arXiv:2611.33081v1 [cs.GT] 24 Nov 2026

突破多面体博弈中线性交换遗憾的平方根障碍

Vasilis Pollatos*[1,2] 和 Alberto Marchesi†[3]

[1]雅典国立卡波迪斯特里亚大学数学系[2]雅典娜研究中心 Archimedes，希腊[3]米兰理工大学

摘要

我们研究自博弈中针对每位玩家策略多面体上所有仿射自映射的线性交换遗憾。针对这一类映射的高效对抗性界按 $\sqrt{T}$ 缩放。对于由线性不等式给出的紧致多面体，我们的确定性非耦合动态可实现 $\widetilde{\mathcal{O}}(\mathrm{poly}(D)T^{1/3})$ 累积遗憾，其中 D 是博弈描述的二进制长度。该界与顶点数量无关。在每位玩家至多有 n 名玩家的完美回忆扩展式博弈中，每位玩家至多有 M 个终端序列，我们针对所有线性偏离获得 $\mathcal{O}(n^2M^4\log T)$遗憾界，将 Anagnostides 等人提出的对数刻度触发遗憾界Anagnostides 等人(2023年)推广至此类偏离。这一类严格包含触发偏离（ Farina 和 Pipis，2023 年），且恰好由无定时通信（UTC）可实现的偏离构成；在这种通信中，玩家沿着一条独立的可行路径交替行动并向调解者发起查询（Zhang 等人，2024年）。证明结合了在紧致矩阵提升上进行的乐观对数障碍学习，以及 Gordon 不动点的稳定性，并在扩展式博弈中利用树丛几何。

1 引言

在博弈中，无遗憾学习会随着比较器类别的表达能力增强而变得更具挑战性。外部遗憾将学习者的表现与固定动作的表现进行比较，并导向粗相关均衡；而交换遗憾允许根据动作进行替换，并导向相关均衡（Foster 和 Vohra，1997；Hart 和 Mas-Colell，2000；Blum 和 Mansour，2007）。 Φ-遗憾框架（Gordon 等人，2008年）将这些概念推广到任意变换类别。在凸策略集上，线性偏离是其仿射自映射。它们定义了线性交换遗憾（也

*vaspoll@math.uoa.gr
†alberto.marchesi@polimi.it

称为线性 Φ-遗憾）和线性相关均衡（Farina 和 Pipis，2023年；Daskalakis 等人，2025 年）。

对于策略空间具有紧凑多面体表示、但顶点数量呈指数级增长的博弈，计算挑战尤为突出。在显式表示的有限博弈中，自博弈中的完全交换遗憾界已从对数刻度（Anagnostides 等人，2022b）改进至次对数（Tsuchiya，2026年），最终达到与时域无关的速率（Mai，2026年）。然而，这些界仍然对显式表示的动作数量呈多项式依赖。完美回忆扩展式博弈（EFG）是这一差异至关重要的典型场景：序列式策略空间可用多项式规模的表示来描述，而标准式则包含指数数量的应变计划（Koller etal.，1996）。因此，快速的完全交换算法可能需要指数于博弈树规模的时间。在 T 轮行动中，针对线性偏离的高效对抗式遗憾界，在一般凸集（Daskalakis 等人，2025 年；Anagnostides 等人，2026b）以及扩展式博弈的序列式多面体（Farina 和 Pipis，2023年）上的增长率均为 $\sqrt{T}$。

Farina 和 Pipis (2023年) 指出了为何基于由效用变化界定的遗憾（RVU）的乐观分析（它是 $o(\sqrt{T})$ 在自博弈中累积触发遗憾界的基础）不能直接推广到线性偏离。Anagnostides 等人(2023年) 的对数遗憾构造利用了这样一个事实：触发固定点具有正系数的有理表示，这些表示继承自局部马尔可夫链的平稳分布。相比之下，一般线性偏离的 Gordon 不动点尚无已知的类似局部刻画。Zhang 等人 (2024 年) 提出的 UTC 表述为优化所有线性偏离提供了一个多项式规模的定义域。然而，要利用乐观遗憾界中的负移动项，就必须控制偏离映射的变化如何影响学习者所采取的相应不动点。这引出了以下问题：

问题一： 在自博弈中，非耦合动态能否在具有简洁 H 表示的多面体上实现 $o(\sqrt{T})$ 线性 Φ-遗憾，且对输入规模呈多项式依赖？

在完美回忆扩展式博弈中，Anagnostides 等人(2023年证明，自博弈中的触发式遗憾可达到对数刻度。触发式会在某个信息集收到特定推荐时，切换到固定的后续行动。因此，经验行动构成一个$\mathcal{O}(\log T/T)$-近似扩展式相关均衡（EFCE）。UTC 偏离还允许行动前沿着一条可行的假设路径向调解者查询。UTC 遗憾等于线性 Φ-遗憾（Zhang 等人，2024年，定理 4.1）。

在von Stengel 和 Forges (2008年）的信号扩展式相关均衡中，Farina 和 Pipis (2023年，例 E.1）展示了一种有利可图的线性偏离。为了将这个均衡示例与无遗憾学习联系起来，考虑一种重复博弈过程：行动依次循环经过四种等概率的联合计划。每个完整循环结束后，博弈经验分布都与均衡分布一致。在底层信号博弈中，学生私下观察自己的类型（好类型或坏类型），并向教授发送两个信号之一；随后教授决定是否接受该学生。教授只接受推荐给好类型的那个信号，因此好类型只要遵从推荐就能确保被接受。触发式偏离只能利用针对学生实际类型的推荐。每种针对坏类型的推荐都会出现在两个等概率计划中。其中一个计划里，教授接受所推荐的信号；另一个计划里，

教授接受相反的信号。因此，遵从和切换带来的被接受概率均为 $1/2$，所以任何类型都无法通过触发式偏离获益。在 UTC 下，即使是坏类型的学生也可以询问调解者会向好类型推荐哪个信号。发送该信号即可确保被接受。在完整循环中，学生的触发式遗憾为零，而这种 UTC 偏离会使学生的预期累计收益从 $7T/2$提高到 $5T$，增加 $3T/2$。

受乐观动态所实现的对数刻度触发式遗憾（Anagnostides 等人，2023年）以及 UTC 偏离能够在没有任何触发式偏离可以获利的情况下获利这一能力的启发，我们转向第二个问题。该问题同样由 Farina 和 Pipis （2023年，第 5 节）和Zhang 等人 （2024年，问题 3）提出：

> **Q2：** 高效的、适用于完美回忆扩展式博弈的非耦合动态，能否在自博弈中实现 O（$\log T$）线性 Φ-遗憾，并且对博弈树规模具有多项式依赖？

我们对问题 Q1 和 Q2 的回答都是肯定的。令 D 表示紧凑博弈描述的二进制编码长度。我们用紧致多面体中的非负矩阵表示每个线性偏离。乐观对数障碍更新使学习者的矩阵保持严格为正，并控制其在轮次之间的变化。对于任意由不等式描述的（H-）多面体，弱重置能够稳定 Gordon 不动点并给出 $T^{1/3}$ 速率；对于扩展式博弈，递归树丛几何能够稳定未重置不动点，并针对每个线性偏离实现对数刻度遗憾。

定理 1（主要结果）。 我们考虑归一化的 n人多重仿射博弈，其策略多面体紧致且有理，并且能够以多项式时间获得全信息反馈。对于完美回忆扩展式博弈， M 表示任一玩家的终端序列数量的最大值。算法 *1*具有确定性的非耦合实现，使得对每位玩家，

$$\mathrm{LinReg}_i(T) = \begin{cases} \tilde{\mathcal{O}}(\mathrm{poly}(D)T^{1/3}), & \textit{arbitrary compact } H\textit{-polytopes}, \\ \mathcal{O}(n^2M^4\log(T+1)), & \textit{perfect-recall EFGs}, \end{cases}$$

这些界与顶点数量无关，且每轮使用 $\mathrm{poly}(D, \log(T+1))$次位运算。

技术概述。 我们的算法框架（在算法 1 中正式定义）结合了两种几何结构：乐观跟随正则化领导者（FTRL）在紧致正提升中进行移动，而 Gordon 约化则将每个偏离转换为实际采取的不动点策略。因此，证明必须将缓慢的偏离移动 转化为缓慢的策略移动。乐观 FTRL 的遗憾界每轮减去两个 Bregman 散度，分别对应预测步骤和校正步骤。多重仿射收益的光滑性将累积的预测误差平方和控制在行动路径长度平方的范围内。随后，不动点稳定性将该路径长度界定为同一组散度之和的某个倍数。定理4 将这种抵消关系单独提炼出来，最终只留下依赖于具体几何结构的稳定性界。

对于任意 H-多面体中，精确不动点可能不稳定。弱重置能使

表 1：归一化 n 人游戏在 $T \geq 2$ 轮次中的累积遗憾。此处，$N \geq 2$ 是标准式中任一玩家拥有的纯策略数的最大值。在多面体博弈中，这些策略可以隐式表示，且 N 可能随 d 呈指数增长，其中 d 是玩家策略多面体的最大内在维数。在整张表中，D 表示紧凑博弈描述的二进制长度，poly(·) 表示多项式因子。近端界固定了 Lipschitz 常数、光滑性常数和策略集直径。

参考	设定 / 表示	比较器	累积遗憾
Mai (2026年)	标准式；自博弈	完全交换	关于(poly(n, N))
Dagan 等人(2024年); Peng 和 Rubinstein (2024年)	标准式 / 扩展式博弈纯策略计划；对抗式	完全交换	$\mathcal{O}(T\log\log(NT)/\log T)$
Anagnostides 等人 (2023年)	完美回忆扩展式博弈；自博弈	触发式	$\mathcal{O}$（poly(D) log T）
Cai 等人（2026年）	光滑凸博弈；自博弈	近端	$\mathcal{O}(\text{poly}(n)T^{1/4})$
Farina 和 Pipis(2023年); Zhang 等人(2024年)	完美回忆扩展式博弈；对抗式	线性 Φ / UTC	$\mathcal{O}(\text{poly}(D)T^{1/2})$
Anagnostides 等人 (2026年b)	John 预条件化凸集；对抗式	线性 Φ	$\mathcal{O}(\text{poly}(d)T^{1/2})$
本文	紧凑的 H-多面体；自博弈	线性 Φ	$\widetilde{\mathcal{O}}(\text{poly}(D)T^{1/3})$
本文	完美回忆扩展式博弈；自博弈	线性 Φ / UTC	$\mathcal{O}$(多项式(D) 对数刻度 T)

Gordon 映射变为压缩映射，但直接扰动会使逆重置强度平方。我们转而将重置不动点表示为再生随机复合的平均终点。KL链式法则只计入期望轨迹长度，因此熵障碍函数与对数障碍函数的比较只会产生一个逆重置因子。平衡这种稳定性代价与重置偏差，即可得到 $T^{1/3}$ 速率。

完美回忆扩展式博弈提供了一条更精确的路径。将其限制在终端序列上的实现计划多面体是由积与连接递归构建的树丛，从而诱导出一种和/最大范数，在该范数下，每个可行偏差矩阵都是非扩张的。预解迹符号在这一递归过程中得以保留。将其应用于正则化 Perron 方程后，该符号使伴随矩阵公式中的标量分母随着可行矩阵族的两个系数变化而保持单调；随后，一维 Bernstein 不等式将分母控制转化为不动点的多项式导数界。这完全消除了重置，并为完整的线性/UTC 类带来对数遗憾。

2 进一步的相关工作

正规式中的快速交换遗憾。 正规式博弈中较早的自博弈保证包括 $T^{1/4}$ Chen 和 Peng 的界 (2020) 以及 Anagnos- tides 等人 (2022a的多重对数界。尺度不变动力学无需预先掌握效用尺度，即可达到对数刻度的交换遗憾

（Tsuchiya 等人，2026年）。对于参与者外部遗憾之和始终非负的博弈，也存在与时域无关的保证（Hait 等人，2025 年）。附录 A讨论了这一限制，以及它与近期常数外部遗憾界之间的区别。

大行动空间下的交换遗憾。Dagan 等（2024）的 Swap Regret 2.0 归约以及 Peng 和 Rubinstein（2024）的同期工作，针对对抗性效用给出了累积完全交换遗憾 O(Tlog log(NT) / log T)，改进了对隐式动作数量 N 的依赖，而该数量可能随多面体维度呈指数增长。对于基集大小为 k 的组合多臂老虎机，Kontogiannis 等人（2026，式 (2) 和定理 4.1）将这种多尺度方法改用于聚合奖励反馈，并针对无适应性对手获得 O(Tlog($k$log T)/ log T) 的交换伪遗憾。即使在扩展式博弈中，面对无适应性对手，要实现逆多项式平均完全交换遗憾，也可能需要超多项式数量的轮次（Daskalakis 等人，2024）。对于定义在显式多面体上的多线性博弈，Anagnostides 等人（2026a，定理 4）证明，在 PCP-for-PPAD 猜想成立的前提下，对于任意预先给定的多项式支持集大小上界，以适当的逆多项式精度计算相关均衡是 PPAD-困难的。对于玩家数量固定的有限时域马尔可夫博弈，Anagnostides 等人（2026c）给出了正规式相关均衡的完全多项式时间近似方案。

其他比较器类与紧凑学习。 在若干结构化博弈中，随机加权多数算法和核化乘法权重避免了纯动作枚举（Beaglehole 等人，2023年；Farina 等人，2022年；Kontogiannis 等人，2025 年），但其均衡保证基于外部遗憾。在 Lipschitz 光滑凸博弈中，Cai 等人 (2026年，定理 4）针对凸函数的近端映射，得到了 $T^{1/4}$ 自博弈速率。其仿射比较涵盖了具有对称线性部分的偏离。

3 预备知识

线性代数与散度。 对于凸集 K，aff K 和 ri K 分别表示其仿射包和相对内部。对于方阵 A，tr A 和 adj A 分别表示其迹和伴随矩阵。对于可微凸函数 F，我们将 $D_F(v,\ u) = F(v) - F(u) - \langle \nabla F(u)\ v - u\rangle$记为其 Bregman 散度。对于概率分布 P，Q、$D_{\mathrm{KL}}(P\|Q)$ 和 TV(P，Q)分别表示 KL 散度和全变差。

多面体博弈。多面体博弈由玩家集合 $[n] := \{1, \ldots, n\}$、每位玩家 i 对应的非空紧致策略多面体 Q_i，以及每个 i 对应的标量损失 $L_i \colon \prod_{j=1}^{n} Q_j \to [-1, 1]$ 构成。每个 L_i 对每位玩家的策略分别是仿射函数。这包括 Farina 等人（2022年）和 Kontogiannis 等人（2025 年）提出的 $0/1$-多面体博弈。我们记 $d_i := \dim(\mathrm{aff}\, Q_i)$。固定对手策略组 x_{-i} 时，令 $\ell_i(x_{-i})$ 表示玩家 i 的损失关于其自身策略的线性部分。因此，对于 $u, v \in Q_i$,

$$L_i(u, x_{-i}) - L_i(v, x_{-i}) = \langle \ell_i(x_{-i}), u - v \rangle.$$

这在 $\mathrm{span}(Q_i - Q_i)$ 上定义了 $\ell_i(x_{-i})$。

我们用如下方式衡量策略之间的距离：取一个取值于 $[0,1]$ 的仿射函数，该函数能够赋予二者的损失差的最大值。对于 $u,\ v \in Q_i$，定义

$$\|u - v\|_i := \sup\Big\{|f(u) - f(v)| :\ f : Q_i \to [0,1] \text{ affine}\Big\}.$$

因此，对于可行策略 u、v，有 $\|u-v\|_i \le 1$。在概率单纯形上，这就是全变差距离 $\frac{1}{2}\|u - v\|_1$。一般而言，如附录 B 所示，它通过齐次性扩展为 $\mathrm{span}(Q_i - Q_i)$ 上的范数。其对偶范数衡量可行策略之间的最大损失差异，$\|g\|_{i,*} = \max_{u,\, v \in Q_i} |g,\ u - v|$。因此，归一化 $L_i \in [-1,1]$ 将单方面损失差异限制在 2 以内。对于任意仿射自映射 $\phi :\ Q_i \to Q_i$，将仿射测试函数 $f :\ Q_i \to [0,1]$ 与 ϕ 复合，可得到另一个同类测试函数。因此，ϕ 在策略范数下是非扩张的：$\|\phi(u) - \phi(v)\|_i \le \|u - v\|_i$ 对所有 u、$v \in Q_i$ 均成立。多重仿射性给出

$$\|\ell_i(x_{-i})\|_{i,*} \le 2, \qquad \|\ell_i(x_{-i}) - \ell_i(y_{-i})\|_{i,*}^2 \le 16(n-1)\sum_{j \ne i} \|x_j - y_j\|_j^2. \tag{1}$$

在线学习与线性 Φ-遗憾。我们研究具有完整信息反馈的重复同时博弈（Farina 等，2022；Kontogiannis 等，2025）。在第 t 轮，每位参与者 i 根据自身过去的策略和损失向量选择 $x_i^t \in Q_i$。参与者选定策略组合 $x^t = (x_1^t, \ldots, x_n^t)$ 后，参与者 i 承受损失 $L_i(x_i^t, x_{-i}^t)$，并观察到 $\ell_i^t := \ell_i(x_{-i}^t)$。该反馈给出了其任意两种策略之间的损失差。因此，每位参与者都面临一个在线线性优化问题。这些动态是互不耦合的。除共同的规模和时域参数外，更新仅使用该参与者自身的策略多面体和反馈历史。我们的保证针对自博弈情形，即所有参与者都遵循规定的学习规则。

参与者 i 的线性偏离是仿射自映射 $\Phi_i = \{\phi :\ Q_i \to Q_i :\ \phi \text{仿射}\}$。仿射映射 $x \mapsto Bx + b$ 在增广坐标 $(x,\ 1)$ 中是线性的。我们采用Daskalakis 等人(2025 年，式（1））中的损失约定。参与者 i 的累积线性 Φ-遗憾为

$$\mathrm{LinReg}_i(T) := \sum_{t=1}^{T} L_i(x^t) - \min_{\phi \in \Phi_i} \sum_{t=1}^{T} L_i\Big(\phi(x_i^t), x_{-i}^t\Big).$$

最小方在事后针对所有 T 轮次选择一个线性偏离，同时将对手的策略固定为其观测值。如果对于任意玩家进行此类偏离所带来的预期损失降低幅度至多为 ε，则策略剖面上的一个分布是ε-线性相关均衡。在对每个轮次赋予权重 $1/T$ 的经验分布下，玩家 i 的最大此类损失降低幅度恰好为 $\mathrm{LinReg}_i(T)/T$。

扩展式博弈。我们采用标准的有限完美回忆模型；详细定义和序列式表示参见 Koller 等人（1996）；Farina 和 Pipis（2023）。记 J_i 为玩家 i 的信息集，A_j 为 j 处的动作，并用 $p(j)$ 表示该玩家紧邻 j 之前的序列。序列集合为 $\Sigma_i := \{\varnothing\} \cup \{ja :\ j \in \mathrm{J}_i,\ a \in \mathrm{A}_j\}$，其中 $\varnothing$ 为空序列，ja 记录在 j 处采取动作 a。我们使用 $\Sigma_i^{\perp} := \{\sigma \in \Sigma_i :\ p(j) \ne$

σ 对于该玩家的每个终端序列 $j \in \mathcal{J}_i\}$ 。对局历史可以在该玩家最后一次决策之后继续。我们设 $m_i := |\Sigma_i^{\perp}|$ ，且 $M := \max_i m_i$。实现计划多面体为

$$Q_i = \left\{ x \geq 0 : x_{\varnothing} = 1,\ \sum_{a \in \mathcal{A}_j} x_{ja} = x_{p(j)}\ (j \in \mathcal{J}_i) \right\}. \tag{2}$$

这里 x_σ 是玩家自身沿途各动作的行动概率乘积 σ。由于 $x_{\varnothing} = 1$，在 Q_i 上的线性偏离可表示为实现计划的线性映射。实现计划由其在 $\Sigma_i^{\perp}$上的坐标确定。仅保留这些坐标，可得到一个仿射同构的树丛，并保留所有线性偏离。在无时序通信（UTC）偏离中，玩家可以向调解者查询其推荐的动作，而无需与游戏中的行动时序相匹配。这些偏离恰好对应实现计划上的线性偏离（Zhang 等人，2024年，定理 4.1）。附录 C给出了递归构造，并说明如何根据终端序列坐标恢复完整实现计划。

