基于残差伪梯度估计的连续博弈零阶学习

Yuanhanqing Huang 和 Jianghai Hu

arXiv:2301.02279v1 [math.OC] 5 Jan 2023

摘要——**许多实际问题都可以建模为多玩家博弈中的决策过程：一群自利型参与者旨在优化各自的局部目标，而这些目标又取决于其他参与者采取的行动。在实现用于求解博弈的算法时，每位参与者的局部梯度信息至关重要，但往往无法获得。本文聚焦于利用赌博机反馈设计多玩家博弈的求解算法，即每位参与者唯一可用的反馈是实际获得的目标函数值。为解决现有单次预言机调用的赌博机学习算法方差较大的问题，我们将残差反馈方案融入单次调用外梯度方法，提出了两种算法。随后，我们证明，若博弈为伪单调加性，则实际博弈序列几乎必然收敛至一个临界点；并刻画了博弈为强伪单调时收敛至该临界点的收敛速率。我们还进一步研究了单调博弈中所生成序列的遍历收敛速率。最后，通过数值示例进一步验证了所提出算法的有效性。**

I. 引言

随着网络化工程系统日益普及，且其中存在对共享资源的竞争，解决具有竞争性利益的多智能体系统中的决策问题，已引起系统与控制领域越来越多的关注 [1]。例如，智能电网管理 [2], [3], 无线与通信网络 [4], [5], 交通运输系统[6], 等。博弈论源于开创性研究 [7], [8], ，为建模和分析自利型参与者之间的交互与动态提供了理论工具和框架。具体而言，在纳什均衡问题（NEP）中，每位参与者独立地从其策略集中选择一个行动，其偏好由局部目标函数体现。局部目标函数不仅取决于该参与者自身的行动，也会受到其他参与者所采取行动的影响。该领域的核心问题之一是设计算法或动态过程，使参与者能够达到纳什均衡（NE）等稳定解。

大多数现有方法（如 [9]–[11] ）都利用局部偏梯度信息进行更新，因此需要一阶预言机。由于偏梯度信息取决于其他博弈参与者采取的行动，计算一阶信息需要所有参与者提供的信息。通信和计算资源方面的实际限制，使得这种集中式一阶预言机无法应用，尤其是在大规模多智能体场景中。因此，研究者开始设计用于网络博弈的算法，以分布式方式计算偏梯度信息，并且仅要求在邻域内进行局部通信 [12]–[14]。然而，每个博弈参与者都需要维护对其他参与者行动的局部估计；当全局策略空间规模较大时，这会带来可扩展性问题。其他问题还包括：参与者不愿透露自己的行动、不了解自身的目标模型，或者更极端地，完全不知道博弈的存在。这促使我们放宽对反馈的假设，考虑老虎机设定：参与者唯一能够观察到的反馈信息是实际获得的目标函数值。具体而言，这组参与者将遵循一种典型的在线学习范式，其流程如下：每次迭代中，每个参与者选择一个行动，观察实际获得的目标函数值，并据此更新自己的行动，然后重复这一过程 [15], [16]。

相关工作：额外梯度（EG）方法在求解变分不等式和寻找博弈中的均衡方面有着悠久的历史，其早期推导可见于文献 [17]–[19]。过去几十年来，人们开发了大量 EG 随机变体，并对其收敛性质进行了详细研究 [20]–[22]。与前向-后向方法 [10] 或镜像下降法 [23], 等其他方法相比，EG 能够保证实际序列收敛，相关的遍历平均序列则以 $O(1/t)$ 的速率收敛。另一方面，EG 每次迭代的一阶信息查询次数和投影操作次数均增加一倍，这可能会显著影响算法的性能，尤其是在大规模问题中。为降低额外步骤带来的查询和计算成本，人们投入了大量精力，通过利用已有信息进行近似，替代其中一次查询和投影操作，以开发 EG 的单次调用变体 [24]–[28]。

在优化领域，零阶（或无导数）方法已得到广泛研究。这类方法通过老虎机预言机返回的扰动函数值来近似缺失的梯度信息 [29]。单点方法 [30]利用一次预言机查询获取梯度估计。尽管方差较大，但这使其在实现上更具吸引力。另一方面，多点方法通过两次

本研究得到美国国家科学基金会资助，资助编号为 2014816 和 2038410。作者任职于美国印第安纳州西拉法叶市普渡大学埃尔莫尔家族电气与计算机工程学院，邮编 47907（电子邮箱：huan1282@purdue.edu；jianghai@purdue.edu）。

本文已提交至 IEEE，供其考虑发表。 版权可能会在不另行通知的情况下转让，届时可能无法再访问此版本。

次或更多次查询来计算梯度估计，从而控制方差，但也使实现更加复杂，尤其是在时变环境中。为兼得两者的优点，Zhang 等人 [31]提出了一种残差反馈方案来控制估计方差。该方案每次迭代仅使用一次查询，性能却与两点零阶方法相当。受极值寻优控制思想的启发，Chen 等人 [32]提出了一种新颖的高通/低通滤波器单点方法，进一步改善了收敛速率对问题维度的依赖关系。

关于带有赌博机反馈的博弈学习文献，Bravo 等人 [33] 通过同时扰动随机逼近方法设计了一个单点赌博机学习过程。已证明，所提出的算法在满足对角严格凹性的博弈中几乎必然收敛；当博弈为强单调时，其渐近收敛速率为 $O(t^{-1/3})$ 。Tatarenko 等人 [34]通过 Tikhonov 正则化将研究范围扩展到仅单调情形，并开发了一种在单时间尺度下运行的无遗憾单点学习算法；为确保收敛，必须适当调整四个递减序列。在一项较新的研究中， [35], Tatarenko 等人提出了单点和双点赌博机学习方法，并证明当博弈为强单调时，这两种方法的收敛速率分别为 $O(t^{-1/2})$和 $O(t^{-1})$。此外，Lin 等人 [36] 致力于改进多玩家博弈的解，以实现最优遗憾和更快的收敛速率。他们开发了一种基于障碍函数的赌博机学习算法族的镜像下降变体，并证明该变体具有无遗憾性，且以 $O(t^{-1/2})$的速率收敛。然而，这项工作的适用范围仅限于强单调博弈。

我们的贡献：鉴于上述讨论，本文旨在探讨能否结合单调用外梯度法与零阶优化残差反馈的优势，从而提升博弈中的单点赌博机学习性能。为此，我们分别将残差反馈的思想融入乐观镜像下降和反射梯度下降，提出了两种算法。首先，我们通过证明在多玩家博弈领域中，残差估计是伪梯度某个平滑版本的无偏估计，补充了 [31] 中的现有结果。随后，我们证明，将残差估计应用于所提出的算法时，其方差具有一致的常数上界。其次，我们证明，若博弈满足伪单调加假设，则所提出算法生成的实际博弈序列几乎必然收敛到该博弈的一个临界点。与 [33]，相比，我们的收敛结果是在更宽松的正则性假设下得到的。此外，我们分析了所提出算法在单调博弈中生成的遍历序列，并证明其以 $O(t^{-(1/2-\epsilon)})$ 的速率收敛，其中正常数 ϵ 可通过适当调整步长和探索半径而任意减小。最后，我们研究了所提出算法在强伪单调博弈中的收敛速度，并证明实际博弈序列能够以 $O(t^{-(1-\epsilon)})$ 的速率收敛到临界点，这

Ref.	正则性 收敛	在.....下的收敛速率 强单调性
[33]	严格单调	$O(t^{-1/3})$
[34]	仅单调	N/A
[35]	强单调	单次查询 $O(t^{-1/2})$ / 两次查询 $O(t^{-1})$
[36]	强单调	$O(t^{-1/2})$
算法 1 和 2	伪单调 plus	$O(t^{-(1-\epsilon)})$

表 I：多玩家博弈的赌博机学习算法：单调性假设与最后迭代收敛率

与现有方法相比，显著加快了学习过程。此外，我们通过 V-C 节中的多建筑热控制问题，对这些求解算法的性能进行了实证比较，结果表明，所提出的算法收敛更快，且估计方差更小。

组织结构： 第二节将正式建立本文研究的多玩家博弈模型，并介绍解概念、一些基本定义及假设。此外，我们还将简要介绍镜像映射及其相关概念，并说明在本文中使用镜像映射的动机。第三节将介绍残差反馈方案，并据此提出两种单点老虎机算法。我们将分析伪梯度估计误差的偏差和方差，这些分析将为后续证明提供重要工具。随后，第四节将陈述并证明所提出算法在不同正则性假设和度量下的三个主要收敛结果。第五节将通过三个数值示例展示理论发现以及所提出算法在实际应用中的有效性：一个投资组合优化问题、一个线性模型中的参数学习问题，以及一个建筑物最优热管理问题。第六节总结全文，并展望潜在的扩展与应用。

基本记号：对于矩阵集合 $\{V_i\}_{i\in S}$，我们用 $\mathrm{blkd}(V_1,\dots,V_{|S|})$ 或 $\mathrm{blkd}(V_i)_{i\in S}$ 表示这些矩阵的对角拼接，用 $[V_1,\dots,V_{|S|}]$ 表示它们的水平堆叠，并用$[V_1;\cdots;V_{|S|}]$ 表示它们的垂直堆叠。对于向量集合 $\{v_i\}_{i\in S}$，$[v_i]_{i\in S}$或 $[v_1;\cdots;v_{|S|}]$ 表示它们的垂直堆叠。对于向量v 和正整数 i， $[v]_i$ 表示 v的第 i个分量。记 $\mathbb{R}_+ := [0,+\infty)$、$\mathbb{R}_{++} := (0,+\infty)$和 $\mathbb{N}_+ := \mathbb{N}\backslash\{0\}$。我们用$\|\cdot\|_2$ 表示欧几里得范数，用 $\|\cdot\|$ 表示一般范数，并用 $\|\cdot\|_*$表示其对偶范数。对于集合 S，令 1_S 表示该集合的示性函数，即若 $x\in S$，则 $1_S(x)=1$，否则为 0。记号$N_S(x)$表示集合 S 在点 x处的法锥 $\subseteq\mathbb{R}^n$：若 $x\in S$，则 $N_S(x) := \{u\in\mathbb{R}^n \mid \sup_{z\in S}\langle u, z-x\rangle \le 0\}$；否则，$N_S(x) := \varnothing$。令 cl(S) 表示集合 S 的闭包，int(S) 表示其内部， ∂S 表示其边界。若 $S\in\mathbb{R}^n$ 是闭凸集，则映射 $\mathrm{Proj}_S:\mathbb{R}^n\to S$ 表示到 S 上的投影，即 $\mathrm{Proj}_S(x) := \mathrm{argmin}_{v\in S}\|v-x\|_2$。

II. 预备知识

A. 问题表述

本文中，我们考虑一个由有限参与者集合 $\mathrm{N} := \{1, \ldots, N\}$ 构成的 N 人博弈 G。每个博弈参与者 i 从其个体策略空间 $\mathrm{X}^i \subseteq \mathbb{R}^{n^i}$ 中选择行动 x^i。对所有参与者进行汇总后，我们用 $\mathrm{X} := \prod_{i\in\mathrm{N}} \mathrm{X}^i \subseteq \mathbb{R}^n$ 和 $x = [x^i]_{i\in\mathcal{N}} \in \mathrm{X}^i$ 分别表示整个博弈的策略空间和行动组合，其中 $n := \sum_{i\in\mathrm{N}} n^i$。为简化后续记号，我们将其他参与者的行动堆叠记作 $x^{-i} := [x^j]_{j\in\mathrm{N}_{-i}} \in \mathrm{X}^{-i} \subseteq \mathbb{R}^{n^{-i}}$，其中 $\mathrm{N}^{-i} := \mathrm{N}\backslash\{i\}$，$\mathrm{X}^{-i} := \prod_{j\in\mathrm{N}_{-i}} \mathrm{X}^j$，且 $n^{-i} := \sum_{j\in\mathrm{N}_{-i}} n^j$。此外，无论参与者的索引如何，我们均记作 $x := [x^i; x^{-i}]$。给定一个特定行动组合 $x \in \mathrm{X}$，每个博弈参与者 i 都会获得一个目标值 $J^i(x^i; x^{-i})$，该目标值决定博弈参与者 i 对不同策略的偏好。从每个利己博弈参与者 i 的角度来看，其目标是求解以下局部优化问题：

$$\text{minimize}_{x^i\in\mathcal{X}^i}\ J^i(x^i; x^{-i}). \tag{1}$$

尽管局部优化问题受限于可行集 $\mathcal{X}^i$，但假设每个博弈参与者 i 均可从行动空间 $\mathcal{X}_a^i$ 中选择行动，并获得相应的目标函数值；其中，行动空间 $\mathcal{X}_a^i$ 可以略大于策略空间 $\mathcal{X}^i$，即 $\mathcal{X}^i \subseteq \mathcal{X}_a^i$。将各个个体行动空间组合起来，得到群体行动空间 $\mathcal{X}_a := \prod_{i\in\mathcal{N}} \mathcal{X}_a^i$。对于目标函数 J^i、策略空间 $\mathcal{X}^i$ 和行动空间 $\mathcal{X}_a^i$ 的正则性，我们作如下总体假设，这些假设在零阶学习或优化文献中十分常见。

假设1. 对于每个博弈参与者 i，局部目标函数 J^i 在行动空间 $\mathcal{X}_a$ 上关于 x 连续可微。此外，其个体策略空间 $\mathcal{X}^i$ 是紧致凸集，而行动空间 $\mathcal{X}_a^i$ 是紧致集且具有非空内部。

B. 博弈正则化与解概念

为便于后续讨论正则性假设，我们引入所谓的伪梯度算子 $F : \mathcal{X}_a \to \mathbb{R}^n$，它刻画了 $\mathcal{G}$ 的一阶信息，并定义为偏梯度的直积：

$$F(x) := \prod_{i\in\mathcal{N}} [\nabla_{x^i} J^i(x^i; x^{-i})]. \tag{2}$$

首先，我们针对 F 作出以下利普希茨性质假设，以应对零阶设定中缺乏一阶信息的问题。

假设2。 伪梯度 F 在 $\mathcal{X}_a$ 上利普希茨连续，常数为 L；即对于任意 $x,\ x' \in \mathcal{X}_a$，有

$$\|F(x) - F(x')\|_* \le L\|x - x'\|. \tag{3}$$

此外，对于每个 i，算子 $\nabla_{x^i} J^i : \mathcal{X}_a \to \mathbb{R}^{n_i}$ 的利普希茨常数 L^i 更小。

纳什均衡（NE）是非合作博弈中的一种标准解概念，其定义为一个决策剖面能够抵御任意单边偏离。形式上，决策剖面 $x_* = [x_*^i; x_*^{-i}] \in \mathcal{X}$ 是博弈 $\mathcal{G}$ 的纳什均衡，当且仅当

$$J^i(x_*^i; x_*^{-i}) \le J^i(x^i; x_*^{-i}),\ \forall x^i \in \mathcal{X}^i,\ \text{for all } i \in \mathcal{N}. \tag{4}$$

接下来，我们关注一种比纳什均衡更宽松的解概念，称为临界点（CP） [37,第2节。2],其定义通过（Stampacchia）变分不等式（VI）以及上文定义的伪梯度 F 来表述。

定义 1.（临界点）若决策剖面 $x_* \in \mathcal{X}$ 是非合作博弈 $\mathcal{G}$ 相关变分不等式（*VI*）的解，即

$$\langle F(x_*), x - x_*\rangle \ge 0,\ \forall x \in \mathcal{X}. \tag{5}$$

我们作出一个总体假设：本文讨论的博弈在 $\mathcal{X}$ 内至少存在一个临界点。一个众所周知的结果是，当 J^i 在 x^i 上对所有 i [38,Sec. 1.4.2]均为凸且连续可微时，CP 与纳什均衡重合。令 $\mathcal{X}_*$ 表示博弈 $\mathcal{G}$ 的临界点集。文献中另一种常用的变分形式概念，可用于衡量候选解 $x_\star = [x_\star^i; x_\star^{-i}] \in \mathcal{X}$ 的误差，称为以下优 merit函数：

$$\mathrm{Err}_{\mathcal{X}}(x_\star) := \max_{x\in\mathcal{X}} \langle F(x), x_\star - x\rangle. \tag{6}$$

大量关于连续博弈的文献研究了伪梯度（强）单调的博弈，这类博弈称为（强）单调博弈 [39]。请注意，对于单调博弈，对所有 $x_* \in \mathcal{X}$，$\mathrm{Err}_{\mathcal{X}}(x_*) \ge 0$；并且 $\mathrm{Err}_{\mathcal{X}}(x_*) = 0$ 当且仅当 x_* 是所研究变分不等式的解。本文将转向不同的研究范围，讨论若干类不同的博弈 [38, Def. 2.3.9], ，其正则性如下所述。

定义2。 博弈 $\mathcal{G}$ 是

(i) 若对于任意行动剖面 $x, y \in X$，$\langle F(y), x - y\rangle \ge 0 \implies \langle F(x), x - y\rangle \ge 0$，则称其为伪单调；*(ii)* 若其为伪单调，且对于任意行动剖面 $x, y \in X$，$\langle F(y), x - y\rangle \ge 0$ *and* $\langle F(x), x - y\rangle = 0 \implies F(x) = F(y)$，则称其为伪单调加性；*(iii)* 若对于任意行动剖面 $x, y \in X$，$\langle F(y), x - y\rangle \ge 0 \implies \langle F(x), x - y\rangle \ge 0$，且当且仅当 $x = y$ 时等号成立，则称其为严格伪单调；*(iv)* 若对于任意行动剖面 $x, y \in X$，$\langle F(y), x - y\rangle \ge 0 \implies \langle F(x), x - y\rangle \ge \mu\|x - y\|^2$，则称其为 μ-强伪单调；*(v)* 若对于任意临界点 $x_* \in X_*$，$\langle F(x), x - x_*\rangle > 0,\ \forall x \in X\backslash X_*$，其中 X_* 为临界点集，则称其为严格相干；*(vi)* 若 F 是伪凸函数 Φ 的梯度，即 $F = \nabla\Phi$，且对于任意 $x, y \in X$，$\langle \nabla\Phi(x), y - x\rangle \ge 0 \implies \Phi(y) \ge \Phi(x)$，则称其具有伪凸势函数 Φ。

