arXiv:2305.02458v1 [math.OC] 3 May 2023

通过相对 Krasnoselskii-Mann 迭代求解不可约随机平均收益博弈和熵博弈

MarianneAkian ✉法国国家信息与自动化研究所（INRIA）和巴黎理工学院数学应用中心（CMAP），巴黎理工学院综合理工学院，巴黎理工学院，法国国家科学研究中心，法国

StéphaneGaubert ✉法国国家信息与自动化研究所（INRIA）和巴黎理工学院数学应用中心（CMAP），巴黎理工学院综合理工学院，巴黎理工学院，法国国家科学研究中心，法国

UlysseNaepels ✉巴黎理工学院综合理工学院，巴黎理工学院，法国

BasileTerver ✉巴黎理工学院综合理工学院，巴黎理工学院，法国

摘要

我们分析了一种求解随机平均收益博弈的算法，该算法结合了相对值迭代与 Krasnoselskii-Mann 阻尼的思想。我们为满足不可约条件的若干类博弈推导出参数化复杂度界。特别地，我们证明，不可约并发随机博弈的博弈值的 ϵ-近似，可以通过 $O(|\log\,\epsilon|)$ 次迭代计算得出，其中 $O(\cdot)$ 中的常数是显式的，且取决于最小的非零转移概率。这一结果可与 Chatterjee 和 Ibsen-Jensen（ICALP 2014）针对同一类博弈得出的 $O(|\epsilon|^{-1}|\log(\epsilon)|)$ 界，以及 Allamigeon、Gaubert、Katz 和 Skomra（ICALP 2022）针对轮流制博弈得出的 $O(|\epsilon|^{-1})$ 界相比较。我们还为熵博弈推导出参数化复杂度界。熵博弈是一类由 Asarin、Cervelle、Degorre、Dima、Horn 和 Kozyakin 提出的矩阵乘法博弈。我们通过变分分析的方法推导出这些结果，并证明相对 Krasnoselskii-Mann 迭代关于希尔伯特半范数具有压缩性质。

2012 年美国计算机协会学科分类 计算理论 → 算法博弈论

关键词与短语 随机平均收益博弈、并发博弈、熵博弈、相对值迭代、Krasnoselskii-Mann 不动点算法、希尔伯特射影度量

1 引言

1.1 动机与背景

随机平均收益博弈是零和博弈中的一个基本类别，存在多种形式。在轮流制博弈中，两名玩家依次进行游戏，轮流行动或选择行动，并知晓另一名玩家先前的决策。具有平均收益且状态空间和行动空间有限的轮流制博弈，是复杂性理论中尚未解决的问题之一：它们属于复杂性类 NP ∩ coNP [14, 44]，但尚不清楚是否存在多项式时间算法。关于不同类别的轮流制博弈的更多信息，请参阅综述 [7]。相比之下，在并发博弈中，两名玩家在每个阶段同时选择一个行动，且不知道对方在同一阶段的选择。轮流制博弈等价于并发博弈的一个子类（在该子类的每个状态中，两名玩家之一是虚设玩家）。Mertens 和 Neyman [30] 的一项著名成果证明了并发随机平均收益博弈的博弈值存在。这一结果建立在 Bewley 和 Kohlberg 早先的研究之上；他们通过令折扣因子趋近于 1，将平均收益并发博弈与折扣并发博弈联系起来，参见 [11]。并发博弈难以精确求解：其博弈值是一个代数数，其

次数可能随着状态数量 [22]呈指数增长。此外，并发可达性博弈是平方根和困难的 [17]。

另一类是熵博弈，由 Asarin、Cervelle、Degorre、Dima、Horn 和 Kozyakin 提出，作为一种有趣的“矩阵乘法博弈” [8]。熵博弈涵盖了多种应用，包括风险敏感控制 [24, 6]，投资组合优化 [4], 增长最大化和种群动力学 [39, 36, 35, 43]。Asarin 等人证明，熵博弈属于 NP ∩ coNP 类，这与轮流制博弈有相似之处。在 [1], 中，Akian、Gaubert、Grand-Clément 和 Guillaud 证明，熵博弈实际上是随机平均收益博弈的特例，其中行动空间是无限集合（单纯形），收益则由 Kullback-Leibler 散度给出。

随机平均收益博弈的一个重要子类，是满足遍历性或不可约性 条件的博弈。这类条件意味着博弈值与初始状态无关。此类条件中最简单的一种要求，两位玩家的每一对策略（位置策略）都会诱导出一个不可约马尔可夫链。此时，求解博弈可归结为求解如下形式的非线性特征值问题 $T(u) = \lambda e + u$：其中， $u \in \mathbb{R}^n$ 是非线性特征向量， λ 是非线性特征值，它给出平均收益博弈的博弈值； e 是 $\mathbb{R}^n$ 的单位向量，而 T 是 $\mathbb{R}^n$ 到自身的自映射，即该博弈的动态规划算子，以下称其为“Shapley”算子。事实上，Shapley 最初提出了这一算子的一个变体，适用于折扣情形 [37]。随后，Gillette 研究了无折扣的平均收益情形 [21]。关于 Shapley 算子和博弈“算子方法”的背景知识，请参阅 [31, 34] ；关于非线性特征值问题的讨论，请参阅 [2] 。

在单人情形中，White [42] 引入了相对值迭代，其过程是在加性常数意义下进行不动点迭代 $\lambda_k \in \mathbb{R}$，即 $x_{k+1} = T(x_k) - \lambda_k e$。在本原性假设下，这可以求解非线性特征值问题 $T(u) = \lambda e + u$。然而，从经典的 Krasnoselkii-Mann 算法 [26, 29],来看，这一假设似乎过于严格；该算法通过构造“阻尼”序列 $x_{k+1} = (1-\theta)T(x_k) + \theta x_k$，其中$0 < \theta < 1$，能够求出有限维赋范空间的非扩张自映射 T 的不动点。事实上，文献 [20]提出将该算法应用于非线性特征值问题 $T(u) = \lambda e + u$，并将其视为商向量空间$\mathbb{R}^n/\mathbb{R}e$中的不动点问题。我们将该算法称为相对 *Krasnoselskii-Mann*值迭代。作为 Baillon 和 Bruck [10], 的一般定理的推论，文献 [20] 推导出了该算法的误差界 $O(1/\sqrt{k})$，并在特殊情形下证明了渐近几何收敛率的存在性。这留下了一个悬而未决的问题：对于特定类别的随机平均收益博弈，能否在“白盒模型”中获得更强的迭代复杂度界。

1.2 贡献

我们应用相对 Krasnoselskii-Mann 价值迭代算法，推导若干类随机博弈的复杂度界。我们特别考虑单链 并发随机平均收益博弈，其中两位玩家的每一对位置策略都会诱导出一个单链转移矩阵（即具有唯一最终类的随机矩阵）。我们将 $p_{\min}$ 定义为模型中最小的非零非对角转移概率。推论 20 表明，经过 $C(\lfloor \log \epsilon \rfloor)$ 次迭代后，相对 Krasnoselskii-Mann 迭代可得到博弈值的 ϵ-近似。因子 C 包含一个形式为 $k\theta^{-k}$的重要项，其中 $k \leqslant n$ 是一个特定的“单链指数”；若所有转移概率均为正，该指数等于 1， $\theta = p_{\min}/(1+p_{\min})$，而 n 表示状态数。接着，我们考虑具有有理转移概率的单链轮流制博弈这一特殊情形，其转移概率

的分母整除 M。定理 23 表明，经过数量级为 M^k 的迭代后即可得到最优策略。其主要工具是定理 19，该定理表明，单链并发博弈的 Shapley 算子的某个适当迭代在希尔伯特半范数下是一个压缩映射。该定理利用变分分析技术证明；具体而言，我们使用了 Mills 的一项经典结果 [32], ，该结果刻画了矩阵博弈值的方向导数，以及非光滑半可微映射的性质。

最后，我们介绍一种适用于熵博弈的相对 Krasnoselkii-Mann 算法变体。定理 26 表明，可在阶为 $(1+\mathcal{A}/m)^k$ 的时间内精确求解不可约熵博弈，其中 $k \leqslant n$ 是某个“不可约性指数”， $m \geqslant 1$ 是非对角转移的最小重数，而$\mathcal{A}$ 是博弈歧义性的一种度量。特别地，有 $W \leqslant \mathcal{A} \leqslant n^{1-1/n}W$，其中 W 是某次转移的最大重数。证明利用了 Birkhoff-Hopf定理，该定理指出，正矩阵在 Hilbert射影度量下是一个压缩映射。

1.3 相关工作

满足不可约条件的随机平均收益博弈的算法方法可追溯至 Hoffman 和 Karp 的工作 [23], ，他们将策略迭代应用于求解轮流制博弈。Chatterjee 和 Ibsen-Jensen [13] 在适当的遍历性条件下，更一般地研究了并发随机平均收益博弈。他们特别证明，博弈值的近似问题属于 FNP，并且对于轮流制遍历博弈，受限的该近似问题至少与简单随机博弈的判定问题同样困难。他们还证明，对于满足不可约条件的并发随机博弈，值迭代经过 $O(\tau\epsilon^{-1}|\log\ \epsilon|)$ 次迭代即可给出博弈值的 ϵ-近似，其中 τ表示在任意策略下任意两个状态之间通行时间的一个上界，见定理 18，同上。Allamigeon、Gaubert、Katz 和 Skomra [5, 关于值迭代的最新“通用上界”（Th. 13]）将该上界改进为 $O(\tau\epsilon^{-1})$。推论 20 进一步将该上界改进为$C|\log\ \epsilon|$。然而，后一个结果要求单链假设，而 [5, Th. 13]的假设则较为宽松。

计算并发折扣随机博弈的博弈值这一问题，Hansen、Koucký、Lauritzen、Miltersen 和 Tsigaridas 已在 [22], 中进行研究。他们利用半代数几何技术证明，若状态数 n 固定，则可在多项式时间内得到一般并发博弈博弈值的 ϵ近似。该多项式的次数为 $O(n)^{n^2}$ 量级，并且他们在 [22]中指出：“如何改善对 n 的依赖关系，是一个非常值得研究的开放问题。”Boros、Gurvich、Elbassioni 和 Makino 研究了并发平均收益博弈的 ϵ遍历性概念，要求两个初始状态的平均收益之差至多为 ϵ。他们提出了一种势函数归约算法，可用于判定 ϵ遍历性，并获得博弈值的 ϵ近似；其对 ϵ 的依赖程度为 $\epsilon^{-O(2^{2n}n_{\max(|A|,|B|)})}$量级，见 [12]。Attia 和 Oliu-Barton 在 [9] 中开发了一种二分算法，其复杂度关于 $|\log\ \epsilon|$ 以及 $|A|^n$ 和 $|B|^n$ 为多项式，其中 A, B 是行动空间。与上述三项研究不同，我们的方法仅适用于单链 并发博弈这一子类，但其复杂度对状态数的依赖更优；特别是，我们的界中的指数至多为 n，执行时间仅随动作数 $|A|$和 $|B|$呈多项式增长。此外，我们的方法适用范围更广，也适用于无限（紧）行动空间（我们只需要一个预言机，能够以给定精度计算可能为无限维的矩阵博弈的博弈值）。

利用压缩技术分析相对值迭代，可追溯到 Federguen、Schweitzer 和 Tijms [18], 关于单人情形及有限行动空间情形的研究，该研究是在本原性条件下进行的。本文的新意在于分析并发双人情形，以及分析 Krasnoselskii-Mann 阻尼的影响；这种阻尼使得我们能够用较弱的单链条件取代先前的本原性条件。此外，即使在单人情形下，我们在定理 19 中给出的收缩率公式也优于 [18] 中的公式（比较见注 37）。

我们在第 9 节关于熵博弈的研究成果受到一系列工作的启发 [8,1, 5]。“无 Despot”熵博弈这一子类可在多项式时间内求解 [1], ，而一般熵博弈能否在多项式时间内求解仍是一个开放问题。 [5] 的方法意味着，可以在 $O(\epsilon^{-1})$次迭代内得到熵博弈博弈值的一个 ϵ-近似，其中 $O(\cdot)$中的因子关于博弈参数呈指数级增长。本文将这一界改进为 $O(|\log\ \epsilon|)$，其中 $O(\cdot)$中的因子取决于一种“模糊性”度量——但我们的方法要求满足不可约性假设。

2 Shapley 算子的初步结果

设 n 为一个整数。若映射 $T\colon\ \mathbb{R}^n \to \mathbb{R}^n$ 满足以下条件，则称其为保序： $\forall x, y \in \mathbb{R}^n$, $x \leqslant y \implies T(x) \leqslant T(y)$，其中 $\leqslant$表示 $\mathbb{R}^n$上的标准偏序。若满足以下条件，则称其为加法齐次： $\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, T(x+\lambda e) = T(x) + \lambda e$ 其中 e 是 $\mathbb{R}^n$ 中每个坐标均为 1 的向量。

▶ **定义 1**。 映射 $T\colon\ \mathbb{R}^n \to \mathbb{R}^n$ 若保序且加法齐次，则称为（抽象）Shapley 算子 。

