arXiv:2611.31149v1 [cs.GT] 23 Nov 2026

矩阵博弈中的最优遗憾下界

Taira Tsuchiya*

摘要

我们研究双人零和博弈（也称为矩阵博弈）中的学习问题。在这类博弈中，两位分别拥有 m 和 n 个行动的玩家，在 T 个轮次中反复选择混合策略。基于乐观 Hedge 的学习动态可将遗憾保证为 $O(\log(mn))$；如果每位玩家都知道对手的行动数量，则可改进为 $O(\sqrt{\log\, m \log\, n})$。然而，此前仅针对特定学习动态得到了相匹配的遗憾下界。我们证明，对于每一种确定性学习动态，都存在一个收益矩阵，使得在 m 和 n 的较大范围内，最大个体遗憾为 $\Omega(\sqrt{\log\, m \log\, n})$。结合上界 $O(\sqrt{\log\, m \log\, n})$，该下界表明，在这一范围内，矩阵博弈中确定性学习动态的极小极大遗憾为 $\Theta(\sqrt{\log\, m \log\, n})$。对于所有 m、n 和 T，我们还证明了下界 $\Omega(\min\{T, \log \min\{m, n\}\})$。即使每位玩家还能观察对手的策略和反馈，这些下界仍然成立。我们的主要技术贡献是对收益矩阵进行自适应构造。在运行两位玩家的算法时，我们在每一轮加入一个秩一矩阵，使前几轮的反馈保持不变，同时提高两位玩家遗憾之和的一个下界。作为推论，我们还得到了平均策略收敛至纳什均衡的速率下界，以及确定性算法为计算近似纳什均衡而对矩阵-向量乘积所需查询复杂度的下界。

1 引言

我们研究双人零和博弈（也称为矩阵博弈）中的学习问题。在这个问题中，两位玩家反复进行由一个未知 $m \times n$ 收益矩阵给定的博弈。在每一 T 轮次中，两位玩家分别在 m 和 n 行动上选择混合策略，随后每位玩家都会观察到自己的每个行动对阵对手策略时的期望收益，并将其作为反馈。利用这些反馈，每位玩家更新自己的策略，目标是将遗憾降至最低。如果两位玩家都采用无遗憾学习算法，那么两位玩家的平均策略将以由其遗憾决定的速度收敛到纳什均衡（Cesa-Bianchi 和 Lugosi，1999；Freund 和 Schapire，2006）。

一种标准的无遗憾算法是 Hedge（Littlestone 和 Warmuth，1994；Freund 和 Schapire，1997），它保证玩家的遗憾为 $O(\sqrt{T}$ 对数 $m)$

m 针对任意对手的行动。如果双方都使用合适的学习算法，这个最坏情况界可以显著改善。Daskalakis 等人 (2011) 提出了首个无遗憾学习算法：当双方都使用该算法时，其遗憾关于 T 呈对数增长，关于 m 和 n 呈多重对数增长。随后，研究表明，基于乐观 Hedge（Rakhlin 和 Sridharan, 2013; Syrgkanis 等人, 2015）且采用常数学习率的学习动态，可以保证双方的遗憾均为 $O(\log(mn))$，与 T 无关。这意味着，平均策略以 $O(\log(mn)/T)$ 的速度收敛到纳什均衡。如果每位玩家还知道对手的行动数量，遗憾还可以进一步降低至 $O(\sqrt{\log\, m \log\, n})$（Tsuchiya, 2026a）。

*东京大学、大阪大学和理化学研究所；tsuchiya@mist.i.u-tokyo.ac.j

与这些上界相比，矩阵博弈中的遗憾下界远未得到充分理解。现有的大多数遗憾下界仅针对特定学习算法，例如 Hedge（Syrgkanis 等人，2015年；Chen 和 Peng，2020）。对于基于乐观 Hedge 的学习动态，已知上述上界中对行动数量 m 和 n 的依赖是紧的（Tsuchiya，2026a），但这并不排除其他学习动态能够取得更小遗憾的可能性。对于强非耦合学习动态，Daskalakis 等人 (2011年) 给出了一个 $\Omega(1/T)$ 的下界，约束了平均策略收敛到纳什均衡的速度，但该下界与 m 和 n 无关。这引出了一个自然的问题：

矩阵博弈中的极小极大最优遗憾是多少？

本文的贡献。 我们通过证明与上述上界相匹配的遗憾下界，回答了确定性学习动态下的这一问题。即使每位玩家还能观察对手在此前所有轮次中的策略和反馈，我们的下界仍然成立。

我们首先证明，对于每一种确定性学习动态，都存在一个收益矩阵，使得最大个体遗憾为 $\Omega(\min\{T, \log \min\{m,n\}\})$（推论 1）。结合乐观 Hedge 的上界可知，当行动数量相同时，确定性学习动态下最大个体遗憾的极小极大值为 $\Theta(\min\{T, \log\ m\})$，即 $m = n$。基于这一遗憾下界的构造，我们进一步将其加强为 $\Omega(\sqrt{\log\ m \log\ n})$，其中 $m \geq n$，$\log\ m \leq n^\alpha$，对于常数 ~~$\alpha \in (0, 1)$~~，且 $T \geq \log_2 m$（推论 3）。对于知道对手行动数量的玩家而言，这一下界与上界 $O(\sqrt{\log\ m \log\ n})$ 相匹配（Tsuchiya，2026a）。条件 $\log m \leq n^\alpha$ 不能完全去除，因为对于 $n = 2$，存在一种确定性学习动态，使其对每个 m 的遗憾都被一个常数所界定（附录 A 中的命题1）。

我们的主要技术贡献是构造这些下界中的收益矩阵。由于学习动态所选的策略会通过反馈依赖于收益矩阵，因此我们在前 k 轮运行两位玩家的算法时，自适应地构造矩阵。在每一轮 $t = 1, \ldots, k$，在算法选定策略 x_t 和 y_t 后，我们定义

$$A_t = A_{t-1} + \frac{1}{3}(u_t - u_{t-1})v_t^\top, \tag{1}$$

其中 $A_0 = 0$ 是零矩阵，u_0 是全 1 向量，而 $u_t \in [-1, 1]^m$ 和 $v_t \in [-1, 1]^n$ 分别是与 $x_1, \ldots, x_t$ 和 $y_1, \ldots, y_t$ 正交的向量。算法根据矩阵 A_t 接收第 t 轮的反馈，而收益矩阵为 $A = A_k$。根据正交性，后续轮次添加的项不会改变这一反馈，因此在矩阵 A 上运行时，确定性算法会选出相同的策略。我们选择这些向量，使得 A 的元素属于 $[-1, 1]$，并且对于某些行动 i_* 和 j_*，坐标 $u_t(i_*)$ 和 $v_t(j_*)$ 直到第 k 轮都等于 1。然后，在矩阵 A 上，行动 i_* 对 y 玩家所采用的每种策略的收益均为零，而每种策略 x_t 对行动 j_* 的收益均为 $-1/3$。这一构造使我们能够对两位玩家在 H 轮后的遗憾总和给出下界 $H/3$，该结论对每个 $H \leq k$ 均成立（第 4 节中的引理 1）。

轮次数量 k 受以下要求限制：坐标 $u_t(i_*)$ 和 $v_t(j_*)$ 在所有轮次中都必须等于 1。在正交约束下，每轮只能保留约一半仍等于 1 的坐标，使其继续等于 1，因此 k 至多约为 $\log_2 \min\{m, n\}$（第 5 节中的引理 3）。这就是我们的下界关于行动数量呈对数增长的原因。对于推论 3 中的下界 $\Omega(\sqrt{\log\ m \log\ n})$，其中 n 可以是 m 的多对数级，我们避免了 y 玩家这一方的这种折半。在每轮中，我们不采用行动 j_*，而是使用 y 玩家的混合策略，并对其进行乘法更新，代价是每轮下界的增幅较小（第 6 节中的引理 4）。

我们的下界对纳什均衡的计算也有影响。它们给出了确定性学习动态的平均策略收敛到纳什均衡的速率下界 $\Omega(\log \min\{m, n\}/T)$

（推论 2），这改进了 Daskalakis 等人(2011) 给出的下界。当 $m = n$ 时，该下界与乐观 Hedge 的收敛速率 $O(\log(mn)/T)$ 相差一个常数因子。作为我们下界的附带结果，我们还证明，确定性算法计算 ε-近似纳什均衡所需的矩阵-向量乘积查询复杂度为 $\Omega(\log \min\{m, n\})$，当$\varepsilon \leq 1/12$ (推论 4)。这可能略高于 Hadiji 等人(2023年)；Kornowski 和 Shamir (2026年) 已知的下界，而图2 在第7节 比较了该下界与已知的上界和下界。

相关工作。在算法方面，乐观 Hedge 是乐观跟随正则化领导者（Rakhlin 和 Sridharan，2013）的一种实例，而乐观预测已成为设计学习动态以实现快速收敛的核心方法（例如，Syrgkanis 等人，2015年；Chen 和 Peng，2020；Daskalakis 等人，2021；Anagnostides 等人，2022；Farina 等人，2022；Soleymani 等人，2025；Tsuchiya，2026b）。在本文所考虑的反馈条件下，采用确定性初始化和参数选择的标准乐观 Hedge 和乐观跟随正则化领导者动态会以确定性方式生成其混合策略，因此适用于我们的下界。

在下界方面，Syrgkanis 等人（2015）证明，在某些双人零和博弈中，使用任意固定学习率的 Hedge 的玩家面对最佳响应对手时，其遗憾为 $\Omega(\sqrt{T})$。Chen 和 Peng（2020）证明，当双方玩家都使用相同固定学习率的 Hedge 时，在双人一般和博弈中，无法保证至少一名玩家的遗憾为 $o(\sqrt{T})$。当双方玩家分别使用学习率为 η 和 η'的乐观 Hedge 时，在某个双人零和博弈中，他们的遗憾可以分别至少达到 $(\log m)/\eta - o(1)$ 和 $(\log n)/\eta' - o(1)$（Tsuchiya，2026a）。Anagnostides 等人（2024）在 PPAD 指数时间假设下，为多人一般和博弈中的 Hedge 和乐观 Hedge 证明了条件性的、依赖维度的下界。这些下界针对特定的学习算法，而我们的下界是无条件的，且适用于所有确定性学习动态。

一个相关问题是矩阵博弈的查询复杂度，也就是计算一个 ε-近似纳什均衡所需的收益矩阵-向量乘积次数（Nemirovski，2004年；Nesterov，2005年；Hadiji 等人，2023年；Kornowski 和 Shamir，2026年；Karmarkar 等人，2026年）。该问题与遗憾最小化不同，因为算法的输出会接受评估。我们针对该问题得到的遗憾下界的一个推论将在第7节中讨论。

将收益矩阵表示为秩一矩阵之和，形式为$(u_t - u_{t-1})v_t^\top$，如(1)所示，也被Kornowski 和 Shamir (2026年)用于证明确定性算法计算矩阵博弈的近似纳什均衡所需的矩阵-向量乘积查询复杂度下界。两种构造采用了这种形式的矩阵，但秩一项中的向量选择方式不同，目的也各不相同。Kornowski 和 Shamir (2026年)通过将高斯向量投影到此前查询的正交补上来构造这些向量，并证明算法经过 N 次矩阵-向量乘积后，其输出仍非最优，差距至少为$\Omega(N^{-3/2})$，忽略对数因子。相比之下，我们构造矩阵时确保每一轮构造中遗憾之和的下界都会增加一个常数；正如下文将要看到的，向量 u_t 和 v_t 是为此目的而选取的。

