arXiv:2302.06652v1 [cs.LG] 13 Feb 2023

A对抗策略性对手，实现更低遗憾

LE CO NG DINH、Tri-Dung Nguyen、Alain Zemkoho、Long Tran-Thanh

我们研究这样一类在线学习问题：学习者对对手的行为有额外了解，即处于博弈论情境中，对手通常会采用某种无外部遗憾学习算法。在此假设下，我们提出两种新的在线学习算法：精确正则化跟随领导者（AFTRL）和 Prod 最佳响应（Prod-BR）。这两种算法会充分利用这些额外信息，同时在额外信息不准确的最坏情况中仍保持无遗憾性质。具体而言，与 Prod-BR 的 $O(\sqrt{T})$ 动态遗憾相比，面对无外部遗憾对手时，AFTRL 可实现 $O(1)$ 外部遗憾或 $O(1)$ 前向遗憾。据我们所知，我们的算法是首个考虑前向遗憾、且能在面对策略性对手时实现 $O(1)$ 遗憾的算法。在使用精确乘法权重更新（AMWU）（AFTRL 的一个特例）进行零和博弈时，我们实现了向纳什均衡的最后一轮收敛。我们还提供了数值实验，进一步支持理论结果。具体来说，与当前最先进的方法（如乘法权重更新（MWU）及其乐观版本 OMWU）相比，我们的方法能实现显著更优的遗憾界和最后一轮收敛速率。

1 引言

无悔算法因其出色的最坏情况性能保证，在在线学习和算法博弈论文献中广受欢迎 [Cesa-Bianchi 和 Lugosi， 2006]。具体而言，使用这些算法选择要采取的策略，可以严格保证：无论遇到何种序列，事后看来，平均收益都不会显著低于最佳固定策略。凭借这一性质，无悔算法常用于与对手博弈以及求解双人零和博弈；在自博弈情境下，最终会实现向纳什均衡（NE）的平均收敛 [Dinh 等人，2021b、Lanctot 等人，2017、Zinkevich 等人， 2007]。然而，为了使遗憾界保持较小，无悔算法（例如乘法权重更新、正则化跟随者、镜像下降 [Abernethy 等，2009、Nemirovskij 和 Yudin， 1983]）需要采用较小的学习率，导致策略组合变化缓慢。这使得无悔算法采取的策略序列具有可预测性，因为每个策略组合都会与其前序策略相关联。因此，面对无悔学习对手时，学习者/玩家每轮遇到的损失序列并非完全任意的对抗序列，因此最坏情况性能保证对学习者而言过于悲观。因此，在与无悔算法（策略性对手）博弈等情境下，理想的做法是开发一种学习算法，既能利用额外的结构，又能在最坏情况下保持无遗憾性质，并回答以下问题：

我们能否利用无悔算法？

除了追求更好的遗憾界之外，我们还关注最后一轮收敛，而不仅仅是向纳什均衡的平均收敛。在双人零和博弈中，乘法权重更新（MWU） [Freund 和 Schapire， 1999] 或跟随正则化领导者（FTRL）等无悔算法只能实现平均收敛，而无法实现最后一轮收敛。事实上， [Bailey 和 Piliouras，2018、Mertikopoulos 等人， 2018] 的最新研究表明，在许多情况下，乘法权重更新和跟随正则化领导者会导致偏离纳什均衡而发散。平均收敛不仅会增加计算和内存开销，而且在使用神经网络进行求解时也会带来困难，因为平均化并不总是可行的 [Bowling 等人，2015]。对于博弈论以及在线优化的现代应用（例如训练生成对抗网络） [Daskalakis 等人， 2017], ，最后一轮收敛在整个过程中发挥着至关重要的作用，因此，开发能够实现最后一轮收敛的算法至关重要。

为探究本文上述两个目标，在在线线性优化的设定下，我们**首先**提出一种新算法——精确正则化跟随领导者（AFTRL），该算法能够利用无外部遗憾对手，实现 $O(1)$ 外部遗憾或 $O(1)$ 前向遗憾，同时在最坏情况下保持最先进的遗憾界 $O\left(\sqrt{\sum_{t=1}^{T}\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q^2}\right)$。我们还将结果推广到另一类在线学习问题，以此展示方法的通用性，并为其提出一种新算法——精确镜像下降（AMD），具有类似的前向遗憾界。据我们所知，我们是首个研究密集利用并针对无外部遗憾对手实现 $O(1)$ 前向遗憾的团队。**其次**，我们探索了 [Sani 等人提出的 (A,B)-Prod 算法思路，并据此提出一种新算法——Prod 最佳响应（Prod-BR），在我们的设定下能够实现更强的性能保证。具体而言，Prod-BR 在针对无外部遗憾对手时，实现 $O(\sqrt{T})$ 动态遗憾，同时在最坏情况下保持 $O(\sqrt{T}\log(T))$ 外部遗憾。**第**三，对于带有熵正则项的 AFTRL 特例——精确乘法权重更新（AMWU），我们证明这种新算法能够在双人零和博弈中实现最后一轮收敛，因此可在许多实际应用中充当高效的博弈求解器。此外，这也为最后一轮收敛相关研究带来了新的贡献

。**最后**，为展示 AMWU 的实际效率，我们表明，在若干随机矩阵博弈和元博弈（例如四子棋或 Disc）中，我们的算法显著优于 MWU 和 OMWU [Daskalakis 和 Panageas, 018a，Rakhlin and Sridharan， 2013]，其平均损失和动态遗憾更低，最后一轮收敛速度也更快，优势显著 [Czarnecki 等人， 2020]。

2 相关工作

针对无悔学习者的在线学习： Deng 等人 [Deng 等人， 2019] 研究了一个类似的设置，其中智能体在重复博弈中与无外部遗憾对手进行博弈。假设智能体了解博弈结构（即收益矩阵、玩家效用）， [Deng 等人，2019] 提出了一种智能体的固定策略（通过求解优化问题），使智能体能够保证获得斯塔克尔伯格值，该值在某些博弈（例如一般和博弈）中是最优的。尽管 [Deng 等人， 2019] 提供了一种针对无外部遗憾对手的规划方案，但它无法应用于许多实际情形：在这些情形中，环境或博弈结构未知（即智能体无法预先计算斯塔克尔伯格策略），或者对手不遵循无悔算法（即无法保证对一般对手的性能）。Chiang 等人 [Chiang 等人， 2012] 和 Rakhlin 等人 [Rakhlin and Sridharan，2013] 研究了另一种设置：智能体在第 t轮做出决策之前可以获得预测 M_t 的 $\boldsymbol{x}_t$。[1] 新算法——乐观正则化领导者跟随算法

（OFTRL）的外部遗憾与{v1}呈线性关系，$\sum_{t=1}^{T} \|\boldsymbol{x}_t - M_t\|_*^2$。然而，如果预测准确（即 $M_t \approx \boldsymbol{x}_t$），人们可以期待比 OFTRL 的无外部遗憾更强的性能保证。直观地说，由于 OFTRL 为预测固定赋予权重 1 M_t [2]，它限制了额外知识在学习过程中的作用。我们的新算法（AFTRL 和 AMD）推广了 [Rakhlin and Sridharan 2013] 的工作，以便在最坏情况下保持无前向遗憾性质 [Saha 等人，2012]的同时，进一步利用学习过程中的额外知识。

最后一轮收敛： 尽管博弈论和在线学习领域已对无遗憾学习动态的平均收敛进行了广泛研究（例如， [Cesa-Bianchi 和 Lugosi，2006，Freund 和 Schapire， 1999]），但最后一轮收敛直到近几年才成为研究课题，这源于它在博弈论和优化中的应用。这项研究始于 [Bailey 和 Piliouras，2018，Mertikopoulos 等人， 2018], 的负面结果：他们证明，在具有内部均衡的博弈中，如果智能体使用乘法权重更新，那么最后一轮的策略会远离纳什均衡，并朝边界移动。最近， [Daskalakis 和 Panageas，018a，Wei 等人， 2020] 证明，在具有唯一纳什均衡的双人零和博弈中，如果双方都采用一种称为乐观乘法权重更新（OMWU）的乘法权重更新变体，那么动态过程将在最后一轮收敛到纳什均衡。在非对称情形下， [Dinh 等人，2021a] 提出了非对称博弈最后一轮收敛算法（LRCA），该算法要求一名智能体掌握极小极大均衡的估计，因此限制了算法的适用范围。在我们的工作中，我们证明，我们的方法 AMWU 无需这一要求，即可在最后一轮收敛到双人零和博弈的纳什均衡，而且其收敛速度快于 OMWU 和乘法权重更新。

3 预备知识

我们考虑在线线性优化设定：在第 t轮次，学习者选择策略 $\boldsymbol{f}_t \in \mathrm{F}$，其中 $\mathrm{F} \subset [0,1]^n$ [3] 是一个凸紧集。与此同时，环境

[1]正如我们在引理 4.1 中证明的那样，与策略性对手对弈可以准确预测 $\boldsymbol{x}_t$。[2]算法 1 中的利用率 α。
[3]对于策略和损失向量的有界域，所有结果仍然成立。

给出损失向量 $\boldsymbol{x}_t \in [0,1]^n$，学习者承受的损失为：$\boldsymbol{f}_t,\ \boldsymbol{x}_t$。学习者的目标是在经过 T轮次后最小化总损失：$\min_{f_1,\ \ldots f_T} \sum_{t=1}^{T} \boldsymbol{f}_t,\ \boldsymbol{x}_t$，这可以转化为最小化以下动态遗憾：

*De*定义 *3.1*（动态遗憾 *[Besbes* 等人，*2015]*）。动态遗憾定义为 ：

$$DR_T := \sum_{t=1}^{T} \left(\langle \boldsymbol{f}_t, \boldsymbol{x}_t \rangle - \arg\min_{\boldsymbol{g} \in \mathcal{F}} \langle \boldsymbol{g}, \boldsymbol{x}_t \rangle \right)$$

在不了解 $\boldsymbol{x}_t$ 的情况下，通常无法实现无动态遗憾。因此，争取实现无外部遗憾更为可行 [Cesa-Bianchi 和 Lugosi, 2006]:

定义 *3.2*（无外部遗憾）。设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ 是环境产生的一系列混合损失。学习者所使用、能够生成一系列混合策略 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ 的算法称为无外部遗憾算法，若满足：

$$\lim_{T \to \infty} \frac{R_T}{T} = 0, \text{ where } R_T := \min_{\boldsymbol{f} \in \mathcal{F}} \sum_{t=1}^{\top} \left(\langle \boldsymbol{f}_t, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}, \boldsymbol{x}_t \rangle \right).$$

在本文中，由于我们假设学习者对对手有额外的了解，因此，与传统的无外部遗憾相比，学习者能够实现一种更强的性能概念，即：

定义 *3.3*（前向遗憾 *[Saha* 等人，*2012]*）。前向遗憾定义为：

$$FR_T := \sum_{t=1}^{T} \left(\langle \boldsymbol{f}_t, \boldsymbol{x}_t \rangle - \langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle \right), \text{where } \boldsymbol{g}_{t+1} = \arg\min_{\boldsymbol{g} \in \mathcal{F}} G_{t+1}(\boldsymbol{g}) = \langle \boldsymbol{g}, \sum_{s=1}^{t} \boldsymbol{x}_s + \boldsymbol{x}_{t+1} \rangle + \frac{R(\boldsymbol{g})}{\eta}.$$

特别是，以下引理表明，如果一个算法具有无前向遗憾性质，那么它也是一种无外部遗憾算法，但反之则不成立 [4]。

引理 3.4。设 $\boldsymbol{g}_t$ 如上所定义，则对于任意 $\boldsymbol{f} \in \mathcal{F}$，以下关系成立：

$$\sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle \leq \langle \boldsymbol{f}, \sum_{t=1}^{T} \boldsymbol{x}_t \rangle + \frac{R(\boldsymbol{f})}{\eta}.$$

在第 6 节中，我们研究一种更简单形式的在线线性优化，其中损失函数具有以下形式：$\boldsymbol{A}^T \boldsymbol{y}$ 其中 $\boldsymbol{y}$ 是单纯形 Δ_m中的一点。我们还令 F 为单纯形 Δ_n，这种博弈通常称为零和矩阵博弈 $\boldsymbol{A}$。双人零和博弈 $\boldsymbol{A}$中的纳什均衡可以由约翰·冯·诺伊曼极小极大定理 [Neumann, 1928]表示为：

$$\max_{\boldsymbol{y} \in \Delta_m} \min_{\boldsymbol{f} \in \Delta_n} \boldsymbol{f}^\top \boldsymbol{A} \boldsymbol{y} = \min_{\boldsymbol{f} \in \Delta_n} \max_{\boldsymbol{y} \in \Delta_m} \boldsymbol{f}^\top \boldsymbol{A} \boldsymbol{y} = v \tag{1}$$

对于某个 $v \in \mathbb{R}$。满足方程 (1) 的点 $(\boldsymbol{f}^*, \boldsymbol{y}^*)$ 是该博弈的纳什均衡。

4 精确跟随正则化领导者

为了具有无（外部）遗憾性质，FTRL 和 OMD 等常用算法需要采用较小的学习率 η（即，参见 [Shalev-Shwartz 等人，2012]）：$\eta = O\left(\frac{1}{\sqrt{T}}\right)$。基于这一观察，~~我~~我们可以证明以下引理，它在我们的分析中发挥着重要作用：

[4]此引理的证明见 B.1。

引理 4.1. 设 f_t、 f_{t+1} 是无外部遗憾算法（即 *FTRL*、*OMD*）的两个连续策略。那么对于任意范数 $\|.\|_q$，我们有：

$$\|f_{t+1} - f_t\|_q = O(\frac{1}{\sqrt{T}}).$$

完整证明见附录 *B.3*。

现在，设 R 是关于 $\|.\|_p$ 范数的 β强凸函数。不失一般性，假设 $\min_{f\in\mathcal{F}} R(f) = 0$。

算法 1： 精确正则化跟随领袖算法

输入： 学习率 $\eta > 0$、利用率 $\alpha \geq 1$、$f_1 = \arg\min_{f\in\mathcal{F}} R(f)$。**输出：** 下一次策略更新

$$f_{t+1} = \arg\min_{f\in\mathcal{F}} F_{t+1}(f) = \langle f, \sum_{s=1}^{t} x_s + \alpha x_t\rangle + \frac{R(f)}{\eta}.$$

精确正则化跟随领袖算法（AFTRL）包含两个重要参数：利用率 α和学习率 η。学习率 η通过避免被利用来稳定策略更新，而利用率 α衡量历史数据 $\sum_{s=1}^{t} x_s$与预测 x_t之间的相对权重。直观地说，在预测准确的情况下 x_t，较大的α将提升 AFTRL 的性能，因为 x_t相比历史数据 $\sum_{s=1}^{t} x_s$能更好地描述下一个损失向量 x_{t+1}。改变 α会得到文献中的不同算法。当 $\alpha = 0$时，该算法将变为经典的跟随正则化领导者（FTRL） [Abernethy 等人， 2009]。当 $\alpha = 1$时，AFTRL 将恢复 Rakhlin and Sridharan， 2013]提出的乐观跟随正则化领导者方法（OFTRL） [。我们可以得到 AFTRL 算法的以下遗憾界：

定理 4.2。 设$\mathcal{F} \subset [0,1]^n$ 是一个凸紧集，且 R 是关于 $\|.\|_p$ 范数的 β-强凸函数，并且$\min_{f\in\mathcal{F}} R(f) = 0$. 记 $\|.\|_q$ 为对偶范数，其中$1/p + 1/q = 1$。那么，面对一般对手时，*AFTRL* 实现的外部遗憾为 $O(1)$，或前向遗憾为 $O\left(\sqrt{\sum_{t=1}^{T}(\|x_t - x_{t-1}\|_q)^2}\right)$。更重要的是，面对无外部遗憾对手（即 *FTRL*、在线镜像下降）时，*AFTRL* 实现的外部遗憾为 $O(1)$，或前向遗憾为 $O(1)$。

Proof 证明概要。我们首先证明，对于任意环境策略，AFTRL 满足：

$$\sum_{t=1}^{T}\langle f_t, x_t\rangle - \frac{1}{\alpha}\langle f', \sum_{t=1}^{T} x_t\rangle - \frac{\alpha-1}{\alpha}\sum_{t=1}^{T}\langle g_t, x_t\rangle \leq \frac{1}{\eta\alpha}R(f') + \frac{\eta\alpha}{\beta}\sum_{t=1}^{T}(||x_t - x_{t-1}||_q)^2. \quad (2)$$

定义 h_{t+1}如下： $h_{t+1} = \arg\min_{f\in\mathcal{F}} H_{t+1}(f) = \langle f, \sum_{s=1}^{t} x_s + \alpha x_{t+1}\rangle + \frac{R(f)}{\eta}$.