关于 (vi)，简要说明如下：当且仅当 $\nabla\Phi$ 在 $\mathcal{X}$ 上伪单调时，可微势函数 Φ 在 $\mathcal{X}$ 上伪凸。 [40,定理 3.1]。

C. 巴拿赫空间之间的镜像映射

为使本文行文更加流畅，本小节将讨论我们借助镜像映射的动机，并介绍相关的基本概念。在面对

优化问题时，如果其目标函数和相关约束集在欧几里得空间中具有良好的性质，我们便可以采用投影梯度下降法，而l_2 范数通常可作为高效的距离度量。然而，在更一般的情形下，更一般的巴拿赫空间$\mathcal{B}$可能更适合作为环境空间；其相关的范数度量并非由内积导出。例如，对于混合策略博弈，每个参与者的局部可行集由概率单纯形描述，此时通常使用l_1 范数。与对偶空间和自身等距的希尔伯特空间不同，直接在（原）巴拿赫空间$\mathcal{B}$ 中实施梯度下降已不再有意义，因为目标函数的梯度属于其对偶$\mathcal{B}^*$。幸运的是，镜像下降和镜像映射的思想最早由 [41],提出，它们可以通过将$\mathcal{B}$ 中的点映射到$\mathcal{B}^*$，在$\mathcal{B}^*$中执行梯度更新，再将更新后的点映射回$\mathcal{B}$，来解决这一不一致问题。关于更详细的讨论和示例，感兴趣的读者可参阅 [42, 第 4] [43]章。

我们首先介绍本文中距离生成函数（DGF）的定义。稍微滥用一下记号，如果函数 $\psi: \mathrm{dom}\,\psi \to \mathbb{R}$，且 $\mathrm{dom}\,\psi \subseteq \mathbb{R}^n$，则称其为环境空间 $\mathrm{B} = (\mathbb{R}^n, \|\cdot\|)$ 中的距离生成函数（DGF），当且仅当其满足以下三个性质：(i) ψ 可微，并且对于某个常数 $\tilde{\mu} > 0$，$\tilde{\mu}$-强凸；(ii) $\nabla\psi(\mathrm{dom}\,\psi) = \mathbb{R}^n$；(iii) $\mathrm{cl}(\mathrm{dom}\,\psi) \supseteq \mathrm{X}$ 且 $\lim_{x\to\partial(\mathrm{dom}\,\psi)} \|\nabla\psi(x)\|_* = +\infty$，即其梯度在 $\mathrm{dom}\,\psi$ 的边界上发散。有了 DGF 后，就可以将从 B^* 到 B 的镜像映射$\nabla\psi^*$ 定义为：

$$\nabla\psi^*(z) = \mathrm{argmax}_{x\in\mathcal{X}}\{\langle z, x\rangle - \psi(x)\}, \tag{7}$$

其中，$\mathcal{X}$ 表示原始可行集；ψ^* 从 $\mathcal{B}^*$ 到 $\mathbb{R}$ 是 ψ的凸共轭，即 $\psi^*(z) = \max_{x\in\mathcal{X}}\{\langle z, x\rangle - \psi(x)\}$；表达式 $\nabla\psi^*$ 表示 ψ^*的次梯度；而（7）由 Danskin 定理推出。由 ψ诱导的伪距离称为布雷格曼散度：

$$D(p,x) = \psi(p) - \psi(x) - \langle\nabla\psi(x), p - x\rangle, \forall p, x \in \mathrm{dom}\,\psi. \tag{8}$$

由于 ψ 的 $\tilde{\mu}$-强凸性，布雷格曼散度可下界为 $D(p,x) \geq \tilde{\mu}/2\|p - x\|^2$。与采用$l_2$ 范数的投影梯度下降类似，近端映射 $P_{x,\mathcal{X}}: \mathcal{B}^* \to$定义域 $\psi \cap \mathcal{X}$由某个固定的 $x \in$定义域 $\psi \cap \mathcal{X}$诱导，并通过布雷格曼散度定义为：

$$P_{x,\mathcal{X}}(y) = \mathrm{argmin}_{x'\in\mathcal{X}}\{\langle y, x - x'\rangle + D(x', x)\}. \tag{9}$$

关于镜像映射和近端映射的性质，请参阅引理 A.1。注意，对于 Lipschitz 连续凸函数 f，可通过以下镜像下降迭代将其最小化：$x_{k+1} \in \nabla\psi^*(\nabla\psi(x_k) - \gamma_k\nabla f(x_k)) = P_{x_k,\mathcal{X}}(-\gamma_k\nabla f(x_k))$，其中 $(\gamma_k)_{k\in\mathbb{N}}$ 是一列适当的步长 [42, 第 4 节2] [43]。

虽然根据$D(p,x) \geq \tilde{\mu}/2\|p - x\|^2$这一事实，$D(p,x_k) \to 0$蕴含 $x_k \to p$，但从上述构造并不能自然地得出反向结论，因此 $D(p,\cdot)$的等值集不足以指示 p [44]的邻域。为使后续论述成立，现对所选的距离生成函数及相应的布雷格曼散度作出以下温和假设。

假设 3。（布雷格曼互反性）所选的距离生成函数 ψ 满足：当 $x_k \to p$时，有 $D(p,x_k) \to 0$。

III. 两种单次调用额外梯度算法及其相应的零阶变体

A. 额外梯度算法族

为获得式（5）所示非合作博弈中的临界点，我们考虑广义额外梯度算法族。在该算法族中，每次迭代 k 时，算法都会持续更新两个状态，即基状态 X_k 和主状态 $X_{k+1/2}$。给定一个步长序列 $(\gamma_k)_{k\in\mathbb{N}}$ 以及两个序列 $(\tilde{F}_k)_{k\in\mathbb{N}}$ 和 $(\tilde{F}_{k+1/2})_{k\in\mathbb{N}}$，它们分别与基状态和主状态的一阶／伪梯度信息相关，额外梯度方案包含以下两个步骤：

$$X_{k+1/2} = P_{X_k,\mathcal{X}_1}(-\gamma_k\tilde{F}_k),\ X_{k+1} = P_{X_k,\mathcal{X}_2}(-\gamma_k\tilde{F}_{k+1/2}). \tag{10}$$

为说明一阶序列以及集合$\mathcal{X}_1$ 和 $\mathcal{X}_2$, 的一些可行选取，我们给出文献中的以下几个具体示例。令 ξ_k 和 $\xi_{k+1/2}$ 表示一些二阶矩有界的无偏随机噪声。

例 *1.*（随机镜像近端法（*SMP*） *[20]*）考虑一般的环境空间（$\mathbb{R}^n, \|\cdot\|$）。令 $\tilde{F}_k = F(X_k) + \xi_k$,$\tilde{F}_{k+1/2} = F(X_{k+1/2}) + \xi_{k+1/2}$，且 $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$，则得到随机镜像近端算法：$X_{k+1/2} = P_{X_k,\mathcal{X}}(-\gamma_k(F(X_k) + \xi_k))$，$X_{k+1} = P_{X_k,\mathcal{X}}(-\gamma_k(F(X_{k+1/2}) + \xi_{k+1/2}))$。

例 *2.*（乐观镜像下降（*OMD*） *[25], [26]*）给定一般的环境空间（$\mathbb{R}^n, \|\cdot\|$），分别令 $\tilde{F}_k = F(X_{k-1/2}) + \xi_{k-1/2}$, $\tilde{F}_{k+1/2} = F(X_{k+1/2}) + \xi_{k+1/2}$, $\mathcal{X}_1 = \mathbb{R}^n$，以及 $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$。于是得到以下更新：$X_{k+1/2} = P_{X_k,\mathcal{X}}(-\gamma_k(F(X_{k-1/2}) + \xi_{k-1/2}))$, $X_{k+1} = P_{X_k,\mathcal{X}}(-\gamma_k(F(X_{k+1/2}) + \xi_{k+1/2}))$。

例3. *(*反射梯度下降 *(RGD) [27], [28])* 考虑欧几里得环境空间 $(\mathbb{R}^n, \|\cdot\|_2)$。令$\tilde{F}_k = 1/\gamma_k(X_{k-1} - X_k)$，$\tilde{F}_{k+1/2} = F(X_{k+1/2}) + \xi_{k+1/2}$、$\mathcal{X}_1 = \mathbb{R}^n$ 和$\mathcal{X}_2 = \mathcal{X}$，我们可以将(10) 具体化为如下显式迭代：$X_{k+1/2} = 2X_k - X_{k-1}$，$X_{k+1} = \mathrm{Proj}_{\mathcal{X}}[X_k - \gamma_k(F(X_{k+1/2}) + \xi_{k+1/2})]$。

其确定性对应方法最早在 [18],中提出，SMP 是研究最广泛的外梯度方法之一。然而，它每次迭代都需要向一阶预言机查询两次，并执行两次近端映射，因此实现成本较高。RGD 利用 X_{k-1} 在 X_k 中的反射，而不是在 X_k 处计算梯度，因此每次迭代只需进行一次欧几里得投影和一次预言机调用。不过，要将这一策略扩展到欧几里得框架之外，则需要对距离生成函数施加额外的正则性假设。OMD 的实现涉及两次近端映射，但只需一次预言机调用，因为它利用上一次迭代的一阶信息更新主状态。在接下来两个小节的开篇，我们指出，本文的零阶学习算法以 RGD 和 OMD 为基础，并在每次迭代中利用主状态 $X_{k+1/2}$ 处查询得到的伪梯度，来估计 k 。

B. 残差伪梯度估计

查询点 $X_{k+1/2}$处的一阶信息，即$F(X_{k+1/2})$，对于 OMD 和 RGD 求得临界点至关重要。然而，这类信息往往无法获得，因此每个博弈参与者 i 都需要根据观测到的目标值，估计自身的偏梯度信息$\nabla_{x^i} J^i(X_{k+1/2})$。为实现赌博机学习，我们让每个博弈参与者在$n^i$维空间的单位球面上随机采样一个查询方向 u_k^i ，并在第 k次迭代时使用查询半径 δ_k 。令 $u_k \coloneqq [u_k^i]_{i\in\mathcal{N}}$。对于博弈参与者 i，第 k次迭代的偏梯度估计候选项定义为

$$\frac{n^i}{\delta_k}\Big(J^i(X_{k+1/2}+\delta_k u_k) - J^i(X_{k-1/2}+\delta_{k-1}u_{k-1})\Big)u_k^i. \tag{11}$$

将（11）式在这组参与者上进行堆叠，可得到所谓的残差伪梯度估计。此外，为确保扰动查询点后仍满足局部可行性，我们假设每个$\mathcal{X}_a^i$ 内都存在一个以 p_i 为中心、半径为 r_i的闭球，并采用以下可行性调整步骤：

$$\begin{aligned}\hat{X}_{k+1/2}^i &= X_{k+1/2}^i + \delta_k(u_k^i - \frac{1}{r^i}(X_{k+1/2}^i - p^i))\\ &= (1-\frac{\delta_k}{r^i})X_{k+1/2}^i + \frac{\delta_k}{r^i}(p^i + r^i u_k^i) = \bar{X}_{k+1/2}^i + \delta_k u_k^i,\end{aligned}$$

其中 $\bar{X}_{k+1/2}^i = (1-\frac{\delta_k}{r^i})X_{k+1/2}^i + \frac{\delta_k}{r^i}p^i$，并令 $\delta_k < r^i$ 对所有 k成立，以确保上述表达式是$\mathcal{X}_a^i$内两点的可行凸组合。将（11）式中的扰动动作替换为调整后的版本，可得以下估计

$$G_k^i \coloneqq \frac{n^i}{\delta_k}\Big(J^i(\hat{X}_{k+1/2}) - J^i(\hat{X}_{k-1/2})\Big)u_k^i, \tag{RPG}$$

其中，对每个 $k\in\mathbb{N}_+$，$\hat{}X_{k+1/2} \coloneqq [\hat{}X_{k+1/2}^i]_{i\in\mathcal{N}}$。为后续刻画残差伪梯度估计（RPG），我们构造如下滤过：对每个 $k\in\mathbb{N}_+$，令$\mathcal{F}_k \coloneqq \sigma\{X_0, u_1, \dots, u_{k-1}\}$。注意，虽然 $X_{k+1/2}$, $\bar{X}_{k+1/2}\in\mathcal{F}_k$，但$\hat{}X_{k+1/2}\notin\mathcal{F}_k$。零阶学习文献中广泛使用的一种函数是 δ-平滑目标函数，其定义如下：

$$\tilde{J}_\delta^i(x^i; x^{-i}) \coloneqq \frac{1}{\mathbb{V}^i}\int_{\delta\mathbb{S}_{-i}}\int_{\delta\mathbb{B}_i} J^i(x^i+\tilde{\tau}^i; x^{-i}+\tau^{-i})d\tilde{\tau}^i d\tau^{-i}, \tag{12}$$

其中 $\mathbb{S}_{-i} \coloneqq \prod_{j\in\mathcal{N}^{-i}}\mathbb{S}_j \subseteq \mathbb{R}^{n^{-i}}$ 中的每个 $\mathbb{S}_j$ 均表示以 $\mathbb{R}^{n_j}$ 内原点为中心的单位球面； $\mathbb{B}_i$ 表示 $\mathbb{R}^{n_i}$ 内以原点为中心的单位球； $\mathbb{V}^i \coloneqq \mathrm{vol}(\delta\mathbb{B}_i)\cdot\mathrm{vol}(\delta\mathbb{S}_{-i})$ 为体积常数。注意，$\tilde{J}_\delta^i(x^i; x^{-i})$可解释为局部目标函数 J^i 在区域 $\delta\mathbb{B}_i\times\delta\mathbb{S}_{-i}$ 内的平均值。形式化地，我们有如下引理来建立（RPG）与函数 $\tilde{J}_\delta^i$ 之间的关系。

引理1. 假设假设*1*成立。那么在每次迭代 $k\in\mathbb{N}_+$ 中，$\nabla_{x^i}\tilde{J}_{\delta_k}^i(\bar{X}_{k+1/2})$ 是条件期望 $\mathbb{E}[G_k^i \mid \mathcal{F}_k]$ 的一个版本，即，

$$\nabla_{x^i}\tilde{J}_{\delta_k}^i(\bar{X}_{k+1/2}) = \mathbb{E}[G_k^i \mid \mathcal{F}_k] \ a.s., \ \forall i\in\mathcal{N}. \tag{13}$$

证明. 见附录 B。 □

换言之，尽管很容易观察到， G_k^i 是点 $\bar{X}_{k+1/2}$ 处真实伪梯度的有偏估计，但在给定 $\mathcal{F}_k$ 的条件下， G_k^i 是 $\bar{X}_{k+1/2}$ 处 δ 平滑函数 $\tilde{J}_{\delta_k}^i$ 的无偏估计。基于上述观察，估计（RPG）引入的误差可分解为系统误差 B_k^i 和随机误差 V_k^i ，即：

$$\begin{aligned}G_k^i &= \nabla_{x^i}J^i(X_{k+1/2}) +\\ &\underbrace{\big(G_k^i - \nabla_{x^i}\tilde{J}_{\delta_k}^i(\bar{X}_{k+1/2})\big)}_{V_k^i} + \underbrace{\big(\nabla_{x^i}\tilde{J}_{\delta_k}^i(\bar{X}_{k+1/2}) - \nabla_{x^i}J^i(X_{k+1/2})\big)}_{B_k^i}.\end{aligned} \tag{14}$$

回顾滤过$(\mathcal{F}_k)_{k\in\mathbb{N}_+}$的构造，系统误差 B_k^i 是$\mathcal{F}_k$-可测的，而随机误差$V_k^i \in \mathcal{F}_{k+1}$ 不是$\mathcal{F}_k$-可测的。令 $G_k \coloneqq [G_k^i]_{i\in\mathcal{N}}$，对于 B_k、 V_k及其他变量也作类似定义。随机变量 B_k 和 V_k 的陪域由以下引理刻画。

引理 2. 假设假设*1*成立。那么对于每次迭代 k，都有 $\|B_k\|_* \le \alpha_B\delta_k$ ，其中常数 α_B 取决于利普希茨常数 $\tilde{L}$ 以及$\mathcal{X}_a$的几何性质。此外，$\|V_k\|_* \le \alpha_V/(\delta_k)^2$，其中常数 α_V取决于目标函数 J^i's以及$\mathcal{X}_a$的几何性质。

Proof. See Appendix B. □

误差 B_k^i 和V_k^i 的陪域具有此类结论，这在单点零阶方法 [33]–[35]中尤为常见。作为铺垫，在下一小节中，我们将说明，通过采用恰当的步长和查询半径序列，可以界定$\|V_k\|_*^2$ 的陪域；这是RPG的一项引人注目的特性所带来的结果。

C. 采用单次调用外梯度方法的零阶学习

现在，我们将 RPG 融入外梯度方法 OMD 和 RGD 中，提出两种零阶学习算法。在介绍算法细节之前，先引入一些记号：设每个博弈参与者 i 都配备一个局部距离生成函数 ψ^i；对于这组参与者，我们另外记$\psi(x) \coloneqq \sum_{i\in\mathcal{N}}\psi^i(x^i)$；$\hat{}J_k^i$ 表示博弈参与者 i 在第 k 次迭代时实现的目标函数值。

1）乐观镜像下降法：将例 2 中的 OMD 学习方案与 RPG 机制相结合，我们可以直接得到每次迭代所采用的对偶向量 k，即分别为 $\tilde{F}_{k+1/2} = [G_k^i]_{i\in\mathcal{N}}$ 和 $\tilde{F}_k = [G_{k-1}^i]_{i\in\mathcal{N}}$。将 OMD 更新按博弈参与者分别进行，可得到算法 1 中相应的基于收益的学习算法。注意，在算法 1 中，行动空间 $\mathcal{X}_a$ 与策略空间 $\mathcal{X}$ 一致。