2 问题设定与预备知识

本节介绍问题设定以及无遗憾学习与纳什均衡之间的关系，并形式化本文所考虑的学习动态类别。对于正整数 n，记$[n] = \{1, \dots, n\}$和 $\Delta_n = \{y \in \mathbb{R}^n \colon y(j) \geq 0 \text{ 对所有 } j \in [n], \ \sum_{j=1}^n y(j) = 1\}$。我们用$\mathbf{1}$ 表示全一向量，用 e_i 表示第 i个标准基向量，其维度由上下文决定。

对于向量 x，记其第 i个坐标为 $x(i)$，其 ℓ_p范数为 $\|x\|_p$。除非另行指定底数，所有对数均为自然对数。

问题设定。在双人零和博弈（亦称矩阵博弈）中进行学习，由收益矩阵 $A \in [-1,1]^{m\times n}$刻画，其中 m 和 n 分别表示 x-玩家和 y-玩家的行动数量。在每一轮 $t = 1, \ldots\ldots, T$中，x-玩家选择策略 $x_t \in \Delta_m$，而 y-玩家同时选择 $y_t \in \Delta_n$。随后，x-玩家观察到期望收益向量 $g_t = Ay_t$，而 y-玩家观察到期望损失向量 $\ell_t = A^\top x_t$。x-玩家的收益和 y-玩家的损失均为 $\langle x_t, Ay_t\rangle$。

每位玩家的目标都是最小化（外部）遗憾。对于任意 $H \in [T]$，x-玩家和y-玩家在前 H 轮中的遗憾为

$$\mathsf{Reg}_H^x = \max_{x^*\in\Delta_m} \sum_{t=1}^{H} \langle x^* - x_t, Ay_t\rangle, \qquad \mathsf{Reg}_H^y = \max_{y^*\in\Delta_n} \sum_{t=1}^{H} \langle y_t - y^*, A^\top x_t\rangle.$$

社会遗憾定义为两位玩家遗憾之和，即

$$\begin{aligned} \mathsf{Reg}_H^x + \mathsf{Reg}_H^y &= \max_{x^*\in\Delta_m,\, y^*\in\Delta_n} \sum_{t=1}^{H} \big(\langle x^*, Ay_t\rangle - \langle x_t, Ay^*\rangle\big) \\ &= H\left(\max_{x\in\Delta_m} \langle x, A\bar{y}_H\rangle - \min_{y\in\Delta_n} \langle \bar{x}_H, Ay\rangle\right), \end{aligned} \tag{2}$$

其中 $\bar{x}_H = \frac{1}{H}\sum_{t=1}^{H} x_t$ 和 $\bar{y}_H = \frac{1}{H}\sum_{t=1}^{H} y_t$ 分别是在 H 轮次中的平均策略。我们将通过在 (2)的最大值中固定一对（x^*，y^*）来证明我们的下界。注意，社会遗憾的下界也给出了最大个体遗憾的下界，因为 $\max\{\mathsf{Reg}_H^x、\mathsf{Reg}_H^y\} \geq (\mathsf{Reg}_H^x + \mathsf{Reg}_H^y)/2$。

无遗憾学习与纳什均衡。如果概率分布对（x^*，y^*）分别定义在行动集$[m]$ 和 $[n]$ 上，则称其为关于 $\varepsilon \geq 0$ 的 ε-近似纳什均衡，如果对于任意分布 $x \in \Delta_m$ 和$y \in \Delta_n$，

$$\langle x, Ay^*\rangle - \varepsilon \leq \langle x^*, Ay^*\rangle \leq \langle x^*, Ay\rangle + \varepsilon.$$

如果一对（x^*，y^*）是 0-近似纳什均衡，则称其为纳什均衡；此时，$\langle x^*, Ay^*\rangle$ 称为博弈值。如果一个策略是标准基向量，则称其为纯策略；如果纳什均衡（x^*，y^*）中的 x^* 和 y^* 均为纯策略，则称该纳什均衡为纯纳什均衡。

众所周知，近似纳什均衡是无遗憾学习动态的结果（Freund 和 Schapire，1999）。根据(2)，平均策略对$(\bar{x}_H,\ \bar{y}_H)$是一个 $\big((\mathsf{Reg}_H^x + \mathsf{Reg}_H^y)/H\big)$-近似纳什均衡。反之，如果$(\bar{x}_H,\ \bar{y}_H)$是一个 ε-近似纳什均衡，那么(2) 给出$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \leq 2\varepsilon H$。

学习动态。我们将一对学习算法（每位玩家各一个）称为学习动态。我们的下界适用于以下一类学习动态，我们称之为确定性学习动态。在确定性学习动态中，x玩家和 y玩家的两个算法都可以在每轮 t 根据 m、n、T以及过去历史中的所有信息 $(x_s,\ y_s,\ g_s,\ \ell_s)_{s=1}^{t-1}$来选择各自的策略。因此，每位玩家也可以使用对手的策略和反馈，但任何算法都没有关于 A 的其他信息，也不使用内部随机化。基于乐观 Hedge 的学习动态中，每位玩家只使用自己的反馈，因此属于确定性学习动态。

3 主要结果

本节介绍我们的遗憾下界。我们的第一个结果是一个下界，它随轮次 H 线性增长，直至参数 k；其中 k 是任意满足 $k \le T$ 且 $k2^k+1 \le \min\{m,n\}$ 的正整数。

定理 1. 假设 k、T、m、n 是满足 $k \le T$ 和 m、$n \ge k2^k+1$ 的正整数。那么，对于每个确定性学习动态，都存在一个固定的收益矩阵 $A \in [-1,1]^{m\times n}$，它具有一个零值纯纳什均衡，并且对于每个 $H \in [T]$,

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \ge \frac{1}{3}\min\{H,k\}, \qquad \max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \ge \frac{1}{6}\min\{H,k\}. \tag{3}$$

证明见 第 4 节和 第 5节。在 定理 1中，矩阵 A取决于学习动态、k以及 T，但同一个矩阵对每个 (3) 都成立： $H \in [T]$。该矩阵具有一个纯纳什均衡，这意味着下界并非源于寻找混合均衡的困难。

备注 1. (3) 中的第二个不等式 给出了两个个体遗憾的最大值的下界，而不可能对两个个体遗憾的最小值 给出类似的遗憾下界。事实上，假设 y玩家每轮都选择均匀策略 $\mathbf{1}/n$，而 x玩家在第一轮选择 $\mathbf{1}/m$，并在之后的所有轮次中选择使第一反馈向量 $g_1 = A\mathbf{1}/n$最大化的行动。那么遗憾$_H^x$ = 最大$_{i\in[m]}\, g_1(i) - \langle \mathbf{1}/m,\ g_1\rangle \le 2$ 对每个 A 和每个 H都成立，它与 k 无关，因此 m，n。

选择 k 与 $\log\min\{m,n\}$ 成正比，可得 定理 1 的以下推论，其证明见 第5.3节。

推论 1. 假设 $T \ge 1$ 和 $m,n \ge 2$。那么，对于每个确定性学习动态，都存在一个固定的收益矩阵 $A \in [-1,1]^{m\times n}$，使得对每个 $H \in [T]$,

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \ge \frac{1}{12}\min\{H,\log_2\min\{m,n\}\}, \quad \max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \ge \frac{1}{24}\min\{H,\log_2\min\{m,n\}\}.$$

当 $m=n$时，推论 1 与上界相差至多一个常数因子 $O(\log(mn))$，该上界适用于乐观 Hedge 在常数学习率下的个体遗憾，对每个 T(Rakhlin 和 Sridharan，2013)。由于每种学习动态的个体遗憾至多为 $2T$，确定性学习动态下最大个体遗憾的极小极大值为 $\Theta(\min\{T,\log m\})$，其中 $m=n$。

推论 1 改进了 Daskalakis 等人(2011) 关于平均策略收敛至纳什均衡的速率的下界。将推论 1 中的社会遗憾下界与 (2) 结合，可得以下推论，其证明见第5.3节。

推论 2。 假设 $T \ge 1$ 和 m、$n \ge 2$。那么，对于每个确定性学习动态，都存在一个固定的收益矩阵 $A \in [-1,1]^{m\times n}$，使得对于每个 $H \in [T]$，平均策略 $(\bar{x}_H$、$\bar{y}_H)$的组合对于任意都不是 ε-近似纳什均衡

$$\varepsilon < \frac{1}{24H}\min\{H,\log_2\min\{m,n\}\}.$$

对于 $T \ge \log_2\min\{m,n\}$，推论 2 中关于 ε 且 $H=T$时的下界为 $\log_2\min\{m,n\}/(24T)$，而 $\Omega(1/T)$ 这一 Daskalakis 等人 (2011) 针对强非耦合学习动态给出的下界则与行动数量无关。当 $m=n$ 时，推论 2 中的下界与$O(\log(mn)/T)$ 这一采用常数学习率的乐观 Hedge 的速率相匹配，差一个常数因子。

接下来，我们考虑两位玩家的行动数量不同的博弈。以下定理将 $n \ge k2^k+1$ 在定理 1中的条件替换为 $n \ge 16k$，代价是下界中多出一个因子 $c \le 1/4$。

定理 2。假设 k、T、m、n 是满足 $k \le T$、$m \ge k2^k + 1$ 和 $n \ge 16k$ 的正整数，并定义

$$c = \min\left\{\frac{1}{4}, \sqrt{\frac{1}{16k}\log\frac{n}{k}}\right\}. \tag{4}$$

那么，对于每一种确定性学习动态，都存在一个固定矩阵 $A \in [-1,1]^{m\times n}$，其纳什均衡的值为零，且其 x-玩家的策略为纯策略，使得对于每个 $H \in [T]$，

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \ge \frac{c}{3}\min\{H,k\}, \qquad \max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \ge \frac{c}{6}\min\{H,k\}. \tag{5}$$

特别地，对于每个满足 $k \le H \le T$ 的 H，

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \ge \frac{1}{12}\min\left\{k, \sqrt{k\log\frac{n}{k}}\right\}, \qquad \max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \ge \frac{1}{24}\min\left\{k, \sqrt{k\log\frac{n}{k}}\right\}. \tag{6}$$

证明见第 6 节。交换玩家并将 A 替换为 $-A^\top$，即可得到对应结果，其中关于 m 和 n 的条件互换，此时 y-玩家的均衡策略为纯策略。在 k 取为与$\log m$ 成比例时，定理 2 给出以下推论，其证明见第 6.2 节。

推论 3。对于每个 $\alpha \in (0,\ 1)$，都存在一个常数 $C_\alpha > 0$，使得以下结论成立。假设 $m \ge n$、$\log m \le n^\alpha$、$T \ge \log_2 m$，并且 n 足够大，且其大小仅取决于 α。那么，对于每一种确定性学习动态，都存在一个固定收益矩阵 $A \in [-1,1]^{m\times n}$，使得

$$\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \ge 2C_\alpha\sqrt{\log m \log n}, \qquad \max\{\mathsf{Reg}_T^x, \mathsf{Reg}_T^y\} \ge C_\alpha\sqrt{\log m \log n}.$$