4 算法框架

紧致仿射提升。 我们用一种坐标表示策略，使得偏离可以作为非负矩阵来学习。对于一般多面体，这些坐标是处处不紧的不等式的松弛量。对于扩展式博弈（EFG），这些坐标是玩家终端序列上的实现权重。在这两种情况下，坐标变换都是从 Q_i 到多面体 $P_i \subseteq \mathbb{R}_+^{m_i}$的仿射双射。这里， m_i 表示一般多面体的松弛坐标数，或扩展式博弈的终端序列数。选取一个有理参考点 $s_i^0 \in P_i$，使其所有坐标都严格为正。可行点与该参考点之间的差 $s - s_i^0$ 张成一个维数为 $d_i = \dim Q_i$的线性空间 W_i 。该表示的仿射方程将 s 限制在 $s_i^0 + W_i$内，而施加非负性后得到

$$P_i = (s_i^0 + W_i) \cap \mathbb{R}_+^{m_i}.$$

如附录 B 所示，我们可以选择正数权重 w_i ，使得 $w_i^{\top} s = 1$ 对于每个$s \in P_i$。这种归一化使我们能够以线性形式表示仿射函数：在 P_i上，仿射函数 $g^{\top} s + b$ 等于 $(g + b w_i)^{\top} s$。

坐标变换也保留了第 3 节中定义的策略距离。每个仿射损失 $Q_i \to [0,1]$ 都对应于一个仿射损失 $P_i \to [0,1]$，且其在对应策略上的取值相同，反之亦然。因此，在任意一种表示中，两个策略之间的最大损失差都相同。

我们继续用 x_i 表示这些坐标中的策略，并选择一个环境向量$\ell_i \in \mathbb{R}^{m_i}$ 来表示在 P_i上的原始单方面损失差异。任意两个这样的选择在 P_i上的损失相差一个常数，因此它们给出相同的偏离比较。

保持仿射切片 $s_i^0 + W_i$ 映射 P_i 自身的非负矩阵。我们将这些矩阵用作学习者的决策集，并在其正数部分上使用逐元素对数障碍。

其中 $A > 0$ 表示逐项严格正性，定义

$$\mathcal{Z}_i = \{A \in \mathbb{R}_+^{m_i \times m_i} : A(s_i^0 + W_i) \subseteq s_i^0 + W_i\}, \qquad R_i(A) = -\sum_{a,b=1}^{m_i} \log A_{ab}. \tag{3}$$

命题 2 (线性偏离的紧凑表示)。 对于每位玩家 i，提升 Z_i 具有以下三个性质。

(i) Z_i 是一个由线性等式和非负性定义的紧致多面体。它包含严格正矩阵 $s_i^0 w_i^\top$。(ii) 对于 $A \in$， $s \mapsto As$ Z_i 的限制恰好是 P_i 的仿射自映射，因此表示 Q_i 上的每个线性偏离。(iii) 每个具有 $A > 0$ 的 $A \in Z_i$ 都有唯一的不动点 $s_A \in P_i$。该不动点严格为正，并满足 $w_i^\top s_A = 1$。

Daskalakis 等人 (2025 年) 给出了在目标多面体由不等式描述且能对源多面体进行高效线性优化时，对仿射映射进行高效成员资格测试的方法。每个$A \in \mathcal{Z}_i$ 都将 P_i 映射到自身，从而通过所选的仿射同构诱导出 Q_i 的仿射自映射。反过来，逐行线性规划对偶表明，Q_i 的每个仿射自映射都由某个 $A \in \mathcal{Z}_i$诱导。每个输出坐标都表示为 P_i上输入坐标的非负组合，并利用 $w_i^\top s = 1$ 吸收常数项。紧致性和 Perron–Frobenius 理论给出了不动点结论（见附录 B）。该提升有两个作用：以多项式规模表示完整的事后比较类，并且其正矩阵提供 Gordon 约化所需的不动点。学习者在矩阵空间中更新，其中 R_i 控制移动；下面的分析将这一控制转移到所采取的策略上。

我们设定 $p_i := d_i(d_i + 1)$，即内在坐标中的仿射映射系数数量，并令 $p := \max_i p_i$。由于单元素策略集的遗憾为零，我们在更新中省略它们。对于一般多面体，我们也使用 $M = \max_i m_i$。一般多面体模式使用 $\lambda > 0$：对于固定的重置点 $q_i^0 \in P_i$ （例如有理锚点 s_i^0），x_A^λ 是满足 $x = (1-\lambda)Ax + \lambda q_i^0$ 的唯一解，而重置会使映射成为压缩映射。扩展式博弈模式将 $\lambda = 0$ 设为.....，并采取 $x_A^0 = s_A$，即经 $w_i^\top s_A = 1$ 归一化的 $As_A = s_A$ 的唯一正数解。第 5.3 节利用树丛几何证明稳定性。

我们将算法 1 *ORBITS* （通过障碍迭代与变换稳定性实现乐观遗憾）称为 ORBITS。参数 λ 和 γ 分别针对一般多面体和扩展式博弈，在第 5.2 节和 5.3 节中选取。我们将预测、行动和反馈整合为一个周期。乐观矩阵更新使用上一次的偏差损失来预测下一次的偏差损失（第 3 行），而 Gordon 的不动点步骤则将该矩阵转换为实际采取的策略（第 4 行）。我们选择拉回成本（第 6 行），使得 $c_i^t(A_i^t) - c_i^t(C) = \langle \ell_i^t, x_i^t - [(1-\lambda)Cx_i^t + \lambda q_i^0] \rangle$ 对每个比较器 $C \in \mathcal{Z}_i$ 均成立。因此，矩阵遗憾本身就是 Gordon 偏离比较：只有显式重置使其与原始偏离有所不同，而在扩展式博弈模式中，该重置会消失。因此，移动和稳定性成为剩余的分析瓶颈。

在第 t 轮开始时，累积的实际成本 $\sum_{\tau=1}^{t-1} c_i^\tau$ 概括了过去，而 c_i^{t-1}则再次用作对下一次偏差损失的预测。因此，我们得到一个

算法 1 ORBITS 玩家 i（采用 Gordon 约化的乐观跟随正则化领导者）

要求： $\mathcal{Z}_i$，重置 q_i^0，参数 $\lambda \in [0,1)$，$\gamma > 0$

1: 初始化偏差损失 c^0 $_i \leftarrow 0$
2: **对于** $t = 1, 2, \ldots$ **执行**
3: $A_i^t \leftarrow$ 取使 $\left\{\gamma R_i(A) + \text{最小} \sum_{\tau=1}^{t-1} c_i^\tau(A) + c_i^{t-1}(A)\right\}$ 的参数 $_{A\in\mathcal{Z}_i,\ A>0}$
4: 找到满足 $x_i^t = (1-\lambda)A_i^t x_i^t + \lambda q_i^0$ 的 $x_i^t \in P_i$
5: 执行 x_i^t 并观察 ℓ_i^t
6: 设置 $c_i^t(A) \leftarrow (1-\lambda)\langle \ell_i^t, Ax_i^t\rangle$ 对于 $A \in \mathcal{Z}_i$
7: **结束循环**

正预测矩阵 A_i^t，它来自障碍正则化最小化（第 3 行）。对数障碍确保每次迭代 A_i^t 都严格为正。遗憾是相对于 $\mathcal{Z}_i$ 中所有非负矩阵来衡量的，包括含有零元素的矩阵。根据命题 2，这涵盖了每一种线性偏离。

矩阵本身无法直接执行，因此我们通过 Gordon 不动点方程（第 4 行）将其转换为可行策略。在一般多面体模式下，重置使该转换具有压缩性；在扩展式博弈模式下 $\lambda = 0$，学习者直接执行 s_A。游戏揭示 ℓ_i^t 后，我们通过 $c_i^t(A) = (1-\lambda)\langle \ell_i^t, Ax_i^t\rangle$ 将损失拉回偏离空间，并将该实际成本加入累积分数（第 5–6 行）。因此，只有当缓慢的矩阵移动在不动点转换后仍然存在时，乐观性才有用；这正是下面的稳定性问题。每一轮（第 3–6 行）需要在多项式规模提升上进行一次障碍正则化凸优化、一次线性不动点求解以及线性反馈更新；附录 I 给出了确定性有限精度实现和位复杂度证明。

5 遗憾分析

对数障碍控制着所选偏离矩阵在轮次之间的移动幅度。然而，玩家损失的变化取决于其对手所采取的策略。这些策略是所选映射的不动点，而映射的微小变化可能会使其不动点发生显著移动。第 5.1 节首先证明在不动点稳定性条件下的一般遗憾界。第 5.2 节和 5.3 节随后分别证明一般多面体在重置情况下以及扩展式博弈在不重置情况下所需的稳定性。

5.1 一般遗憾界

乐观跟随正则化领导者控制偏离矩阵的移动，而损失变化取决于所采取策略的移动。不动点稳定性将这两个量联系起来。令 $_A_{c,\ i}$ 解析中心和 A_i^t 仅根据实际成本形成的校正领导者，

不包含额外的预测：

$$A_{c,i} := \arg\min_{A\in\mathcal{Z}_i,\,A>0} R_i(A), \qquad \bar{A}_i^t := \arg\min_{A\in\mathcal{Z}_i,\,A>0} \{\gamma R_i(A) + \textstyle\sum_{\tau=1}^{t-1} c_i^\tau(A)\}.$$

我们通过以下方式衡量矩阵移动和策略移动：

$$E_i := \sum_{t=1}^{T}\Big[D_{R_i}(A_i^t, \bar{A}_i^t) + D_{R_i}(\bar{A}_i^{t+1}, A_i^t)\Big], \qquad V_i := \sum_{t=2}^{T} \|x_i^t - x_i^{t-1}\|_i^2.$$

能量 E_i 记录了预测与校正之间的两次有向布雷格曼步；V_i 是实际策略的平方路径长度。我们通过两个可行偏离之间损失差自上一轮以来的最大变化来衡量预测误差，

$$\Delta_i^t := \max_{B,C\in\mathcal{Z}_i} \Big|(c_i^t(B) - c_i^t(C)) - (c_i^{t-1}(B) - c_i^{t-1}(C))\Big|. \tag{4}$$

标准的乐观型 FTRL/RVU 界（Syrgkanis 等人，2015年，定理 19，命题 7）以有向布雷格曼形式给出：对于每个内部比较器 C，

$$\sum_{t=1}^{T}[c_i^t(A_i^t) - c_i^t(C)] \le \gamma D_{R_i}(C, A_{c,i}) + \gamma^{-1}\sum_{t=1}^{T}(\Delta_i^t)^2 - \tfrac{\gamma}{2}E_i. \tag{5}$$

这三项分别是比较器半径、正的预测误差和负的障碍能量。我们通过策略路径长度 V_i来界定预测误差，然后利用不动点稳定性，通过 E_i界定每个 V_i。由此可以使两个移动项相互抵消。

下一个引理描述了预测误差、所采取的策略以及偏离矩阵的移动之间的关系。

引理 3（自我博弈变差与不动点稳定性）。（i）预测误差。在自博弈中，

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(\Delta_i^t)^2 \le 4n + 32n^2\sum_i V_i.$$

（ii）策略移动。假设对于玩家 i 和所有正数 A、$B\in\mathcal{Z}_i$，其实际采用的不动点满足 $\|x_B^\lambda - x_A^\lambda\|_i^2 \le C_* D_{R_i}(B, A)$，其中某个常数为 $C_* > 0$。那么

$$V_i \le 2C_* E_i.$$

证明概要。对于 $t \ge 2$，任意两个比较器 B、$C\in\mathcal{Z}_i$ 的损失差变化为

$$(c_i^t - c_i^{t-1})(B - C) = (1-\lambda)\Big[\langle \ell_i^t - \ell_i^{t-1}, (B-C)x_i^t\rangle + \langle \ell_i^{t-1}, (B-C)(x_i^t - x_i^{t-1})\rangle\Big].$$

公式（1）通过对手移动控制第一项；B 和 C在策略范数 $\|\cdot\|_i$ 以及$\|\ell_i^{t-1}\|_{i,\,*}\le 2$ 下的非扩张性将第二项界定为 $4\|x_i^t - x_i^{t-1}\|_i$。平方并求和即可得到第一个不等式（初始化轮次贡献为 $4n$）。为了比较

连续的行动，我们首先在 A_i^{t-1}和 $\bar{A}_i^t$之间应用不动点稳定性，然后在 $\bar{A}_i^t$和 A_i^t之间应用。三角不等式以及 $(a+b)^2 \le 2(a^2+b^2)$ 在对轮次求和后给出 $V_i \le 2C_*E_i$。关键在于，该桥接过程产生的两个有向散度，恰好就是（5）中已有的、带负号的两项。附录H 给出了完整计算过程。□

定理 4（从稳定性到遗憾）。假设某个公共 $C_* > 0$ 满足 $\|x_B^\lambda - x_A^\lambda\|_i^2 \le C_* D_{R_i}(B, A)$ 对每个参与者 i 以及所有正数 A、$B \in \mathcal{Z}_i$均成立。若 $\gamma^2 \ge 128n^2C_*$，则精确迭代点满足

$$\sum_i \mathrm{LinReg}_i(T) \le \gamma\Big(\sum_i m_i^2\Big)\log(T+1) + \frac{4n}{\gamma} + 2n(\lambda T + 1). \tag{6}$$

对于每个 i，右侧相同的项也界定了线性遗憾$_i(T)$，因为恒等映射偏离不会产生损失，因此每个玩家的遗憾都是非负的。

证明概要。对（5）关于参与者求和后，我们应用引理 3（i）来界定预测误差。在定理的不动点稳定性假设下，(ii) 部分进一步给出

$$\frac{32n^2}{\gamma}\sum_i V_i \le \frac{32n^2}{\gamma}\sum_i 2C_*E_i = \frac{64n^2C_*}{\gamma}\sum_i E_i.$$

第一步使用 $V_i \le 2C_*E_i$，第二步仅对常数进行汇总。对于负的$-(\gamma/2)\sum_i E_i$ 项，净能量系数因此为 $64n^2C_*/\gamma - \gamma/2 \le 0$。这就是全部的自博弈抵消。事后最优比较器 C 可能位于$\mathcal{Z}_i$的边界上，此时对数障碍会发散。为应用（5），我们用严格正的比较器 $C_\varepsilon = (1-\varepsilon)C + \varepsilon A_{c_i}$ 近似 C，并且满足 $\varepsilon = 1/(T+1)$。逐坐标正性使障碍成本至多为 $m_i^2\log(T+1)$，而归一化将累计平滑损失界定为 2。最后，精确的 Gordon 恒等式将矩阵遗憾转换回线性 Φ遗憾，用原始比较器替换重置比较器的代价至多为 $2\lambda T$。这些项给出（6）；有向散度的计算见附录 H。□

该定理将在线优化与博弈几何区分开来：除去比较器半径因子，选择 γ 数量级的步长倒数 $n\sqrt{C_*}$，会以数量级为$n\sqrt{C_*(\sum_i m_i^2)\log(T+1)}$ 的障碍函数项 $2n\lambda T$，换取重置成本 $2n\lambda T$。因此，任何对所有线性偏离都具有紧致提升，并满足定理 4 所假设的不动点稳定性界的类别，都能继承快速的自博弈保证，而无需新的遗憾分解。正性只约束对数障碍函数的迭代点；平滑可恢复每个边界比较器。接下来我们给出 C_*的两个取值：一个适用于任意多面体且依赖于重置，另一个适用于树丛且与时域无关。

5.2 任意 H-多面体：弱重置

我们省略玩家索引，并分别用 P、$\mathcal{Z}$ 和 R 表示相应的正切片、提升多胞体和元素对数障碍；$\|\cdot\|_P$ 表示 P 上的内在策略范数。令 $\Phi(P)$ 表示 P 的仿射自映射集合，并定义 $\pi\colon \mathcal{Z} \to \Phi(P)$，其中 $\pi(A)(s) = As$ 对于 $s \in P$。

为比较重置不动点，我们将每个内部仿射自映射 $z \in \Phi(P)$ 与仿射自映射上的一个概率分布 $\mathbb{P}_z$ 关联起来，该概率分布的系数均值等于 z 的系数。因此，对于每个 $x \in P$，都有 $\mathbb{E}_{\phi\sim\mathbb{P}_z}[\phi(x)] = z(x)$。从 q^0 开始，我们在每一步之前以 λ 的概率停止，否则从 $\mathbb{P}_z$ 中独立抽取一个映射并应用。期望终点是 z 的重置不动点，因此我们可以通过采样映射序列的概率分布来比较两个重置不动点。根据独立性，它们的 KL 散度等于单次抽样的散度乘以期望抽样次数，即 $(1-\lambda)/\lambda$。我们对概率分布的选择使得单次抽样散度成为均值映射之间的 Bregman 散度，而算法的元素对数障碍可以控制该散度。

增大 λ 因而可以通过缩短序列来改善稳定性界，但这也会赋予 q^0 更大的权重，从而产生量级为 λT 的累积重置代价。调节学习率后，稳定性贡献的量级为 $\lambda^{-1/2}$，另有对数因子和博弈规模因子。平衡 T 的幂次可得 $\lambda \asymp T^{-2/3}$。

在内在坐标中，用矩阵和位移表示仿射映射。令 μ 为系数体 $\Phi(P)$ 上的归一化勒贝格测度。对于玩家 i，该系数体的维数为 $p_i \le p$，并定义

$$\Psi(\theta) = \log \mathbb{E}_{\phi\sim\mu} e^{\langle\theta,\phi\rangle}, \qquad H(z) = \Psi^*(z) := \sup_{\theta}\{\langle\theta, z\rangle - \Psi(\theta)\}.$$

对于每个内点 $z \in \Phi(P)$，令 $\mathbb{P}_z$ 的密度为

$$p_z(\phi) = e^{\langle\nabla H(z),\phi\rangle - \Psi(\nabla H(z))}$$

相对于 μ。该指数族概率分布的均值为 z，其 KL 散度与均值之间的 Bregman 散度一致：

$$\mathbb{E}_{\mathbb{P}_z}[\phi] = z, \qquad D_H(z', z) = D_{\mathrm{KL}}(\mathbb{P}_{z'} \| \mathbb{P}_z). \tag{7}$$

这些概率分布仅用于分析：算法既不会从中采样，也不会计算 Ψ 或 H。第一个引理比较这一辅助熵几何与显式对数障碍函数；第二个引理则将其转化为不动点稳定性。

引理 5（熵与对数障碍函数比较）。对于正数 A，$B \in \mathcal{Z}$，$D_H(\pi(B)\ \pi(A)) \le 4(p+1)^2 D_R(B,\ A)$。

证明概要。我们将非零可行切向量 V 缩放，使得 $\sum_{a,b} V_{ab}^2/A_{ab}^2 = 1$，并设 $f(t) = H(\pi(A+tV))$。于是，当 $|t|<1$ 时，$A+tV$ 保持为正数。自协调性和障碍函数梯度不等式给出 $-p/(1+t) \le f'(t) \le p/(1-t)$，以及 $f''(t) \ge f''(0)/(1+\sqrt{f''(0)}|t|)^2$。在 $[-1/2, 1/2]$ 上积分可得 $\sqrt{f''(0)}/(1+\sqrt{f''(0)}/2) \le 4p$，因此 $\sqrt{f''(0)} \le 2p+2$，且 $\nabla^2(H\circ\pi) \preceq 4(p+1)^2\nabla^2 R$。沿从 A 到 B 的线段再次积分即可得到有向散度界；附录 D 记录了端点和零方向细节。

□

引理 6 (重置稳定性)。 对于 $0 < \lambda < 1$ 和正数 $A, B \in Z$,

$$\|x_B^\lambda - x_A^\lambda\|_P^2 \le \frac{2(1-\lambda)(p+1)^2}{\lambda} D_R(B, A). \tag{8}$$

证明概要。 从 q^0 出发，连续复合从 $\mathbb{P}_{\pi(A)}$ 中抽取的几何分布数量 K 个独立仿射映射，其中 $\Pr(K = k) = \lambda(1-\lambda)^k$。仿射性和独立性意味着平均终点满足重置方程，因此它等于 x_A^λ。

为比较 A 和 B，在给定 $K = k$的条件下进行分析。KL链式法则给出两组采样映射记录之间的 $kD_H(\pi(B)\ \pi(A))$。对它们共同的几何分布长度取平均，得到总KL散度

$$\mathbb{E}[K] D_H(\pi(B), \pi(A)) = \frac{1-\lambda}{\lambda} D_H(\pi(B), \pi(A)).$$

记录 $(K, \phi_1, \ldots, \phi_K)$ 通过依次将采样映射作用于 q^0来确定终点。根据数据处理不等式，在 B和 A下的端点分布之间的 KL 散度因此至多为 $\frac{1-\lambda}{\lambda} D_H(\pi(B)\ \pi(A))$。对于任意仿射 $f: P \to [0, 1]$，其在两个端点分布下的期望之差的绝对值不超过它们的全变差距离。按照第 3 节的做法，对 f 取上确界并应用 Pinsker不等式，得到

$$\|x_B^\lambda - x_A^\lambda\|_P^2 \le \frac{1-\lambda}{2\lambda} D_H(\pi(B), \pi(A)).$$

引理5 由此得到（8）。 □

关键特征是只有一个因子 $1/\lambda$。直接相减重置方程得到 $[I - (1-\lambda)B](x_B^\lambda - x_A^\lambda) = (1-\lambda)(B - A)x_A^\lambda$。非扩张性将逆算子界定为 $1/\lambda$，因此直接的平方敏感度论证需要付出 $1/\lambda^2$。再生式/KL证明则只需付出期望轨迹长度（$1-\lambda$）$/\lambda$。在主界中，这给出$C_* = \mathcal{O}(p^2/\lambda)$ 和 $\gamma \asymp n\sqrt{C_*}$，因此我们在 $\lambda \asymp T^{-2/3}$处平衡主导项 $\gamma \log T + \lambda T$，并得到 $T^{1/3}$ 遗憾。