2）反射镜像下降：将例3中用于 RGD 的欧几里得设置推广后，我们考虑在一般范数空间 $(\mathbb{R}^n, \|\cdot\|)$中应用反射策略。因此，原空间中的向量差$1/\gamma_k(X_{k-1} - X_k)$ 不再适用，应将其替换为对偶空间中的某种反射代理。一种可能的候选形式是$\tilde{F}_k = 1/\gamma_k(\nabla\psi(X_{k-1}) - \nabla\psi(X_k))$，而主状态按 $X_{k+1/2} = P_{X_k,\mathbb{R}^n}(-\gamma_k\tilde{F}_k)$更新。近端映射 $P_{X_k,\mathbb{R}^n}$ 可化为目标函数强凸的无约束问题，这类问题通常具有闭式解，或者至少可得到无投影解，并具有指数收敛

算法 1： 基于
乐观镜像下降的纳什均衡零阶学习（博弈参与者 i）

初始化： $X_0^i = X_{1/2}^i = X_1^i \in \mathcal{X}^i \cap$ 定义域 ψ^i 任意选择；
$\hat{J}_0^i = J^i(X_{1/2}^i; X_{1/2}^{-i})$；$G_0^i = \mathbf{0}_{n^i}$；$p^i$，$r^i$ 作为中心，并集合 $\mathcal{X}^i$ 内任意一个球的半径；
在第 k 次迭代（$k \in \mathbb{N}_+$）：
  $X_{k+1/2}^i \leftarrow P_{X_k^i, \mathcal{X}^i}(-\gamma_k G_{k-1}^i)$；
  从 $\mathbb{S}^{n^i}$ 随机采样方向 u_k^i；
  $\hat{X}_{k+1/2}^i \leftarrow (1 - \frac{\delta_k}{r^i})X_{k+1/2}^i + \frac{\delta_k}{r^i}(p^i + r^i u_k^i)$；
  采取行动 $\hat{X}_{k+1/2}^i$ 并观察实际得到的目标函数值 $\hat{J}_k^i := J^i(\hat{X}_{k+1/2}^i; \hat{X}_{k+1/2}^{-i})$；
  $G_k^i \leftarrow \frac{n^i}{\delta_k}(\hat{J}_k^i - \hat{J}_{k-1}^i)u_k^i$；
  $X_{k+1}^i \leftarrow P_{X_k^i, \mathcal{X}^i}(-\gamma_k G_k^i)$；
返回： $\{\hat{X}_{k+1/2}^i\}_{i \in \mathcal{N}}$。

算法 2： 基于
反射镜像下降的纳什均衡零阶学习（博弈参与者 i）

初始化： $X_0^i = X_{1/2}^i = X_1^i \in \mathcal{X}^i \cap$ 定义域 ψ^i 任意选取；
$\hat{J}_0^i = J^i(X_{1/2}^i; X_{1/2}^{-i})$；$p^i$，$r^i$ 分别作为反射集 $\mathcal{X}_a^i$ 内任意一个球的中心和半径；
在第 k 次迭代（$k \in \mathbb{N}_+$）：
  $X_{k+1/2}^i \leftarrow P_{X_k^i, \mathbb{R}^{n^i}}(-(\nabla\psi^i(X_{k-1}^i) - \nabla\psi^i(X_k^i)))$；
  从 $\mathbb{S}^{n^i}$ 随机采样方向 u_k^i；
  $\hat{X}_{k+1/2}^i \leftarrow (1 - \frac{\delta_k}{r^i})X_{k+1/2}^i + \frac{\delta_k}{r^i}(p^i + r^i u_k^i)$；
  采取行动 $\hat{X}_{k+1/2}^i$ 并观察实现的目标函数值 $\hat{J}_k^i := J^i(\hat{X}_{k+1/2}^i; \hat{X}_{k+1/2}^{-i})$；
  $G_k^i \leftarrow \frac{n^i}{\delta_k}(\hat{J}_k^i - \hat{J}_{k-1}^i)u_k^i$；
  $X_{k+1}^i \leftarrow P_{X_k^i, \mathcal{X}^i}(-\gamma_k G_k^i)$；
返回： $\{\hat{X}_{k+1/2}^i\}_{i \in \mathcal{N}}$。

速率。为便于后续分析收敛性，我们施加额外的正则性条件，要求群距离生成函数 ψ 为范数型。

假设4。 群距离生成函数 ψ 在 $\mathcal{X}_a$ 上是 $\tilde{L}$-光滑的，即对于 $\mathcal{X}_a$ 中的任意 x_a 和 x_b，

$$\psi(x_a) \le \psi(x_b) + \langle \nabla\psi(x_b), x_a - x_b \rangle + \frac{\tilde{L}}{2}\|x_a - x_b\|^2.$$

等价条件是 $\nabla\psi: \mathcal{X}_a \to \mathbb{R}^n$ 是 $\tilde{L}$-*Lipschitz* 连续的：

$$\langle \nabla\psi(x_a) - \nabla\psi(x_b), x_a - x_b \rangle \le \tilde{L}\|x_a - x_b\|^2.$$

因此，对于每个博弈参与者 $i \in \mathcal{N}$，其距离生成函数 ψ^i 都是 $\tilde{L}^i$-光滑的，常数为 $\tilde{L}^i \le \tilde{L}$。

采用反射镜像下降（RMD）的零阶学习算法总结于算法 2。我们用 X_R 表示 G 的反射空间，即算法 2 中 $X_{k+1/2}$ 的取值范围。下面简要说明 X、X_a 和 X_R 之间的关系。在 RMD 框架中，前导行动 $X_{k+1/2} \in \mathrm{X}_R$ 可能落在策略空间 X 之外，但始终应位于 X_a 内，在此处必要的正则性假设成立。因此，与 OMD 相比，RMD 中的近似策略计算效率更高，但对正则性的要求也更高。前导状态的更新最优性条件表明，$\nabla\psi(X_{k+1/2}) - \nabla\psi(X_k) = \nabla\psi(X_k) - \nabla\psi(X_{k-1})$。根据假设4中假定的 $\tilde{\mu}$-强凸性和 $\tilde{L}$-光滑性，$\tilde{\mu}\|X_{k+1/2} - X_k\| \le \tilde{L}\|X_k - X_{k-1}\|$ 满足 X_k 且 $X_{k-1} \in \mathrm{X}$，这意味着反射集 X_R 有界。此外，对于任意迭代 k，$\|X_{k+1/2} - X_k\| \le L\gamma_{k-1}/\tilde{\mu} \cdot \|G_{k-1}\|$，其中随机变量 $\|G_k\|_*$ 的有界性将在引理 3 中证明，并且通过将 γ_k 取得足够小，X_R 可以任意接近 X。综上，应有 $\mathrm{X} \subseteq \mathrm{X}_R \subseteq \mathrm{X}_a$，且 $\mathrm{X}_R \subseteq \mathrm{cl}(\mathrm{dom}\,\psi)$。

下面的引理表明，通过适当调整步长的衰减速率，使其略快于查询半径的衰减速率，所提出的算法可以使随机误差 V_k 的对偶范数保持有界。

引理 3. 假设假设*1*成立。此外，单调递减的步长序列 $(\gamma_k)_{k \in \mathbb{N}_+}$ 和查询半径序列 $(\delta_k)_{k \in \mathbb{N}_+}$ 满足：$\lim_{k\to\infty}\gamma_k = 0$，$\lim_{k\to\infty}\delta_k = 0$，$\sum_{k\in\mathbb{N}}\gamma_k = \infty$，且 $\lim_{k\to\infty}\gamma_k/\delta_k = 0$。考虑由算法*1*和算法*2*生成的 $RPG(G_k)^{k\in\mathbb{N}_+}$，其中算法*2*在满足假设*4*的条件下执行 ψ。那么，对于每次迭代 k，$\|V_k\|_*^2 \le C_V$，其中某个常数为 C_V。

证明。 参见附录 B。 □

IV. 所提出算法的收敛性质

本节将在若干不同的正则性条件下，确定算法1和算法2在非合作博弈中的临界点收敛性质。分析将围绕渐近收敛性质、遍历收敛率以及实际行动序列的收敛速率展开。

A. 伪单调加情形下所提出算法的几乎必然收敛

我们首先分析具有伪单调加伪梯度的博弈的渐近收敛性质。

定理1。 考虑一个伪单调加博弈 $\mathcal{G}$。假设博弈参与者在 $\mathcal{G}$ 中遵循算法*1*，且假设*1*至*3*成立；或者所有参与者都执行算法*2*，且假设*1*至*4*成立。此外，步长 $(\gamma_k)_{k\in\mathbb{N}_+}$ 和查询半径 $(\delta_k)_{k\in\mathbb{N}_+}$ 单调递减并满足

$$\sum_{k\in\mathbb{N}_+}\gamma_k = \infty,\ \sum_{k\in\mathbb{N}_+}\gamma_k^2 < \infty,\quad \sum_{k\in\mathbb{N}_+}\gamma_k\delta_k < \infty,\ \lim_{k\to\infty}\gamma_k/\delta_k = 0. \tag{15}$$

那么，博弈过程序列 $(\hat{X}_{k+1/2})_{k\in\mathbb{N}_+}$ 几乎必然收敛到某个纳什均衡 x_*。

证明。 参见附录C。 □

受 [21]，中关于 SMP 的讨论启发，我们将定理1中的收敛结果扩展到定义2所给出的伪单调加博弈的若干变体，具体结果将在下述推论中正式陈述。在继续之前，我们简要说明定义2中各类之间的相互关系：(v) 和 (vi) 既不是 (ii)（即伪单调加类）的子类，也不是其超类，

而 (*iii*) 和 (*iv*) 具有更强的正则性，分别是 (*ii*) 和 (*v*)的子类。

推论1。 考虑一个博弈 $\mathcal{G}$ ，其伪梯度 F满足定义*2* (*iii*)-(*vi*) 中的任意一项，其他设置与定理*1*中的相同。则博弈过程序列 $(\hat{X}_{k+1/2})^{k\in\mathbb{N}_+}$ 几乎必然收敛到某个纳什均衡 x_* 。

证明。 见附录C。 □

B. 仅单调情形下遍历平均的收敛速率

上一小节中的收敛性质在本质上是渐近的。为刻画所提出迭代的收敛速度，我们将关注所生成的遍历序列，并借助度量函数（6）量化其误差，这与单调变分不等式的遍历收敛分析思路一致。迭代中实际博弈过程序列的遍历平均定义如下:

$$\check{X}_k \coloneqq \sum_{t=1}^{k} \gamma_t \hat{X}_{t+1/2} / \sum_{t=1}^{k} \gamma_t. \tag{16}$$

定理2。 考虑一个仅单调博弈 $\mathcal{G}$。其他设置与定理*1*中的相同。则遍历平均 $\check{X}_k$ 满足:

$$\mathbb{E}[Err_{\mathcal{X}}(\check{X}_k)] \le \frac{\mathbb{E}[\max_{p\in\mathcal{X}} D(p, X_1)] + M}{\sum_{t=1}^{k} \gamma_t}, \ \forall k \in \mathbb{N}_+ \tag{17}$$

其中， M 是一个常数，取决于$\mathcal{G}$的性质以及 γ_k 和 δ_k的具体取值。

证明。 见附录D。 □

C. 所提出的算法在强伪单调情形下的$(1/k^{1-\epsilon})$ 收敛速率

为研究实际博弈序列的收敛速率，除前述假设外，我们还对F 施加强伪单调性要求。已有证明表明，强伪单调变分不等式存在唯一解 x_*（见 [45,定理 21] ），因此可以利用序列与该唯一解之间的距离来衡量收敛速率。此外，我们将情形限定为布雷格曼散度是范数型的，即假设4成立。在这些条件下，我们得到算法 1 和算法 2 的如下全局收敛结果。

定理 3。 考虑一个强伪单调博弈$\mathcal{G}$。假设假设*1*至*4*成立，且$\mathcal{G}$ 的所有参与者均采用算法 *1* 或算法 *2*，并分别将步长和查询半径选为 $\gamma_k = c_\gamma/(k+b_\gamma)^{a_\gamma}$ 和 $\delta_k = c_\delta/(k+b_\delta)^{a_\delta}$。此外，若 a_γ 和 a_δ 满足 $0 < a_\delta < a_\gamma < 1$和 $a_\gamma + a_\delta > 1$，则博弈行动序列具有如下收敛速率:

$$\mathbb{E}[\|\hat{X}_{k+1/2} - x_*\|^2] \le \frac{M_1}{k^{a_\gamma + a_\delta - 1}} + \frac{M_2}{k}, \forall k > K \tag{18}$$

其中， x_* 是$\mathcal{G}$的唯一纳什均衡； M_1、 M_2和 K 表示一些常数，它们取决于$\mathcal{G}$ 的性质以及 γ_k 和 δ_k的具体取值。

证明。 参见附录 E。 □

定理 3 给出了实现动作的主要显式收敛速率分析结果。从理论上讲，当a_γ 从左侧趋近1 ，且 a_δ 和 a_γ 均从左侧趋近时，所提出的算法便可达到收敛速率$\mathbb{E}[\|\hat{X}_{k+1/2} - x_*\|^2] = O(1/k^{1-\epsilon})$，其中某个 ϵ可以任意接近0。在实际应用中，建议对 a_γ 和 a_δ 采用比上述讨论更为保守的取值，以避免由不理想的 M_1、 M_2和 K导致不良的暂态行为。V-C 节中的数值实验为这一结论提供了经验支持。

V. 案例研究与数值模拟

A. 投资组合优化

在单智能体投资组合优化问题中，智能体选择最优投资组合或资产配置，并旨在最大化其预期收益 [46]。稍微滥用一下符号，假设存在 N 种资产，其收益率记为 $\xi \coloneqq [\xi_1; \cdots; \xi_N]$，且服从正态分布，即 $\xi \sim \mathrm{N}(\mu, \Sigma)$，均值向量为 $\mu \coloneqq [\mu_1; \cdots; \mu_N] \in \mathbb{R}_+^N$，协方差矩阵为 $\Sigma \in \mathbb{S}_{++}^N$。令 $z \coloneqq [z_1; \cdots; z_N]$表示智能体的策略或资产配置，相应地，总收益率服从 $\eta = \xi^T z \sim \mathrm{N}(\mu^T z, z^T \Sigma z)$。给定预期收益率 $r > 0$，智能体试图求解：最大化$_z$ $P(\eta \ge r) = \Phi(\frac{\mu^T z - r}{\sqrt{z^T \Sigma z}})$，约束条件为 $\mathbf{1}_N^T z = 1$，
$\mu^T z - r \ge 0$ $0 \le z \le 1$ Φ a $d\tau$ ，以及 ，其中 $(\ ) \coloneqq \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp(-\frac{\tau^2}{2})$ 表示标准正态分布的分布函数。注意到 Φ 单调递增，并进行简单的坐标变换 φ: $\mathbb{R}^{N-1} \to \mathbb{R}^N$，其中 φ： $x \mapsto [x_1, \cdots, x_{N-1}, 1 - \mathbf{1}_{N-1}^T x]$，即可进一步重构该问题，从而得到一个具有非空内部的可行集。形式化地，重构后的问题可描述如下:

$$\begin{aligned} &\text{minimize}_x \ J(x) = \frac{r - \mu^T \varphi(x)}{\sqrt{\varphi(x)^T \Sigma \varphi(x)}} \\ &\text{subject to } x \in \mathcal{X} \coloneqq \{x \in \mathbb{R}^{N-1} \mid 0 \le x \le 1, \mathbf{1}_{N-1}^T x \le 1, \\ &\qquad [\mu_N - \mu_1; \cdots; \mu_N - \mu_{N-1}]^T x \le \mu_N - r\}. \end{aligned} \tag{19}$$

我们声称，目标函数 $J(x) = J_n(x)/J_d(x)$ 在 X 上是伪凸的，其中分子和分母$J_n(x) = r - \mu^T \varphi(x)$ 以及 $J_d(x) = \sqrt{\varphi(x)^T \Sigma \varphi(x)}$ 均为凸函数。对于满足 $\langle \nabla J(x'), x - x' \rangle \ge 0$ 的任意 $x \in \mathrm{X}$ 和 $x' \in \mathrm{X}$，我们有$\langle {J_d}^{-1}(x') (\nabla J_n(x') - J_n(x') J_d(x') \nabla J_d(x')), x - x' \rangle \ge 0$。由于$J_n(x) - J_n(x') \ge \langle \nabla J_n(x'), x - x' \rangle$， $J_d(x) - J_d(x') \ge \langle \nabla J_d(x'), x - x' \rangle$， J_n 在 X 上始终为非正，而 J_d 始终为正，我们最终推得 $J(x) \ge J(x')$，且 J 在 X 上是伪凸的。

在数值模拟中，智能体有 $N = 6$ 项资产可供投资，其均值 μ 和协方差 Σ 通过随机采样得到，并展示于图 1 (a) 和 (b) 中。预期收益率 r 设为 μ_i 的平均值。我们将步长和查询半径分别设为 $\gamma_k = 1/(k+2\times10^3)^{a_\gamma}$ 和 $\delta_k = 1/(k+2\times10^3)^{a_\delta}$。性能指标包括相对更新距离 $\|X_{k+1} - X_k\|_2$，以及当前值 $J(X_k)$ 与通过 [21] 得到的最优值 J^* 之间的差值。图1中，我们用实线展示这些指标的滚动平均值（窗口大小为100 ），并用半透明曲线展示原始波动。请注意，