直观地说，策略 h_{t+1} 的表现会远优于普通的跟随正则化领导者，因为智能体可以提前一步观察对手的策略。注意，我们可以将智能体的总损失分解如下

$$\sum_{t=1}^{T}\langle f_t, x_t\rangle = \sum_{t=1}^{T}\langle f_t - h_t, x_t - x_{t-1}\rangle \sum_{t=1}^{T}\langle f_t - h_t, x_{t-1}\rangle + \sum_{t=1}^{T}\langle h_t, x_t\rangle. \quad (3)$$

证明的关键步骤是，我们可以通过数学归纳法证明：

$$\sum_{t=1}^{T}\langle f_t - h_t, x_{t-1}\rangle + \sum_{t=1}^{T}\langle h_t, x_t\rangle \leq \frac{1}{\alpha}\langle f', \sum_{t=1}^{T} x_t\rangle + \frac{\alpha-1}{\alpha}\sum_{t=1}^{T}\langle g_t, x_t\rangle + \frac{1}{\eta\alpha}R(f'),\ \forall f' \in \mathcal{F}. \quad (4)$$

此外，利用 β-强凸函数的性质，我们可以推导出：

$$\begin{aligned} &\|\boldsymbol{x}_{t-1}-\boldsymbol{x}_t\|_q \geq \frac{\beta}{\eta\alpha}\|\boldsymbol{h}_t-\boldsymbol{f}_t\|_p \\ \implies &\sum_{t=1}^{T}\langle \boldsymbol{f}_t-\boldsymbol{h}_t, \boldsymbol{x}_t-\boldsymbol{x}_{t-1}\rangle \leq \sum_{t=1}^{T}||\boldsymbol{f}_t-\boldsymbol{h}_t||_p||\boldsymbol{x}_t-\boldsymbol{x}_{t-1}||_q \leq \frac{\eta\alpha}{\beta}\sum_{t=1}^{T}(||\boldsymbol{x}_{t-1}-\boldsymbol{x}_t||_q)^2. \end{aligned} \tag{5}$$

利用等式（3）中的不等式（4）和（5），我们可以推导出不等式（2）。

现在，针对一般对手，如果 $\sum_{t=1}^{T}\boldsymbol{f}_t,\ \boldsymbol{x}_t-\boldsymbol{f}, \sum_{t=1}^{T}\boldsymbol{x}_t \leq 0$ 那么根据定义，AFTRL 具有 $O(1)$ 外部遗憾。在 $\sum_{t=1}^{T}\boldsymbol{f}_t,\ \boldsymbol{x}_t-\boldsymbol{f}, \sum_{t=1}^{T}\boldsymbol{x}_t \geq 0$ 的情况下，利用不等式（2）并设置 $\eta\alpha=\sqrt{\beta R(\boldsymbol{f}')/(\sum_{t=1}^{T}(\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q)^2)}$，我们有：

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t,\boldsymbol{x}_t\rangle-\sum_{t=1}^{T}\langle \boldsymbol{g}_t,\boldsymbol{x}_t\rangle \leq \frac{\alpha}{\alpha-1}\sqrt{R(\boldsymbol{f}')\sum_{t=1}^{T}(\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q)^2/\beta}=O(\sum_{t=1}^{T}(\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q)^2).$$

对于未知界 $\sum_{t=1}^{T}(\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q)^2$,，我们可以使用附录 B.2 中所示的倍增技巧，以获得类似的遗憾界。

面对无外部遗憾对手，利用引理 4.1，我们得到：

$$\sum_{t=1}^{T}(\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q)^2=\sum_{t=1}^{T}(O(\frac{1}{\sqrt{T}}))^2=O(1).$$

因此，不等式（2）变为：

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t,\boldsymbol{x}_t\rangle-\frac{1}{\alpha}\langle \boldsymbol{f}',\sum_{t=1}^{T}\boldsymbol{x}_t\rangle-\frac{\alpha-1}{\alpha}\sum_{t=1}^{T}\langle \boldsymbol{g}_t,\boldsymbol{x}_t\rangle \leq \frac{1}{\eta\alpha}R(\boldsymbol{f}')+\frac{\eta\alpha}{\beta}O(1)=O(1).$$

采用类似的推理分析一般对手时，AFTRL 面对无外部遗憾对手可达到 $O(1)$ 外部遗憾或 $O(1)$ 前向遗憾。完整证明见附录 B.6。□

备注（**AFTRL 与 OFTRL 的比较**）。尽管 *AFTRL* 和 *OFTRL* 都采用了利用“可预测序列”的相同思路，但二者存在显著差异。首先，*OFTRL* 对可预测序列的依赖程度固定为1，而 *AFTRL* 则允许灵活控制对可预测序列的依赖程度（即通过参数 α）。因此，与 *OFTRL* 相比，在预测准确的情况下，*AFTRL* 能取得好得多的表现，图 *1* 中的实验结果也证实了这一点。其次，在最坏情况下，*AFTRL* 能保证比 *[Rakhlin and Sridharan, 2013]* 中 *OFTRL* 的外部遗憾界更强的前向遗憾界。

我们的方法可以扩展到另一类算法，例如镜像下降[Shalev- Shwartz 等人， 2012]。我们提出精确镜像下降（AMD）[5]，其遗憾界与 AFTRL 类似：

定理 4.3。设 $\mathcal{F}$ 是巴拿赫空间 $\mathcal{B}$ 中的一个凸集。设 $\mathcal{R}:\mathcal{B}\to\mathbb{R}$ 是 $\mathcal{F}$ 上的一个 β-强凸函数，且相对于某个范数 $\|.\|_p$。记 $\|.\|_q$ 其对偶范数为 $1/p+1/q=1$。对于任意环境策略和任意 $\boldsymbol{f}'\in\mathcal{F}$，*AMD* 给出

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t,\boldsymbol{x}_t\rangle-\frac{1}{\alpha}\langle \boldsymbol{f}',\boldsymbol{x}_t\rangle-\frac{\alpha-1}{\alpha}\langle \boldsymbol{g}_{t+1},\boldsymbol{x}_t\rangle \leq \frac{\eta\alpha}{2\beta}\sum_{t=1}^{T}\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_q^2+\frac{\mathcal{R}_{max}^2}{\eta\alpha},$$

其中 $\mathrm{R}_{max}^2=\max_{f\in\mathrm{F}}\mathrm{R}(f)-\min_{f\in\mathrm{F}}\mathrm{R}(f)$。

[5]伪代码见附录 A.2

证明。我们将 $\boldsymbol{h}_{t+1}$ 定义如下：

$$\boldsymbol{h}_{t+1} = \arg\min_{\boldsymbol{h}\in\mathcal{F}} H_{t+1}(\boldsymbol{h}) = \eta\langle \boldsymbol{h}, \alpha\boldsymbol{x}_t\rangle + D_{\mathcal{R}}(\boldsymbol{h}, \boldsymbol{g}_t).$$

对于任意 $\boldsymbol{f}' \in \mathrm{F}$,

$$\begin{aligned}
&\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \frac{1}{\alpha}\langle \boldsymbol{f}', \boldsymbol{x}_t\rangle - \frac{\alpha-1}{\alpha}\langle \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle \\
&= \langle \boldsymbol{f}_t - \boldsymbol{h}_{t+1}, \boldsymbol{x}_t - \boldsymbol{x}_{t-1}\rangle + \langle \boldsymbol{f}_t - \boldsymbol{h}_{t+1}, \boldsymbol{x}_{t-1}\rangle \\
&+ \langle \boldsymbol{h}_{t+1} - \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle + \frac{1}{\alpha}\langle \boldsymbol{g}_{t+1} - \boldsymbol{f}', \boldsymbol{x}_t\rangle.
\end{aligned} \tag{6}$$

利用对偶范数的性质，我们推导得出

$$\begin{aligned}
\langle \boldsymbol{f}_t - \boldsymbol{h}_{t+1}, \boldsymbol{x}_t - \boldsymbol{x}_{t-1}\rangle &\le \|\boldsymbol{f}_t - \boldsymbol{h}_{t+1}\|_p \|\boldsymbol{x}_{t-1} - \boldsymbol{x}_t\|_q \\
&\le \frac{\beta}{2\eta\alpha}\|\boldsymbol{f}_t - \boldsymbol{h}_{t+1}\|_p^2 + \frac{\eta\alpha}{2\beta}\|\boldsymbol{x}_{t-1} - \boldsymbol{x}_t\|_q^2.
\end{aligned} \tag{7}$$

注意，对于任意 $\boldsymbol{g} \in \mathrm{F}$ 和 $\boldsymbol{f} = \arg\min_{f\in\mathcal{F}} \boldsymbol{f},\ \boldsymbol{x} + D_{\mathcal{R}}(\boldsymbol{f},\ \boldsymbol{c})$，我们有以下不等式（例如参见 [Beck 和 Teboulle, 2003]）：

$$\langle \boldsymbol{f} - \boldsymbol{g}, \boldsymbol{x}\rangle \le D_{\mathcal{R}}(\boldsymbol{g}, \boldsymbol{c}) - D_{\mathcal{R}}(\boldsymbol{g}, \boldsymbol{f}) - D_{\mathcal{R}}(\boldsymbol{f}, \boldsymbol{c}).$$

由此可得

$$\begin{aligned}
\langle \boldsymbol{f}_t - \boldsymbol{h}_{t+1}, \boldsymbol{x}_{t-1}\rangle &\le \frac{1}{\eta\alpha}(D_{\mathcal{R}}(\boldsymbol{h}_{t+1}, \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{h}_{t+1}, \boldsymbol{f}_t) - D_{\mathcal{R}}(\boldsymbol{f}_t, \boldsymbol{g}_t)), \\
\langle \boldsymbol{h}_{t+1} - \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle &\le \frac{1}{\eta\alpha}(D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{h}_{t+1}) - D_{\mathcal{R}}(\boldsymbol{h}_{t+1}, \boldsymbol{g}_t)), \\
\langle \boldsymbol{g}_{t+1} - \boldsymbol{f}', \boldsymbol{x}_t\rangle &\le \frac{1}{\eta}(D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_{t+1}) - D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{g}_t)).
\end{aligned}$$

将上述不等式相加可得

$$\begin{aligned}
&\langle \boldsymbol{f}_t - \boldsymbol{h}_{t+1}, \boldsymbol{x}_{t-1}\rangle + \langle \boldsymbol{h}_{t+1} - \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle + \frac{1}{\alpha}\langle \boldsymbol{g}_{t+1} - \boldsymbol{f}', \boldsymbol{x}_t\rangle \\
&\le \frac{1}{\eta\alpha}(D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_{t+1}) - D_{\mathcal{R}}\left(\boldsymbol{h}_{t+1}, \boldsymbol{f}_t\right) - D_{\mathcal{R}}(\boldsymbol{f}_t, \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{h}_{t+1})).
\end{aligned} \tag{8}$$

利用强凸函数的性质，我们得到

$$D_{\mathcal{R}}(\boldsymbol{h}_{t+1}, \boldsymbol{f}_t) \ge \frac{\beta}{2}\|\boldsymbol{h}_{t+1} - \boldsymbol{f}_t\|_p^2; \quad D_{\mathcal{R}}(\boldsymbol{f}_t, \boldsymbol{g}_t) \ge \frac{\beta}{2}\|\boldsymbol{f}_t - \boldsymbol{g}_t\|_p^2. \tag{9}$$

将不等式 (7)、(8) 和 (9) 代入等式 (6)，可得

$$\begin{aligned}
\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \frac{1}{\alpha}\langle \boldsymbol{f}', \boldsymbol{x}_t\rangle - \frac{\alpha-1}{\alpha}\langle \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle &\le \frac{\eta\alpha}{2\beta}\|\boldsymbol{x}_{t-1} - \boldsymbol{x}_t\|_q^2 \\
+ \frac{1}{\eta\alpha}(D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_{t+1})) - \frac{\beta}{2\eta\alpha}\|\boldsymbol{f}_t - \boldsymbol{g}_t\|_p^2
\end{aligned}$$

对 $t = 1, \ldots\ldots, T$ 求和可得，对于任意 $f' \in \mathrm{F}$,

$$\begin{aligned}
&\sum_{t=1}^{T}\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \frac{1}{\alpha}\langle \boldsymbol{f}', \boldsymbol{x}_t\rangle - \frac{\alpha-1}{\alpha}\langle \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle \\
&\leq \frac{\eta\alpha}{2\beta}\sum_{t=1}^{T}\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q^2 + \frac{\mathcal{R}_{max}^2}{\eta\alpha} - \frac{\beta}{2\eta\alpha}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{g}_t\|_p^2 \\
&\leq \frac{\eta\alpha}{2\beta}\sum_{t=1}^{T}\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q^2 + \frac{\mathcal{R}_{max}^2}{\eta\alpha}.
\end{aligned}$$

其中 $\mathrm{R}_{max}^2 = \max_{f\in\mathrm{F}} \mathrm{R}(f) - \min_{f\in\mathrm{F}} \mathrm{R}(f)$.

利用以下不等式，对于任意给定的 $f, \in \mathrm{F}$ 和 $g_{t+1} = \arg\min_{g\in\mathcal{F}} \eta\langle g, x_t\rangle + D_{\mathcal{R}}(g, g_t)$（例如，参见 Beck 和 Teboulle， 2003]）：

$$\eta\langle \boldsymbol{g}_{t+1} - \boldsymbol{f}', \boldsymbol{x}_t\rangle \leq D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{f}', \boldsymbol{g}_{t+1}) - D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{g}_t)$$

我们可以推导出，对于任意 $f' \in \mathrm{F}$,

$$\sum_{i=1}^{T}\langle \boldsymbol{g}_{t+1}, \boldsymbol{x}_t\rangle \leq \sum_{i=1}^{T}\langle \boldsymbol{f}', \boldsymbol{x}_t\rangle + \frac{\mathcal{R}_{max}^2}{\eta} - \frac{\beta}{2\eta}\sum_{t=1}^{T}\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|^2.$$

因此，相对于 $\sum_{i=1}^{T} g_{t+1}, x_t$ 的遗憾（即 AMD 的前向遗憾）强于相对于 $\sum_{i=1}^{T} f, x_t \ \forall f, \in \mathrm{F}$ 的（外部）遗憾。□

5 采用最佳响应的 PROD

尽管 AFTRL 能保证无前向遗憾，但我们不禁会想：如果智能体拥有额外知识，能否取得更好的性能（例如，无动态遗憾）？ 本节将介绍 Prod 与最佳响应算法（Prod-BR），使智能体在面对无外部遗憾对手时能够实现无动态遗憾，同时在最坏情况下维持无外部遗憾性能。我们的 Prod-BR 算法变体的灵感来自 [Sani 等人 2014],提出的 (A,B)-Prod 算法；我们注意到，基于当前反馈的最佳响应策略可以利用无外部遗憾对手。Prod-BR 在主算法内部运行两个独立的算法（即跟随正则化领导者和 BR）。直观地说，跟随正则化领导者针对最坏情况维持性能保证，而 BR 算法则利用无外部遗憾对手所具有的额外结构，从而提升 Prod-BR 算法的效率。Prod-BR 能够在准确和不准确的额外知识之间取得平衡，使智能体在面对无外部遗憾对手时实现 $O(\sqrt{T})$ 动态遗憾，同时在最坏情况下维持 $\overline{O}(\sqrt{T}\log(T))$ 外部遗憾。我们首先证明，当对手遵循无外部遗憾

算法 2： Prod-Best Response 算法

输入： 学习率 $\eta > 0, \eta_1 \in (0, 1]$, 初始权重 $w_{1,R}, w_{1,BR}$, 正则化函数 $R(.)$。

$f_{t+1} = \arg\min_{f\in\mathcal{F}} F_{t+1}(f) = f, \sum_{s=1}^{t} x_s + {}^{R}(f)\eta; \quad BR_{t+1} = \arg\min_{f\in\mathcal{F}} f, x_t$

输出： 下一次策略更新 g_{t+1} 和下一个权重 $w_{t+1,R}$:

$$g_{t+1} = \frac{w_{t,R}}{w_{t,R} + w_{1,BR}} f_{t+1} + \frac{w_{1,BR}}{w_{t,FTRL} + w_{1,BR}} BR_{t+1}; \quad w_{t+1,R} = w_{t,R}(1 + \eta_1\langle BR_{t+1} - f_{t+1}, x_{t+1}\rangle).$$

算法（即采用最优学习率的跟随正则化领导者、在线镜像下降）时，针对先前反馈的最佳响应可以保证智能体获得以下结果：

引理 5.1. 设 x_t、 x_{t+1} 是无外部遗憾算法（即跟随正则化领导者、在线镜像下降）的两个连续策略。那么，我们有

$$\langle \boldsymbol{b}, \boldsymbol{x}_{t+1}\rangle - \langle \boldsymbol{c}, \boldsymbol{x}_{t+1}\rangle = O(\frac{1}{\sqrt{T}}),\ \textit{where}\ \boldsymbol{b} = \arg\min_{f\in\mathcal{F}}\langle \boldsymbol{f}, \boldsymbol{x}_t\rangle\ ,\ \boldsymbol{c} = \arg\min_{f\in\mathcal{F}}\langle \boldsymbol{f}, \boldsymbol{x}_{t+1}\rangle.$$

完整证明见附录 *B.7*。

接下来，我们可以证明关于 Prod-BR 算法性能的以下定理：

定理 5.2. 设智能体按照算法 *2* 执行 *Prod-BR*，参数为 $\eta = n/\sqrt{2T}$、 $\eta_1 = \overline{1/2}.\sqrt{\log(T)/T}$以及 $w_{1,BR} = 1-w_{1,R} = 1-\eta_1$。那么，面对一般对手时，它可实现 $O(\sqrt{T\log(T)})$外部遗憾，同时面对无外部遗憾对手时保持 $O(\sqrt{T})$动态遗憾。

证明。按照推论 1 中 (A,B)-Prod 的遗憾界分析，依据 [Sani 等人， 2014] 我们有

$$\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle \le \sum_{t=1}^{T}\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle + 2\sqrt{T\log(T)} \text{ and} \tag{10a}$$

$$\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle \le \sum_{t=1}^{T}\langle BR_t, \boldsymbol{x}_t\rangle + 2\log(2). \tag{10b}$$

由于智能体在算法 2 中对跟随正则化领导者采用了最优学习率，根据跟随正则化领导者的遗憾界分析（即参见 [Shalev-Shwartz 等人, 2012]），我们有

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \sum_{t=1}^{T}\langle \boldsymbol{f}, \boldsymbol{x}_t\rangle \le n\sqrt{2T}\ \ \forall \boldsymbol{f}\in\mathcal{F}.$$

结合不等式 (10a)，我们有

$$\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle - \sum_{t=1}^{T}\langle \boldsymbol{f}, \boldsymbol{x}_t\rangle \le 2\sqrt{T\log(T)} + n\sqrt{2T} = O(\sqrt{T\log(T)})\ \ \forall \boldsymbol{f}\in\mathcal{F},$$

因此，面对一般对手，$\overline{\text{Prod-BR}}$ 的外部遗憾为 $O(\sqrt{T\log(T)})$。对于证明的第二部分，利用不等式 (10b) 并结合引理 5.1，我们有

$$\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle - \arg\min_{f\in\mathcal{F}}\langle \boldsymbol{f}, \boldsymbol{x}_t\rangle \le \sum_{t=1}^{T}\langle BR_t, \boldsymbol{x}_t\rangle - \arg\min_{f\in\mathcal{F}}\langle \boldsymbol{f}, \boldsymbol{x}_t\rangle + 2\log(2)$$
$$= \sum_{t=1}^{T} O(\frac{1}{\sqrt{T}}) + 2\log(2) = O(\sqrt{T}),$$

因此，面对无外部遗憾对手，Prod-BR 的动态遗憾为 $O(\sqrt{T})$。 □

备注（Prod-BR 与 AFTRL）。在最坏情况下，*AFTRL* 相比 *Prod-BR* 提供了更好的性能保证（$O(\sqrt{T})$对比 $\overline{O(\sqrt{T\log(T)})}$）。然而，面对无外部遗憾对手时，与 *AFTRL* 的无前向遗憾相比，*Prod-BR* 提供了强得多的性能保证概念（无动态遗憾）。请注意，*Prod-BR* 和 *AFTRL* 都依赖于对手两种连续策略之间的距离较小这一条件。虽然对于许多无外部遗憾算法（如引理 *4.1* 中所述）这一条件成立，但也存在一些无外部遗憾算法（例如 *AdaHedge [De Rooij* 等人， *2014]*），其连续策略之间的距离具有如下形式： $\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_q = O(1/\sqrt{t})$ 其中 t

表示当前迭代。在这种情况下，沿用相同的论证，*AFTRL* 可实现 $O(1)$外部遗憾或 $O(\log(T))$前向遗憾，而 *Prod-BR* 则维持 $O(\sqrt{T})$动态遗憾。