参数选择与遗憾。对于已知时域 T，设 $a_T = \lceil (T+1)^{1/3} \rceil$ 并选择

$$\lambda = a_T^{-2}, \qquad \gamma = 32n(p+1)a_T. \tag{9}$$

于是 $C_* \le 2(p+1)^2 a_T^2$，因此定理4 给出定理1的第一行。所选参数满足 $\gamma^2 \ge 128 n^2 C_*$，代入后得到

$$\sum_i \mathrm{LinReg}_i(T) \le 32n(p+1)a_T \left(\sum_i m_i^2 \right) \log(T+1) + \frac{1}{8(p+1)a_T} + 2n(T/a_T^2 + 1),$$

即 $\tilde{\mathcal{O}}(\mathrm{poly}(D) T^{1/3})$。除以 T后得到：在经验分布下，线性偏离所导致的最大期望损失降低为 $\tilde{\mathcal{O}}(\mathrm{poly}(D) T^{-2/3})$。

5.3 完美回忆扩展式博弈：无重置时的稳定性

对数刻度界利用了策略树丛的一项性质，而不仅仅是其偏离表示的多项式规模。在终端序列坐标中，树丛带有一种递归范数，在该范数下，所有可行矩阵都是非扩张的。我们利用这一递归证明下面的迹不等式。对于其正切片表示 P，令 $\mathcal{N}(P)$ 包含非负质量泛函，即为每个可行计划分配总质量一的权重。我们利用这些权重定义该范数

$$\mathcal{N}(P) = \{\ell \in \mathbb{R}^m_+ : \ell^\top x = 1 \ \forall x \in P\}, \qquad \|z\|_{X_P} = \max_{\ell \in \mathcal{N}(P)} \ell^\top |z|,$$

其中 $|z|$ 是按坐标取值的。这就是 Fan 等（2024，第4.2引理）针对终端序列定义的树丛 ℓ_1 范数，如下方的递归所示。它作用于完整的坐标空间。策略范数 $\|\cdot\|_P$ 作用于可行差的张成空间。在概率单纯形上，$\mathbf{N}(P) = \{1\}$，因此，可行差上的两个范数分别为 $\|\cdot\|_1$ 和 $\frac{1}{2}\|\cdot\|_1$。非负性以及 $A^\top \mathbf{N}(P) \subseteq \mathbf{N}(P)$ 意味着，对于每个可行偏差矩阵 A，都有 $\|Az\|_{X_P} \le \|z\|_{X_P}$。乘积与连接满足

$$\|z\|_{X_{P_1 \times \cdots \times P_b}} = \max_j \|z_j\|_{X_{P_j}}, \qquad \|z\|_{X_{\mathrm{join}(P_1,\ldots,P_b)}} = \sum_j \|z_j\|_{X_{P_j}}. \tag{10}$$

这个精确的求和/取最大值递归使符号能够沿整棵树传播，而无需乘以一个依赖于深度的敏感性常数。

根据 Perron–Frobenius 定理，每个逐项为正的方阵 A 都有一个简单正特征值，其值大于其他所有特征值绝对值。这个特征值称为*Perron* 根，并且等于 A的谱半径。它的正右特征向量和左特征向量称为*Perron* 向量，各自在相差一个缩放因子的意义下唯一（Horn 和 Johnson，2013年，第8.2节）。对于一个可行的 $A > 0$，取一个右 Perron 向量 $v > 0$。其特征值为正，因此等于 $\|Av\|_{X_P}/\|v\|_{X_P}$。非扩张性将这个比值限制在1以内。反向不等式由命题2（iii）给出的不动点 $s \in P$ 保证。由于 $w^\top s = 1$，恒等映射 $As = s$给出了一个特征值为1的非零特征向量。因此， A 的谱半径恰好为1。我们将 s_A 和 l_A 记作右 Perron 向量和左 Perron 向量，其中

$$As_A = s_A, \quad w^\top s_A = 1, \qquad l_A^\top A = l_A^\top, \quad l_A^\top s_A = 1.$$

学习者 plays s_A，因此，稳定性关注的是这个归一化右特征向量如何随 A 变化 .

对预解式的直接界可能取决于谱间隙或最小转移概率，而这些量可能任意小。定点论证只需要单个迹的符号。乘积和连接递归保持该符号不变。

引理 7（预解式迹不等式）。 对于通过 （10）中的两种运算，从绝对值递归构造出的任意范数，记 $\|T\|_{\mathrm{op}} := \sup_{\|z\| \le 1} \|Tz\|$ 为其诱导算子范数。如果 $\|U\|_{\mathrm{op}} < 1$且 $\|V\|_{\mathrm{op}} \le 1$，则 $\mathrm{tr}((I-U)^{-1}(I-V)) \ge 0$。

证明概要。 在标量叶节点处，迹为 $(1-v)\,/\,(1-u)\;\geq 0$。对于乘积或连接，通过舒尔补消去互补块，每次保留一个子节点。和范数或最大范数可确保约化后的 U 仍是严格压缩映射，相应的约化后 V 是非扩张的。随后，分块逆矩阵公式将原迹表示为子节点上同形式迹的和；在最大范数情形下，这一步利用了循环性。根据归纳法，每一项均非负。附录 F 给出了分块公式和范数界。 □

对于正的可行 A 和满足齐次提升等式的非零方向 V，我们设 $\eta=\max_{a,b}|V_{ab}|/A_{ab}$, $A^{\pm}=A\pm V/\eta$, 以及 $L(\alpha,\beta)=\alpha(I-A^{+})+\beta(I-A^{-})$。矩阵 $A^{\pm}$ 根据 η 的选择均为非负，且其和为正数 $2A$，因此它们的加权$\tilde{}$平均值 $\tilde{A}=(\alpha A^{+}+\beta A^{-})/(\alpha+\beta)$对于 $\alpha,\beta>0$为正。由于 $L=(\alpha+\beta)(I-\tilde{A})$，$\tilde{}$adj L 的列和行分别属于 $\tilde{A}$的右、左佩龙特征空间。因此，adj L 是 $s_{\tilde{A}}l_{\tilde{A}}^{\top}$的倍数，而其迹为一。这使得 tr(adj L)成为不动点公式（附录 G）中的归一化因子。

对于触发偏差，Anagnostides 等人（2023年）采用具有正系数的有理不动点公式。对于线性偏离，只需证明分母关于 L的每个系数均为非递减。下一个引理从引理 7.

引理 8 (单调佩龙分母)。 对于 $\alpha,\beta>0$, $a(\alpha,\beta):=\operatorname{tr}(\operatorname{adj}L(\alpha,\beta))$ 为正数，且满足 $\partial_{\alpha}a$, $\partial_{\beta}a\geq 0$。

证明概要。对于 $\varepsilon>0$，令 $c=\alpha+\beta+\varepsilon$和 $U=(\alpha A^{+}+\beta A^{-})/c$。由于 $A^{\pm}\in\mathsf{Z}$，两者在树复形范数 $\|\cdot\|_{X_P}$ 下均为非扩张，如前文 (10) 所示。因此 $\|U\|_{\mathrm{op}}\leq(\alpha+\beta)/c<1$。于是，对于 $0\leq t\leq 1$，$I-tU$ 可逆，因为 $x=tUx$ 蕴含 $x=0$。其行列式从 $t=0$ 处的 1 连续变化且始终不为零，因此 $L+\varepsilon I=c(I-U)$ 的行列式为正数。

对于一个 行列式为正数的可微矩阵 $B(t)$，由雅可比公式可得 s

$$\frac{d}{dt}\log\det B(t)=\operatorname{tr}\Big(B(t)^{-1}B'(t)\Big).$$

这里我们改变 α，令 β和 ε保持不变。于是 $\partial_{\alpha}(L+\varepsilon I)=I-A^{+}$以及 $(L+\varepsilon I)^{-1}=c^{-1}(I-U)^{-1}$。代入这些表达式，并应用引理7 与 $V=A^{+}$ 可得

$$\partial_{\alpha}\log\det(L+\varepsilon I)=c^{-1}\operatorname{tr}\Big((I-U)^{-1}(I-A^{+})\Big)\geq 0.$$

同样的论证也适用于 β。因此$\det(L+\varepsilon I)/\varepsilon$ 关于每个系数都是非递减的，$\tilde{}$并且当 $\varepsilon\to 0^{+}$ 时，其极限为 $a(\alpha,\ \beta)$。正矩阵 A 具有单重 Perron 根1，因此 L 具有一个单重零特征值，且该极限严格为正。附录F 给出了行列式论证。 □

定理 9 （树丛不动点稳定性）。 对于每个具有由终端序列索引的 m 坐标的树丛，以及正数 A、 $B\in Z$,

$$\|s_B-s_A\|_P^2\leq 200m^4D_R(B,A). \tag{11}$$

证明概要。对于 $m \geq 2$，令 $r = m-1$ 并设 $s(t) = s_{A+tV/\eta}$ 表示扰动矩阵的不动点。我们对 $s'(0)$ 进行界定，以控制所采用策略在该方向上的敏感性。记作 $L_t = L(1+t, 1-t)$。如附录 G所示，对于 $|t| < 1$，归一化不动点满足

$$s(t) = \frac{\operatorname{adj}(L_t)s^0}{\rho(t)}, \qquad \rho(t) := \operatorname{tr}(\operatorname{adj}(L_t)) = a(1+t, 1-t) > 0.$$

$\operatorname{adj}(L_t)$ 的每个元素都是 L_t 的带符号 $r \times r$ 子式。由于 t 中的 L_t 是仿射的，这些元素及其迹 $\rho(t)$ 都是次数至多为 r 的多项式。对于任意中心化仿射测试函数 $f\colon P \to [-1/2, 1/2]$，记 $f(s) = g^\top s + b$。于是

$$f(s(t)) = \frac{F(t)}{\rho(t)}, \qquad F(t) := g^\top \operatorname{adj}(L_t)s^0 + b\rho(t), \qquad \deg F \leq r.$$

由引理 8 的单调性以及 a 的 r 次齐次性可知，$|\rho'(0)| \leq r\rho(0)$，且对于 $|t| \leq 1/(2r)$，有 $\rho(t) \leq 2\rho(0)$。由于 $|f| \leq 1/2$，在该区间上，$|F(t)|$ 也因此被 $\rho(0)$ 所界定。对于次数至多为 r 的多项式 p，原点处的伯恩斯坦不等式为 $|p'(0)| \leq r \max_{|u|\leq 1} |p(u)|$ (Queffélec 和 Zarouf, 2019)。取 $p(u) = F(u/(2r))$，得到 $|F'(0)| \leq 2r^2\rho(0)$。商法则中的公因子 $\rho(0)$ 可以约去。对中心化仿射测试函数取上确界，可得路径 $s(t) = s_{A+tV/\eta}$ 上的 $\|s'(0)\|_P \leq \frac{5}{2}(m-1)^2$。由于 $s_{A+tV} = s(\eta t)$，其沿原方向 V 的导数为 $\eta s'(0)$。取范数并代入 η 的定义，得到

$$\left\| \frac{d}{dt} s_{A+tV} \Big|_{t=0} \right\|_P \leq \tfrac{5}{2}(m-1)^2 \max_{a,b} \frac{|V_{ab}|}{A_{ab}}. \tag{12}$$

对于 $D_R(B, A) < 1/16$，逐元素对数障碍控制着从 A 到 B 的线段上的这些相对变化。对（12）进行积分可得（11）。对于更大的散度，策略直径界已经足够。附录 G 给出了完整证明。

□

参数选择与遗憾。 稳定性界给出了 $C_* = 200M^4$，无需重置。选择

$$\lambda = 0, \qquad \gamma = 320nM^2. \tag{13}$$

这些参数满足定理 4，并给出 线性遗憾$_i(T) \leq 320n^2M^4\log(T+1) + 1/(80M^2) + 2n$。因此，博弈经验分布以如下速率趋近于线性相关均衡$\mathcal{O}(\operatorname{poly}(D)\log(T+1)/T)$。

这种与一般论证的对比在结构上有所不同，而不仅仅是定量上的差异。在一般情况下，稳定性是通过重置来构造的，其代价通过$1/\lambda$显式体现；对于树丛，精确的无重置映射具有 $C_* = \mathcal{O}(M^4)$稳定性。树递归在每一层都保持迹不等式，从而得到一个对所有正的可行偏离矩阵一致成立的稳定性界。当 $\lambda = 0$时，这个一致界使我们能够选择步长倒数 γ，且不依赖于 T。于是，式（6）中唯一随 T 增长的项就是比较项，其与 $\log(T+1)$成正比。定理 4 也适用于其他紧致表示，只要能够建立可比的稳定性估计。

UTC 实现。 终端序列矩阵提升 $\mathcal{Z}_i$ 通过其在实现计划上的作用来描述偏离。Zhang 等人的 UTC 多面体 (2024年) 描述相同的线性偏离，但依据的是玩家选择行动还是查询调解者。其局部流约束可以直接从博弈树中读出。ORBITS 可以在 $\widehat{Z_i}$ 上使用逐元素对数障碍 $R_i(z)$ 进行优化：$= -\sum_e \log z_e$，并对由此诱导的终端序列映射 $C_i(z)$取不动点并进行行动。两种障碍可能选择不同的迭代点，因此 UTC 更新需要其自身的稳定性界。

推论 10 （ORBITS 的 UTC 实现）。 令 q_i 表示移除固定根行后，玩家 i 的 *UTC* 对数障碍中的坐标数量。对于正数 z、$z' \in Z_i$，

$$\|s_{C_i(z')} - s_{C_i(z)}\|_i^2 \le 200M^4 D_{\widehat{R}_i}(z', z).$$

UTC 更新可与每个线性偏离进行竞争。如果 $\gamma^2 \ge 128n^2$（$200M^4$），其精确迭代满足

$$\sum_i \mathrm{LinReg}_i(T) \le \gamma\Big(\sum_i q_i\Big)\log(T+1) + 4n/\gamma + 2n.$$

消除单动作信息集可得 $q_i \le 5m_i^2$，并恢复 $\mathcal{O}$（$n^2M^4\log$（$T+1$）） 速率，其中 （13）。

证明概要。 $C_i(z)$ 的每个元素都是 UTC 坐标的非负线性组合。因此，可行方向 h 满足

$$\max_{a,b} \frac{|C_i(h)_{ab}|}{C_i(z)_{ab}} \le \max_e \frac{|h_e|}{z_e} \le \sqrt{\nabla^2 \widehat{R}_i(z)[h,h]}.$$

因此，UTC 障碍函数控制了（12）中的相对矩阵扰动。对该导数界进行积分，可得 UTC 坐标下的稳定性。每个线性偏离都有一个 UTC 表示，因此相同的遗憾论证可应用于 q_i 障碍项。附录 K 给出了详细内容。 □

两种表示给出相同的事后比较类。使用 UTC 还使 ORBITS 和 UTC-CFR/RM 能够在实验中基于相同的行动与查询变量进行学习，其中公共预言机计算事后最优线性偏离（附录 L）。第 5.4 节比较了两种 ORBITS 实现的每轮成本。

5.4 每次迭代的复杂度

偏离更新（第 3 行）求解一个对数障碍正则化凸规划。对于一般多面体，我们使用矩阵提升 $\mathcal{Z}_i$。在扩展式博弈中，终端序列矩阵提升具有 m_i^2 个障碍项。UTC 实现使用 $q_i =$ （$|\Sigma_i| - 1$）$|\Sigma_i| + |\mathcal{J}_i|^2$ 个行动与查询坐标（附录K）。其局部流方程为我们在实验中使用的牛顿求解器提供了一个稀疏约束矩阵（附录L）。两种实现都满足下面的多项式每轮比特复杂度界。

策略步骤（第 4 行）求解 $[I-(1-\lambda)A]x=\lambda q^{0}$ 并进行重置，或求解（$I-A$）$x=0$ 而$w_i^{\top}x=1$ 不进行重置。全信息反馈在当前紧致策略剖面处计算，并映射回偏离空间。成本更新（第 6 行）及其与累积分数的加和在线性提升坐标中完成。

对于任一域，令 k_i 表示障碍项的数量（m_i^2 或 q_i），并令 $u^0>0$ 为一个有理可行锚点。预测分数的范围至多为 $2(T+1)$，因此由一阶最优性可将精确极小值点的第 e 个分量下界为 $u_e^0/\lfloor k_i+2(T+1)/\gamma\rfloor$。该下界具有多项式编码长度。对于精确得分 b 及极小值点 u^{*}，正的可行 $\hat{u}$ 的目标差恰好为 $\gamma D_R(\hat{u},u^{*})$。附录 I 利用这一恒等式控制反馈舍入和近似优化对实际采用的策略剖面的影响。按照所述调度方案二者的总成本在遗憾总和中小于 1。每一轮使用 $\mathrm{poly}(D,\log(T+1))$ 次位运算。附录 J 讨论未知时域。

6 结论

ORBITS 为全信息自我博弈中最小化线性 Φ-遗憾提供了高效的非耦合动态。在紧致的 H-多面体上，ORBITS 将学习到的偏离与常值映射混合，得到一个内点策略。该重置限制了偏离变化时实际采用的定点能够移动的幅度。在其稳定作用与重置造成的损失之间进行权衡，可针对每个仿射自映射获得 $\widetilde{\mathcal{O}}(\mathrm{poly}(D)T^{1/3})$ 的累积遗憾。每一轮所需时间关于紧致描述长度和 $\log(T+1)$ 为多项式。

在完美回忆扩展式博弈中，序列形式多面体的递归结构使我们能够直接控制精确不动点。由此得到的$\mathcal{O}(n^2M^4\log(T+1))$遗憾界，将对数触发遗憾保证推广到了所有线性/UTC 偏离。因此，对于固定游戏，在博弈经验分布下，任意此类偏离所能带来的最大期望收益为$\mathcal{O}(\log(T+1)/T)$。高效的非耦合动态能否在每个紧致的 H-多面体上实现多对数级的线性 Φ-遗憾，且对其描述长度呈多项式依赖，仍是一个悬而未决的问题。

致谢

Alberto Marchesi 获得意大利大学与研究部根据 2025年11月12日第 18010 号部长令通过 FIS 3 项目征集提供的资助。FIS-2024-05736 项目（初创研究资助），题目为：“迈向机器学习流程中可信的数据战略性应用”（STRATDATA）。统一项目代码：D53C25002380001。Vasilis Pollatos 感谢雅典娜研究中心阿基米德研究单元为其博士研究提供的资金支持。

参考文献

Omar Abbadi、Rida Laraki 和 Panayotis Mertikopoulos。 通过高阶乐观实现一般博弈中的常数遗憾。 *arXiv:2609.04113*，2026年。

Ioannis Anagnostides、Constantinos Daskalakis、Gabriele Farina、Maxwell Fishelson、Noah Golowich 和 Tuomas Sandholm。多人一般和博弈中相关均衡的近最优无遗憾学习。发表于*ACM*计算理论研讨会，第 736–749 页，2022年。arXiv:2111.06008。

Ioannis Anagnostides、Gabriele Farina、Christian Kroer、Chung-Wei Lee、Haipeng Luo 和 Tuomas Sandholm。多人博弈中的$\mathcal{O}(\log T)$ 交换遗憾的非耦合学习动态。发表于神经信息处理系统进展，35:3292–3304，2022年。

Ioannis Anagnostides、Gabriele Farina 和 Tuomas Sandholm。近最优 Φ遗憾学习在扩展式博弈中的应用。发表于国际机器学习会议，PMLR202:814–839，2023年。

Ioannis Anagnostides、Constantinos Daskalakis、Gabriele Farina、Noah Golowich、Tuomas Sandholm 和 Brian Hu Zhang。超越正规式博弈的相关均衡复杂性。arXiv:2605.17665，2026年。STOC 2027 投稿 #988，2026年11月2日。

Ioannis Anagnostides、Gabriele Farina、Maxwell Fishelson、Haipeng Luo 和 Jon Schneider。通过基于响应的可接近性实现交换遗憾最小化。载于学习理论会议，PMLR 336:195–223，2026年。

Ioannis Anagnostides、ConstantinosDaskalakis、Gabriele Farina、Noah Golowich、Tuomas Sandholm 和 Brian Hu Zhang。马尔可夫博弈中的正规式相关性。arXiv:2610.03621，2026年。

Daniel Beaglehole、Max Hopkins、Daniel Kane、Sihan Liu 和 Shachar Lovett。均衡采样：结构化博弈中的快速无遗憾学习。载于*ACM–SIAM*离散算法研讨会，第3817–3855页，2023年。arXiv:2201.10758。