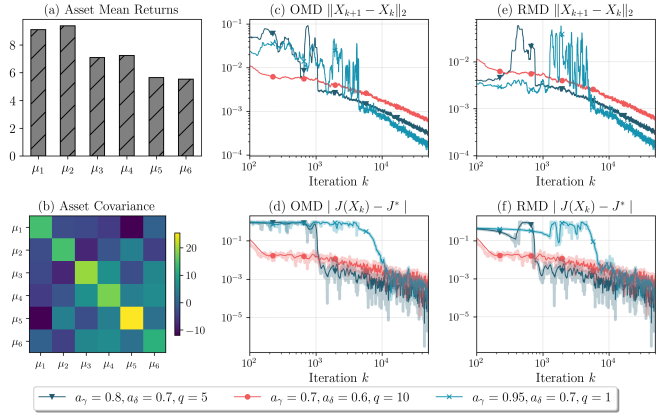

图1：算法1和算法2在投资组合优化中的表现

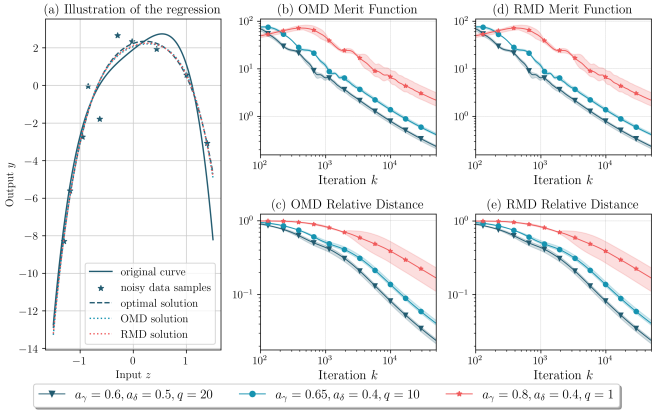

图2：线性模型线性最小二乘中算法1和算法2的性能

在实验中，RPG 采用不同的查询次数计算，即 $q = 1,5,10$；而在前文中，我们重点分析单次查询（$q = 1$）情形。

B. 线性模型中的最小二乘估计

考虑一个包含数据样本$\{(z_j, y_j)\}_{j=1}^M$的数据集，其中 $z_j \in \mathbb{R}^N$ 表示输入向量，$y_j \in \mathbb{R}$ 表示输出标签。此外，y_j和z_j通过线性模型相关联，即$y_j = w_0 + w^T z_j + \xi_j$，其中 $w_0 \in \mathbb{R}$、$w \in \mathbb{R}^N$和 ξ_j表示某种随机噪声。在本例中，我们假设，尽管 w_0 和 w 的确切值未知，需要根据数据样本进行估计，但已知 w_0 以及 w 的每个元素均位于某个有界区间 $[-\bar{w}, \bar{w}]$ 内，其中 $\bar{w} \in \mathbb{R}_{++}$足够大。为简单起见，令每个 $\tilde{z}_j = [1; z_j]$、$\tilde{Z} = [\tilde{z}_1, \ldots \tilde{z}_M]$、$y = [y_1; \cdots; y_M]$，且 $\tilde{w} = [w_0; w]$。掌握这些信息后，优化问题可写为：

$$\underset{-\tilde{w} \le \tilde{w} \le \bar{w}}{\text{minimize}} \ \frac{1}{2}\|\tilde{Z}^T \tilde{w} - y\|_2^2 \tag{20}$$

与 [47,第 VI 节], 类似，我们将上述优化问题重新表述为一个双人零和博弈。引入辅助变量 $\lambda \in \mathbb{R}^M$，从而得到以下等价关系：$\frac{1}{2}\|\tilde{Z}^T \tilde{w} - y\|_2^2 = \max_{\lambda \in \mathbb{R}^M} \lambda^T (\tilde{Z}^T \tilde{w} - y) - \frac{1}{2}\|\lambda\|_2^2 = \max_{\lambda \in \mathbb{R}^M} J(\tilde{w}, \lambda)$。此外，$\tilde{w}$的有界性意味着 λ 有界，我们还手动设定 $-\bar{\lambda} \le \lambda \le \bar{\lambda}$，其中 $\bar{\lambda}$足够大。记局部目标函数 $J^1(x^1; x^2) = J(x^1, x^2)$以及 $J^2(x^2; x^1) = -J(x^1, x^2)$。那么，双人零和博弈可表示如下：

$$\text{Player 1: } \underset{-\tilde{w} \le x^1 \le \tilde{w}}{\text{minimize}} \ J^1(x^1; x^2), \quad \text{Player 2: } \underset{-\bar{\lambda} \le x^2 \le \bar{\lambda}}{\text{minimize}} \ J^2(x^2; x^1).$$

相应的伪梯度为

$$F\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \mapsto : \ , \qquad M_{\text{lin}}\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} 0 \\ y \end{pmatrix} \qquad M_{\text{lin}} := \begin{pmatrix} 0 & \tilde{Z} \\ -\tilde{Z}^T & I \end{pmatrix}$$

其中矩阵为 。 F的单调性由 M_{lin}的半正定性直接推出。为更直观地展示该问题，我们考虑多项式回归情形，其中输入向量 z_j的第 m个分量是 $[z_j]_1$的 m次幂。我们选取 $N = 5$、$M = 10$、$\bar{w} = \bar{\lambda} = 5$，分别从紧区间 $[-1。5, \ 1。5]$ 和 $[-2, 2]$ 中随机采样每个 $[z_j]_1$ 和 ξ_j，并通过数值计算确认 M_{lin} 满秩，因此 F 的临界点 x^* 在内部唯一。

设置步长 $\gamma_k = 1/(k + 10^4)^{a_\gamma}$，查询半径设为 $\delta_k = 1/(k + 10^2)^{a_\delta}$。OMD 和 RMD 的实验使用相同的随机样本路径。在图2(a)中，我们绘制了待拟合的原始曲线、所用的含噪数据样本，以及通过 OMD 和 RMD 得到的两个遍历解。用于衡量性能的一个指标是优 merit函数 (6)，结果在图2(b)和(d)中以三组不同参数展示。我们用实线表示在相同参数选择下三次不同运行的指标平均值，并用半透明包络表示实际波动。到唯一最优解 x_* 的相对距离，即 $\|\bar{X}_k - x_*\|_2/\|x_*\|_2$，报告于图2(c)和(e)。仿真中我们观察到，图2(b)和(d)中的收敛速率与定理2的结果相符。如果参与者每次迭代只有一次观测，则应适当提高步长的衰减速率，并降低查询半径的衰减速率，以应对估计方差。此外，我们注意到，当 OMD 和 RMD 仅在前瞻更新步骤上不同时，两者的遍历结果也彼此相似。

C. 建筑物热控制

在此示例中，我们考虑一个由 N栋建筑物组成的负荷聚合商，记为$\mathcal{N} := \{1, \ldots, N\}$。给定 [48],中定义的内部定价机制，我们希望考察所提出的赌博机在线纳什均衡学习算法在强伪单调问题中的收敛速率。给定时间范围$\mathcal{T} := \{1, \ldots, T\}$，令 x_t^i表示建筑物 i 在时段 $t \in \mathcal{T}$的功耗，$x^i := [x_t^i]_{t \in \mathcal{T}}$表示建筑物 i 在所有时段的功率曲线，$x := [x^i]_{i \in \mathcal{N}}$ 表示所有参与者的能源曲线的拼接。

假设聚合商以 $p_e \in \mathbb{R}_{++}^T$ 的能源价格从批发能源市场购电。此外，还存在费率为 $p_d \in \mathbb{R}_{++}$ 的任意时段需量电费，用于惩罚聚合商在所考虑时间范围内的峰值用电量。峰值用电量由参与建筑物的团集刻画，记为 $C := \{C_1, \ldots, C_{n_c}\}$，其中每个 $C_j \subseteq \mathcal{N}$满足 $j = 1, \ldots, n_c$。我们用 y_t^i 表示——

建筑物 i 在第 t 个时段的温度，其动态由受控线性时不变系统$r_t^i = a^i r_{t-1}^i + b^i x_t^i$, $y_t^i = c^i r_t^i$描述。理想的能源配置曲线应在将室内温度控制在舒适区间 $[y_t^i, \bar{y}_t^i]$ 内与降低能源成本之间取得平衡。此外，每栋建筑物的空调功率均不得超过系统容量。在此设定下，给定其他建筑物的功率曲线 x^{-i}，每栋建筑物都旨在确定最优功率控制策略，可表示如下：

$$
\begin{aligned}
&\text{minimize}_{x^i \in \mathcal{X}^i} \ (p_e)^T x^i + Q^i(x^i) + p_d \cdot R^i(x) \\
&\text{subject to } r_t^i = a^i r_{t-1}^i + b^i x_t^i, \forall t \in \mathcal{T} \\
&\qquad y_t^i = c^i r_t^i, \forall t \in \mathcal{T} \\
&\qquad \underline{y}_t^i \le y_t^i \le \bar{y}_t^i, \forall t \in \mathcal{T} \\
&\qquad 0 \le x_t^i \le \bar{x}^i, \forall t \in \mathcal{T}.
\end{aligned} \tag{21}
$$

其中， Q^i 是为比较收敛速率而人为添加的强凸二次函数，R^i表示建筑物 i在聚合商峰值需求中所占的份额。我们采用 Shapley 值的近似版本，在参与者之间分摊集体需求费用：

$$
R^i(x) = \sum_{C_j : i \in C_j} \frac{(N - |C_j|)!(|C_j| - 1)!}{N!} \Big(V(C_j, x) - V(C_j \backslash \{i\}, x) \Big),
$$

$$
\text{where } V(C_j, x) = \frac{1}{C} \log \Big(\sum_{t \in \mathcal{T}} \exp \big(\sum_{l \in C_j} C x_t^l \big) \Big) \approx \max_{t \in \mathcal{T}} \Big\{ \sum_{l \in C_j} x_t^l \Big\}.
$$

在上述定义中， C 是一个人为选定的参数，用于控制近似精度。我们还注意到，这组参与者的目标函数存在如下势函数：

$$
\begin{aligned}
\Phi(x) = &\sum_{i \in N} \Big((p_e)^T x^i + Q^i(x^i) \Big) \\
&+ p_d \cdot \sum_{C_j \in C} \frac{(N - |C_j|)!(|C_j| - 1)!}{N!} V(C_j, x).
\end{aligned}
$$

假设有十栋建筑物（$N = 10$）参与该博弈，并进行两组仿真，其中每栋建筑物i 需要分别针对 2 个时段（$T = 2$）和 4 个时段（$T = 4$）确定其能源策略。二次项 $Q^i(x^i) = (x^i)^T \text{diag}(\lambda_{i1}, \ldots, \lambda_{in^i}) x^i$的每个对角元素$\lambda_{ij}$均从 $[0.04, 0.06]$. 中随机采样。我们采用以下指标衡量各方法的性能：纳什均衡与扰动动作之间的相对距离，$\| \hat{X}_{k+1/2} - x_* \|_2 / \|x_*\|_2$, 以及势函数最优值与扰动动作处的值之差， $\Phi(\hat{X}_{k+1/2}) - \Phi_*$。 $T = 2$ 的仿真结果见图 3， $T = 4$ 的仿真结果见图 4；图中将所提方法的收敛速度与 [33], [35], [36], [49] 中已有学习算法的收敛速度进行了比较。步长和查询半径分别设为$\gamma_k = \alpha_\gamma / (k + K_\gamma)^{a_\gamma}$ 和 $\delta_k = \alpha_\delta / (k + K_\delta)^{a_\delta}$的形式，幂参数的具体取值列于图例中。现有算法的参数分别根据 [33, Thm. 5.2], [49, Thm. 3], [35,Thm. 2]和 [36, Thm. 2], 中的结果选取。每种方法在不同随机样本路径下运行三次，并取其平均值

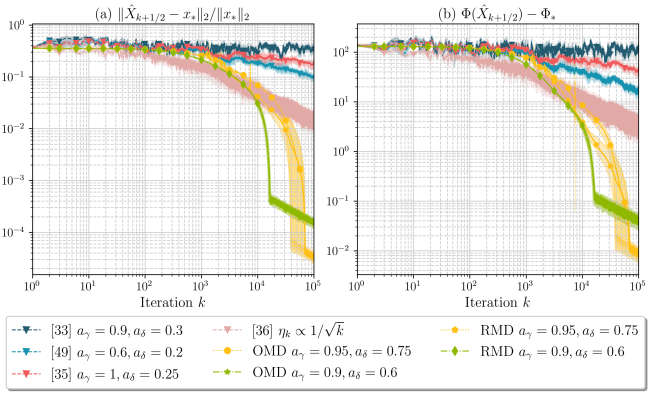

图3：热控制问题中算法1和算法2的性能（$T = 2$）

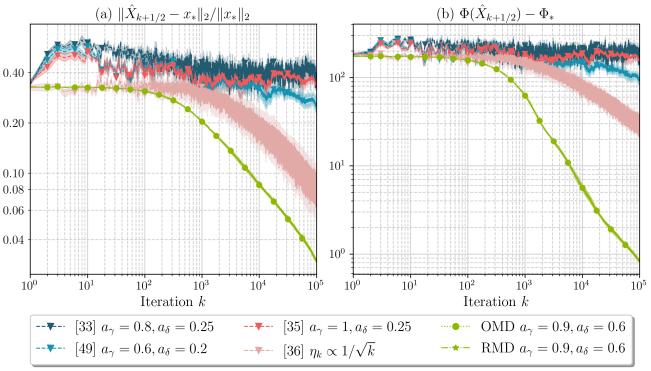

图4：热控制问题中算法1和算法2的性能（$T = 4$）

结果以实线/虚线表示，波动则通过半透明包络线呈现。

从图3和图4可以看出，算法1和算法2能够降低估计方差，并显著提高收敛速率。在图3中，我们考虑两组参数：参数组（a）更新较为激进（$a_\gamma = 0.95$, $a_\delta = 0.75$），参数组（b）则较为保守（$a_\gamma = 0.9$, $a_\delta = 0.6$）。在前 2×10^4 次迭代中，参数组（b）的表现优于参数组（a），这可能是因为参数组（b）的常数 M_1 和 M_2 较小，且起始迭代次数 K 符合定理 3 中的说明。然而，随着算法继续运行，参数组（a）逐渐超过参数组（b），这与定理 3 中的衰减速率结果一致。当 T 从2 增加到4时，常数 M_1、 M_2和K 显著增大，因此建议选择相对保守的参数组，以获得理想的学习动态。

变分不等式。结论与未来方向

本文研究多玩家连续博弈中的赌博机学习，并提出两种学习算法。这些算法将残差伪梯度估计分别与两种单调用外梯度方案相结合，即乐观镜像下降和反射镜像下降。我们证明，所提出算法的实际博弈序列收敛至

伪单调加博弈的一个临界点，且几乎必然收敛。此外，对于强伪单调博弈，所提出的算法能够达到最优收敛速率 $O(1/t^{1-\epsilon})$，从而显著提升仅需一次预言机调用的学习方案的收敛速度。仍有若干未解决问题。在问题表述中，我们隐含地假设实际获得的目标函数值是准确的，而当这些值受到随机噪声干扰时，如何达到相同或相近的收敛速率仍是一个未解决的问题。另一个可能的未来研究方向是，在不采用 Tikhonov 正则化的情况下，将实际博弈序列几乎必然收敛的结果推广到更一般的博弈类别，例如仅单调博弈。我们计划在未来的工作中探讨这些问题。

附录

A. 预备知识

引理 A.1. 考虑赋范巴拿赫空间 $\mathcal{B}$，其范数为 $\|\|$，并设有一个闭凸可行集 $\mathcal{X} \subseteq cl(\text{定义域 } \psi) \subseteq \mathcal{B}$。假设 ψ: 定义域 $\psi \to \mathbb{R}$ 是一个距离生成函数，则对于所有 $x \in \text{定义域} \psi \cap \mathcal{X}$ 和 $y \in \mathcal{B}^*$，以下关系成立：

(i) $x = \nabla\psi^*(y) \Longleftrightarrow y \in \nabla\psi(x) + N_{\mathcal{X}}(x)$; *(ii)* 以下条件等价：$x^+ = P_{x,\psi}(y) \Longleftrightarrow \nabla\psi(x) + y \in \nabla\psi(x^+) + N_{\mathcal{X}}(x) \Longleftrightarrow x^+ = \nabla\psi^*(\nabla\psi(x) + y)$; *(iii)* $x = \nabla\psi^*(y)$ 和 $p \in \mathcal{X} \Longrightarrow \langle \nabla\psi(x), x - p\rangle \le \langle y, x - p\rangle$; *(iv)* 镜像映射 $\nabla\psi^*$ 和近端映射 $P_{x,\psi}$ 是 $1/\tilde{\mu}$-利普希茨连续的，即对于任意 y_1 和 $y_2 \in \mathcal{B}^*$，$\|\nabla\psi^*(y_1) - \nabla\psi^*(y_2)\| \le \frac{1}{\tilde{\mu}}\|y_1 - y_2\|_*$ 和 $\|P_{x,\psi}(y_1) - P_{x,\psi}(y_2)\| \le \frac{1}{\tilde{\mu}}\|y_1 - y_2\|_*$;

(v) $\nabla\psi^* \circ \nabla\psi$: 定义域 $\psi \cap \mathcal{X} \to$ 定义域 $\psi \cap \mathcal{X}$ 是一个恒等映射。

证明。*(i)* 反向结论直接由约束优化的一阶最优性条件得出，即若对于某个 $x \in \text{定义域} \psi \cap \mathcal{X}$ 和 $y \in \mathrm{B}^*$，零包含关系 $0 \in \nabla\psi(x) - y + N_{\mathcal{X}}(x)$ 成立，则 $x = \operatorname{argmax}_{x \in \mathcal{X}}\{\langle y, x\rangle - \psi(x)\}$。至于另一个方向，重点在于说明，我们可以将解的讨论范围从 $\mathrm{X} = \mathrm{cl}(\mathrm{dom}\,\psi) \cap \mathrm{X}$ 限定到定义域 $\psi \cap \mathrm{X}$。给定任意 $y \in \mathrm{B}^*$，最大化问题 $\nabla\psi^*(y)$ 的一个解 x 满足 $0 \in \nabla\psi(x) - y + N_{\mathcal{X}}(x)$。由于 $\nabla\psi(\text{定义域 } \psi) = \mathbb{R}^n$ 且对所有 $x \in \mathrm{X}$ 都有 $N_{\mathcal{X}}(x) \subseteq \mathbb{R}^n$，因此总存在一个 $x^\star \in \text{定义域} \psi \cap \mathrm{X}$，使上述零包含关系成立。此外，$-\psi$ 的强凹性表明，$\nabla\psi^*(y)$ 存在唯一解 $x^\star$。

(ii) 该等价关系也是由*(i)*中讨论的一阶最优性条件直接得出的，出于同样的原因，我们可以断言 $x^+ \in \text{定义域} \psi \cap \mathcal{X}$。

(iii) 注意到 $y - \nabla\psi(x) \in N_{\mathcal{X}}(x)$ 以及法锥的定义，即若 $\langle \phi, p - x\rangle \le 0$ 对所有 $p \in \mathcal{X}$ 均成立，则有 $\phi \in N_{\mathcal{X}}(x)$，我们可以得到关系 $\langle y - \nabla\psi(x),\ p - x\rangle \le 0$。