推论 3 与最大个体遗憾的上界相差一个常数因子 $O(\sqrt{\text{对数 } m\text{对数 } n})$；当每位玩家都知道对手的行动数时，乐观 Hedge 可达到该上界（Tsuchiya，2026a，定理 9）。因此，在推论 3的假设下，确定性学习动态的最大个体遗憾的极小极大值为

$$\inf_{(\mathsf{Alg}^x,\mathsf{Alg}^y)} \sup_{A\in[-1,1]^{m\times n}} \max\{\mathsf{Reg}_T^x, \mathsf{Reg}_T^y\} = \Theta\big(\sqrt{\log m \log n}\big),$$

其中，下确界取遍所有确定性学习动态（Alg^x，Alg^y），且 Θ中的常数仅取决于 α。

注 2. 推论 3 中的下界并不适用于所有 $m \ge n$。事实上，当 y玩家的行动数为 $n = 2$时，存在一种确定性学习动态，使其个体遗憾始终都不超过4，对任意 $A \in [-1,1]^{m\times 2}$ 以及任意 H 均成立（见附录A中的命题1），而 $\sqrt{\text{对数 } m\text{ 对数 } n}$ 则随m增长。

4 收益矩阵的构造

本节给出收益矩阵的构造，该构造适用于定理 1和2的证明。我们根据每轮选取的向量 u_t 和 v_t（用于(1)）构造矩阵（算法 1），并证明只要这些向量满足三个条件，下界就成立（引理 1）。随后，在 u_t、v_t 中选取这些向量：第 5 节中用于定理 1，而第 6 节中用于定理 2。

算法 1： 构造收益矩阵的模板

输入：一个确定性学习动态 $(\text{Alg}^x, \text{Alg}^y)$、行动数 m, n、时域 T 和轮次数 $k \le T$**1** 设置 $u_0 = \mathbf{1}$ 和 $A_0 = 0 \in \mathbb{R}^{m\times n}$； **2** 对于 $t = 1, \ldots, k$，执行 **3** 令算法 $(\text{Alg}^x, \text{Alg}^y)$ 从 $(x_s, y_s, g_s, \ell_s)_{s=1}^{t-1}$ 中选择 $x_t \in \Delta_m$ 和 $y_t \in \Delta_n$； **4** 选择满足对所有 $s \in [t]$ 均有 $\langle x_s, u_t\rangle = 0$ 的 $u_t \in [-1,1]^m$； **5** 选择满足对所有 $s \in [t]$ 均有 $\langle y_s, v_t\rangle = 0$ 的 $v_t \in [-1,1]^n$； **6** 按照 **(1)** 设置 $A_t = A_{t-1} + \frac{1}{3}(u_t - u_{t-1})v_t^\top$； **7** 将反馈 $g_t = A_t y_t$ 和 $\ell_t = A_t^\top x_t$ 提供给 $(\text{Alg}^x, \text{Alg}^y)$；

O

utput: $A = A_k$

收益矩阵的必要性质。根据(2)，每个行动 $i_* \in [m]$ 和每个分布 $y^* \in \Delta_n$均满足

$$\text{Reg}_H^x + \text{Reg}_H^y \ge \sum_{t=1}^{H}\big(\langle e_{i_*}, Ay_t\rangle - \langle x_t, Ay^*\rangle\big). \tag{7}$$

我们将构造一个矩阵 A，并选择 i_*、y^* 和一个常数 $c > 0$，使其满足

$$e_{i_*}^\top A = 0, \qquad Ay^* \le 0, \qquad \langle x_t, Ay^*\rangle = -\frac{c}{3} \quad \text{for all } t \le k, \tag{8}$$

其中不等 式逐坐标成立。在这些条件下，接着从(7)可得 e

$$\text{Reg}_H^x + \text{Reg}_H^y \ge -\sum_{t=1}^{H}\langle x_t, Ay^*\rangle = -\sum_{t=1}^{\min\{H,k\}}\langle x_t, Ay^*\rangle - \sum_{t=\min\{H,k\}+1}^{H}\langle x_t, Ay^*\rangle \ge \frac{c}{3}\min\{H,k\}, \tag{9}$$

这给出了定理 1中的社会遗憾下界，其中 $c = 1$，而在定理 2中亦然。**(8)** 中的前两个性质(8) 还意味着，(e_{i_*}, y^*) 是一个零值纳什均衡。难点在于，x_t 中(8) 的策略由算法选定，并通过反馈依赖于 A。

构造。由于策略取决于 A，我们在运行两个算法进行 k 轮次的过程中，自适应地构造 A，如算法 **1** 所述。在第 t 轮，两个玩家的算法根据策略和前几轮的反馈，选出 (x_t, y_t)（第 **3** 行）。随后，构造过程选择与迄今选出的策略正交的向量$u_t \in [-1,1]^m$ 和 $v_t \in [-1,1]^n$（第 **4** 行和第 **5** 行）。接着，将项 **1 3** $(u_t - u_{t-1})v_t^\top$添加到矩阵中（第 **6** 行），并根据得到的矩阵 A_t 向算法提供反馈（第 **7** 行）。收益矩阵即为最终矩阵 $A = A_k$。

因此，第 t 轮的反馈是根据 A_t 计算的，而遗憾则是针对在固定矩阵 A 上运行的算法定义的。第 **4** 行和第 **5** 行中的正交性是为了确保第 t 轮之后添加的项不会改变第 t 轮的反馈，这意味着确定性算法在 A 上运行时会选择相同的策略。式 **(1)** 中的差值 $u_t - u_{t-1}$ 用于推导式 **(8)** 中的性质。若对于所有 t，$u_t(i_*) = 1$，则每一项的行 i_* 均为零；若对于所有 t，$\langle y^*, v_t\rangle = c$，则 $Ay^* = \frac{c}{3}\sum_{t=1}^{k}(u_t - u_{t-1}) = \frac{c}{3}(u_k - \mathbf{1})$。由于 $u_k \le \mathbf{1}$，且第 **4** 行中的正交性给出对于所有 $t \le k$，$\langle x_t, u_k\rangle = 0$，该恒等式意味着 $Ay^* \le 0$，并且对于所有 $t \le k$，$\langle x_t, Ay^*\rangle = -\frac{c}{3}$。下面的引理表明，在满足这两个条件以及对于每个 $i \in [m]$ 均有 $\sum_{t=1}^{k}|u_t(i) - u_{t-1}(i)| \le 3$ 的条件下，下界仍然成立；该条件确保各项

表1： 向量条件的摘要 u_t 以及 v_t 和矩阵性质的摘要 $A = A_k$，其中每个条件所推出的性质为 $i_* \in [m]$，$y^* \in \Delta_n$，以及 $c > 0$。条件 (a) 是 算法 1第 4行和第 5行所要求的条件，而条件 (b)–(d) 是 引理 1的假设。在条件 (a) 下，第 (c) 行和第 (d) 行中的性质给出 (8)。

	向量 u_t 和 v_t的条件	收益矩阵 A的性质
(a)	$\langle x_s, u_t\rangle = 0$ 和 $\langle y_s, v_t\rangle = 0$ 对所有 $1 \le s \le t \le k$	$Ay_s = g_s$ 和 $A^\top x_s = \ell_s$ 对所有 $s \in [k]$
(b)	$\sum_{t=1}^{k} \lvert u_t(i) - u_{t-1}(i)\rvert \le 3$ 对所有 $i \in [m]$	$A \in [-1,1]^{m\times n}$
(c)	$u_t(i_*) = 1$ 对所有 $t \in \{0,\dots,k\}$	$e_{i_*}^\top A = 0$
(d)	$\langle y^*, v_t\rangle = c$ 对所有 $t \in [k]$	$Ay^* = \frac{c}{3}(u_k - \mathbf{1})$

属于 A 中的 $[-1,1]$。表1 总结了向量 u_t 和 v_t 的条件，以及 A 由各条件所推出的性质。算法 1 并未对向量 u_t 和 v_t 作出正交性之外的规定；该引理适用于所有满足其假设的选择。

引理 1. 假设 $k \le T$ 且前 k 轮次由 算法 1 在某个确定性学习动态下生成。我们还假设 $\sum_{t=1}^{k} \lvert u_t(i) - u_{t-1}(i)\rvert \le 3$ 对每个 $i \in [m]$都成立。则有以下结论：

(i) 对每个 $s \in [k]$，$A_k y_s = g_s$ 和 $A_k^\top x_s = \ell_s$ 均成立。此外，$A = A_k$ 属于集合 $[-1,1]^{m\times n}$，并且在固定博弈 A 上运行的算法会选择 $x_1,\dots,x_k$ 和 $y_1,\dots,y_k$，并在前 k 轮次中收到 $g_1,\dots,g_k$ 和 $\ell_1,\dots,\ell_k$。

(ii) 进一步假设，存在 $i_* \in [m]$，使得对所有 $t \in \{0,\dots,k\}$，$u_t(i_*) = 1$；并且存在 $y^* \in \Delta_n$ 和 $c > 0$，使得对所有 $t \in [k]$，$\langle y^*, v_t\rangle = c$。那么，在固定博弈 A 上运行的算法，其遗憾满足：对于每个 $H \in [T]$，

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \ge \frac{c}{3}\min\{H,k\}, \qquad \max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \ge \frac{c}{6}\min\{H,k\},$$

并且 (e_{i_*}, y^*) 是 A 的一个纳什均衡，其值为零。

证明。我们分三步证明 (i)，然后证明(ii)。

我们首先证明，对于每个 $s \in [k]$，都有 $A_k y_s = g_s$ 和 $A_k^\top x_s = \ell_s$。固定 $s \in [k]$。根据 算法 1 的第 6 行和第 7行，

$$g_s = A_s y_s, \qquad \ell_s = A_s^\top x_s, \qquad A_k - A_s = \frac{1}{3}\sum_{r=s+1}^{k}(u_r - u_{r-1})v_r^\top.$$

对于每个 $r \in \{s+1,\dots,k\}$，我们有

$$(u_r - u_{r-1})v_r^\top y_s = 0, \qquad v_r(u_r - u_{r-1})^\top x_s = \big(\langle u_r, x_s\rangle - \langle u_{r-1}, x_s\rangle\big)v_r = 0,$$

其中，第一个等式由 $\langle y_s, v_r\rangle = 0$ (第5行) 得出，第二个等式则由 $\langle x_s, u_r\rangle = \langle x_s, u_{r-1}\rangle = 0$ (第 4 行 在轮次 $t = r$ 和 $t = r-1$ 得出，因为 $s \le r-1$）。因此 $(A_k - A_s)y_s = 0$ 且 $(A_k - A_s)^\top x_s = 0$，由此可得 $A_k y_s = A_s y_s = g_s$ 和 $A_k^\top x_s = A_s^\top x_s = \ell_s$。

接下来我们证明 $A_k \in [-1,1]^{m\times n}$。对于所有 $i \in [m]$ 和 $j \in [n]$,

$$|A_k(i,j)| \le \frac{1}{3}\sum_{r=1}^{k} |u_r(i) - u_{r-1}(i)|\,|v_r(j)| \le \frac{1}{3}\sum_{r=1}^{k} |u_r(i) - u_{r-1}(i)| \le 1,$$