我们将在下一节说明“Shapley 算子”这一术语的由来，届时会给出具体示例，这些示例是不同类别零和重复博弈的动态规划算子。我们令 $[n] := \{1, \dots n\}$。对于任意 $x \in \mathbb{R}^n$，记 $\mathrm{t}(x) := \max_{i\in[n]} x_i$， $\mathrm{b}(x) := \min^{i\in[n]} x_i$ （读作“顶部”和“底部”）。我们将 x 的希尔伯特半范数定义为： $\|x\|_{\mathrm{H}} = \mathrm{t}(x) - \mathrm{b}(x)$。由于 $\|x\|_{\mathrm{H}} = 0$ 当且仅当 $x \in \mathbb{R}e$，因此 $\|\cdot\|_{\mathrm{H}}$ 实际上是商向量空间 $\mathbb{R}^n/\mathbb{R}e$上的一个范数。我们还注意到： $\|x\|_\infty = \inf\{\lambda \in \mathbb{R}_+ \colon\ -\lambda e \leqslant x \leqslant \lambda e\}$，以及： $\|x\|_H = \inf\{\beta - \alpha \in \mathbb{R}_+ \colon\ \alpha, \beta \in \mathbb{R}, \alpha e \leqslant x \leqslant \beta e\}$。利用这些表达式不难证明，Shapley 算子 T 关于 $\|\cdot\|_{\mathrm{H}}$ 和 $\|\cdot\|_\infty$均为非扩张的（即 1-Lipschitz）。由此，它在商向量空间 $\mathbb{R}^n/\mathbb{R}e$上诱导出一个自映射 T，将等价类 $x + \mathbb{R}e$ 映射为 $T(x) + \mathbb{R}e$，且该映射为非扩张。

▶ 定义*2*。我们将 *Shapley* 算子 T 的逃逸率 $\chi(T)$ 定义为 $\lim_{k\to\infty} k^{-1}T^k(v) \in \mathbb{R}^n$，其中 v 属于 $\mathbb{R}^n$。下逃逸率和上逃逸率分别定义为 $\underline{\chi}(T) = \lim_k k^{-1}\mathrm{b}(T^k(v))$ 和 $\overline{\chi}(T) = \lim_k k^{-1}\mathrm{t}(T^k(v))$。

由于 T 在上确界范数下是非扩张的，因此这些极限的存在性及其值均与 $v \in \mathbb{R}^n$的选择无关。一般而言，逃逸率 $\chi(T) = \lim_{k\to\infty} k^{-1}T^k(v)$ 不一定存在，但次可加论证表明，下逃逸率和上逃逸率始终存在，参见例如 [19]。确立逃逸率存在性的一项基本工具是考虑以下遍历方程。

▶ **定义3.** 我们称遍历方程 有解，当存在 $\lambda \in \mathbb{R}$ 以及满足以下条件的 $u \in \mathbb{R}^n\colon\ T(u) = \lambda e + u$。

▶ **观察 4**。 如果上述遍历方程可解，则 $\chi(T) = \lambda e$。更一般地，如果 $\alpha e + v \leqslant T(v) \leqslant \beta e + v$ 对某些 $v \in \mathbb{R}^n$ 和 $\alpha, \beta \in \mathbb{R}$成立，则 $\alpha \leqslant \underline{\chi}(T) \leqslant \overline{\chi}(T) \leqslant \beta$。

证明。 通过直接归纳法，并且由于 T 是保序且加法齐次的，我们有： $k\alpha e + v \leqslant T^k(v) \leqslant k\beta e + v$。于是， $k\alpha + \mathbf{b}(v) \leqslant \mathbf{b}(T^k(v)) \leqslant \mathbf{t}(T^k(v)) \leqslant k\beta + \mathbf{t}(v)$。两边除以 k 并令 k 趋于无穷大，我们得到第二个结论。◀

我们受到不动点理论中的以下观察启发，该观察已在附录 B 中得到证明。

▶ **观察 5.** 现在假设 T^q 在希尔伯特半范数 $k \cdot k_H$ 下是 γ-压缩映射，其中 $q \geqslant 1$ 和 $0 < \gamma < 1$ 满足相应条件。那么，遍历方程有解。

Shapley 算子包括（有限维）马尔可夫算子，其形式为$T(x) = Mx$，其中 M 是一个 $n \times n$ 随机矩阵（即 M 的元素非负且行和为一）。在这种情况下，收缩率有一个精确公式。事实上，可以考虑 M 的算子范数，将其视为作用于商向量空间 $\mathbb{R}^n/\mathbb{R}e$ 上的线性映射，$\|M\|_{\mathrm{H}} = \sup_{u \notin \mathbb{R}e} \|Mu\|_{\mathrm{H}} \|u\|_{\mathrm{H}}$。

▶ **定理 6** （[15]的推论）。 $\|M\|_{\mathrm{H}} = \delta(M) := 1 - \min_{1 \leqslant i < j \leqslant n} \left\{ \sum_{k \in [n]} \min(M_{ik}, M_{jk}) \right\}$。

附录 C 中回顾了由 [15] 得出的推导。术语 $\delta(M)$ 称为 *Dobrushin 遍历系数*。

3 两类零和双人重复博弈

下面回顾有限状态空间下两类零和双人博弈的定义和基本性质。关于随机博弈的更多细节，参见 [31]；关于熵博弈的更多细节，参见 [8, 1]。

3.1 并行重复零和双人随机博弈

我们假设状态空间等于 $[n] = \{1, \ldots, n\}$。我们称两位玩家为“Min”和“Max”。该博弈由以下数据规定。对于每个状态 $i \in [n]$，给定两个非空紧集 $A(i)$ 和 $B(i)$，分别表示玩家 Min 和 Max 的可行动作。对于每个 $i \in [n]$ 以及每种动作选择 $(a, b) \in A(i) \times B(i)$，给定一个实数 r_i^{ab}，表示即时支付；并给定一个随机向量 $P^{a,\ i\ b} = (P_{ij}^{ab})_{j \in [n]}$，即 $P_{ij}^{ab} \geqslant 0$ 且 $\sum_{j \in [n]} P_{ij}^{ab} = 1$。我们假设函数 $(a, b) \mapsto r_i^{ab}$ 和 $(a, b) \mapsto P^{a,\ i\ b}$ 连续。

并发博弈从已知的初始状态 i_0 开始，分阶段连续进行。在阶段 k，我们分别用 a_k 和 b_k 表示 Min玩家和 Max 选择的动作，用 i_k 表示该阶段的状态。到阶段 k 为止的历史由序列 $H_k = ((i_\ell, a_\ell, b_\ell)_{0 \leqslant \ell < k}, i_k)$ 组成。Min玩家（分别为 Max）的随机化策略是一组可测函数，为每段历史 H_k 指定紧集 $A(i_k)$（分别为 $B(i_k)$）上的概率测度 α_k（分别为 β_k）。在阶段 k，Min玩家获知截至该阶段的历史 H_k 后，按照概率测度 α_k 抽取一个随机动作 a_k；同样，Max 按照概率测度 β_k 抽取一个随机动作 b_k。随后，Min玩家向 Max 支付即时支付 $r^{a_k, b_k}_{i_k}$，并按照状态空间 $[n]$ 上的概率测度 $(P^{a_k b_k}_{i_k j})_{j \in [n]}$ 随机抽取下一个状态 i_{k+1}；也就是说，给定历史 H_k 和动作 a_k, b_k 时，$i_{k+1} = j$ 的条件概率为 $P^{a_k b_k}_{i_k j}$。若玩家的动作是历史的确定性函数，则称该策略为纯策略或确定性策略。我们用 σ_k（分别为 τ_k）表示 Min玩家（分别为 Max）在阶段 k 的策略，并用 σ 和 τ 表示序列 $(\sigma_k)_{k \geqslant 0}$ 和 $(\tau_k)_{k \geqslant 0}$。这样，对于任意初始状态 $i_0 \in [n]$ 以及任意一对的

两位玩家的策略（σ, τ）对应于一个无限随机序列（i_k, a_k, b_k）$_{k\geqslant 0}$。我们用 $\mathbb{E}^{\sigma}_{\cdot,i_0}{}^{\tau}$ 表示相对于该过程的期望值。

我们需要考虑一些特殊类别的策略。马尔可夫策略或位置策略是这样一种策略：σ_k（分别为 τ_k）仅取决于当前状态 i_k。这种位置策略 σ_k 或 τ_k 也称为策略。如果该策略与 k无关，则称其为平稳策略。我们将用 $\Delta_{A(i)}$（分别为 $\Delta_{B(i)}$）表示 A（i）上的概率测度集合（分别为B（i））。我们用 $\Pi_{\mathrm{P}}^{\mathrm{Min}}$（分别为 $\Pi_{\mathrm{R}}^{\mathrm{Min}}$）表示 Min 的纯策略（分别为随机策略）集合。它与 $\prod_{i\in[n]} A$（i）（分别为 $\prod_{i\in[n]} \Delta_{A(i)}$）一一对应（并将二者视为相同对象）。事实上，如果$\sigma \in \Pi_{\mathrm{P}}^{\mathrm{Min}}$，则 σ可标识为（$\sigma(1) \ldots, \sigma(n)$），其中 $\sigma(i) \in A(i)$ 是 Min 根据策略 σ在状态 i 所选择的动作。同样地，如果 $\alpha \in \Pi_{\mathrm{R}}^{\mathrm{Min}}$，则 α 可标识为（$\alpha_1, \ldots, \alpha_n$），其中对于 $i \in [n]$，且 $a \in A(i)$，$d\alpha_i(a)$ 是 Min 根据策略 α在状态 i选择动作 a 的概率。最后，Min 的纯策略 σ 可标识为由狄拉克测度组成的随机策略，因此 $\Pi_{\mathrm{P}}^{\mathrm{Min}} \subset \Pi_{\mathrm{R}}^{\mathrm{Min}}$。对于 Max，我们也使用相同的记号和对应关系。我们还将用 $\Pi_{\mathrm{P}} = \Pi_{\mathrm{P}}^{\mathrm{Min}} \times \Pi_{\mathrm{P}}^{\mathrm{Max}}$和 $\Pi_{\mathrm{R}} = \Pi_{\mathrm{R}}^{\mathrm{Min}} \times \Pi_{\mathrm{R}}^{\mathrm{Max}}$表示策略对的空间。

给定初始状态 i_0 以及两位玩家的一对策略（σ，τ），玩家 Max 在时域 N 内获得的期望收益定义为

$$J_{i_0}^N(\sigma, \tau) \coloneqq \mathbb{E}_{i_0}^{\sigma,\tau}\left[\sum_{k=0}^{N-1} r_{i_k}^{a_k,b_k}\right] .$$

我们将具有上述 i_0项（对每个 $i_0 \in [n]$）的 $\mathbb{R}^n$向量记为 J^N（σ，τ）。有限时域博弈有一个博弈值 $v^N \in \mathbb{R}^n$，并且存在一对最优（随机化）策略（σ^*，τ^*），这意味着

$$J^N(\sigma^*, \tau) \leqslant v^N = J^N(\sigma^*, \tau^*) \leqslant J^N(\sigma, \tau^*) , \tag{1}$$

对于所有策略对（σ，τ）均成立，见 [31]。此外，可以选择一对位置型最优策略（σ^*，τ^*），也就是说，对于所有 $k \leqslant N$，都有（σ_k^*，τ_k^*）$\in \Pi_{\mathrm{R}}$（但通常取决于 k 和 N）。这些最优策略可以通过该博弈的动态规划方程求得，如下所示。

对于任意 i，$j \in [n]$、$\alpha_i \in \Delta_{A(i)}$ 和 $\beta_i \in \Delta_{B(i)}$，记

$$r_i^{\alpha_i,\beta_i} = \int_{A(i)\times B(i)} r_i^{a,b} d\alpha_i(a) d\beta_i(b) \quad \text{and} \quad P_{i,j}^{\alpha_i,\beta_i} = \int_{A(i)\times B(i)} P_{i,j}^{a,b} d\alpha_i(a) d\beta_i(b) \ . \tag{2}$$

这将函数 $(a, b) \mapsto r^{a}_{\cdot,i}{}^{b}$和 $(a, b) \mapsto P^{a}_{\cdot,i}{}^{b}{}_{,j}$从 $A(i) \times B(i)$ 扩展到 $\Delta_{A(i)} \times \Delta_{B(i)}$。随后，我们将并发博弈的 Shapley 算子 T 定义为映射 $T\colon \mathbb{R}^n \to \mathbb{R}^n$，使得

$$T_i(v) = \min_{\alpha_i\in\Delta_{A(i)}} \max_{\beta_i\in\Delta_{B(i)}} \left(r_i^{\alpha_i,\beta_i} + \sum_{j\in[n]} P_{ij}^{\alpha_i,\beta_i} v_j\right), \ \text{for } i \in [n],\ v \in \mathbb{R}^n \ . \tag{3}$$

注意，在上述表达式中，下确界和上确界可以交换次序，这是由于行动空间的紧性以及函数 $(a, b) \mapsto r^{a}_{\cdot,i}{}^{b}$ 和 $(a, b) \mapsto P_{ij}^{ab}$的连续性假设所致（这可由 Sion 极小极大定理推出）。此外，算子 T满足定义 1 的性质。

然后，通过递推方程可得有限时域并发博弈的博弈值：$v^0 = 0$，$v^N = T(v^{N-1})$。此外，当剩余时域为 $k < N$（或处于第 $N - k$ 阶段）时，通过选择最优策略 α and可得到该博弈的最优策略。

β 相对于向量 v^k，也就是说，使得 α_i 和 β_i 在 (3) 中的 $T_i(v^k)$ 表达式里为最优。

现在我们描述平均收益博弈，它是通过考察收益在时域 N 趋于无穷时的 Cesàro 极限得到的。更确切地说，我们定义：

$$\chi^+_{i_0}(\sigma,\tau) \coloneqq \limsup_{N\to\infty} N^{-1} J^N_{i_0}(\sigma,\tau) \quad \chi^-_{i_0}(\sigma,\tau) \coloneqq \liminf_{N\to\infty} N^{-1} J^N_{i_0}(\sigma,\tau) \ .$$