(iv) 我们只需证明 $\nabla\psi^*$ 的利普希茨连续性，而 P_x 的利普希茨连续性可直接由 *(ii)* 中给出的等价关系推出。对于任意 $y_1, y_2 \in \mathrm{B}^*$，令 $x_1 = \nabla\psi^*(y_1)$ 和 $x_2 = \nabla\psi^*(y_2)$，则 *(iii)* 给出 $\langle \nabla\psi(x_1), x_1 - x_2\rangle \le \langle y_1, x_1 - x_2\rangle$ 以及 $\langle \nabla\psi(x_2), x_2 - x_1\rangle \le \langle y_2, x_2 - x_1\rangle$。结合两者可得 $\langle \nabla\psi(x_1) - \nabla\psi(x_2), x_1 - x_2\rangle \le \langle y_1 - y_2, x_1 - x_2\rangle$。根据柯西-施瓦茨不等式，$\langle y_1 - y_2, x_1 - x_2\rangle \le \|y_1 - y_2\|_* \cdot \|x_1 - x_2\|$。由 ψ 的 $\tilde{\mu}$-强凸性，即 $\langle \nabla\psi(x_1) - \nabla\psi(x_2), x_1 - x_2\rangle \ge \tilde{\mu}\|x_1 - x_2\|^2$，即可得出结论。

(v) 令 $x^+ = \nabla\psi^*(\nabla\psi(x))$。显然，$\nabla\psi(x) \in \nabla\psi(x^+) + N_{\mathcal{X}}(x^+)$ 且 $0 \in N_{\mathcal{X}}(x^+)$，两者共同推出 $x^+ = x$，且其唯一确定。□

引理 A.2。 设 ψ 是定义在 $\mathcal{S}_2 \subseteq \mathcal{S}_1 \subseteq cl(\text{定义域} \psi)$ 上的 $\tilde{\mu}$-强凸距离生成函数，D 是其关联的布雷格曼散度。对于任意 $p \in \mathcal{S}_2$ 和 $x \in$ 定义域 ψ，令 $x_1^+ = P_{x,\mathcal{S}_1}(y_1)$ 以及 $x_2^+ = P_{x,\mathcal{S}_2}(y_2)$。则下列不等式成立

$$D(p, x_2^+) \le D(p, x) + \langle y_2, x_1^+ - p\rangle + \langle y_1 - y_2, x_1^+ - x_2^+\rangle - D(x_2^+, x_1^+) - D(x_1^+, x) \tag{A.1}$$

$$\le D(p, x) + \langle y_2, x_1^+ - p\rangle + \frac{1}{2\tilde{\mu}}\|y_2 - y_1\|_*^2 - \frac{\tilde{\mu}}{2}\|x_1^+ - x\|^2. \tag{A.2}$$

证明。根据 Bregman 散度的“三点恒等式”，我们可以将 x_2^+ 与 x 联系起来，如下所示：

$$D(p, x_2^+) = D(p, x) - D(x_2^+, x) + \langle \nabla\psi(x_2^+) - \nabla\psi(x), x_2^+ - p\rangle \overset{(a)}{\le} D(p, x) - D(x_2^+, x) + \langle y_2, x_2^+ - p\rangle,$$

其中，在 *(a)*中，我们利用了以下事实：$x_2^+ = \nabla\psi^*(\nabla\psi(x) + y_2)$ 在集合 $\mathcal{S}_2$ 上成立，且 $p \in \mathcal{S}_2$，，这意味着 $\langle \nabla\psi(x_2^+), x_2^+ - p\rangle \le \langle \nabla\psi(x) + y_2, x_2^+ - p\rangle$，由引理 A.1*(iii)*可得。由于 $x_2^+ \in \mathcal{S}_2 \subseteq \mathcal{S}_1$，，同样根据上述类似的论证，我们有：

$$D(x_2^+, x_1^+) \le D(x_2^+, x) - D(x_1^+, x) + \langle y_1, x_1^+ - x_2^+\rangle.$$

结合上述两个不等式即可得到 (A.1)。根据 ψ 的 $\tilde{\mu}$-强单调性所蕴含的 D 下界，可得 $D(x_2^+, x_1^+) \ge \frac{\tilde{\mu}}{2}\|x_2^+ - x_1^+\|$ 和 $D(x_1^+, x) \ge \frac{\tilde{\mu}}{2}\|x_1^+ - x\|^2$。此外，根据柯西-施瓦茨不等式，$\langle y_1 - y_2, x_1^+ - x_2^+\rangle \le \frac{1}{2\tilde{\mu}}\|y_1 - y_2\|^2 + \frac{\tilde{\mu}}{2}\|x_1^+ - x_2^+\|$。综合以上各点，可得 (A.2) 中所需的界。□

引理 A.3。 考虑两个非负序列 $(\gamma_k)_{k \in \mathbb{N}_+} \in \mathbb{R}^{\mathbb{N}}$ 和 $(\delta_k)_{k \in \mathbb{N}_+} \in \mathbb{R}^{\mathbb{N}}$. 假设 $\lim_{k\to\infty} \gamma_k = 0$ 且 $\lim_{k\to\infty} \delta_k = 0$. 此外，设 $\sum_{k \in \mathbb{N}} \gamma_k = +\infty$ 且 $\lim_{k\to\infty} \gamma_k/\delta_k = 0$. 那么，对于任意固定的 $K \in \mathbb{N}_+$，都有 $\lim_{k\to\infty} \gamma_{k-K}/\delta_k = 0$。

证明。为了研究 γ_{k-K}/δ_k，注意到 $\gamma_k/\delta_k + (\gamma_{k-K} - \gamma_k)/\delta_k$，因此只需证明第二部分收敛到零。为此，我们通过望远镜求和考虑 $\sum_{k=K}^{\infty}(\gamma_{k-K} - \gamma_k) = \lim_{n\to\infty}\sum_{k=K}^{n}(\gamma_{k-K} - \gamma_k) = \lim_{n\to\infty}(\sum_{k=0}^{K-1}\gamma_k - \sum_{k=n-K+1}^{n}\gamma_k)$。此外，$\sum_{k=0}^{K-1}\gamma_k$ 只是一个常数，且 $\lim_{n\to\infty}\sum_{k=n-K+1}^{n}\gamma_k = 0$。因此 $\sum_{k=K}^{\infty}(\gamma_{k-K} - \gamma_k) < \infty$，并且尾部的 $(\gamma_k - \gamma_{k+K})_{k\in\mathbb{N}}$ 比 $(\gamma_k)_{k\in\mathbb{N}}$ 衰减得更快，由此即可得到所需结论。□

B. 残差伪梯度估计的结果

证明。（引理1的证明）首先注意到，右端可以简化如下。

$$\begin{aligned}\mathbb{E}[G_k^i \mid \mathcal{F}_k] &= \frac{n^i}{\delta_k}\mathbb{E}\left[\left(J^i(\hat{X}_{k+1/2}) - J^i(\hat{X}_{k-1/2})\right)u_k^i \mid \mathcal{F}_k\right] \\ &= \frac{n^i}{\delta_k}\mathbb{E}[J^i(\hat{X}_{k+1/2})u_k^i \mid \mathcal{F}_k] - \frac{n^i}{\delta_k}J^i(\hat{X}_{k-1/2})\mathbb{E}[u_k^i] \\ &= \frac{n^i}{\delta_k}\mathbb{E}[J^i(\hat{X}_{k+1/2})u_k^i \mid \mathcal{F}_k].\end{aligned}$$

接下来，我们将证明 $\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})$ 是 $\frac{n^i}{\delta_k}\mathbb{E}[J^i(\hat{X}_{k+1/2}) u^i_k \mid \mathcal{F}_k]$ 的一个版本。$\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})$ 的完整表达式如下：

$$\frac{1}{\mathbb{V}^i}\nabla_{x^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i} J^i(\bar{X}^i_{k+1/2}+\tilde{\tau}^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})d\tilde{\tau}^i d\tau^{-i}.$$

为考察积分与导数能否互换，我们逐项检查上面的等式，即先分析 $\partial_{[x^i]_l}$ 对于 $l\in\{1,\dots,n^i\}$ 的情形，再以类似方式处理其他项。由于假设 J^i 关于 x^i 可微，根据中值定理，对于任意 $x\in\mathcal{X}$，

$$\frac{1}{\varepsilon}\big(J^i(x+\varepsilon e_l)-J^i(x)\big)=\partial_{[x^i]_l}J^i(x+\alpha\cdot\varepsilon e_l),$$

其中 ϵ 表示一个小系数；α 是$[0,1]$上的一个适当常数。在当前的正则性设定下，偏导数$\partial_{[x^i]_l}J^i$局部有界，其界取决于考察点$\bar{X}_{k+1/2}$及其邻域，这进一步意味着

$$\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i}\mid\partial_{[x^i]_l}J^i(\bar{X}^i_{k+1/2}(\omega)+\tilde{\tau}^i;\bar{X}^{-i}_{k+1/2}(\omega)+\tau^{-i})\mid d\tilde{\tau}^i d\tau^{-i}<\infty$$

对于任意可能的随机样本路径 $\omega\in\Omega$，根据控制收敛定理，我们可以交换 ε 趋于0 的极限与关于 $\delta_k\mathbb{S}_{-i}\times\delta_k\mathbb{B}_i$ 的积分，得到：

$$\begin{aligned}
&\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})\\
&=\frac{1}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i}\nabla_{x^i}J^i(\bar{X}^i_{k+1/2}+\tilde{\tau}^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})d\tilde{\tau}^i d\tau^{-i}\\
&\overset{(a)}{=}\frac{1}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{\mathcal{N}}}J^i(\bar{X}^i_{k+1/2}+\tau^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})\frac{\tau^i}{\|\tau^i\|}d\tau\\
&\overset{(b)}{=}\frac{n^i/\delta_k}{\mathrm{vol}(\delta_k\mathbb{S}_{\mathcal{N}})}\int_{\delta_k\mathbb{S}_{\mathcal{N}}}J^i(\bar{X}^i_{k+1/2}+\tau^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})\frac{\tau^i}{\delta_k}d\tau\\
&\overset{(c)}{=}\frac{n^i/\delta_k}{\mathrm{vol}(\mathbb{S}_{\mathcal{N}})}\int_{\mathbb{S}_{\mathcal{N}}}J^i(\bar{X}_{k+1/2}+\delta_k\tau)\tau^i d\tau,
\end{aligned}$$

其中，在(a)中，我们令 $\mathbb{S}_{\mathcal{N}}\coloneqq\prod_{i\in\mathcal{N}}\mathbb{S}_i$；等号由 Stokes 定理 [50, 定理 9.3.1] 得出，因为 J^i 在 x^i中是 C^1；(b) 直接由 $\mathrm{vol}(\delta_k\mathbb{B}_i)=(\delta_k/n_i)\mathrm{vol}(\delta_k\mathbb{S}_i)$ 这一事实推出；对于 (c)，我们只需应用变量替换。对于任意事件 $E\in\mathcal{F}_k$，我们可以验证：

$$\begin{aligned}
&\mathbb{E}[\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})\mathbb{1}_E]\\
&=\frac{n^i/\delta_k}{\mathrm{vol}(\mathbb{S}_{\mathcal{N}})}\cdot\int_E\int_{\mathbb{S}_{\mathcal{N}}}J^i(\bar{X}_{k+1/2}(\omega)+\delta_k\tau)\tau^i d\tau\mathcal{P}(d\omega)\\
&\overset{(a)}{=}\frac{n^i}{\delta_k}\int_{\bar{X}^i_{k+1/2}(E)}\int_{\mathbb{S}_{\mathcal{N}}}J^i(x+\delta_k\tau)\tau^i\tilde{\mu}_k(d\tau)\cdot\nu_{k+1/2}(dx)\\
&\overset{(b)}{=}\frac{n^i}{\delta_k}\int_{\bar{X}^i_{k+1/2}(E)\times\mathbb{S}_{\mathcal{N}}}J^i(x+\delta_k\tau)\tau^i(\tilde{\mu}_k\times\nu_{k+1/2})(dx,d\tau)\\
&\overset{(c)}{=}\frac{n^i}{\delta_k}\int_{E\times\Omega}J^i(\bar{X}_{k+1/2}(\omega_1)+\delta_k u_k(\omega_2))u^i_k(\omega_2)(\mathcal{P}\times\mathcal{P})(d\omega_1,d\omega_2)\\
&\overset{(d)}{=}\frac{n^i}{\delta_k}\mathbb{E}[J^i(\bar{X}_{k+1/2}+\delta_k u_k)u^i_k\mathbb{1}_E],
\end{aligned}$$

其中，在 (a) 中，我们令 $\tilde{\mu}_k\coloneqq\mathcal{P}\circ(u_k)^{-1}$ 表示 u_k 的概率测度，它对应于 $\mathbb{S}_{\mathcal{N}}$ 上的均匀分布；而 $\nu_{k+1/2}$ 表示 $\bar{X}_{k+1/2}$ 的概率测度。(a) 中的等式由变量替换公式得出，该公式用于计算 $\bar{X}^i_{k+1/2}$ 的期望值。为得到 (b) 中的关系，我们首先注意到 $\bar{X}^i_{k+1}/2(E)+\delta_k\mathbb{S}_{\mathcal{N}}$ 是 X_a 的子集，根据假设和构造，X_a 有界。因此，我们有可积性 $\int_{\bar{X}^i_{k+1}/2(E)\times\mathbb{S}\mathrm{N}}\mid J^i(x+\tau)\mid(\tilde{\mu}_k\times\nu_{k+1/2})(dx,d\tau)\leq\max_{x\in\mathcal{X}}J^i(x)<\infty$，并可在此应用 Fubini 定理得到该等式。对于 (c)，我们再次使用变量替换公式来计算涉及 $\bar{X}^i_{k+1}/2$ 和 u_k 的期望值。最后，(d) 成立，这是因为 u_k 与 $\bar{X}^i_{k+1}/2$ 和 $\mathbb{1}_E$ 相互独立。由此，根据定义 [51, Sec. 4.1]，完成了证明：$\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})$ 是条件期望 $\mathbb{E}[G^i_k\mid\mathrm{F}_k]$ 的一个版本。□

证明。（引理 2 的证明）

系统误差可以如下分解：

$$\|B^i_k\|_*=\|\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})-\nabla_{x^i}J^i(X_{k+1/2})\|_*\leq\underbrace{\|\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+\frac{1}{2}})-\nabla_{x^i}J^i(\bar{X}_{k+\frac{1}{2}})\|_*}_{(i)}+\underbrace{\|\nabla_{x^i}J^i(\bar{X}_{k+\frac{1}{2}})-\nabla_{x^i}J^i(X_{k+\frac{1}{2}})\|_*}_{(ii)}.$$

对于 (i)，我们首先应用引理1证明中的相同论证，对计算 $\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+\frac{1}{2}})$ 时交换导数 ∇_{x^i} 与积分，得到

$$\begin{aligned}
(i)&=\Big\|\frac{1}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i}\big(\nabla_{x^i}J^i(\bar{X}^i_{k+1/2}+\tilde{\tau}^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})\\
&\qquad-\nabla_{x^i}J^i(\bar{X}^i_{k+1/2};\bar{X}^{-i}_{k+1/2})\big)d\tilde{\tau}^i d\tau^{-i}\Big\|_*\\
&\overset{(a)}{\leq}\frac{1}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i}\big\|\nabla_{x^i}J^i(\bar{X}^i_{k+1/2}+\tilde{\tau}^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})\\
&\qquad-\nabla_{x^i}J^i(\bar{X}^i_{k+1/2};\bar{X}^{-i}_{k+1/2})\big\|_*d\tilde{\tau}^i d\tau^{-i}\\
&\overset{(b)}{\leq}\frac{L^i}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i}\|[\tilde{\tau}^i;\tau^{-i}]\|d\tilde{\tau}^i d\tau^{-i}\\
&\overset{(c)}{\leq}L^i\bar{u}_{\mathcal{N}}\cdot\delta_k,
\end{aligned}$$

其中，(a) 直接由三角不等式得到；对于 (b)，我们使用了 $\nabla_{x^i}J^i$ 的 L^i-利普希茨连续性；最后，对于 (c)，我们将被积函数替换为常数 $\bar{u}_{\mathcal{N}}$，其中对于某个 $u\in\mathbb{S}_{\mathcal{N}}\coloneqq\prod_{i\epsilon}$，$\bar{u}_{\mathcal{N}}\coloneqq\|u\|_{\mathcal{N}}\mathbb{S}_{i_\circ}$

通过再次使用 $\nabla_{x^i}J^i$ 的利普希茨连续性，可以直接构造出 (ii) 的上界，如下所示：

$$(ii)\leq L^i\cdot\Big\|\frac{\delta_k}{r^i}(X_{k+1/2}-p)\Big\|\leq\frac{L^i D_{\mathcal{X}_a}}{r^i}\delta_k,$$

其中，$D_{\mathcal{X}_a}$ 表示全局动作空间$\mathcal{X}_a$ 在给定范数 $\|\cdot\|$下的直径。于是，系统误差的叠加 $\|B_k\|_*$满足上界 $\|B_k\|_*\leq\alpha_B\delta_k$，其中 $\alpha_B=\sum_{i\in\mathcal{N}}L^i(\bar{u}_{\mathcal{N}}+D_{\mathcal{X}_a}/r^i)$。

对于随机误差 V_k，根据三角不等式，有

$$\|V^i_k\|_*\leq\underbrace{\|G^i_k\|_*}_{(i)}+\underbrace{\|\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})\|_*}_{(ii)}.$$

对于(i)，注意到 J^i 在紧集$\mathcal{X}_a$上连续且其最大值存在，可得

$$\|G^i_k\|_*\leq\frac{n^i}{\delta_k}\quad 2\quad\overline{\ }_{x\in\mathcal{X}_a}\mid\cdot\max()\cdot,$$

$J^i\ x\mid\|u^i_k\|_*\qquad\|u^i_k\|_*\qquad\|\ \|_*$

其中 是从 $\mathbb{S}^{n_i}$ 中采样的单位向量的·范数。令 $\bar{u}^i_*\coloneqq\|u^i_k\|_{*\circ}$ 对于(ii)，我们再次利用引理1证明中的论证来交换