其中，第二个不等式由 $v_r \in [-1,1]^n$ 得出，最后一个不等式则由假设 $\sum_{r=1}^{k} |u_r(i) - u_{r-1}(i)| \le 3$ 对每个 $i \in [m]$ 都成立得出。

最后，我们通过对 $s \in [k]$ 进行归纳，证明在固定博弈 $A = A_k$ 上运行的算法会在每一轮 $r \le s$ 选择 (x_r, y_r) 并收到 (g_r, ℓ_r)。对于 $s = 1$，算法在第 1 轮根据 m、n 和 T 选择策略，因此在 A 上的第 1 轮会选择 (x_1, y_1)。根据第一步，在 A 上第 1 轮的反馈为 $Ay_1 = g_1$ 和 $A^\top x_1 = \ell_1$，因此该结论对 $s = 1$ 成立。令 $s \in \{2, \dots, k\}$，并假设该结论对 $s-1$ 成立，这是归纳假设。根据归纳假设，A 上第 1 轮至第 $s-1$ 轮的策略和反馈与算法 1 相同。由于算法在第 s 轮根据 m、n、T 以及第 1 轮至第 $s-1$ 轮的策略和反馈来选择策略，因此它们在 A 上的第 s 轮会选择 (x_s, y_s)。根据第一步，在 A 上第 s 轮的反馈为 $Ay_s = g_s$ 和 $A^\top x_s = \ell_s$，因此该结论对 s 成立。这就证明了第 (i) 部分。

对于第 (ii) 部分，A 的定义以及假设 $u_r(i_*) = 1$ 对于 $r \in \{0, \dots, k\}$ 以及 $\langle y^*, v_r\rangle = c$ 对于 $r \in [k]$ 给出

$$e_{i_*}^\top A = \frac{1}{3}\sum_{r=1}^{k} \big(u_r(i_*) - u_{r-1}(i_*)\big) v_r^\top = 0, \tag{10}$$

$$Ay^* = \frac{1}{3}\sum_{r=1}^{k} \langle v_r, y^*\rangle (u_r - u_{r-1}) = \frac{c}{3}\sum_{r=1}^{k} (u_r - u_{r-1}) = \frac{c}{3}(u_k - \mathbf{1}). \tag{11}$$

由于 $u_k \le \mathbf{1}$ 在各坐标上成立，(11) 给出 $Ay^* \le 0$ 在各坐标上成立，因此

$$\langle x, Ay^*\rangle \le 0 = \langle e_{i_*}, Ay^*\rangle = \langle e_{i_*}, Ay\rangle \qquad \text{for all } x \in \Delta_m,\ y \in \Delta_n,$$

其中两个等式由 (10) 得出。这表明 $(e_{i_*},\ y^*)$ 是 A 的一个纳什均衡，其值为零。

还需证明遗憾的下界。对于算法在 $t \in [T]$ 上运行的每一轮 A，由 (10) 和 (11)可知，(7) 右侧的每一项都可以写成

$$\langle e_{i_*}, Ay_t\rangle - \langle x_t, Ay^*\rangle = 0 - \Big\langle x_t, \frac{c}{3}(u_k - \mathbf{1})\Big\rangle = \frac{c}{3}\big(1 - \langle x_t, u_k\rangle\big) \ge 0, \tag{12}$$

其中我们用了 $\langle x_t,\ \mathbf{1}\rangle = 1$，最后一个不等式由 $x_t \in \Delta_m$ 和 $u_k \le \mathbf{1}$得出。式 (12)中的量对于每个 $t \le k$ 都等于 $c/3$，因为此时 x_t 是第算法 1 中由第 (i)部分选出的策略，并且 第 4 行的 算法 1 给出 $\langle x_t,\ u_k\rangle = 0$。因此，利用这两个性质，对于每个 $H \in [T]$，我们有

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \ge \sum_{t=1}^{H}\big(\langle e_{i_*}, Ay_t\rangle - \langle x_t, Ay^*\rangle\big) = \frac{c}{3}\sum_{t=1}^{H}\big(1 - \langle x_t, u_k\rangle\big) \ge \frac{c}{3}\min\{H, k\},$$

其中，第一个不等式由(7)得出，等式由(12)得出。最大个体遗憾的下界由$\max\{$遗憾$_H^x$，遗憾$_H^y\} \ge ($遗憾$_H^x$ +遗憾$_H^y)/2$推出，至此证明完成。 □

根据 引理 1，接下来只需选择向量 u_t 和 v_t，分别放在 第 4 行和第 5 行的算法 1中，使得对于某个 i_*，y^*和 c，该引理的假设均成立。第 5 节给出了满足 $c = 1$ 的这类向量，其中$m,\ n \ge k2^k + 1$；而第 6 节给出了向量 v_t，使得式 c 中的常数(4) 在 $n \ge 16k$时成立。

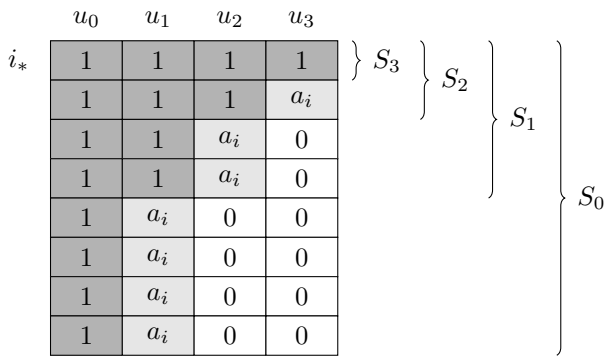


图1： 向量示意图 u_0、……、u_3 构造于 第 5 节 中，用于 $k = 3$。每一行显示一个坐标 $u_0(i)$ 、……、 $u_3(i)$ 的序列，其中 i，且 $a_i \in [-1, 1)$ 。集合 S_t 由以下坐标组成：它们在 1 中 u_0、……、 u_t中的值等于 1，且坐标 i_* 属于 S_3。

5 定理 1 的证明

本节构造向量 u_t 和 v_t，将其放在算法 1 的第 4 行和第 5 行，并证明定理 1。我们首先概述本节内容。向量 u_t 必须满足：对于所有 $1 \le s \le t \le k$，均有 $\langle x_s, u_t \rangle = 0$，如第 4 行所要求；而引理 1 假设，对于每个 $i \in [m]$， $\sum_{t=1}^{k} |u_t(i) - u_{t-1}(i)| \le 3$，并且对于某个 $i_* \in [m]$，对所有 $t \in \{0, \ldots, k\}$ 均有$u_t(i_*) = 1$。

为满足条件 $\sum_{t=1}^{k} |u_t(i) - u_{t-1}(i)| \le 3$，我们令每个坐标至多改变两次。 u_t 中的所有坐标不可能都等于 1，因为 $\langle x_t, u_t \rangle = 0$ 和 $x_t \in \Delta_m$。因此，坐标必须从 1 变为其他值，并且一旦某个坐标取值为 $a \in [-1, 1)$，我们就在之后的所有轮次中将其设为零。于是，每个序列 $u_0(i)$ ……、$u_k(i)$ 要么恒为 1，要么具有如下形式：1、……、1、 a、0、……、0，如图1所示；在后一种情况下， $u_t(i)$ 的累积变差的上界为

$$\sum_{t=1}^{k} |u_t(i) - u_{t-1}(i)| \le |a - 1| + |0 - a| \le 3.$$

还需要确保每一轮中都有某个坐标等于 1。我们定义集合 s

$$S_0 = [m], \qquad S_t = \{i \in S_{t-1} : u_t(i) = 1\} \quad \text{for } t \in [k],$$

该集合由在 $u_0, \ldots, u_t$ 中取值为 1 的坐标组成。在第 t 轮， u_t 中不属于S_{t-1} 的坐标均为零，而对于 $i \in S_{t-1}$， 数值 $u_t(i)$ 必须满足 t 线性约束 $\langle x_s, u_t \rangle = 0$。在第5.1节中，我们将证明，可以在 $[-1, 1]$ 中选取 $i \in S_{t-1}$ 对应的数值 $u_t(i)$，使这些 t 约束成立，且至少有 $(|S_{t-1}| - t)/2$ 个数值等于 1（引理 2），从而得到 $|S_t| \ge (|S_{t-1}| - t)/2$。根据归纳法， $|S_k| \ge m/2^k - k$，因此，如果 $m \ge k2^k + 1$，则 S_k 非空；这一点将在第 5.2 节的引理 3 中证明。每一轮都将规模减半，这就是定理 1 中的下界仅在约 $\log_2 \min\{m, n\}$ 轮内增长的原因。向量 v_t 是通过对 y-玩家的策略应用相同的构造得到的；定理 1 和推论 1和推论2将在第5.3节中证明。

5.1 线性约束下立方体中的向量

向量 u_t 的构造基于以下引理，该引理给出了 $[-1, 1]^d$ 中一个与给定向量正交且有许多坐标等于 1 的向量。

引理 2。 设 $a_1, \ldots, a_r \in \mathbb{R}^d$。则存在 $z \in [-1,1]^d$，使得 $\langle a_s, z\rangle = 0$ 对所有 $s \in [r]$ 都成立，并且 z 中严格介于 -1 和 1 之间的坐标至多有 r 个。因此，存在这样一个向量，其至少有$(d-r)/2$ 个坐标等于 1。

证明。 令 $P = \{z \in [-1,1]^d \colon \langle a_s, z\rangle = 0$ 对所有 $s \in [r]\}$成立。由于 P包含零向量且是紧集，因此存在

$$z^* \in \arg\max_{z \in P} \|z\|_2^2.$$

我们证明 z^* 满足第一个结论。定义

$$F = \{i \in [d] : -1 < z^*(i) < 1\},$$

并反设 $|F| > r$。那么， $|F|$ 向量 $(a_1(i) \ldots, a_r(i)) \in \mathbb{R}^r$ 关于 $i \in F$线性相关，因为 $|F| > r$。因此存在一个非零向量 $h \in \mathbb{R}^d$，使得

$$h(i) = 0 \quad \text{for all } i \notin F, \qquad \langle a_s, h\rangle = \sum_{i \in F} h(i)\, a_s(i) = 0 \quad \text{for all } s \in [r].$$

由于 $-1 < z^*(i) < 1$ 对 $i \in F$ 成立，且 $h(i) = 0$ 对 $i \notin F$成立，因此存在 $\delta > 0$，使得 $z^* + \delta h$ 和 $z^* - \delta h$ 都属于 $[-1,1]^d$。这两个向量属于 P，因为对所有 $s \in [r]$， $\langle a_s, z^* \pm \delta h\rangle = 0$ 成立，而且它们满足

$$\|z^* + \delta h\|_2^2 + \|z^* - \delta h\|_2^2 = 2\|z^*\|_2^2 + 2\delta^2 \|h\|_2^2 > 2\|z^*\|_2^2.$$

