.62	831.01	9.08e-01	0.0
pMC	1280000	100	3839995	12.90	902.52	4,747.43[a]	1,396.51	1,507.77	3.32e+00	0.0
pMC	1280000	1000	3839995	12.79	902.67	37,078.12[a]	1,550.45	1,617.27	3.18e+00	0.0
pMC	1280000	10000	3839995	超时[b]	—	—	—	—	—	—
prMC	200	10	595	3.36	0.93	0.03	0.02	0.03	-3.02e-04	-0.0
prMC	800	100	2395	26.44	4.13	0.49	0.09	0.11	-1.57e-03	-0.0
prMC	5000	50	14995	141.55	23.25	1.67	0.59	0.66	-1.88e-02	-0.0
prMC	5000	100	14995	136.07	23.46	3.55	0.60	1.58	-1.26e-02	-0.0
prMC	5000	921	14995	138.74	29.82	25.23	0.85	1.09	-4.44e-03	-0.0
prMC	20000	100	59995	2,789.77	1,276.43	15.68	2.40	2.70	-4.96e-01	-0.1
prMC	20000	1000	59995	2,258.41	339.96	159.70	3.53	4.09	-9.51e-02	-0.0
prMC	80000	100	239995	超时[b]	—	—	—	—	—	—
prMC	80000	1000	239995	超时[b]	—	—	—	—	—	—
prMC	80000	10000	239995	超时[b]	—	—	—	—	—	—

[a] 根据参数为 10 时的运行时间推算至所有 $|V|$ 参数。[b] 验证 p(r)MC 时发生了超时（1 小时），而计算导数时没有。

表 3：除网格世界外所有基准测试的模型规模和运行时间。对于参数数量至少为 10 的模型，我们仅针对 $k = 10$ 个参数求解问题 3，即 $|V| \geq 10$。