Avrim Blum 和 Yishay Mansour。 从外部遗憾到内部遗憾。 机器学习研究杂志，8：1307–1324，2007 年。

Sébastien Bubeck 和 Ronen Eldan。 熵障碍：一种简单且最优的通用自协调障碍。载于学习理论会议，PMLR 40，2015年。

Yang Cai、Constantinos Daskalakis、Haipeng Luo、Chen-Yu Wei 和 Weiqiang Zheng。近端遗憾与近端相关均衡：在线学习与博弈的一种新型可计算解概念。载于*ACM*计算理论研讨会，2026。arXiv:2511.01852。

Xi Chen 和 Binghui Peng。 博弈中的对冲：外部遗憾与交换遗憾的更快收敛。载于神经信息处理系统进展，33:18990–18999，2020。arXiv:2006.04953。Sinho Chewi。 熵障碍是 n-自协调的。 *arXiv:2112.10947*，2021。Yuval Dagan、Constantinos Daskalakis、Maxwell Fishelson 和 Noah Golowich。 从外部遗憾到交换遗憾 2.0：针对大型行动空间的高效归约。载于 *ACM*计算理论研讨会，2024。 arXiv:2310.19786。Constantinos Daskalakis、Gabriele Farina、Noah Golowich、Tuomas Sandholm 和 Brian Hu Zhang。 扩展式博弈中交换遗憾的下界。 arXiv:2406.13116，2024。Constantinos Daskalakis、Gabriele Farina、Maxwell Fishelson、Charilaos Pipis 和 Jon Schneider。 一般凸博弈中线性相关均衡的高效学习与计算。载于 *ACM*计算理论研讨会，2025。Zhiyuan Fan、Christian Kroer 和 Gabriele Farina。 膨胀熵的最优性及扩展式博弈在线学习的下界。载于 神经信息处理系统进展，第37卷，2024年。

Gabriele Farina 和 Charilaos Pipis。 不完全信息序贯博弈中的多项式时间线性交换遗憾最小化。载于神经信息处理系统进展，36:28255–28280，2023年。

Gabriele Farina、Chung-Wei Lee、Haipeng Luo 和 Christian Kroer。 用于 0/1-多面体博弈的核化乘法权重：弥合扩展式博弈与正规式博弈中学习之间的差距。载于 国际机器学习会议，PMLR 162:6337–6357，2022年。arXiv:2202.00237。

Dean P. Foster andRakesh V. Vohra. 校准学习与相关均衡。 《博弈与经济行为》，21(1–2):40–55，1997。Geoffrey J. Gordon、Amy Greenwald 和 Casey Marks。 凸博弈中的无遗憾学习。载于 国际机器学习会议，第 360–367 页，2008年。Martin Grötschel、László Lovász 和 Alexander Schrijver。 《几何算法与组合优化》。施普林格，1988年。Junsoo Ha。 一般和博弈中乐观 Hedge 的时域无关遗憾界。arXiv:2609.22839，2026年。Soumita Hait、Ping Li、Haipeng Luo 和 Mengxiao Zhang。 比较器自适应 Φ遗憾：更优界、更简单的算法及其在博弈中的应用。载于神经信息处理系统进展，第 38 卷，2025 年。 arXiv:2505.17277。

Sergiu Hart 和 Andreu Mas-Colell。导向相关均衡的简单自适应程序。《计量经济学》，68(5):1127–1150，2000。Samid Hoda、Andrew Gilpin、Javier Peña 和 Tuomas Sandholm。计算序贯博弈纳什均衡的平滑技术。《运筹学数学》，35(2):494–512，2010。Roger A. Horn 和 Charles R. Johnson。《矩阵分析》。剑桥大学出版社，第 2 版，2013。Daphne Koller、Nimrod Megiddo 和 Bernhard von Stengel。扩展式双人博弈均衡的高效计算。《博弈与经济行为》，14(2):247–259，1996。H. W. Kuhn。一种简化的双人扑克。载于 H. W. Kuhn 和 A. W. Tucker（编），《博弈论贡献》，第 1 卷，第 97–103 页。普林斯顿大学出版社，1950。Andreas Kontogiannis、Vasilis Pollatos、Gabriele Farina、Panayotis Mertikopoulos 和 Ioannis Panageas。超越完全信息的多面体博弈高效核化学习：从 Blotto 上校博弈到拥塞博弈。载于《神经信息处理系统进展》，第 38 卷，2025 年。arXiv:2509.20919。Andreas Kontogiannis、Vasilis Pollatos、Panayotis Mertikopoulos 和 Ioannis Panageas。组合多臂老虎机中的高效交换遗憾最小化。载于人工智能与统计国际会议，PMLR 300:1963–1971，2026。会议论文集版本。Mingyang Liu、Gabriele Farina 和 Asuman Ozdaglar。一般博弈中的常数个体遗憾。arXiv:2608.31166，2026年。Tung Mai。通过乐观转移矩阵实现一般和博弈中的常数交换遗憾。arXiv:2609.16751，2026年。Binghui Peng 和 Aviad Rubinstein。快速交换遗憾最小化及其在近似相关均衡中的应用。载于 ACM计算理论研讨会，2024年。arXiv:2310.19647。Hervé Queffélec 和 Rachid Zarouf。关于多项式的伯恩斯坦不等式。arXiv:1903.10801，2019年。Ashkan Soleymani 和 Georgios Piliouras。博弈中的乘法乐观与常数遗憾。arXiv:2609.21976，2026年。Finnegan Southey、Michael Bowling、Bryce Larson、Carmelo Piccione、Neil Burch、Darse Billings 和 Chris Rayner。贝叶斯虚张声势：扑克中的对手建模。载于第二十一届人工智能不确定性会议论文集，第 550–558 页，2005年。

Vasilis Syrgkanis、Alekh Agarwal、Haipeng Luo 和 Robert E. Schapire。 博弈中正则化学习的快速收敛。载于 神经信息处理系统进展，28，2015年。

Taira Tsuchiya。 通过混合正则化实现多人一般和博弈中的次对数交换遗憾。 arXiv:2608.04149，2026年。

Taira Tsuchiya、Haipeng Luo 和 Shinji Ito。 博弈中的尺度不变快速收敛。 arXiv:2602.11857，2026年。

Bernhard von Stengel 和 Françoise Forges。 扩展式相关均衡：定义与计算复杂性。《运筹学数学》，33(4)：1002–1022，2008年。

Martin J. Wainwright 和 Michael I. Jordan。 图模型、指数族与变分推断。 机器学习基础与趋势，1(1–2)：1–305，2008年。

Brian Hu Zhang、Gabriele Farina 和 Tuomas Sandholm。一般序贯博弈中线性相关均衡的调解者解释与更快的学习算法。载于国际学习表征会议，2024年。

A 扩展相关工作

近似常数和常数交换遗憾。表 1 中的对抗式完全交换一行给出了对 N 和 T 一致成立的界。Dagan 等人（2024，推论 1.2）所需的轮次数具有 $(c\log(N)/\varepsilon^2)^{c/\varepsilon}$ 的形式，其中 c 为普适常数；Peng 和 Rubinstein（2024）也得到了相同的权衡。为反解该式，令 $L=\log T$，并令 $J=\log\log(eNT)$。对于 $\varepsilon = CJ/L < 1$，有 $\log(c\log(N)/\varepsilon^2) = \mathrm{O}(J)$，因为 $\log\log N \le J$ 且 $\log L \le J$。取足够大的普适常数 C，即可使所需轮次数至多为 T。当 $CJ/L \ge 1$ 时，平凡的线性界已足够。对于 N、$T \ge 2$，这给出了所示的 $\mathrm{O}(T\log\log(NT)/\log T)$。

正规形式方面的最新进展进一步揭示了 Blum–Mansour 约简与乐观学习之间的相互作用。 Tsuchiya (2026) 将熵与对数障碍正则化相结合，把对数时域因子改进为其平方根。Mai (2026) 则通过双尺度高阶乐观方法控制偏离收益的预测误差，在考虑转移矩阵平稳分布变化的同时，获得了与时域无关的上界。这项工作还给出了一种稳健变体，将自博弈的常数上界保持在一个通用因子以内，并保证对任意有界收益序列的 $\mathcal{O}(\sqrt{NT}\log N)$ 交换遗憾。其对抗式保证仍随显式动作数呈多项式依赖，因此并不与第 2 节讨论的紧致域下界相矛盾。

Hait 等人 提出的比较器自适应结果 (2025) 提供了另一条较早实现常数遗憾的途径，但要求对游戏施加限制：对于每个行动序列，所有玩家未截断外部后悔 之和必须非负。其上界可具体化为完全交换遗憾的 $\mathcal{O}((nN+n^2)\log N)$。这一条件在任意一般和博弈中并不成立，也不应与交换遗憾本身的非负性混为一谈。 Tsuchiya 等人(2026) 处理的是另一个问题：通过完全尺度自由、尺度不变的动力学，实现对数刻度交换遗憾。

常数外部遗憾。近期与时域无关的外部遗憾结果包括高阶乐观动力学（Liu 等人，2026年；Abbadi 等人，2026年）、乘法式乐观（Soleymani 和 Piliouras，2026年），以及对乐观 Hedge 的时域无关分析（Ha，2026年）。后一个界限隐含依赖于有限游戏的维度。这些保证针对外部比较器类；它们本身并不能得到常数交换遗憾或用于线性 Φ遗憾的紧凑算法。将外部遗憾方法分别应用于 Blum–Mansour 转移矩阵的各行，需要额外的分析，因为平稳分布将各行学习器的反馈耦合起来（Mai，2026年）。

其他紧凑方法。 结构化外部遗憾方法利用采样和核化，在若干博弈族中避免纯动作枚举（Beaglehole 等人，2023；Farina 等人，2022年；Kontogiannis 等人，2025 年）。它们的均衡保证基于外部遗憾，因此不一定能控制相对于每个线性偏离的遗憾。Cai 等人（2026年，定理 4）在 Lipschitz 光滑凸博弈中界定了个体近端遗憾。为了控制相对于对称仿射偏离的遗憾，他们将其与恒等映射混合以获得近端映射，并对所得遗憾界进行缩放。在

一般多面体证明中使用的Bubeck和Eldan（2015年）给出了熵几何；其精确的自协调参数由 Chewi （2021）确立。

B 多面体预备知识与内在几何

B.1 仿射包约化与编码几何

省略玩家索引，记为 $Q=\{x\in\mathbb{R}^d:\ Cx\le b\}$，其中 $C\in\mathbb{R}^{m\times d}$。不失一般性，我们假设 Q 是满维的，且没有不等式在整个Q上处处取等。

事实上，对于任意非空有理紧多面体，可通过线性规划（LP）分别最大化其上的每个松弛量。最大值为零的行在整个多面体上都是等式。它们的公共等式空间就是其仿射包：对每个剩余行取一个可行正松弛见证点并求平均，即可得到一个使所有这些行均严格成立的点。因此，该点在等式空间中的一个邻域是可行的。如果没有剩余行，则可行集本身就是该等式空间。有理参数化 $x=x^0+Uy$ 该空间即可消除这些等式。在新坐标中，损失差使用 $U^\top\ell$ ，而通过该参数化进行共轭会给出仿射自映射之间的双射。因此，每个偏离比较以及遗憾都得以保持。这些约化使用多项式时间线性规划和有理线性代数（Grötschel 等人，1988年）。零维策略集产生的遗憾为零。

将 $c_j^\top$记为 C的第 j 行。逐坐标线性规划给出 $Q\subseteq[-R,R]^d$，其有理 $R\ge1$ 具有多项式编码长度。内接盒可由以下方式得到

$$\max_{q,r}\ r\quad\text{subject to}\quad c_j^\top q+r\|c_j\|_1\le b_j\quad(j\in[m]),\qquad 0\le r\le1.$$

这些约束精确地表示 $q+r[-1,1]^d\subseteq Q$。满维性保证存在正数最优值 r ，以及一组具有多项式编码长度的有理最优解 $(q,\ r)$。

将仿射映射写为 $x\mapsto Gx+h$。若 G 和 $h-q$ 的每个系数的绝对值都不超过 $r/(dR+1)$，则每个 $x\in Q$ 的像都属于该内接盒。因此， $\Phi(Q)$在所有 $d(d+1)$系数方向上都包含一个盒子。对于外界，在 q和 $q+re_j$处计算任意自映射，其中 e_j是坐标向量 j。两个像都位于 $[-R,R]^d$ 中，因此 Ge_j的每个元素的绝对值都不超过 $2R/r$。恒等映射 $h=\phi(q)-Gq$ 由此给出 $\|h\|_\infty\le R+2dR^2/r$的界。这些论证给出了具有多项式编码长度的内半径和外半径；即使 r 小到呈指数级，也足以满足要求。

B.2 A仿射映射不等式与紧致非负提升

逐行线性规划对偶刻画原坐标中的仿射映射 :

$$(G,h)\in\Phi(Q)\quad\Longleftrightarrow\quad\exists\Lambda\in\mathbb{R}_+^{m\times m}:\ \Lambda C=CG,\quad\Lambda b+Ch\le b.\tag{14}$$

确实，第 j 行要求最大化 $_{Cx\le b}\,c_j^\top Gx+c_j^\top h\le b_j$。其对偶问题是在满足 $C^\top v_j=G^\top c_j$ 的 $v_j\ge0$ 上最小化 $b^\top v_j+c_j^\top h$ 。原问题可行且有界，因此对偶最优解存在。

堆叠其转置得到 Λ。尽管 $\Phi(Q)$ 是紧集，乘子纤维仍可能无界。

对于在线障碍集 $s = b - Cx$、 $s^0 = b - Cq > 0$ 和 $W = \mathrm{range}(-C)$。若 $Cv = 0$with $v \neq 0$，则经过任意可行点的直线都会位于 Q，这与有界性矛盾。因此，松弛映射是到 $P = (s^0 + W) \cap \mathbb{R}^m_+$ 上的仿射同构，其中 dim $W = d$and aff $P = s^0 + W$。此外 $W \cap \mathbb{R}^m_+ = \{0\}$，因为该交集中任何非零向量都会从 s^0 引出一条无界射线。因此，严格择一性定理给出 $w > 0$perpendicular to W。将其缩放至 $w^\top s^0 = 1$，于是 $w^\top s = 1$ on aff P。通过在满足 $C^\top w = 0$、 $w^\top s^0 = 1$，且对每个 a 均有 $w_a \geq t \geq 0$ 的条件下最大化 t，即可得到一个有理数解。

对于一个扩展式博弈，附录 C 构造了一个非负线性映射，可根据终端序列重建完整计划。在终端投影的仿射包上，重建出的坐标因仿射性而满足序列式等式。非负的终端坐标会重建出非负计划，并还原给定的终端项。因此，终端多面体也等于其仿射包与非负正交象限的交集。均匀的正数行为概率给出一个正数参考点。

命题 2 的证明。（i）紧致性。设 F 的行张成 $W^\perp$， B 的列张成 W，并令 $h_0 = Fs^0$。于是 $P = \{s \geq 0 : Fs = h_0\}$。保持 aff P 等价于$FAs^0 = h_0$和 $FAB = 0$。这些等式以及 $A \geq 0$定义了式 (3) 的闭提升集 Z $=$Z(P)。由于 $As^0 \in P$，非负性意味着

$$A_{ab} s^0_b \leq (As^0)_a \leq \max_{s \in P} s_a.$$

因此每个坐标都有界。矩阵 $s^0 w^\top$ 严格为正，并将整个仿射包映射到 s^0。

(ii) 表示。对于仿射自映射的每个输出坐标，选择其在 P 上的环境仿射表示 $g^\top s + \beta$。其最小值是非负的。根据强线性规划对偶性，存在无约束乘子 y，使得 $F^\top y \leq g$，并且

$$\kappa := h_0^\top y + \beta = \min_{Fs = h_0,\ s \geq 0} (g^\top s + \beta) \geq 0.$$

在 P 上，该输出等于 $(g - F^\top y + \kappa w)^\top s$，其系数向量是非负的。将各行叠加起来，得到一个在 P 上与该映射一致的非负矩阵。其线性性以及 $A(P) \subseteq P$ 蕴含 $A(\mathrm{aff}\ P) \subseteq \mathrm{aff}\ P$，如所需。反之，每个可行矩阵都保持仿射切片和非负性，因此将 P 映射到 P 中。

(iii) 不动点。布劳威尔不动点定理保证每个可行的 A 在非空紧凸集 P中都有一个不动点。P 中的每个点都满足 $w^\top s = 1$，且非零。此外，对于每个整数 $k \geq 1$， A^k都属于有界提升，因此没有特征值的模大于一。对于 $A > 0$，佩龙–弗罗贝尼乌斯定理保证存在一个单重特征值一、严格为正的右特征向量和左特征向量，并且其他所有特征值的模都严格小于一。可行的不动点是经 $w^\top s_A = 1$归一化的唯一右 Perron 向量 s_A。将左特征向量 l_A按 $l_A^\top s_A = 1$归一化。对于任意 $s \in P$，集合 $\{x \in P : l_A^\top x = l_A^\top s\}$非空、紧且凸。由于 $l_A^\top A = l_A^\top$， A将该集合映射到自身。根据布劳威尔不动点定理

该集合包含一个不动点，根据唯一性，该不动点必为 s_A。因此，对于每个$s \in P$，都有 $l_A^\top s = l_A^\top s_A = 1$。附录G将用到这个恒等式。 □

记 $\pi:\ \mathcal{Z} \to \Phi(P)$ 为在仿射映射坐标中限制到 $\operatorname{aff} P$所得的映射。根据 (ii)，它是一个仿射满射。其线性部分将提升的方向空间映到 $\Phi(P)$ 的方向空间，因为它将仿射包映到仿射包。任意正的 A 的一个充分小的相对邻域都包含于 $\mathcal{Z}$。有限维空间中的满射线性映射会将相对邻域映到其像的邻域，因此 $\pi(A)$ 是系数体的内点。这正是熵比较中所用的内点性质。

B.3 非扩张性与收益归一化

仿射测试函数 $f:\ P \to [0,1]$ 的线性部分 g 满足：对所有 $u, v \in P$，都有 $|g(u-v)| \le 1$。反过来，W 上满足这些不等式的每个线性泛函都定义了这样一个测试函数：从 $s \mapsto g(s - s^0)$ 在 P 上的取最小值中减去该泛函。因此，测试泛函的集合是 W 中 $P - P$ 的极集。差集是紧、凸、对称的，并且在 W 中包含零的一个邻域。根据双极定理，它是策略范数的单位球；该范数在 W 上的公式为

$$\|v\|_P = \inf\{a > 0 : v = a(x - y),\ x, y \in P\}.$$

对于 $v \neq 0$，由紧致性可知下确界可以取得。若 $v = a(x-y)$，则任意仿射自映射的线性部分将 v 映射到 $a(\phi(x) - \phi(y))$，因此其算子范数至多为一。因此，$(1-\lambda)\phi + \lambda q^0$ 的利普希茨常数对于 $0 < \lambda < 1$ 至多为 $1 - \lambda$，并且在 P中具有唯一不动点。同样，在 P 中的一点处计算两个可行偏离之差，会得到 $P - P$中的一个元素。

由于$\|g\|_{P,\,*} = \max_{u,\ v \in P} |\langle g,\ u - v\rangle|$，$[-1,1]$ 中的损失给出 $\|\ell_i(x_{-i})\|_{i,\,*} \le 2$。要比较两个对手的策略组合，每次只改变一个对手的 j。若 $a = \|x_j - y_j\|_j > 0$，则将$x_j - y_j = a(u_j - v_j)$写作某个 $u_j,\ v_j \in P_j$。对于两个自身策略 $u_i,\ v_i$，它们的损失差异所引起的变化为 a倍

$$L_i(u_i, u_j, \ldots) - L_i(v_i, u_j, \ldots) - L_i(u_i, v_j, \ldots) + L_i(v_i, v_j, \ldots)$$

由多重仿射性可知，在固定其他玩家的情况下。所示表达式的绝对值至多为四。差值为零时，该项贡献为零。对 u_i, v_i 取上确界，对对手求和，并应用柯西–施瓦茨不等式，得到

$$\|\ell_i(x_{-i}) - \ell_i(y_{-i})\|_{i,*} \le 4 \sum_{j \neq i} \|x_j - y_j\|_j \le 4 \sqrt{(n-1) \sum_{j \neq i} \|x_j - y_j\|_j^2}.$$

这证明了（1），也包括求和为空的一名玩家情形。

C　扩展式与 Treeplex 表示

我们采用常见的显式有限完美回忆扩展式博弈模型。在每个信息集上，所有节点的行动集都相同。玩家自身的序列记录其先前的信息集及

动作。完美回忆使得每个信息集处的父自身序列 $p(j)$ 都有明确定义 j。实现约束为 (2)；对于父序列实现值非零的情况，可得到行为概率 $x_{ja}/x_{p(j)}$；而在父序列实现值为零的信息集上，行为可以任意选择。反过来，将各自身序列上的行为概率相乘，即可得到一个可行的实现计划。完美回忆保证了与混合策略的实现等价性（Koller 等人，1996）。机会因素和其他参与者的到达概率体现在收益中，而非玩家的实现约束中。