我们称带有平均收益的博弈具有博弈值 $\chi^* \in \mathbb{R}^n$，如果对于所有 $\varepsilon > 0$，都存在两位玩家的策略 $\sigma^\varepsilon, \tau^\varepsilon$，使其为 ε-最优；这意味着，对于任意策略σ和 τ，$-\varepsilon e + \chi^+(\sigma^\varepsilon,\tau) \leqslant \chi^* \leqslant \chi^-(\sigma,\tau^\varepsilon) + \varepsilon e$。Mertens 和 Neyman [30], 基于 Bewley 和 Kohlberg 的一项结果 [11], 证明，当行动空间 $A(i)$ 和 $B(i)$ 有限时，平均收益博弈具有博弈值（实际上，这在更强的一致 意义下成立）。此外，该博弈值与 Shapley 算子的逃逸率一致，即 $\chi^* = \lim_k T^k(0)/k$。Vigeral 的一个反例表明，这些性质并不适用于一般的紧致行动空间 [41]。

一个特别值得关注的情形是遍历方程可解时，也就是说，存在 $\lambda \in \mathbb{R}$ 和 $v \in \mathbb{R}^n$，使得 $T(v) = \lambda e + v$。在这种情形下，$\chi^* = \lambda e$，并且存在分别属于两位玩家的最优随机策略，且这两个策略均为位置型和平稳策略。通过选择一对关于 v 的最优策略 (α, β)，即可得到这样一对策略；也就是说，这对策略在 (3) 中 $T(v)$ 的表达式下是最优的。我们将看到，即使在行动空间紧致的情况下，单链条件也总能保证遍历方程可解（定理 11）。

一个值得注意的并发博弈子类是轮流行动的博弈。此时，动作空间 $A(i)$ 和 $B(i)$ 必须是有限的，并且对于每个状态 $i \in [n]$，我们假设 $A(i)$ 或 $B(i)$ 中至少有一个是单元素集。换句话说，状态集存在一个二分划分 $[n] = I_{\mathrm{Min}} \uplus I_{\mathrm{Max}}$，使得在每个状态 $i \in I_{\mathrm{Min}}$ （分别地，I_{Max}）中，Min（分别地，Max）是唯一需要做出决策的玩家。于是，博弈的 Shapley 算子化简为$T_i(v) = \min_{a\in A(i)} \max_{b\in B(i)} \left(r^{a,\ b}_{i} + \sum_j P^{a,\ i\ \ b}_{\ \ ,\ j} v_j\right)$，其中 $i \in [n]$; 再次地，min 和 max 可以交换，因为对于每个 $i \in [n]$，min 或 max 都是在一个缩减为单元素集的集合上取值。因此，相对于向量 v 的最优策略集包含纯策略，可通过在$T_i(v)$ 的前述表达式中取 $\sigma(i) = a$ 和 $\tau(i) = b$ 最优值而得。由此可得有限时域轮流行动随机博弈的纯最优策略，以及遍历方程有解的平均收益轮流行动随机博弈的纯最优策略。Liggett 和 Lippman [28] 证明了轮流行动平均收益随机博弈存在纯最优策略。附录 A 给出了一个说明性示例。

3.2 熵博弈

熵博弈在 [8] 中首次提出。这里我们采用 [1, 5], 中稍微一般化的模型，相关背景请参阅该文。熵博弈是一种轮流行动的博弈，在一个（有限）有向图（$\mathcal{V}$, $\mathcal{E}$）上进行，有两名玩家，分别称为“Despot”和“Tribune”，以及一名额外的非确定性玩家，称为“People”。我们假设顶点集 $\mathcal{V}$ 有一个非平凡划分：$\mathcal{V} = \mathcal{V}_D \uplus \mathcal{V}_T \uplus \mathcal{V}_P$。Despot、Tribune 和 People 玩家分别控制 $\mathcal{V}_D$、$\mathcal{V}_T$ 和 $\mathcal{V}_P$ 中的状态，并交替行动，即 $\mathcal{E} \subset (\mathcal{V}_D \times \mathcal{V}_T) \cup (\mathcal{V}_T \times \mathcal{V}_P) \cup (\mathcal{V}_P \times \mathcal{V}_D)$。我们假设每条边 $(p,\ d) \in \mathcal{E}$，其中 $p \in \mathcal{V}_P$ 且 $d \in \mathcal{V}_D$，都配有一个重数 m_{pd}，它是一个（正）自然数。为简化表述，这里我们仅用纯策略（平稳位置策略）定义熵博弈的博弈值。更准确地说，Despot 的策略 σ 是一个映射，它为每个节点 $d \in \mathcal{V}_D$ 指定一个节点 t，使得 $(d,\ t) \in \mathcal{E}$。类似地，Tribune 的策略 τ 是一个映射，它为每个

节点 $t \in \mathcal{V}_T$ 指定一个节点 $p \in \mathcal{V}_P$。我们用 n 表示 $\mathcal{V}_D$ 的基数。这样一对策略确定一个 $n \times n$ 矩阵 $M^{\sigma,\tau}$，使得 $M^{\sigma,\tau}_{d,d'} = m_{\tau(\sigma(d)),d'}$。给定一个初始状态 $d \in \mathcal{V}_D$，我们用极限 $R(\sigma,\tau) := \lim_{k\to\infty} \left((M^{\sigma,\tau})^k e\right)^{1/k}_{\bar{d}}$ 来衡量玩家 People 的“自由度”。一对策略确定一个子图 $G^{\sigma,\tau}$，其构造方式是：对于 $\mathcal{V}_D$ 中的每个节点，仅保留 σ 指定的后继节点；对 τ 和 $\mathcal{V}_T$ 也同样处理。于是，People 的“自由度”恰好等于从节点 d 出发、长度为 $3k$ 的路径数量的几何增长率，路径数按重数计，且 $k \to \infty$。一般而言，图 $G^{\sigma,\tau}$ 可能有多个强连通分量。据 [5] 所述，$R(\sigma,\tau)$ 等于矩阵 $M^{\sigma,\tau}$ 的对角块中最大谱半径；这些对角块对应于 $G^{\sigma,\tau}$ 中初始状态 d 能到达的强连通分量。在熵博弈中，Despot 希望最小化 People 的自由度，而 Tribune（名称取自罗马共和国的行政官）则希望将其最大化。 [1] 证明了熵博弈在位置策略空间中存在博弈值，也就是说，存在策略σ^*, τ^*，使得对所有策略 σ,τ，均有 $R(\sigma^*,\tau) \leqslant R(\sigma^*,\tau^*) \leqslant R(\sigma,\tau^*)$。

熵博弈的动态规划算子是 F 到 $\mathbb{R}^n_{>0}$ 的自映射，由 $F_d(x) = \min_{t\in\mathcal{V}_T,(d,t)\in\mathcal{E}} \max_{p\in\mathcal{V}_P,(t,p)\in\mathcal{E}} \sum_{d'\in\mathcal{V}_D,(p,d')\in\mathcal{E}} m_{p,d'} x_{d'}$给出，其中 $d \in \mathcal{V}_D$。于是，算子 $T := \log \circ F \circ \exp$ 是一个 Shapley 算子。文献 [1] 证明，初始状态为$\bar{d}$的熵博弈的博弈值由极限 $\lim_{k\to\infty} [F^k(e)]^{1/k}_{\bar{d}}$ 给出。

4 单链性质

回顾一下，对于每个 $n \times n$ 非负矩阵 M，都可关联一个节点集为$[n]$的有向图，使得当且仅当 $M_{ij} > 0$时，存在一条从 i 到 j 的弧。若该有向图是强连通的，则该矩阵是不可约 的。若该有向图有唯一的终端强连通分量（若从某个强连通分量出发的任意路径都留在该分量内，则称该强连通分量为终端 ），则该矩阵是单链 的。单链性质有时也称为遍历性，因为随机矩阵是单链的，当且仅当它只有一个不变测度；等价地，唯一的调和向量 （即满足 v 的 $Mv = v$的解）是常向量，参见文献 [2],中定理 1.1 的讨论及其中的参考文献。

给定一对纯策略 $(\sigma, \tau) \in \Pi_P$，我们定义随机矩阵： $P^{\sigma,\tau} = (P^{\sigma(i)\tau(i)}_{i,j})_{i,j\in[n]}$。

▶**定义 7**。如果对于所有纯策略对 σ、τ，矩阵 $P^{\sigma,\tau}$都是单链（分别为不可约），则称该博弈为单链（分别为不可约）。

▶**定义 8**。如果从状态 $s \in S$出发，并按照策略 σ和 τ行动，下一状态仍属于 S，则称状态的子集 S 在矩阵$P^{\sigma,\tau}$的作用下封闭。

▶ 注 9。如果 S是单链矩阵 $P^{\sigma,\tau}$作用下的一个封闭集合，那么 S包含该矩阵的最终类。

▶ 注 10。对于我们中的策略对 σ，τ，最终类不一定相同。 r
单链博弈的定义。
以下定理在附录 E 中得到证明，论述了单链博弈情形下遍历方程解的存在性问题。

▶**定理 11.** 设 T 为单链并发随机博弈的 *Shapley* 算子。则存在一个向量 $v \in \mathbb{R}^n$ 和 $\lambda \in \mathbb{R}$，使得 $T(v) = \lambda e + v$。此外，对于每个状态 $i \in [n]$，通过选择在 $[T(v)]_i$的表达式中分别取得最小值和最大值的动作，可以得到一对最优（随机化）位置策略。

5 相对值迭代

相对值迭代最初用于求解单人随机平均收益博弈（即平均成本马尔可夫决策过程），见 [42]。“原始”值迭代算法通过从 $x_0 = 0$开始计算序列 $x_{k+1} = T(x_k)$ 来实现。随后， x_k 给出博弈在时域 k内的博弈值向量，因此我们预期当 $k \to \infty$时， x_k 趋于无穷大。相对值的思路是用加法常数对序列进行归一化。算法 1 给出了相对值迭代的一般版本，并允许使用近似动态规划预言机。这使我们能够在图灵计算模型下获得复杂度结果：针对给定的有理向量 x 和给定精度，计算 Shapley 算子 $T(x)$ 的有理近似值。

算法 1 近似算术中的相对值迭代

1：输入：最终要求的数值精度 $\epsilon > 0$ 和参数 $0 < \eta \leqslant \epsilon/3$。一个˜预言机 T，它提供 Shapley 算子 T的上确界范数下的 η-近似。2： $x \coloneqq 0 \in \mathbb{R}^n$3：重复˜ ˜4：$x \coloneqq T(x) - \mathsf{t}(T(x))e$˜5：直到 $\mathsf{k}x - T(x)\mathsf{k}_{\mathrm{H}} \leqslant \epsilon/3$˜ ˜6： $\alpha \coloneqq \mathsf{b}(T(x) - x)$; $\beta \coloneqq \mathsf{t}(T(x) - x)$ 7：返回 x, α, β ▷ T的下逃逸率和上逃逸率包含在区间$[\alpha - \epsilon/3, \beta + \epsilon/3]$中，其宽度至多为 ϵ

▶**定理 12.** 设 T 是一个 *Shapley* 算子。那么，

1. 算法 *1* 终止时，将返回一个宽度至多为 ϵ 的有效区间，其中包含 T的下逃逸率和上逃逸率。

2. 如果存在整数 q 和标量 $0 < \gamma < 1$，使得 T^q 是希尔伯特半范数下的一个 γ-压缩映射，并且 η 的选取足够小，使得 $\eta(12 + 24q/(1-\gamma)) \leqslant \epsilon$，那么算法 *1* 至多经过 $q(\log \|T(0)\|_H + \log 6 + |\log \epsilon|)/|\log \gamma|$ 次迭代即可终止。

证明见 §F，其中利用了算子 T 在希尔伯特半范数下的非扩张性。

6 Krasnoselskii-Mann 阻尼

我们将看到，对于轮流行动博弈或并发博弈，用 Krasnoselskii-Mann 阻尼版本的 Shapley 算子替代原始 Shapley 算子是有益的。这样，相对值迭代算法便能在较弱的条件下收敛。

▶ **定义 13**。 若 T是一个 *Shapley* 算子，我们定义 $T_\theta = \theta I + (1-\theta)T$ 其中 I 是恒等算子。

我们将 *Krasnoselskii-Mann* 算子 称为 T_θ 算子。很容易证明，它也是一个 Shapley 算子。下面的观察将阻尼 Shapley 算子的遍历常数与原始 Shapley 算子的遍历常数联系起来。

▶ **引理 14.** 设 T为 *Shapley* 算子， $u \in \mathbb{R}^n$ 且 $\lambda \in \mathbb{R}$。那么， $T(u) = \lambda e + u$ 当且仅当 $T_\theta(u) = (1-\theta)\lambda e + u$。特别地，只要遍历方程 $T(u) = \lambda e + u$ 有解， $\chi(T) = (1-\theta)^{-1}\chi(T_\theta)$ 就成立。

证明。 等价性显而易见，且 $\chi(T) = (1-\theta)^{-1}\chi(T_\theta)$ 由观察 4 得出。◀

10 相对 Krasnoselskii-Mann 迭代

我们考虑迭代 $x_{k+1} = T_\theta(x_k) - \mathrm{t}(T_\theta(x_k))e$，它是通过对 Krasnoselskii-Mann 算子 T_θ 应用相对值迭代（如算法 1 所示）得到的，初始条件 $x_0 \in \mathbb{R}^n$ 任意。Ishikawa 证明，对于有限维赋范空间上的非扩张自映射，只要存在不动点，普通的 Krasnoselskii-Mann 迭代就会收敛 [25]。由此得到以下结果。

▶ **定理 15** (与 [20] 相比较。 设 $T\colon \mathbb{R}^n \to \mathbb{R}^n$ 为一个 *Shapley* 算子，且 $0 < \theta < 1$。那么，通过对 *Krasnoselskii-Mann* 算子 T_θ 应用相对值迭代所得到的序列 x_k 收敛，当且仅当 $\exists u \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$, $T(u) = \lambda e + u$。

这一结果的乘法变体已在 [20] 的定理 11 中得到证明。为完整起见，我们在 §G 中给出证明。

7 单链博弈在纯策略下的压缩性质

我们定义以下参数，表示非零非对角线转移概率的最小博弈值，$p_{\min} = \min_{i,\ j\in[n],\ i\neq j,\ (a,b)\in A\times B}\{P^a_{i,j}(b) : P^a_{i,j}(b) > 0\}$，并设 $\theta := p_{\min}/(1+p_{\min})$. 对于两个玩家的每一策略对 σ, τ，我们设定 $Q^{\sigma,\tau} := \theta I + (1-\theta)P^{\sigma,\tau}$. 对于任意一列纯策略对 $\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k$，我们对所有 $i \in [n]$,$S_i(\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k)$ 定义：$= \{j \mid [Q^{\sigma_1,\tau_1} \ldots Q^{\sigma_k,\tau_k}]_{ij} > 0\}$.

