多尺度重放：一种针对具有马尔可夫缓冲区的随机变分不等式的稳健算法

Milind Nakul*、Tianjiao Li‡, ◇、Ashwin Pananjady*, †

佐治亚理工学院，工业与系统工程学院* 与电气与计算机工程†麻省理工学院，斯隆管理学院‡

2026年1月6日

arXiv:2601.01502v1 [math.OC] 4 Jan 2026

摘要

我们提出了多尺度经验重放（MER）算法，用于解决一类随机变分不等式（VI）。在这类问题中，样本由马尔可夫链生成，并且我们可以使用记忆缓冲区存储这些样本。MER不再从缓冲区中均匀采样，而是采用多尺度采样方案，以模拟针对独立同分布样本设计的变分不等式算法的行为，克服实际串行方案中的偏差，从而加速收敛。值得注意的是，与标准的样本跳过型串行算法不同，MER具有稳健性：只要可能，它便能在无需了解马尔可夫链混合时间的情况下，实现迭代复杂度上的加速。我们还讨论了MER的应用，尤其是在使用时序差分学习进行策略评估，以及在相关数据下训练广义线性模型中的应用。

1 引言

变分不等式（VI）为对各个领域中广泛问题进行建模提供了一个通用的数学框架，例如博弈论、经济学、算子理论以及现代《机器学习》（Browder，1966；Rockafellar，1969；Bach 等，2008；Chambolle 和 Pock，2011；Ben-Tal 等，2009；Goodfellow 等，2020）。其一大优点在于能够统一处理优化问题和均衡问题。

给定一个闭凸可行域 $X \subseteq \mathbb{R}^n$ 和一个算子 $F\colon X \mapsto \mathbb{R}^n$，VI 问题旨在寻找一个 $x^* \in X$，使得

$$\langle F(x^*), x - x^* \rangle \geq 0. \quad \text{for all } x \in X. \tag{1}$$

在本文中，我们假设问题（ x^* 1）存在解。我们还假设 F 是一个 L-利普希茨连续算子，即对于某个 $L > 0$,

$$\|F(x_1) - F(x_2)\| \leq L\|x_1 - x_2\|, \quad \text{for all } x_1, x_2 \in X. \tag{2}$$

除非另有说明，本文中的记号 $\|\cdot\|$ 均表示欧几里得范数。

此外，我们假设 F 是一个广义强单调算子，即，对于某个 $\mu > 0$,

$$\langle F(x), x - x^* \rangle \geq \mu\|x - x^*\|^2, \quad \text{for all } x \in X. \tag{3}$$

显然，条件（3）在 F 是强单调的情况下成立，即：

$$\langle F(x_1) - F(x_2), x_1 - x_2 \rangle \geq \mu\|x_1 - x_2\|^2, \quad \text{for all } x_1, x_2 \in X.$$

◇Tianjiao Li 在这项工作的大部分开展期间就读于佐治亚理工学院。

请注意，我们假设的广义强单调性条件（3）弱于全局强单调性，但已经足以保证满足不等式（ x^* （1）的解是唯一的。

变分不等式的广泛应用推动了过去几十年里大量算法研究。求解单调变分不等式的经典算法包括（但不限于）梯度投影法（Sibony，1970；Bertsekas，1997）、超梯度法（Korpelevich，1976）和近端点算法（Rockafellar，1976）。大约二十年前，Nemirovski（2004）提出了镜像近端算法，这一突破激发了大量研究，致力于在确定性情形（Auslender 和 Teboulle，2005；Nesterov 和 Scrimali，2006；Nesterov，2007；Monteiro 和 Svaiter，2010；Malitsky，2015；Dang 和 Lan，2015）和随机情形（Juditsky 等人，2011；Nemirovski 等人，2009；Chen 等人，2017；Iusem 等人，2017；Yousefian 等人，2017；Cui 和 Shanbhag，2021；Kotsalis 等人，2022a,b）下求解变分不等式问题。

1.1 具有相关数据的随机变分不等式

在本研究中，我们采用后一种随机情形，此时只能获得关于算子 F 的不精确信息。具体而言，我们假设存在一个随机预言机，当~给定任意查询点 $x \in X$时，它会生成一个随机估计量 $F(x, \xi) \in \mathbb{R}^n$，其中 $\xi \sim \pi$是定义在某个概率空间 Ξ上的随机变量。该估计量满足以下性质：对于任意固定点 x，其边际上无偏，即

$$F(x) = \mathbb{E}[\widetilde{F}(x, \xi)] = \int_{\xi \in \Xi} \widetilde{F}(x, \xi) d\pi(\xi). \tag{4}$$

随机优化文献中的一个典型假设是，我们可以获得一系列样本（ξ_1、ξ_2，.....），这些样本独立同分布地取自 π。然而，在马尔可夫决策过程（MDP）和强化学习（RL）等重要的现实场景中，这一假设并不成立（Sutton 和 Barto，2018）。在这类情形下，常常会遇到更具挑战性的马尔可夫噪声情形，其中（ξ_t： $t = 1$，2，.....）是定义在状态空间 Ξ上的马尔可夫过程，而 π 表示该过程唯一的平稳分布。近年来，具有马尔可夫噪声的随机变分不等式在机器学习、统计学、优化和控制等领域的应用中受到了广泛关注，因为数据往往呈现序列相关性（Harold 等，1997；Spall，2005；Benveniste 等，2012；Meyn 和 Tweedie，2012）。这一情形下的主要挑战是，样本相关性会导致算子 F的估计量有偏。具体而言，令 $\mathcal{F}_{t-1}$ 表示由历史信息（ξ_1，.....，ξ_{t-1}）生成的~σ-域，则有 $\mathbb{E}[F(x, \xi_t)|\mathcal{F}_{t-1}] \neq F(x)$。条件~偏差 $\mathbb{E}[F(x, \xi_t)|\mathcal{F}_{t-1}] - F(x)$给快速收敛算法的开发带来了挑战。

尽管如此，针对相关数据的随机优化已得到广泛研究。我们无法涵盖这方面浩如烟海的文献，关于经典进展的讨论，可参阅 Duchi 等人 (2012) 及其中所引用的文献。与马尔可夫噪声下的变分不等式尤其相关的是强化学习中的策略评估文献；策略评估是一项基础且受到广泛研究的问题。其中最流行的算法称为时序差分（TD）学习，可视为针对某个经过适当构造的变分不等式进行的串行随机逼近。该算法最早由 Sutton (1988) 提出，多年来，TD 学习的理论分析取得了显著进展。Tsitsiklis 和 Van Roy (1996) 给出了最早的一批结果之一，证明了带函数逼近的 TD 学习具有渐近收敛性。其他关于渐近收敛的经典结果包括 Borkar 和 Meyn (2000) 以及 Borkar(2008) 的研究，他们使用常微分方程（ODE）框架分析了 TD 学习的长期行为。近期，Bhandari 等人(2018) 建立了有限时间性能