偏导数与积分，并得到

$$\begin{aligned}\|\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})\|_* &= \frac{n^i/\delta_k}{\mathrm{vol}(\mathbb{S}_N)}\|\int_{\mathbb{S}_N} J^i(\bar{X}_{k+1/2}+\delta_k\tau)\tau^i d\tau\|_* \\ &\le \frac{n^i/\delta_k}{\mathrm{vol}(\mathbb{S}_N)}\int_{\mathbb{S}_N} \mid J^i(\bar{X}_{k+1/2}+\delta_k\tau)\mid \|\tau^i\|_* d\tau \\ &\le \frac{n^i}{\delta_k}\cdot\max_{x\in\mathcal{X}_a}\mid J^i(x)\mid\cdot\bar{u}^i_*.\end{aligned}$$

综上，堆叠随机误差向量为$\|V_k\|_* \le \alpha_V/\delta_k$，其中 $\alpha_V = \sum_{i\in\mathcal{N}} 3n^i$ 最大值$_{x\in\mathcal{X}_a}\mid J^i(x)\mid \bar{u}^i_*$。 □

证明。 （引理 3 的证明）
类似地，我们首先将平方范数分离，即

$$\|V_k\|^2_* \le 2\|G^i_k\|^2_* + 2\|\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})\|^2_*.$$

对于第二部分，可以得到一个常数上界：

$$\begin{aligned}&\|\nabla_{x^i}\tilde{J}^i_{\delta_k}(\bar{X}_{k+1/2})\|^2_* \\ &\overset{(a)}{\le} \frac{1}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i}\|\nabla_{x^i}J^i(\bar{X}^i_{k+1/2}+\tilde{\tau}^i;\bar{X}^{-i}_{k+1/2}+\tau^{-i})\|^2_* d\tilde{\tau}^i d\tau^{-i} \\ &\overset{(b)}{\le} \big(\max_{x\in\mathcal{X}_a}\|\nabla_{x^i}J^i(x)\|^2_*\big)\cdot\frac{1}{\mathbb{V}^i}\int_{\delta_k\mathbb{S}_{-i}}\int_{\delta_k\mathbb{B}_i} 1 d\tilde{\tau}^i d\tau^{-i} \\ &= \max_{x\in\mathcal{X}_a}\|\nabla_{x^i}J^i(x)\|^2_*\end{aligned}$$

其中，我们通过应用三角不等式和詹森不等式得到 (a) ；由于根据假设， $\|\nabla_{x^i}J^i(\cdot)\|^2_*$ 是定义域为 $\mathcal{X}_a$ 上的连续函数，最大值存在于 (b) </style}中，因此它具有一个常数上界。对于第一部分，我们利用 J^i 在 x 中的可微性以及均值定理得到

$$\begin{aligned}\|G^i_k\|^2_* &\overset{(a)}{=} \Big(\frac{n^i}{\delta_k}\Big)^2\Big(\langle\nabla_x J^i(\tilde{X}),\hat{X}_{k+1/2}-\hat{X}_{k-1/2}\rangle\Big)^2\|u^i_k\|^2_* \\ &\overset{(b)}{\le} \Big(\frac{n^i}{\delta_k}\Big)^2\|\nabla_x J^i(\tilde{X})\|^2_*\cdot\|\hat{X}_{k+1/2}-\hat{X}_{k-1/2}\|^2\cdot\|u^i_k\|^2_* \\ &\overset{(c)}{\le} \Big(\frac{n^i}{\delta_k}\Big)^2\bar{\nabla}^2_i(\bar{u}^i_*)^2\cdot\|\hat{X}_{k+1/2}-\hat{X}_{k-1/2}\|^2.\end{aligned}$$

在（a）中， $\tilde{X}$ 是 $\hat{}\,X_{k+1/2}$ 和 $\hat{}\,X_{k-1/2}$ 的某个凸组合，其仍然依赖于随机样本 ω。尽管如此，由于对于所有 $\omega\in\Omega$， $\tilde{X}(\omega)\in\mathrm{X}_a$，如（$b$）所示，我们可以为梯度的对偶范数找到一个常数上界。我们在（c）中将 $\bar{\nabla}^2_i :=$最大值$_{x\in\mathcal{X}_a}\|\nabla_x J^i(x)\|^2_*$ 记为。为简洁起见，令 $C_g :=$ $(\sum_{i\in\mathrm{N}} n^i\bar{\nabla}_i\bar{u}^i_*)^2$ 和 $\|G_k\|^2_* \le C_g/(\delta_k)^2\cdot\|\hat{}X_{k+1/2}-\hat{}X_{k-1/2}\|^2$。

接下来，我们考察 $\|\hat{}X_{k+1/2}-\hat{}X_{k-1/2}\|^2$ 的性质：

$$\begin{aligned}&\|\hat{X}_{k+1/2}-\hat{X}_{k-1/2}\|^2 = \|X_{k+1/2}-X_{k-1/2}+ \\ &\quad\delta_k\underbrace{R^{-1}(p-X_{k+1/2}+Ru_k)}_{\varphi_k} - \delta_{k-1}\underbrace{R^{-1}(p-X_{k-1/2}+Ru_{k-1})}_{\varphi_{k-1}}\|^2 \\ &\le (1+\alpha_1)\|X_{k+1/2}-X_{k-1/2}\|^2 + (1+\frac{1}{\alpha_1})\delta^2_k\|\varphi_k-\frac{\delta_{k-1}}{\delta_k}\varphi_{k-1}\|^2 \\ &\le (1+\alpha_1)\|X_{k+1/2}-X_{k-1/2}\|^2 + (1+\frac{1}{\alpha_1})\delta^2_k\bar{\varphi}^2,\end{aligned}$$

其中 $p := [p_i]_{i\in\mathcal{N}}$， $R :=\mathrm{blkd}(\{r^i I_{n^i}\}_{i\in\mathcal{N}})$ 是某个任意常数，而 $\alpha_1 > 0$， $\bar{\varphi}^2$ 表示$\|\varphi_k - \frac{\delta_{k-1}}{\delta_k}\varphi_{k-1}\|^2$ 的一个常数上界。对于 $\overline{\mathrm{OMD}}$，将所有博弈参与者的变量堆叠起来，算法 1 所建议的迭代包含以下两个主要更新步骤：

$$X_{k+1/2} = P_{X_k,\mathcal{X}}(-\gamma_k G_{k-1}),\ X_{k+1} = P_{X_k,\mathcal{X}}(-\gamma_k G_k). \tag{B.1}$$

因此，我们得到以下不等式：

$$\begin{aligned}&\|X_{k+1/2}-X_{k-1/2}\|^2 = \|X_{k+1/2}-X_k+X_k-X_{k-1/2}\|^2 \\ &\le 2\|X_{k+1/2}-X_k\|^2 + 2\|X_k-X_{k-1/2}\|^2 \\ &\overset{(a)}{\le} 2\|\nabla\psi^*(\nabla\psi(X_k)-\gamma_k G_{k-1})-\nabla\psi^*(\nabla\psi(X_k))\|^2 + \\ &2\|\nabla\psi^*(\nabla\psi(X_{k-1})-\gamma_{k-1}G_{k-1})-\nabla\psi^*(\nabla\psi(X_{k-1})-\gamma_{k-1}G_{k-2})\|^2 \\ &\overset{(b)}{\le} \frac{2}{\tilde{\mu}^2}\gamma^2_k\|G_{k-1}\|^2_* + \frac{2}{\tilde{\mu}^2}\gamma^2_{k-1}\|G_{k-1}-G_{k-2}\|^2_* \\ &\le \frac{6}{\tilde{\mu}^2}\gamma^2_{k-1}\|G_{k-1}\|^2_* + \frac{4}{\tilde{\mu}^2}\gamma^2_{k-1}\|G_{k-2}\|^2_*.\end{aligned}$$

在上述方程中，我们不再具体说明可行集，因为在 OMD 中，两个近端映射都关于策略空间 $\mathcal{X}$；(a) 和 (b) 是直接应用引理 A.1 (ii)、(v)以及(iv)中的利普希茨连续性得到的结果。结合上述结论可得，对所有 $k\ge 2$，

$$\begin{aligned}\|G_k\|^2_* &\le \frac{C_g}{\delta^2_k}\Big((1+\alpha_1)\|X_{k+1/2}-X_{k-1/2}\|^2 + (1+\frac{1}{\alpha_1})\delta^2_k\bar{\varphi}^2\Big) \\ &\le \frac{2C_g(1+\alpha_1)}{\tilde{\mu}^2}\Big(\frac{\gamma_{k-1}}{\delta_k}\Big)^2\big(3\|G_{k-1}\|^2_* + 2\|G_{k-2}\|^2_*\big) + C_g(1+\frac{1}{\alpha_1})\bar{\varphi}^2.\end{aligned}$$

由引理 A.3 可知 $\lim_{k\to\infty}\gamma_{k-1}/\delta_k = 0$。对于任意 $\varepsilon>0$，存在一个常数指标 K，使得对所有 $k > K$，

$\frac{2C_g(1+\alpha_1)}{\tilde{\mu}^2}\Big(\frac{\gamma_{k-1}}{\delta_k}\Big)^2 < \varepsilon$ 且 $\|G_k\|^2_* \le 3\varepsilon\|G_{k-1}\|^2_* + 2\varepsilon\|G_{k-2}\|^2_* + C_{\tilde{g}}$，其中 $C_{\tilde{g}} := C_g(1+\frac{1}{\alpha_1})\bar{\varphi}^2$。对于 $k > K$，根据 Jury 判据以及上述线性离散时间系统的特征多项式 $Q(\lambda) = \lambda^2 - 3\varepsilon\lambda - 2\varepsilon$，当 $Q(1) > 0$、$Q(-1) > 0$ 且 $|-2\varepsilon| < 1$ 时，系统稳定；这些条件共同表明，只需满足 $\varepsilon < 1/5$。对于 $k\le K$，$\|G_k\|_* \le \|F(X_{k+1/2})\|_* + \alpha_B\delta_k + \alpha_V/(\delta_K)^2$。因此，我们有 $\sup_{k\in\mathbb{N}_+}\|G_k\|^2_* < \infty$，并且存在一个常数 C_V，使得 $\sup_{k\in\mathbb{N}_+}\|V_k\|^2_* \le C_V$。

对于 RMD，算法 2 的紧致形式可写为

$$\begin{aligned}X_{k+1/2} &= P_{X_k,\mathbb{R}^n}(-(\nabla\psi(X_{k-1})-\nabla\psi(X_k))), \\ X_{k+1} &= P_{X_k,\mathcal{X}}(-\gamma_k G_k).\end{aligned} \tag{B.2}$$

由于 RMD 中的两个近端映射是针对两个不同集合实现的，因此与 OMD 相比，我们对 $\|X_{k+1/2}-X_{k-1/2}\|^2$ 考虑一个更宽松的上界：

$$\begin{aligned}&\|X_{k+1/2}-X_{k-1/2}\|^2 = \|X_{k+1/2}-X_k+X_k-X_{k-1}+X_{k-1}-X_{k-1/2}\|^2 \\ &= 4\|X_{k+1/2}-X_k\|^2 + 4\|X_k-X_{k-1}\|^2 + 2\|X_{k-1}-X_{k-1/2}\|^2 \\ &\overset{(a)}{\le} (\tilde{L}/\tilde{\mu})^2\cdot(8\|X_k-X_{k-1}\|^2 + 2\|X_{k-1}-X_{k-2}\|^2),\end{aligned}$$

其中 (a) 由全局距离生成函数 ψ 具有 $\tilde{\mu}$-强单调性和 $\tilde{L}$-光滑性这一事实可得。接着，我们利用反射梯度中的第二个更新，得到：

$$\|X_k-X_{k-1}\|^2 \le \frac{\gamma^2_{k-1}}{\tilde{\mu}^2}\|G_{k-1}\|^2_*, \|X_{k-1}-X_{k-2}\|^2 \le \frac{\gamma^2_{k-2}}{\tilde{\mu}^2}\|G_{k-2}\|^2_*.$$

将 $\|\hat{}X_{k+1/2}-\hat{}X_{k-1/2}\|^2$ 代入 $\|G_k\|^2_*$ 的上界，可得以下关系：

$$\|G_k\|^2_* \le \frac{2C_g\tilde{L}^2(1+\alpha_1)}{\tilde{\mu}^4}\Big(\frac{\gamma_{k-2}}{\delta_k}\Big)^2\big(4\|G_{k-1}\|^2_* + \|G_{k-2}\|^2_*\big) + C_{\tilde{g}}.$$

同理，由引理 A.3 可知，$\lim_{k\to\infty}\gamma_{k-2}/\delta_k = 0$，对任意$\varepsilon > 0$均成立。因此，存在一个常数指标 K，使得

对所有 $k > K$ $\|G_k\|_*^2 \le 4\varepsilon\|G_{k-1}\|_*^2 +$, $\varepsilon\|G_{k-2}\|_*^2 + C_{\tilde{g}}$ —— $2C_g L^2(\frac{1+\alpha_1}{\tilde{\mu}^4}\left(\frac{\gamma_{k-2}}{\delta_k}\right)^2 < \varepsilon^)$，且 . 同样，由 Jury稳定性判据以及特征多项式 $Q(\lambda) = \lambda^2 - 4\varepsilon\lambda - \varepsilon$，条件$\varepsilon < 1/5$ 可确保 $Q(1) > 0, Q(-1) > 0$, 以及 $|-\varepsilon| < 1$。其余论证与 OMD 中的相同，由此得到所需结论：对于 RMD，存在一个常数 C_V，使得 $\sup_{k\in\mathbb{N}_+}\|V_k\|_*^2 \le C_V$。□

C. 伪单调加博弈中所提出算法的几乎必然收敛

证明。（定理1的证明）

算法 1. 结合引理 A.2 中的 (B.1) 和 (A.2) 所给出的 OMD 紧致形式，我们得到以下递推关系：

$$D(p, X_{k+1}) \le D(p, X_k) - \gamma_k\langle G_k, X_{k+1/2} - p\rangle + \frac{\gamma_k^2}{2\tilde{\mu}}\|G_k - G_{k-1}\|_*^2 - \frac{\tilde{\mu}}{2}\|X_{k+1/2} - X_k\|^2,$$

对任意 $p \in \mathcal{X}$。回顾一下，在当前参数选择 (15) 下，引理 3 已证明 $\|G_k\|_*$ $k \in \mathbb{N}$ $\frac{\gamma_k^2}{2\tilde{\mu}}\|G_k - G_{k-1}\|_*^2 \le$ $\frac{\gamma_k^2}{\tilde{\mu}}\|G_k\|_*^2 + \frac{\gamma_k^2}{\tilde{\mu}}\|G_{k-1}\|_*^2$ — γ_k^2 对所有 都有界。因此，衰减速度与 相同，我们可以推断它是可求和的。另请注意，给定 σ-域$\mathcal{F}_k := \sigma\{X_0, u_1, \dots, u_{k-1}\}$， X_{k+1}、G_k 和 V_k 是上述变量中仅有的三个非$\mathcal{F}_k$-可测变量。因此，取条件期望 $\mathbb{E}[\cdot \mid \mathcal{F}_k]$，可得：

$$\begin{aligned}\mathbb{E}[D(p, X_{k+1}) \mid \mathcal{F}_k] &\overset{(a)}{\le} D(p, X_k) - \gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - p\rangle \\ &\quad + \gamma_k\|B_k\|_* \cdot \|X_{k+1/2} - p\| - \frac{\tilde{\mu}}{2}\|X_{k+1/2} - X_k\|^2 \\ &\quad + \frac{\gamma_k^2}{\tilde{\mu}}\mathbb{E}[\|G_k\|_*^2 \mid \mathcal{F}_k] + \frac{\gamma_k^2}{\tilde{\mu}}\|G_{k-1}\|_*^2 \\ &\overset{(b)}{\le} D(p, X_k) - \gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - p\rangle - \frac{\tilde{\mu}}{2}\|X_{k+1/2} - X_k\|^2 \\ &\quad + \alpha_B D_{\mathcal{X}}\gamma_k\delta_k + \frac{\gamma_k^2}{\tilde{\mu}}\mathbb{E}[\|G_k\|_*^2 \mid \mathcal{F}_k] + \frac{\gamma_k^2}{\tilde{\mu}}\|G_{k-1}\|_*^2,\end{aligned} \tag{C.1}$$

其中(a) 是由分解 $G_k = F(X_{k+1/2}) + B_k + V_k$ 得出的结果，等式 $\mathbb{E}[\langle V_k, X_{k+1/2} - p\rangle \mid \mathcal{F}_k] = \langle \mathbb{E}[V_k \mid \mathcal{F}_k], X_{k+1/2} - p\rangle = 0$，以及对 $\langle B_k, X_{k+1/2} - p\rangle$ 应用柯西-施瓦茨不等式；对于(b)，我们利用了如下事实：$X_k, X_{k-1/2} \in \mathcal{X}$，且随机变量 $\|B_k\|_*$ 和 $\|V_k\|_*^2$ 的取值范围分别满足 $\|B_k\|_* \le \alpha_B\delta_k$ 和 $\|V_k\|_*^2 \le C_V$。

然后，我们将 p 替换为所研究博弈 $\mathcal{G}$ 的任意临界点 x_*，因此，根据伪单调性，$\langle F(X_{k+1/2}), X_{k+1/2} - x_*\rangle \ge 0$ 对所有可能的 $X_{k+1/2}$ 值均成立。利用 Robbins-Siegmund（R-S）定理 [52, Thm. 1], ，我们可以得出以下结论：

(i) $(D(x_*, X_k))_{k\in\mathbb{N}}$ 收敛到一个几乎处处有限的极限；(ii) $\sum_{k\in\mathbb{N}_+}\tilde{\mu}/2\|X_{k+1/2} - X_k\|^2 < \infty$ 几乎处处成立；(iii) $\sum_{k\in\mathbb{N}}\gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - x_*\rangle < \infty$ 几乎处处成立。