6 精确乘法权重更新与最后一轮收敛

算法 3：精确乘法权重更新

输入： 学习率 $\eta > 0$，利用率 $\alpha > 0$，

$\boldsymbol{f}_1 = \boldsymbol{f}_2 = [1/n、\ldots\ldots、1/n]$.**输出：** 下一次更新

$$\boldsymbol{f}_{t+1}(i) = \frac{\boldsymbol{f}_t(i)e^{\eta((\alpha+1)\boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{y}_t - \alpha \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{y}_{t-1})}}{\sum_j \boldsymbol{f}_t(j)e^{\eta((\alpha+1)\boldsymbol{e}_j^\top \boldsymbol{A}\boldsymbol{y}_t - \alpha \boldsymbol{e}_j^\top \boldsymbol{A}\boldsymbol{y}_{t-1})}}, \tag{11}$$

$\boldsymbol{e}_i$ 表示在第 i个分量上的权重为 1 的单位向量。

现在我们转向第二组贡献，即：利用这一新的算法框架确保最后一轮收敛。我们证明，如果两位玩家都遵循精确乘法权重更新（AMWU）这一算法——它是带熵正则项的 AFTRL 的一个特例——那么在具有唯一纳什均衡的零和博弈中，动态过程将在最后一轮收敛至纳什均衡。[6]

这里需要指出，唯一纳什均衡的假设在以下意义下是一般成立的：由于具有非唯一均衡的零和博弈集合的勒贝格测度为零 [Van Damme,1991], 如果 $\boldsymbol{A}$ 的各个元素是从某个连续分布中独立抽样得到的，那么该博弈以概率 1 有唯一纳什均衡。我们将唯一性假设的放宽留待未来工作。我们的主要最后一轮收敛结果如下：

定理 6.1（**AMWU 的最后一轮收敛**）。设 $(\boldsymbol{f}^*, \boldsymbol{y}^*)$ 是矩阵博弈 $\boldsymbol{A}$ 的唯一纳什均衡。那么，对于 $b \in (0, 1]$ 且 η 足够小的情形，*AMWU* 的动态最后一轮收敛到该博弈的纳什均衡：$\lim_{t\to\infty}(\boldsymbol{f}_t, \boldsymbol{y}_t) = (\boldsymbol{f}^*, \boldsymbol{y}^*)$。

证明概要。我们将证明分为三个主要部分。首先，在第 6.1 节中，我们证明 K-L 散度 [Kullback 和 Leibler, 1951] 在第 t 个策略 $(\boldsymbol{f}_t, \boldsymbol{y}_t)$ 与 $(\boldsymbol{f}^*, \boldsymbol{y}^*)$ 之间会缩小至原来的 η^{2+b}，除非策略 $(\boldsymbol{f}_t, \boldsymbol{y}_t)$ 与 O 相差不超过 $\eta^{b/3}$:

$$RE((\boldsymbol{f}^*, \boldsymbol{y}^*)||(\boldsymbol{f}_{t+1}, \boldsymbol{y}_{t+1})) \le RE((\boldsymbol{f}^*, \boldsymbol{y}^*)||(\boldsymbol{f}_t, \boldsymbol{y}_t)) - \Omega(\eta^{b+2}).$$

关键的一步是观察到，量 $\boldsymbol{f}_{t-1}^\top \boldsymbol{A}\boldsymbol{y}_t - \boldsymbol{f}_t^\top \boldsymbol{A}\boldsymbol{y}_{t-1}$ 可以界定为：

$$\eta \boldsymbol{f}_{t-1}^\top \boldsymbol{A}\boldsymbol{y}_t - \eta \boldsymbol{f}_t^\top \boldsymbol{A}\boldsymbol{y}_{t-1} = -\eta^2 \sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t - \boldsymbol{e}_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha \boldsymbol{y}_{t-1}))^2$$

$$- \eta^2 \sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t - \boldsymbol{e}_i)^\top \boldsymbol{A}^\top((\alpha \boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 + O(\eta^{2+b}).$$

由上述结果可知，若起始点为均匀分布（即 $\boldsymbol{f}_1 = (1/n, \ldots\ldots, 1/n)$ 且 $\boldsymbol{y}_1 = (1/m, \ldots\ldots, 1/m)$），则 AMWU 至多经过：$O\left(\frac{\log(nm)}{\eta^{2+b}}\right)$迭代步数即可达到距 $O(\eta^{b/3})$ 足够近的状态。

其次，在第6.2节中，我们证明，当 $\eta^{b/3}$足够接近时，在学习率 η足够小的情况下，该点也意味着接近纳什均衡。证明与 Daskalakis 和 Panageas 的定理 3.2 [018a] 的证明密切相关。因此，对于任何各元素均非零的起始策略，只要学习率 η足够小，AMWU 就能任意接近纳什均衡。

最后，在第6.3节中，通过证明唯一极小极大均衡的谱半径小于一，我们表明 AMWU 的更新动力学在纳什均衡处局部收敛。

[6]稍微滥用一下记号，在本节中，我们同时使用 $\boldsymbol{f}(i)$ 和 $\boldsymbol{f}_i$ 来表示向量 $\boldsymbol{f}$ 的第 i 个元素。[7]我们稍后会在附录A.3中对其作出严格定义

这意味着，如果动态过程到达纳什均衡邻域内的某一点，就会收敛到纳什均衡。现在，将第一点和第二点应用于初始策略中各元素均非零的 AMWU算法的动态过程，可知当学习率 η 足够小时，AMWU 将任意接近纳什均衡（$\boldsymbol{f}^*$，$\boldsymbol{y}^*$）。接着，利用 AMWU 的局部收敛性质，即可直接推出定理 6.1 中的最后一轮收敛结果。

所有缺失的证明都可以在附录B.1中找到。

□

现在，我们给出上述三个关键步骤的证明。

6.1 减小 K-L距离

在本小节中，我们的部分分析基于 AMWU 的线性变体，其更新规则如下：

$$\boldsymbol{f}_{t+1}(i) = \frac{\boldsymbol{f}_t(i)(1+\eta((\alpha+1)e_i{}^\top \boldsymbol{A}\boldsymbol{y}_t - \alpha e_i{}^\top \boldsymbol{A}\boldsymbol{y}_{t-1}))}{\sum_j \boldsymbol{f}_t(j)(1+\eta((\alpha+1)e_j{}^\top \boldsymbol{A}\boldsymbol{y}_t - \alpha e_j{}^\top \boldsymbol{A}\boldsymbol{y}_{t-1}))}.$$

由于该变体的更新规则不包含指数部分，因此降低了分析的复杂度。我们首先通过以下引理量化 AMWU 两次连续更新之间的距离：

引理 6.2。 设 $\boldsymbol{f} \in \Delta_n$ 为最大化玩家的向量，$\boldsymbol{w}, \boldsymbol{z} \in \Delta_m$ 满足 $\|\boldsymbol{w}-\boldsymbol{z}\|_1 = O(\eta)$,$\eta\alpha = O(1)$，并假设 $\boldsymbol{f}\prime,\boldsymbol{f}\prime\prime$ 分别是 *AMWU* 及其线性变体在当前向量 $\boldsymbol{f}$ 以及最小化玩家的向量 $\boldsymbol{w}, \boldsymbol{z}$ 下的下一次迭代。则有

$$\|\boldsymbol{f}' - \boldsymbol{f}''\|_1 \text{ is } O(\eta^2) \text{ and } \|\boldsymbol{f}' - \boldsymbol{f}\|_1,\ \|\boldsymbol{f}'' - \boldsymbol{f}\|_1 \text{ are } O(\eta).$$

类似地，对于最小化玩家的向量 $\boldsymbol{y} \in \Delta_m$ 及其下一次迭代，该结论同样成立。（完整证明见附录*B.8*。）

分析第 t 个策略 $(\boldsymbol{f}_t, \boldsymbol{y}_t)$ 与 $(\boldsymbol{f}^*, \boldsymbol{y}^*)$ 之间的 K-L 散度时，我们会遇到量 $\boldsymbol{f}_{t-1}^\top \boldsymbol{A}\boldsymbol{y}_t - \boldsymbol{f}_t^\top \boldsymbol{A}\boldsymbol{y}_{t-1}$。为了界定这个量，我们需要以下引理：

引理6.3。 设 $\boldsymbol{f} \in \Delta_n$ 为最大化玩家的向量，$\boldsymbol{w}$、$\boldsymbol{z} \in \Delta_m$ 满足 $\|\boldsymbol{w}-\boldsymbol{z}\|_1 = O(\eta)$，$\eta\alpha = O(1)$，并假设 $\boldsymbol{f}\prime,\boldsymbol{f}\prime\prime$ 分别是 *AMWU* 及其线性变体在当前向量 $\boldsymbol{f}$ 以及最小化玩家的向量 $\boldsymbol{w}$、$\boldsymbol{z}$ 下的下一次迭代。则（当 η 足够小时）有

$$\begin{aligned}
&\eta(\boldsymbol{f}' - \boldsymbol{f})^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha\boldsymbol{z}) \\
&= \eta(\boldsymbol{f}'' - \boldsymbol{f})^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha\boldsymbol{z}) - O(\eta^3) \\
&= (1-O(\eta))\eta^2 \sum_i \boldsymbol{f}_i((\boldsymbol{f} - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha\boldsymbol{z}))^2 - O(\eta^3) \\
&= (1-O(\eta))\eta^2 \sum_i \boldsymbol{f}_i'((\boldsymbol{f}' - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha\boldsymbol{z}))^2 - O(\eta^3).
\end{aligned}$$

证明。根据引理 6.2，我们只需证明第二个等式。令 $\boldsymbol{B} = (\mathbb{1}_n \mathbb{1}_m^\top + \eta \boldsymbol{A})$。根据线性 AMWU 的定义，有 ~~$f_m = f_i (\boldsymbol{B}((\alpha+1)\boldsymbol{w}) - \alpha z) f^\top \boldsymbol{B} \quad \alpha+1 \quad \boldsymbol{w} - \alpha z$~~ $^{i}((()))$。由此可得

$$
\begin{aligned}
&(f''^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z).(f^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) \\
&= \sum_{ij} \boldsymbol{B}_{ij} f_i''((\alpha+1)\boldsymbol{w} - \alpha z)_j.(f^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) \\
&= \sum_{ij} \boldsymbol{B}_{ij} f_i \frac{(\boldsymbol{B}((\alpha+1)\boldsymbol{w}) - \alpha z)_i}{f^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w}) - \alpha z)}((\alpha+1)\boldsymbol{w} - \alpha z)_j.(f^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) \\
&= \sum_{ij} \boldsymbol{B}_{ij} f_i (\boldsymbol{B}((\alpha+1)\boldsymbol{w}) - \alpha z)_i ((\alpha+1)\boldsymbol{w} - \alpha z)_j \\
&= \sum_{i} f_i (\boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z)_i)^2 \\
&= (f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z))^2 + \sum_i f_i (f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) - (\boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z))_i)^2.
\end{aligned}
$$

因此，我们有：

$$
\begin{aligned}
&(f''^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z).(f^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) \\
&= (f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z))^2 + \sum_i f_i (f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) - (\boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z))_i)^2.
\end{aligned} \tag{12}
$$

由于我们的假设是 $\|\boldsymbol{w} - z\|_1 = O(\eta)$ 和 $\eta\alpha = O(1)$，因此我们有：

$$
\|\boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha z)\| = \|\alpha \boldsymbol{A}(\boldsymbol{w} - z) + \boldsymbol{A}\boldsymbol{w}\| = O(\alpha\eta) + O(1) = O(1).
$$

因此，我们还有：

$$
f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) = 1 \pm O(\eta).
$$

将方程（12）的两边除以 $f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z)$，可得：

$$
\begin{aligned}
&(f''^\top \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) \\
&= (f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z)) + (1 - O(\eta)) \sum_i f_i (f^T \boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z) - (\boldsymbol{B}((\alpha+1)\boldsymbol{w} - \alpha z))_i)^2 \\
&= \eta f^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha z) + (1 - O(\eta))\eta^2 \sum_i f_i ((f - e_i)^T \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha z))^2.
\end{aligned} \tag{13}
$$

因此，第二个等式得证。其他等式可直接由引理 6.2 得出。 □

此外，根据上述引理，若我们施加以下条件：

$$
\eta\alpha = \eta^b,
$$

其中 b 属于 $(0, 1]$。注意，该条件与 $\eta\alpha = O(1)$。并不矛盾。因此，根据上述引理，我们有：

$$
\begin{aligned}
&\eta(f' - f)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha z) \\
&= \eta(f' - f)^\top \boldsymbol{A}\boldsymbol{w} + \eta\alpha(f' - f)^\top \boldsymbol{A}(\boldsymbol{w} - z) \\
&= \eta(f' - f)^\top \boldsymbol{A}\boldsymbol{w} + \eta^b O(\eta^2) \\
&\implies \eta(f' - f)^\top \boldsymbol{A}\boldsymbol{w} = (1 - O(\eta))\eta^2 \sum_i f_i ((f - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha z))^2 - \eta^b O(\eta^2) \\
&= (1 - O(\eta))\eta^2 \sum_i f_i' ((f' - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{w} - \alpha z))^2 - \eta^b O(\eta^2).
\end{aligned}
$$

类似地，对于最小化玩家，我们有以下引理：

引理 6.4。 设 $\boldsymbol{y} \in \Delta_m, \boldsymbol{w}, \boldsymbol{z} \in \Delta_n$，并假设 $\boldsymbol{y}'$ 是 $AMWU$ 的下一次迭代，其当前向量为$\boldsymbol{y}$，输入为 $\boldsymbol{w}, \boldsymbol{z}$。此外，假设 $\|\boldsymbol{w} - \boldsymbol{z}\|_1 = O(\eta)$，且对于 $0 \le b \le 1$，有 $\eta\alpha = \eta^b$。则（当 η 充分小时）：

$$\begin{aligned}&\eta(\boldsymbol{y}' - \boldsymbol{y})^\top \boldsymbol{A}^\top(-\boldsymbol{w})\\&= (1 - O(\eta))\eta^2 \sum_i \boldsymbol{y}'_i((\boldsymbol{y}' - e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{z} - (\alpha+1)\boldsymbol{w}))^2 - \eta^b O(\eta^2).\end{aligned}$$

接下来，我们可以证明以下引理：

引理 6.5。 设 $(\boldsymbol{f}_t, \boldsymbol{y}_t)$ 为 $AMWU$ 动力系统的第 t 次迭代。对于每个时间步 $t \ge 2$，都有

$$\begin{aligned}&\eta \boldsymbol{f}_{t-1}^\top \boldsymbol{A}\boldsymbol{y}_t - \eta \boldsymbol{f}_t^\top \boldsymbol{A}\boldsymbol{y}_{t-1} = -\eta^2 \sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2\\&- \eta^2 \sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t - e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 + O(\eta^{2+b}).\end{aligned}$$

证明。

$$\begin{aligned}&\eta \boldsymbol{f}_{t-1}^\top \boldsymbol{A}\boldsymbol{y}_t - \eta \boldsymbol{f}_t^\top \boldsymbol{A}\boldsymbol{y}_{t-1}\\&\le -(1 - O(\eta))\eta^2 \sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_{t-1} - \alpha\boldsymbol{y}_{t-2}))^2 +\\&- (1 - O(\eta))\eta^2 \sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t - e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-2} - (\alpha+1)\boldsymbol{y}_{t-1}))^2 + \eta^b O(\eta^2)\\&= -(1 - O(\eta))\eta^2 \sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2 - (1 - O(\eta))\eta^2\eta^{2b} +\\&- (1 - O(\eta))\eta^2 \sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t - e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 - (1 - O(\eta))\eta^2\eta^{2b} + \eta^b O(\eta^2)\\&= -(1 - O(\eta))\eta^2 \sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t - e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2 +\\&- (1 - O(\eta))\eta^2 \sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t - e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 + \eta^b O(\eta^2)\end{aligned} \tag{14}$$

□

根据引理 6.5，我们可以推导出主定理：

定理 6.6。 设$(\boldsymbol{f}^*, \boldsymbol{y}^*)$为唯一最优极小极大均衡，且 η 足够小。假设 $\alpha = \eta^{b-1}$，其中 $b \in (0, 1]$。那么 $RE((\boldsymbol{f}^*, \boldsymbol{y}^*)\|(\boldsymbol{f}_t, \boldsymbol{y}_t))$随时间 t 递减，递减幅度为 η^{2+b}，除非$(\boldsymbol{f}_t, \boldsymbol{y}_t)$与$(\eta^{b/3})$相差不超过 O。

证明概要。根据相对熵的定义和以下不等式

$$\boldsymbol{f}^{*\top} \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}) \ge \boldsymbol{f}^{*\top} \boldsymbol{A}\boldsymbol{y}^*,$$

我们可以推导出以下关系

$$\begin{aligned}&RE((\boldsymbol{f}^*, \boldsymbol{y}^*)\|(\boldsymbol{f}_{t+1}, \boldsymbol{y}_{t+1})) - RE((\boldsymbol{f}^*, \boldsymbol{y}^*)\|(\boldsymbol{f}_t, \boldsymbol{y}_t))\\&\le \log\left(\sum_i \boldsymbol{f}_t(i) e^{\eta((e_i - \boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))}\right) + \log\left(\sum_i \boldsymbol{y}_t(i) e^{\eta((-(\alpha+1)\boldsymbol{f}_t + \alpha\boldsymbol{f}_{t-1})^\top \boldsymbol{A}(e_i - \boldsymbol{y}_t))}\right)\\&+ \eta^b(\boldsymbol{f}_{t-1}^\top \boldsymbol{A}\boldsymbol{y}_t - \boldsymbol{f}_t^\top \boldsymbol{A}\boldsymbol{y}_{t-1}).\end{aligned}$$

结合引理 6.5 与 $\eta^{b/3}$-接近的性质即可得到该结果（完整证明见附录 B.10）。□

备注。 根据上述定理，如果起始点是均匀分布（即 $\boldsymbol{f}_1 = (1/n, \ldots, 1/n)$ 和$\boldsymbol{y}_1 = (1/m, \ldots, 1/m)$），*AMWU* 至多经过： $O\left(\log\binom{nm}{\eta^{2+b}}\right)$迭代步数后达到 $O(\eta^{b/3})$~~——close~~。

6.2 $\eta^{b/3}$-接近性意味着接近最优解

我们首先需要以下引理:

引理 6.7。 设 $i \in \text{Supp}(\boldsymbol{f}^*)$ 和 $j \in \text{Supp}(\boldsymbol{y}^*)$。只要满足以下条件，就有 $x_T(i) \geq \frac{1}{2}\eta^{b/3}$ 且 $y_T(i) \geq \frac{1}{2}\eta^{b/3}$

$$\eta^{b/3} \leq \min_{s \in \text{Supp}(\boldsymbol{f}^*)} \frac{1}{(nm)^{1/\boldsymbol{f}^*(s)}}, \min_{s \in \text{Supp}(\boldsymbol{y}^*)} \frac{1}{(nm)^{1/\boldsymbol{y}^*(s)}}.$$

证明。根据 T 的定义，对于 $2 \leq t \leq T-1$，K-L 散度递减，因此

$$RE((\boldsymbol{f}^*, \boldsymbol{y}^*)||(\boldsymbol{f}_{T-1}, \boldsymbol{y}_{T-1})) < RE((\boldsymbol{f}^*, \boldsymbol{y}^*)||(\boldsymbol{f}_1, \boldsymbol{y}_1)).$$