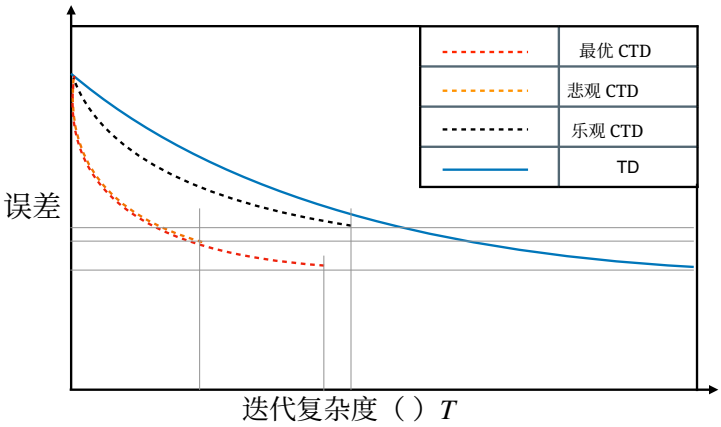


图 1. 展示误差随迭代次数变化的示意图，其中 T 表示可用样本数。蓝色曲线表示运行串行随机逼近算法的结果，我们将其标记为 TD。红色曲线表示 CTD 算法（Kotsalis 等人，2022b），其跳过参数设置正确，并且混合时间信息准确。该曲线在达到与 TD 相同的最终误差的同时，加快了迭代复杂度。橙色曲线表示一种悲观的 CTD 算法，其跳过参数设置得远大于实际混合时间。尽管该曲线在少数几次迭代中与最优 CTD 曲线重合，但它会早得多停止，并产生较大的最终误差，因为它会更早耗尽所有可用样本。黑色曲线表示 CTD 的一种乐观版本，其跳过参数小于混合时间。尽管它运行的迭代次数更多，但通常无法达到与 TD 相当的最终误差。

保证。其他值得关注的贡献包括 Srikant 和 Ying(2019)；Chen 等人 (2024)；Durmus 等人 (2021)；Min 等人 (2021)；Li 等人 (2024)；Wu 等人 (2025)。

广义而言，相关文献中的保证将误差分解为 $\|x_T - x^*\|^2$ 在第T 次迭代时的两个组成部分：其中一个部分刻画了即使没有噪声时算法的进展，称为确定性误差，通常会随 T呈指数衰减；已有多种方法被提出，用于加快该收敛过程中指数的变化速度，具体取决于问题的条件数（例如参见 Li 等人 （2023）及其中的参考文献）。另一个部分称为随机误差，用于刻画随机性对收敛的影响——在有噪声的情况下，这通常是衰减较慢的误差项，也是本文关注的重点。在此背景下，为了阐明本文的研究动机，下面更具体地介绍该领域中两个具有代表性的结果。

一方面，对于求解（投影）线性不动点方程的问题——这是线性算子变分不等式的一种特殊情形——Mou 等人(2024) 最近分析了采用迭代平均、基于马尔可夫数据的经典串行随机逼近算法。他们证明，随机误差的收敛率为 $\mathcal{O}(\bar{\tau}\sigma^2/T)$[1]，其中 σ^2 是问题中适当定义的噪声水平， $\bar{\tau}$ 是底层马尔可夫链的混合时间，而 T 是迭代总次数；由于该算法按串行方式执行，这也就是处理的样本总数。

[1]这里， σ^2 随 d线性增长，其中 d 是问题的维度。

另一方面，Kotsalis 等人(2022b) 提出了条件时序差分（CTD）算法，用于求解基于马尔可夫数据的随机变分不等式问题。他们通过跳过样本来处理数据相关性，并证明经过 T 次迭代后，随机误差可被界定为 $\mathcal{O}(\sigma^2/T)$，从而加速迭代复杂度，并获得与独立同分布样本情形相当的保证。然而，CTD 需要预先了解马尔可夫链的混合性质，以决定每次迭代要跳过多少个样本；遗憾的是，在现实应用中，这些性质几乎不可能事先完全掌握。这也并非纯粹的理论要求；在实现这些算法时，跳过参数设定过高或过低都可能导致不理想的收敛行为，图 1示意了这些问题。因此，一个关键的开放问题是，CTD 的加速收敛能否在不知道链的混合性质的情况下实现。事实上，这类问题通常归入自适应无参数算法这一广泛范畴，近期在优化和强化学习领域都受到了一定关注（例如，Dorfman 和 Levy，2022；Ju 和 Lan，2025）。

1.2 记忆缓冲区和经验回放

在本文中，我们利用了一种实践中通常可用的额外资源，它有助于设计出无需了解混合性质即可快速收敛的算法，即记忆缓冲区。具体而言，假设我们可以将样本存储在大小为 B的缓冲区中，并使用此前收集的样本向量启动算法

$$\xi^B = (\xi_t)_{t=1}^B, \tag{5}$$

并在（如果）收集到更多样本时更新该缓冲区。TD 等串行随机逼近算法仅按照因果顺序处理缓冲区中的所有样本，而 CTD 等基于跳过的算法则旨在处理缓冲区中的每τ 个样本。相反，我们能否在算法的每次迭代中策略性地从缓冲区中选择样本？需要注意的是，由于这些样本已被存储，算法可以灵活地按照任意顺序处理缓冲区中的条目。

记忆缓冲区在强化学习（RL）中得到广泛使用，由这一范式产生的一种非常流行的算法称为经验回放（ER），最初由 Lin（1992）提出。广义而言，大多数经验回放算法通过均匀随机采样数据点来处理缓冲区；当训练强化学习智能体并由其在线收集样本时，这些技术因能够打破数据中的时间依赖关系而广受欢迎（Mnih 等（2015）；Wang 等（2017）；Zhang 和 Sutton，2017；Fedus 等（2020）；Kumar 和 Nagaraj，2024）。尽管经验回放在实践中取得了成功，但人们对迭代方法中经验回放的理论理解仍处于起步阶段（例如，参见 Nagaraj 等（2020，2022）；Kowshik 等（2021）；Lazic 等（2021））。例如，Nagaraj 等（2022）利用“反向经验回放”的超鞅结构，推导了 Q 学习的样本复杂度界。Kowshik 等（2021）提出了一种利用反向经验记忆的算法，以解决线性系统辨识问题；而 Lazic 等（2021）则针对平均奖励马尔可夫决策过程，为带有回放缓冲区的正则化策略迭代提供了遗憾界。此外，Di-Castro 等（2022）深入分析了回放缓冲区的随机性质，确立了强化学习中演员-评论家算法的渐近收敛性。