断言 (ii) 表明 $\lim_{k\to\infty}\|X_{k+1/2} - X_k\|^2 = 0$ 几乎必然。此外，由于 $(\gamma_k)^{k\in\mathbb{N}_+}$ 不可求和，根据断言 $\liminf_{k\to\infty}\langle F(X_{k+1/2}), X_{k+1/2} - x_*\rangle(\omega) = 0$，对任意$\omega \in \tilde{\Omega}$，我们得到，其中 $\tilde{\Omega} \subseteq \Omega$ 以概率 1 成立。换言之，沿着子序列 $(k_m)_{m\in\mathbb{N}} \subseteq \mathbb{N}$，有 $\langle F(X_{k_m+1/2}), X_{k_m+1/2} - x_*\rangle(\omega) \to 0$。由于 $(X_k(\omega))_{k\in\mathbb{N}} \in \mathrm{X}$ 是有界序列，因此存在一个子子序列，其满足 $(l_m)_{m\in\mathbb{N}} \subseteq (k_m)_{m\in\mathbb{N}}$，使得 $(X_{\ell_m}(\omega))_{m\in\mathbb{N}}$ 收敛到点 $X^\dagger(\omega) \in \mathrm{X}$。断言 (ii) 进一步意味着 $X_{\ell_m+1/2}(\omega) \to X^\dagger(\omega)$。由$F$的连续性可得：

$$\begin{aligned}&\lim_{m\to\infty}\langle F(X_{\ell_m+1/2}(\omega)), X_{\ell_m+1/2}(\omega) - x_*\rangle \\ &= \langle \lim_{m\to\infty} F(X_{\ell_m+1/2}(\omega)), \lim_{m\to\infty}(X_{\ell_m+1/2}(\omega) - x_*)\rangle \\ &= \langle F(X^\dagger(\omega)), X^\dagger(\omega) - x_*\rangle = 0\end{aligned}$$

由于 x_* 是相应变分不等式的解，我们有$\langle F(x_*), X^\dagger(\omega) - x_*\rangle \ge 0$。结合上述结果，根据假设 F 具有伪单调加性质，可得 $F(x_*) = F(X^\dagger(\omega))$，进而表明 $\forall x \in \mathcal{X}$：

$$\begin{aligned}&\langle F(X^\dagger(\omega)), x - X^\dagger(\omega)\rangle = \langle F(X^\dagger(\omega)), x - x_* + x_* - X^\dagger(\omega)\rangle \\ &= \langle F(x_*), x - x_*\rangle + \langle F(X^\dagger(\omega)), x_* - X^\dagger(\omega)\rangle \ge 0,\end{aligned}$$

也就是说，$X^\dagger(\omega)$ 是相应变分不等式的解，因此是所研究原博弈的一个临界点。于是，我们可以在递归不等式中用 $X^\dagger(\omega)$ 替换 x_*。结合布雷格曼互易性假设，以及存在一个子序列 $(X_{\ell_m})_{m\in\mathbb{N}}$ 使得 $D(X^\dagger(\omega), X_{\ell_m}(\omega)) \to 0$，并且 $(D(X^\dagger(\omega), X_{\ell_m}(\omega)))_{k\in\mathbb{N}}$ 存在有限极限这一事实，我们可以得出，对于整个序列，$D(X^\dagger(\omega), X_k(\omega)) \to 0$，因此 $X_k(\omega) \to X^\dagger(\omega)$。最后，回想一下，$\hat{} X_{k+1/2} = (1 - \delta_k R^{-1})X_{k+1/2} + \delta_k R^{-1}(p + Ru_k)$，其中 $p := [p_i]_{i\in\mathcal{N}}$ 且 $R := \mathrm{blkd}(\{r^i I_{n^i}\}_{i\in\mathcal{N}})$，因此我们有 $\hat{} X_{k+1/2}$几乎必然收敛→ $X_{k+1/2}$，当$\delta_k \to 0$时成立，由此完成了对实际行动序列 $(\hat{} X_{k+1/2})^{k\in\mathbb{N}_+}$ 几乎必然收敛的证明。

算法 2. 回顾式 (B.2) 中 RMD 的紧致形式。对于任意 $p \in \mathcal{X}$，将式 (A.2) 中的变量替换为上述原始变量和对偶变量，可得：

$$D(p, X_{k+1}) \le D(p, X_k) - \gamma_k\langle G_k, X_{k+1/2} - p\rangle - \frac{\tilde{\mu}}{2}\|X_{k+1/2} - X_k\|^2 + \frac{1}{2\tilde{\mu}}\|-\gamma_k G_k + (\nabla\psi(X_{k-1}) - \nabla\psi(X_k))\|_*^2. \tag{C.2}$$

为了应用 R-S定理并证明几乎必然收敛，我们对上述不等式中的最后一个对偶范数进行拆解，并证明其可求和性可由步长 γ_k 良好控制：

$$\begin{aligned}&\frac{1}{2\tilde{\mu}}\|-\gamma_k G_k + (\nabla\psi(X_{k-1}) - \nabla\psi(X_k))\|_*^2 \\ &\le \frac{(\gamma_k)^2}{\tilde{\mu}}\|G_k\|_*^2 + \frac{1}{\tilde{\mu}}\|\nabla\psi(X_{k-1}) - \nabla\psi(X_k)\|_*^2 \\ &\overset{(a)}{\le} \frac{(\gamma_k)^2}{\tilde{\mu}}\|G_k\|_*^2 + \frac{\tilde{L}^2}{\tilde{\mu}}\|X_{k-1} - X_k\|^2 \overset{(b)}{=} \frac{(\gamma_k)^2}{\tilde{\mu}}\|G_k\|_*^2 \\ &\quad + \frac{\tilde{L}^2}{\tilde{\mu}}\|\nabla\psi^*(\nabla\psi(X_{k-1}) - \gamma_{k-1}G_{k-1}) - \nabla\psi^*(\nabla\psi(X_{k-1}))\|^2 \\ &\overset{(c)}{\le} \frac{(\gamma_k)^2}{\tilde{\mu}}\|G_k\|_*^2 + \frac{(\tilde{L}\gamma_{k-1})^2}{\tilde{\mu}^3}\|G_{k-1}\|_*^2,\end{aligned}$$

其中，(a) 直接由 $\tilde{L}$ 关于 $\nabla\psi$ 的利普希茨连续性得到；在 (b) 中，我们应用引理 A.1(iii) ，将 X_k 表示为镜像映射 $\nabla\psi^*$ 的形式，并注意到对于任意 $x \in \mathcal{X}$，$(\nabla\psi^* \circ \nabla\psi)(x) = x$；$(c)$ 是 $\nabla\psi^*$ 的$(1/\tilde{\mu})$-利普希茨连续性的结果。回顾这一分解

$G_k = F(X_{k+1/2}) + V_k + B_k$，以及引理 3 中证明的 $\|G_k\|_*$ 对所有 $k \in \mathbb{N}_+$ 有界。(C.2) 中最后一个对偶范数的可求和性继承自$(\gamma_k^2)^{k\in\mathbb{N}_+}$的可求和性。令滤过定义为 $\mathcal{F}_k := \sigma\{X_0, u_1, \ldots, u_{k-1}\}$，我们在 (C.2) 两边取条件期望 $\mathbb{E}[\cdot \mid \mathcal{F}_k]$，得到：

$$\begin{aligned}\mathbb{E}[D(p, X_{k+1}) \mid \mathcal{F}_k] \le D(p, X_k) - \gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - p\rangle \\ - \tilde{\mu}/2 \cdot \|X_{k+1/2} - X_k\|^2 + \gamma_k\|B_k\|_*\|X_{k+1/2} - p\| \\ + (\gamma_k)^2/\tilde{\mu} \cdot \mathbb{E}[\|G_k\|_*^2 \mid \mathcal{F}_k] + (\tilde{L}\gamma_{k-1})^2/\tilde{\mu}^3 \cdot \|G_{k-1}\|_*^2.\end{aligned} \tag{C.3}$$

然而，对于 $F(X_{k+1/2})$ 与 $X_{k+1/2} - p$ 之间的内积，$X_{k+1/2} \in \mathrm{X}_{\boldsymbol{R}}$ 可以位于 X 之外。为了利用定义 1 中的正则性并确保内积为正，我们更容易改为处理 $\langle F(X_k), X_k - p\rangle$，其中$X_k \in \mathrm{X}$。为此，我们对式 (C.3) 中的内积推导如下上界：$-\gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - p\rangle = -\gamma_k\langle F(X_k), X_k - p\rangle + \gamma_k\langle F(X_{k+1/2}) - F(X_k), p\rangle + \gamma_k\langle F(X_k), X_k\rangle - \gamma_k\langle F(X_{k+1/2}), X_{k+1/2}\rangle \le -\gamma_k\langle F(X_k), X_k - p\rangle + \gamma_k L\|X_{k+1/2} - X_k\|\cdot\|p\| + \gamma_k\langle F(X_k) - F(X_{k+1/2}), X_k\rangle - \gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - X_k\rangle \le -\gamma_k\langle F(X_k), X_k - x_*\rangle + \Delta_k'$，其中 $\Delta_k' := \gamma_k\|X_{k+1/2} - X_k\|\cdot(L\|p\| + L\|X_k\| + \|F(X_{k+1/2})\|_*)$。结合关系 $\|X_{k+1/2} - X_k\| \le \frac{\tilde{L}}{\tilde{\mu}}\|X_k - X_{k-1}\| \le \frac{\tilde{L}\gamma_{k-1}}{\tilde{\mu}}\|G_{k-1}\|_*$和 ($L\|x_*\| + L\|X_k\| + \|F(X_{k+1/2})\|_*$) 的有界性，我们断言 Δ_k'可求和，并得到以下递归不等式

$$\begin{aligned}\mathbb{E}[D(p, X_{k+1}) \mid \mathcal{F}_k] \le D(p, X_k) - \gamma_k\langle F(X_k), X_k - p\rangle \\ - \tilde{\mu}/2 \cdot \|X_{k+1/2} - X_k\|^2 + \Delta_k' + \alpha_B D_{\mathcal{X}}\gamma_k\delta_k \\ + (\gamma_k)^2/\tilde{\mu} \cdot \mathbb{E}[\|G_k\|_*^2 \mid \mathcal{F}_k] + (\tilde{L}\gamma_{k-1})^2/\tilde{\mu}^3 \cdot \|G_{k-1}\|_*^2.\end{aligned} \tag{C.4}$$

再次，我们用临界点 $x_* \in \mathcal{X}$ 替换 p，并应用 R-S定理得出以下结论：

(i) $(D(x_*, X_k))_{k\in\mathbb{N}}$ 收敛到一个几乎必然有限的极限；
(ii) $\sum_{k\in\mathbb{N}_+}\tilde{\mu}/2\|X_{k+1/2} - X_k\|^2 < \infty$ 几乎必然；
(iii) $\sum_{k\in\mathbb{N}}\gamma_k\langle F(X_k), X_k - x_*\rangle < \infty$几乎必然。

在算子 F 为伪单调加性的假设下，OMD 的其余论证可直接沿用至此处的讨论；随后即可证明 $(\hat{X}_{k+1/2})^{k\in\mathbb{N}_+}$ 几乎必然收敛的相同结论。□

证明。（推论1的证明）对于 OMD 和 RMD，我们采用定理1中的相同论证来得到这三个结论，以及应用 R-S定理后（子）子序列 $(\ell_m)_{m\in\mathbb{N}}$ 的存在性。沿着 $(\ell_m)_{m\in\mathbb{N}}$，我们有 $X_{\ell_m+1}/2(\omega) \to X^\dagger(\omega)$、$X_{\ell_m}(\omega) \to X^\dagger(\omega)$，且极限点 $X^\dagger(\omega)$ 满足$\langle F(X^\dagger(\omega))$、$X^\dagger(\omega) - x_*\rangle = 0$。若 $X^\dagger(\omega)$ 是一个临界点，则还可再次应用定理1的其余结论，从而得出 $(\hat{X}_{k+1/2})^{k\in\mathbb{N}_+}$ 几乎必然收敛。在余下的证明中，我们将证明，根据定义 2，$X^\dagger(\omega)$ 是 G 的临界点（*iii*）–（*vi*）。

对于定义2 (*v*)，由于 $\langle F(x), x - x_*\rangle > 0$对于所有 $x \in \mathcal{X}\backslash\mathcal{X}_*$，而当$\langle F(X^\dagger(\omega)), X^\dagger(\omega) - x_*\rangle = 0$时，我们有 $X^\dagger(\omega) \in \mathcal{X}_*$。

对于定义2 (*iii*) 和(*iv*)，$X^\dagger(\omega) \in \mathcal{X}_*$ 可直接由以下事实推出：严格伪单调性或强伪单调性均是严格相干性的充分条件。

根据定义2（*vi*），由于 x_* 是一个临界点，因此对所有 $x \in$X, 我们有$\Phi(x_*) \le \Phi(x)$，即它是一个全局最小值。关系 $\langle F(X^\dagger(\omega)), x_* - X^\dagger(\omega)\rangle = 0$进一步表明 $\Phi(x_*) \ge \Phi(X^\dagger(\omega))$。因此，我们可以得出结论：$X^\dagger(\omega)$ 是 Φ 的全局最小值，因而是一个临界点。□

D. 仅单调博弈中的遍历收敛速率

证明。（定理2的证明）

算法 1。在假设 1 和 2 下，我们可以得到递归不等式 (C.1)。根据引理 3 的证明，存在一个常数 $\bar{g}$，使得 $\sup_{k\in\mathbb{N}_+,\ \omega\in\Omega}\|G_k(\omega)\|_* \le \bar{g}$且 $\bar{\nabla} := \max_{x\in\mathcal{X}}\|F(x)\|_*$。对 (C.1) 两边取期望可得：

$$\mathbb{E}[\gamma_k\langle F(X_{k+1/2}), X_{k+1/2} - p\rangle] \le \mathbb{E}[D(p, X_k)] - \mathbb{E}[D(p, X_{k+1})] + \Delta_k,$$

其中，为简化记号，我们令 $\Delta_k := \alpha_B D_{\mathcal{X}}\gamma_k\delta_k + 2\gamma_k^2\bar{g}^2/\tilde{\mu}$。由于 F 是定义在 $\mathcal{X}$ 上的单调算子，因此有$\langle F(p), \hat{X}_{k+1/2} - p\rangle \le \langle F(p), X_{k+1/2} - p\rangle + \|F(p)\|_*\|\hat{X}_{k+1/2} - X_{k+1/2}\| \le \langle F(X_{k+1/2}), X_{k+1/2} - p\rangle + \bar{\nabla}\alpha_{\mathcal{X}}\delta_k$，其中 $\alpha_{\mathcal{X}}$表示一个取决于 $\mathcal{X}$ 几何性质的常数。令 $\Delta_k^\star := \Delta_k + \bar{\nabla}\alpha_{\mathcal{X}}\delta_k\gamma_k$。现在，从 $t = 1$ 到 k作简单的望远镜求和，可得：

$$\frac{1}{\sum_{t=1}^k\gamma_t}\sum_{t=1}^k\mathbb{E}[\langle F(p), \gamma_t(\hat{X}_{t+1/2} - p)\rangle] \le \frac{\mathbb{E}[D(p, X_1)] + \sum_{t=1}^k\Delta_t^\star}{\sum_{t=1}^k\gamma_t}.$$

在假设 (15) 中的 γ_k 和 δ_k 成立时，存在一个常数 M，使得对所有 $k \in \mathbb{N}_+$，增量序列为 $\sum_{t=1}^k\Delta_t^\star \le M$。最后，应用 (6) 和 (16) 中的定义，并令 $p^* := \operatorname{argmax}_{p\in\mathcal{X}}\langle F(p), \check{X}_k - p\rangle$，即可得到：

$$\mathbb{E}[\mathrm{Err}_{\mathcal{X}}(\check{X}_k)] \le \frac{\mathbb{E}[D(p^*, X_1)] + M}{\sum_{t=1}^k\gamma_t} \le \frac{\mathbb{E}[\max_{p\in\mathcal{X}} D(p, X_1)] + M}{\sum_{t=1}^k\gamma_t},$$

至此，关于 OMD 的证明完成。

算法 2。与定理1的证明类似，我们可以直接由 (C.4) 得到如下用于望远镜求和的基本结果：

$$\mathbb{E}[\gamma_k\langle F(X_k), X_k - p\rangle] \le \mathbb{E}[D(p, X_k)] - \mathbb{E}[D(p, X_{k+1})] + \Delta_k,$$

其中 $\Delta_k := \alpha_B D_{\mathcal{X}_\alpha}\gamma_k\delta_k + (\gamma_k\bar{g})^2/\tilde{\mu} + (\tilde{L}\gamma_{k-1}\bar{g})^2/\tilde{\mu}^3 + \Delta_k'$。由 F 在 X 上的单调性可得 $\langle F(p), \hat{X}_{k+1/2} - p\rangle \le \|F(p)\|_*(\|\hat{X}_{k+1/2} - X_{k+1/2}\| + \|X_{k+1/2} - X_k\|) + \langle F(p), X_k - p\rangle \le \bar{\nabla}\alpha_{\mathcal{X}}\delta_k + \bar{\nabla}\tilde{L}\bar{g}\gamma_{k-1}/\tilde{\mu} + \langle F(X_k), X_k - p\rangle$，其中 $\alpha_{\mathcal{X}}$ 的定义如上。令 $\Delta_k^\star := \Delta_k + \bar{\nabla}\alpha_{\mathcal{X}}\delta_k\gamma_k + \bar{\nabla}\tilde{L}\bar{g}\gamma_k\gamma_{k-1}/\tilde{\mu}$。同样，在 (15) 成立时，存在一个常数 M，使得对所有 $k \in \mathbb{N}_+$，增量序列为$\sum_{t=1}^k\Delta_t^\star \le M$。其余论证与上述 OMD 证明类似，区别仅在于此处常数 M 取值不同。□