这意味着:

$$\begin{aligned}
&\boldsymbol{f}^*(i) \log\left(\frac{1}{\boldsymbol{f}_{T-1}}(i)\right) \leq \sum_j \boldsymbol{f}^*(j) \log\left(\frac{1}{\boldsymbol{f}_{T-1}}(j)\right) \\
&\leq \sum_i \boldsymbol{f}^*(i) \log\left(\frac{1}{\boldsymbol{f}_1(i)}\right) + \sum_i \boldsymbol{y}^*(i) \log\left(\frac{1}{\boldsymbol{y}_1(i)}\right) = \log(nm) \\
&\implies \boldsymbol{f}_T(i) > \frac{1}{(mn)^{1/\boldsymbol{f}^*(i)}} \geq \eta^{b/3}.
\end{aligned}$$

由于 $|\boldsymbol{f}_T(i) - \boldsymbol{f}_{T-1}(i)|$是 $O(\eta)$，结论得证。 □

利用上述引理，我们可以采用定理 3.2 中 Daskalakis 和 Panageas [018a] 的相同论证，证明以下定理:

定理 6.8（$\eta^{b/3}$-接近意味着接近最优解）。假设（$\boldsymbol{f}^*, \boldsymbol{y}^*$）是该问题的唯一最优解。令 T 为 *KL* 散度首次不再以 $\Omega(\eta^{b+2})$ 递减的时间。由此可得，当 $\eta \to 0$ 时， $\eta^{b/3}$-接近点（$\boldsymbol{f}_T, \boldsymbol{y}_T$）与（$\boldsymbol{f}^*, \boldsymbol{y}^*$）之间的距离趋于零:

$$\lim_{\eta \to 0} \|(\boldsymbol{f}^*, \boldsymbol{y}^*) - (\boldsymbol{f}_T, \boldsymbol{y}_T)\|_1 = 0.$$

为使论文内容完整，我们在附录 B.12 中给出完整证明。

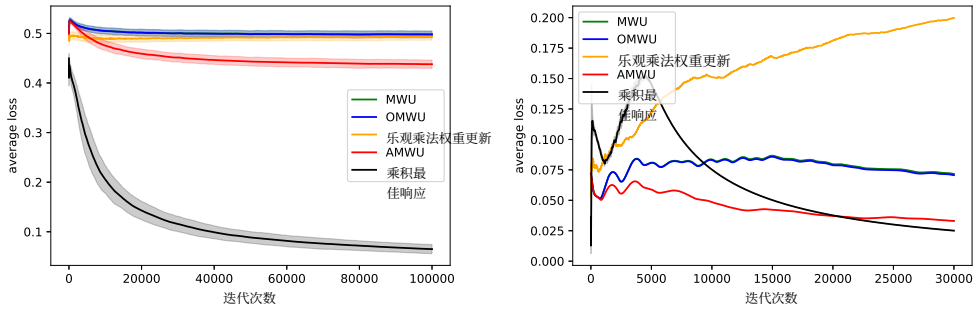
6.3 局部收敛的证明

我们使用 动力系统中的以下著名结论来证明局部收敛:

命题 6.9（参见 [Galor, 2007]）。如果连续可微更新规则w 在不动点 $\boldsymbol{z}$ 处的雅可比矩阵的谱半径小于 *1*，那么存在一个邻域 U，围绕 $\boldsymbol{z}$，使得对于所有 $\boldsymbol{x} \in U$，动态都会收敛到 $\boldsymbol{z}$。

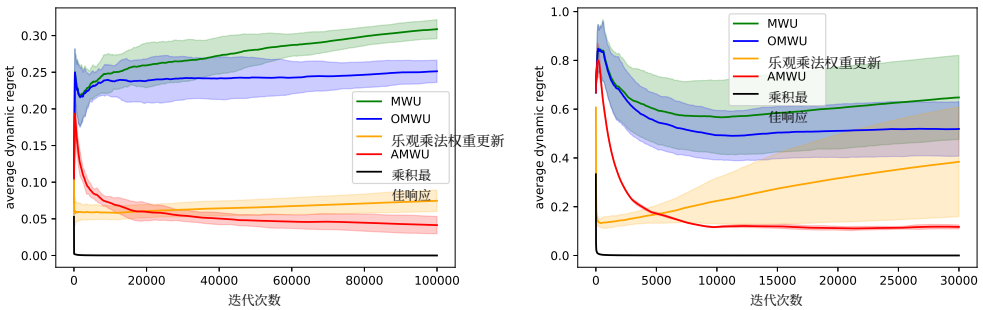
据此，我们的局部收敛定理表述如下:

定理 6.10。 设$(\boldsymbol{f}^*, \boldsymbol{y}^*)$是博弈 $\boldsymbol{A}$ 的唯一极小极大均衡。存在$(\boldsymbol{f}^*, \boldsymbol{y}^*)$的一个邻域，使得 *AMWU* 动力学收敛。



(a) 随机博弈中学习率为 0.5 的乘法权重更新对手 (b) 四子棋中学习率为 0.5 的乘法权重更新对手

图 1. 对无记忆乘法权重更新对手的平均损失



(a) 随机博弈中的非无知的 MWU 对手 (b) 四子棋中的非无知的 MWU 对手

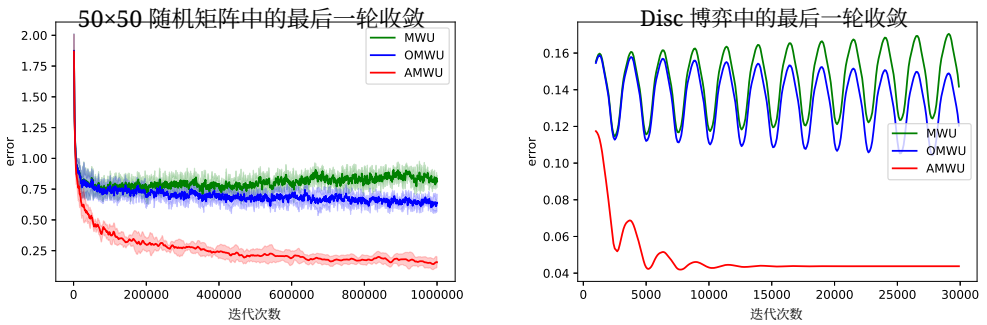
图 2. 对非无知的 MWU 对手的平均损失

证明概要。我们首先根据公式 (11) 构造 AMWU 更新的动力系统，其中变量是两个连续策略（例如，参见公式 (24)）。容易证明，$(\boldsymbol{f}^*, \boldsymbol{y}^*, \boldsymbol{f}^*, \boldsymbol{y}^*)$ 是该动力系统的一个不动点。接着，根据命题 B.14，要证明局部收敛性质，我们只需证明在 $(\boldsymbol{f}^*, \boldsymbol{y}^*, \boldsymbol{f}^*, \boldsymbol{y}^*)$ 处计算得到的该动力系统的雅可比矩阵的谱半径小于 1，即在 $(\boldsymbol{f}^*, \boldsymbol{y}^*, \boldsymbol{f}^*, \boldsymbol{y}^*)$ 处计算得到的雅可比矩阵的每个特征值都小于 1。完整证明见附录 B.15。□

7 实验

本节中，我们在多种设定下测试算法 AMWU 和 Prod-BR 的表现。首先，我们考虑一个无知的无外部遗憾对手，并在随机博弈和元博弈中，衡量我们的算法相较于基线方法（即 MWU、OMWU）的平均损失表现 [Czarnecki 等人， 2020]。其次，我们测试 AMWU 和 Prod-BR 对抗一个非无知的无外部遗憾对手时的表现，并衡量平均动态遗憾表现。最后，我们测试 AMWU 在自博弈设定中的表现，并衡量其纳什均衡的最后一轮收敛率。

对无知对手的表现：为了公平比较 AMWU、Prod-BR 与基线方法的平均损失表现，我们考虑无知乘法权重更新对手：智能体的历史策略不会影响乘法权重更新对手的策略。为了构造这种



（a） 50×50 随机博弈 (b) 圆盘元博弈

图 3. 最后一轮收敛

非无知对手，我们假设对手采用乘法权重更新，与另一个对手而非该智能体进行博弈，因此智能体的策略不会影响对手的行为。[8]。如图 1 所示，AMWU 和 Prod-BR 的表现远胜于其他基线方法。具体而言，与 AMWU 和其他基线方法相比，Prod-BR 达到了最低的平均损失。直观来说，由于智能体与无知对手进行博弈，AMWU 和 Prod-BR 更好的理论遗憾保证意味着更好的平均损失表现，正如我们在本实验中所展示的。因此，遗憾界指标最佳（即动态遗憾）的 Prod-BR 表现最佳，其次是具有前向遗憾保证的 AMWU。有趣的是，在我们的设定中，乘法权重更新与采用相同学习率的乐观乘法权重更新表现几乎完全相同，这印证了乐观乘法权重更新并未充分利用额外信息这一观点。

对抗非无知对手的表现： 我们现在测试算法对抗非无知对手（即智能体的行为可能改变对手的策略），并回答以下问题：AMWU 和 Prod-BR 更优的理论遗憾界，能否在实践中带来更好的对抗无外部遗憾对手的遗憾表现？如图 2 所示，与基线方法相比，AMWU 和 Prod-BR 的平均动态遗憾要小得多。这进一步印证了我们的理论结果：与基线方法相比，AMWU 和 Prod-BR 对抗无外部遗憾对手时具有更优的遗憾界保证，因此在实践中也能取得更好的遗憾界。

最后一轮收敛：我们比较 AMWU 与 OWMU 和 MWU 的收敛速度。为公平比较，三个算法均使用相同的学习率 $\mu = 0.01$ [9]。如图 3 所示，在收敛至纳什均衡方面，AMWU 大幅优于 OMWU 和 MWU。有趣的是，在四子棋和 Disc 元博弈中，AMWU 呈现出明显的收敛趋势，而在相同设置下，OWMU 和 MWU 则出现波动（图 3b）。

AMWU与OMWU的比较：为了突出 AMWU 与 OMWU 之间的差异，我们使用与 AMWU 相同的可预测序列 $\boldsymbol{x}_{t-1}$ 与正则化项R之间的相对权重（即 $\eta_{OMWU} = \eta_{AMWU} \times \alpha_{AMWU}$）来测试 OWMU_1。正如我们在图 1 中清楚看到的那样，在我们考察的每个博弈中，AMWU 都优于 OWMU_1。我们可以确认，AMWU 和 OMWU 是两种截然不同的算法，因为它们利用额外知识的程度不同。

[8]具体设置见附录C。[9] [1] 其他取值的结果也呈现出类似的整体趋势。更多详情见附录C.2。

8 结论

我们研究学习者掌握有关对手行为的额外知识（即无外部遗憾对手）时的在线学习问题。在这种设置下，我们的算法 AFTRL 和 Prod-BR 能够充分利用这些额外知识，分别实现 $O(1)$ 的前向遗憾和 $O(\sqrt{T})$ 的动态遗憾。此外，即使额外知识不准确，AFTRL 和 Prod-BR 在最坏情况下仍保持无遗憾性质。最后，我们证明，作为 AFTRL 特例的 AMWU 能够在具有唯一纳什均衡的双人零和博弈中实现最后一轮收敛。

参考文献

Jacob D Abernethy、Elad Hazan 和 Alexander Rakhlin。2009年。《在黑暗中竞争：一种高效的老虎机线性优化算法》。载于学习理论会议。J. P. Bailey 和 G. Piliouras。2018年。零和博弈中的乘法权重更新。*ACM EC*（2018），321–338。Amir Beck 和 Marc Teboulle。2003。用于凸优化的镜像下降与非线性投影次梯度法。《运筹学研究快报》31，3（2003），167–175。Omar Besbes、Yonatan Gur 和 Assaf Zeevi。2015。非平稳随机优化。《运筹学》63，5（2015），1227–1244。Michael Bowling、Neil Burch、Michael Johanson 和 Oskari Tammelin。2015。单挑限注德州扑克已被攻克。《科学》347，6218（2015），145–149。Nicolo Cesa-Bianchi 和 Gábor Lugosi。2006。《预测、学习与博弈》。剑桥大学出版社。Chao-Kai Chiang、Tianbao Yang、Chia-Jung Lee、Mehrdad Mahdavi、Chi-Jen Lu、Rong Jin 和 Shenghuo Zhu。2012。具有渐进变化的在线优化。载于学习理论会议。6.1–6.20。Wojciech Marian Czarnecki、Gauthier Gidel、Brendan Tracey、Karl Tuyls、Shayegan Omidshafiei、David Balduzzi 和 Max Jaderberg。2020年。现实世界的博弈看起来像陀螺。*arXiv* 预印本 *arXiv:2004.09468*（2020年）。Constantinos Daskalakis、Andrew Ilyas、Vasilis Syrgkanis 和 Haoyang Zeng。2017年。以乐观策略训练生成对抗网络。*arXiv* 预印本 *arXiv:1711.00141*（2017年）。C. Daskalakis 和 I. Panageas。2018a。最后迭代收敛：零和博弈与约束极小极大优化。*arXiv* 预印本 *arXiv:1807.04252*（2018a）。Steven De Rooij、Tim Van Erven、Peter D Grünwald 和 Wouter M Koolen。2014。能跟随领先者时就跟随，否则就采用对冲。《机器学习研究期刊》15，1（2014），1281–1316。Yuan Deng、Jon Schneider 和 Balasubramanian Sivan。2019。针对无遗憾学习者制定策略。《神经信息处理系统进展》32（2019）。Le Cong Dinh、Tri-Dung Nguyen、Alain B Zemhoho 和 Long Tran-Thanh。2021a。非对称重复博弈中的最后一轮收敛与无动态遗憾。载于算法学习理论。PMLR，553–577。Le Cong Dinh、Yaodong Yang、Zheng Tian、Nicolas Perez Nieves、Oliver Slumbers、David Henry Mguni、Haitham Bou Ammar 和 Jun Wang。2021b。在线双重预言机。*arXiv* 预印本 *arXiv:2103.07780*（2021年）。Yoav Freund 和 Robert E Schapire。1999。运用乘法权重进行自适应博弈。《博弈与经济行为》29，1-2（1999），79–103。Oded Galor。2007年。《离散动力系统》。Springer Science & Business Media。Solomon Kullback 和 Richard A Leibler。1951。论信息与充分性。《数学统计年刊》22，1（1951），79–86。Marc Lanctot、Vinicius Zambaldi、Audrunas Gruslys、Angeliki Lazaridou、Karl Tuyls、Julien Pérolat、David Silver 和 Thore Graepel。2017年。多智能体强化学习的统一博弈论方法。*arXiv* 预印本 *arXiv:1711.00832*（2017年）。Ruta Mehta、Ioannis Panageas、Georgios Piliouras、Prasad Tetali 和 Vijay V Vazirani。2017年。动态环境中的突变、有性繁殖与生存。载于第*8*届理论计算机科学创新会议（*ITCS 2017*）。Schloss Dagstuhl-Leibniz-Zentrum fur Informatik GmbH，Dagstuhl 出版社。P. Mertikopoulos、C. Papadimitriou 和 G. Piliouras。2018。对抗性正则化学习中的周期。*SODA*（2018）。Dragoslav S Mitrinovic 和 Petar M Vasic。1970年。《解析不等式》。第61卷。Springer。Arkadij Semenovič Nemirovskij 和 David Borisovich Yudin。1983年。《优化中的问题复杂度与方法效率》。约翰·威立父子出版公司。J v Neumann。1928。论博弈论。*Math. Ann.* 100，1（1928），295–320。Alexander Rakhlin 和 Karthik Sridharan。2013。基于可预测序列的在线学习。载于学习理论会议。PMLR，993–1019。Ankan Saha、Prateek Jain 和 Ambuj Tewari。2012。在线学习中稳定性与遗憾的相互作用。*arXiv* 预印本 *arXiv:1211.6158*（2012年）。

Amir Sani、Gergely Neu 和 Alessandro Lazaric。2014。在线优化中利用易数据。《神经信息处理系统进展》27（2014）。Shai Shalev-Shwartz 等人。2012。在线学习与在线凸优化。*Foundations and Trends® in Machine Learning* 4，2（2012），107–194。Eric Van Damme。1991年。《纳什均衡的稳定性与完美性》。第339卷。Springer。Chen-Yu Wei、Chung-Wei Lee、Mengxiao Zhang 和 Haipeng Luo。2020年。约束鞍点优化中的线性最后迭代收敛。*arXiv* 预印本 *arXiv:2006.09517*（2020年）。Martin Zinkevich、Michael Johanson、Michael Bowling 和 Carmelo Piccione。2007年。不完全信息博弈中的遗憾最小化。《神经信息处理系统进展》20（2007），1729–1736。

A 缺失算法与定义

A.1 正则化跟随者

正则化跟随者 [Abernethy 等人提出的 2009], 一种广为人知的无外部遗憾算法，在我们主算法的分析中发挥着重要作用:

算法 4: 正则化跟随者

输入: 学习率 $\eta > 0, f_1 = \arg\min_{f \in \mathcal{F}} R(f)$。

输出: 下一次策略更新

$$f_{t+1} = \arg\min_{f \in \mathcal{F}} F_{t+1}(f) = \langle f, \sum_{s=1}^{t} x_s \rangle + \frac{R(f)}{\eta}$$

A.2 精确镜像下降

现在，我们将精确预测技术应用于另一种著名的无悔算法——镜像下降 [Nemirovskij 和 Yudin， 1983]。设 R 是关于范数 $\|.\|_p$ 的 β-强凸函数，令 $D_{\mathcal{R}(.,.)}$表示关于 R 的 Breg man 散度。设 $\|.\|_q$与 $\|.\|_p$互为对偶。那么，精确镜像下降（AMD）算法可描述如下:

算法 5: 精确镜像下降

输入: 学习率 $\eta > 0$， 利用率 $\alpha \geq 1$,
$f_1 = g_1 = \arg\min_{f \in \mathcal{F}} \mathrm{R}(f)$。

输出: 下一次策略更新

$$g_{t+1} = \arg\min_{g \in \mathcal{F}} G_{t+1}(g) = \eta \langle g, x_t \rangle + D_{\mathcal{R}}(g, g_t)$$

$$f_{t+1} = \arg\min_{f \in \mathcal{F}} F_{t+1}(f) = \eta \langle f, \alpha M_{t+1} \rangle + D_{\mathcal{R}}(f, g_{t+1})$$

其中，正则化项 $R(f)$ 是相对于 l_p 范数的 β-强凸函数， $p \geq 1$。

A.3 缺失的定义

定义 *A.1*（β-接近性 *[Mehta* 等人， *2017]*）。假设 $\beta > 0$。点 $(f, y) \in \Delta_n \times \Delta_m$ 是 β-接近的，当且仅当对每个 $i \in [n]$，都有 $f_i \leq \beta$ 或 $|f^\top A y - (Ay)_i| \leq \beta$；并且对每个 $j \in [m]$，都有 $y_i \leq \beta$ 或 $|f^\top A y - (A^\top x)_i| \leq \beta$。