尽管经验回放主要局限于强化学习领域，但这些思想尚未广泛应用于随机优化。此外，据我们所知，为了达到最优收敛速率，如何对缓冲区中的条目进行最优采样这一问题尚未得到深入研究

。本文旨在填补这一重要空白，致力于开发变分不等式算法，并将这些见解应用于变分不等式的具体实例（包括强化学习中的实例）。

1.3 贡献与文章结构

我们的主要贡献是提出并分析一种新算法，称为多尺度经验回放（MER），用于在具有记忆缓冲区的情况下，针对马尔可夫数据求解变分不等式中的随机优化问题。具体而言：

1. 我们的算法采用精心设计的采样方案，从记忆缓冲区中选取用于执行算法更新的条目。具体来说，算法按轮次运行，在每一轮中，使用彼此间隔一个回放间隔的样本更新迭代序列，而回放间隔会随轮次递减。重要的是，实现该方案无需任何关于马尔可夫链混合时间的信息。在标准假设下，我们证明了该算法的迭代复杂度界，表明其加速效果与最优地跳过样本相当，如 CTD 算法所示。

2. 除了建立误差上界外，我们还表明，只要马尔可夫链在 1-瓦瑟斯坦距离下混合，且随机算子满足适当的正则性，那么在最初几轮中，我们的算法模拟了在独立同分布数据上运行的对应迭代的行为。尽管已有非正式论述指出经验回放具有这种优势，我们通过严格建立一个双侧界，证明了我们算法的误差与其在独立同分布样本上运行的对应算法的误差保持一致。

3. 最后，我们将 MER 应用于广义线性模型以及强化学习中的策略评估问题。对于这两个问题，我们证明了其收敛速率与当前最先进的保证相当，但所用算法无需了解链的混合时间。我们还建立了关于模拟独立同分布设置的类似结果。

本文其余部分安排如下。在第2节中，我们介绍本文关注的变分不等式的若干示例。第3 节介绍用于变分不等式的 MER 算法。第4节介绍一般设置下的假设和主要结果。第5节讨论这些结果对示例的影响。第6节介绍验证理论结果的数值实验，第7 节汇总所有结论的证明。最后，我们在第8节讨论尚未解决的问题。

2 示例

回顾前文的设定，我们希望在闭凸可行域 x^* X 中找到一点，使其满足不等式（1）中的条件，其中样本存储在形式为（5）的马尔可夫缓冲区中。本节将介绍两个具体示例。

2.1 广义线性模型（GLMs）

考虑以下广义线性回归模型（也称为单指标模型），其中未知信号为 $x^* \in X$，观测值按照以下方式生成：

$$y_t = f(a_t^\top x^*) + v_t. \tag{6}$$

这里， $a_t \in \mathbb{R}^d$ 表示协变量的第 t 个样本， $y_t \in \mathbb{R}$ 表示相应的响应变量，标量 v_t 表示未观测到的零均值独立同分布噪声，且满足 $\mathbb{E}[v_t^2] = \sigma_v^2$。假设 f： $\mathbb{R} \to \mathbb{R}$ 是一个利普希茨连续且强单调的链接函数，并且对于 f 定义域中的所有 x、y，以下两组不等式成立： $|f(x) - f(y)|/|x - y| \le L_f$以及 $(f(x) - f(y))(x - y) \ge \mu_f(x - y)^2$。单调广义线性模型应用广泛，涵盖数据分析和信号恢复等领域（例如参见 Juditsky 和 Nemirovski，2020；Lou 等，2025；Pananjady 和 Foster，2021；Ilandarideva 等，2024；Zhu 等，2025；Haque 和 Maguluri，2024）。

现在假设协变量 $(a_t)_{t=1}^B$ 由具有有限混合时间的遍历马尔可夫链生成，且 a_t几乎必然有界，即对所有t，都有 $\|a_t\| \le D_a$ 几乎必然成立。此外，假设 a_1 取自该链的平稳分布 π ，并令$\Lambda := \mathbb{E}[a_t a_t^\top]$ 表示平稳协方差，满足 $\Lambda \succcurlyeq \kappa I$ ，其中 $\kappa > 0$为某个常数。

对于每个样本协变量 a_t，相应的观测值 y_t 由式（6）中的广义线性模型生成。为在我们的框架中表述信号恢复问题，注意到第 t个样本为元组 $\xi_t =$ （a_t， y_t），而我们的马尔可夫缓冲区收集了 B 个样本（a_t， y_t）$_{t=1}^B$。我们的目标是利用该缓冲区中的样本估计底层信号 x^* 。为求解 x^*，我们将其表述为一个随机变分不等式问题，其算子定义如下：

$$F(x) = \mathbb{E}_{a\sim\pi}\left[af(a^T x)\right] - \mathbb{E}_{a\sim\pi}\left[af(a^T x^*)\right] \text{ and } \widetilde{F}(x, \xi_t) = a_t f(a_t^T x) - a_t y_t. \tag{7}$$

根据定义，没有 意到 $F(x^*) = 0$，这意味着 Ineq. (1) 中的 VI 条件得到满足 d.

验证算子 F 的条件。注意

$$\begin{aligned}
\|F(x) - F(y)\| &= \|\mathbb{E}_{a\sim\pi}\left[af(a^T x) - af(a^T y)\right]\| \\
&\le \mathbb{E}_{a\sim\pi}\left[\|a\||f(a^T x) - f(a^T y)|\right] \\
&\overset{\text{(i)}}{\le} L_f \mathbb{E}_{a\sim\pi}\left[\|a\|\|a^T(x - y)\|\right] \\
&\overset{\text{(ii)}}{\le} L_f \mathbb{E}_{a\sim\pi}\left[\|a\|^2\right]\|x - y\| \\
&\le L_f \cdot D_a^2 \cdot \|x - y\|,
\end{aligned}$$

其中，步骤 (i) 源于链接函数 f 的利普希茨性，步骤 (ii) 则源于柯西–施瓦茨不等式。因此，算子 F 是利普希茨的，其参数为 $L_f \cdot D_a^2$。
我们还有

$$\begin{aligned}
\langle F(x), x - x^*\rangle &= \mathbb{E}_{a\sim\pi}\left[\langle af(a^T x) - af(a^T x^*), x - x^*\rangle\right] \\
&= \mathbb{E}_{a\sim\pi}\left[(f(a^T x) - f(a^T x^*)a^T(x - x^*))\right] \\
&\overset{\text{(i)}}{\ge} \mu_f \mathbb{E}_{a\sim\pi}\left[(a^T(x - x^*))^2\right] \\
&= \mu_f (x - x^*)^T \mathbb{E}_{a\sim\pi}[aa^T](x - x^*) \overset{\text{(ii)}}{\ge} \mu_f \cdot \kappa \cdot \|x - x^*\|^2,
\end{aligned}$$