在流入质量为 1 的玩家选择节点， t_a 是分配给动作 a的实现权重。如果 a 之后的延续策略属于 P_a，其坐标将按 t_a缩放。在互不相交的坐标块中，这就构成了凸连接

$$\mathrm{join}(P_1,\dots,P_b)=\{(t_1z_1,\dots,t_bz_b): t\in\Delta_b,\ z_j\in P_j\ (j\in[b])\},$$

其中 Δ_b 是概率单纯形。在自己的序列之后，可能经由机会因素或对手的动作到达不同的后续信息集。一个计划会指定每个此类信息集上的延续策略，因此相应的坐标块构成笛卡尔积。从标量叶节点 $\{1\}$ 开始，便得到递归的树丛表示（Hoda 等人，2010；Fan 等人，2024年）。

投影后，剩余坐标以终端序列为索引。例如，如果两个信息集都可能接续同一条权重为 t 的自身序列，那么它们各自的后续动作块之和都等于 t。任一求和结果都能还原出 t ，它们的平均值也可以。一般来说，要还原被舍弃的流入权重，可取一个具有质量泛函 $w_j^\top s=1$ 的子节点，并将其作用于缩放后的坐标块，得到 $w_j^\top(t_jz_j)=t_j$。在乘积处，对正数子块质量泛函取平均；在连接处，则将它们相加。递归进行即可通过线性映射还原完整实现计划。因此，限制到终端序列不会改变任何线性偏离及其遗憾。

本文中的比较器类是实现计划上的所有仿射自映射 $\phi_i:\ Q_i\to Q_i$ 。加入根坐标 $x_\varnothing=1$ 后，每个此类映射都会变为线性映射；反过来，仿射切片上的线性自同态也是仿射自映射。因此，这一表示捕获了序列形式中的完整线性/UTC 偏离类，同时仍不同于直接定义在纯应变计划上的任意函数。

C.1 归一化对偶性与共同环境范数

为完整起见，我们重述 EFG 论证中使用的唯一辅助范数。对于以终端序列为索引的正切片表示 $P\subset\mathbb{R}^m_+$ ，令

$$\mathcal{N}(P)=\{\ell\ge 0:\ell^\top s=1\text{ for all }s\in P\},\qquad \|z\|_{X_P}=\max_{\ell\in\mathcal{N}(P)}\ell^\top|z|,$$

其中 $|z|$按坐标逐项取值。由于仿射包 $P=s^0+W$，非负权重 ℓ恰当属于$\mathcal{N}(P)$，当且仅当 $\ell\in W^\perp$ 且 $\ell^\top s^0=1$。 s^0的正性意味着，对于每个 $\ell\in\mathcal{N}(P)$，都有 $\ell_j\le 1/s_j^0$；而上文的质量泛函提供了一个正数成员。此外， $A\in\mathcal{Z}(P)$意味着，对于每个 $\ell\in\mathcal{N}(P)$，都有$A^\top\ell\in\mathcal{N}(P)$。因此

$$\|Az\|_{X_P}\le\max_{\ell\in\mathcal{N}(P)}(A^\top\ell)^\top|z|\le\|z\|_{X_P}.$$

这里使用的是同一个完整提升；它不会扩展策略多面体，也不会改变其偏离。

对于笛卡尔积 $P = \prod_{j=1}^{b} P_j$，我们记作 $\Delta_b := \{t \in \mathbb{R}_+^b : \sum_{j=1}^{b} t_j = 1\}$。每次改变一个分量可知，$\ell_j^\top s_j = t_j$ 在每个块上都是常数。这些常数非负，且总和为一。若 $t_j > 0$，则除以 t_j 以得到子块质量泛函。若 $t_j = 0$，对严格正的子节点锚点进行检验会迫使整个非负块 ℓ_j 消失。反过来，任意选择权重为这些值且归一化的各块，都会使总质量为一。因此

$$\mathcal{N}(P) = \{(t_1\ell_1, \dots, t_b\ell_b) : t \in \Delta_b,\ \ell_j \in \mathcal{N}(P_j) \text{ for } j = 1, \dots, b\}\,.$$

对于凸连接 $P = \text{连接}(P_1, \dots, P_b)$，对仅支撑在单个子节点上的点进行检验可得

$$\mathcal{N}(P) = \prod_{j=1}^{b} \mathcal{N}(P_j).$$

在这两个显式公式中取 $\ell^\top |z|$ 的最大值，得到 (10)。在标量点处，范数就是绝对值。递归可以有任意深度；下面的舒尔论证传递的是符号，而非依赖于维数的因子。

D Pro熵与对数障碍函数比较

下面的引理以稍微一般化的形式陈述，以便将新的比较结果与后文所用的特定熵函数区分开来。

回顾一下，在满维凸体 K 上的 ν-自协调障碍函数是一个凸C^3 函数 $F\colon \operatorname{int} K \to \mathbb{R}$，其海森矩阵正定，并且当 x 从内部趋近∂K时，$F(x) \to +\infty$；此外，对任意直线限制 $f(t) := F(x + tv)$，都有

$$|f'''(t)| \le 2f''(t)^{3/2}, \qquad f'(t)^2 \le \nu f''(t)$$

当 $x + tv \in \operatorname{int} K$成立时。这些分别是自协调性不等式和障碍函数梯度不等式。具有非空内部的凸体在 $\mathbb{R}^k$ 上的熵障碍函数满足 $\nu = k$(Bubeck和Eldan，2015年；Chewi，2021）。中的加性归一化常数不会影响这两个不等式。

引理 11（正提升上的比较）。考虑紧集 $Z = \{z \ge 0 : Ez = b\}$，其中含有一个正点；一个仿射满射 $\pi :\ Z \to K$ 映到一个满维紧凸体上；以及一个定义在 K 上的 ν-自协调障碍函数 F，满足 $\nu \ge 1$。记作$R(z) := -\sum_a \log z_a$。对于正数 z、$z' \in Z$，

$$D_F(\pi(z'), \pi(z)) \le 4(\nu + 1)^2 D_R(z', z).$$

证明。我们固定一个正数 z 和一个方向 v，满足 $Ev = 0$。如果 $v = 0$，结论显然成立。我们对 v 进行缩放，使得 $\sum_a v_a^2/z_a^2 = 1$。由于这意味着对每个坐标都有 $|v_a| \le z_a$，因此点$z + tv$对每个 $|t| < 1$都是正数且可行。它属于 $\operatorname{ri} Z$，因此根据仿射

满射性（附录中的邻域论证 B）。记 $f(t) := F(\pi(z + tv))$；若投影方向为零，则无需证明，因此我们假设 $f''(0) > 0$。

我们首先推导端点梯度界

$$-\frac{\nu}{1+t} \le f'(t) \le \frac{\nu}{1-t} \qquad (-1 < t < 1). \tag{15}$$

如果 $f'(t) > 0$，凸性会使其在右侧为正。障碍梯度不等式给出 $(1/f')' = -f''/(f')^2 \le -1/\nu$。如果 $f'(t) > \nu/(1-t)$，则积分会迫使 $1/f'$ 在严格小于 1 之前变为非正，这与定义域内的有限性和可微性相矛盾。这就证明了上界，包括平凡的非正情形。将直线反射即可证明下界。

自协调性还给出 $|(f''^{-1/2})'| \le 1$。我们将这一微分估计与以下内容结合：

$$\begin{aligned} f''(t) &\ge \frac{f''(0)}{\left(1 + \sqrt{f''(0)}|t|\right)^2}, \\ \frac{f''(0)}{1 + \sqrt{f''(0)}/2} &\le f'(1/2) - f'(-1/2) \le 4\nu, \\ \sqrt{f''(0)} &\le \nu + \sqrt{\nu^2 + 4\nu} \le 2\nu + 2. \end{aligned}$$

第一行通过对 $|(f''^{-1/2})'| \le 1$ 进行积分得到。第二行对 f'' 在 $[-1/2, 1/2]$ 上积分，并应用 (15)；最后一行求解所得的二次不等式。恢复 v 的尺度即可证明逐点 Hessian 支配，

$$\nabla^2(F \circ \pi)(z)[v, v] \le 4(\nu + 1)^2 \nabla^2 R(z)[v, v]. \tag{16}$$

对于 $v = z' - z$，沿正的可行线段对（16）积分。由于 π 是仿射的 e,

$$D_F(\pi(z'), \pi(z)) = \int_0^1 (1-t)\, \nabla^2(F \circ \pi)(z + tv)[v, v] dt.$$

这就证明了所述有向方向下的结果。 □

引理 5 的证明。我们将引理 11应用于紧致全仿射提升 $\mathcal{Z}$、其在实际仿射映射系数上的投影，以及 $\Phi(P)$上的熵障碍 $\nu = p_i = d_i(d_i + 1)$，其中玩家 i对应的系数体具有非空内部，见附录 B。由于 $p_i \le p$，一致比较因子为 $4(p+1)^2$。 □

几何推论。 元素对数障碍 R 与熵障碍 H 诱导出不同的几何结构，且通常具有不同的极小值点。引理 11 利用显式矩阵散度控制投影障碍迭代点之间的熵散度，其多项式损失取决于仿射映射维数。此比较给出了引理6中使用的一般重置稳定性估计。

E 再生式不动点稳定性

E.1 熵/KL恒等式

我们考虑一个由仿射映射系数构成的满维紧凸体 $\Phi \subset \mathbb{R}^k$，使用其均匀基准分布 μ，并定义

$$\Psi(\theta) = \log \mathbb{E}_{\phi\sim\mu}\left[e^{\langle\theta,\phi\rangle}\right], \qquad H = \Psi^*.$$

在积分号下求导可得倾斜分布的均值为 $\nabla\Psi(\theta)$，其协方差为 $\nabla^2\Psi(\theta)$。由于密度在一个满维集合上为正数，后者是正定的。为说明每个内点均值 z 都能取到，取满足 $z + 2\rho B_2 \subseteq \Phi$ 的 $\rho > 0$，其中 B_2 是欧几里得单位球。对于 $\theta \neq 0$，在以 $z + \rho\theta/\|\theta\|_2$ 为中心、半径为 $\rho/2$ 的球上积分。其μ质量 $\alpha > 0$ 与 θ 的方向无关，因此

$$\Psi(\theta) - \langle\theta, z\rangle \;\geq\; \tfrac{\rho}{2}\|\theta\|_2 + \log\alpha.$$

因此，严格凸目标函数有唯一极小值点 θ_z，满足 $\nabla\Psi(\theta_z) = z$。凸共轭给出 $\theta_z = \nabla H(z)$ 以及 $H(z) = \langle\theta_z, z\rangle - \Psi(\theta_z)$ (Bubeck 和 Eldan，2015 年；Wainwright 和 Jordan，2008 年)。因此，对于每个内点 z，逐点密度

$$p_z(\phi) := \exp(\langle\nabla H(z), \phi\rangle - \Psi(\nabla H(z))), \qquad \phi \in \Phi,$$

相对于 μ 进行了归一化，且均值为 z。我们将通过以 p_z 对 μ 进行加权而得到的概率分布记为 $\mathbb{P}_z$。对于所有内部 z，z'，直接代入可得

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{z'}\|\mathbb{P}_z) &= \langle\nabla H(z') - \nabla H(z), z'\rangle - \Psi(\nabla H(z')) + \Psi(\nabla H(z)) \\ &= D_H(z', z). \end{aligned}$$

这就证明了（7）。概率分布 $\mathbb{P}_z$ 仅用于分析；算法从不从中采样。

E.2 引理 6的证明

证明。 我们固定一个内部均值仿射映射 z，并将 $x = (1-\lambda)z(x) + \lambda q^0$ 的唯一解记为 x_z^λ。从 q^0 出发，我们抽取一个满足以下条件的随机数 K：

$$\Pr(K = k) = \lambda(1-\lambda)^k, \qquad k = 0, 1, \ldots.$$

在给定 $K = k$ 的条件下，我们抽取独立仿射映射 ϕ_1，……，$\phi_k \sim \mathbb{P}_z$，并令

$$X_z := \phi_k \circ \cdots \circ \phi_1(q^0),$$

当 $k = 0$ 时为 $X_z = q^0$。令 $u_k = \mathbb{E}[X_z \mid K = k]$。给定 $K = k$，最后一个映射独立于前面的 $k-1$ 个映射；先对最后一个映射取平均，再对其输入取平均，得到 $u_k = z(u_{k-1})$，其中 $u_0 = q^0$。因此，几何混合的均值为

$$\mathbb{E}X_z = \lambda q^0 + (1-\lambda)\sum_{k\geq 0}\lambda(1-\lambda)^k z(u_k) = \lambda q^0 + (1-\lambda)z(\mathbb{E}X_z).$$

该级数收敛，因为 P 是紧的；最后一个等式利用了仿射性以及权重之和为一这一事实。重置不动点的唯一性给出 $\mathbb{E}X_z = x_z^\lambda$。

对于两个均值映射 z、z'，我们以公共事件 $K = k$ 为条件。独立性和 KL链式法则表明，从 z'-记录分布到 z-记录分布的条件KL散度为 $k\, D_{\mathrm{KL}}(\mathbb{P}_{z'} \| \mathbb{P}_z) = k\, D_H(z', z)$，其中最后一个等式为（7）。该记录包含 K，其边缘分布在两个实验中相同。KL链式法则和 $\mathbb{E}K = (1-\lambda)/\lambda$ 因此给出完整轨迹记录之间的散度 $(1-\lambda)D_H(z', z)/\lambda$。我们将终端点 X_z 的分布记为 Q_z。由于 X_z 是轨迹记录的函数，数据处理给出

$$D_{\mathrm{KL}}(Q_{z'} \| Q_z) \le \frac{1-\lambda}{\lambda} D_H(z', z).$$

对于每个仿射测试函数 $f\colon\ P \to [0,1]$，仿射性给出 $f(x_z^\lambda) = \mathbb{E} f(X_z)$，而有界性给出$|\mathbb{E} f(X_{z'}) - \mathbb{E} f(X_z)| \le \mathrm{TV}(Q_{z'}, Q_z)$。对所有此类 f 取上确界，并应用 Pinsker不等式，可得

$$\left\| x_{z'}^\lambda - x_z^\lambda \right\|_P^2 \le \frac{1-\lambda}{2\lambda} D_H(z', z).$$

最后，我们令 $z' = \pi(B)$ 以及 $z = \pi(A)$，并应用引理 5。 □

F 预解式迹不等式

引理 7 的证明。 我们通过对范数的递归和/最大值构造进行归纳法来论证。对于实数轴，$|u| < 1$ 和 $|v| \le 1$ 意味着 $(1-v)/(1-u) \ge 0$。一元和可以省略。对于归纳步骤，X 是一个 ℓ_1 或 ℓ_∞ 由较小空间组成的和 $X_1, \ldots\ldots, X_b$ 且断言在这些空间上成立。记 $r = \|U\|_{\mathrm{op}} < 1$。固定一个块 i，并将其余块归为 J，采用它们继承的和范数或最大范数。算子上的下标依次指定其输出块和输入块。坐标嵌入和坐标投影的算子范数均为一，因此 $\|U_{JJ}\|_{\mathrm{op}} \le r$。

对于 $W = U$ 或 $W = U_{JJ}$，界 $\left\|W^k\right\|_{\text{算子}} \le r^k$ 意味着矩阵几何级数 $\sum_{k=0}^{\infty} W^k$ 在算子范数下收敛。在其任一侧与 $I - W$ 相乘，可得其部分和

$$(I-W)\sum_{k=0}^{K} W^k = I - W^{K+1} = \left(\sum_{k=0}^{K} W^k\right)(I-W).$$

由于 $W^{K+1} \to 0$ 如 $K \to \infty$，这证明了 Neumann级数公式 $(I-W)^{-1} = \sum_{k=0}^{\infty} W^k$。因此，我们可以定义

$$B = (I-U)^{-1}, \qquad D_i = (I - U_{JJ})^{-1}, \qquad U^{[i]} = U_{ii} + U_{iJ} D_i U_{Ji}.$$

首先在 $(I-U)(x,\ y) = (h,\ k)$ 中消去 J-块，得到 $y = D_i(k + U_{Ji}x)$，进而得到 $(I - U^{[i]})x = h + U_{iJ} D_i k$。读出解算子可得

$$B_{ii} = (I - U^{[i]})^{-1}, \quad B_{iJ} = B_{ii} U_{iJ} D_i, \quad B_{Ji} = D_i U_{Ji} B_{ii}. \tag{17}$$

我们首先证明 $\left\|U^{[i]}\right\|_{\text{算子}} \le r$。对于 $x \in X_i$，令 $y = D_i U_{Ji} x$。那么 $U(x,\ y) = (U^{[i]}x,\ y)$。在一个 ℓ_1 求和中，

$$\left\|U^{[i]}x\right\| + \|y\| \le r(\|x\| + \|y\|),$$

由此即可得到结论。在 ℓ_∞ 和中，压缩映射 $y \mapsto U_{Ji}x + U_{JJ}y$ 将半径为 $\|x\|$ 的球映射到自身，因为完整输入（$x,\ y$）的范数至多为 $\|x\|$。其不动点满足 $\|y\| \le \|x\|$，而 U 的第 i 个块行给出 $\left\|U^{[i]}x\right\| \le r\,\|x\|$。

对于 ℓ_1 和，令

$$V^{[i]} = V_{ii} + U_{iJ} D_i V_{Ji}.$$

我们将 U 的第 i 个块列替换为 V 的第 i 个块列，得到 Z。每个块列都是从其输入块到 X 的压缩映射。事实上，对于 $z = (z_k)_k$，由三角不等式可得 $\|Zz\| \le \sum_k \|Z_{\cdot k} z_k\| \le \sum_k \|z_k\| = \|z\|$。取 $y = D_i V_{Ji} x$ 可得 $Z(x,\ y) = (V^{[i]}x,\ y)$，因此

$$\left\|V^{[i]}x\right\| + \|y\| \le \|x\| + \|y\|\,.$$

因此 $\left\|V^{[i]}\right\|_{\text{op}} \le 1$。公式（17）给出

$$[B(I - V)]_{ii} = (I - U^{[i]})^{-1}(I - V^{[i]}).$$

根据归纳法，其迹非负。对角块的迹求和即可证明 ℓ_1 情形下的结论。

对于 ℓ_∞ 和式，则改为定义

$$V^{[i]} = V_{ii} + V_{iJ} D_i U_{Ji}.$$

调和延拓 $y = D_i U_{Ji} x$ 的范数至多为 $\|x\|$，因此应用 V 的块行 i 即可证明 $\left\|V^{[i]}x\right\| \le \|x\|$。根据循环性，$\mathrm{tr}(B(I - V)) = \mathrm{tr}((I - V)B)$。后一个乘积的第 i 个对角块为

$$[(I - V)B]_{ii} = (I - V^{[i]})(I - U^{[i]})^{-1}.$$

其迹等于 X_i 上的归纳所得非负表达式。求和即可证明结论。不存在依赖于求和/取最大值层数的因子。 □

F.1 引理 8 的证明

证明。 我们固定 $\alpha,\ \beta > 0$ 并定义

$$\widetilde{A}(\alpha,\beta) = \frac{\alpha A^{+} + \beta A^{-}}{\alpha + \beta}, \qquad L(\alpha,\beta) = (\alpha + \beta)\Big(I - \widetilde{A}(\alpha,\beta)\Big).$$

矩阵 $A^{\pm}$ 可行且非负。由于 $A^{+} + A^{-} = 2A > 0$ 逐元素为正，且两个~系数均为正数，它们的凸组合 $A(\alpha,\ \beta)$ 严格为正。它是可行的，

因此命题 2 使得 1 成为其简单的佩龙特征值。因此 $L(\alpha,\ \beta)$ 有一个简单的零特征值，并且

$$a(\alpha,\beta)=\operatorname{tr}(\operatorname{adj} L(\alpha,\beta))\neq 0.$$

我们首先在不对奇异矩阵 $L(\alpha,\beta)$ 求导的情况下，确定其偏导数的符号。

对于 $\varepsilon>0$,

$$L(\alpha,\beta)+\varepsilon I=(\alpha+\beta+\varepsilon)\left(I-\frac{\alpha+\beta}{\alpha+\beta+\varepsilon}\widetilde{A}(\alpha,\beta)\right).$$

~标量 $(\alpha+\beta)/(\alpha+\beta+\varepsilon)$ 属于 $(0,1)$。由于 $A(\alpha,\beta)$ 为正数，其佩龙特征值为 1 ，且其他所有特征值的模都严格小于 1。因此，每个实因子$1-\frac{\alpha+\beta}{\alpha+\beta+\varepsilon}\lambda_j$ 均为正数，而每一对非实共轭特征值的乘积为 $\left|1-\frac{\alpha+\beta}{\alpha+\beta+\varepsilon}\lambda_j\right|^2>0$。因此 $\det\Big(L(\alpha,\beta)+\varepsilon I\Big)>0$。由于每个可行偏差矩阵在树复形范数下都是非扩张的，将引理 7 应用于

$$U=\frac{\alpha+\beta}{\alpha+\beta+\varepsilon}\widetilde{A}(\alpha,\beta),\qquad V=A^{+}$$