因此，这两个向量中至少有一个的范数大于 z^*，这与 z^* 的定义矛盾。因此$|F| \le r$，这就证明了第一个结论。

对于第二个结论，根据第一个结论，至少有 $d-r$ 个 z^* 的坐标等于 1 或 -1。如果其中至少一半等于 1，那么 z^* 满足第二个结论。否则，其中至少一半等于 -1，且$-z^*$ 满足第二个结论，因为对于所有 $s \in [r]$，都有 $\langle a_s, -z^*\rangle = 0$，并且 $-z^*$ 中严格介于-1 和 1 之间的坐标就是 F中的那些坐标。证明至此完成。

□

5.2 向量 u_t的构造

下面的引理给出了满足 u_t 所要求的正交性第 4 行中的算法 1以及引理 1的假设的向量。每一轮都会对之前所有轮次中都等于1 的坐标应用引理 2来构造这些向量。

引理 3. 假设 k 和 m 是满足 $m \ge k2^k + 1$的正整数，并令 $u_0 = \mathbf{1}$。那么，对于任意依次给定的 $x_1, \ldots, x_k \in \Delta_m$，可以为 $t \in [k]$选择 $u_t \in [-1,1]^m$，其中 u_t 仅取决于$x_1, \ldots, x_t$，以及一个坐标 $i_* \in [m]$，使得

$$\langle x_s, u_t\rangle = 0 \qquad \textit{for all } 1 \le s \le t \le k, \tag{13}$$

$$\sum_{t=1}^{k} |u_t(i) - u_{t-1}(i)| \le 3 \qquad \textit{for all } i \in [m], \tag{14}$$

$$u_t(i_*) = 1 \qquad \textit{for all } t \in \{0, \ldots, k\}. \tag{15}$$

证明。 我们递归地构造向量 u_t 以及集合 $S_t \subseteq [m]$，其中 $t = 1, \ldots, k$，然后证明陈述中的三个性质。

设 $S_0 = [m]$。令 $t \in [k]$，并假设 u_{t-1} 和 S_{t-1} 已由 $x_1, \ldots, x_{t-1}$ 构造得到。我们对 $i \notin S_{t-1}$，设定 $u_t(i) = 0$。对于 $i \in S_{t-1}$，我们根据引理 2 的第二个断言，取值 $u_t(i) \in [-1, 1]$；应用该断言时，使用 $d = |S_{t-1}|$ 和 $r = t$，并将其应用于向量 $(x_s(i))_{i \in S_{t-1}}$，其中 $s \in [t]$。这些向量是 $x_1, \ldots, x_t$ 在坐标 S_{t-1} 上的限制。这些值满足

$$\sum_{i \in S_{t-1}} x_s(i) u_t(i) = 0 \qquad \text{for all } s \in [t], \tag{16}$$

并且其中至少有 $(|S_{t-1}| - t)/2$ 个等于 1。接着我们定义

$$S_t = \{i \in S_{t-1} : u_t(i) = 1\}.$$

然后，等于 1 的值的数量下界给出

$$|S_t| \geq \frac{|S_{t-1}| - t}{2}. \tag{17}$$

根据构造，u_t 和 S_t 仅依赖于 x_1、……、x_t 以及 $S_0 \supseteq S_1 \supseteq \cdots \supseteq S_k$。

正交性 (13) 直接由上述构造得出。对于 $1 \leq s \leq t \leq k$，由于 $u_t(i) = 0$ 当 $i \notin S_{t-1}$ 时，(16) 给出

$$\langle x_s, u_t \rangle = \sum_{i \in S_{t-1}} x_s(i) u_t(i) = 0.$$

W接下来证明，对于某个 (15)，存在 $i_* \in [m]$。为证明这一点，我们首先对 t 使用归纳法证明

$$|S_t| \geq \frac{m}{2^t} - t \qquad \text{for every } t \in \{0, \ldots, k\}. \tag{18}$$

对于 $t = 0$，(18) 成立，因为 $|S_0| = m$。设 $t \in [k]$，并假设归纳假设 $|S_{t-1}| \geq m/2^{t-1} - (t-1)$ 成立。那么

$$|S_t| \geq \frac{|S_{t-1}| - t}{2} \geq \frac{1}{2}\left(\frac{m}{2^{t-1}} - (t-1) - t\right) = \frac{m}{2^t} - t + \frac{1}{2} \geq \frac{m}{2^t} - t,$$

其中，第一个不等式由(17)得出，第二个不等式则由归纳假设得出。因此，(18)成立，而假设 $m \geq k2^k + 1$ 给出

$$|S_k| \geq \frac{m}{2^k} - k \geq 2^{-k} > 0.$$

固定 $i_* \in S_k$。由于 $S_k \subseteq S_t$，对于每个 $t \in [k]$，都有 $i_* \in S_t$；而根据 S_t 的定义，对于每个 $t \in [k]$，都有 $u_t(i_*) = 1$。由于 $u_0 = \mathbf{1}$，坐标 i_* 对所有 $t \in \{0, \ldots, k\}$ 满足 $u_t(i_*) = 1$。

剩下需要证明(14)。固定 $i \in [m]$。若 $i \in S_k$，则 $u_t(i) = 1$ 对所有 $t \in \{0, \ldots, k\}$ 都成立，理由与 i_*的论证相同，且(14) 中的和为零。若 $i \notin S_k$，定义 $p = \min\{t \in [k] : i \notin S_t\}$。那么 $i \in S_t$ 对$t < p$成立，并且 $i \notin S_t$ 对 $t \geq p$成立，因此

$$u_t(i) = 1 \quad \text{for } t < p, \qquad u_p(i) \in [-1, 1), \qquad u_t(i) = 0 \quad \text{for } t > p,$$

其中，第一个关系由 $u_0 = \mathbf{1}$ 和 S_t的定义得出，第二个关系由 $i \in S_{p-1} \setminus S_p$ 和 S_p的定义得出，第三个关系则由 $i \notin S_{t-1}$ 和 u_t 在 S_{t-1}之外的定义得出。因此，序列$u_0(i) \ldots, u_k(i)$ 的形式为 $\mathbf{1}, \ldots, \mathbf{1}, u_p(i)\, \mathbf{0}, \ldots, \mathbf{0}$，并且

$$\sum_{t=1}^{k} |u_t(i) - u_{t-1}(i)| \leq |u_p(i) - 1| + |0 - u_p(i)| \leq 3,$$

至此，证明完成。 □

5.3 定理 1 及 推论 1 和 2 的证明

要应用 引理 1，还需为 v_t 在 第5行 的 算法 1 中选取向量 y^* 以及一个分布$\langle y^*, v_t\rangle = c$ 使得对所有 $t \in [k]$ 均成立。对于 定理 1，我们通过将 引理 3 应用于y玩家的策略，并取 $y^* = e_{j_*}$ 作为坐标 j_*，使得 $v_t(j_*) = 1$对所有 t均成立，从而得到 $c = 1$。

定理 1 的证明。固定任意一个确定性学习动态，并运行算法 1 共 k 轮。我们通过将引理 3 应用于 $x_1, \ldots, x_t$，在第 4 行选取u_t；并通过将引理 3 应用于以 m替代n 的 $y_1, \ldots, y_t$，在第5行选取 v_t。这是可行的，因为 $m, n \geq k2^k + 1$，且向量 u_t 和 v_t 根据引理 3 中的 (13) 满足第 4 行和第5行的要求。这两项选择都仅使用截至第 t 轮所选的策略，因此，第 7 行第 t 轮的反馈在算法于第 $t + 1$轮选择策略之前便已确定。

我们验证引理 1 的假设。根据引理 3 中的 (14)，对于每个 $i \in [m]$，都有 $\sum_{t=1}^{k} |u_t(i) - u_{t-1}(i)| \leq 3$。根据两次应用中引理 3 的 (15)，存在 $i_* \in [m]$ 和 $j_* \in [n]$，使得对于每个 $t \in \{0, \ldots, k\}$，$u_t(i_*) = 1$ 且 $v_t(j_*) = 1$，其中 $v_0 = \mathbf{1}$。结合 $y^* = e_{j_*}$和 $c = 1$，对于每个 $t \in [k]$，我们有$\langle y^*, v_t\rangle = v_t(j_*) = c$。因此，引理 1 的所有假设均成立。

因此，根据 引理 1，矩阵 $A = A_k$属于 $[-1, 1]^{m\times n}$，有序对 (e_{i_*}, e_{j_*}) 是 A 的一个纯纳什均衡，其值为零；在固定博弈 A 上运行的算法的遗憾满足所需的遗憾下界 (3) 对每个 $H \in [T]$都成立，由此完成证明。 □

推论 推论 *1*的证明。 记 $d = \min\{m, n\}$。由于 $\max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \geq (\mathsf{Reg}_H^x + \mathsf{Reg}_H^y)/2$，只需证明第一个不等式。首先假设 $d \geq 4$ 并应用 定理 1 取

$$k = \min\{T, \lfloor \tfrac{1}{2} \log_2 d \rfloor\},$$

满足 $1 \leq k \leq T$。由于 $2^k \geq k + 1$ 和 $2k \leq \log_2 d$，我们有 $k2^k + 1 \leq (k + 1)2^k \leq 4^k \leq d$，因此 $m, n \geq k2^k + 1$。此外，对于 $z \geq 1$，$\lfloor z \rfloor \geq z/2$ 给出 $\lfloor \frac{1}{2} \log_2 d \rfloor \geq \frac{1}{4} \log_2 d$，因此，无论 $k = T$还是 $k = \lfloor \frac{1}{2} \log_2 d \rfloor$，

$$\min\{H, k\} \geq \tfrac{1}{4} \min\{H, \log_2 d\} \qquad \text{for every } H \in [T]. \tag{19}$$

由(3) 中定理 1 的(19)的第一个不等式可得

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \geq \frac{1}{3} \min\{H, k\} \geq \frac{1}{12} \min\{H, \log_2 d\} \qquad \text{for every } H \in [T],$$

这就证明了 $d \geq 4$的结论。

下面考虑 $d \in \{2, 3\}$的情形。在这种情况下，我们构造一个矩阵，使得对每个 $H \in [T]$，都有 $\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \geq 1$，这大于所述界，因为 $\frac{1}{12} \log_2 3 < 1$。由于第一个策略 x_1不依赖于 A，我们可以选择 $i \in [m]$，使得 $x_1(i) \leq 1/m$，并定义 $A = (2e_i - \mathbf{1})\mathbf{1}^\top \in [-1, 1]^{m\times n}$。于是，对每个 $y \in \Delta_n$，都有 $Ay = 2e_i - \mathbf{1}$，因此由 (7)和 $i_* = i$ 可得，对每个 $H \in [T]$，$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \geq \sum_{t=1}^{H} (1 - \langle x_t, 2e_i - \mathbf{1}\rangle) = \sum_{t=1}^{H} 2(1 - x_t(i)) \geq 2(1 - 1/m) \geq 1$，从而完成证明。 □

证明推论 *2*。令 A 为推论 1中的矩阵，并固定 $H \in [T]$。根据(2) 以及推论 1中的第一个不等式，我们有

$$\max_{x\in\Delta_m} \langle x, A\bar{y}_H\rangle - \min_{y\in\Delta_n} \langle \bar{x}_H, Ay\rangle = \frac{\mathsf{Reg}_H^x + \mathsf{Reg}_H^y}{H} \geq \frac{1}{12H} \min\{H, \log_2 \min\{m, n\}\}.$$