其中，步骤 (i) 是由 f 的强单调性得出的，步骤 (ii) 则由假设$\Lambda \succcurlyeq \kappa I$得出。因此，算子 F 是参数为 $\mu_f \cdot \kappa$的广义强单调算子。

2.2 强化学习中的策略评估

策略评估通过马尔可夫奖励过程（MRP）来表述，该过程由元组 $(\mathcal{S}, P, R, \gamma)$ 描述，其中 $\mathcal{S}$ 表示状态空间，P 是转移矩阵，R 是奖励函数，$\gamma \in (0, 1)$ 是折扣因子；更多详情和背景信息请参见 Puterman (2014)；Bertsekas (2019) 等文献。在每个时间步，系统以概率 $P(s, s')$ 从当前状态 $s \in \mathcal{S}$ 转移到某个新状态 $s' \in \mathcal{S}$，同时智能体因这次转移获得奖励 $R(s, s')$。状态 s 的期望即时奖励定义为 $r(s) = \sum_{s' \in S} R(s, s')P(s, s')$。某个状态的价值函数定义为从该状态出发所获得的无限时域折扣奖励

$$v^*(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, s_{t+1}) | s_0 = s\right],$$

策略评估的目标是求解 v^*。令 $D = |\mathcal{S}|$ 表示状态数量，价值函数 v^* 和奖励函数 r 是 D 维实向量，并满足以下贝尔曼方程

$$v^* = r + \gamma P v^*. \tag{8}$$

我们假设马尔可夫链是遍历的，因此存在唯一的平稳分布 π，其各项均严格为正，并满足 $\pi P = \pi$。令 $\Pi := \text{diag}(\pi_1, \pi_2, \ldots, \pi_D) \in \mathbb{R}^{D \times D}$。

在状态空间较大的现代变体问题中，通常需要寻求贝尔曼方程（8）的近似解。具体而言，我们选择一个 d 维子空间，其定义为 $\mathbb{S} := \text{span}\{\psi_1, \ldots, \psi_d\}$，其中包含 d 个线性无关的基向量，并尝试通过 $v \in \mathbb{S}$ 投影贝尔曼方程求解，该方程可用矩阵向量记法写成如下形式。

定义矩阵 $\Psi := [\psi_1, \ldots, \psi_d]^T$，该方程可写为

$$\Psi\Pi(\bar{v} - r - \gamma P\bar{v}) = 0,$$

并且由于 $\bar{v} := \Psi^T\bar{\theta}$ 对于某个 θ 成立，投影贝尔曼方程可以等价地写成 $\mathbb{R}^d$：

$$\Psi\Pi\Psi^T\bar{\theta} = \Psi\Pi r + \Psi\Pi\gamma P\Psi^T\bar{\theta}. \tag{9}$$

为简单起见，我们为子空间 $\mathbb{S}$ 定义一组正交归一基向量 $\Phi \equiv [\phi_1, \ldots, \phi_d] := Q^{-1/2}\Psi$ 其中 $Q_{i,j} := \langle\psi_i\psi_j\rangle$。记 $\lambda_{\max}(Q)$ 和 $\lambda_{\min}(Q)$ 分别为矩阵 Q 的最大和最小特征值。我们可以将投影贝尔曼不等式（9）视为具有以下算子的变分不等式来求解：

$$F(\theta) = \Psi\Pi(\Psi^T\theta - r - \gamma P\Psi^T\theta) \text{ and } \widetilde{F}(\theta, \xi_k) = (\langle\psi(s_k), \theta\rangle - R(s_k, s_{k+1}) - \gamma\langle\psi(s_{k+1}), \theta\rangle)\psi(s_k).$$

根据上述算子 F 的定义，我们注意到，对于 θ，即投影贝尔曼不等式（9）的解，$F(\theta) = 0$。这意味着，对于 θ，不等式（1）中的变分不等式条件得以满足。

在马尔可夫设定下，给定来自马尔可夫奖励过程 $(s_t, s_{t+1}, R(s_t, s_{t+1}))_{t \geq 0}$ 的一条轨迹，其中 s_0 服从平稳分布。基于这些数据，我们构建一个包含样本的马尔可夫缓冲区

$$\xi^B = (\xi_t = (\psi(s_t), \psi(s_{t+1}), R(s_t, s_{t+1})))_{t=1}^B. \tag{10}$$

我们的目标是利用马尔可夫缓冲区中的样本估计 θ。

验证算子 F的条件。 可以很容易地验证， F是一个（$1+\gamma$）$\cdot\lambda_{max}$（Q）-利普希茨连续且（$1-\gamma$）$\cdot\lambda_{min}$（Q）-强单调算子（参见例如Li 等人，2023）。

3 多尺度经验回放

介绍了激励我们开展这项工作的示例后，我们现在将介绍用于求解带有马尔可夫缓冲区的随机变分不等式的主要算法。作为引出我们方法的铺垫，回顾一下，采用样本跳过的标准随机逼近（SA）（在Kotsalis 等人(2022b中称为 CTD）可表示为

$$x_{t+1}=\arg\min_{x\in X}\left\{\eta\langle\widetilde{F}(x_t,\xi_{t\bar{\tau}}),x\rangle+\frac{1}{2}\|x_t-x\|^2\right\},\ \text{for } t=1,2,\ldots. \tag{11}$$

这次更新使用间隔 $\bar{\tau}$ 步的样本；理想情况下，该值应与马尔可夫链的混合时间成正比。然而，如前所述并如图 1 所示，跳过参数 $\bar{\tau}$ 设置不当可能导致统计和计算效率低下。

为解决这一问题，我们提出以下多尺度经验回放（MER）算法。给定一个大小为 B的记忆缓冲区，MER 算法会谨慎地选择该缓冲区中的元素，使随机逼近更新所用的“活跃”样本彼此尽可能远离。该方法的正式描述见算法1。

算法1 多尺度经验回放

输入：记忆缓冲区 $\{\xi_1,\xi_2,\ldots,\xi_B\}$，总轮数 K。**对于** $k=1,\ldots,K$ **执行**独立于过去重新初始化变量 $x_1^{(k)}\in X$。定义回放间隔 $\tau_k=\frac{B}{2^k}$ 以及第 k 轮使用的样本量 $T_k=2^k$。**对于** $t=1,\ldots,T_k$ **执行**使用样本 $\xi_{t\tau_k}$ 执行一次随机逼近：

$$x_{t+1}^{(k)}=\arg\min_{x\in X}\eta_k\langle\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}),x\rangle+\frac{1}{2}\|x-x_t^{(k)}\|^2. \tag{12}$$