▶ **引理 16.** 假设一个并发博弈是单链的。那么，存在一个整数 $k \leqslant n$，使得对于所有 i_1, $i_2 \in [n]$，以及所有纯策略对序列 $\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k$,$S_{i_1}(\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k) \cap S_{i_2}(\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k) \neq \varnothing$。

我们称博弈的单链指数 为 k_{uni}，并将满足引理 16 所述性质的最小整数记为 k。类似地，我们称不可约博弈的不可约性指数为 k_{irr}，并将满足以下条件的最小整数记为 k：对于任意纯策略序列$\sigma_1, \tau_1, \cdots \ \sigma_k, \tau_k$，矩阵 $Q^{\sigma_1,\tau_1} \cdots Q^{\sigma_k,\tau_k}$ 为正矩阵。我们有 $1 \leqslant k_{\text{uni}} \leqslant k_{\text{irr}}$。

以下结果将使我们能够获得几何收缩率。该定理及下一条命题的证明尤其表明，$k_{\text{uni}} \leqslant n$ 如果博弈是单链，并且 $k_{\text{irr}} \leqslant n$ 如果博弈是不可约。

▶**定理 17**。 假设一个具有 n 个状态的并发博弈是单链的，其单链指数为 $k = k_{\text{uni}}$。那么，对于两个玩家的纯策略对的所有序列 $\sigma_1, \tau_1, \cdots \ \sigma_k, \tau_k$，$kQ^{\sigma_1,\tau_1} \ldots Q^{\sigma_k,\tau_k}k_{\mathrm{H}} \leqslant 1 - \theta^k$。

下述命题改进了定理17给出的收缩率界，适用于不可约博弈这一特殊情形。

▶**命题 18**。 假设一个具有 n 个状态的并发博弈是不可约的，且 $k = k_{\text{irr}}$ 是该博弈的不可约性指数。那么，对于两个玩家的纯策略对的所有序列 $\sigma_1, \tau_1, \cdots \ \sigma_k, \tau_k$，$kQ^{\sigma_1,\tau_1} \ldots Q^{\sigma_k,\tau_k}k_{\mathrm{H}} \leqslant 1 - n\theta^k$。

定理 17 和命题 18 的证明见附录 H。

8 通过相对 Krasnoselskii-Mann 迭代求解并发博弈和轮流制博弈

我们首先为并发单链博弈建立一个一般界。回顾一下，$\theta = p_{\min}/(1+p_{\min})$。

▶**定理 19.** 设 T_θ 是一个并发单链博弈的 *Krasnoselskii-Mann* 算子，且那么， T_θ^k在希尔伯特半范数下是一个压缩映射，其收敛率有界于 $1-\theta^{k_{\mathrm{uni}}}$。此外，如果该博弈是不可约的，则相同的收敛率有界于 $1-n\theta^{k_{\mathrm{irr}}}$。

该定理在附录 I 中得到证明。结合这一结果与定理 12，我们得到以下结论，其中我们用 $\|r\|_\infty := \max_{i,a,b}|r_i^{ab}|$ 表示支付函数的上确界范数。

▶**推论 20.** 设 T 为一个并发单链博弈的 *Shapley* 算子，且 $\epsilon \in (0,1)$。将算法 *1* 应用于 *Krasnoselskii-Mann* 算子 T_θ，并采用定理 *12* 中规定的精度 η，最多经过 $(|\log(\epsilon)|+|\log(1-\theta)+\log 12+\log\|r\|_\infty)k_{\mathrm{uni}}\theta^{-k_{\mathrm{uni}}}$ 次迭代，即可得到该博弈值的 ϵ-近似值。

现在我们考虑轮流制博弈的特殊情形。此时，博弈值是有理数，并且存在最优纯策略。现在，我们应用我们的方法来精确计算博弈值并找到最优纯策略。

▶ **假设21**。 现在我们假设概率 $P_{i,j}^{a,b}$ 均为有理数，其公共分母记为 M。我们还假设支付 $r_i^{a,b}$ 均为整数。

▶ **引理22 (** [38]的推论**)**。 设 P为一个 $n\times n$ 单链矩阵，其元素为具有公共分母 M的有理数。那么， P 的唯一不变测度的各元素均为有理数，且分母至多为 nM^{n-1}。

当算法 1 停止并返回向量 $x\in\mathbb{R}^n$时，我们选择两个相对于 x最优的纯策略 σ^* 和τ^*，这意味着对于 $i\in[n]$，有：

$$T_i(x)=\max_{b\in B(i)}\left(r_i^{\sigma^*(i),b}+\sum_{j\in[n]}P_{ij}^{\sigma^*(i),b}x_j\right)=\min_{a\in A(i)}\left(r_i^{a,\tau^*(i)}+\sum_{j\in[n]}P_{ij}^{a,\tau^*(i)}x_j\right)\ . \tag{4}$$

▶**定理 23**。 考虑一个满足假设*21*的单链轮流行动随机博弈。选择 $\epsilon=(1-\theta)(n^2M^{2(n-1)})^{-1}$，使得将算法 *1* 应用于 T_θ 时，其运行次数至多为

$$(\log(1-\theta)+2\log n+2(n-1)\log M+\log 12+\log\|r\|_\infty)\theta^{-k_{\mathrm{uni}}}k_{\mathrm{uni}} \tag{5}$$

迭代。设 x^* 为算法返回的向量。选择纯策略 σ^* 和τ^*，使其分别达到表达式 $T(x^*)$中的最小值和最大值，如式 (4) 所示。然后，这些策略是最优的。

该定理的证明见附录 K。

9 将乘法 Krasnoselskii-Mann 阻尼应用于熵博弈

对于熵博弈，针对第 3.2 节中定义的算子 F，遍历特征值问题是求解 $u\in\mathbb{R}^n$ 和 $\lambda\in\mathbb{R}$，使得 $\exp(\lambda)\exp(u)=F(u)$。等价地，$\lambda e+u=T(u)$，其中 $T=\log\circ F\circ\exp$。若该方程有解，则对所有初始状态 $d\in\mathcal{V}_D$，$\exp(\lambda)$ 都是该熵博弈的博弈值。为求解该方程，我们固定一个正数 $\vartheta>0$，并考虑 Krasnoselskii-Mann 算子的以下“乘法”变体：

$$[T_{\mathrm{m},\vartheta}(v)]_d=\log\min_{t\in\mathcal{V}_T,(d,t)\in\mathcal{E}}\ \max_{p\in\mathcal{V}_P,(t,p)\in\mathcal{E}}\left(\vartheta\exp(v_d)+\sum_{d'\in\mathcal{V}_D,(p,d')\in\mathcal{E}}m_{p,d'}\exp(v_{d'})\right)\ .$$

12 相对 Krasnoselskii-Mann 迭代

与加法情形不同，我们不对恒等映射和 Shapley 算子进行“凸组合”，而只添加“对角项” ϑ $\exp(v_d)$，其中 ϑ 仍可解释为“阻尼强度”，不过是在乘法意义下。如果 $T(u) = \lambda e + u$，那么容易验证 $T_{\mathrm{m},\vartheta}(u) = \mu e + u$，其中 $\mu = \log(\vartheta + \exp(\lambda))$，反之亦然，因此 T 和 $T_{\mathrm{m},\vartheta}$ 的非线性特征值问题是等价的。与加法情形一样，必须调整阻尼强度以优化复杂度界。若在博弈的图中不存在路径 $d' \to t \to p \to d'$，我们称重数 $m_{p,d'}$ 为非对角。等价地，对于两位玩家的任意策略选择 σ, τ，元素 $m_{p,d'}$ 不会出现在第 3.2 节定义的矩阵 $M^{\sigma,\tau}$ 的对角线上。于是，我们用m表示非对角重数的最小值，并注意到 m 恰好是与所有策略对相关的矩阵 $M^{\sigma,\tau}$ 的所有非对角元素中的最小值。我们令 $\vartheta := m$。

如果对于每一对策略 σ, τ，矩阵 $M^{\sigma,\tau}$ 都是不可约的，我们就称熵博弈是不可约的。不可约熵博弈的不可约性指数 k_{irr} 是满足以下条件的最小整数 k：对于所有策略 $\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k$，矩阵 $M^{\sigma_1\tau_1} \ldots M^{\sigma_k\tau_k}$ 都具有正元素。与随机并发博弈的情形类似，我们可以得出，只要博弈是不可约的，就有$k_{\mathrm{irr}} \leqslant n$。我们定义熵博弈 A_l 的 l-歧义度为：$= \max_{d,d' \in \mathcal{V}_D} \max_{\sigma_1,\tau_1,\ldots,\sigma_l,\tau_l} (M^{\sigma_1\tau_1} \ldots M^{\sigma_l\tau_l})_{d,d'}$。注意，$(M^{\sigma_1\tau_1} \ldots M^{\sigma_l\tau_l})_{d,d'}$ 是由策略 $\sigma_1, \tau_1, \ldots, \sigma_l, \tau_l$ 所诱导的有限时域博弈中，从 d 到 d' 的路径数（按重数计），这也解释了“l-歧义度”这一术语的由来。若博弈是不可约的，我们将该博弈的歧义度定义为 $\mathsf{A} := \max_{1 \leqslant l \leqslant k_{\mathrm{irr}}} \mathsf{A}_l^{1/l}$。令$W := \max_{(p,d) \in \mathcal{E} \cap (\mathcal{V}_P \times \mathcal{V}_D)} m_{p,d}$，并注意到 $W \leqslant \mathsf{A} \leqslant n^{1-1/k_{\mathrm{irr}}} W$。

▶**定理 24**。设 $T_{m,\ \vartheta}$ 为一个不可约熵博弈的乘法 *Krasnoselskii-Mann* 算子。则，$T^{k_{\mathrm{irr}}}{}_{m\ \vartheta}$ 在希尔伯特半范数下是一个压缩映射，其收缩率的上界为 $\bar{\mathcal{M}}^{-1}\mathcal{M}+1$，其中 $\bar{M} := (1 + A/m)^{k_{\mathrm{irr}}}$。

该结果在附录 L 中证明。我们回顾以下分离界。

▶**定理 25 (** [5]的推论）。设在具有 n *Despot* 状态的熵博弈中，两对策略产生不同的博弈值。那么，这些博弈值至少相差 ν_n^{-1}，其中

$$\nu_n := 2^n (n+1)^{8n} n^{2n^2+n+1} e^{4n^2} \max(1, W/2)^{4n^2} .$$

然后，利用定理 12，我们推得：

▶**定理 26**。考虑一个不可约熵博弈，其不可约性指数为 k_{irr}。令我们选择 $\epsilon = (1 + (m + W)\nu_n)^{-1}$，使得将算法 *1* 应用于 $T_{m,\ \vartheta}$ 至多运行 $\bar{}(\log(1 + (m + W)\nu_n) + \log 6)k_{\mathrm{irr}}M/2$-次迭代。此外，设 x^* 为算法返回的向量。选取纯策略 σ^* 和 τ^*，分别在 $T_{m,\ \vartheta}(x^*)$ 的表达式中达到最小值和最大值。则这些策略是最优的。

10 结语

我们已为应用于若干满足不可约条件的随机博弈类别的相对值迭代，建立了参数化复杂度界。这些界依赖于希尔伯特半范数中的压缩性质。值得探讨的是，能否也利用这些压缩性质为策略迭代推导复杂度界，而非为值迭代推导。