得到

$$\operatorname{tr}\left[\left(I-\frac{\alpha+\beta}{\alpha+\beta+\varepsilon}\widetilde{A}(\alpha,\beta)\right)^{-1}(I-A^{+})\right]\geq 0.$$

固定 β 和 ε 后，$\partial_\alpha(L+\varepsilon I)=I-A^{+}$。雅可比公式给出 $\partial_\alpha \log\det(L+\varepsilon I)=\operatorname{tr}((L+\varepsilon I)^{-1}(I-A^{+}))$。代入逆算子并利用上述迹界，可得

$$\frac{\partial}{\partial\alpha}\log\det\Big(L(\alpha,\beta)+\varepsilon I\Big)=\frac{1}{\alpha+\beta+\varepsilon}\operatorname{tr}\left[\left(I-\frac{\alpha+\beta}{\alpha+\beta+\varepsilon}\widetilde{A}(\alpha,\beta)\right)^{-1}(I-A^{+})\right]\geq 0. \quad (18)$$

采用相同的论证并结合 A^{-} 可得

$$\frac{\partial}{\partial\beta}\log\det\Big(L(\alpha,\beta)+\varepsilon I\Big)\geq 0. \quad (19)$$

对于每个固定的 $\varepsilon>0$，行列式为正数，因此这两个不等式证明了$\varepsilon^{-1}\det(L+\varepsilon I)$ 关于每个系数都是非递减的。行列式是关于 ε 的多项式。其常数项为零，而 ε 的系数是主（$m-1$）-子式之和，等于 $\operatorname{tr}(\operatorname{adj} L)$。因此

$$\lim_{\varepsilon\to 0^{+}}\frac{\det(L+\varepsilon I)}{\varepsilon}=a(\alpha,\beta).$$

该极限非负且不会为零，因为零特征值是单重的。对于任意两个正的 α 值，在 β 固定时，正则化行列式的序关系会传递到极限。这证明了 a 关于 α 的单调性；固定 α 则可得其关于 β 的单调性。由于 a 是多项式，其两个偏导数在整个开正象限内均非负。

□

对和/最大值范数结构的依赖。引理中列拼接和调和延拓的步骤7 严格使用递归和/最大值范数。在任意范数下共同满足压缩性并不足够。例如，

$$U = \begin{pmatrix} 1/20 & 3/4 \\ 39/40 & 1/20 \end{pmatrix}, \qquad V = \begin{pmatrix} 97/100 & 591/10000 \\ 0 & 97/100 \end{pmatrix}$$

均为逐元素非负的欧几里得压缩映射。要检验它们的范数， $I - U^\top U$ 的左上角元素和行列式分别为 $3/64$ 和 $8289/640000$。对于 $I - V^\top V$，它们分别为$591/10000$ 和0。第一个对称矩阵是正定的，第二个是半正定的，因此对于诱导的欧几里得算子范数，有 $\|U\|_2 < 1$ 和 $\|V\|_2 \le 1$。直接求逆得到$\mathbf{tr}(I-U)^{-1}(I-V)) = -249/68500$。因此，树丛证明使用了附录$\ell_1/\ell_\infty$ 中确立的递归C结构。

G 从分母单调性到不动点稳定性

定理9的证明。当 $m = 1$时， $w^\top s = 1$ 固定唯一的坐标，策略移动为零。以下设 $m \ge 2$。在提升中固定 $A > 0$，并取满足其齐次仿射约束的非零矩阵 V。令

$$\eta = \max_{a,b} \frac{|V_{ab}|}{A_{ab}}, \qquad A^\pm = A \pm V/\eta.$$

A^+ 和 A^- 均为可行且非负。对于 α， $\beta > 0$，定义

$$L(\alpha,\beta) = \alpha(I - A^+) + \beta(I - A^-), \qquad \widetilde{A}(\alpha,\beta) = \frac{\alpha A^+ + \beta A^-}{\alpha + \beta}.$$

矩阵 $\widetilde{A}(\alpha,\beta)$ 为正数矩阵。记其分别经过归一化的右 Perron 向量和左 Perron 向量为 $s_{\alpha,\beta}$ 和 $l_{\alpha,\beta}$，其中归一化条件为

$$w^\top s_{\alpha,\beta} = 1, \qquad l_{\alpha,\beta}^\top s_{\alpha,\beta} = 1.$$

对于 $s \in P$，每个 $\widetilde{A}^k s$ 都是可行的，并且具有单位 w-质量。佩龙–弗罗贝尼乌斯收敛给出 $\lim_{k\to\infty} \widetilde{A}^k s = s_{\alpha,\beta}(l_{\alpha,\beta}^\top s)$。对该极限取 w-质量即可证明 $l_{\alpha,\beta}^\top s = 1$，特别地，对于 $s = s^0$也成立。该步骤仅在可行策略上使用质量恒等式。

令

$$a(\alpha,\beta) = \operatorname{tr}(\operatorname{adj} L(\alpha,\beta)), \qquad \mathbf{u}(\alpha,\beta) = \operatorname{adj} L(\alpha,\beta) s^0.$$

由于 $L(\alpha,\beta)s_{\alpha,\beta} = (\alpha+\beta)(I - \widetilde{A}(\alpha,\beta))$ 的秩为 $m-1$，$\operatorname{adj} L(\alpha,\beta)$ 的每一列都位于 $\ker L(\alpha,\beta)$ 的张成空间$\{s_{\alpha,\beta}\}$中，且每一行都位于左核的张成空间$\{l_{\alpha,\beta}^\top\}$中。因此，$\operatorname{adj} L(\alpha,\beta)$ 对于某个标量成立。取迹并利用可得 (α,β)，所以

$$s_{\alpha,\beta} = \frac{\mathbf{u}(\alpha,\beta)}{a(\alpha,\beta)}.$$

每个伴随矩阵元素都是一个矩阵的带符号$(m-1)$子式，而该矩阵关于$(\alpha,\ \beta)$是线性且齐次的。因此 a 以及$\mathbf{u}$ 的每个非零坐标都是 $r=m-1$次齐次的。由于 $a>0$在正象限上为正数，上述商沿我们使用的每条线段均可微。引理8 给出

$$\frac{\partial a(\alpha,\beta)}{\partial \alpha}\geq 0,\qquad \frac{\partial a(\alpha,\beta)}{\partial \beta}\geq 0.$$

记

$$\rho(t)=a(1+t,1-t).$$

欧拉恒等式、逐坐标单调性和齐次性给出了以下逐步界

$$\begin{aligned}|\rho'(0)| &\leq \left.\frac{\partial a}{\partial \alpha}\right|_{(1,1)}+\left.\frac{\partial a}{\partial \beta}\right|_{(1,1)}=r\rho(0),\\ \rho(t) &\leq a(1+|t|,1+|t|)=(1+|t|)^r\rho(0)<2\rho(0),\qquad |t|\leq \frac{1}{2r}.\end{aligned}$$

第一行使用了欧拉恒等式以及非负偏导数；第二行先使用单调性，随后使用齐次性和界 $(1+1/(2r))^r<2$。

固定一个仿射测试函数 $P\to[0,1]$，如第 3 节所述。减去 $1/2$ 会保留 P上的所有差值。由于质量泛函在 $w^\top s=1$ 处满足这一点，因此，中心化的仿射测试函数可以表示为一个向量 f，满足

$$|f^\top s|\leq \frac{1}{2}\qquad (s\in P).$$

我们定义 $F(t):=f^\top\mathbf{u}(1+t,\ 1-t)$，这是一个次数至多为 r 的多项式。对于 $|t|\leq 1/(2r)$，f 和 ρ 上的界给出

$$|F(t)|=\rho(t)\,|f^\top s_{1+t,1-t}|\leq \tfrac{1}{2}\rho(t)<\rho(0).$$

对于每个次数至多为 p 的三角多项式 r，伯恩斯坦不等式（QueffélecandZarouf，2019）给出

$$\max_{\theta\in\mathbb{R}}|p'(\theta)|\leq r\max_{\theta\in\mathbb{R}}|p(\theta)|.$$

将其应用于 $G(\theta)=F(\sin\theta/(2r))$。每个幂次 $(\sin\theta)^j$ 的频率绝对值至多为 j，因此 G 的三角次数至多为 r。对 F 的界给出 $\max_\theta|G(\theta)|\leq\rho(0)$，因为 $|\sin\theta|\leq 1$。根据链式法则和伯恩斯坦不等式，

$$\frac{|F'(0)|}{2r}=|G'(0)|\leq r\max_\theta|G(\theta)|\leq r\rho(0).$$

因此 $|F'(0)|\leq 2r^2\rho(0)$。将其与 $|\rho'(0)|\leq r\rho(0)$ 及 $|F(0)|\leq\rho(0)/2$结合可得

$$\left|\left(\frac{F}{\rho}\right)'(0)\right|\leq\frac{|F'(0)|}{\rho(0)}+\frac{|F(0)|\,|\rho'(0)|}{\rho(0)^2}\leq 2r^2+\frac{r}{2}\leq\frac{5}{2}r^2.$$

由于

$$\tilde{A}(1+t,1-t)=A+tV/\eta,$$

对导数重新缩放会使前面的界乘以 η。导数位于 P的方向空间中，因为不动点沿着这条直线始终处于 P 中。因此，对中心化仿射测试函数取上确界，便得到局部敏感度界

$$\left\| \frac{d}{dt} s_{A+tV} \Big|_{t=0} \right\|_P \le \frac{5}{2}(m-1)^2 \max_{a,b} \frac{|V_{ab}|}{A_{ab}} \le \frac{5}{2}(m-1)^2 \left(\sum_{a,b} \frac{V_{ab}^2}{A_{ab}^2} \right)^{1/2}. \tag{20}$$

对于 $0 \le t \le 1$，还需沿 $A_t = (1-t)A + tB$ 积分。 $A = B$ 的情形显然成立。若 $D_R(B,A) \ge 1/16$，则 $\|s_B - s_A\|_P^2 \le 1 \le 16 D_R(B,A)$。否则，令 $g(u) = u - 1 - \log u$。散度中的每一项 $g(B_{ab}/A_{ab})$ 都是非负的。该函数在 1 之前递减，之后递增，且 $g(1/2)\ g(2) > 1/16$。因此，每个比值都落在 $(1/2, 2)$ 中。记 $u_{ab} := B_{ab}/A_{ab}$。由于 $R(A) = -\sum_{a,b} \log A_{ab}$，其梯度的各分量为 $(\nabla R(A))_{ab} = -1/A_{ab}$。因此

$$\begin{aligned} D_R(B,A) &= R(B) - R(A) - \langle \nabla R(A), B - A \rangle \\ &= \sum_{a,b} \left(-\log B_{ab} + \log A_{ab} + \frac{B_{ab} - A_{ab}}{A_{ab}} \right) \\ &= \sum_{a,b} \Big(u_{ab} - 1 - \log u_{ab} \Big) = \sum_{a,b} g(u_{ab}). \end{aligned}$$

对于 $u \in [1/2, 2]$，在 1 处应用泰勒定理，并结合 $g(1) = g'(1) = 0$，可得 $g(u) = \frac{1}{2} g''(\xi)(u-1)^2$，其中某个 ξ 介于 1 和 u 之间。由于 $g''(\xi)$ 在此区间上 $= 1/\xi^2 \ge 1/4$，因此 $g(u) \ge (u-1)^2/8$。将此不等式应用于每个 u_{ab} 并求和，得到

$$\sum_{a,b} (u_{ab} - 1)^2 \le 8 \sum_{a,b} g(u_{ab}) = 8 D_R(B,A).$$

接下来，我们沿着 $A_t = (1-t)A + tB$界定每个条目的相对变化。对于每个$t \in [0,1]$，由界限 $u_{ab} \ge 1/2$ 可得

$$(A_t)_{ab} = A_{ab}(1 - t + t u_{ab}) \ge A_{ab}(1 - t/2) \ge A_{ab}/2.$$

因此，对于所有 t，均有

$$\begin{aligned} \sum_{a,b} \frac{(B_{ab} - A_{ab})^2}{(A_t)_{ab}^2} &\le 4 \sum_{a,b} \frac{(B_{ab} - A_{ab})^2}{A_{ab}^2} \\ &= 4 \sum_{a,b} (u_{ab} - 1)^2 \le 32 D_R(B,A). \end{aligned}$$

每个 A_t 都是正数且可行，该线段的方向为 $B - A$。沿此方向在20处应用（A_t，得到

$$\left\| \frac{d}{dt} s_{A_t} \right\|_P \le \tfrac{5}{2}(m-1)^2 \left(\sum_{a,b} \frac{(B_{ab} - A_{ab})^2}{(A_t)_{ab}^2} \right)^{1/2}$$

$$\leq \tfrac{5}{2}(m-1)^2\sqrt{32D_R(B,A)}.$$

由于 $A_0 = A$ 和 $A_1 = B$，由微积分基本定理和三角不等式可得

$$\begin{aligned}\|s_B - s_A\|_P &= \left\| \int_0^1 \frac{d}{dt} s_{A_t}\, dt \right\|_P \\ &\leq \int_0^1 \left\| \frac{d}{dt} s_{A_t} \right\|_P dt \leq \tfrac{5}{2}(m-1)^2\sqrt{32D_R(B,A)}.\end{aligned}$$

将两边平方并利用 $(5/2)^2 \cdot 32 = 200$ ，可得

$$\|s_B - s_A\|_P^2 \leq 200(m-1)^4 D_R(B,A) \leq 200m^4 D_R(B,A).$$

□

H 从不动点稳定性到自博弈遗憾

H.1 得分与近端形式的等价性

固定一个玩家并省略其索引。 $\mathcal{Z}$ 的仿射包由提升等式给出。元素对数障碍在其相对边界处趋于无穷，并且在切向方向上具有正定海森矩阵。因此，算法 1中的每个障碍加线性目标函数都有唯一的¯正极小化解。在 A^t 其方向导数在整个切空间上均为零。因此，在 $\mathcal{Z}$ 上，函数

$$A \longmapsto \gamma R(A) + \sum_{\tau<t} c^\tau(A) - \gamma D_R(A, \bar{A}^t)$$

¯为常数：其线性部分为 $\gamma\nabla R(A^t) + \sum_{\tau<t} c^\tau$，它对所有可行差均为零。因此，无论加上预测还是实际成本，都会得到

$$\begin{aligned}A^t &= \arg\min_{A\in\mathcal{Z},\, A>0} \{c^{t-1}(A) + \gamma D_R(A, \bar{A}^t)\}, \\ \bar{A}^{t+1} &= \arg\min_{A\in\mathcal{Z},\, A>0} \{c^t(A) + \gamma D_R(A, \bar{A}^t)\}.\end{aligned}$$

¯两种更新都从 A^t 开始。第一种在观察第 t 轮的损失之前决定行动。第二种是在观察到该损失后用于分析的领导者。在 $t = 1$, $\bar{A}^1 = A^1 = A_c$ 因为$c^0 = 0$。

H.2 有向散度与强凸性

对于线性代价 g和正数 U，令 Y最小化 $g+\gamma D_R(\cdot\, U)$。对于任意正数比较器X，沿着 $X - Y$ 的一阶最优性以及 D_R 的定义可得

$$g(Y) - g(X) = \gamma[D_R(X,U) - D_R(X,Y) - D_R(Y,U)].$$

先将此恒等映射应用于 $(g, U, Y, X) = (c^t, A^t, A^{t+1}, C)$，再应用于 $(c^{t-1}, A^t, A^t, A^{t+1})$。将这两个恒等式相加，并加上 $A^t - A^{t+1}$ 上实际代价与预测代价之差。$D_R(A^{t+1}, A^t)$项相互抵消，得到

$$\begin{aligned} c^t(A^t) - c^t(C) = \gamma[D_R(C, \bar{A}^t) - D_R(C, \bar{A}^{t+1})] \\ - \gamma D_R(A^t, \bar{A}^t) - \gamma D_R(\bar{A}^{t+1}, A^t) \\ + (c^t - c^{t-1})(A^t - \bar{A}^{t+1}). \end{aligned}$$

Here 矩阵差上的线性代价表示两个代价之差 .

在此证明中，记 $\|V\|_{\mathcal{Z}-\mathcal{Z}} :=$ 下确界$\{r > 0 : \ V \in r(\mathcal{Z} - \mathcal{Z})\}$ 定义在可行差空间上。这与附录 B 中用于策略的差体范数相同，只是现在定义在提升空间上。若 $V \neq 0$ 在正数 A处于切空间中，令

$$h_A(V) := \left(\sum_{a,b} V_{ab}^2 / A_{ab}^2 \right)^{1/2}.$$

两者 $A \pm V/h_A(V)$ 均满足提升等式和非负性。它们的差为 $2V/h_A(V)$，因此 $2\|V\|_{\mathcal{Z}-\mathcal{Z}} \leq h_A(V)$。对于 $V = 0$，同样的不等式显然成立。对正数 A，B 令 $V = B - A$ 并使用积分余项公式可得

$$D_R(B, A) = \int_0^1 (1-u)\, h_{A+uV}(V)^2 \, du \geq 2\|B - A\|_{\mathcal{Z}-\mathcal{Z}}^2.$$

特别地，下面较弱的界 $\|B - A\|_{\mathcal{Z}-\mathcal{Z}} \leq \sqrt{2D_R(B, A)}$ 已经足够。Δ^t 的定义意味着，对于 U，$V \in \mathcal{Z}$，有 $|(c^t - c^{t-1})(U - V)| \leq \Delta^t$。通过缩放可得，在切空间上 $|(c^t - c^{t-1})(H)| \leq \Delta^t \|H\|_{\mathcal{Z}-\mathcal{Z}}$。取 $H = A^t - \bar{A}^{t+1}$ 应用杨氏不等式：

$$|(c^t - c^{t-1})(H)| \leq \Delta^t \sqrt{2D_R(\bar{A}^{t+1}, A^t)} \leq \frac{(\Delta^t)^2}{\gamma} + \frac{\gamma}{2} D_R(\bar{A}^{t+1}, A^t).$$

对 t 求和后，比较器散度发生望远镜消去。最后一项 $-\gamma D_R(C, \bar{A}^{T+1})$ 非正。两个移动散度的系数均至多为 $-\gamma/2$，这证明了（5）。

H.3 损失变化与共同速率定理

恢复玩家索引。对于任意两个比较器映射 B，C 和 $t \geq 2$，偏差成本的线性性给出

$$\begin{aligned} &(c_i^t - c_i^{t-1})(B - C) \\ &= (1-\lambda)\Big[\big\langle \ell_i^t - \ell_i^{t-1}, (B-C)x_i^t \big\rangle + \big\langle \ell_i^{t-1}, (B-C)(x_i^t - x_i^{t-1}) \big\rangle\Big]. \end{aligned}$$

向量 $(B - C)x_i^t$ 是两个可行策略之差，因此其策略范数至多为一。记 $v_j^t = \|x_j^t - x_j^{t-1}\|_j$。由于 B 和 C 是非扩张的，并且 $\|\ell_i^{t-1}\|_{i,\,*} \leq 2$，

第二个内积的绝对值至多为 $4v_i^t$。根据（1），将和平方并利用$(a+b)^2 \le 2a^2+2b^2$ 可得

$$(\Delta_i^t)^2 \le 32(n-1)\sum_{j\neq i}(v_j^t)^2 + 32(v_i^t)^2.$$

对 i 求和时，每个 $(v_j^t)^2$ 都会在对手求和中出现 $n-1$ 次。在 $t=1$ 处，$c_i^0=0$ 和 $\Delta_i^1 \le 2$。因此

$$\sum_{i,t}(\Delta_i^t)^2 \le 4n + 32[1+(n-1)^2]\sum_i V_i \le 4n + 32n^2\sum_i V_i.$$

对于引理 3的第二部分，将 $\bar{A}_i^t$ 的不动点插入 x_i^{t-1}与 x_i^t之间。由平方三角不等式和假设的有向稳定性可得

$$(v_i^t)^2 \le 2C_*\Big[D_{R_i}(\bar{A}_i^t, A_i^{t-1}) + D_{R_i}(A_i^t, \bar{A}_i^t)\Big].$$

第一个散度是 $t-1$ 处的校正项，第二个是 t 处的预测项。从 2 到 T 求和时，这些项各自至多使用一次，得到 $V_i \le 2C_*E_i$。

为证明定理 4，任取比较器 $C \in \mathcal{Z}_i$。由于 $x_i^t = (1-\lambda)A_i^t x_i^t + \lambda q_i^0$，

$$\left\langle \ell_i^t, x_i^t - Cx_i^t \right\rangle = c_i^t(A_i^t) - c_i^t(C) + \lambda\left\langle \ell_i^t, q_i^0 - Cx_i^t \right\rangle.$$

最后一个内积的上界为二。为允许比较器位于边界上，令 $\varepsilon = 1/(T+1)$ 且 $C_\varepsilon = (1-\varepsilon)C + \varepsilon A_{c,\,i}$。$A_{c,\,i}$ 的正性和平稳性给出

$$D_{R_i}(C_\varepsilon, A_{c,i}) = R_i(C_\varepsilon) - R_i(A_{c,i}) \le m_i^2\log(1/\varepsilon),$$