删除样本 $\xi_{t\tau_k}$，并将新样本追加到缓冲区末尾。

结束循环

结束循环

在实际操作中，记忆缓冲区会经过多轮遍历（轮次）；每轮开始时，我们都会重新初始化迭代变量。在第 k 轮中，回放间隔定义为 $\tau_k=B/2^k$，它决定了该轮内用于更新迭代变量的样本之间的间隔。因此，第 k 轮使用的样本总数为 $T_k=2^k$。回放间隔 τ_k 逐渐缩小算法运行的尺度。如果 $\tau_k=\bar{\tau}$ 对于某个 k，则算法会在该轮执行被跳过的随机逼近更新（11）。这一多尺度过程会持续进行，直到最后一轮恢复标准的串行随机逼近更新，此时不再跳过任何样本。

请注意，如所述，算法 1 适用于在线设置，在这种设置下，我们会持续观测到新的样本 ξ_{B+1}，ξ_{B+2}，.. 。在进行迭代时，我们希望将这些样本存入缓冲区。在每次

迭代中，算法处理当前样本并将其从缓冲区中移除，同时将新观测到的样本追加到缓冲区末尾。只需省略删除当前样本这一步，即可轻松地将该算法调整为适用于没有新样本可用的离线设置。在这种情况下，缓冲区保持静态，从而可以在多个周期中重复使用相同的数据点。我们的所有结果也都适用于静态设置，

4 理论结果

本节将给出算法 1中提出的 MER 方法的收敛保证。首先，我们将介绍并讨论分析所依赖的假设。

4.1 假设与简写符号

我们首先对最优解 x^* 作出一个温和的假设。

假设 1. 最优解 x^* 位于一个半径已知为 $D/2$ 的 ℓ_2 球内，即 $x^* \in \mathbb{B}_2(D/2)$。因此，我们在每个轮次开始时独立于过去，将迭代序列初始化在此球内，从而使得 $x_1^{(k)} \in \mathbb{B}_2(D/2)$。

~现在，我们对随机算子 F 作出以下温和的假设，并将在第 5 节中验证这些假设对于我们在第 2 节介绍的两个具体示例均成立。

~**假设 2**。 随机算子 F 在其第一个参数上是利普希茨连续映射，~即存在一个 $L_1 > 0$，使得对于任意 $\xi \in \Xi$，都有

$$\|\widetilde{F}(x_1, \xi) - \widetilde{F}(x_2, \xi)\| \leq \widetilde{L}_1 \|x_1 - x_2\|. \tag{13}$$

假设 3（随机算子方差）。 存在正标量 σ、$\zeta \geq 0$，使得对于每个点 $x \in X$ 和每一对自然数 $t' \geq t$，以概率1有：

$$\mathbb{E}\left[\left\|\widetilde{F}(x, \xi_{t'}) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}]\right\|^2 \middle| \mathcal{F}_{t-1}\right] \leq \frac{\sigma^2}{2} + \frac{\zeta^2}{2}\|x - x^*\|^2. \tag{14}$$

为了说明接下来的两个假设，注意，如果 ξ_t 独立于 x，且从平稳分布中抽取，则随机算子 F 是 F的无偏估计量，满足~$\mathbb{E}[F(x, \xi_{t'})] = F(x)$。然而，由于在我们的设定中噪声具有马尔可夫性质，随机算子 ~在给定过去的条件下是有偏的，即对于 $t' \geq t$，有 $\mathbb{E}[F(x, \xi_{t'})|\mathcal{F}_{t-1}] \neq F(x)$。接下来的两个假设确保马尔可夫链混合得足够快，从而能够适当控制最优点x^* 处的偏差，同时也能控制两个算子评估之差的条件偏差。

假设 4（最优解处的偏差）。 存在常数 $C_M > 0$ 和 $\rho_1 \in (0, 1)$，使得对于任意一对自然数 $t' \geq t$，以下不等式以概率 *1* 成立：

$$\|F(x^*) - \mathbb{E}[\widetilde{F}(x^*, \xi_{t'})|\mathcal{F}_{t-1}]\| \leq C_M \cdot \rho_1^{t'-t}. \tag{15}$$

假设 5（差值的偏差）。 存在常数 $C_B > 0$ 和 $\rho_2 \in$ （0，1），使得对于任意一对点 x、$y \in X$ 以及一对自然数 $t' \geq t$，以下结论以概率1成立：

$$\|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}] - F(y) + \mathbb{E}[\widetilde{F}(y, \xi_{t'})|\mathcal{F}_{t-1}]\| \leq C_B \cdot \rho_2^{t'-t} \cdot \|x - y\|. \tag{16}$$

鉴于它们与马尔可夫链混合性质密切相关，我们将 ρ_1 和ρ_2 在上述假设中称为混合参数。我们定义 $\rho := \max\{\rho_1, \rho_2\} \in (0,1)$，从而假设4 和假设5 均由公共混合参数 $\rho \in (0,1)$满足。我们还定义

$$C := \frac{C_M}{40} + C_B, \tag{17}$$

这组 合并了假设4 和假设5中的乘法混合常数。

陈述完假设后，我们接着定义一些额外的记号，以便阐述主定理及后续讨论。首先定义以下时间尺度参数 τ_M，它刻画了使算子 F 的偏差充分减小所需的时间步数，

$$\tau_M := \frac{\log(18C/\mu)}{\log(1/\rho)}. \tag{18}$$

该参数与马尔可夫链的混合速度有关，也涉及强单调性标量 μ。在本文中，我们将 τ_M 称为“有效混合时间”。我们还定义

$$\alpha_k := \frac{\tau_M}{\tau_k}. \tag{19}$$

请注意，参数 α_k 会随 k 单调演化。为了建立直观认识，将这一演化划分为以下两个阶段是很有帮助的：

- （$\tau_k > \tau_M$）当回放间隔大于有效混合时间时，我们有 $\alpha_k < 1$。粗略地说，在此轮次中使用的样本 k 应被视为近似独立同分布，这是由于假设 4和假设5中的混合条件。

- （$\tau_k \leq \tau_M$）当回放间隔小于有效混合时间时，我们有 $\alpha_k \geq 1$。此时的样本不可避免地存在相关性，并且这种相关性会随着 τ_k的减小而增强。

因此，参数 α_k 概括了随着周期推进，样本从（近似）独立样本向相关样本的转变。

为了定义最后几项记号，回顾假设 D 中的参数 1 并定义

$$\begin{aligned} \bar{L} &:= L + \widetilde{L}_1, \; M := \max\left\{2 + \frac{4C}{\mu}, \frac{40C_M}{\mu}\right\} \\ p_k &:= 1 + (\log T_k)^{-1} \cdot \log\left(\frac{\mu^2 M \max\{1, D^2\}}{3(6\sigma^2 + 4\|F(x^*)\|^2)}\right). \end{aligned} \tag{20}$$