E. $O(1/k^{1-\epsilon})$ 强伪单调博弈中所提出的算法的收敛速率

引理 E.1. 设$(a_k)^{k\in\mathbb{N}_+}$ 是一个非负序列。幂常数 s, t 满足 $0 < s < t < 1$ 和 $s + t > 1$。若序列 $(a_k)^{k\in\mathbb{N}_+}$ 满足递归线性不等式

$$a_{k+1} \le (1 - \frac{c}{k^s})a_k + \frac{d}{k^{t+s}},$$

对于某些正常数 c 和 d，则

$$a_k \le \frac{c_\star}{k^{t+s-1}} + \frac{d_\star}{k}, \forall k \ge K,$$

其中，常数指标 K 满足 $K > cK^{1-s} \ge 1$；两个系数选为：$c_\star := \frac{d}{\lfloor cK^{1-s}\rfloor - (t+s-1)}$以及 $d_\star := \max\{0, \frac{K}{(K-1)\tilde{a}_0}\}$，其中 $\tilde{a}_0 := a_K - \frac{c_\star}{K^{s+t-1}}$。

证明。以下内容主要借鉴了 [53,引理 1]。我们可以找到一个 K，使得 $K > cK^{1-s} \geq 1 > t+s-1$。为简单起见，令 $p \coloneqq t+s-1$ 和 $\tilde{c} \coloneqq \lfloor cK^{1-s} \rfloor$，其中 $\tilde{c}-p>0$和$\tilde{c}/K < 1$。从指标 K开始，递归不等式可放宽为 $a_{k+1} \leq (1-\frac{\tilde{c}}{k})a_k + \frac{d}{k^{t+s}}$。此外，注意到 $\frac{1}{k^p} - \frac{1}{(k+1)^p} \leq \frac{p}{k^{p+1}}$ $\frac{\tilde{c}}{k^{p+1}} - \frac{1}{k^p} + \frac{1}{(k+1)^p} \geq \frac{\tilde{c}-p}{k^{p+1}}$ $\frac{1}{(k+1)^p} - \frac{\tilde{c}}{k}\frac{1}{k^p} = 0$，这意味着 $0 \leq 0$。将 γ替换为上述结果可得：

$$a_{k+1} - \frac{d}{\tilde{c}-p}\cdot\frac{1}{(k+1)^p} \leq (1-\frac{\tilde{c}}{k})(a_k - \frac{d}{\tilde{c}-p}\cdot\frac{1}{k^p}).$$

若 $a_K \leq \frac{d}{\tilde{c}-p}\cdot\frac{1}{K^p}$，则对所有 $k \geq K$, $a_k \leq \frac{d}{\tilde{c}-p}\cdot\frac{1}{k^p}$ 均成立。否则，记 $\tilde{a}_k = a_{k+K} - \frac{d}{\tilde{c}-p}\frac{1}{(k+K)^p}$ $k \in \mathbb{N}$ $\tilde{a}_{k+1} \leq 0$为，并且我们有 $\frac{k+K-\tilde{c}}{k+K}\tilde{a}_k$ $\tilde{a}_k \leq \prod_{i=1}^{\tilde{c}\wedge k}\frac{K-\tilde{c}-1+i}{K+k-i}\tilde{a}_0 \leq \frac{(K-1)^{\tilde{a}_0}}{(k+K-\tilde{c})}$ 通过望远镜求和，$\frac{1}{K}\frac{1}{(K-1)}$ $\leq 0_{K-\tilde{c}}\tilde{a}_0 \cdot \frac{1}{k+K}$。结合上述所有结果，我们可以得到引理中所述的关于 a_k 的目标界。□

注 1. 与 *[53, 引理 4]* 和 *[54, 引理 5]*, 相比，这里的结果更为保守；根据后两者，我们可以推知，引理 *E.1* 中考虑的序列 $(a_k)^{k\in\mathbb{N}_+}$ 具有渐近收敛速率 $O(1/k^t)$。然而， *[53], [54]* 的结果是在“limsup”意义下得出的，因此不太有助于明确刻画收敛性。鉴于此，我们采用引理 *E.1* 中的结果；经过某个固定迭代次数 K后，该序列将以明确的速率收敛。

证明。 （定理 3 的证明）
算法 1。 由伪梯度 F 的强伪单调性可得

$$\langle F(X_{k+1/2}), X_{k+1/2} - x_*\rangle \geq \mu\|X_{k+1/2} - x_*\|^2 \geq$$
$$\frac{\mu}{2}\|X_k - x_*\|^2 - \mu\|X_{k+1/2} - X_k\|^2 \overset{(a)}{\geq} \frac{\mu}{\tilde{L}}D(x_*, X_k) - \mu\|X_{k+1/2} - X_k\|^2,$$

其中 (a) 由假设 4 中的类范数限制得到。然后，根据镜像映射的利普希茨连续性，平方范数 $\|X_{k+1/2} - X_k\|^2$ 可界定如下：

$$\|X_{k+1/2} - X_k\|^2 \leq \|\nabla\psi^*(\nabla\psi(X_k) - \gamma_k G_{k-1}) - \nabla\psi^*(\nabla\psi(X_k))\|^2$$
$$\leq \frac{1}{\tilde{\mu}^2}\gamma_k^2\|G_{k-1}\|_*^2.$$

将上述关系与 (C.1) 结合，并对不等式两边取期望，可得

$$\mathbb{E}[D(x_*, X_{k+1})] \leq (1-\frac{\mu\gamma_k}{\tilde{L}})\mathbb{E}[D(x_*, X_k)] + C_{e,1}\gamma_k\delta_k,$$

其中，最后一项 $C_{e,1}\gamma_k\delta_k$ 用于处理 (C.1) 中以更快速度衰减的其他项，假设 $\gamma_k = c_\gamma/(k+b_\gamma)^{a_\gamma}$ 且 $\delta_k = c_\delta/(k+b_\delta)^{a_\delta}$，其中 $0 < a_\delta < a_\gamma < 1$ 和 $a_\gamma + a_\delta > 1$。根据引理 E.1， $\mathbb{E}[D(x_*, X_k)]$ 从某次迭代 K 开始以 $1/k^{a_\gamma+a_\delta-1}$ 的速率衰减。回顾上述分析以及扰动动作的构造可知，对于某些常数 C_e,2 和 C_e,3, 有 $\|X_{k+1/2} - X_k\|^2 \leq \frac{1}{\tilde{\mu}^2}\gamma_k^2\|G_{k-1}\|_*^2 = C_{e,2}\gamma_k^2$ 且 $\|\hat{X}_{k+1/2} - X_{k+1/2}\|^2 = \|\delta_k u_k + \delta_k R^{-1}(p - X_{k+1/2})\|^2 = C_{e,3}\delta_k^2$。由于对于所有 p 和 x, $D(p,x) \geq \tilde{\mu}/2\|p-x\|^2$，因此我们立即得到，对于所有 $k > K$, $\mathbb{E}[\|\hat{X}_{k+1/2} - x_*\|^2] = M_1/k^{a_\gamma+a_\delta-1} + M_2/k$；其中 M_1, M_2 和 K 表示由 G 的性质以及所选取的 γ_k 和 δ_k 决定的某些常数。

算法 2. 利用引理 3 中的中间结果，我们首先注意到 $\langle F(X_k), X_k - x_*\rangle \geq \mu\|X_k - x_*\|^2 \geq \mu/\tilde{L}D(x_*, X_k)$。与 OMD 的思路相同，我们结合 (C.4)、 F 的强伪单调性以及上述关系，得到

$$\mathbb{E}[D(x_*, X_{k+1})] \leq (1-\frac{\mu\gamma_k}{\tilde{L}})\mathbb{E}[D(x_*, X_k)] + C_{e,1}\gamma_k\delta_k,$$

其中 $C_{e,1}\gamma_k\delta_k$ 处理所有衰减速度不慢于 $\gamma_k\delta_k$ 阶的项。沿用 OMD 的论证，我们也可以得出 RMD 的相同结论。□

参考文献

[1] T. Li、G. Peng、Q. Zhu 和 T. Başar, 《网络、博弈与学习的交汇：用于网络多智能体决策的博弈论框架》, IEEE 控制系统杂志，第 42 卷，第 4 期，第 35–67 页，2022 年。[2] S. Maharjan、Q. Zhu、Y. Zhang、S. Gjessing 和 T. Basar, 《智能电网中的可靠需求响应管理：一种斯塔克尔伯格博弈方法》, IEEE 智能电网汇刊，第 4 卷，132，2013 1，120第期，第 – 页。[3] Q. Zhu、Z. Han 和 T. Başar, 《智能电网分布式需求侧管理的微分博弈方法》, 载于《2012 IEEE 国际通信会议（ICC）》。IEEE，2012 年，第 3345–3350 页。[4] Z. Han、D. Niyato、W. Saad、T. Başar 和 A. Hjørungnes, 《无线与通信网络中的博弈论：理论、模型与应用》。剑桥大学出版社，2012 年。[5] Q. Zhu、Z. Yuan、J. B. Song、Z. Han 和 T. Basar, 《认知无线电多跳网络中的干扰感知路由博弈》, IEEE 通信领域精选期刊，第 30 卷，2015，2012 10，2006第期，第 – 页。[6] D. Q. Vu、K. Antonakopoulos 和 P. Mertikopoulos, 《不确定条件下的快速路由：通过指数权重在拥塞博弈中进行自适应学习》, 《神经信息处理系统进展》, 第 34 卷，第 14 708–14 720 页，2021 年。[7] J. F. Nash Jr, 《n人博弈中的均衡点》, 《美国国家科学院院刊》, 第 36 卷，第 1 期，第 48–49 页，1950 年。[8] J. V. Neumann 和 O. Morgenstern, 《博弈论与经济行为》（纪念版）。美国新泽西州普林斯顿：普林斯顿大学出版社，2007 年。[9] P. Mertikopoulos 和 Z. Zhou, 《在连续动作集和未知收益函数下的博弈学习》, 《数学规划》, 第 173 卷，第 1 期，第 465–507 页，2019 年。[10] P. Yi 和 L. Pavel, 《一种用于分布式广义纳什均衡计算的算子分裂方法》, Automatica，第 102 卷，第 111–121 页，2019 年。[11] T. Tatarenko、W. Shi 和 A. Nedić, 《用于分布式纳什均衡求解的梯度博弈算法的几何收敛性》, IEEE Transactions on Automatic Control，第 66 卷，第 11 期，第 5342–5353 页，2020 年。[12] L. Pavel, 《通过双增广算子分裂方法，在网络上利用部分决策信息求解分布式广义纳什均衡》, IEEE Transactions on Automatic Control，第 65 卷，1597，2019 4，1584第期，第 – 页。[13] M. Bianchi、G. Belgioioso 和 S. Grammatico, 《利用部分决策信息快速求解广义纳什均衡》, Automatica，第 136 卷，文章编号 110080，2022 年。[14] Y. Huang 和 J. Hu, 《利用部分信息计算分布式随机广义纳什均衡：一种增广最佳响应方法》, IEEE 网络系统控制汇刊，2022 年。[15] N. Cesa-Bianchi 和 G. Lugosi, 《预测、学习与博弈》。剑桥大学出版社，2006 年。[16] E. Hazan 等, 《在线凸优化导论》, 《优化领域基础与趋势》, 第 2 卷，第 3-4 期，第 157–325 页，2016 年。[17] G. M. Korpelevich, 《用于求解鞍点及其他问题的外梯度法》, Matecon，第 12 卷，第 747–756 页，1976 年。[18] A. Nemirovski, 《用于具有利普希茨连续单调算子的变分不等式和光滑凸凹鞍点问题的收敛速率为 O(1/t) 的近端方法》, SIAM 优化期刊，第 15 卷，第 1 期，第 229–251 页，2004 年。[19] Y. Nesterov, 《对偶外推及其在求解变分不等式和相关问题中的应用》, 《数学规划》, 第 109 卷，第 2 期，第 319–344 页，2007 年。[20] A. Juditsky、A. Nemirovski 和 C. Tauvel, 《利用随机镜像近端算法求解变分不等式》, 《随机系统》, 第 1 卷，第 1 期，第 17–58 页，2011 年。

[21] A. Kannan 和 U. V. Shanbhag，《用于伪单调随机变分不等式问题及其变体的最优随机外梯度方案》，《计算优化与应用》，第 74 卷，第 3 期，第 779–820 页，2019 年。[22] A. N. Iusem、A. Jofré、R. I. Oliveira 和 P. Thompson，《用于随机变分不等式的方差缩减外梯度法》，SIAM 优化期刊，第 27 卷，第 2 期，第 686–724 页，2017 年。[23] Z. Zhou、P. Mertikopoulos、A. L. Moustakas、N. Bambos 和 P. Glynn，《连续博弈中的镜像下降学习》，载于《2017 IEEE 第 56 届决策与控制年会（CDC）》。IEEE，2017 年，第 5776–5783 页。[24] Y.-G. Hsieh、F. Iutzeler、J. Malick 和 P. Mertikopoulos，《关于单次调用随机外梯度方法的收敛性》，《神经信息处理系统进展》，第 32 卷，2019 年。[25] G. Gidel、H. Berard、G. Vignoud、P. Vincent 和 S. Lacoste-Julien，《生成对抗网络的变分不等式视角》，arXiv 预印本 arXiv:1802.10551，2018 年。[26] W. Azizian、F. Iutzeler、J. Malick 和 P. Mertikopoulos，《随机变分不等式中乐观镜像下降的最后迭代收敛率》，载于《学习理论会议》。PMLR，2021 年，第 326–358 页。[27] Y. Malitsky，《用于单调变分不等式的投影反射梯度方法》，SIAM 优化期刊，第 25 卷，第 1 期，第 502–520 页，2015 年。[28] S. Cui 和 U. V. Shanbhag，《关于单调随机变分不等式问题的反射梯度法与分裂方法分析》，载于《2016 IEEE 第 55 届决策与控制会议（CDC）》。IEEE，2016 年，第 4510–4515 页。[29] A. D. Flaxman、A. T. Kalai 和 H. B. McMahan，《老虎机设定下的在线凸优化：无梯度的梯度下降》，载于《第十六届 ACM-SIAM 离散算法年会论文集》，2005 年，第 385–394 页。[30] A. V. Gasnikov、E. A. Krymova、A. A. Lagunovskaya、I. N. Usmanova 和 F. A. Fedorenko，《随机在线优化：单点与多点非线性多臂老虎机；凸情形与强凸情形》，《自动化与遥控》，第 78 卷，第 2 期，第 224–234 页，2017 年。[31] Y. Zhang、Y. Zhou、K. Ji 和 M. M. Zavlanos，《用于黑箱学习与控制的新型单点残差反馈预言机》，Automatica，第 136 卷，文章编号 110006，2022 年。[32] X. Chen、Y. Tang 和 N. Li，《利用高通和低通滤波器改进单点零阶优化》，载于《国际机器学习会议》。PMLR，2022 年，第 3603–3620 页。[33] M. Bravo、D. Leslie 和 P. Mertikopoulos，《凹N人博弈中的赌博机学习》，《神经信息处理系统进展》，第 31 卷，2018 年。[34] T. Tatarenko 和 M. Kamgarpour，《单调博弈中纳什均衡的赌博机在线学习》，arXiv 预印本 arXiv:2009.04258，2020 年。[35] ——，《关于基于收益的算法收敛至强单调博弈中纳什均衡的速率》，arXiv 预印本 arXiv:2202.11147，2022 年。[36] T. Lin、Z. Zhou、W. Ba 和 J. Zhang，《具有赌博机反馈的强单调博弈中的最优无悔学习》，arXiv 预印本 arXiv:2112.02856，2021 年。[37] P. Mertikopoulos、Y.-P. Hsieh 和 V. Cevher，《从随机逼近视角看博弈学习》，arXiv 预印本 arXiv:2206 03922，2022 年。[38] F. Facchinei 和 J.-S. Pang，《有限维变分不等式与互补问题》。Springer，2003 年。[39] G. Scutari、D. P. Palomar、F. Facchinei 和 J.-S. Pang，《凸优化、博弈论与变分不等式理论》，IEEE信号处理杂志，第 27 卷，第 3 期，第 35–49 页，2010 年。[40] S. Karamardian，《锥上的互补问题：具有单调和伪单调映射的情形》，《优化理论与应用期刊》，第 18 卷，第 4 期，第 445–454 页，1976 年。[41] A. S. Nemirovskij 和 D. B. Yudin，《优化中的问题复杂度与方法效率》，1983 年。[42] S. Bubeck，《机器学习的凸优化理论》，arXiv 预印本 arXiv:1405.4980，第 15 卷，2014 年。[43] A. Juditsky、J. Kwon 和 É. Moulines，《统一镜像下降与对偶平均》，《数学规划》，第 1–38 页，2022 年。[44] P. Mertikopoulos、B. Lecouat、H. Zenati、C.-S. Foo、V. Chandrasekhar 和 G. Piliouras，《鞍点问题中的乐观镜像下降：再多走一步（外梯度）》，载于《国际学习表征会议》，2019 年。[45] D. S. Kim、P. T. Vuong 和 P. D. Khanh，《强伪单调变分不等式的定性性质》，《优化快报》，第 10 卷，第 8 期，第 1669–1679 页，2016 年。[46] Q. Liu、Z. Guo 和 J. Wang，《用于约束伪凸优化及其动态投资组合优化应用的单层循环神经网络》，*NeuralNetworks*，第 26 卷，第 99–109 页，2012 年。[47] B. Gao 和 L. Pavel，《单调凹博弈中的连续时间折扣镜像下降动力学》，*IEEE Transactions on Automatic Control*，第 66 卷，第 11 期，第 5451–5458 页，2020 年。[48] Z. Jiang 和 J. Cai，《需求响应聚合商热负荷的博弈论控制》，载于*2021 American Control Conference (ACC)*。IEEE，2021 年，第 4141–4147 页。[49] T. Tatarenko 和 M. Kamgarpour，《一类凸博弈中的广义纳什均衡学习》，*IEEE Transactions on Automatic Control*，第 64 卷，第 4 期，第 1426–1439 页，2018 年。[50] A. Galbis 和 M. Maestre，《向量分析与向量微积分》。Springer Science & Business Media，2012 年。[51] R. Durrett，《概率论：理论与例题》。剑桥大学出版社，2019 年，第 49 卷。[52] H. Robbins 和 D. Siegmund，《非负几乎超鞅的收敛定理及其若干应用》，载于统计优化方法。Elsevier，1971 年，第 233–257 页。[53] K. L. Chung，《论一种随机逼近方法》，*The Annals of Mathematical Statistics*，第 463–483 页，1954 年。[54] B. T. Polyak，《优化导论，1987 年》，*Optimization Software, Inc,* 纽约。