模型统计信息					验证		所有导数。	k 最高阶导数	
实例	Type	$\|S\|$	$\|V\|$	转移数	求解$_{(R)}$[u]	验证 [s]	显式 [s]	$k=1$ [s]	$k=10$ [s]
BRP (16,2)	pMC	613	2	803	0.10	0.11	0.17	0.20	—
BRP (32,3)	pMC	1638	2	2179	0.04	0.29	0.46	0.51	—
BRP (64,4)	pMC	4103	2	5507	0.02	0.74	1.13	1.18	—
BRP (512,5)	pMC	39432	2	53251	0.02	7.77	11.66	12.12	—
BRP (1024,6)	pMC	92169	2	124931	0.01	20.66	29.69	31.76	—
人群 (3,5)	pMC	1367	2	2027	0.87	0.25	0.36	0.40	—
人群 (6,5)	pMC	14	1	16	0.25	0.00	0.01	0.03	—
人群 (10,5)	pMC	104512	2	246082	0.83	27.86	36.02	36.28	—
与非门 (2,4)	pMC	326	2	435	0.71	0.06	0.10	0.14	—
与非门 (5,10)	pMC	8112	2	11577	0.57	1.60	2.35	2.42	—
与非门 (10,15)	pMC	104412	2	156247	0.51	26.37	36.55	37.02	—
病毒	pMC	761	14	5009	88.89	0.18	0.57	0.60	0.60
WLAN0	pMC	2711	15	4877	65,886.52	0.47	1.02	1.06	1.07
载波侦听多路访问 (2,4)	pMC	7958	26	10594	98.58	1.39	3.05	3.07	3.09
抛硬币 (4)	pMC	22656	4	74957	0.92	4.25	7.71	7.80	—
Maze	pMC	1303	590	2658	79.81	0.26	1.76	0.54	0.54
无人机 (mem1)	pMC	4179	1053	9414	0.11	0.88	12.34	1.44	1.52
无人机 (mem5)	pMC	32403	12286	70099	0.11	8.58	16,592.78[a]	12.73	14.02
卫星 (36,5)	pMC	31325	2555	156924	0.00	10.68	402.14	65.83	66.30
卫星 (36,65)	pMC	217561	10042	615433	0.00	96.48	203,804.52[a]	179.62	199.52
BRP (16,2)	prMC	613	2	803	0.24	1.17	0.02	0.03	—
BRP (16,2)	prMC	613	190	803	0.24	1.16	0.19	0.03	0.03
BRP (32,3)	prMC	1638	2	2179	0.13	3.09	0.04	0.05	—
BRP (32,3)	prMC	1638	541	2179	0.13	3.10	1.00	0.05	0.06
BRP (64,4)	prMC	4103	2	5507	0.07	7.97	0.11	0.12	—
BRP (64,4)	prMC	4103	1404	5507	0.07	8.00	5.75	0.13	0.17
BRP (512,5)	prMC	39432	2	53251	0.14	76.16	1.11	1.69	—
BRP (512,5)	prMC	39432	13819	53251	0.14	76.34	746.47	1.71	1.59
BRP (1024,6)	prMC	92169	2	124931	0.07	181.33	2.66	5.08	—
BRP (1024,6)	prMC	92169	32762	124931	0.08	180.37	4,625.87[a]	5.27	3.33
人群 (3,5)	prMC	1367	2	2027	0.91	2.28	0.03	0.04	—
人群 (3,5)	prMC	1367	495	2027	0.92	2.84	1.68	0.05	0.42
人群 (6,5)	prMC	14	2	16	0.29	0.02	0.00	0.00	—
人群 (10,5)	prMC	104512	2	246082	0.88	183.58	4.12	2.95	—
人群 (10,5)	prMC	104512	38610	246082	0.88	389.26	36,477.85[a]	5.19	—
与非门 (2,4)	prMC	326	2	435	0.94	0.60	0.01	0.01	—
与非门 (2,4)	prMC	326	109	435	0.95	0.72	0.09	0.02	0.07
与非门 (5,10)	prMC	8112	2	11577	0.99	15.91	0.38	0.22	—
与非门 (5,10)	prMC	8112	3465	11577	1.00	18.27	158.02	0.41	2.82
与非门 (10,15)	prMC	104412	2	156247	1.00	341.76	37.58	3.62	—
与非门 (10,15)	prMC	104412	51835	156247	1.00	764.31	44,289.51[a]	5.79	—
病毒	prMC	761	14	5009	151.30	2.26	0.06	0.04	0.31
病毒	prMC	761	558	5009	151.53	2.29	0.85	0.04	0.05
WLAN0	prMC	2711	45	4877	66.7×10^6	3.13	0.18	0.04	0.55
载波侦听多路访问 (2,4)	prMC	7958	1	10594	100.11	7.56	0.04	0.07	—
载波侦听多路访问 (2,4)	prMC	7958	202	10594	113.48	9.80	1.68	0.18	0.16
抛硬币 (4)	prMC	22656	4	74957	0.97	39.66	1.00	0.69	—
抛硬币 (4)	prMC	22656	6214	74957	0.97	38.56	876.16	0.71	6.49
Maze	prMC	1303	590	2658	177.80	5.11	2.36	0.13	0.14
Maze	prMC	1303	1300	2658	207.22	5.14	10.78	0.13	0.93
无人机 (mem1)	prMC	4179	1053	9414	0.20	9.54	19.30	0.23	0.34
无人机 (mem1)	prMC	4179	4176	9414	0.39	17.64	1,018.50	0.82	7.63
无人机 (mem5)	prMC	32403	12286	70099	0.20	844.53	10,800.88	2.95	4.59
无人机（mem5）	prMC	32403	32401	70099	0.39	1,808.07	173,662.85[a]	4.51	—
卫星（36,5）	prMC	31325	50	156924	0.00	43.04	83.81	0.95	60.26
卫星 (36,5)	prMC	31325	30040	156924	0.12	609.86	186,404.70[a]	4.64	232.28
卫星 (36,65)	prMC	217561	10042	615433	超时[b]	—	—	—	—

[a] 根据 10 个参数的运行时间外推至全部 $|V|$ 个参数。[b] 验证 p(r)MC 时发生了超时（1 小时），计算导数时未发生超时。