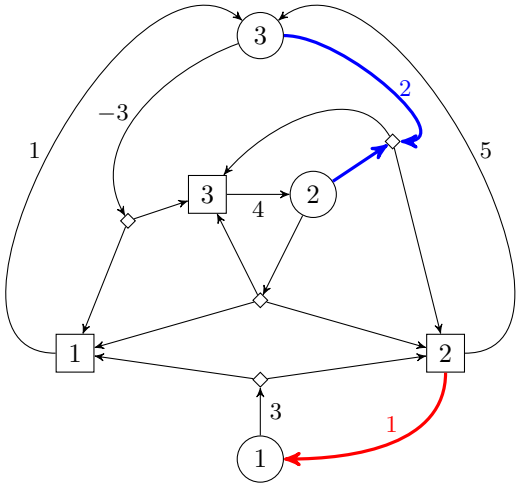
参考文献

1 M. Akian、S. Gaubert、J. Grand-Clément 和 J. Guillaud。熵博弈的算子方法。《计算系统理论》，63:1089–1130，2019。 2 M. Akian、S. Gaubert 和 A. Hochart。零和博弈的遍历性条件。《离散与连续动力系统》，35(9):3901–3931，2015。 3 M. Akian、S. Gaubert 和 R. Nussbaum。非扩张半可微映射不动点的唯一性。《美国数学会汇刊》，368(2):1271–1320，2016年2月。 4 M. Akian、A. Sulem 和 M. I. Taksar。具有交易成本和对数效用的投资组合长期增长率的动态优化。《数学金融》，11(2):153–188，2001年4月。 5 X. Allamigeon、S. Gaubert、R. D. Katz 和 M. Skomra。基于值迭代的通用复杂度界及其在熵博弈中的应用。载于 Mikołaj Bojańczyk、Emanuela Merelli 和 David P. Woodruff 编，《第49届自动机、语言与编程国际研讨会（ICALP 2022）》，莱布尼茨国际信息学会议论文集（LIPIcs）第229卷，第110:1–110:20页，德国达格施图尔，2022年。达格施图尔城堡—莱布尼茨信息学中心。 6 V. Anantharam 和 V. S. Borkar。风险敏感收益的变分公式。《SIAM控制与优化杂志》，55(2):961–988，2017。arXiv:1501.00676。 7 D. Andersson 和 P. B. Miltersen。求解图上的随机博弈的复杂度。载于《第20届算法与计算国际研讨会（ISAAC）论文集》，计算机科学讲义第5878卷，第112–121页。施普林格，2009。 8 E. Asarin、J. Cervelle、A. Degorre、C. Dima、F. Horn 和 V. Kozyakin。熵博弈与矩阵乘法博弈。载于《第33届计算机科学理论问题国际研讨会（STACS）论文集》，LIPIcs第47卷。莱布尼茨国际信息学会议论文集，第11:1–11:14页，瓦登，2016年。达格施图尔城堡—莱布尼茨信息学中心。 9 L. Attia 和 M. Oliu-Barton。随机博弈博弈值的一个公式。《美国国家科学院院刊》，52(116):26435–26443，2019。 10 J. B. Baillon 和 R. E. Bruck。平均非扩张映射渐近正则性的最优速率。载于 K. K. Tan 编，《第二届不动点理论与应用国际会议论文集》，第27–66页。世界科学出版社，1992。 11 T. Bewley 和 E. Kohlberg。随机博弈的渐近理论。《数学运筹学》，1(3):197–208，1976。 12 E. Boros、Kh. Elbassioni、V. Gurvich 和 K. Makino。双人零和平均收益随机博弈的势函数约简算法。《动态博弈与应用》，8(1):22–41，2018年7月。 13 K. Chatterjee 和 R. Ibsen-Jensen。遍历平均收益博弈的复杂度。发表于 ICALP 会议论文集的论文扩展版，2014年。 arXiv:1404.5734。 14 A. Condon。随机博弈的复杂度。《信息与计算》，96(2):203–224，1992。 15 R. L. Dobrushin。非平稳马尔可夫链的中心极限定理。I。《概率论及其应用》，1(1):65–80，1956年1月。 16 R. L. Dobrushin。非平稳马尔可夫链的中心极限定理。II。《概率论及其应用》，1(4):329–383，1956年1月。 17 K. Etessami 和 M. Yannakakis。递归并发随机博弈。《计算机科学逻辑方法》，4(4)，2008年11月。 18 A Federgruen、P.J Schweitzer 和 H.C Tijms。无折扣马尔可夫决策问题中的压缩映射。《数学分析与应用杂志》，65(3):711–730，1978。 19 S. Gaubert 和 J. Gunawardena。齐次单调函数的佩龙–弗罗贝尼乌斯定理。《美国数学会汇刊》，356(12):4931–4950，2004。 20 S. Gaubert 和 N. Stott。用于计算非负矩阵联合谱半径的收敛非线性特征值问题层级。《数学控制与相关领域》，10(3):573–590，2020。

14 相对 Krasnoselskii-Mann 迭代

21 D. Gillette。《零停止概率随机博弈》，第 3 卷。普林斯顿大学出版社，1957。22 K. Arnsfelt Hansen、M. Koucky、N. Lauritzen、P. Bro Miltersen 和 E. P. Tsigaridas。求解随机博弈的精确算法。载于 STOC 2011，2011。23 A. J. Hoffman 和 R. M. Karp。论非终止随机博弈。《管理科学》，12(5):359–370，1966。24 R. A. Howard 和 J. E. Matheson。风险敏感马尔可夫决策过程。《管理科学》，18(7):356–369，1972。25 S. Ishikawa。巴拿赫空间中非扩张映射的不动点与迭代。《美国数学会会刊》，59(1):65–71，1976。26 M. A. Krasnosel'skiĭ。关于逐次逼近法的两点说明。《数学科学进展》，10:123–127，1955。27 B. Lemmens 和 R. Nussbaum。《非线性 Perron-Frobenius 理论》，剑桥数学丛书第 189 卷。剑桥大学出版社，2012 年 5 月。28 T. M. Liggett 和 S. A. Lippman。具有完全信息和时间平均收益的随机博弈。《SIAM 评论》，11:604–607，1969。29 W. R. Mann。迭代中的均值方法。《美国数学会会刊》，4:506–510，1953。30 J.-F. Mertens 和 A. Neyman。随机博弈。《国际博弈论杂志》，10(2):53–66，1981。31 J.-F. Mertens、S. Sorin 和 S. Zamir。《重复博弈》，经济学会专著丛书第 55 卷。剑桥大学出版社，剑桥，2015。32 H.D. Mills。矩阵博弈与线性规划的边际值。载于 H. W. Kuhn 和 A. W. Tucker 编，《线性不等式及相关系统》，《数学研究年刊》第 38 卷，第 183–194 页。普林斯顿大学出版社，1956。33 R.T. Rockafellar 和 R.J.-B. Wets。《变分分析》。施普林格出版社，纽约，1997。34 D. Rosenberg 和 S. Sorin。零和重复博弈的算子方法。《以色列数学杂志》，121(1):221–246，2001。35 U. G. Rothblum。乘法马尔可夫决策链。《运筹学数学》，9(1):6–24，1984。36 U. G. Rothblum 和 P. Whittle。分枝马尔可夫决策链的增长最优性。《运筹学数学》，7(4):582–601，1982。37 L. S. Shapley。随机博弈。《美国国家科学院院刊》，39(10):1095–1100，1953。38 M. Skomra。有限马尔可夫链中平稳分布位数的最优界。2021。39 K. Sladký。关于乘法马尔可夫决策链的动态规划递推，第 216–226 页。施普林格柏林海德堡，柏林，海德堡，1976。doi:10.1007/BFb0120753。40 J. N. Tsitsiklis。判定平均成本马尔可夫决策过程的单链条件是否成立是 NP困难的。《运筹学快报》，35(3):319–323，2007。41 G. Vigeral。一个具有紧行动集且没有渐近值的零和随机博弈。《动态博弈与应用》，3(2):172–186，2013 年 1 月。doi:10.1007/s13235-013-0073-z。42 D.J White。动态规划、马尔可夫链与逐次逼近法。《数学分析与应用杂志》，6(3):373–376，1963。43 W. H. M. Zijm。具有一般非负矩阵的动态规划递推的渐近展开。《优化理论与应用杂志》，54(1):157–191，1987。 doi:10.1007/BF00940410。44 U. Zwick 和 M. Paterson。图上平均收益博弈的复杂性。《理论计算机科学》，158(1–2):343–359，1996。 doi:10.1016/0304-3975(95)00188-3。

图 1 轮流行动随机平均收益博弈示例。Min 状态以方形表示；Max 状态以圆形表示；自然状态以小菱形表示。弧上标出了 Min 向 Max 支付的收益。对于每个自然状态，下一个状态按照后继状态中的均匀分布选取。Min 和 Max 的最优策略以粗体显示（分别为红色和蓝色弧）。



轮流行动随机平均收益博弈示例

图 1 展示了一个轮流行动随机平均收益博弈的示例。该博弈的 Shapley 算子是映射 $T\colon \mathbb{R}^3 \to \mathbb{R}^3$，其定义为

$$
\begin{aligned}
T_1(x) &= 1 + \max\left(2 + \frac{x_2 + x_3}{2}, -3 + \frac{x_1 + x_3}{2}\right),\\
T_2(x) &= \min\left(5 + \max\left(2 + \frac{x_2 + x_3}{2}, -3 + \frac{x_1 + x_3}{2}\right), 1 + 3 + \frac{x_1 + x_2}{2}\right),\\
T_3(x) &= 4 + \max\left(\frac{x_2 + x_3}{2}, \frac{x_1 + x_2 + x_3}{3}\right).
\end{aligned}
$$

第 7 节中定义的单链指数为 $k_{\mathrm{uni}} = 1$。事实上，对于所有策略对 (σ_1, τ_1)，都有 $S_1(\sigma_1,\tau_1) \supset \{1,3\}$，$S_3(\sigma_1,\tau_1) \supset \{2,3\}$，并且若 σ_1 将 Min 的状态 2 送至 Max 的状态 3，则 $S_2(\sigma_1,\tau_1) \supset \{2,3\}$；若 σ_1 将 Min 的状态 2 送至 Max 的状态 1，则 $S_2(\sigma_1,\tau_1) = \{1,2\}$。在所有情况下，对于 $i \neq j$，都有 $S_i(\sigma_1,\tau_1) \cap S_j(\sigma_1,\tau_1) \neq \varnothing$。我们有 $p_{\min} = 1/3$，且 $\theta = p_{\min}/(1 + p_{\min}) = 1/4$。由定理 19 可知，阻尼 Shapley 算子 T_θ 是收缩率为 $3/4$ 的压缩映射。根据定理 11，我们知道遍历特征值问题有解。应用算法 1 后，我们得到 $T(u) = \lambda e + u$，其中 $(-1, -0.5, 0)$，且 $\lambda = 3.75$。经过仅 15 次迭代，便得到上确界范数下精度为 $< 10^{-8}$ 的 u 近似值；相比之下，同样迭代次数下，理论上界给出的精度为 $(3/4)^{15}$，10^{-2} 量级。因此，实际收敛速度可能快于推论 20 所示的速度。由 $T(u) = \lambda e + u$ 可推知，无论初始状态如何，平均收益博弈的博弈值均为 3.75。通过选择使 $T(u)$ 表达式达到最小值或最大值的动作，可以得到两位玩家的最优策略 σ 和 τ。这些最优策略中的非平凡动作如下：

从 Min 的状态 2（右下角的方块）前往 Max 的状态 3（顶层的圆圈）；从 Max 的状态 2（中间层的圆圈）出发；此外，从 Max 的状态 3 前往自然的右上方状态（菱形）。这些动作如图 1 所示。与这些策略相关的随机矩阵 $P^{\sigma,\tau}$ 和支付向量 $r^{\sigma,\tau}$ 为

$$r^{\sigma,\tau} = \begin{pmatrix} 3 \\ 4 \\ 4 \end{pmatrix}, \qquad P^{\sigma,\tau} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix} .$$

矩阵 $P^{\sigma,\tau}$ 的唯一不变测度为 $\pi = (1/4, 1/2, 1/4)$，并且我们有 $\pi r^{\sigma,\tau} = 15/4 = 3.75$，与先前求得的平均收益博弈值一致。

B 观察 5 的证明

令 $[x] \coloneqq x + \mathbb{R}e$ 表示 $\mathbb{R}^n/\mathbb{R}e$ 中 x 的等价类。于是，诱导映射 $T\colon \mathbb{R}^n/\mathbb{R}e \to \mathbb{R}^n/\mathbb{R}e,\ [x] \mapsto [T(x)]$，满足：$T^q$ 是关于 $\|\cdot\|_H$ 的 γ-压缩映射。此外，赋予 $\|\cdot\|_H$ 的 $\mathbb{R}^n/\mathbb{R}e$ 是一个巴拿赫空间。因此，T^q 有唯一不动点 $[u]$。此外，$T^q(T([u])) = T(T^q([u])) = T([u])$，并且根据 T^q 的不动点唯一性，我们可推出 $T([u]) = [u]$。这恰好意味着，对于某个 $\lambda \in \mathbb{R}$，有 $T(u) = \lambda e + u$。

C 定理 6 的推导

Dobrushin 的原始表述见 [15, 第 1.4 节、（1.5”）和（1.19）]，，并在 [16], 中证明了 $\delta(M)$ 是作用于空间 M^T 的算子范数，该空间为 $H = \{x \in \mathbb{R}^n \mid e \cdot x = 0\}$，配备 ℓ_1-范数 $\|\cdot\|_1$。空间 $(\mathbb{R}^n/\mathbb{R}e, \|\cdot\|_{\mathrm{H}})$ 是 $(H, \|\cdot\|_1)$ 的对偶空间。由于两个巴拿赫空间之间的算子及其作用于对偶空间上的伴随算子具有相同的算子范数，因此可得 $\|M\|_{\mathrm{H}} = \delta(M)$。

D 关于随机零和博弈的补充内容

对于任意一对随机策略 $(\alpha, \beta) \in \Pi_{\mathrm{R}}$，我们根据 (2) 定义向量 $r^{\alpha,\beta} = (r^{\alpha_i}_{,i\ \ \beta_i})_{i\in[n]} \in \mathbb{R}^n$ 和随机矩阵 $P^{\alpha,\beta} = (P^{\alpha_i}_{,ij\ \ \beta_i})_{i,j\in[n]}$。我们还定义映射 $T^{\alpha,\beta}$、T^{α} 和 ${}^{\beta}T$，它们从 $\mathbb{R}^n$ 映射到自身，使得对于所有 $v \in \mathbb{R}^n$，都有

$$\begin{aligned} T^{\alpha,\beta}(v) &= r^{\alpha,\beta} + P^{\alpha,\beta} v \\ T^{\alpha}(v)_i &= \sup_{\beta_i \in \Delta_{B(i)}} \left(r_i^{\alpha_i,\beta_i} + \sum_{j\in[n]} P_{ij}^{\alpha_i,\beta_i} v_j \right) \quad \text{for } i \in [n] \ , \\ {}^{\beta}T(v) &= \inf_{\alpha_i \in \Delta_{A(i)}} \left(r_i^{\alpha_i,\beta_i} + \sum_{j\in[n]} P_{ij}^{\alpha_i,\beta_i} v_j \right) \quad \text{for } i \in [n] \ . \end{aligned}$$