最后，为了分析矩阵博弈 A 中的动态，我们将使用 Kullback-Leibler 散度来理解博弈参与者策略的行为。

定义 *A.2*（*[Kullback* 和 *Leibler*， *1951]*）。向量 x_1和 x_2在 Δ_n中的相对熵或 Kull-Leibler（K-L）散度定义为 $RE(x_1 \| x_2) = \sum_{i=1}^{n} x_1(i)$ 对数 $\left(\frac{x_1(i)}{x_2(i)}\right)$。

这始终非负。我们还可以证明， $RE(x_1 \| x_2) = 0$ 当且仅当 $x_1 = x_2$ 几乎处处成立 [Mitrinovic 和 Vasic， 1970]。

B 缺失的证明

引理 B.1（引理 3.4）。设 $\boldsymbol{g}_t$ 定义如上，则对于任意$\boldsymbol{f} \in \mathcal{F}$，以下关系成立：

$$\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle \leq \langle \boldsymbol{f}, \sum_{t=1}^{T}\boldsymbol{x}_t\rangle + \frac{R(\boldsymbol{f})}{\eta}.$$

引理 3.4 的证明。我们通过数学归纳法证明。对于 $t = 1$:

$$\langle \boldsymbol{g}_1, \boldsymbol{x}_1\rangle \leq \langle \boldsymbol{g}_1, \boldsymbol{x}_1\rangle + \frac{R(\boldsymbol{g}_1)}{\eta} \leq \langle \boldsymbol{f}, \boldsymbol{x}_1\rangle + \frac{R(\boldsymbol{f})}{\eta} \ \forall \boldsymbol{f} \in \mathcal{F}.$$

假设该命题对 T 成立，其中

$$\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle \leq \langle \boldsymbol{f}, \sum_{t=1}^{T}\boldsymbol{x}_t\rangle + \frac{R(\boldsymbol{f})}{\eta} \ \forall \boldsymbol{f} \in \mathcal{F}.$$

在两边加上 $\boldsymbol{g}_{T+1}$， $\boldsymbol{x}_{T+1}$，可得

$$\begin{aligned}\sum_{t=1}^{T+1}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle &\leq \langle \boldsymbol{f}, \sum_{t=1}^{T}\boldsymbol{x}_t\rangle + \frac{R(\boldsymbol{f})}{\eta} + \langle \boldsymbol{g}_{T+1}, \boldsymbol{x}_{T+1}\rangle \ \forall \boldsymbol{f} \in \mathcal{F} \\ &\leq \langle \boldsymbol{g}_{T+1}, \sum_{t=1}^{T}\boldsymbol{x}_t\rangle + \frac{R(\boldsymbol{g}_{T+1})}{\eta} + \langle \boldsymbol{g}_{T+1}, \boldsymbol{x}_{T+1}\rangle \\ &\leq \langle \boldsymbol{f}, \sum_{t=1}^{T+1}\boldsymbol{x}_t\rangle + \frac{R(\boldsymbol{f})}{\eta} \ \forall \boldsymbol{f} \in \mathcal{F}.\end{aligned}$$

因此，该命题对 $T + 1$成立。

根据上述不等式，如果一个算法是无前向遗憾的，即:

$$\sum_{t=1}^{\top}(\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle) = o(T),$$

那么我们还有:

$$\begin{aligned}&\min_{\boldsymbol{f}\in\mathcal{F}}\sum_{t=1}^{\top}(\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \langle \boldsymbol{f}, \boldsymbol{x}_t\rangle) \leq \sum_{t=1}^{\top}(\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle) + \frac{R(\boldsymbol{f})}{\eta} \\ &= o(T) + \frac{R(\boldsymbol{f})}{\eta} = o(T).\end{aligned}$$

因此，该算法也是一种无外部遗憾算法。 □

引理 B.2（倍增技巧）。倍增技巧的思路是将时间区间划分为不同的阶段，并在每个阶段重新启动算法（即 *AFTRL*）。我们将证明，在过程中考虑不同阶段后，*AFTRL* 仍能保持遗憾界 $O\left(\sqrt{\sum_{t=1}^{T}(\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q)^2}\right)$。

利用引理 *3.4*，可推导出公式 *2* 中的遗憾界:

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \langle \boldsymbol{f}^*, \sum_{t=1}^{T}\boldsymbol{x}_t\rangle \leq \frac{\alpha}{\eta\alpha}R(\boldsymbol{f}^*) + \frac{\eta\alpha}{\beta}\sum_{t=1}^{T}(\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q)^2 \ \forall \boldsymbol{f}^* \in \mathcal{F}.$$

现在，我们将时间区间 T 划分为不同的阶段，其中阶段 i 具有恒定的学习率 $\eta_i = \eta_0 2^{-i}$。定义阶段 $i+1$ 的起始点如下：

$$s_{i+1} = \min\{\tau : \frac{\eta_i \alpha}{\beta} \sum_{t=s_i}^{\tau} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_t^*)^2 > \frac{\alpha}{\eta_i \alpha} R(\boldsymbol{f}^*)\}.$$

以及 $s_1 = 1$。令 N 为博弈的最后一个阶段，并令 $s_{N+1} = T+1$。于是我们有：

$$\begin{aligned}\sum_{t=1}^{T} \langle \boldsymbol{f}_t, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}^*, \sum_{t=1}^{T} \boldsymbol{x}_t \rangle &\leq \sum_{i=1}^{N} \frac{\alpha}{\eta_i \alpha} R(\boldsymbol{f}^*) + \frac{\eta_i \alpha}{\beta} \sum_{t=s_i}^{s_{i+1}-1} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_t^*)^2 \\ &\leq 2 \sum_{i=1}^{N} \frac{\alpha}{\eta_i \alpha} R(\boldsymbol{f}^*) \leq \frac{2^{N+2}}{\eta_0} R(\boldsymbol{f}^*),\end{aligned}$$

其中，不等式来自 s_i 的定义。注意，我们有：

$$\begin{aligned}\frac{1}{\eta_0} = \frac{1}{\eta_{N-1} 2^{N-1}} &\leq \frac{1}{2^{N-1}} \sqrt{\sum_{t=s_{N-1}}^{s_N} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q)^2} \sqrt{\frac{\alpha}{\beta R(\boldsymbol{f}^*)}} \\ &\leq \frac{1}{2^{N-1}} \sqrt{\sum_{t=1}^{T} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q)^2} \sqrt{\frac{\alpha}{\beta R(\boldsymbol{f}^*)}}.\end{aligned}$$

因此，我们有：

$$\begin{aligned}\sum_{t=1}^{T} \langle \boldsymbol{f}_t, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}^*, \sum_{t=1}^{T} \boldsymbol{x}_t \rangle &\leq \frac{2^{N+2}}{\eta_0} R(\boldsymbol{f}^*) \\ &\leq 2^{N+2} \frac{1}{2^{N-1}} \sqrt{\sum_{t=1}^{T} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q)^2} \sqrt{\frac{\alpha}{\beta R(\boldsymbol{f}^*)}} R(\boldsymbol{f}^*) \\ &= 8 \sqrt{\sum_{t=1}^{T} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q)^2} \sqrt{\frac{\alpha R(\boldsymbol{f}^*)}{\beta}} = O\left(\sqrt{\sum_{t=1}^{T} (\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_q} \right).\end{aligned}$$

因此，我们得出该结果。

引理 B.3（引理 4.1）。令 $\boldsymbol{f}_t$、$\boldsymbol{f}_{t+1}$ 为无悔算法（即跟随正则化领导者、在线镜像下降）的两个连续策略。则对于任意范数 $\|.\|_q$，我们有：

$$\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_q = O(\frac{1}{\sqrt{T}}).$$

为了证明引理 4.1，我们首先需要以下关于跟随正则化领导者和在线镜像下降的两个连续策略之间距离的引理：

引理 B.4。设 $\boldsymbol{f}_t, \boldsymbol{f}_{t+1}$ 为 *FTRL* 算法的两个连续策略。则有：

$$\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p \leq \eta \frac{2n}{\beta}, \ \textit{where } \|.\|_p \textit{ denotes } l_p \textit{ norm.}$$

证明。根据 β-强凸函数的性质，我们有：

$$\begin{aligned}
&F_t(\boldsymbol{f}_{t+1}) - F_t(\boldsymbol{f}_t) \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2 \\
&\iff \langle \boldsymbol{f}_{t+1}, \sum_{s=1}^{t} \boldsymbol{x}_s \rangle + \frac{R(\boldsymbol{f}_{t+1})}{\eta} - \langle \boldsymbol{f}_{t+1}, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}_t, \sum_{s=1}^{t-1} \boldsymbol{x}_s \rangle - \frac{R(\boldsymbol{f}_t)}{\eta} \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2 \\
&\iff F_{t+1}(\boldsymbol{f}_{t+1}) - \langle \boldsymbol{f}_{t+1}, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}_t, \sum_{s=1}^{t-1} \boldsymbol{x}_s \rangle - \frac{R(\boldsymbol{f}_t)}{\eta} \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2.
\end{aligned}$$

根据定义，我们有 $F_{t+1}(\boldsymbol{f}_{t+1}) \leq F_{t+1}(\boldsymbol{f}_t)$。因此，将其代入上述不等式可得：

$$\begin{aligned}
&F_{t+1}(\boldsymbol{f}_t) - \langle \boldsymbol{f}_{t+1}, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}_t, \sum_{s=1}^{t-1} \boldsymbol{x}_s \rangle - \frac{R(\boldsymbol{f}_t)}{\eta} \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2 \\
&\iff \langle \boldsymbol{f}_t, \sum_{s=1}^{t} \boldsymbol{x}_s \rangle + \frac{R(\boldsymbol{f}_t)}{\eta} - \langle \boldsymbol{f}_{t+1}, \boldsymbol{x}_t \rangle - \langle \boldsymbol{f}_t, \sum_{s=1}^{t-1} \boldsymbol{x}_s \rangle - \frac{R(\boldsymbol{f}_t)}{\eta} \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2 \\
&\iff \langle \boldsymbol{f}_t - \boldsymbol{f}_{t+1}, \boldsymbol{x}_t \rangle \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2 \\
&\implies \|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p \|\boldsymbol{x}_t\|_q \geq \frac{\beta}{2\eta}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p^2 \\
&\implies \frac{2\eta n}{\beta} \geq \|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p,
\end{aligned}$$

由于 $\boldsymbol{x}_t \in [0,1]^n$，因此 $\|\boldsymbol{x}_t\|_q \leq n^{1/q} = n^{1-1/p} \leq n$。由此可得结果。

□

其他无悔算法也具有类似性质，例如在线镜像下降：引理 B.5。设 $\boldsymbol{g}_t$，$\boldsymbol{g}_{t+1}$ 是 *OMD*算法的两个连续策略。则有：

$$\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p \leq \frac{\eta}{\beta}$$

证明。根据 β-强凸函数的性质，我们有：

$$\begin{aligned}
&G_{t+1}(\boldsymbol{g}_t) - G_{t+1}(\boldsymbol{g}_{t+1}) \geq \frac{\beta}{2}\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p^2 \\
&\iff \eta\langle \boldsymbol{g}_t - \boldsymbol{g}_{t+1}, \boldsymbol{x}_t \rangle + D_{\mathcal{R}}(\boldsymbol{g}_t, \boldsymbol{g}_t) - D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{g}_t) \geq \frac{\beta}{2}\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p^2 \\
&\iff \eta\langle \boldsymbol{g}_t - \boldsymbol{g}_{t+1}, \boldsymbol{x}_t \rangle \geq D_{\mathcal{R}}(\boldsymbol{g}_{t+1}, \boldsymbol{g}_t) + \frac{\beta}{2}\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p^2 \\
&\implies \eta\langle \boldsymbol{g}_t - \boldsymbol{g}_{t+1}, \boldsymbol{x}_t \rangle \geq \frac{\beta}{2}\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p^2 + \frac{\beta}{2}\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p^2 \\
&\implies \eta\|\boldsymbol{g}_t - \boldsymbol{g}_{t+1}\|_p \|\boldsymbol{x}_t\|_q \geq \beta\|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p^2 \\
&\implies \frac{\eta}{\beta} n \geq \|\boldsymbol{g}_{t+1} - \boldsymbol{g}_t\|_p,
\end{aligned}$$

由于 $D_{\mathcal{R}}(\boldsymbol{g}_t, \boldsymbol{g}_t) = 0$ 且 $\boldsymbol{x}_t \in [0,1]^n$。 □

现在我们可以证明引理 4.1：

引理 4.1 的证明。根据引理 B.4 和引理 B.5，以及无悔算法的性质，例如 $\eta = O(\frac{1}{\sqrt{T}})$，我们有：

$$\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p = O(\frac{1}{\sqrt{T}}).$$

现在，对于 $q > p$，很容易证明：

$$\begin{aligned} &\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_q \leq \|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p \\ \implies &\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_q = O(\frac{1}{\sqrt{T}}). \end{aligned}$$

对于 $q < p$，利用霍尔德不等式，我们得到：

$$\begin{aligned} &\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_q \leq n^{1/q-1/p}\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_p = n^{1/q-1/p}O(\frac{1}{\sqrt{T}}) \\ \implies &\|\boldsymbol{f}_{t+1} - \boldsymbol{f}_t\|_q = O(\frac{1}{\sqrt{T}}). \end{aligned}$$

证明完毕。 □

定理 B.6（定理 4.2）。设 $\mathcal{F} \subset \mathbb{R}^n$ 是一个凸紧集，且 R 是一个 β-强凸函数，并满足 $\min_{f\in\mathcal{F}} R(\boldsymbol{f}) = 0$. 对于任意环境策略，*AFTRL* 算法满足：

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \frac{1}{\alpha}\langle \boldsymbol{f}^*, \sum_{t=1}^{T}\boldsymbol{x}_t\rangle - \frac{\alpha-1}{\alpha}\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle \leq \frac{1}{\eta\alpha}R(\boldsymbol{f}^*) + \frac{\eta\alpha}{\beta}\sum_{t=1}^{T}(||\boldsymbol{x}_t - \boldsymbol{x}_{t-1}||_q)^2.$$

定理 4.2 的证明。我们首先将 $\boldsymbol{h}_{t+1}$ 定义如下

$$\boldsymbol{h}_{t+1} = \arg\min_{\boldsymbol{f}\in\mathcal{F}} H_{t+1}(\boldsymbol{f}) = \langle \boldsymbol{f}, \sum_{s=1}^{t}\boldsymbol{x}_s + \alpha\boldsymbol{x}_{t+1}\rangle + \frac{R(\boldsymbol{f})}{\eta}.$$

注意，对于任意 $\boldsymbol{f}_t \in \mathrm{F}$ 序列，

$$\begin{aligned} \sum_{t=1}^{T}\langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle &= \sum_{t=1}^{T}\langle \boldsymbol{f}_t - \boldsymbol{h}_t, \boldsymbol{x}_t - \boldsymbol{x}_{t-1}\rangle \\ &+ \sum_{t=1}^{T}\langle \boldsymbol{f}_t - \boldsymbol{h}_t, \boldsymbol{x}_{t-1}\rangle + \sum_{t=1}^{T}\langle \boldsymbol{h}_t, \boldsymbol{x}_t\rangle. \end{aligned}$$

现在我们用数学归纳法证明

$$\begin{aligned} &\sum_{t=1}^{T}\langle \boldsymbol{f}_t - \boldsymbol{h}_t, \boldsymbol{x}_{t-1}\rangle + \sum_{t=1}^{T}\langle \boldsymbol{h}_t, \boldsymbol{x}_t\rangle \leq \\ &\frac{1}{\alpha}\langle \boldsymbol{f}', \sum_{t=1}^{T}\boldsymbol{x}_t\rangle + \frac{\alpha-1}{\alpha}\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle + \frac{1}{\eta\alpha}R(\boldsymbol{f}'),\ \forall \boldsymbol{f}' \in \mathcal{F}. \end{aligned} \tag{15}$$

对于 $t = 1$，$M_1 = 0$，我们有

$$\langle \boldsymbol{h}_1, \boldsymbol{x}_1\rangle + \frac{R(\boldsymbol{h}_1)}{\eta\alpha} \leq \langle \boldsymbol{f}', \boldsymbol{x}_1\rangle + \frac{R(\boldsymbol{f}')}{\eta\alpha}, \implies \frac{1}{\alpha}\langle \boldsymbol{h}_1, \boldsymbol{x}_1\rangle \leq \frac{1}{\alpha}\langle \boldsymbol{f}', \boldsymbol{x}_1\rangle + \frac{R(\boldsymbol{f}')}{\eta\alpha},$$

因为 $\alpha \geq 1$和 $R(\boldsymbol{f}) \geq 0 \ \forall \boldsymbol{f} \in \mathrm{F}$。我们还有

$$\begin{aligned}
&\langle \boldsymbol{h}_1, \alpha \boldsymbol{x}_t \rangle + \frac{R(\boldsymbol{h}_1)}{\eta} \leq \langle \boldsymbol{g}_1, \alpha \boldsymbol{x}_t \rangle + \frac{R(\boldsymbol{g}_1)}{\eta} \\
&= \langle \boldsymbol{g}_1, \boldsymbol{x}_t \rangle + \frac{R(\boldsymbol{g}_1)}{\eta} + (\alpha - 1)\langle \boldsymbol{g}_1, \boldsymbol{x}_t \rangle \\
&\leq \langle \boldsymbol{h}_1, \boldsymbol{x}_t \rangle + \frac{R(\boldsymbol{h}_1)}{\eta} + (\alpha - 1)\langle \boldsymbol{g}_1, \boldsymbol{x}_t \rangle \\
&\implies \langle \boldsymbol{h}_1, \boldsymbol{x}_1 \rangle \leq \langle \boldsymbol{g}_1, \boldsymbol{x}_1 \rangle.
\end{aligned}$$

于是，我们有：

$$\langle \boldsymbol{h}_1, \boldsymbol{x}_1 \rangle \leq \frac{1}{\alpha}\langle \boldsymbol{f}', \boldsymbol{x}_1 \rangle + \frac{R(\boldsymbol{f}')}{\eta\alpha} + \frac{\alpha - 1}{\alpha}\langle \boldsymbol{g}_1, \boldsymbol{x}_t \rangle.$$