如果（$\bar{x}_H$，$\bar{y}_H$）是 A 的 ε-近似纳什均衡，那么左侧至多为 2ε，因此$\varepsilon \geq \frac{1}{24H} \min\{H, \log_2 \min\{m, n\}\}$，从而证明完毕。 □

6 定理 2 的证明定理 2

本节证明定理 2。我们首先概述本节的证明。向量 u_t按照第 5 节中的方式构造。对于向量 v_t，当 $n < k2^k + 1$ 时，不一定存在一个坐标 j_*，使得对所有 $t \in [k]$ 都有 $v_t(j_*) = 1$。不过，注意到引理 1 中的假设 $\langle y^*, v_t \rangle = c$ 并不要求 y^*是纯策略，我们可以构造向量 v_t 以及一个分布 $y^* \in \Delta_n$，使其对于某个常数 $c \in (0, 1/4]$满足该假设。然后，根据引理 1，我们可以得到社会遗憾下界$\frac{c}{3}\min\{H, k\}$。

分布 y^* 必须满足 $\langle y^*, v_t \rangle = c$ 对所有 $t \in [k]$均成立，而每个 v_t 必须在第 t轮根据 y_1、……、y_t确定，因为反馈 $\ell_t = A_t^\top x_t$取决于 v_t。因此，我们递归地构造 y^*。在每一轮 t，我们结合 v_t构造一个满足 $\langle y_t^*, v_s \rangle = c$ 对所有 $s \le t$均成立的分布 $y_t^* \in \Delta_n$，并令 $y^* = y_k^*$。从均匀分布 $y_0^* = \mathbf{1}/n$开始，我们选择

$$y_t^*(j) = y_{t-1}^*(j)\big(1 + \eta_t\, v_t(j)\big) \qquad \text{for } j \in [n],$$

其中 $\eta_t > 0$ 是步长。由于 $|v_t(j)| \le 1$，当$\eta_t < 1$时，该乘法更新会保持所有坐标为正。

接下来需要选择向量 v_t 和步长 η_t。根据引理 2，在 $2t$个线性约束下选择向量 v_t，这些约束包括第5行算法 1 中要求其与 y_1、……、y_t 正交的条件，以及另外 t个约束，在这些约束下，y_t^*的坐标之和为 1，并且对所有 $s < t$均有 $\langle y_t^*, v_s \rangle = c$ 成立。随后选择步长 η_t，使得 $\langle y_t^*, v_t \rangle = c$。我们将证明，如果 $c \le 1/4$ 且$n \ge 8k\exp(4kc^2)$，则对所有 $t \in [k]$ 均有 $\eta_t \le 1/2$ 成立；这正是第 6.1 节中引理 4 的假设。第二个条件等价于 $4kc^2 \le \log(n/(8k))$，这也是公式 (4) 中的常数 c 以及定理 2 中的下界分别包含 $\sqrt{\log(n/k)/k}$ 和 $\sqrt{k\log(n/k)}$项的原因。~~公式 (4) 中~~的常数 c 在 $n \ge 16k$时同时满足这两个条件，于是定理 2 和推论 3 在第 6.2 节中得到证明。

6.1 向量 v_t 和分布 y^* 的构造

下面的引理给出了上述向量 v_t 和分布 y^*，它们满足 算法 1 的第5行 所要求的正交性，以及假设 $\langle y^*, v_t \rangle = c$ 见引理 1。

引理 4。 假设 k、n 是正整数，且 c 是满足 $0 < c \le 1/4$ 和 $n \ge 8k\exp(4kc^2)$ 的实数。那么，对于任意依次给定的 y_1、…、$y_k \in \Delta_n$，可以选择仅取决于 y_1、…、y_t 的 $v_t \in [-1, 1]^n$，并且在第 k 轮之后，存在一个分布 $y^* \in \Delta_n$，使得

$$\langle y_s, v_t \rangle = 0 \qquad \textit{for all } 1 \le s \le t \le k, \tag{20}$$
$$\langle y^*, v_t \rangle = c \qquad \textit{for all } t \in [k]. \tag{21}$$

证明。 我们递归地构造向量 v_1、…、v_k 和分布 y_0^*、…、y_k^*，并通过对 $t \in \{0, \ldots, k\}$ 进行归纳法证明

$$\text{(a) } y_t^* \in \Delta_n, \quad \text{(b) } \langle y_t^*, v_s \rangle = c \text{ for all } s \in [t], \quad \text{(c) } \|y_t^*\|_2^2 \le \frac{\exp(4tc^2)}{n}. \tag{22}$$

对于 $t = k$，条件(22a)和(22b) 说明 $y^* = y_k^*$ 属于 Δ_n 并满足(21)，而(22c) 仅用于归纳步骤。

对于 $t = 0$，我们设定 $y_0^* = \mathbf{1}/n$。那么(22a) 成立，(22b)不包含任何条件，并且(22c) 成立，因为$\|y_0^*\|_2^2 = 1/n$。

设 $t \in [k]$，并假设归纳假设 v_1、...、 v_{t-1} 和 y^*_{t-1} 已由y_1、...、 y_{t-1} 构造而成，并且(22) 在以 $t-1$ 代替 t时成立。我们定义 y^*_t如下：

$$y^*_t(j) = y^*_{t-1}(j)\big(1 + \eta_t\, v_t(j)\big) \qquad \text{for } j \in [n], \tag{23}$$

其中向量 $v_t \in [-1,1]^n$ 和步长 $\eta_t > 0$ 将在下文中选取。对于任意给定的 v_t 和η_t，定义（23）以及归纳假设 (22a)和 (22b) 给出

$$\sum_{j=1}^n y^*_t(j) = 1 + \eta_t \langle y^*_{t-1}, v_t\rangle, \tag{24}$$

$$\langle y^*_t, v_s\rangle = c + \eta_t \sum_{j=1}^n y^*_{t-1}(j) v_s(j) v_t(j) \qquad \text{for all } 1 \le s < t, \tag{25}$$

$$\langle y^*_t, v_t\rangle = \langle y^*_{t-1}, v_t\rangle + \eta_t \sum_{j=1}^n y^*_{t-1}(j) v_t(j)^2, \tag{26}$$

$$\|y^*_t\|_2^2 = \|y^*_{t-1}\|_2^2 + 2\eta_t \sum_{j=1}^n y^*_{t-1}(j)^2 v_t(j) + \eta_t^2 \sum_{j=1}^n y^*_{t-1}(j)^2 v_t(j)^2. \tag{27}$$

我们首先选择 v_t 作为 $[-1,1]^n$ 中的一个向量，使其满足以下 $2t$ 线性约束：

$$\langle y_s, v_t\rangle = 0 \qquad \text{for all } s \in [t], \tag{28}$$

$$\langle y^*_{t-1}, v_t\rangle = 0, \tag{29}$$

$$\sum_{j=1}^n y^*_{t-1}(j) v_s(j) v_t(j) = 0 \qquad \text{for all } 1 \le s < t. \tag{30}$$

约束（28）正是（20）中要求的正交性。在约束（29）和（30）下，恒等式（24）和（25）给出 $\sum_{j=1}^n y^*_t(j) = 1$ 和 $\langle y^*_t, v_s\rangle = c$，对所有 $s < t$ 均成立，这符合第 t 轮次中（ 22a ）和（ 22b ）的要求。约束（ 28 ）–（30）要求 v_t 与 $2t$ 个向量 $y_1, \ldots, y_t, y^*_{t-1}$ 以及 $(y^*_{t-1}(j)v_s(j))_{j\in[n]}$ 对于 $1 \le s < t$ 正交。因此，引理 2 的第一个断言结合 $d = n$ 和 $r = 2t$，给出一个满足（28）–（30）的向量 $v_t \in [-1,1]^n$，且其严格位于 -1 与 1 之间的坐标至多有 $2t$ 个。由于$-v_t$ 也具有这两个性质，我们可以选择 v_t 的符号，使得

$$\sum_{j=1}^n y^*_{t-1}(j)^2 v_t(j) \le 0, \tag{31}$$

这使得项 $2\eta_t \sum_{j=1}^n y^*_{t-1}(j)^2 v_t(j)$ 在 (27) 中非正。

接下来我们选择 η_t 使得 $\langle y^*_t, v_t\rangle = c$。由 (26) 和 (29)，可得 $\langle y^*_t, v_t\rangle = \eta_t \sum_{j=1}^n y^*_{t-1}(j) v_t(j)^2$，并且我们首先对这个和进行下界估计。由归纳假设 (22c)，可得

$$\|y^*_{t-1}\|_2^2 \le \frac{\exp(4(t-1)c^2)}{n} \le \frac{\exp(4kc^2)}{n} \le \frac{1}{8k} \le \frac{1}{8t},$$

其中，第三个不等式由假设 $n \ge 8k\exp(4kc^2)$ 得出，第二个和第四个不等式则由$t \le k$得出。因此

$$\sum_{j=1}^n y^*_{t-1}(j) v_t(j)^2 \ge \sum_{j:\,|v_t(j)|=1} y^*_{t-1}(j) = 1 - \sum_{j:\,|v_t(j)|<1} y^*_{t-1}(j) \ge 1 - \sqrt{2t}\,\|y^*_{t-1}\|_2 \ge \frac{1}{2}, \tag{32}$$

其中，第一个不等式由 $y^*_{t-1}(j) \geq 0$ 得出，等式由 $\sum_{j=1}^n y^*_{t-1}(j) = 1$ 得出，第二个不等式由柯西–施瓦茨不等式以及至多有 $2t$ 个坐标满足 $|v_t(j)| < 1$ 这一事实得出，最后一个不等式则由 $\|y^*_{t-1}\|_2^2 \leq 1/(8t)$得出。我们定义

$$\eta_t = \frac{c}{\sum_{j=1}^n y^*_{t-1}(j) v_t(j)^2}. \tag{33}$$

由 (32) 以及假设 $c \leq 1/4$可知，步长满足 $0 < \eta_t \leq 2c \leq 1/2$。采用这些 v_t 和η_t后，(23) 定义了 y^*_t，且 v_t 和 y^*_t仅依赖于 y_1、……、 y_t。

接下来只需证明 (22) 在第 t 轮成立。由 (29)、(30) 和 (33)，恒等式 (24)–(26) 得到$\sum_{j=1}^n y^*_t(j) = 1$ 以及 $\langle y^*_t, v_s \rangle = c$，对所有 $s \in [t]$ 均成立。每个坐标都满足 $y^*_t(j) \geq y^*_{t-1}(j)(1-\eta_t) \geq 0$，这是由$y^*_{t-1}(j) \geq 0, |v_t(j)| \leq 1$ 以及 $\eta_t \leq 1/2$ 得出的。因此 (22a) 和 (22b) 成立。对于 (22c)，恒等式 (27) 给出

$$\|y^*_t\|_2^2 \leq \left(1 + \eta_t^2\right) \|y^*_{t-1}\|_2^2 \leq \exp(4c^2) \|y^*_{t-1}\|_2^2 \leq \frac{\exp(4tc^2)}{n},$$

其中，第一个不等式由 (31) 和 $|v_t(j)| \leq 1$得出，第二个由 $\eta_t \leq 2c$ 和 $1 + z \leq \exp(z)$得出，第三个则由归纳假设 (22c)得出。因此，对任意 (22) 都成立 $t \in \{0, \ldots, k\}$。