那么，对于所有 $v \in \mathbb{R}^n$，存在 $(\alpha, \beta) \in \Pi_{\mathrm{R}}$ 使得 $T(v) = T^{\alpha}(v) = {}^{\beta}T(v) = T^{\alpha,\beta}(v)$。此类策略称为关于 v 的最优策略，其集合记作：

$$\mathrm{Opt}_{\mathrm{Min}}(v) = \{\alpha \in \Pi_{\mathrm{R}}^{\mathrm{Min}} \mid T(v) = T^{\alpha}(v)\} \tag{6a}$$
$$\mathrm{Opt}_{\mathrm{Max}}(v) = \{\beta \in \Pi_{\mathrm{R}}^{\mathrm{Max}} \mid T(x) = {}^{\beta}T(v)\} \ . \tag{6b}$$

▶ 注释27。将上述记号应用于纯策略 α, β，并按照第3.1节将其对应为元素$\sigma \in \prod_{i\in[n]} A(i)$ 和 $\tau \in \prod_{i\in[n]} B(i)$，可得对于所有 $(\sigma, \tau) \in \Pi_{\mathrm{P}}$，以及所有$i, j \in [n]$，均有

$$r_i^{\sigma,\tau} = r_i^{\sigma(i),\tau(i)} \quad \text{and} \quad P_{i,j}^{\sigma,\tau} = P_{i,j}^{\sigma(i),\tau(i)} \ .$$

每一对随机策略 $(\alpha, \beta) \in \Pi_{\mathrm{R}}$ 都会在纯策略对的空间上诱导一个乘积概率测度 $\nu^{\alpha,\beta}(\sigma, \tau) \in \Pi_{\mathrm{P}}$，其定义为 $\nu^{\alpha,\beta} = (\alpha_1 \otimes \beta_1) \otimes \cdots \otimes (\alpha_n \otimes \beta_n)$。换言之，两个玩家所采取的动作 $\sigma(1)\ \tau(1) \ldots, \sigma(n)\ \tau(n)$ 被视为相互独立的随机变量，其概率分布为 $\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n$。于是，下面的恒等式是独立性的结果：

$$r^{\alpha,\beta} = \int_{\Pi_{\mathrm{P}}} r^{\sigma,\tau} d\nu^{\alpha,\beta}(\sigma,\tau) \qquad \text{and} \qquad P^{\alpha,\beta} = \int_{\Pi_{\mathrm{P}}} P^{\sigma,\tau} d\nu^{\alpha,\beta}(\sigma,\tau) \ , \tag{7}$$

这意味着，与随机策略对 (α, β) 相关联的支付向量和转移矩阵，是与纯策略对相关联的支付向量和转移矩阵的凸组合。

▶注释28。根据米尔曼的克莱因–米尔曼定理逆定理，(7) 中的第二个等式意味着，紧凸矩阵集 $\{P^{\alpha,\beta} \mid (\alpha, \beta) \in \Pi_{\mathrm{R}}\}$ 的极点集包含于与纯策略相关联的转移矩阵集$\{P^{\sigma,\tau} \mid (\sigma, \tau) \in \Pi_{\mathrm{P}}\}$之中。

E 定理 11 的证明

令 $\hat{T}(x) := \lim_{s\to\infty} s^{-1}T(sx)$ 表示 Shapley 算子的 衰退函数 T。我们依据 [2] 的定理3.1（在（iii）中使用了 $g = 0$），该定理表明，T 的遍历特征值问题可解，当且仅当 $\hat{T}$ 仅有平凡不动点，也就是说，若 $\hat{T}(\eta) = \eta$ 且 $\eta \in \mathbb{R}^n$，则 $\eta \in \mathbb{R}e$。由（3）可推得

$$[\hat{T}(\eta)]_i = \min_{\alpha_i\in\Delta_{A(i)}} \max_{\beta_i\in\Delta_{B(i)}} \sum_{j\in[n]} P_{ij}^{\alpha_i,\beta_i}\eta_j \ .$$

由于行动空间是紧的，且转移概率关于动作连续，因此我们具有选择性质，这意味着，对于所有 $\eta \in \mathbb{R}^n$，存在随机策略 α, β使得 $\hat{T}(\eta) = P^{\alpha,\beta}\eta$。因此，如果 $\hat{T}(\eta) = \eta$，$\eta = P^{\alpha,\beta}\eta$是 $P^{\alpha,\beta}$的调和向量。由于根据假设，$P^{\alpha,\beta}$ 是单链，调和向量 η必为常数。因此，根据 [2], 的定理3.1，存在 $u \in \mathbb{R}^n$ 和 $\lambda \in \mathbb{R}$，使得$T(u) = \lambda e + u$。随后，我们选择在 $T(u)$ 的表达式中分别达到最小值和最大值的随机策略（见（3））。这些策略在随机平均收益博弈中是最优的。

F 定理 12 的证明

定理 12 的证明依赖于两个初等引理。我们设 $R(x) := x - \mathbf{t}(x)$.

▶ **引理 29.** 对于所有 x、$y \in \mathbb{R}^n$，$\|R(x) - R(y)\|_H = \|x - y\|_H$。此外，对于所有 $x \in \mathbb{R}^n$ 和 $\ell \geqslant 1$，$\|(R \circ \tilde{T})^\ell(x) - (R \circ T)^\ell(x)\|_H \leqslant 2\ell\eta$。

证明。第一个性质显而易见。我们通过对 ℓ 进行归纳法来证明不等式 $\|(R \circ \tilde{T})^\ell(x) - (R \circ T)^\ell(x)\|_H \leqslant \ell\eta$。对于 $\ell = 1$，这由 R 在希尔伯特 $\tilde{}$ $\tilde{}$ 半范数下的非扩张性，以及 $\|T(x) - T(x)\|_H \leqslant 2\|T(x) - T(x)\|_\infty \leqslant 2\eta$ 得出。现在假设 $\|(R \circ$

$\tilde{T})^{\ell}(x) - (R \circ T)^{\ell}(x)\|_H \leqslant 2\ell\eta$。然后，利用 R 和 T 的非扩张性，我们得到 $\|(R \circ \tilde{T})^{\ell+1}(x) - (R \circ T)^{\ell+1}(x)\|_H \leqslant \|(R \circ \tilde{T})^{\ell+1}(x) - (R \circ T) \circ (R \circ \tilde{T}^{\ell}(x)\|_H + \|(R \circ T) \circ (R \circ \tilde{T})^{\ell}(x) - (R \circ T)^{\ell+1}(x)\|_H \leqslant 2\eta + 2\ell\eta = 2(\ell+1)\eta$。◀

▶**引理 30.** 假设 $T : \mathbb{R}^n \to \mathbb{R}^n$ 在希尔伯特半范数下是非扩张的，且 T^q 是一个 γ-压缩映射。令 $x_k := T^k(x_0)$ 满足 $x_0 \in \mathbb{R}^n$。则，$\|x_{k+1} - x_k\|_{\mathrm{H}} \leqslant \gamma^{\lfloor k/q \rfloor}\|T(x_0) - x_0\|_{\mathrm{H}}$。

证明. 对于所有 $k \geqslant 0$ 和 $r \geqslant 0$，我们有 $\|T^{kq+r+1}(x_0) - T^{qk+r}(x_0)\|_{\mathrm{H}} \leqslant \gamma^k\|T^{r+1}(x_0) - T^r(x_0)\|_{\mathrm{H}} \leqslant \|T(x_0) - x_0\|_{\mathrm{H}}$。◀

定理 12 的证明. 假设算法终止。我们有

$$T(x) \leqslant \eta e + \tilde{T}(x) \leqslant x + (\eta + \beta)e \ , \tag{8}$$

类似地，$T(x) \geqslant x + (\alpha - \eta)e$。由于 $\eta \leqslant \epsilon/3$，利用观察 4，这给出了所有初始状态的博弈值的上述界限。此外，由于 $\beta - \alpha = \|x - T(x)\|_H \leqslant \epsilon/3$，区间 $[\alpha - \epsilon/3, \beta + \epsilon/3]$ 的宽度为 $\beta - \alpha + 2\epsilon/3 \leqslant \epsilon$。这证明了该算法的正确性。

注意到，$(R \circ T)^q(x)$ 与 $T^q(x)$ 仅相差一个加法常数。因此，$(R \circ T)^q$ 也是关于希尔伯特半范数的一个 γ-压缩映射。根据引理 30，这意味着序列 $x_k := (R \circ T)^k(0)$ 满足

$$\|x_{k+1} - x_k\|_H \leqslant \gamma^{\lfloor k/q \rfloor}\|T(0)\|_H \ . \tag{9}$$

~~我们将此序列 x_k 与通过使用不精确预言机 T 的算法 1 构造的序列 $\tilde{x}_0 = 0$，$\tilde{x}_{k+1} = R \circ T(\tilde{x}_k)$~~进行比较。我们首先对 $k \geqslant 0$ 使用归纳法证明：

$$\|\tilde{x}_{qk} - x_{qk}\|_H \leqslant 2q\eta/(1-\gamma) \tag{10}$$

当 $k = 0$ 时，该性质显然成立。若 $\|\tilde{x}_{qk} - x_{qk}\|_H \leqslant 2q\eta/(1-\gamma)$，则利用引理 29 及 ~~$\ell = q$~~，可得 $\|\tilde{x}_{qk+q} - x_{qk+q}\|_H = \|(R \circ T)^q(\tilde{x}_{qk}) - (R \circ T)^q(x_{qk})\|_H \leqslant \|(R \circ T)^q(\tilde{x}_{qk}) - (R \circ T)^q(\tilde{x}_{qk})\|_H + \|(R \circ T)^q(\tilde{x}_{qk}) - (R \circ T)^q(x_{qk})\|_H \leqslant 2\eta q + \gamma 2\eta q/(1-\gamma) = 2\eta q/(1-\gamma)$，因为 $(R \circ T)^q$ 在希尔伯特半范数下是一个 γ-压缩映射，由此可得式 (10)。

此外，利用 T 的非扩张性，并利用 (10)，

$$\begin{aligned}
\|\tilde{x}_{qk} - \tilde{T}(\tilde{x}_{qk})\|_{\mathrm{H}} &\leqslant \|\tilde{x}_{qk} - x_{qk}\|_{\mathrm{H}} + \|x_{qk} - T(x_{qk})\|_{\mathrm{H}} \\
&\quad + \|T(x_{qk}) - T(\tilde{x}_{qk})\|_{\mathrm{H}} + \|T(\tilde{x}_{qk}) - \tilde{T}(\tilde{x}_{qk})\|_{\mathrm{H}} \\
&\leqslant \frac{2\eta q}{1-\gamma} + \gamma^k\|T(0)\|_{\mathrm{H}} + \frac{2\eta q}{1-\gamma} + 2\eta \\
&= \eta(2 + \frac{4q}{1-\gamma}) + \gamma^k\|T(0)\|_{\mathrm{H}} \ .
\end{aligned}$$

因此，只要条件 $\eta(2 + \frac{4q}{1-\gamma}) + \gamma^k\|T(0)\|_H \leqslant \epsilon/3$ 满足，算法 1 就会终止。若定理的假设 $\eta(12 + \frac{24q}{1-\gamma}) \leqslant \epsilon$ 成立，则当 $\epsilon/6 + \gamma^k\|T(0)\|_H \leqslant \epsilon/3$ 时，前述条件即得到满足，也就是说，当 $k \geqslant (\log\|T(0)\|_H + \log 6 + |\log \epsilon|)/|\log \gamma|$ 时。◀

G 定理 15 的证明

假设 $T(u) = \lambda e + u$。那么 $T_\theta(u) = (1-\theta)\lambda e + u$。此外，$T_\theta$ 与常数的加法作用可交换，并且关于 $\|\cdot\|_{\mathrm{H}}$ 是非扩张的。因此，T_θ 诱导出一个商。

˜算子 T_θ： $\mathbb{R}^n/\mathbb{R}e \to \mathbb{R}^n/\mathbb{R}e$。此外，$(\mathbb{R}^n/\mathbb{R}e,\ \|\cdot\|_{\mathrm{H}})$ 继承了有限维赋范˜向量空间结构，在该结构上，T_θ 是非扩张的。

令 $\bar{u} = \{u+\alpha e, \alpha \in \mathbb{R}\}$ 为 u 的等价类。则 $\bar{u}$ 是商算子的不动点。记 $\{x_k, k \in \mathbb{N}\}$ 为对 T_θ 应用算法 1 所产生的序列。根据石川定理，$\{\bar{x}_k, k \in \mathbb{N}\}$ 在 $\mathbb{R}^n/\mathbb{R}e$ 中收敛。由于 $\mathbf{t}(x_k) = 0$，这意味着 x^k 在 $\mathbb{R}^n$中收敛。

反之，假设序列 $\{x_k, k \in \mathbb{N}\}$ 在 $\mathbb{R}^n$ 中收敛，记其极限为 u。那么 $\|T(u) - u\|_{\mathrm{H}} = 0$，因此 $T(u) - u \in \mathbb{R}e$，这意味着存在 $\lambda \in \mathbb{R}$ 使得 $T(u) = \lambda e + u$。 ◀

H 定理 17 和命题 18 的证明

定理 17 的证明。 考虑一列纯策略对 $\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k, \ldots$。对所有 $i \in [n]$，为简化记号，定义 $S_i(k) := S_i(\sigma_1, \tau_1, \ldots, \sigma_k, \tau_k) = \{j \ [Q^{\sigma_1,\tau_1} \ldots Q^{\sigma_k,\tau_k}]_{ij} > 0\}$。由于所有矩阵 $Q^{\sigma,\tau}$ 的对角系数都严格为正，因此有 $S_i(k) \subset S_i(k+1)$。