因此，关于 $t = 1$ 的数学归纳法第一步是正确的。

为进行数学归纳法，假设上述不等式对 $\tau = T - 1$ 成立。利用 $\boldsymbol{f}_{\prime} = \boldsymbol{f}_T$，并在两边加上 $\boldsymbol{f}_T - \boldsymbol{h}_T, \boldsymbol{x}_{t-1}$ $+\boldsymbol{h}_T, \boldsymbol{x}_{t-1}$，我们得到

$$\begin{aligned}
&\sum_{t=1}^{T}\langle \boldsymbol{f}_t - \boldsymbol{h}_t, \boldsymbol{x}_{t-1} \rangle + \sum_{t=1}^{T}\langle \boldsymbol{h}_t, \boldsymbol{x}_t \rangle \\
&\leq \frac{1}{\alpha}\langle \boldsymbol{f}_T, \sum_{t=1}^{T-1}\boldsymbol{x}_t \rangle + \frac{\alpha - 1}{\alpha}\sum_{t=1}^{T-1}\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle + \frac{1}{\eta\alpha}R(\boldsymbol{f}_T) + \langle \boldsymbol{f}_T - h_T, \boldsymbol{x}_{t-1} \rangle + \langle h_T, \boldsymbol{x}_t \rangle \\
&= \frac{1}{\alpha}(\langle \boldsymbol{f}_T, \sum_{t=1}^{T-1}\boldsymbol{x}_t + \alpha \boldsymbol{x}_{t-1} \rangle + \frac{1}{\eta}R(\boldsymbol{f}_T)) + \frac{\alpha - 1}{\alpha}\sum_{t=1}^{T-1}\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle + \langle h_T, \boldsymbol{x}_t - \boldsymbol{x}_{t-1} \rangle \\
&\leq \frac{1}{\alpha}(\langle h_T, \sum_{t=1}^{T-1}\boldsymbol{x}_t + \alpha \boldsymbol{x}_{t-1} \rangle + \frac{1}{\eta}R(h_T)) + \frac{\alpha - 1}{\alpha}\sum_{t=1}^{T-1}\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle + \langle h_T, \boldsymbol{x}_t - \boldsymbol{x}_{t-1} \rangle \qquad (16) \\
&= \frac{1}{\alpha}(\langle h_T, \sum_{t=1}^{T-1}\boldsymbol{x}_t + \alpha \boldsymbol{x}_t \rangle + \frac{1}{\eta}R(h_T)) + \frac{\alpha - 1}{\alpha}\sum_{t=1}^{T-1}\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle \\
&\leq \frac{1}{\alpha}(\langle g_T, \sum_{t=1}^{T-1}\boldsymbol{x}_t + \alpha \boldsymbol{x}_t \rangle + \frac{1}{\eta}R(g_T)) + \frac{\alpha - 1}{\alpha}\sum_{t=1}^{T-1}\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle \\
&\leq \frac{1}{\alpha}(\langle \boldsymbol{f}', \sum_{t=1}^{T-1}\boldsymbol{x}_t + \boldsymbol{x}_t \rangle + \frac{1}{\eta}R(\boldsymbol{f}')) + \frac{\alpha - 1}{\alpha}\sum_{t=1}^{T}\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle \ \forall \boldsymbol{f}'.
\end{aligned}$$

该证明由 $\boldsymbol{f}_t, \boldsymbol{g}_t$ 和 $\boldsymbol{h}_t$ 的最优性推导而来。至此，归纳论证完成。现在，我们只需对总和进行界定：

$$\sum_{t=1}^{T}\langle \boldsymbol{f}_t - \boldsymbol{h}_t, \boldsymbol{x}_t - \boldsymbol{x}_{t-1} \rangle.$$

利用强凸函数的性质，我们有：

$$\begin{aligned} F_t(\boldsymbol{h}_t) - F_t(\boldsymbol{f}_t) &\geq \frac{\beta}{2\eta}||\boldsymbol{h}_t - \boldsymbol{f}_t||_p^2 \\ H_t(\boldsymbol{f}_t) - H_t(\boldsymbol{h}_t) &\geq \frac{\beta}{2\eta}||\boldsymbol{h}_t - \boldsymbol{f}_t||_p^2 \\ \implies \alpha\langle \boldsymbol{h}_t - \boldsymbol{f}_t, \boldsymbol{x}_{t-1} - \boldsymbol{x}_t\rangle &\geq \frac{\beta}{\eta}||\boldsymbol{h}_t - \boldsymbol{f}_t||_p^2 \\ \implies ||\boldsymbol{h}_t - \boldsymbol{f}_t||_p||\boldsymbol{x}_{t-1} - \boldsymbol{x}_t||_q &\geq \frac{\beta}{\eta\alpha}||\boldsymbol{h}_t - \boldsymbol{f}_t||_p^2 \\ \implies ||\boldsymbol{x}_{t-1} - \boldsymbol{x}_t||_q &\geq \frac{\beta}{\eta\alpha}||\boldsymbol{h}_t - \boldsymbol{f}_t||_p. \end{aligned}$$

因此，我们有

$$\begin{aligned} \sum_{t=1}^{T}\langle \boldsymbol{f}_t - \boldsymbol{h}_t, \boldsymbol{x}_t - \boldsymbol{x}_{t-1}\rangle &\leq \sum_{t=1}^{T}||\boldsymbol{f}_t - \boldsymbol{h}_t||p||\boldsymbol{x}_t - \boldsymbol{x}_{t-1}||_q \\ &\leq \frac{\eta\alpha}{\beta}\sum_{t=1}^{T}(||\boldsymbol{x}_{t-1} - \boldsymbol{x}_t||_q)^2. \end{aligned}$$

结合不等式 15 即可得出完整证明。 □

引理 B.7（引理 5.1）。设 $\boldsymbol{x}_t$、$\boldsymbol{x}_{t+1}$ 是某个无悔算法（即跟随正则化领导者、在线镜像下降）的两个连续策略。那么，我们有：

$$\langle \boldsymbol{b}, \boldsymbol{x}_{t+1}\rangle - \langle \boldsymbol{c}, \boldsymbol{x}_{t+1}\rangle = O(\frac{1}{\sqrt{T}}),\ \textit{where}\ \boldsymbol{b} = \arg\min_{f\in\mathcal{F}}\langle \boldsymbol{f}, \boldsymbol{x}_t\rangle\ ,\ \boldsymbol{c} = \arg\min_{f\in\mathcal{F}}\langle \boldsymbol{f}, \boldsymbol{x}_{t+1}\rangle.$$

引理 5.1 的证明。由于 $\boldsymbol{b} = \arg\min_{f\in\mathcal{F}} \boldsymbol{f},\ \boldsymbol{x}_t$，因此我们有：$\boldsymbol{b},\ \boldsymbol{x}_t \leq \boldsymbol{c},\ \boldsymbol{x}_t$。因此，我们可以推导出：

$$\begin{aligned} &\langle \boldsymbol{b}, \boldsymbol{x}_{t+1}\rangle - \langle \boldsymbol{c}, \boldsymbol{x}_{t+1}\rangle = \langle \boldsymbol{b}, \boldsymbol{x}_{t+1}\rangle - \langle \boldsymbol{b}, \boldsymbol{x}_t\rangle + \langle \boldsymbol{b}, \boldsymbol{x}_t\rangle - \langle \boldsymbol{c}, \boldsymbol{x}_{t+1}\rangle \\ &\leq \langle \boldsymbol{b}, \boldsymbol{x}_{t+1} - \boldsymbol{x}_t\rangle + \langle \boldsymbol{c}, \boldsymbol{x}_t\rangle - \langle \boldsymbol{c}, \boldsymbol{x}_{t+1}\rangle = \langle \boldsymbol{b}, \boldsymbol{x}_{t+1} - \boldsymbol{x}_t\rangle + \langle \boldsymbol{c}, \boldsymbol{x}_t - \boldsymbol{x}_{t+1}\rangle. \end{aligned}$$

利用引理 4.1，可得 $\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|_1 = O(\frac{1}{\sqrt{T}})$，且 $\boldsymbol{b},\ \boldsymbol{c} \in [0,1]^n$，因此我们有：

$$\begin{aligned} &\langle \boldsymbol{b}, \boldsymbol{x}_{t+1}\rangle - \langle \boldsymbol{c}, \boldsymbol{x}_{t+1}\rangle = \langle \boldsymbol{b}, \boldsymbol{x}_{t+1} - \boldsymbol{x}_t\rangle + \langle \boldsymbol{c}, \boldsymbol{x}_t - \boldsymbol{x}_{t+1}\rangle \\ &\leq \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|_1 + \|\boldsymbol{x}_t - \boldsymbol{x}_{t+1}\|_1 \leq 2O(\frac{1}{\sqrt{T}}) = O(\frac{1}{\sqrt{T}}). \end{aligned}$$

证明完毕。 □

B.1 AMWU 最后一轮收敛的缺失证明

B.1.1 降低 *K-L* 距离。 在本小节中，我们的部分分析基于 AMWU 的线性变体及其以下更新规则：

$$\boldsymbol{f}_{t+1}(i) = \frac{\boldsymbol{f}_t(i)(1 + \eta((\alpha+1)e_i{}^\top \boldsymbol{A}\boldsymbol{y}_t - \alpha e_i{}^\top \boldsymbol{A}\boldsymbol{y}_{t-1}))}{\sum_j \boldsymbol{f}_t(j)(1 + \eta((\alpha+1)e_j{}^\top \boldsymbol{A}\boldsymbol{y}_t - \alpha e_j{}^\top \boldsymbol{A}\boldsymbol{y}_{t-1}))}.$$

由于该变体的更新规则不包含指数部分，因此降低了分析的复杂度。我们首先通过以下引理量化 AMWU 两次连续更新之间的距离：

引理 B.8。 设 $f \in \Delta_n$ 为最大化玩家的向量， w、 $z \in \Delta_m$ 满足 $\|w - z\|_1 = O(\eta)$、$\eta\alpha = O(1)$，并设 f', f'' 分别为 $AMWU$ 及其线性变体在当前向量 f 以及最小化玩家的向量 w、 z 下的下一次迭代。则有

$$\|f' - f''\|_1 \text{ is } O(\eta^2) \text{ and } \|f' - f\|_1, \|f'' - f\|_1 \text{ are } O(\eta).$$

类似地，对于最小化玩家的向量 $y \in \Delta_m$ 及其后续迭代，该结论也成立。

证明。令 η 足够小（小于 A中绝对值最大的元素）。根据假设， $\|w - z\|_1 = O(\eta)$和 $O(\eta\alpha) = O(1)$，我们有：

$$(\alpha + 1)(Aw)_i - \alpha(Az)_i = (Aw)_i + O(1).$$

因此，我们可以推导出以下等式：

$$\begin{aligned}
|f_i' - f_i''| &= f_i \left| \frac{e^{\eta((\alpha+1)(Aw)_i - \alpha(Az)_i)}}{\sum_j f_j e^{\eta((\alpha+1)(Aw)_j - \alpha(Az)_j)}} - \frac{1 + \eta((\alpha+1)(Aw)_i - \alpha(Az)_i)}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j))} \right| \\
&= f_i \left| \frac{1 + \eta((\alpha+1)(Aw)_i - \alpha(Az)_i) \pm O(\eta^2)}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j)) \pm O(\eta^2)} - \frac{1 + \eta((\alpha+1)(Aw)_i - \alpha(Az)_i)}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j))} \right| \\
&= f_i O(\eta^2).
\end{aligned}$$

因此， $\|f' - f''\|_1$ 是 $O(\eta^2)$。此外，我们有

$$\begin{aligned}
|f_i - f_i''| &= f_i \left| 1 - \frac{1 + \eta((\alpha+1)(Aw)_i - \alpha(Az)_i)}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j))} \right| \\
&= f_i \left| \frac{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j)) - (1 + \eta((\alpha+1)(Aw)_i - \alpha(Az)_i))}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j))} \right| \\
&= f_i \left| \frac{\sum_j f_j (\eta((\alpha+1)(Aw)_j - \alpha(Az)_j)) - \eta((\alpha+1)(Aw)_i - \alpha(Az)_i)}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j))} \right| \\
&= f_i \left| \frac{\eta \left(\sum_j f_j ((\alpha+1)(Aw)_j - \alpha(Az)_j) - ((\alpha+1)(Aw)_i - \alpha(Az)_i) \right)}{\sum_j f_j (1 + \eta((\alpha+1)(Aw)_j - \alpha(Az)_j))} \right| \\
&= f_i O(\eta).
\end{aligned}$$

结合上述两个证明，利用三角不等式即可得到引理的第三部分。□

Lemma B.9. 令 (f_t, y_t) 表示 $AMWU$ 动力学的第 t 次迭代。对于 $t \geq 2$，有 t

$$\begin{aligned}
f^{*\top} A((\alpha+1)y_t - \alpha y_{t-1}) &\geq f^{*\top} A y^* \text{ and} \\
((\alpha+1)f_t - \alpha f_{t-1})^\top A y^* &\leq f^{*\top} A y^*
\end{aligned}$$

证明。只需证明 $((\alpha+1)y_t - \alpha y_{t-1}) \in \Delta_m$ 且 $((\alpha+1)f_t - \alpha f_{t-1}) \in \Delta_n$。

根据引理 6.2，我们有 $f_t(i) = (1 - O(\eta)) f_{t-1}(i)$。因此，为了证明 $((\alpha+1)f_t(i) - \alpha f_{t-1}(i)) \geq 0$，我们需要证明：

$$\begin{aligned}
&(1 - O(\eta)) \geq \frac{\alpha}{\alpha + 1} \\
\iff\ &1 \geq (\alpha + 1) O(\eta),
\end{aligned} \tag{17}$$

这是成立的，因为 $\alpha\eta = \eta^b$， $b \in [0, 1]$ 并且 η 足够小。 □

定理 B.10。 设 (f^*, y^*) 为唯一最优极小极大均衡，且 η 足够小。假设 $\alpha\eta = \eta^b$，其中 $b \in [0, 1]$。那么

$$RE((f^*, y^*)||(f_t, y_t))$$

随时间 t 递减 η^{2+b}，除非 (f_t, y_t) 是 $O(\eta^{b/3})$ $-$ $close$。

证明。我们计算两个相连策略之间相对熵距离的差:

$$\begin{aligned}
&RE((f^*, y^*)||(f_{t+1}, y_{t+1})) - RE((f^*, y^*)||(f_t, y_t)) \\
&= -\left(\sum_i f^*(i)\log(\frac{f_{t+1}(i)}{f_t(i)}) + \sum_i y^*(i)\log(\frac{y_{t+1}(i)}{y_t(i)})\right) \\
&= -\left(\sum_i f^*(i)\log(e^{\eta((\alpha+1)Ay_t - \alpha Ay_{t-1})(i)}) + \sum_i y^*(i)\log(e^{\eta(-(\alpha+1)A^\top f_t + \alpha A^\top f_{t-1})(i)})\right) \\
&+ \log\left(\sum_i f_t(i)e^{\eta((\alpha+1)Ay_t - \alpha Ay_{t-1})(i)}\right) + \log\left(\sum_i y_t(i)e^{\eta(-(\alpha+1)A^\top f_t + \alpha A^\top f_{t-1})(i)}\right) \\
&= -\eta x^{*\top} A((\alpha+1)y_t - \alpha y_{t-1}) - \eta y^{*\top} A^\top(-(\alpha+1)f_t + \alpha f_{t-1}) + \\
&\log\left(\sum_i f_t(i)e^{\eta((\alpha+1)Ay_t - \alpha Ay_{t-1})(i)}\right) + \log\left(\sum_i y_t(i)e^{\eta(-(\alpha+1)A^\top f_t + \alpha A^\top f_{t-1})(i)}\right).
\end{aligned} \tag{18}$$

根据引理 B.9，我们有:

$$-\eta f^{*\top} A((\alpha+1)y_t - \alpha y_{t-1}) - \eta y^{*\top} A^\top(-(\alpha+1)f_t + \alpha f_{t-1}) \le 0.$$

因此，我们有:

$$\begin{aligned}
&RE((f^*, y^*)||(f_{t+1}, y_{t+1})) - RE((f^*, y^*)||(f_t, y_t)) \\
&\le \log\left(\sum_i f_t(i)e^{\eta((\alpha+1)Ay_t - \alpha Ay_{t-1})(i)}\right) + \log\left(\sum_i y_t(i)e^{\eta(-(\alpha+1)A^\top f_t + \alpha A^\top f_{t-1})(i)}\right) \\
&= \log\left(\sum_i f_t(i)e^{\eta((\alpha+1)((Ay_t)(i) - f_t^\top Ay_t) - \alpha((Ay_{t-1})(i) - f_t^\top Ay_{t-1}))}\right) \\
&+ \log\left(\sum_i y_t(i)e^{\eta(-(\alpha+1)((A^\top f_t)(i) - f_t^\top Ay_t) + \alpha((A^\top f_{t-1})(i) - f_{t-1}^\top Ay_t))}\right) + \alpha\eta(f_{t-1}^\top Ay_t - f_t^\top Ay_{t-1}) \\
&= \log\left(\sum_i f_t(i)e^{\eta((e_i - f_t)^\top A((\alpha+1)y_t - \alpha y_{t-1}))}\right) + \log\left(\sum_i y_t(i)e^{\eta((-(\alpha+1)f_t + \alpha f_{t-1})^\top A(e_i - y_t))}\right) \\
&+ \eta^b(f_{t-1}^\top Ay_t - f_t^\top Ay_{t-1}).
\end{aligned} \tag{19}$$

对函数 e^x（即 $e^x = 1 + x + \frac{1}{2}x^2$）进行泰勒近似（$\eta$ 足够小），并对 $\log(1+x) < x$ 在 $x > 0$ 时进行泰勒近似，我们便有：

$$\begin{aligned}
&= \log\left(\sum_i \boldsymbol{f}_t(i) e^{\eta((e_i-\boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))}\right) \\
&\le \log(\sum_i x_t(i)(1+\eta((e_i-\boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1})))+ \\
&\sum_i \boldsymbol{f}_t(i)((\frac{1}{2}+O(\eta^b))\eta^2((e_i-\boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2) \\
&= \log\left(1+\sum_i \boldsymbol{f}_t(i)((\frac{1}{2}+O(\eta^b))\eta^2((e_i-\boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2\right) \\
&\le \sum_i \boldsymbol{f}_t(i)((\frac{1}{2}+O(\eta^b))\eta^2((e_i-\boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2.
\end{aligned} \tag{20}$$

结合引理 6.5，我们得到：

$$\begin{aligned}
&RE((\boldsymbol{f}^*,\boldsymbol{y}^*)||(\boldsymbol{f}_{t+1},\boldsymbol{y}_{t+1})) - RE((\boldsymbol{f}^*,\boldsymbol{y}^*)||(\boldsymbol{f}_t,\boldsymbol{y}_t)) \\
&\sum_i(\frac{1}{2}+O(\eta^b))\eta^2 \boldsymbol{f}_t(i)((e_i-\boldsymbol{f}_t)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2+ \\
&\sum_i(\frac{1}{2}+O(\eta^b))\eta^2 \boldsymbol{y}_t(i)((\boldsymbol{y}_t-e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 \\
&-\frac{\eta^b}{\eta}(1-O(\eta))\eta^2\sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t-e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2- \\
&\frac{\eta^b}{\eta}(1-O(\eta))\eta^2\sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t-e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 + \frac{\eta^b}{\eta}\eta^b O(\eta^2) \\
&\le -(\frac{1}{2}-O(\eta^b))\eta^2\sum_i \boldsymbol{f}_t(i)((\boldsymbol{f}_t-e_i)^\top \boldsymbol{A}((\alpha+1)\boldsymbol{y}_t - \alpha\boldsymbol{y}_{t-1}))^2- \\
&(\frac{1}{2}-O(\eta^b))\eta^2\sum_i \boldsymbol{y}_t(i)((\boldsymbol{y}_t-e_i)^\top \boldsymbol{A}^\top((\alpha\boldsymbol{y}_{t-1} - (\alpha+1)\boldsymbol{y}_t))^2 + \eta^b O(\eta^2).
\end{aligned} \tag{21}$$