向量 v_t 根据 (28)满足 (20)，而分布 $y^* = y^*_k$ 属于 Δ_n，并根据 (22a)和 (22b) 满足 (21)，其中 $t = k$，由此完成证明。 □

6.2 定理 2 和 推论 3的证明

我们首先验证式 c 中的常数 (4) 满足引理 4 的假设，然后像证明定理 1时那样应用引理 1。

定理 2的证明。 引理 4 假设 $0 < c \leq 1/4$ 且 $n \geq 8k\exp(4kc^2)$。第一个假设由(4) 和 $n \geq 16k$得出。第二个假设等价于 $4kc^2 \leq \log(n/(8k))$，这成立，因为

$$4kc^2 \leq \frac{1}{4} \log \frac{n}{k} \leq \log \frac{n}{k} - \log 8 = \log \frac{n}{8k},$$

其中，第一个不等式由(4) 得出，第二个不等式则由 $n \geq 16k$ 以及3 4 $\log\ 16 = \log\ 8$得出。

固定任意确定性学习动态，并运行算法 1 k 轮次。我们通过对 $x_1, \ldots, x_t$应用引理 3，在第 4 行选择 u_t；并通过对 $y_1, \ldots, y_t$应用引理 4，在第5行选择 v_t。引理 3适用，因为$m \geq k2^k + 1$；根据引理 3中的(13)和引理 4中的(20)，向量 u_t 和 v_t 满足第 4 行和第5行的要求。这两次选择仅使用截至第 t轮所选的策略，因此，第 7 行中第 t轮的反馈在算法于第 $t+1$轮选择策略之前就已确定。

现在验证引理 1的假设。根据引理 3中的(14)，对于每个 $i \in [m]$，都有 $\sum_{t=1}^k |u_t(i) - u_{t-1}(i)| \leq 3$。根据引理 3中的(15)，存在 $i_* \in [m]$，使得对于每个 $t \in \{0, \ldots, k\}$，都有 $u_t(i_*) = 1$。根据引理 4中的(21)，存在一个分布 $y^* \in \Delta_n$，使得对于每个 $t \in [k]$，都有 $\langle y^*, v_t \rangle = c$。因此，引理 1的所有假设均成立。

因此，根据 引理 1，矩阵 $A = A_k$属于 $[-1, 1]^{m \times n}$，对(e_{i_*}, y^*) 是 A 的一个纳什均衡，其值为零，并且在固定博弈 A 上运行的算法的遗憾满足 (5) 对每个 $H \in [T]$均成立。对每个 H 满足 $k \leq H \leq T$的情形，(4) 和 (5) 给出

$$\mathsf{Reg}_H^x + \mathsf{Reg}_H^y \geq \frac{ck}{3} = \frac{1}{12} \min\left\{k, \sqrt{k \log \frac{n}{k}}\right\}, \ \max\{\mathsf{Reg}_H^x, \mathsf{Reg}_H^y\} \geq \frac{ck}{6} = \frac{1}{24} \min\left\{k, \sqrt{k \log \frac{n}{k}}\right\},$$

即为(6) 并完成证明。 □

推论 *3* 的证明推论 *3*。 我们取陈述中的条件 $n \geq 16^{1/(1-\alpha)}$，即 n 充分大，并通过应用 定理 2 （取 $k = \lfloor \frac{1}{2} \log_2 m \rfloor$）来证明该结论。由于 $m \geq n \geq 16$，我们有 $\frac{1}{2} \log_2 m \geq 1$，而 $z/2 \leq \lfloor z \rfloor \leq z$ 对于 $z \geq 1$ 给出 $\frac{1}{4} \log_2 m \leq k \leq \frac{1}{2} \log_2 m$。

我们首先验证定理 2 的假设 $k \leq T$、$m \geq k2^k + 1$ 和 $n \geq 16k$。由于 $2k \leq \log_2 m$，因此有 $k \leq \log_2 m \leq T$ 和 $m \geq 4^k \geq k2^k + 1$，其中最后一个不等式由 $2^k \geq k + 1$ 得出。此外，$2 \log 2 \geq 1$，且假设 $\log m \leq n^\alpha$ 给出 $k \leq \frac{1}{2} \log_2 m = \log_2 \sqrt{m} \leq \log \sqrt{m} \leq n^\alpha$，从而有 $n/k \geq n^{1-\alpha} \geq 16$。因此 $n \geq 16k$，定理 2 适用。

接下来需要对(6) 中定理 2 的最小值里的两项给出下界。由于 $k \geq \frac{1}{4} \log_2 m = \frac{\log m}{4 \log 2}$，$m \geq n$，且 $n/k \geq n^{1-\alpha}$，我们有

$$k \geq \frac{\log m}{4 \log 2} \geq \frac{\sqrt{\log m \log n}}{4 \log 2}, \qquad \sqrt{k \log \frac{n}{k}} \geq \sqrt{\frac{\log m}{4 \log 2} (1-\alpha) \log n} = \sqrt{\frac{1-\alpha}{4 \log 2}} \sqrt{\log m \log n}.$$

在定理 2中应用(6)并结合 $H = T \geq k$，因此可得结论，其中 $C_\alpha = \text{最小值}\{1, \sqrt{4(1-\alpha) \log 2\}/(96 \log 2)}$，证明至此完成。 □

7 讨论

本节讨论定理 1对近似纳什均衡查询复杂度的一个推论，以及一个尚未解决的问题。

近似纳什均衡的查询复杂度。在第2节中，每位玩家的目标是最小化其遗憾，而遗憾取决于所有轮次中选定的策略。另一个不同的问题是，用较少的查询计算一个未知收益矩阵 $A \in [-1, 1]^{m \times n}$ 的近似纳什均衡。每次查询中，算法选择一对策略 $(x_t, y_t) \in \Delta_m \times \Delta_n$ 并观察乘积 Ay_t 和 $A^\top x_t$，它们是在第2节中的反馈向量。最后一次查询后，算法输出一对策略，且只有这一输出需要构成一个近似纳什均衡。

该模型中的查询复杂度由 Hadiji 等人（2023年）和Karmarkar 等人（2026年）进行了研究。Kornowski 和 Shamir （2026年）研究了查询可以是任意向量的模型。两种模型的查询数量最多相差两倍，因为与任意向量的乘积可以表示为与两个策略的乘积的线性组合。在其他设置下，针对观察纯策略组合收益的算法（Fearnley 等人，2015年）以及针对欧氏球上的双线性鞍点问题的一阶方法（Ouyang 和 Xu，2021）已有下界。这些下界并不能直接推出上述查询模型中的下界。

目前已知的最佳上界为 $O(\sqrt{\log m \log n/\varepsilon})$ 次查询（Nemirovski，2004年；Nesterov，2005年）以及 $\widetilde{O}(\varepsilon^{-2/3})$ 次查询（Karmarkar 等人，2026年），其中 $\widetilde{O}$ 隐藏了关于 m、n 和 $1/\varepsilon$ 的多对数因子。此外，$\min\{m, n\}$ 次查询即可确定 A，因此足以应对每个 ε。至于下界，Hadiji等人（2023年）证明，当 ε 至多为 d^{-4} 量级时，对于 $d \times d$ 个矩阵，至少需要 $\Omega(\min\{d, \log(1/(d\varepsilon))/\log d\})$ 次查询；Kornowski和Shamir（2026年）则证明，当 m 和 n 相对于 $1/\varepsilon$ 均为多项式规模时，确定性算法需要 $\widetilde{\Omega}(\varepsilon^{-2/3})$ 次查询。Kornowski 和 Shamir （ 2026 年）的下界在相差多对数因子的意义下，与Karmarkar等人（2026年）的上界相匹配；然而，当 ε 为常数时，这一下界和Hadiji等人（2023年）的下界都不会随行动数量增加。定理 1 推出以下针对确定性算法的下界，该下界会随行动数量增加。

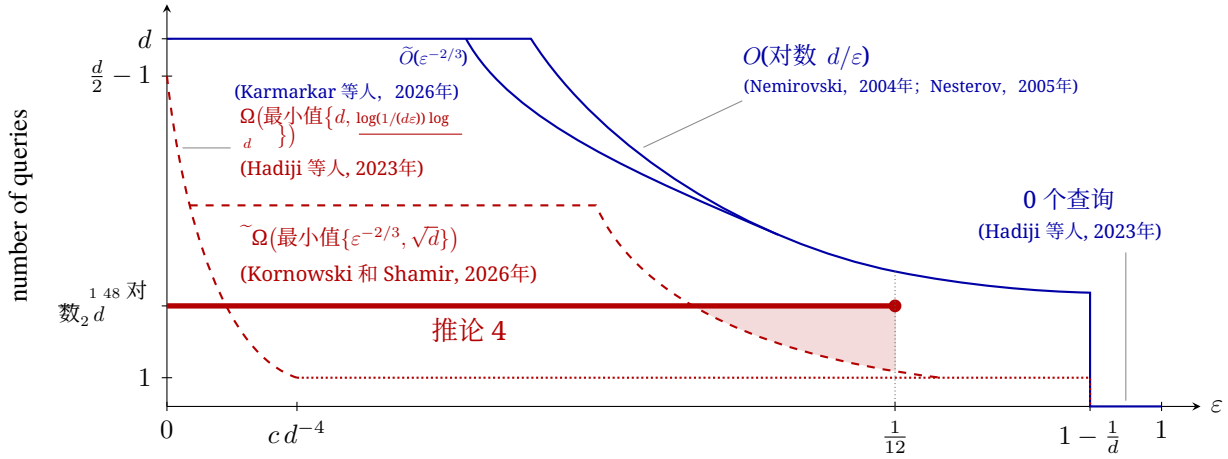


图2： 确定性算法输出一对$(\widetilde{x},\ \widetilde{y})$ 且满足$\max_{x\in\Delta_d}\langle x,\ A\widetilde{y}\rangle - \min_{y\in\Delta_d}\langle\widetilde{x},\ Ay\rangle \le 2\varepsilon$ 对每个 $A\in[-1,1]^{d\times d}$所需查询次数的上界（蓝色实线）和下界（红色实线、虚线和点线）。该图改编自 Hadiji 等人(2023年）中的图1。红色粗线是推论 4的下界，阴影区域表示该下界高于其他下界的范围。数值 c是一个绝对常数。

推论 4. 设 m、$n\ge 4$ 和 $\varepsilon\in[0,1/6]$。考虑一个确定性算法，对于每个 $A\in[-1,1]^{m\times n}$，依次选择 N 对 $(x_t, y_t)\in\Delta_m\times\Delta_n$，其中每一对都取决于此前选择的各对以及$(Ay_s, A^\top x_s)_{s<t}$，随后输出 $(\widetilde{x},\widetilde{y})\in\Delta_m\times\Delta_n$，使其满足

$$\max_{x\in\Delta_m}\langle x, A\widetilde{y}\rangle - \min_{y\in\Delta_n}\langle\widetilde{x}, Ay\rangle \le \varepsilon.$$