这些参数将简化定理陈述中的符号。就实际意义而言，参数

$\bar{L}$ 结合了算子 F 和 F的 Lipschitz 常数。

4.2 MER 的收敛保证

在第 4.1 节陈述了假设并定义了参数后，我们便可以为 MER 方法提供收敛保证。

定理 1。 假设 *1*-*5*成立。如果缓冲区大小满足下界 $B \geq 32p_1\tau_M(\zeta^2+16\bar{L}^2)3\mu^2\log\left(32p_1\tau_M(\zeta^2+16\bar{L}^2)3\mu^2\right)$，并且我们采用步长序列

$$\eta_k := \min\left\{\frac{3\mu}{16(\zeta^2+16\bar{L}^2)}, \frac{p_k\log T_k}{\mu T_k}\right\} \qquad for\ k=1,\ldots,K, \tag{21}$$

then

$$\begin{aligned}\mathbb{E}\|x_{T_k+1}^{(k)}-x^*\|^2 \leq M&\left(1+\frac{3\mu^2}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)}\right)^{-T_k}\left(D^2+1\right)\\ &+\frac{20C_M\rho^{\tau_M+\tau_k-1}}{\mu}\left(1+\frac{3}{128\left(\alpha_k+1\right)}\right)\\ &+\frac{3(2+p_k\log T_k)}{\mu^2T_k}\left(\alpha_k+1\right)\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right).\end{aligned} \tag{22}$$

详见第 7.2节，了解该定理的详细证明。

这里有几点说明。首先，请注意，为了设定公式（21）中的步长序列，我们需要获取一些参数[2]，这些参数取决于总体算子 F，但关键是不取决于 ρ 或 τ_M，后两者定义了链的混合性质。这与该领域中的其他算法（例如Kotsalis 等人，2022b）形成了鲜明对比，也揭示了在我们所处的设置中能够使用记忆缓冲区的一项关键优势。

接下来讨论保证（22）中的误差项。不等式（22）右侧的第一项呈线性收敛，对应于 MER 方法的确定性误差——即使问题中不存在随机性，这一项也依然存在。通过应用条件数 L/μ 上的算子外推技术（Kotsalis 等人（2022a，b），这一项对条件数的依赖关系很可能可以得到改善；如前所述，实现确定性误差的最优加速收敛超出了本文的研究范围。不等式（22）右侧的第二项源于最优点处的偏差 x^*。由于该项随 τ_M 指数衰减，因此通常可以忽略。

如前所述，我们关注的是不等式（22）右侧的最后一项，它是随机误差的上界。具体而言，当 B 至少达到 $T_k\tau_M$的量级时，或者等价地，当$\alpha_k = \mathcal{O}$（1）时，该随机误差项可简化为

$$\mathcal{O}\left(\frac{3(2+p\log T_k)}{\mu^2T_k}\left(6\sigma^2+4\|F(x^*)\|^2\right)\right), \tag{23}$$

这几乎与处理 T_k 独立同分布样本时的随机误差项相匹配。反之，当 $\alpha_k \gg 1$时，随机误差（23）会被放大α_k倍，具体取决于有效混合时间与回放间隔之比。不过请注意，我们始终有 $\alpha_k \leq \tau_M$，因此该放大倍数至多与有效混合时间同阶。事实上，除 $\alpha_k = \tau_M$时的最后一轮次外，MER 在 Kotsalis 等人（2022b）的标准马尔可夫随机逼近算法基础上，性能提升了回放间隔 τ_k倍。当最后一轮次的回放间隔满足 $\tau_k = 1$时，我们也能得到Kotsalis 等人（2022b）中的保证。

[2]存在一些依赖 取决于这一对（C，C_M），但与 ρ和 τ_M相比，这些量对混合并没有那么关键。

4.3 跟踪独立同分布误差

在上一节中，我们证明了，当回放间隔 τ_k 大于有效混合时间 τ_M时，MER 算法的收敛速率类似于使用 T_k 个样本进行独立同分布采样的情形。尽管这一结果令人鼓舞，但这仅仅是对上界的比较，并未回答 MER 是否确实模拟了独立同分布行为这一问题。本节将分析 MER 算法的误差——尤其是在最初几个轮次中——是否以及如何跟踪在独立同分布样本序列上运行简单随机逼近时的误差。

在陈述这一保证之前，我们需要一些额外的定义和假设。首先，在度量空间 (Ξ, r) 上定义 Wasserstein 距离。对于定义在 Ξ 上的一对概率分布 (μ, μ')，令 $\mathcal{P}(\mu, \mu')$ 表示 μ 和 μ' 的所有可能耦合构成的空间。对于任意 $q \geq 1$， μ 和 μ' 之间的 Wasserstein-q 距离定义为

$$\mathcal{W}_{q,r}(\mu, \mu') = \left\{ \inf_{\gamma \in \mathcal{P}(\mu,\mu')} \int_{\Xi \times \Xi} r(x,y)^q d\gamma(x,y) \right\}^{1/q}. \tag{24}$$

除非另有说明，本文始终取度量 r 为欧几里得距离。现在我们作出如下假设：该链在 Wasserstein 距离下具有快速混合性质。

假设 6 （Wasserstein 混合）。 设有效混合时间 τ_M 定义于公式 （18）中，对于任意 x， $y \in X$，存在通用正常数 c_0 使得

$$\mathcal{W}_1(\delta_x P^{\tau_M}, \delta_y P^{\tau_M}) \leq c_0 \|x - y\|, \tag{25}$$

其中， δ_x 表示将全部概率质量集中在 x 上的概率分布，并回顾一下， P 是该马尔可夫链的转移矩阵。

我们还要求随机~算子 F的第二个参数满足额外的利普希茨连续性假设。

~**假设 7**。 随机算子 F 关于其第二个参数是一个利普希茨连续映射，~也就是说，存在一个正数常量 L_2，使得对任意 $x \in X$，都有

$$\|\widetilde{F}(x, \xi_1) - \widetilde{F}(x, \xi_2)\| \leq \widetilde{L}_2 \|\xi_1 - \xi_2\|. \tag{26}$$

在最初的几个轮次中，回放间隔 τ_k 很大。具体而言，为了陈述下面的保证，假设对于某个 $\beta > 1$，有 $\tau_k = \beta \tau_M$。令 x_{T+1}表示从某个初始化 T出发，在第 12 轮应用 MER 更新（式 k）并经过 x_1步后得到的迭代点。特别地，在给定 x_1的条件下， x_{T+1} 中的随机性完全来自更新中使用的马尔可夫样本 $(\xi_{k\beta\tau_M})_{k=1}^{T}$。为了将该迭代点与独立同分布对应项进行比较，定义 $\widetilde{x}_{T+1}$为从相同的初始化 x_1 出发并应用 T步更新后得到的迭代点

$$\widetilde{x}_{t+1} = \arg\min_{x \in X} \left\{ \eta \langle \widetilde{F}(\widetilde{x}_t, \widetilde{\xi}_{t\beta\tau_M}), x \rangle + \frac{1}{2} \|\widetilde{x}_t - x\|^2 \right\}, \text{ for } t = 1, 2, \ldots,$$