由于 $\frac{\eta^b}{\eta} > 1$。现在显然，只要 $(\boldsymbol{f}_t,\ \boldsymbol{y}_t)$ 且 $(\boldsymbol{f}_{t-1},\ \boldsymbol{y}_{t-1})$ 不与 $O(\eta^{b/3})$ 足够接近，根据上述不等式可得：

$$RE((\boldsymbol{f}^*,\boldsymbol{y}^*)||(\boldsymbol{f}_{t+1},\boldsymbol{y}_{t+1})) - RE((\boldsymbol{f}^*,\boldsymbol{y}^*)||(\boldsymbol{f}_t,\boldsymbol{y}_t)) \le -\Omega(\eta^{b+2}),$$

或者相对熵距离至少按 η^{b+2} 的因子递减，结论便得证。 □

B.1.2 $\eta^{b/3}$-接近蕴含接近最优解。我们首先需要以下引理：

引理 B.11。设 $i \in \mathrm{Supp}(\boldsymbol{f}^*)$ 和 $j \in \mathrm{Supp}(\boldsymbol{y}^*)$。有 $x_T(i) \ge \frac{1}{2}\eta^{b/3}$ 且 $y_T(i) \ge \frac{1}{2}\eta^{b/3}$，只要

$$\eta^{b/3} \le \min_{s\in\mathrm{Supp}(\boldsymbol{f}^*)} \frac{1}{(nm)^{1/\boldsymbol{f}^*(s)}},\ \min_{s\in\mathrm{Supp}(\boldsymbol{y}^*)} \frac{1}{(nm)^{1/\boldsymbol{y}^*(s)}}.$$

利用上述引理，我们可以沿用 [Daskalakis 和 Panageas, 018a] 中定理 3.2 的相同论证，证明以下定理：

定理 B.12. 假设(f^*, y^*)是该问题的唯一最优解。令 T 为 KL 散度不再下降 $\Omega(\eta^{b+2})$ 的最早时间。由此可得，当 $\eta \to 0$ 时， $\eta^{b/3}$-接近点(f_T, y_T)与(f^*, y^*)之间的距离趋于零：

$$\lim_{\eta \to 0} \|(f^*, y^*) - (f_T, y_T)\|_1 = 0.$$

证明。根据引理 6.7 和 T的定义，对于支撑集 f^*中的 i， $|(Ay_T)_i - f_T{}^\top A y_T|$为 $O(\eta^{1/3})$；对于支撑集 y^*中的 j， $|(f_T^\top A)_j - f_T^\top A y_T|$为 $O(\eta^{1/3})$。我们通过移除质量小于 $\frac{1}{2}\eta^{b/3}$的所有坐标并对其重新缩放，来考虑 (w_T, z_T) 对 (f_T, y_T) 的投影。我们有如下关系：

$$\lim_{\eta \to 0} \|(f_T, y_T) - (w_T, z_T)\| = 0. \tag{22}$$

由于对于 w和 z中的所有坐标，都有 $|(\hat{A}z_T)_i - w_T{}^\top \hat{A} z_T|$，且 $|(w_T^\top \hat{A})_j - w_T^\top \hat{A} z_T|$为$O(\eta^{b/3})$，因此 (w, z) 是博弈 $\hat{A}$的 $O(\eta^{b/3})$-近似解。使用以下引理：

引理 B.13（ [Daskalakis 和 Panageas，018a]中的命题 3.5）。设（x^*、 y^*）是该博弈的唯一最优解。对于每个 $\epsilon > 0$，都存在一个 γ，使得对于每个 γ-近似解（x、 y），对所有 $i \in [n]$，均有 $|x_i - x_1^*| < \epsilon$。对于玩家 y，类似结论也成立。

利用上述引理并取足够小的 η，我们可知，对于所有不在 x^* 支撑集中的 i，均有 $|w_i| < \eta^{b/3}$。由于子博弈 A 包含博弈 $\hat{A}$ 的纳什均衡支持集中的所有纯策略，因此子博弈 A 也将具有唯一纳什均衡，且权重与博弈 A 中的相同。因此，我们有：

$$\lim_{\eta \to 0} \|(w_T, z_T) - (f^*, y^*)\| = 0. \tag{23}$$

结合公式 (22) 和 (23) 即可得证。 □

B.1.3 局部收敛证明。我们利用动力系统中以下广为人知的事实来证明局部收敛：

命题 B.14（参见 [Galor， 2007]）。如果连续可微更新规则w 在不动点 z 处的雅可比矩阵的谱半径小于 *1*，那么存在一个邻域 U，围绕着 z，使得对于所有 $x \in U$，动态都会收敛到 z。

据此，我们的局部收敛定理如下：

定理 B.15。设(f^*, y^*)为博弈 A的唯一极小极大均衡。存在(f^*, y^*)的一个邻域，使得 *E-OMWU* 动力学收敛。

证明。AMWU 的更新规则可描述为以下动力系统：

$$\begin{aligned}
&g(f, y, z, w) := (g_1(f, y, z, w), g_2(f, y, z, w), g_3(f, y, z, w) g_4(f, y, z, w)) \\
&g_{1,i}(f, y, z, w) := (g_1(f, y, z, w))_i := f_i \frac{e^{\eta((\alpha+1)e_i{}^\top A y - \alpha e_i{}^\top A w)}}{\sum_i f_i e^{\eta((\alpha+1)e_i{}^\top A y - \alpha e_i{}^\top A w)}} \ \forall i \in [n] \\
&g_{2,i}(f, y, z, w) := (g_2(f, y, z, w))_i := y_i \frac{e^{-\eta((\alpha+1)e_i{}^\top A x - \alpha e_i{}^\top A z)}}{\sum_i y_i e^{-\eta((\alpha+1)e_i{}^\top A x - \alpha e_i{}^\top A z)}} \ \forall i \in [m] \\
&g_3(f, y, z, w) := I_{n \times n} f \\
&g_4(f, y, z, w) := I_{m \times m} y.
\end{aligned} \tag{24}$$

很容易证明，(f^*, y^*, f^*, y^*) 是驻点。根据命题 B.14，只需证明 g 在 (f^*, y^*, f^*, y^*) 处的雅可比矩阵的特征值小于 1。

现在，我们计算 g 在点 $(\boldsymbol{f}^*,\ \boldsymbol{y}^*,\ \boldsymbol{f}^*,\ \boldsymbol{y}^*)$ 处的雅可比矩阵，并证明其谱半径小于 1。我们研究在驻点 $(\boldsymbol{f}^*,\ \boldsymbol{y}^*,\ \boldsymbol{f}^*,\ \boldsymbol{y}^*)$ 处计算得到的雅可比矩阵。

令 v为博弈值，$\boldsymbol{f}^*,\ \boldsymbol{y}^*$为唯一的极小极大均衡（即 $\boldsymbol{f}^{*\top}\boldsymbol{A}\boldsymbol{y}^* = v$）。对于 $i \notin Supp(\boldsymbol{f}^*)$（例如 $\boldsymbol{f}_i^* = 0$），我们有

$$\frac{\partial g_{1,i}}{\partial f_i} = \frac{e^{\eta(\boldsymbol{A}\boldsymbol{y}^*)_i}}{\sum_t \boldsymbol{f}_t^* e^{\eta(\boldsymbol{A}\boldsymbol{y}^*)_t}} = \frac{e^{\eta(\boldsymbol{A}\boldsymbol{y}^*)(i)}}{e^{\eta v}}$$

且其他偏导数均为零。因此，$\frac{e^{\eta(\boldsymbol{A}\boldsymbol{y}_i^*)}}{e^{\eta v}} < 1$ 是由最优解处计算得到的雅可比矩阵的一个特征值（例如，由于唯一性，$\boldsymbol{A}\boldsymbol{y}_i^* < v$）。类似地，对于$j \notin Supp(\boldsymbol{y}^*)$，$\frac{\partial \boldsymbol{g}_{2\,j}}{\partial \boldsymbol{y}_j} = \frac{e^{-\eta(\boldsymbol{A}^\top \boldsymbol{x}^*)_j}}{e^{-\eta v}} < 1$ 是雅可比矩阵的一个特征值。删除对应于上述特征值的行和列后，我们构造矩阵 J，其中仅包含 $\boldsymbol{f}^*$ 和 $\boldsymbol{y}^*$的支撑集中的元素。由此可知，雅可比矩阵的谱半径小于 1，当且仅当新矩阵 J的谱半径小于 1。记 $D_x,\ D_y$ 为分别包含 $\boldsymbol{f}^*$ 和 $\boldsymbol{y}^*$ 非零元素的对角矩阵。令 $\boldsymbol{B}$为与 $\boldsymbol{f}^*,\ \boldsymbol{y}^*$的非零元素相对应的收益$\boldsymbol{A}$的子矩阵。于是，矩阵 J 如下：

$$A = \begin{bmatrix} \boldsymbol{I}_{k_1\times k_1} - D_x 1_k 1_k^T & \eta(\alpha+1)D_x(\boldsymbol{B} - v1_{k_1}1_{k_2}^\top) & 0_{k1\times k1} & -\eta\alpha D_x(\boldsymbol{B} - v1_{k_1}1_{k_2}^T) \\ (\alpha+1)\eta D_y(v1_{k_2}1_{k_1}^\top - \boldsymbol{B}^\top) & \boldsymbol{I}_{k_2\times k_2} - D_y 1_{k_2}1_{k_2}^\top & -\eta\alpha D_y(v1_{k_2}1_{k_1}^\top - \boldsymbol{B}^\top) & 0_{k_2\times k_2} \\ \boldsymbol{I}_{k_1\times k_1} & 0_{k_1\times k_2} & 0_{k_1\times k_1} & 0_{k1\times k_2} \\ 0_{k_2\times k_1} & \boldsymbol{I}_{k_2\times k_2} & 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{bmatrix}$$

显然，$(1_{k_1},\ 0_{k_2},\ 0_{k_1},\ 0_{k_2})$、$(0_{k_1},\ 1_{k_2},\ 0_{k_1},\ 0_{k_2})$ 是特征值为零的左特征向量，因此，任何具有非零特征值的右特征向量 $(\boldsymbol{f},\ \boldsymbol{y},\ \boldsymbol{z},\ \boldsymbol{w})$ 都具有性质 $\boldsymbol{f}^\top 1_{k_1} = 0$和$\boldsymbol{y}^\top 1_{k_2} = 0$。因此，上述矩阵的每个非零特征值都是下列矩阵的特征值：

$$J_{new} = \begin{bmatrix} \boldsymbol{I}_{k_1\times k_1} & \eta(\alpha+1)D_x\boldsymbol{B} & 0_{k1\times k1} & -\eta\alpha D_x\boldsymbol{B} \\ -(\alpha+1)\eta D_y\boldsymbol{B}^\top & \boldsymbol{I}_{k_2\times k_2} & \eta\alpha D_y\boldsymbol{B}^\top & 0_{k_2\times k_2} \\ \boldsymbol{I}_{k_1\times k_1} & 0_{k_1\times k_2} & 0_{k_1\times k_1} & 0_{k1\times k_2} \\ 0_{k_2\times k_1} & \boldsymbol{I}_{k_2\times k_2} & 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{bmatrix}$$

利用 t he 分块矩阵的行列式，我们得到该矩阵的特征多项式：

$$J_{new} = (-1)^k \det\left(\begin{bmatrix} \lambda(1-\lambda)\boldsymbol{I}_{k_1\times k_1} & \eta(\lambda(\alpha+1)-\alpha)D_x\boldsymbol{B} \\ -\eta(\lambda(\alpha+1)-\alpha)D_y\boldsymbol{B}^\top & \lambda(1-\lambda)\boldsymbol{I}_{k_2\times k_2} \end{bmatrix}\right)$$

这等价于

$$(\alpha - (\alpha+1)\lambda)^k q\left(\frac{\lambda(\lambda-1)}{(\alpha+1)\lambda-\alpha}\right),$$

其中 $q(\lambda)$ 是下列矩阵的特征多项式

$$J_{small} = \left(\begin{bmatrix} 0_{k_1\times k_1} & \eta D_x\boldsymbol{B} \\ -\eta D_y\boldsymbol{B}^\top & 0_{k_2\times k_2} \end{bmatrix}\right)$$

根据 [Daskalakis 和 Panageas, 018a]，中的引理 B.6 J_{small}，我们有 $\pm i\eta\tau$ 的特征值形式为 $\tau \in \mathrm{R}$。记 $\sigma := \eta\tau$，因此 σ 和 $\sigma\alpha$ 的绝对值可以足够小。由此可得，矩阵 J 的任意非零特征值 λ 都满足：

$$\begin{aligned} &\frac{\lambda(\lambda-1)}{(\alpha+1)\lambda-\alpha} = i\sigma \\ &\iff \lambda^2 - \lambda(1 + i\sigma(\alpha+1)) + i\sigma\alpha = 0 \\ &\lambda = \frac{1 + i\sigma(\alpha+1) \pm \sqrt{1 - \sigma^2(\alpha+1)^2 - i2\sigma(\alpha-1)}}{2}. \end{aligned}$$

假设 $\sqrt{1-\sigma^2(\alpha+1)^2-i2\sigma(\alpha-1)}=x+iy$，那么我们可以推导出，为了在 σ 相对较小时使 λ 的幅度最大，我们有：

$$x=\sqrt{\frac{1-\sigma^2(\alpha+1)^2+\sqrt{(1-\sigma^2(\alpha+1)^2)^2+4\sigma^2(\alpha-1)^2}}{2}},\ y=\frac{-\sigma(\alpha-1)}{x}$$

因此，λ 的模平方为：

$$\frac{(1+x)^2+(\sigma(\alpha+1)+y)^2}{4}$$

注意，当 σ足够小时：

$$\begin{aligned}x&=\sqrt{\frac{1-\sigma^2(\alpha+1)^2+\sqrt{(1+\sigma^2(\alpha+1)^2)^2-16\sigma^2\alpha}}{2}}\\&\le\sqrt{\frac{1-\sigma^2(\alpha+1)^2+(1+\sigma^2(\alpha+1)^2)-2\sigma^2\alpha}{2}}\\&=\sqrt{1-\sigma^2\alpha}\end{aligned}$$

此外，我们有：

$$\begin{aligned}x&=\sqrt{\frac{1-\sigma^2(\alpha+1)^2+\sqrt{(1+\sigma^2(\alpha+1)^2)^2-16\sigma^2\alpha}}{2}}\\&\ge\sqrt{\frac{1-\sigma^2(\alpha+1)^2+(1+\sigma^2(\alpha+1)^2)-8\sigma^2\alpha}{2}}\\&=\sqrt{1-4\sigma^2\alpha}.\end{aligned}$$

由于 $\sqrt{1-4\sigma^2\alpha}\le x\le 1$，我们有：

$$\frac{-\sigma(\alpha-1)}{\sqrt{1-4\sigma^2\alpha}}\le y=\frac{-\sigma(\alpha-1)}{x}\le-\sigma(\alpha-1).$$

我们将证明：

$$\begin{aligned}&\sigma(\alpha+1)+\frac{-\sigma(\alpha-1)}{\sqrt{1-4\sigma^2\alpha}}\ge 0\\&\iff(\alpha+1)\ge\frac{(\alpha-1)}{\sqrt{1-4\sigma^2\alpha}}\\&\iff(\alpha^2+2\alpha+1)(1-4\sigma^2\alpha)\ge(\alpha-1),\end{aligned}$$

由于 σ和 $\sigma\alpha$可以设得足够小，因此该结论成立。于是我们有：

$$0\le\sigma(\alpha+1)+y\le 2\sigma$$

于是我们有：

$$\begin{aligned}&\frac{(1+x)^2+(\sigma(\alpha+1)+y)^2}{4}\le\frac{(1+\sqrt{1-4\sigma^2\alpha})^2+(2\sigma)^2}{4}\\&\le\frac{2+2\sqrt{1-4\sigma^2\alpha}-4\sigma^2\alpha+4\sigma^2}{4}\le 1,\end{aligned}$$

由于 $\alpha\ge 1$，且等式仅在 $\sigma=0$时成立。对于 $\sigma=0$，这意味着 J_{new} 有一个等于 1 的特征值。设 $(\hat{\boldsymbol{x}},\ \hat{\boldsymbol{y}},\ \hat{z},\ \hat{\boldsymbol{w}})$ 是对应的特征向量。于是我们有 $\boldsymbol{I}\hat{\boldsymbol{x}}-\boldsymbol{I}\hat{z}=0$ 和 $\boldsymbol{I}\hat{\boldsymbol{y}}-\boldsymbol{I}\hat{\boldsymbol{w}}=0$，因此可得：$\hat{\boldsymbol{x}}=\hat{z}$和$\hat{\boldsymbol{y}}=\hat{\boldsymbol{w}}$。此外，我们还有：$D_x\boldsymbol{B}\hat{\boldsymbol{x}}=0$和 $D_y\boldsymbol{B}^\top\hat{\boldsymbol{y}}=0$，因此有 $\boldsymbol{B}\hat{\boldsymbol{x}}=0$和 $\boldsymbol{B}^\top\hat{\boldsymbol{y}}=0$。由前面的论证，我们

也有：$\hat{\boldsymbol{x}}^\top 1_{k_1} = 0$ 和 $\hat{\boldsymbol{y}}^\top 1_{k_2} = 0$。因此，策略 $(\boldsymbol{x}^*,\ \boldsymbol{y}^*) + t(\hat{\boldsymbol{x}},\ \hat{\boldsymbol{y}})$ 也是一个最优策略，只要 t 足够小，就能使每个元素都非负。根据唯一性假设，我们有 $\hat{\boldsymbol{x}} = 0,\ \hat{\boldsymbol{y}} = 0$，这与假设矛盾。因此，矩阵 J 的每个特征值的模都小于 1。证明完毕。□