于是 $N\ge\frac{1}{48}\log_2\min\{m,n\}$。

证明。我们通过将定理 1应用于这样一种学习动态来证明该结论：它先选取算法的 N 对，然后输出该算法。记 $d=\min\{m,n\}$，并令 $k=\lfloor\frac{1}{2}\log_2 d\rfloor$，由 $k\ge 1$ 可知$d\ge 4$。与推论 1的证明类似，我们有 $d\ge 4^k\ge k2^k+1$ 以及 $k\ge\frac{1}{4}\log_2 d$。若 $N\ge k$，则$N\ge\frac{1}{4}\log_2 d$，结论成立。因此我们假设 $N<k$。

考虑如下包含 $T=k$ 轮次的学习动态。在每一轮 $t\le N$，两位玩家选择算法的第 t 个配对 (x_t, y_t)；在每一轮 $t>N$，他们选择算法的输出 $(\widetilde{x},\widetilde{y})$。学习动态的反馈为 $(g_s,\ell_s)=(Ay_s, A^\top x_s)$，因此，这些策略是历史 $(x_s, y_s, g_s, \ell_s)_{s=1}^{t-1}$ 的函数。

根据 定理 1 以及 $T=k$，存在一个矩阵 $A\in[-1,1]^{m\times n}$，使得该学习动态满足遗憾$_k^x$ +遗憾$_k^y\ge k/3$。接下来，我们为这个矩阵给出社会遗憾的上界。对于所有 $x^*\in\Delta_m$ 和 $y^*\in\Delta_n$，有

$$\langle x^*, Ay_t\rangle - \langle x_t, Ay^*\rangle \le \begin{cases} 2 & \text{for } t\le N,\\ \varepsilon & \text{for } t>N,\end{cases}$$

其中，第一种情况由 $A\in[-1,1]^{m\times n}$ 得出，第二种情况由 $(x_t,\ y_t)=(\widetilde{x},\ \widetilde{y})$ 以及关于算法输出的假设得出。因此我们有

$$\frac{k}{3}\le \mathsf{Reg}_k^x+\mathsf{Reg}_k^y=\max_{x^*\in\Delta_m,\,y^*\in\Delta_n}\sum_{t=1}^{k}\big(\langle x^*, Ay_t\rangle-\langle x_t, Ay^*\rangle\big)\le 2N+(k-N)\varepsilon\le 2N+\frac{k}{6},$$

其中，第一个不等式由 定理 1 得出，等式由 (2) 得出，最后一个不等式由 $\varepsilon \le 1/6$ 得出。因此 $N \ge k/12 \ge \frac{1}{48}\log_2 d$，证明完毕。 □

一个 ε-近似纳什均衡 $(\widetilde{x}, \widetilde{y})$ 满足推论 4 中的不等式，只需将 ε 替换为 2ε，因此该推论适用于输出具有 $\varepsilon \le 1/12$ 的 ε-近似纳什均衡的算法。对于常数 $\varepsilon \in (0, 1/12]$ 和 $m = n$，下界与上界 $O(\sqrt{\log\, m \log\, n/\varepsilon})$ 相差一个常数因子。下界不依赖于 ε，因此当 ε 较小时，它远小于上界。图2 展示了 $m = n = d$ 对应的这些界。

开放问题。 我们的下界适用于确定性学习动态，而它们能否推广到使用内部随机化的学习动态，仍是一个开放问题。在 算法 1 中，矩阵 A 由两个算法选取的策略构造而成。如果算法使用内部随机化，该矩阵便取决于随机化的结果，因此这种构造无法得到一个固定矩阵，使得其期望遗憾满足相同的下界。Kornowski 和 Shamir (2026年) 关于矩阵-向量乘积数量的下界也仅适用于确定性算法。

人工智能使用声明

我们下界中收益矩阵的最终构造借助 GPT-6 Astra 完成。作者对这一构造及其表述进行了大幅修改，使其更加清晰。修订和整理稿件时使用了 Claude Fable 5.1、5.2 和 Opus 5.5。作者核验了所有证明，并对本文内容负责。

致谢

TT 获得日本学术振兴会科学研究费助成事业项目（编号 JP26K21297）和日本科学技术振兴机构 BOOST 项目（编号 JP- MJBY25D4）的资助。

参考文献

Ioannis Anagnostides、Gabriele Farina、Christian Kroer、Chung-Wei Lee、Haipeng Luo 和 Tuomas Sandholm。具有 $O(\log T)$ 交换遗憾的多人博弈中的无耦合学习动态。载于 神经信息处理系统进展，第 35 卷，第 3292–3304 页。Curran Associates, Inc.，2022 年。 3

Ioannis Anagnostides、Alkis Kalavasis、Tuomas Sandholm 和 Manolis Zampetakis。《博弈中计算稀疏均衡的复杂性以及无遗憾学习的下界》。载于 第 *15* 届理论计算机科学创新会议，第 287 卷，第 5:1–5:24 页，2024 年。 3

Nicolo Cesa-Bianchi 和 Gábor Lugosi。 预测、学习与博弈。剑桥大学出版社，2006 年。 1

Xi Chen 和 Binghui Peng。《博弈中的对冲：更快地收敛外部遗憾和交换遗憾》。载于神经信息处理系统进展，第 33 卷，第 18990–18999 页。Curran Associates, Inc.，2020年。 2，3

Constantinos Daskalakis、Alan Deckelbaum 和 Anthony Kim。《零和博弈的近乎最优无遗憾算法》。载于第二十二届 ACM-SIAM 离散算法年度研讨会论文集，第 235–254 页。工业与应用数学学会，2011年。1、2、3、5 Constantinos Daskalakis、Maxwell Fishelson 和 Noah Golowich。《一般博弈中的近乎最优无遗憾学习》。载于神经信息处理系统进展，第 34 卷，第 27604–27616 页，2021年。3 Gabriele Farina、Ioannis Anagnostides、Haipeng Luo、Chung-Wei Lee、Christian Kroer 和 Tuomas Sandholm。《一般凸博弈的近乎最优无遗憾学习动态》。载于神经信息处理系统进展，第 35 卷，第 39076–39089 页。Curran Associates, Inc.，2022年。3 John Fearnley、Martin Gairing、Paul W. Goldberg 和 Rahul Savani。《通过收益查询学习博弈均衡》。机器学习研究期刊，16(39):1305–1344，2015年。17 Yoav Freund 和 Robert E Schapire。《在线学习的决策理论推广及其在提升法中的应用》。计算机与系统科学期刊，55(1):119–139，1997年。1 Yoav Freund 和 Robert E. Schapire。《使用乘法权重的自适应博弈》。博弈与经济行为，29(1):79–103，1999年。1、4 Hedi Hadiji、Sarah Sachs、Tim van Erven 和 Wouter M Koolen。《迈向刻画零和矩阵博弈中学习（近似）纳什均衡的一阶查询复杂度》。载于神经信息处理系统进展，第 36 卷，第 13356–13373 页。Curran Associates, Inc.，2023年。3、17、18 Ishani Karmarkar、Liam O’Carroll 和 Aaron Sidford。《以近乎最优的矩阵向量乘法复杂度求解矩阵博弈》。arXiv 预印本 arXiv:2601.02347，2026年。3、17、18 Guy Kornowski 和 Ohad Shamir。《基于单纯形的矩阵博弈的预言机复杂度》。arXiv 预印本 arXiv:2412.06990v4，2026年。初步版本发表于 2025年学习理论第 38 届年度会议论文集。3、17、18、19 Nick Littlestone 和 Manfred K Warmuth。《加权多数算法》。信息与计算，108(2):212–261，1994年。1 Arkadi Nemirovski。《具有收敛速率 $O(1/t)$的近端方法：用于具有 Lipschitz 连续单调算子和光滑凸凹鞍点问题的变分不等式》。SIAM 优化期刊，15(1):229–251，2004年。3、17、18 Yurii Nesterov。《非光滑函数的光滑最小化》。数学规划，103(1):127–152，2005年。3、17、18 Yuyuan Ouyang 和 Yangyang Xu。《凸凹双线性鞍点问题的一阶方法复杂度下界》。数学规划，185(1):1–35，2021年。17 Sasha Rakhlin 和 Karthik Sridharan。《利用可预测序列进行优化、学习和博弈》。载于神经信息处理系统进展，第 26 卷，第 3066–3074 页。Curran Associates, Inc.，2013年。1、3、5 Ashkan Soleymani、Georgios Piliouras 和 Gabriele Farina。《谨慎乐观：一种在一般博弈中实现近乎常数遗憾的元算法》。载于第 26 届 ACM 经济学与计算会议论文集，第 870 页，2025年。3

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，第 28 卷，第 2989–2997 页。Curran Associates, Inc.，2015年。 1，2，3

Taira Tsuchiya。乐观 Hedge 在双人零和博弈中的紧致遗憾上下界。载于第*29*届人工智能与统计国际会议论文集，第 300 卷，第 4870–4878 页。机器学习研究论文集，2026a。 1，2，3，6

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* 预印本 *arXiv:2608.04149*，2026年。3

一个具有两列的博弈

推论 3 假设 $\log m \le n^{\alpha}$ 以及 n 足够大。以下命题表明，这些条件无法完全去除，因为它给出了一个针对 $n = 2$ 的确定性学习动态，其遗憾至多为 4，对每个 m而言，而 $\sqrt{\log m \log n}$ 会随 m增长。

命题1。 设 T、$m \ge 1$ 和 $n = 2$。那么存在一种确定性学习动态，使得对于每个 $A \in [-1,1]^{m\times 2}$ 和每个 $H \in [T]$,

$$\mathsf{Reg}_H^x \le 4, \qquad \mathsf{Reg}_H^y \le 4.$$

证明。 在第1轮和第2轮，玩家选择 $x_1 = x_2 = e_1$、$y_1 = e_1$和 $y_2 = e_2$。这几轮中x-玩家的反馈由 A的两列 $g_1 = Ae_1$ 和 $g_2 = Ae_2$ 组成；由于每位玩家都可以使用对手的反馈，因此双方在每一轮 $t \ge 3$都可以使用这些列。对于每个 $t \ge 3$，玩家选择 $x_t = x^*$ 和 $y_t = y^*$，其中（x^*，y^*）是 A 的一个纳什均衡，双方通过固定的确定性过程根据 g_1 和g_2 计算得出。

对于所有 $x \in \Delta_m$ 和 $y \in \Delta_2$，我们有

$$\langle x - x_t, Ay_t\rangle \le \begin{cases} 2 & \text{for } t \le 2, \\ 0 & \text{for } t \ge 3, \end{cases} \qquad \langle y_t - y, A^\top x_t\rangle \le \begin{cases} 2 & \text{for } t \le 2, \\ 0 & \text{for } t \ge 3, \end{cases}$$

其中，$t \le 2$ 的界由 $A \in [-1,1]^{m\times 2}$推出，而 $t \ge 3$ 的界由（x_t，y_t）$=$（x^*，y^*）以及$\langle x,\ Ay^*\rangle \le \langle x^*,\ Ay^*\rangle \le \langle x^*,\ Ay\rangle$推出。因此，对于任意 $H \in [T]$,

$$\mathsf{Reg}_H^x = \max_{x\in\Delta_m} \sum_{t=1}^{H} \langle x - x_t, Ay_t\rangle \le 4, \qquad \mathsf{Reg}_H^y = \max_{y\in\Delta_2} \sum_{t=1}^{H} \langle y_t - y, A^\top x_t\rangle \le 4,$$

这就完成了证明。 □