因为$(C_\varepsilon)_{ab} \ge \varepsilon(A_{c,\,i})_{ab}$ 对每个分量都成立。此外，$|c_i^t(C_\varepsilon) - c_i^t(C)| \le 2(1-\lambda)\varepsilon$，因此整个平滑代价至多为二。将 (5) 应用于 C_ε 并使用前述重置恒等式。这些估计对 C一致。命题 2 提升了每位玩家的最优事后偏差，因此对相应不等式求和可得

$$\sum_i \mathrm{LinReg}_i(T) \le \gamma\left(\sum_i m_i^2\right)\log(T+1) + 2n(\lambda T+1) + \frac{4n}{\gamma} + \left(\frac{64n^2C_*}{\gamma} - \frac{\gamma}{2}\right)\sum_i E_i.$$

当 $\gamma^2 \ge 128n^2C_*$时，移动系数非正，从而证明了 (6)。在每位玩家的事后最小化问题中选择恒等映射作为候选，表明最小值至多为该玩家已承受的损失。因此，每位玩家的遗憾均非负，遗憾总和的上界也同时约束每位玩家的个体遗憾。

对于一般多面体调度，$a_T = \lceil (T+1)^{1/3}\rceil$，$\lambda = a_T^{-2}$和 $C_* \le 2(p+1)^2a_T^2$。代入可得

$$\sum_i \mathrm{LinReg}_i(T) \le 32n(p+1)a_T\left(\sum_i m_i^2\right)\log(T+1) + \frac{1}{8(p+1)a_T} + 2n(T/a_T^2+1).$$

a_T和 T/a_T^2均为$\mathcal{O}((T+1)^{1/3})$，从而得到所声称的时域依赖关系。在扩展式博弈模式下，$\lambda=0$、$C_* = 200M^4$和 $\gamma = 320nM^2$。使用 $\sum_i m_i^2 \le nM^2$可得到 $320n^2M^4\log(T+1) + 1/(80M^2) + 2n$ 的矩阵提升。

I 确定性有限精度实现

本节证明显式对数学习器在两种模式下的位数结论。计算原语包括在有理正提升上的弱凸优化，以及针对不动点系统的有理线性代数。输入多面体和博弈描述均为有理的。全信息反馈过程会以请求的多项式位精度返回自身的损失系数；在显式有理扩展式博弈中，这些系数通过在当前有理策略组合处对终端贡献求和得到。

I.1 正坐标下界与确定性优化

我们固定一个有理正可行锚点 B_0。在 $\gamma R + b$ 的极小值点 A 处，沿着 $B_0 - A$ 的最优性给出

$$\gamma \left(m_i^2 - \sum_{ab} \frac{(B_0)_{ab}}{A_{ab}} \right) + b(B_0 - A) = 0.$$

因此，每个坐标都满足

$$A_{ab} \geq \frac{(B_0)_{ab}}{m_i^2 + (\max_{\mathcal{Z}_i} b - \min_{\mathcal{Z}_i} b)/\gamma}.$$

在时域 T 内，由于每个得分至多是 $T+1$ 个单轮损失之和，而在可行矩阵之间，每个损失最多相差 2，因此 $\max_{\mathcal{Z}_i} b - \min_{\mathcal{Z}_i} b \leq 2(T+1)$。对得分进行舍入时，我们将此界保守地增加一个单位。我们将优化限制在各项至少为所得下界一半的范围内。这个有理截断包含相关的精确极小值点，并且在 B_0 周围具有一个半径为多项式编码的相对内点球。仿射等式和坐标上界的位长度也都是多项式级的。在此定义域上，障碍函数梯度是一个有理倒数向量，对数目标函数值可确定性地用有理数包络到任意指定精度。导数的大小以及编码后的外半径与内半径之比可能很大，但它们的对数相对于 D 和 $\log(T+1)$ 都是多项式级的。

因此，借助分离预言机进行确定性弱凸优化，可在 $\mathrm{poly}(D, \log(T+1))$ 位时间内，找到一个正的有理可行点，使其达到任意指定的逆多项式目标差（Grötschel 等人，1988年）。计算在仿射包的一组有理基下进行；通过先向内点锚点取整、再进行坐标取整，可以保留精确可行性。此优化步骤仅取决于有理提升几何和所要求的目标精度。

I.2 有限精度迭代点的遗憾

我们将共同的平方稳定性界记为 C_*，在一般情况下还包括其对 T 的依赖。两种选择（9）和（13）均满足 $\gamma^2 \geq 512 n^2 C_*$。我们设

$$\delta_T = \frac{1}{4096 n^2 (1 + C_*)(T+1)^2}. \tag{21}$$

我们用 $\widehat{A}_i^t \in \mathcal{Z}_i$ 表示有限精度优化程序实际返回的正可行矩阵，并令 $\widehat{x}_i^t := x^{\lambda}_{\widehat{A}_i^t}$ 为其精确不动点。在分析中，我们根据这些实际实施的策略组合所产生的真实成本，定义精确预测引导项 A_i^t 和校正引导项 A_i^t：

$$c_i^t(A) = (1-\lambda)\left\langle \ell_i(\widehat{x}_{-i}^t), A\widehat{x}_i^t \right\rangle.$$

因此 A_i^t 和 A_i^t 是在实际实施的反馈序列上计算的、与第5节中的预测更新和校正更新完全对应的项。记

$$x_i^{*,t} := x^{\lambda}_{A_i^t}, \qquad \widehat{V}_i := \sum_{t=2}^{T} \left\| \widehat{x}_i^t - \widehat{x}_i^{t-1} \right\|_i^2, \qquad V_i^* := \sum_{t=2}^{T} \left\| x_i^{*,t} - x_i^{*,t-1} \right\|_i^2,$$

并令 E_i 表示精确影子引导项的障碍能量，其定义见第 5节。影子引导项用于比较实际实施的迭代点与精确近端解。对于预测近端算子或校正近端算子，我们用 b 表示精确线性评分，用 b 表示其舍入后的实现。对于评分误差 $e := b - b$，我们要求

$$\max_{\mathcal{Z}_i} e - \min_{\mathcal{Z}_i} e \le \frac{\gamma\delta_T}{2}, \qquad (\gamma R_i + \widehat{b})(\widehat{A}) - \min_{A\in\mathcal{Z}_i,\, A\succ 0} (\gamma R_i + \widehat{b})(A) \le \frac{\gamma\delta_T}{2}.$$

若 A^* 表示评分 b对应的精确极小值点，则将这两个界合并可得

$$\begin{aligned}(\gamma R_i + b)(\widehat{A}) - (\gamma R_i + b)(A^*) &\le \max_{\mathcal{Z}_i} e - \min_{\mathcal{Z}_i} e + \frac{\gamma\delta_T}{2} \\ &\le \gamma\delta_T.\end{aligned}$$

第一行将评分舍入误差与优化误差分开；第二行则对两者各采用规定的一半预算。正极小化解 A^* 处的一阶条件使得左侧恰好等于 $\gamma D_{R_i}(A,\ A^*)$。将其应用于预测领导者，可得

$$D_{R_i}(\widehat{A}_i^t, A_i^t) \le \delta_T.$$

舍入后的矩阵仍为正且可行。其重置不动点可通过有理线性代数求得。对于 $\lambda = 0$，我们求解带质量归一化的有理零空间方程；佩龙特征值是单根。对于 $\lambda > 0$，压缩线性系统非奇异。行列式比特长度界保证存在编码长度为多项式的有理数解，即使中间计算中的条件数很大也不例外。

稳定性给出 $\left\| \widehat{x}_i^t - x^{*,\ t}_{\ i} \right\|_i^2 \le C_*\delta_T$。因此，实际路径长度满足

$$\widehat{V}_i \le 3V_i^* + 6TC_*\delta_T, \qquad V_i^* \le 2C_*E_i.$$

实际策略组合变化界现在给出的能量系数至多为

$$\frac{192n^2C_*}{\gamma} - \frac{\gamma}{2} \le -\frac{\gamma}{8}.$$

额外变动成本至多为 $192n^3TC_*\delta_T/\gamma$。上述强凸性估计以及 $D_{R_i}(\widehat{A}_i^t, A_i^t) \le \delta_T$意味着

$$\inf\{r>0 : \widehat{A}_i^t - A_i^t \in r(\mathcal{Z}_i - \mathcal{Z}_i)\} \le \sqrt{2\delta_T}.$$

因此，近似戈登恒等式至多贡献

$$\sum_{i=1}^{n}\sum_{t=1}^{T} 2\inf\{r>0 : \widehat{A}_i^t - A_i^t \in r(\mathcal{Z}_i - \mathcal{Z}_i)\} \le 2nT\sqrt{2\delta_T}.$$

事实上，实现策略是其实现映射的精确不动点，因此差异为 $(1-\lambda)\ell_i^{t\top}(\widehat{A}_i^t - A_i^t)\widehat{x}_i^t$。我们将这两个误差合并为

$$2nT\sqrt{2\delta_T} + \frac{192n^3TC_*\delta_T}{\gamma} < 1,$$

其中，严格界由 (21) 给出，$C_* \ge 1$，以及 $\gamma \ge n$。(6) 中的比较器项和重置项保持不变。

最后，得分舍入和近似反馈可以满足所述误差预算。紧致性和附录B 为每个矩阵坐标给出了一个有理数编码的界。如果每个成本系数的扰动至多为 ξ，则任意两个可行矩阵之间成本差异所产生的误差至多为

$$2m_i^2\xi \max_{A\in\mathcal{Z}_i}\max_{a,b} A_{ab}.$$

将每轮误差设为关于 T、C_* 的逆多项式，并结合该编码界，即可得到所需的总预算。它们的比特长度关于 D、$\log(T+1)$ 为多项式级，因为 λ^{-1}和 C_* 关于 T 及紧致描述均为多项式级。这证明了确定性实现的结论及其实际遗憾保证。

J 未知时域与一般性边界

对于一般多面体模式，预期长度为 1、2、4、.....的公共时期可保持$\widetilde{\mathcal{O}}(\mathrm{poly}(D)T^{1/3})$ 的速率。时域为H 的实例所对应的界也适用于每个前缀，因为平滑、参数和误差预算都是针对 H选定的。对于任意固定的线性偏离，累积遗憾至多为逐时期比较器最大值之和。若 s 是最后一个活跃时期，则几何级数 $\sum_{k\le s} 2^{k/3} = \mathcal{O}(T^{1/3})$ 和 $\sum_{k\le s}(k+1)2^{k/3} = \mathcal{O}(T^{1/3}\log(T+1))$ 可证明该结论。

纯对数型扩展式博弈算法使用与时域无关的 γ，且无需重置，因此其精确的对数定理无需重启即可成立。其所述的有限比特实现指定了时域。随时可用的有限比特实现可以采用可求和的时变精度预算，并在每一轮重新优化累积的有理数得分。

一般定理适用于具有多项式规模描述的任意紧致 H 多面体，而扩展式博弈定理适用于任意深度的完美回忆树丛，

包括不完全信息博弈。扩展式博弈更优的收敛速率源于未重置稳定性背后的和范数/最大范数恒等式。对于一般的 H-多面体，引理 5 则将再生估计转移到显式障碍函数上，同时保留逆重置代价以及由此产生的 $T^{1/3}$ 平衡。

K UTC 实现扩展式博弈算法

本附录证明推论 10。我们给出博弈与查询流约束，证明诱导映射涵盖所有线性偏离，并将局部树丛稳定性估计迁移至 UTC 对数障碍。由此得到附录 L 中使用的坐标更新。

K.1 非负坐标表示下的稳定性

该迁移使用了已在（20）中证明的局部相对扰动估计；它并不比较两个不同的全局矩阵障碍散度。

引理 12（非负线性表示下的稳定性）。 我们考虑一个树丛P，其 m 坐标由终端序列索引，以及一个紧致多面体 $Z = \{z \in \mathbb{R}^q_+ : Bz = b_U\}$，该多面体包含一个严格正点。线性映射 $C: \mathbb{R}^q \to \mathbb{R}^{m\times m}$ 满足$C(Z) \subseteq Z(P)$；$C(z)$ 的每个条目都是 z坐标的非负线性组合，并且当 $z > 0$ 时有 $C(z) > 0$。令 $\widehat{R}(z) := -\sum_{e=1}^q$ 对数 z_e，则每个正的 z、$z' \in Z$ 满足

$$\left\| s_{C(z')} - s_{C(z)} \right\|_P^2 \le 200 m^4 D_{\widehat{R}}(z', z).$$

如果映射 $s \mapsto C(z)s$、$z \in Z$涵盖了 P 的所有仿射自映射，那么通用遗憾分析适用于 Z，并在比较器半径项中以 q 替代 m^2。

证明。 单元素情形显然成立，因此我们假设 $m \ge 2$。记 $C(z)_{ab} = \sum_{e=1}^q c_{ab,e} z_e$，其中所有系数 $c_{ab,e}$ 均为非负数。对于任意可行方向 h 和正数 z，我们将相对坐标估计与（20）结合，得到

$$\frac{|\sum_{e=1}^q c_{ab,e} h_e|}{C(z)_{ab}} \le \frac{\sum_{e=1}^q c_{ab,e}|h_e|}{\sum_{e=1}^q c_{ab,e} z_e} \le \max_{1\le e\le q} \frac{|h_e|}{z_e},$$

$$\left\| \frac{d}{dt} s_{C(z+th)}\Big|_{t=0} \right\|_P \le \frac{5}{2}(m-1)^2 \max_{1\le e\le q} \frac{|h_e|}{z_e}.$$

第一行利用了系数 $c_{ab,e}$ 的非负性，第二行则是将局部矩阵估计应用于 $C(z)$和 $C(h)$。

如果 $D_{\widehat{R}}(z', z) \ge 1/16$，则内在直径界可证明该结论。否则，每个比值$z'_e/z_e$ 都属于区间 $(1/2, 2)$，并且沿该线段有

$$D_{\widehat{R}}(z', z) = \sum_{e=1}^q \left(\frac{z'_e}{z_e} - 1 - \log \frac{z'_e}{z_e} \right),$$

$$\sum_{e=1}^{q}\left(\frac{z'_e}{z_e}-1\right)^2 \le 8D_{\widehat{R}}(z',z),$$

$$\max_{1\le e\le q}\frac{|z'_e-z_e|}{(1-t)z_e+tz'_e} \le 2\left(\sum_{e=1}^{q}\left(\frac{z'_e}{z_e}-1\right)^2\right)^{1/2} \le 2\sqrt{8D_{\widehat{R}}(z',z)},$$

$$\left\|s_{C(z')}-s_{C(z)}\right\|_P^2 \le 200(m-1)^4 D_{\widehat{R}}(z',z) \le 200m^4 D_{\widehat{R}}(z',z).$$

第一行是 UTC 元素对数障碍的坐标公式。第二行再次利用了 $u-1-$ 在 u 上的曲率，其中 $(1/2,\ 2)$；第三行则使用线段下界$(1-t)z_e+tz'_e \ge z_e/2$。最后一行通过对前述导数估计积分并平方得到。在引理最后一句的满射性假设下，每个仿射自映射在 Z中都有一个代表。由于代价仅取决于诱导映射，在 Z 中的两点上取最大值，与在线性偏离上取最大值，所得损失差完全相同，均为式（4）中的损失差。Gordon 恒等式、对数障碍强凸性以及解析中心比较器平滑同样适用于正仿射切片 Z，正如附录 H中一样。其障碍函数包含 q 对数刻度项，因此可进行所述的比较器半径替换。 □

K.2 UTC 流约束及其诱导映射

省略玩家索引。用 $A(z)$ 表示 UTC 表述中的序列对矩阵，用 $H(z)$ 表示其以信息集对为索引的辅助矩阵（Zhang 等人，2024年，式 (2)）。只有 $A(z)$ 会作用于策略。其根行由 $A_{\varnothing,\sigma}=\mathbf{1}_{\{\sigma=\varnothing\}}$ 固定。向量z 由 A 的所有其他元素以及 H的所有元素组成，且均为非负。对于 $\sigma\neq\varnothing$，令 $j(\sigma)$ 表示 σ 最后一个动作所采取的信息集。记 $\mathcal{C}(\sigma)=\{k:\ p(k)=\sigma\}$，并将 $H_{j,j(\varnothing)}$ 视为零。在 (j,σ) 处的流方程为

$$A_{p(j),\sigma}+H_{j,j(\sigma)}=\sum_{a\in\mathcal{A}_j}A_{ja,\sigma}+\sum_{k\in\mathcal{C}(\sigma)}H_{j,k}.$$

其右端列出了偏离者可采取的行动：在 j 处执行一个动作，或在当前查询序列 σ的后继 k 处查询调解者。代入根行后，这些方程的矩阵形式为 $Bz=b_{\mathrm{U}}$。因此，进入障碍函数的坐标数量为

$$q=(|\Sigma|-1)|\Sigma|+|\mathcal{J}|^2.$$

施加流等式后，一旦存在正数可行点，仿射维数就是 $q-$ 秩 B。

为构造这样一个点，按照行动和查询序列总深度递增的顺序访问各对，并将流入质量等分给右侧的所有条目。每个输出条目恰好分配给一对。在 $(j,\ \varnothing)$ 处，仅通过行动动作即可从一个根信息集到达 j。固定 j 后，沿着 σ 进行查询，接下来即可到达 $(j,\ \sigma)$。初始流入质量为 1，因此每条这样的路径都承载正数质量。所以，每一对都会被到达，每个存储坐标也都是正数。同样的深度排序带来了紧致性：在每一对处，每个输出至多是其此前至多两个有界输入之和

所得的有界有理多面体是有界的。通过线性规划最大化其坐标的公共下界，即可等价地找到一个具有多项式编码长度的有理正锚点。

令 S 选取终端序列，并令 T 为附录 C 中的非负重构。特别地，其在终端序列上的行构成恒等映射。于是

$$\mathsf{ST} = I, \qquad \mathsf{TS}x = x \quad (x \in Q), \qquad C(z) = \mathsf{S}A(z)\mathsf{T}.$$

根据 Zhang 等人（2024年，定理 4.1），映射 $x \mapsto A(z)x$ 在 Q 上涵盖所有线性偏离。因此，$C(z)$ 将映射 P 到其自身，保持其仿射包，并表示 P的每个仿射自映射。对于具有决策的玩家，每个终端输出行都是 A的非根行。因此，$C(z)$ 的每个元素都是存储坐标的非负线性组合。对于终端a，b 的恒等行给出 $C(z)_{ab} \geq A(z)_a$，只要 $z > 0$。具有单元素策略集的玩家的遗憾为零。这证明了正在更新的玩家满足引理 12 的所有假设。

如果玩家 i 有 q_i 个已存储的 UTC 坐标，那么精确的 UTC 更新因此满足

$$\sum_{i=1}^{n} \mathrm{LinReg}_i(T) \leq \gamma \left(\sum_{i=1}^{n} q_i \right) \log(T+1) + \frac{4n}{\gamma} + 2n, \qquad \gamma^2 \geq 128n^2(200M^4).$$

终端序列矩阵提升已经给出了 M^4 定理 1中的界。若要得到仅以 M 表示的 UTC 界，首先忽略单动作信息集：其唯一子节点序列的权重与父节点相同，因此合并这些坐标并重新接上后续信息集会得到一个仿射同构，且只会减少终端序列的数量。在由此得到的自身序列树中，将每个 ja 连接到 $p(j)$。每个非叶节点至少有两个子节点，因此 $|\Sigma_i| \leq 2m_i - 1$。此外，$|\Sigma_i| - 1 = \sum_{j \in \mathcal{J}_i} \lfloor \mathcal{A}_j \rfloor \geq 2|\mathcal{J}_i|$。因此$q_i \leq 5m_i^2$，得到 $\mathcal{O}(n^2 M^4 \log(T+1))$，适用于采用式（13）所示调度的 UTC 更新。保留强制行动时，可直接应用 q_i 中展示的界。

K.3 精确 UTC 坐标更新

令 $s^t \in P$ 表示终端坐标中的策略，$x^t = \mathsf{T}s^t \in Q$ 表示其完整实现计划。用 ℓ^t 表示终端坐标中的损失向量。如果反馈是完整坐标向量 g^t，则取 $\ell^t = \mathsf{T}^\top g^t$。设 $c^0 = 0$，并令 c^t 为线性代价 $z \mapsto \langle \ell^t, C(z)s^t \rangle$ 的系数向量。使用完整坐标代价 $\langle g^t, A(z)x^t \rangle$，在可行的 z 上得到的值相同，因为 $\mathsf{T}C(z)s^t = A(z)x^t$。UTC 更新为

$$z^t = \arg\min_{z \in Z,\, z > 0} \left\{ \gamma \widehat{R}(z) + \left\langle \sum_{\tau=1}^{t-1} c^\tau + c^{t-1}, z \right\rangle \right\}, \qquad (I - C(z^t))s^t = 0, \quad w^\top s^t = 1.$$