导数计算
集合 $S_{\boldsymbol{x}} = \sum_i f_i e^{\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{w})}$ 和 $S_{\boldsymbol{y}} = \sum_i \boldsymbol{y}_i e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})}$。在购彩平台彩票平台彩票吗。$(\boldsymbol{f}^*, \boldsymbol{y}^*, \boldsymbol{f}^*, \boldsymbol{y}^*)$ 如下：

$$
\begin{aligned}
&\frac{\partial g_{1,i}}{\partial f_i} = \frac{e^{\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{w})}}{S_{\boldsymbol{x}}} - f_i \frac{e^{2\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{w})}}{{S_{\boldsymbol{x}}}^2} \quad \forall i \in [n],\\
&\frac{\partial g_{1,i}}{\partial \boldsymbol{x}_j} = f_i e^{\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{w})} \frac{-e^{\eta((\alpha+1)e_j^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_j^\top \boldsymbol{A}\boldsymbol{w})}}{{S_{\boldsymbol{x}}}^2} \quad \forall i \in [n], j \in [m], j \neq i,\\
&\frac{\partial g_{1,i}}{\partial \boldsymbol{y}_j} = f_i e^{\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{w})} \frac{\eta(\alpha+1)\boldsymbol{A}_{i,j} S_{\boldsymbol{x}} - \eta(\alpha+1)\sum_t \boldsymbol{A}_{tj}\boldsymbol{x}_t e^{\eta((\alpha+1)e_t^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_t^\top \boldsymbol{A}\boldsymbol{w})}}{S_{\boldsymbol{x}}^2} \quad \forall i \in [n], j = i,\\
&\frac{\partial g_{1,i}}{\partial \boldsymbol{z}_j} = 0 \quad \forall i, j \in [n],\\
&\frac{\partial g_{1,i}}{\partial \boldsymbol{w}_j} = f_i e^{\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{w})} \frac{-\alpha\eta \boldsymbol{A}_{ij} S_{\boldsymbol{x}} + \eta\alpha \sum_t \boldsymbol{A}_{tj}\boldsymbol{x}_t e^{\eta((\alpha+1)e_t^\top \boldsymbol{A}\boldsymbol{y} - \alpha e_t^\top \boldsymbol{A}\boldsymbol{w})}}{S_{\boldsymbol{x}}^2} \quad \forall i \in [n], j \in [m].
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
&\frac{\partial g_{2,i}}{\partial \boldsymbol{y}_i} = \frac{e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})}}{S_{\boldsymbol{y}}} - \boldsymbol{y}_i \frac{e^{-2\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})}}{{S_{\boldsymbol{y}}}^2} \quad \forall i \in [m],\\
&\frac{\partial g_{2,i}}{\partial \boldsymbol{y}_j} = \boldsymbol{y}_i e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})} \frac{-e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})}}{{S_{\boldsymbol{y}}}^2} \quad \forall i \in [n], j \in [m], j \neq i,\\
&\frac{\partial g_{2,i}}{\partial \boldsymbol{x}_j} = \boldsymbol{y}_i e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})} \frac{-\eta(\alpha+1)\boldsymbol{A}_{i,j} S_{\boldsymbol{y}} + \eta(\alpha+1)\sum_t \boldsymbol{A}_{tj}\boldsymbol{y}_t e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})}}{S_{\boldsymbol{y}}^2} \quad \forall i \in [m], j \in [n],\\
&\frac{\partial g_{2,i}}{\partial \boldsymbol{z}_j} = \boldsymbol{y}_i e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})} \frac{\eta\alpha \boldsymbol{A}_{i,j} S_{\boldsymbol{y}} - \eta\alpha \sum_t \boldsymbol{A}_{tj}\boldsymbol{x}_t e^{-\eta((\alpha+1)e_i^\top \boldsymbol{A}\boldsymbol{x} - \alpha e_i^\top \boldsymbol{A}\boldsymbol{z})}}{S_{\boldsymbol{y}}^2} \quad \forall i \in [m], j \in [n],\\
&\frac{\partial g_{2,i}}{\partial \boldsymbol{w}_j} = 0 \quad \forall i, j \in [m],\\
&\frac{\partial g_{3,i}}{\partial f_i} = 1 \quad \forall i \in [n],\\
&\frac{\partial g_{4,i}}{\partial \boldsymbol{y}_i} = 1 \quad \forall i \in [m].
\end{aligned}
\tag{26}
$$

C 实验

C.1 无知对手

我们将实验设置说明如下。在选定的随机矩阵博弈中，首先让智能体使用固定的乘法权重更新，对抗使用集合 $[0.5, 0.45, 0.4, \ldots, 0.05]$ [10] 中选定学习率的乘法权重更新的对手。然后，我们记录对手每一轮的策略，并将其视为无知对手。为突出 AMWU 与乐观乘法权重更新之间的差异，

[10]每个学习率都会产生不同的无知对手。

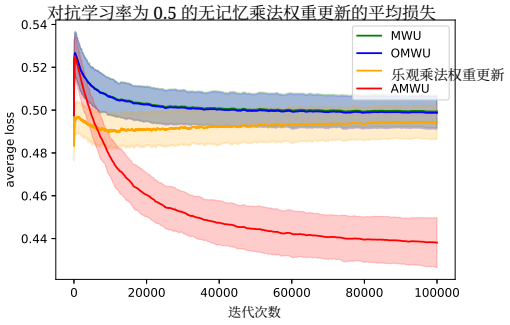
我们还测试了学习率为 $\eta = 1$ 时 OWMU 的性能。对于随机博弈，我们针对每种矩阵规模使用 5 个随机种子进行测试。对于元博弈，我们让算法与 5 个不同的无知对手（即学习率取自 $[0.5, 0.4, 0.3, 0.2, 0.1]$ 的乘法权重更新）进行对抗，并报告平均性能及标准差。

对无知对手的平均性能：我们报告 AMWU 和其他基线方法面对不同无知对手时的性能，即采用不同学习率 $[0.5, 0.45, 0.4. \ldots, 0.05]$ 的乘法权重更新对手。如图 4 和图 5 所示，在随机矩阵博弈的所有对手设置中，AMWU 的表现都大幅优于其他基线方法。图 6 中的四子棋和 Disc 实验也呈现出类似趋势。

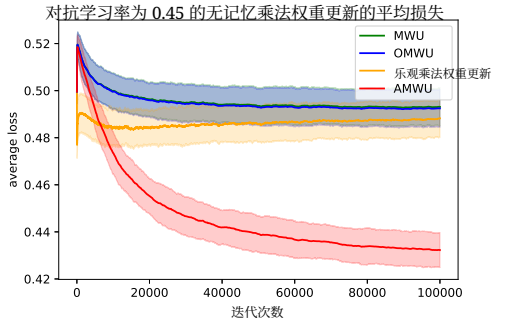
C.2 AMWU 的最后一轮收敛

为公平比较，我们为算法 AMWU 以及基线方法乘法权重更新和乐观乘法权重更新设定相同的学习率。在平均性能实验中，我们首先设定共同学习率 $\eta = 0.01$ 和利用率 $\alpha = 100$。为了凸显 AMWU 与乐观乘法权重更新之间的差异，我们还测试了学习率为 $\eta = 1$ 的乐观乘法权重更新的性能。也就是说，该 OWMU 对可预测序列 $\boldsymbol{x}_{t-1}$ 和正则化项 R 的相对权重与 AMWU 相同（$\boldsymbol{f}$）（即 $\eta_{OMWU} = \eta_{AMWU} \times \alpha_{AMWU}$）。在最后一轮收敛实验中，我们改变共同学习率 η（即 $\eta = [0.01, 0.025, 0.05]$），以观察所见的收敛趋势是否不受学习率影响。这里我们聚焦于随机矩阵博弈（20×20 和 50×50 维），因为其具有纳什均衡唯一这一良好性质，而 AMWU 和 OWMU 的收敛需要这一性质。由于不能保证学习率较大（例如 $\eta = 1$）时 OWMU 会收敛，因此在本实验中，我们不将 $OMWU_1$ 视为基线方法。

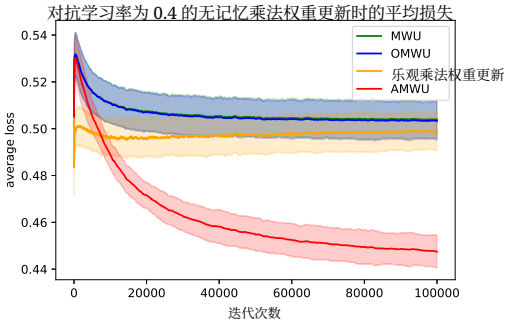
自博弈中的最后一轮收敛：我们报告了 AMWU 和其他基线在自博弈设置中的表现。如图 7、图 8 和图 9 所示，在三种不同的学习率设置下，AMWU 的表现都显著优于 OWMU 和乘法权重更新。正如 [Bailey andPiliouras, 2018] 所预期的那样，乘法权重更新在最后一轮收敛中表现出发散。在图 10 的四子棋和 Disc 实验中也能观察到类似趋势。



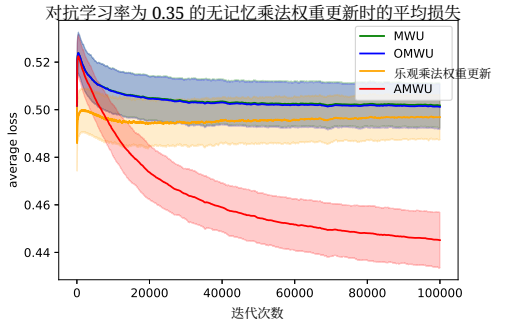
（a）随机博弈中学习率为 0.5 的乘法权重更新对手



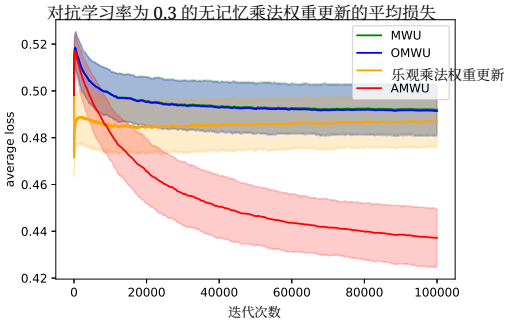
（b）学习率为 0.45 的乘法权重更新对手



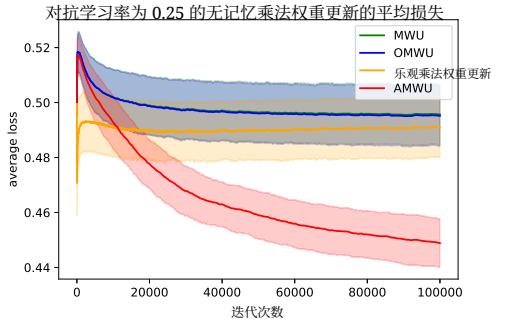
（c）学习率为 0.4 的乘法权重更新对手



（d）学习率为 0.35 的乘法权重更新对手

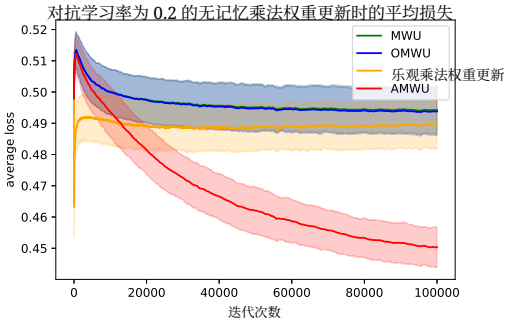


（e）学习率为 0.3 的乘法权重更新对手

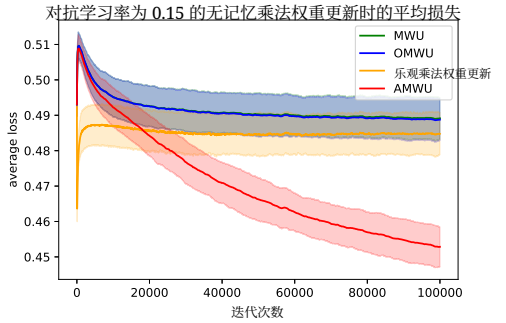


(f) 学习率为 0.25 的乘法权重更新对手

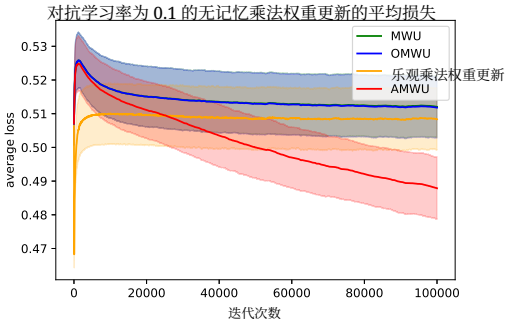
图 4. 在随机博弈中对抗不同的无知 MWU 对手



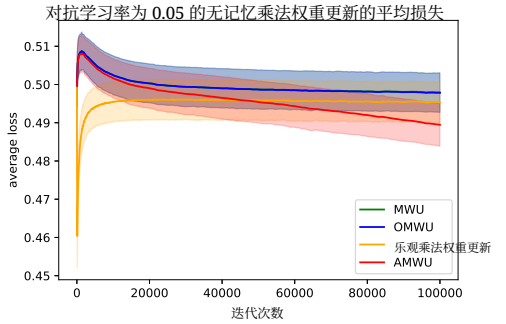
(a) 学习率为 0.2 的乘法权重更新对手



(b) 学习率为 0.15 的乘法权重更新对手

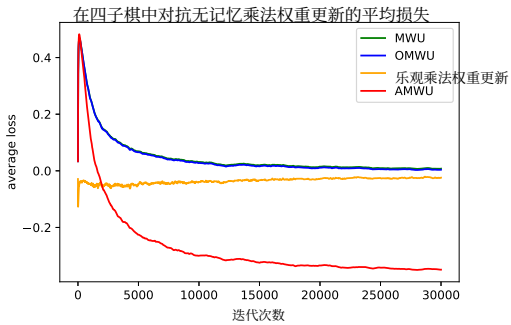


(c) 学习率为 0.1 的乘法权重更新对手

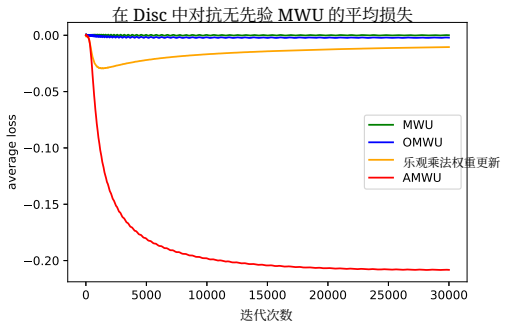


(d) 学习率为 0.05 的乘法权重更新对手

图 5. 在随机博弈中对抗不同的无知 MWU 对手

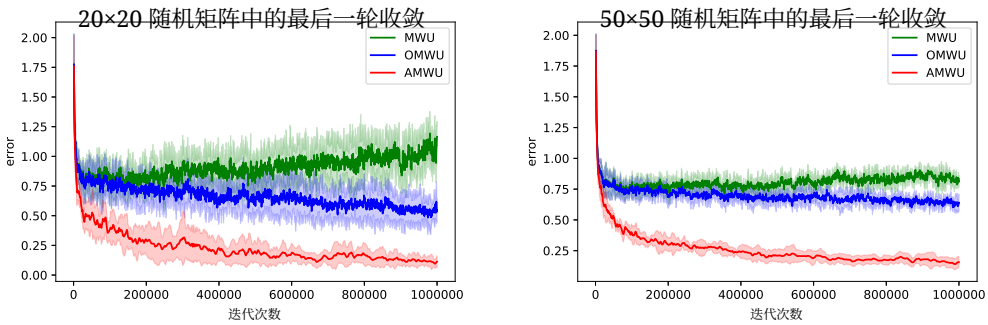


(a) 在四子棋中对抗乘法权重更新对手



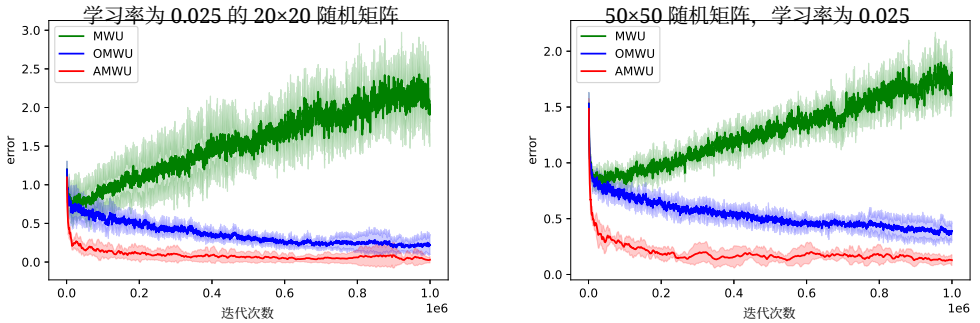
(b) 在 Disc 中对抗乘法权重更新对手

图 6. 在元博弈中对抗无知 MWU 对手



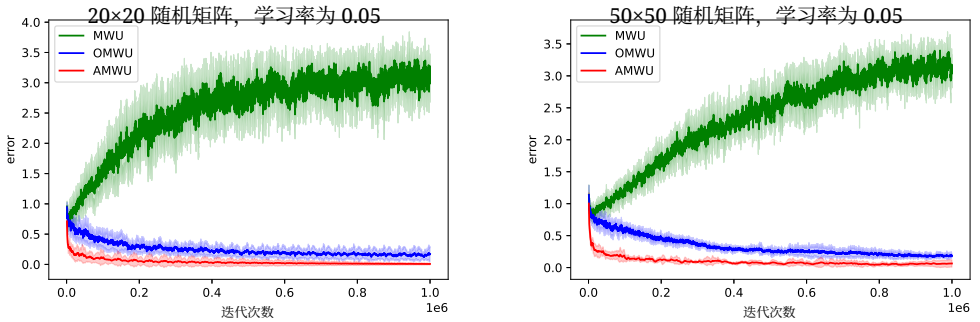
(a) 20×20 随机博弈 (b) 50×50 随机博弈

图 7. 最后一 轮随机博弈中的收敛情况，学习率为 0.01 率



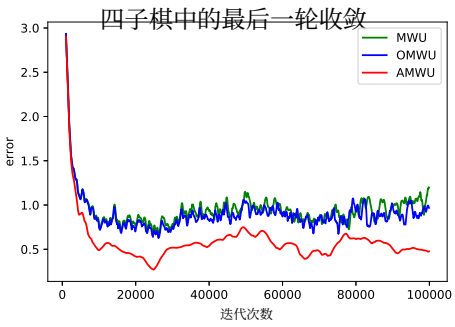
(a) 20×20 随机博弈 (b) 50×50 随机博弈

图 8. 学习率为 0.025 时随机博弈中的最后一轮收敛

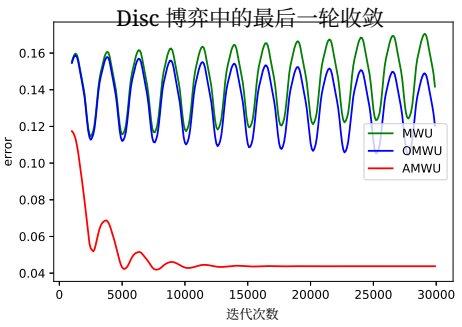


(a) 20×20 随机博弈 (b) 50×50 随机博弈

图 9. 学习率为 0.05 时随机博弈中的最后一轮收敛



（a）四子棋元博弈



(b) Disc 元博弈

图10. 元博弈中的最后一轮收敛 s