其中 $(\widetilde{\xi}_{k\beta\tau_M})_{k=1}^{T}$ 表示从平稳分布 π 中抽取的独立同分布样本序列。~将两种情形下的误差分别定义为 $\Delta_{T+1} := x_{T+1} - x^*$和 $\Delta_{T+1} := \widetilde{x}_{T+1} - x^*$。

定理 2. 假设 2,6和 7 成立。若步长参数的选取满足$\eta \leq \frac{1}{\widetilde{L}_1 T}$，则

$$\left|\mathbb{E}\|\Delta_{T+1}\| - \mathbb{E}\|\widetilde{\Delta}_{T+1}\|\right| \leq 3c_0 \cdot \frac{\widetilde{L}_2}{\widetilde{L}_1} \cdot \sqrt{T} \cdot e^{-\beta}. \tag{27}$$

因此，若 $\beta = \log T^{3/2}$，则算法误差之间的差异是低阶项，并满足

$$\left|\mathbb{E}\|\Delta_{T+1}\| - \mathbb{E}\|\widetilde{\Delta}_{T+1}\|\right| \leq \frac{3c_0\widetilde{L}_2}{T\widetilde{L}_1}. \tag{28}$$

证明见第 7.4节。

定理 2 从定性角度表明，当回放间隔 τ_k 足够大时（具体来说，当 β 较大时），我们在马尔可夫数据缓冲区上使用 MER 算法所得到的误差，与标准随机逼近在独立同分布样本序列上的误差十分相似。这一发现直接证实了任何经验回放策略所追求的核心优势——即使面对相关数据，它也能模拟独立同分布的行为。因此，只要标准随机逼近算法在独立同分布序列上达到的收敛速率快于最坏情况下的马尔可夫界，那么只要轮数 k 不过大，MER 算法就能复现这一较快的速率。

5 特定示例的推论

在确立了一般性结果之后，我们现在将其应用于第2节讨论的广义线性模型和策略评估示例。在这些示例中，通常假设马尔可夫链在全变差（TV）距离下混合。概率分布 μ'和 μ在 Ξ上的全变差距离定义如下:

$$d_{\mathsf{TV}}(\mu, \mu') = \frac{1}{2}\int_{\xi\in\Xi} |d\mu(\xi) - d\mu'(\xi)|. \tag{29}$$

对于具有唯一平稳分布 π 的遍历马尔可夫链，记给定 $\mathcal{F}_{t-1}$ 时，随机变量 $\xi_{t'} \in \Xi$ 在时间步 t' 的条件分布为 $P_{t-1}^{t'}$。回顾一下， $\mathcal{F}_t$ 表示由随机变量 $\{\xi_k\}_{k=1}^t$ 生成的 σ- 代数。一种常用的全变差混合假设如下:

假设 8. 对于 $t' \geq t$，存在正常数 C_{TV} 和 $\rho_{TV} \in (0,\ 1)$，使得

$$d_{\mathsf{TV}}(\pi, P_{t-1}^{t'}) \leq C_{TV} \cdot \rho_{TV}^{t'-t} \quad w.p.\ 1.$$

5.1 广义线性模型

重新审视第 2.1 节中的广义线性模型示例，我们的目标是利用给定的马尔可夫缓冲区 x^*估计底层信号 ξ^B。定理1 给出了广义线性模型的以下推论。

推论 1。 假设 *1* 成立。此外，假设生成观测值的马尔可夫链 $(a_t, y_t)_{t\geq 1}$ 在全变差度量下呈几何混合，因此满足假设 *8*。选择定理*1*规定的步长，我们有

$$\mathbb{E}\|x_{T_k+1}^{(k)} - x^*\|^2 \leq M\left(1 + \frac{3\mu_f^2\kappa^2}{(\alpha_k+1)\,8(\zeta^2+64L_f^2D_a^4)}\right)^{-T_k}\left(D^2+1\right) + \frac{18(2+p_k\log T_k)}{\mu_f^2\kappa^2 T_k}\left(\alpha_k+1\right)\sigma^2.$$

这里是 $\sigma^2 = 6D_a^2\sigma_v^2$ 和 $\zeta^2 = 12L_f^2D_a^4$。

我们在第 1节7.5证明推论。

与先前的结果相比，我们注意到界中的确定性误差项以线性速率收敛，这与 Nagaraj 等人（2020） [定理 1]中给出的下界相符。对于有效混合时间 τ_M （通过 α_k）的依赖关系，就该项而言是信息论最优的。关于对问题维度 d的依赖关系，我们观察到参数 σ^2 通常随 d线性增长。因此，MER 算法用于广义线性模型的随机误差形式为

$$\widetilde{\mathcal{O}}\left(\frac{d\sigma_d^2}{T_k}\left(\frac{\tau_M}{\tau_k}+1\right)\right),$$

其中 σ_d^2 表示问题中与维度无关的噪声水平（参见 Nagaraj 等人，2020，定理 3）。此外，推论 1 意味着，为了在 ϵ年第 k轮（即 $\mathbb{E}\|x_{T_k+1}^{(k)} - x^*\|^2 \leq \epsilon$）中求得广义线性模型信号估计的 ϵ-最优解，所需的迭代复杂度为:

$$\mathcal{O}\left(\max\left(\frac{(\alpha_k+1)\bar{L}^2}{\mu_f^2}\log\frac{(D^2+1)}{\epsilon}, \frac{\sigma^2(\alpha_k+1)}{\epsilon\mu_f^2}\log\frac{1}{\epsilon}\right)\right).$$

该迭代复杂度上界在关于依赖于 ϵ（Kotsalis等人，2022b）的意义上，除一个对数因子外，几乎是最优的。

现在我们来说明 MER 算法对广义线性模型的独立同分布误差的跟踪效果。我们沿用前面的记号： Δ_{T+1}表示在回放间隔为 $\tau_k = \beta\cdot\tau_M$的情况下，使用马尔可夫样本运行 MER 算法 T步、针对 $\beta = \log\, T^{3/2}$的误差； $\widetilde{\Delta}_{T+1}$表示使用独立同分布样本运行标准随机逼近 T 步的误差。