现在让我们考虑 $[n]$ 中两个不同的元素 i_1，i_2。反设 $S_{i_1}(n) \cap S_{i_2}(n) = \varnothing$。那么，对于所有 $k < n$，$S_{i_1}(k) \cap S_{i_2}(k) = \varnothing$。考虑由 $S_{i_1}(1) \cup S_{i_2}(1)$、$S_{i_1}(2) \cup S_{i_2}(2), \cdots$ 以及 $S_{i_1}(n) \cup S_{i_2}(n)$ 构成的 $[n]$ 的子集序列。由于这是一个非递增序列，且 $S_{i_1}(1) \cup S_{i_2}(1)$ 的基数大于 2，因此存在一个整数 $m < n$，使得：$S_{i_1}(m-1) = S_{i_1}(m)$ 且 $S_{i_2}(m-1) = S_{i_2}(m)$。

因此，$S_{i_1}(m-1)$ 和 $S_{i_2}(m-1)$ 在矩阵 Q^{σ_m}下$^{\tau_m}$的作用下是闭合的。由于该矩阵是单链，在其作用下闭合的每个集合都必须包含该矩阵的最终类。由此可知，$S_{i_1}(m-1) \cap S_{i_2}(m-1)$ 包含这个最终类，这与$S_{i_1}(m-1) \cap S_{i_2}(m-1) = \varnothing$矛盾。

令 $k \leqslant n$ 表示满足以下条件的最小整数：对所有 i_1，i_2，$S_{i_1}(k) \cap S_{i_2}(k) \neq \varnothing$ 均成立；并令 $Q := Q^{\sigma_1,\tau_1} \cdots Q^{\sigma_k,\tau_k}$。于是，$\min_{i<j} \sum_l \min(Q_{i,l}, Q_{j,l}) \geqslant \min_{i<j} \sum_{l \in F} \min(Q_{i,l}, Q_{j,l}) > 0$，因此 $\min_{i<j} \sum_l \min(Q_{i,l}, Q_{j,l}) \geqslant \theta^k$，并利用定理 6，

$$\|Q^{\sigma_1,\tau_1} \ldots Q^{\sigma_k,\tau_k}\|_{\mathrm{H}} = 1 - \min_{i<j} \sum_l \min(Q_{i,l}, Q_{j,l}) \leqslant 1 - \theta^k \ .$$

◀

命题 18 的证明。 在不可约 $n \times n$随机矩阵的作用下，唯一的不变集合是 $[n]$。因此，定理 17 的证明表明，在这种情况下，存在一个指标 $k \leqslant n$，使得矩阵 $Q^{\sigma_1,\tau_1} \cdots Q^{\sigma_k,\tau_k}$ 为正。由于该矩阵的所有元素均有下界 θ^k，应用定理 6，即可推出命题 18。 ◀

► 注 31。定理 17 的证明借鉴了 [18] 定理 4 的 (4) $\Rightarrow$(S) 的证明思路，并修正了其中的一个疏漏。事实上，[18] 的作者更一般地考虑了两列不同的策略序列。沿用上文的记法，令 $S_{i_1}(k)$ 表示从 i_1 出发，应用第一列策略序列时至多经过 k 步可到达的状态集合；令 $W_{i_2}(k)$ 表示从 i_2 出发并应用第二列策略序列时对应的集合。[18] 中证明了，存在一个整数 k，使得 $S_{i_1}(k)$ 和 $W_{i_2}(k)$ 都是闭集，其中“闭”指的是我们所说的在随机矩阵作用下不变。然而，它们是在不同随机矩阵作用下的闭集，因此我们无法得出 $S_{i_1}(k) \cap W_{i_2}(k) \neq \varnothing$，尽管 [18] 的定理 4 证明最后一行声称如此。不过，我们将在此采用一种新方法，利用变分分析的性质来计算 Shapley 算子的收缩率，

要应用我们的方法，只需考虑一列策略序列，而非两列，并比较 $S_{i_1}(k)$ 与 $S_{i_2}(k)$；在这个特殊情形下，上述论证是成立的。

▶ 注 32. Tsitsiklis 指出， [40] 在单人情形下，判断所有策略是否都是单链是 co-NP 完全问题。

I 定理 19 的证明

回顾一下，映射 $T\colon \mathbb{R}^n \to \mathbb{R}^p$ 在点 $x \in \mathbb{R}^n$ 处的（单侧）方向导数 是映射 $T'_x\colon \mathbb{R}^n \to \mathbb{R}^p$，定义如下:

$$T'_x(h) \coloneqq \lim_{s\to 0^+} s^{-1}(T(x+sh) - T(x)) \ . \tag{11}$$

以下是 Mills 的一个经典结果的推广， [32],该结果由 [34] 推出。

▶ **命题 33** （见 [34, 命题 4]）。 并发博弈的 *Shapley* 算子确实存在方向导数，其表达式为:

$$T'_x(h) = \sup_{\alpha\in \mathrm{Opt}_{\mathrm{Max}}(v)} \inf_{\beta\in \mathrm{Opt}_{\mathrm{Min}}(v)} P^{\alpha,\beta}h, \qquad \forall h \in \mathbb{R}^n, \forall x \in \mathbb{R}^n \ .$$

此外，对于所有 x、 h，都存在随机策略 α 和 β，使得 $T'_x(h) = P^{\alpha,\ \beta}h$。

然而，我们需要一种比仅仅存在方向导数更强的概念。如果式 (11) 中的极限在 h 上一致，其中 h 遍历单位球面，那么称映射 T 在点 x 处半可微。等价地，要求存在一个一阶“类似泰勒”的扩展 $T(x+h) = T(x) + G_x(h) + o(\|h\|)$，其中 G_x 是一个连续且 1 次正齐次的映射。于是， $G_x = T'_x(h)$。换言之，半可微类似于可微，但半微分映射 T'_x 可以是非线性的。关于这一概念的背景介绍，请参阅 [33] 。如果 T 是 Lipschitz 连续的，且如果 T 在点 x 处沿所有方向都存在方向导数，那么它在该点处半可微，参见例如 [3, 引理 3.2]。由命题 33 可知，并发博弈的 Shapley 算子在每一点处都半可微。半可微映射也满足链式法则。如果 $F\colon \mathbb{R}^n \to \mathbb{R}^p$ 在点 $x \in \mathbb{R}^n$ 处半可微，且 $G\colon \mathbb{R}^p \to \mathbb{R}^q$ 在点 $F(x)$ 处半可微，那么 $G \circ F$ 在点 x 处半可微，其中

$$(G \circ F)'_x(h) = G'_{F(x)} \circ F'_x(h) \ . \tag{12}$$

参见例如 [3, 引理 3.4]。

现在，对于 x， $y \in \mathbb{R}^n$，记

$$f : t \to T^k(x + t(y - x))$$

根据拉德马赫定理，Lipschitz 连续映射 $\mathbb{R} \to \mathbb{R}^n$ 几乎处处可微，且绝对连续。因此

$$T^k(y) - T^k(x) = \int_0^1 f'(s)ds \ , \tag{13}$$

其中， $f'(s)$ 表示 f 在点 s处的导数，该导数几乎处处有定义。当该导数存在时，它与 f'_s (1) 一致，即映射 f 在点 s沿方向 $1 \in \mathbb{R}$的半微分。由此，我们得到以下结果。

▶ **引理 34**。 对于所有 $s \in (0,\ 1)$，存在随机策略$(\alpha_s^1, \beta_s^1) \ldots,\ (\alpha_s^k, \beta_s^k) \in \Pi_{\mathrm{R}}$使得：

$$f'(s) = P^{\alpha_s^1,\beta_s^1} \ldots P^{\alpha_s^k,\beta_s^k}(y-x)\ . \tag{14}$$

证明。 利用链式法则 (12)，我们得到，对于所有 $s \in (0,1)$,

$$f'(s) = T'_{T^{k-1}(x+s(y-x))} \circ \cdots \circ T'_{x+s(y-x)}(y-x)$$

接着，结论由命题 33 中的选择结果得出。 ◀

考虑以下量

$$D_k := \max_{(\alpha_1,\beta_1),\ldots,(\alpha_k,\beta_k)\in\Pi_{\mathrm{R}}} \|P^{\alpha_1\beta_1} \ldots P^{\alpha_k,\beta_k}\|_{\mathrm{H}}\ .$$

注意，最大值是在（随机化策略的）紧集的笛卡尔积上取得的，并且被最大化的表达式关于 $(\alpha_1, \beta_1) \ldots, (\alpha_k, \beta_k)$ 连续。因此，最大值可以取到。

▶ **引理 35**。 对于所有 x、 $y \in \mathbb{R}^n$ 和 $k \in \mathbb{N}$

$$\|T^k(x) - T^k(y)\|_{\mathrm{H}} \leqslant D_k \|y-x\|_{\mathrm{H}}$$

证明。 由引理 34 和式 (13) 可得

$$\|T^k(x) - T^k(y)\|_{\mathrm{H}} \leqslant \int_0^1 \|f'(s)\|_{\mathrm{H}} ds \leqslant D_k \|y-x\|_{\mathrm{H}}\ .$$

◀

下面的引理表明， D_k 中的最大值实际上由纯 策略序列达到。

▶ **引理 36**。 设 $((\alpha_1, \beta_1) \ldots, \ldots, (\alpha_k, \beta_k)) \in \Pi_{\mathrm{R}}^k$。 则，

$$\|P^{\alpha_1,\beta_1} \ldots P^{\alpha_k,\beta_k}\|_{\mathrm{H}} \leqslant \max_{(\sigma_1,\tau_1),\ldots,(\sigma_k,\tau_k)\in\Pi_{\mathrm{P}}^k} \|P^{(\sigma_1,\tau_1)} \ldots P^{(\sigma_k,\tau_k)}\|_{\mathrm{H}}\ . \tag{15}$$

证明。 利用积分表示式 (7)，可得

$$P^{\alpha_1,\beta_1} \ldots P^{\alpha_k,\beta_k} = \int_{\Pi_{\mathrm{P}}^k} P^{\sigma_1,\tau_1} \ldots P^{\sigma_k,\tau_k} d\nu^{\alpha_1,\beta_1}(\sigma_1,\tau_1) \ldots d\nu^{\alpha_k,\beta_k}(\sigma_k,\tau_k)\ .$$

由于 $\nu^{\alpha_1,\ \beta_1} \otimes \cdots \otimes \nu^{\alpha_k,\ \beta_k}$是 Π_{P}^k上的概率测度，利用算子范数$\|\cdot\|_H$的凸性，可得式 (15)。此外，紧性论证表明，式 (15) 中的最大值可以达到。

◀

现在，结合引理 36、引理 35、定理 17 和命题 18，即可得到定理 19。◀

▶ 注 37。结合引理 35 和引理 36，我们可以看出，并发博弈的 Shapley 算子 T 在希尔伯特半范数下是一个压缩映射，其收敛率上界为

$$\max_{(\sigma_1,\tau_1),\ldots,(\sigma_k,\tau_k)\in\Pi_{\mathrm{P}}^k} \|P^{(\sigma_1,\tau_1)} \ldots P^{(\sigma_k,\tau_k)}\|_{\mathrm{H}}\ . \tag{16}$$

这应与 [18], 公式 (3.6) 进行比较，该公式适用于单人情形。与 (16) 不同，公式 (3.6) 涉及由单个玩家的策略对产生的、大小为 $2n \times n$ 的矩形矩阵的 Dobrushin 遍历系数。上述证明基于不同的原理（半可微性质和范数的凸性），得出了改进的估计——表明无需考虑同一玩家的策略对。

J 推论 20 的证明

由定理 19 可知， T_θ^k是希尔伯特半范数下收缩率为 $1-\theta^k$ 的压缩映射。于是，根据定理 12，由对数函数的凹性可得迭代次数的上界；对所有 $0<\delta<1$，有 $|\log(1-\delta)|^{-1}\leqslant\delta^{-1}$，由$\|T(0)\|_{\mathrm{H}}\leqslant 2\|r\|_\infty$，并且由于 $\chi(T_\theta)=(1-\theta)\chi(T)$，我们需要以 $(1-\theta)\epsilon$ 的精度应用算法 1 于 T_θ，从而使原博弈的博弈值达到 ϵ 的最终精度。◀

K 定理 23 的证明

设 x^*,α,β 为对算子 T_θ 应用算法 1 后返回的向量和标量。设 μ 为算子 T 对应的轮流行动平均收益博弈的博弈值。由于该博弈是单链的，因此存在一个向量 $u\in\mathbb{R}^n$，使得 $T(u)=\mu e+u$，从而 $T_\theta(u)=\lambda e+u$，其中 $\lambda=(1-\theta)\mu$。于是，根据 (8)，

$$T_\theta(x^*)\leqslant(\beta+\eta)e+x^*\ ,$$

和 $\beta+\eta\leqslant\lambda+\epsilon$。

我们将使用以下引理。

▶**引理38**。 设 σ',τ'和 σ'',τ''表示两对策略，它们产生不同的博弈值 μ',μ''。则，

$$|\mu'-\mu''|>(nM^{n-1})^{-2}\ . \tag{17}$$

证明。 记 π 为矩阵 $P^{\sigma',\tau'}$的唯一不变测度，则有 $\mu=_{\pi r^{\sigma',\tau'}}$。由引理22并利用假设21可得， μ 是一个有理数，其分母至多为 nM^{n-1}。注意到，两个分母至多为 q 的有理数之差至少为 $1/q(q-1)>1/q^2$，于是得到（17）。◀

现在，根据 σ^*的定义，对玩家 Max 的所有纯策略 τ，

$$r^{\sigma^*\tau}+\theta x^*+(1-\theta)P^{\sigma^*\tau}x^*\leqslant(\beta+\eta)e+x^*\ . \tag{18}$$