正性 $C(z^t)$ 和命题2 给出了 P中的唯一归一化解；其重构计划由 $A(z^t)$固定。也可以像实验中那样，使用 $x^t_\varnothing = 1$求解完整系统。具体而言，将 A 的非根块写作 $[v\ G]$。通过重构得到的正可行不动计划，其非根部分为 $y > 0$，并且满足 $Gy = y - v < y$，因为 $v > 0$。在

加权最大范数 $\|u\| = \max_k |u_k|/y_k$中，非负矩阵 G 的算子范数至多为$\max_k (Gy)_k/y_k < 1$。因此，$(I - G)y = v$ 有唯一解。两种实现都需要进行一次障碍函数优化，并对一个有理可行迭代点进行一次有理线性求解。附录I 中的有限精度证明适用于 z，其中包含 q 对数刻度项以及刚刚证明的稳定性估计。

L 实验协议与结果

我们在四人库恩扑克、五人库恩扑克和三人 Leduc 上，实现了推论 1 中的算法 10的 UTC 坐标扩展式博弈专用版本；每位玩家分别具有 81、193 和 457 个序列坐标，未约简的纯计划数量级分别为 10^{12}, 10^{29}, 和 10^{68}。在针对完整线性/UTC 类、采用相同后见之明预言机的匹配时域下，与 UTC-CFR/RM 相比，ORBITS 将遗憾分别降低了 + 、17.1%、25.4% 和 8.4%。本附录报告终点、协议、轨迹、诊断信息和参数。

实验采用附录 K中确立的 UTC 坐标 EFG 算法。本附录说明基准实例、理论更新的数值求解、评估协议、诊断信息和轨迹。

L.1 基准测试与数值实现

实验使用四人五牌库恩扑克、五人六牌库恩扑克和三人单注 Leduc。Kuhn 和 Leduc 是标准基准系列（Kuhn，1950；Southey 等，2005；Zhang 等，2024）；五人 Kuhn 实例是同一系列中规模更大的成员。在基准构造器中，这些博弈分别包含 3,960、58,320 和 4,500 个终局历史，其中 Leduc 聚合了花色等价的随机结果。损失分别是负的净筹码收益除以 6、8 和 14。每位玩家分别拥有 81、193 和 457 个序列坐标，包括根节点。相应的内在维度分别为 40、96 和 228。每个信息集有两个动作，并在相对内部贡献一个独立的策略参数。因此，未约简的纯计划数量为 2^{40}, 2^{96},和 2^{228}。每位玩家的 UTC 障碍坐标数量分别为 8,080、46,272 和 260,376。相应的 UTC 流量等式数量分别为 3,240、18,528 和 104,196。

算法 1的数值实现。 初始 UTC 点是正域的解析中心，通过单独的零得分求解计算得到。随后每一轮都会对附录 K中所述的精确累积得分障碍问题进行数值求解，然后计算相应的未重置不动点。累积得分始终根据实际损失计算；热启动仅影响数值求解器，不会改变所执行的更新。

数值优化与可行性。 对于累积分数 b，缩放目标函数为 $-\sum_e \log z_e + \langle b/\gamma,\ z\rangle$。若 η是其（缩放目标函数的）等式乘子，则对偶松弛量为

$$s = b/\gamma + B^\top \eta > 0, \qquad z_e = 1/s_e.$$

阻尼牛顿步求解对偶方程 $Bz = b_{\mathrm{U}}$。线性系统在四人库恩扑克中采用稀疏分解，在较大实例中采用预条件共轭梯度法。UTC 反向递归得到的对偶势只会将 b 平移一个属于 B行空间的向量，因此原始目标函数只改变一个常数。前一个对偶解可用作热启动，但每个目标函数都由累积的观测成本构成。

行动前，将每个 UTC 节点的流出权重转换为正比例，并通过 UTC 图向前传播流入质量。这样可以得到一个除浮点舍入误差外均可行的正数可行点。随后，在不静默截断策略坐标的情况下，同时检查策略约束和不动点残差。如果数值求解未通过指定的诊断检查，运行将以错误终止。

对于正数原始候选解 z 和对偶松弛量 $s = b/\gamma + B^{\top}\eta > 0$，定义

$$\mathrm{gap}(z, s) := \gamma \sum_{e} \Big(s_e z_e - 1 - \log(s_e z_e) \Big).$$

在原始目标函数单位下，从原始目标函数中减去对偶值可得 $\mathrm{gap}(z,\ s) + \gamma\eta^{\top}(b_{\mathrm{U}} - Bz)$。因此，当 $Bz = b_{\mathrm{U}}$严格成立时，$\mathrm{gap}(z,\ s)$给出了原始目标函数超过其最小值的量的上界。对于原始牛顿点 $z_e = 1/s_e$，无论流残差为何，间隙表达式都为零。这就是我们单独检查可行性的原因。

我们使用 `log1p`评估间隙。停止容差为：缩放间隙 $\mathrm{gap}(z,\ s)$的 $10^{-5}/\gamma$，以及重构前 UTC 流残差绝对值最大值的 10^{-8}。完成前向流重构后，我们使用数值求解得到的对偶松弛量，在重构后的原始点处评估 $\mathrm{gap}(z,\ s)$。我们还记录 UTC 流、序列式流和不动点方程的残差绝对值最大值。

参数与评估协议。 我们将 T 固定为 1，000、100 和 20，并使用逆步长参数

$$\gamma_{\mathrm{K4}} = 3 \times 10^{-5}, \qquad \gamma_{\mathrm{K5}} = 5 \times 10^{-6}, \qquad \gamma_{\mathrm{Leduc}} = 5 \times 10^{-6}.$$

这些步长在评估前固定，不会对报告的轨迹进行搜索。这些参数是经验性选择，并非定理中的充分最坏情形调度。动力学、博弈遍历和求解器均为确定性的；所有轮次都会被记录，且不会根据一次确定性运行推断统计误差条。

比较器与基线。 报告的指标是 最大值$_i$线性遗憾$_i(t)/t$，根据实际采用的策略进行评估。一个独立的 UTC 后向线性优化递归会在事后计算每位玩家的最佳固定线性偏离；它不会重复使用障碍求解器。UTC-CFR/RM+ 在每个 UTC 决策节点使用正部遗憾匹配，并采用不动点对局（Zhang 等人，2024年）。其幂迭代使用 10^{-12} 相邻迭代容差，且设有10，000次迭代上限，并对所得不动点残差进行显式检查。

UTC-CFR/RM+ 基线在每个游戏中均以与 ORBITS 相同的时域进行评估，因此两种方法使用相同的完整线性/UTC 比较器类和相同的轮次数进行比较。

L.2 数值结果与验证检查

表 2：匹配时域下的最大平均完整线性/UTC 遗憾；数值越低越好。

Game	T	ORBITS	UTC-CFR/RM+
四人 Kuhn	1,000	$\mathbf{5.289\times10^{-4}}$	6.383×10^{-4}
五人 Kuhn	100	$\mathbf{3.423\times10^{-3}}$	4.590×10^{-3}
三人 Leduc	20	$\mathbf{4.357\times10^{-2}}$	4.758×10^{-2}

表 2 包含全部三个端点比较。遗憾曲线使用每个记录的轮次，不进行平滑。图 1–4 报告完整轨迹；图 2 还放大显示了四人库恩扑克的后期轮次。

表 3：所有 ORBITS 预测求解和参与者中的最大浮点诊断值。计算间隙是重构后的原始点处的 gap(z, s)，单位为原始目标函数单位。其他各列给出了方程残差的最大绝对值。

Game	计算间隙	UTC 流	策略流	不动点
四人 Kuhn	7.51×10^{-17}	1.33×10^{-15}	1.47×10^{-14}	2.44×10^{-15}
五人 Kuhn	3.66×10^{-17}	2.33×10^{-15}	2.59×10^{-13}	2.55×10^{-15}
三人 Leduc	9.27×10^{-17}	4.44×10^{-16}	1.09×10^{-14}	5.11×10^{-15}

零得分解析中心解会单独检查，并连同其残差一并保存。前向重构可改善原始可行性，同时不改变累积得分的定义。

独立检查。在双动作示例中，设 u 为查询调解者的概率。在解析中心处，对称性使两个即时动作条目均分，并使每项推荐之后的每对动作也均分。七个坐标的乘积与$(1-u)^2u^5$成正比，因此 $u=5/7$，两个即时动作条目为 $1/7$，四个响应条目各为 $5/14$。求解器恢复出的中心在每个坐标上的误差均小于 10^{-10} 。通过六项独立比较，使用通用线性规划求解器检验了玩具实例和四人库恩扑克上的 UTC 后见之明预言机。在三个基准测试中，另行检查了$\mathsf{ST}=I$、可行实现计划的恢复、诱导映射的正性，以及两种坐标表示的一致性。已进行策略组合上的不动点残差单独记录。这些检查属于数值实现诊断；表示与稳定性保证则在附录K中另行证明。

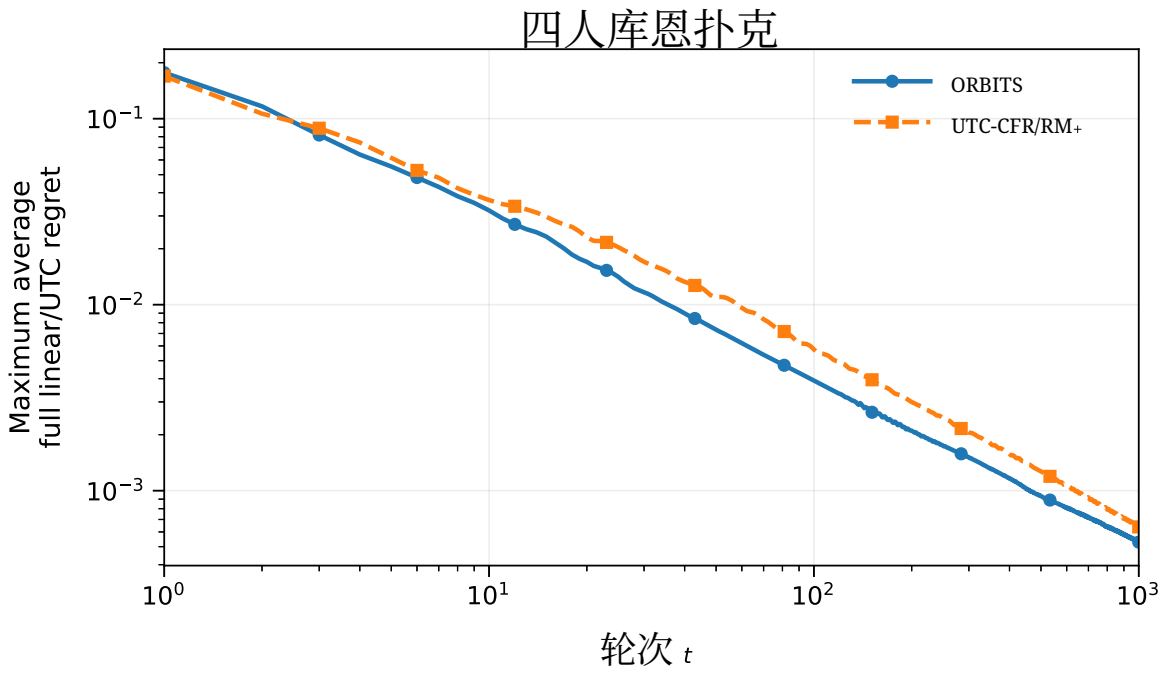

图1：四人库恩扑克，全部 1，000 轮次。两个坐标轴均采用对数刻度；完整线性/UTC 遗憾越低越好。

容差敏感性。附加运行采用相同的步长，并将缩放间隙 10^{-7} 的容差收紧至$\mathrm{gap}(z,\ s)/\gamma$，同时将重构前 UTC 流残差绝对值最大值的容差收紧至 10^{-10} 。比较范围分别涵盖 Kuhn-4、Kuhn-5 和 Leduc 的前100、20和3 轮。表4 比较了这些前缀上的最大平均遗憾轨迹。较长的评估时域未在更严格的容差下重复运行，因此这项检查无法验证它们在完整时域上的数值误差。

表4：短前缀容差检查。最后一列是所示前缀范围内，报告的最大平均遗憾曲线之间的最大绝对差异。

Game	前缀	标准端点	更紧的端点	最大差异
四人 Kuhn	100	3.911×10^{-3}	3.911×10^{-3}	2.38×10^{-8}
五人 Kuhn	20	1.495×10^{-2}	1.495×10^{-2}	6.46×10^{-11}
三人 Leduc	3	1.192×10^{-1}	1.192×10^{-1}	3.72×10^{-11}

计算环境。 实验使用 Python 3.13.5、NumPy 2.3.5、SciPy 1.17.0 和 Numba 0.65.1，运行于配备 Intel Xeon Platinum 8272CL 处理器的主机上，CPU 配额为四核，内存限制为 4 GiB。每次 BLAS 调用最多使用一个线程。ORBITS 在两个 Kuhn 实例上使用一个玩家工作线程，在 Leduc 上使用三个玩家工作线程。独立作业共享该主机。

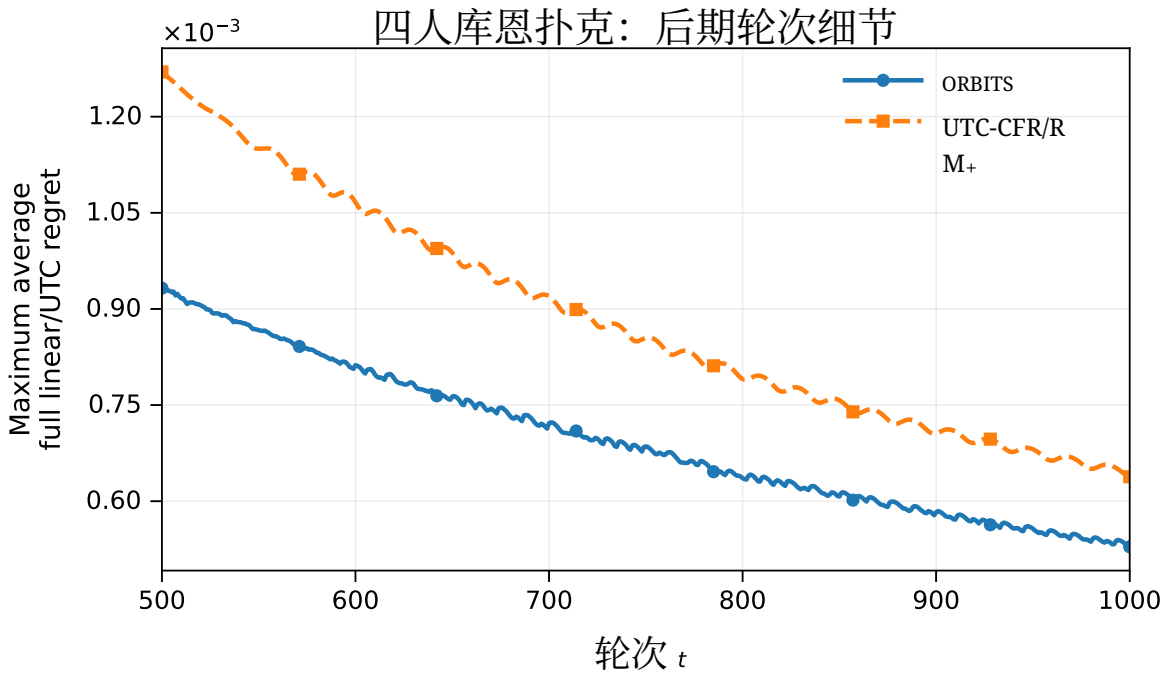

图2：四人库恩扑克，轮次 500–1，000。此后期轮次细节中，两个坐标轴均为线性；在+端点处，ORBITS 仍低于 UTC-CFR/RM。完整轨迹见图 1。

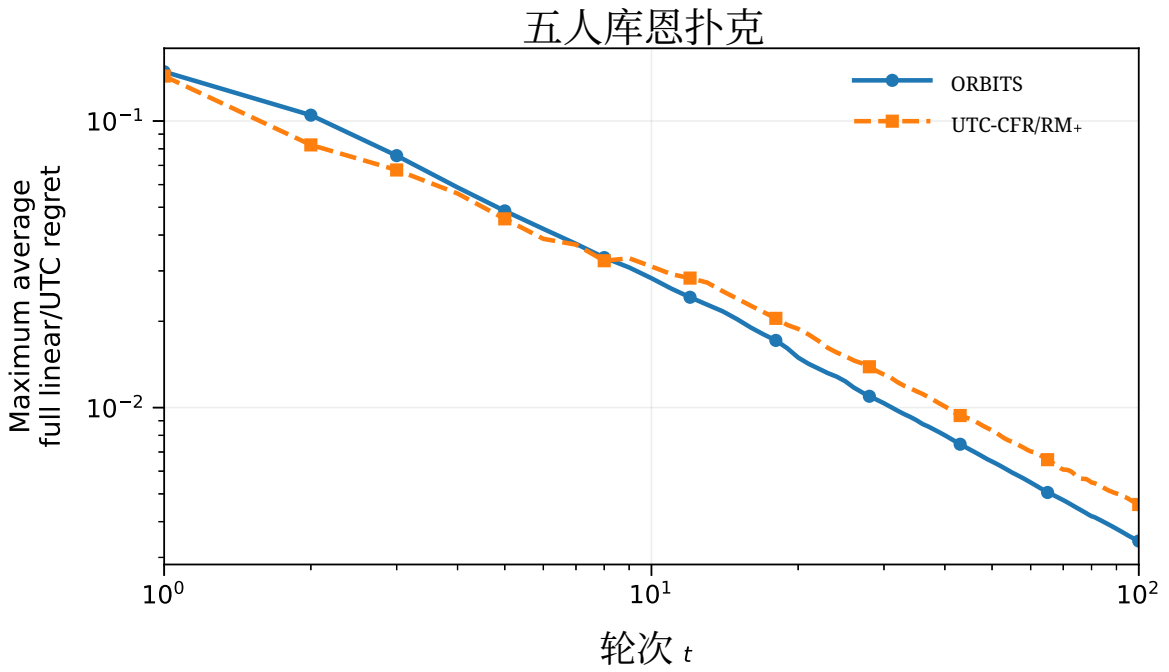

图3：五人库恩扑克，共100轮次。两个坐标轴均采用对数刻度；两种方法均使用相同的完整线性/UTC 比较器类进行评估。

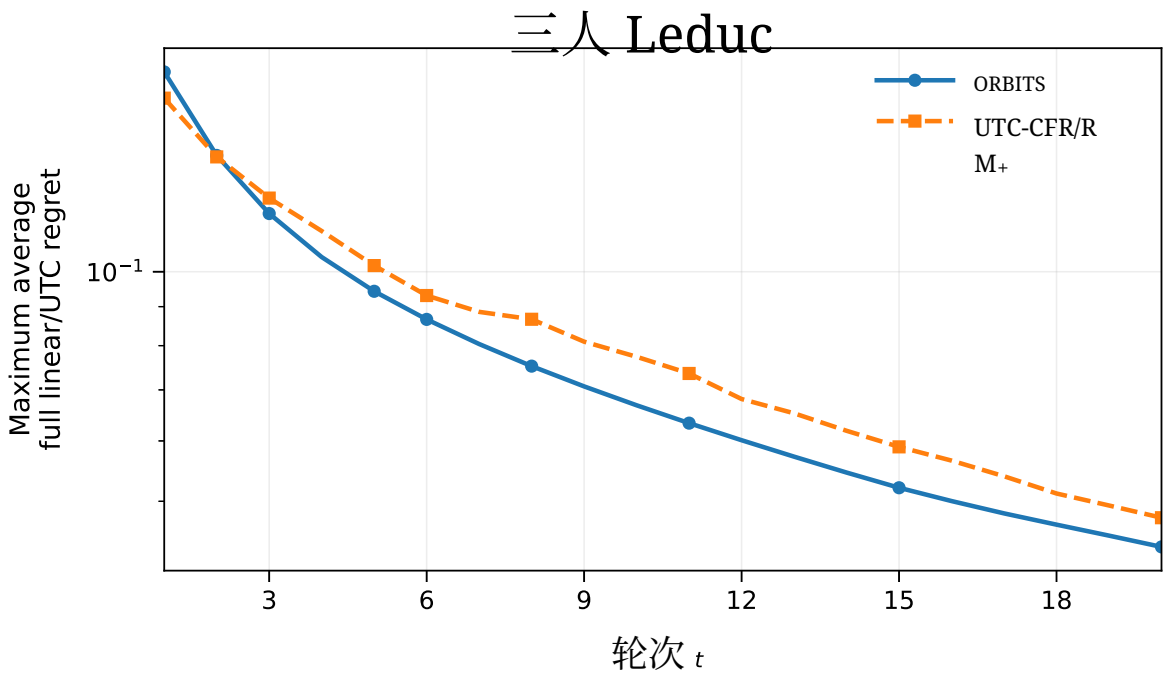

图 4：三人 Leduc，共 20 轮次。轮次轴为线性刻度，遗憾轴为对数刻度。该短期比较并非对渐近遗憾指数的实证检验。