设 μ'表示与策略对 σ^*,τ, $\lambda'=(1-\theta)\mu'$相关联的平均收益，并设 π'表示 $P^{\sigma^*\tau}$的唯一不变测度，因此 $\pi'P^{\sigma^*\tau}=\pi'$且 $\pi'e=1$。在（18）左乘 π'，并注意到 x^* 的贡献相互抵消，可推得 $\lambda'=(1-\theta)\mu'\leqslant\beta+\eta\leqslant\lambda+\epsilon=(1-\theta)\mu+\epsilon$。由于我们选择了 $\epsilon=(1-\theta)(nM^{n-1})^{-2}$，由（17）可知 $\mu'\leqslant\mu$，这表明 σ^* 是 Min 的最优策略。对偶论证可证明 τ^* 是玩家 Max 的最优策略。

最后，结合定理 12 和定理 19，可得迭代次数的估计式 (5)。

L 定理 24 的证明

我们先证明以下引理。

▶ **引理 39.** 令 A_l 表示该博弈的 l-歧义度，并令 $A=\max_{1\leqslant l\leqslant k_{\mathrm{irr}}}A_l^{1/l}$。那么，$A_l\leqslant n^{l-1}W^l$，且 $A\leqslant n^{1-1/k}W$，其中 $k:=k_{\mathrm{irr}}$。此外，对于所有 $s>0$，以及所有策略 $\sigma_1,\tau_1,\ldots,\sigma_k,\tau_k$，矩阵的各项

$$\mathcal{M}(s):=(sI+M^{\sigma_1,\tau_1})\ldots(sI+M^{\sigma_k,\tau_k})$$

不超过 $(s+A)^k$。

证明。 令 J 表示 $n \times n$ 其所有元素均为 1 的矩阵，并注意到$J^2 = nJ$。于是，
$M^{\sigma_1\tau_1} \dots M^{\sigma_l\tau_l} \leqslant (WJ)^l = n^{l-1}JW^l$。由此可得 $\mathcal{A}_l \leqslant n^{l-1}W^l$ 以及$\mathcal{A} \leqslant n^{1-1/k}W$。展开定义 $\mathcal{M}(s)$所得的乘积，可得

$$\begin{aligned}(\mathcal{M}(s))_{d,d'} &= \sum_{0\leqslant l\leqslant k} \sum_{i_1<i_2<\dots<i_l} s^{k-l}[M^{\sigma_{i_1},\tau_{i_1}} \dots M^{\sigma_{i_l},\tau_{i_l}}]_{d,d'} \\ &\leqslant \sum_{0\leqslant l\leqslant k} \binom{k}{l} s^{k-l}\mathcal{A}_l \leqslant \sum_{0\leqslant l\leqslant k} \binom{k}{l} s^{k-l}\mathcal{A}^l = (s+\mathcal{A})^k \ .\end{aligned}$$

◀

定义如下会很方便:

$$[F_{\mathrm{m},\vartheta}(x)]_d = \min_{t\in\mathcal{V}_T,(d,t)\in\mathcal{E}} \max_{p\in\mathcal{V}_P,(t,p)\in\mathcal{E}} \big(\vartheta x_d + \sum_{d'\in\mathcal{V}D,(p,d')\in\mathcal{E}} m_{p,d'}x_{d'}\big), \qquad \forall d\in\mathcal{V}_D \ ,$$

使得 $T_{\mathrm{m},\ \vartheta} = \mathsf{log}\circ F_{\mathrm{m},\ \vartheta}\circ \mathsf{exp}$，其中 log 和 exp 运算逐项进行。由此可得 $T^k{}_{\mathrm{m}\ \vartheta} = \mathsf{log}\circ F^k_{\mathrm{m},\ \vartheta}\circ \mathsf{exp}$。

首先注意到，$F_{\mathrm{m},\ \vartheta}$ 和 $T_{\mathrm{m},\ \vartheta}$ 均为半可微。设 u 和 h 是 $\mathbb{R}^n$中的两个向量。由于 $\mathsf{log}'_u(h) = \mathsf{diag}(u)^{-1}h$ 且 $\mathsf{exp}'_u(h) = \mathsf{diag}(e^u)h$，因此:

$$(T^k_{\mathrm{m},\vartheta})'_u(h) = \mathrm{diag}(F_{\mathrm{m},\vartheta}(e^u))^{-1}[(F^k_{\mathrm{m},\vartheta})'_{e^u}\,\mathrm{diag}(e^u)h]$$

此外，我们知道，对于所有向量 v 和 h':

$$(F^k_{\mathrm{m},\vartheta})'_v h' = (F'_{\mathrm{m},\vartheta})_{F^{k-1}_{\mathrm{m},\vartheta}(v)} \circ \dots \circ (F'_{\mathrm{m},\vartheta})_v(h')$$

此外，对于所有 v、h'，都存在 σ和 τ使得：$(F'_{\mathrm{m}\ \vartheta})_v(h') = (mI + M^{\sigma,\ \tau})h'$。因此，存在 σ_1，τ_1，$\cdots$ σ_k，τ_k 使得:

$$(F^k_{\mathrm{m},\vartheta})'_{e^u}\,\mathrm{diag}(e^u)h = (\vartheta I + M^{\sigma_1,\tau_1})\cdots(\vartheta I + M^{\sigma_k,\tau_k})\,\mathrm{diag}(e^u)h$$

我们记 $\mathcal{M} = (\vartheta I + M^{\sigma_1,\tau_1})\cdots(\vartheta I + M^{\sigma_k,\tau_k})$，那么：$(T^k{}_{\mathrm{m}\ \vartheta})'_u(h) = \mathsf{diag}(F_{\mathrm{m},\vartheta}(e^u))^{-1}[\mathcal{M}\,\mathsf{diag}(e^u)h]$。此外，$F^k{}_{\mathrm{m}\ \vartheta}(e^u) = \mathcal{M}e^u$。因此:

$$(T^k_{\mathrm{m},\vartheta})'_u(h) = \mathrm{diag}(\mathcal{M}(e^u))^{-1}[\mathcal{M}\,\mathrm{diag}(e^u)h] \ .$$

回顾一下，希尔伯特射影度量定义为

$$d_{\mathrm{H}}(x,y) \coloneqq \|\log x - \log y\|_H, \qquad \forall x,y\in\mathbb{R}^n_{>0} \ .$$

由于矩阵 $\mathcal{M}$ 的元素均为正，Birkhoff-Hopf定理给出

$$d_{\mathrm{H}}(\mathcal{M}x,\mathcal{M}y) \leqslant \tanh(\frac{\Delta}{4})d_{\mathrm{H}}(x,y) \ ,$$

其中

$$\Delta = \max_{i,j} \|\log \mathcal{M}_{i\cdot} - \log \mathcal{M}_{j\cdot}\|_H \ ,$$

是集合 $\mathcal{M}(\mathbb{R}^n_{>0})$ 在 Hilbert射影度量下的直径，更多信息请参见 [27,附录 A]。因此:

$$\|\log(\mathcal{M}e^u) - \log(\mathcal{M}e^v)\|_{\mathrm{H}} \leqslant \tanh(\frac{\Delta}{4})\|u-v\|_{\mathrm{H}}$$

利用以下事实

$$(\log\circ\mathcal{M}\circ\exp)'_u h = \lim_{t\to 0}\frac{\log\circ\mathcal{M}\circ\exp(u+th)-\log\circ\mathcal{M}\circ\exp(u)}{t}$$

我们有:

$$\|(\log\circ\mathcal{M}\circ\exp)'_u h\|_{\mathrm{H}} \leqslant \tanh(\frac{\Delta}{4})\|h\|_{\mathrm{H}}.$$

因此,

$$\|(T^k_{\mathrm{m},\vartheta})'_u(h)\|_{\mathrm{H}} \leqslant \tanh(\frac{\Delta}{4})\|h\|_{\mathrm{H}} \ .$$

由于 $\mathcal{M}$ 的每个非零元素均不小于 $\vartheta = m$，因此$\mathcal{M}$ 的最小元素不小于 ϑ^k。此外，根据引理 39，$\mathcal{M}$ 的每个元素均不大于 $(\vartheta+\mathcal{A})^k$。由此可得

$$e^{\Delta} = \max_{i,j,i',j'}\frac{\mathcal{M}_{i,j}\mathcal{M}_{i',j'}}{\mathcal{M}_{i',j}\mathcal{M}_{i,j'}} \leqslant (1+\mathcal{A}/\vartheta)^{2k} = \bar{\mathcal{M}}^2 \ .$$

那么 $\tanh\frac{\Delta}{4} = \frac{e^{\Delta/2}-1}{e^{\Delta/2}+1} \leqslant \frac{\bar{\mathcal{M}}-1}{\bar{\mathcal{M}}+1}$。因此,

$$\|T^k_{\mathrm{m},\vartheta}(x)-T^k_{\mathrm{m},\vartheta}(y)\|_{\mathrm{H}} \leqslant \frac{\bar{\mathcal{M}}-1}{\bar{\mathcal{M}}+1}\|x-y\|_{\mathrm{H}}.$$

▶ 备注 40。我们注意到，算法 1 可以通过考虑迭代 $\tilde{y} := F_{\mathrm{m},\vartheta}(y)/\mathrm{t}(F_{\mathrm{m},\vartheta}(y))$ 来实现，其中 $F_{\mathrm{m},\vartheta}$ 是算子 $F_{\mathrm{m},\vartheta}$ 的近似，使得 $\tilde{\mathrm{k}}\log F_{\mathrm{m},\vartheta}(x) - \log F_{\mathrm{m},\vartheta}(x)\mathrm{k}_\infty \leqslant \eta$，从而避免在每次迭代中不必要地计算 log 和 exp。然后，应将停止条件替换为 $\mathrm{k}\log(y) - \log(F_{\mathrm{m},\vartheta}(y))\mathrm{k}_{\mathrm{H}} \leqslant \epsilon$。

M 定理 26 的证明

我们采用与定理 23 的证明相同的方法，但现在利用定理 25 的分离界。

ˆ 如果一个熵博弈是不可约的，那么，衰退函数 T 只有属于对角线 $\mathbb{R}e$的不动点，因此，由 [19, 定理 9 和 13] 可知，特征值问题$T(u) = \nu e + u$（其中 $\nu\in\mathbb{R}$ 且 $u\in\mathbb{R}^n$）有解。此外，$\nu = \log\mu$，其中 μ 是该熵博弈的博弈值。于是，特征值问题$T_{\mathrm{m},\vartheta}(u) = \lambda e + u$ 的非线性特征值解为 $\log(m+\mu)$。此外，将算法 1 应用于算子$T_{\mathrm{m},\vartheta}$，会得到一个向量 x^* 和标量 α,β，使得

$$(\alpha-\eta)e + x^* \leqslant T_{\mathrm{m},\vartheta}(x^*) \leqslant (\beta+\eta)e + x^*$$

其中

$$\log(\underline{m}+\mu) - \epsilon \leqslant \alpha-\eta \leqslant \beta+\eta \leqslant \log(\underline{m}+\mu)+\epsilon$$

现在我们注意到，如果在一个熵博弈中，两对策略产生不同的博弈值 μ' 和 μ''，那么，应用带精确余项的泰勒公式，并结合定理 25，

$$|\log(\underline{m}+\mu') - \log(\underline{m}+\mu'')| \geqslant ((\underline{m}+W)\nu_n)^{-1} \ . \tag{19}$$

令 σ^* 表示通过对每个节点 $d \in \mathcal{V}_D$ 在表达式中选择一个最小化动作而得到的 Despot 策略

$$\min_{t\in\mathcal{V}_T,(d,t)\in\mathcal{E}} \max_{p\in\mathcal{V}_P,(t,p)\in\mathcal{E}} \big(\vartheta x_d^* + \sum_{d'\in\mathcal{V}_D,(p,d')\in\mathcal{E}} m_{p,d'} \exp(x_{d'}^*)\big),$$

类似地，令 τ^* 表示这样一种执政官策略：对每个节点 $t \in \mathcal{V}_T$，从上述内层表达式中选取一个最大化动作。

对于所有策略对 σ、τ，定义映射

$$T^{\sigma,\tau}_{\mathrm{m},\vartheta}(x) := \log\big((\vartheta I + M^{\sigma,\tau})\exp(x)\big) \ .$$

由 $T_{\mathrm{m}、\ \vartheta}(x^*) \leqslant (\beta+\eta)e + x^*$，可推出 $T^{\sigma^*、\tau\,\mathrm{m}\ \vartheta}(x^*) \leqslant (\beta+\eta)e + x^*$；若 μ' 是策略对（σ^*、τ）的博弈值，则这意味着 $\log(m+\mu') \leqslant \beta+\eta \leqslant \log(m+\mu) + \epsilon$。由于我们选择了 $\epsilon = (1+(m+W)\nu_n)^{-1} < ((m+W)\nu_n)^{-1}$，利用公式（19），可推出 σ^* 保证 Despot 获得不超过 μ 的博弈值。对偶空间中的论证表明，τ^* 保证 Tribune 获得至少为 μ 的博弈值。因此，σ^* 和 τ^* 是最优策略。

最后，定理 26 中给出的迭代次数是通过应用定理 12 得到的，其中 $\gamma = (\bar{\mathcal{M}}-1)/(\bar{\mathcal{M}}+1)$，注意到 $(\bar{\mathcal{M}}-1)/(\bar{\mathcal{M}}+1) \leqslant 1/(1+2\,\bar{\mathcal{M}}^{-1})$，并利用对数映射的凹性推得 $1/|\log\,\gamma| \leqslant 1/\log(1+2\,\bar{\mathcal{M}}^{-1}) \leqslant 1/(2\,\bar{\mathcal{M}}^{-1}) = \bar{\mathcal{M}}/2$。