我们还假设 X 的 ℓ_2 直径至多为 D_x ，且观测空间有界，即对所有 t，都有 $|y_t| \leq D_y$ 。由于 f 是利普希茨连续的，且协变量几乎必然有界，因此存在某个 $D_f \geq 0$ ，使得

$$\sup_{\substack{a:\|a\|\leq D_a \\ x\in X}} |f(a^\top x)| \leq D_f. \tag{30}$$

推论 2。假设上述假设成立。此外，假设生成观测值的马尔可夫链（a_t， y_t）$_{t\geq 1}$ 在 *Wasserstein*-1 度量下是混合的，因此满足假设*6*。采用定理*2*中的参数选择，我们有

$$\left|\mathbb{E}\|\Delta_{T+1}\| - \mathbb{E}\|\widetilde{\Delta}_{T+1}\|\right| \leq \frac{3c_0\max(L_fD_aD_x + D_f + D_y, D_a)}{TL_fD_a^2}.$$

我们在第2 节7.6中证明推论。 据我们所知，这是首个针对广义线性模型，在马尔可夫设置和独立同分布设置下证明随机逼近迭代序列彼此接近的结果。

5.2 强化学习中的策略评估

回顾第2.2节中强化学习策略评估的例子，我们希望找到投影贝尔曼方程（9）的固定点。定理1 给出了该问题的以下推论。

推论 3. 假设 *1* 成立，且特征向量有界，即对于所有 $s \in \mathcal{S}$，$\|\psi(s)\| \leq D_\psi$，并且奖励函数有界，即对于所有 s，$s' \in \mathcal{S}$，$R(s, s') \leq \bar{R}$。此外，我们假设底层马尔可夫链在全变差度量下几何混合，因此满足假设 *8* 。在与定理*1*相同的前提下，我们有

$$\begin{aligned}\mathbb{E}\|\theta_{T_k+1}^{(k)} - \bar{\theta}\|^2 \leq M &\left(1 + \frac{3(1-\gamma)^2\lambda_{min}(Q)^2}{(\alpha_k+1)\,8(\zeta^2 + 16(1+\gamma)^2(D_\psi^2 + \lambda_{max}(Q))^2)}\right)^{-T_k}(D^2+1) \\ &+ \frac{20C_M\rho^{\tau_M+\tau_k-1}}{(1-\gamma)\lambda_{min}(Q)}\left(1 + \frac{3}{128\,(\alpha_k+1)}\right) \\ &+ \frac{3(2+p_k\log T_k)}{(1-\gamma)^2\lambda_{min}(Q)^2T_k}\,(\alpha_k+1)\,(6\sigma^2 + 25C_M^2\rho^{2(\tau_k-1)} + 4\|F(\bar{\theta})\|^2).\end{aligned}$$

这里 $\sigma^2 = 48(1+\gamma)^2D_\psi^4D^2 + 48D_\psi^2\bar{R}^2$ 和 $\zeta^2 = 12(1+\gamma)^2D_\psi^4$。

我们在第 3 节证明推论 7.7。

与前文一样，忽略第二项后，在第 k 轮次中，为策略评估找到一个 ϵ-最优解所需的样本复杂度为，使得 $\mathbb{E}\left[\|\theta_{T_k+1}^{(k)} - \bar{\theta}\|^2\right] \leq \epsilon$

$$\mathcal{O}\left(\max\left(\frac{\alpha_k+1}{(1-\gamma)^2}\log\frac{D^2+1}{\epsilon}, \frac{\bar{\sigma}^2(\alpha_k+1)}{(1-\gamma)^2\epsilon}\log\frac{1}{(1-\gamma)^2\epsilon}\right)\right),$$

其中 $\bar{\sigma}^2 = 6\sigma^2 + 25C_M^2\rho^{2(\tau_k-1)} + 4\|F(\bar{\theta})\|^2$。复杂度界中的第一项对应确定性误差，其对折扣因子 $^{1}{}_{(1-\gamma)^2}$，呈二次依赖关系，因而并非最优。进一步地，可以采用Li 等人(2023) 中的加速技术，以实现对折扣因子的最优依赖关系。复杂度界中主要关注的是最后一项，即随机误差；就其对 ϵ的依赖关系而言，除一个对数因子外，该项已近乎最优。对折扣因子的依赖关系也可以通过方差缩减技术进一步改善；例如，参见 Khamaru 等人 (2021)；Li 等人(2023)。然而，关键在于，给出的收敛速率无需了解混合时间即可达到。

下一个推　论表明，MER 算法在初始轮次跟踪独立同分布误差　　s.

推论 4。 假设底层马尔可夫链在 *Wasserstein*-1 度量下是混合的，因此满足假设 *6* 。另假设特征向量有界，即对于所有$s \in \mathcal{S}$，$\|\psi(s)\| \leq D_\psi$且奖励函数有界，即对于所有 s，$s' \in \mathcal{S}$ $R(s, s')$ $\leq R$，并采用与定理 *2*和 $\beta = \log T^{3/2}$相同的参数选择。则有如下结论

$$\left|\mathbb{E}\|\Delta_{T+1}\| - \mathbb{E}\|\widetilde{\Delta}_{T+1}\|\right| \leq \frac{3c_0\left((2+\gamma)\cdot D_\psi\cdot D_\theta + \bar{R}\right)}{T(1+\gamma)D_\psi^2}.$$

我们在第4 节7.8中证明推论。与广义线性模型的情形一样，我们认为这是文献中的首个此类结果。

6 数值结果

本节报告 MER 算法的数值实验，并将其性能与标准串行SA和采用样本跳过的SA进行比较。我们考察逻辑回归问题和策略评估问题。此外，我们还报告了SA算法在独立同分布数据上的结果，以评估各方法模拟独立同分布设置的效果。

6.1 带有马尔可夫噪声的策略评估

我们首先考虑策略评估问题，并将 MER 算法与 TD 和 CTD 进行比较；在此设置下，它们分别对应于不跳过和跳过的随机逼近。我们考虑参数为 m 的马尔可夫奖励过程，该参数控制底层马尔可夫链的混合时间，而所有运行中的折扣因子固定为 $\gamma = 0.8$ 。给定状态空间 $\mathcal{S}$，转移核 P 定义为

$$P(s'|s) = \begin{cases} \frac{2m-1}{m}, & s' = s \\ \frac{1-m}{(|\mathcal{S}|-1)m}, & s' \neq s \end{cases}.$$

奖励函数为 $r(s) = \mathbb{I}\{s \geq d\}$，其中 d 是特征向量的维度。该链的平稳分布在所有状态上均匀分布，且 P 的谱隙为 $\frac{|\mathcal{S}|}{(1-m)\ (|\mathcal{S}|-1)m}$。因此，有效混合时间 τ_M 的阶为 $\frac{m}{1-m}$：随着 ~~m 接~~近 $1/2$，该链会快速混合并趋近于独立同分布状态；而较大的 m 值会导致混合速度变慢，马尔可夫依赖也更加明显。当 $m = 1$ 时，该链完全停滞，以概率 1 保持在同一状态。

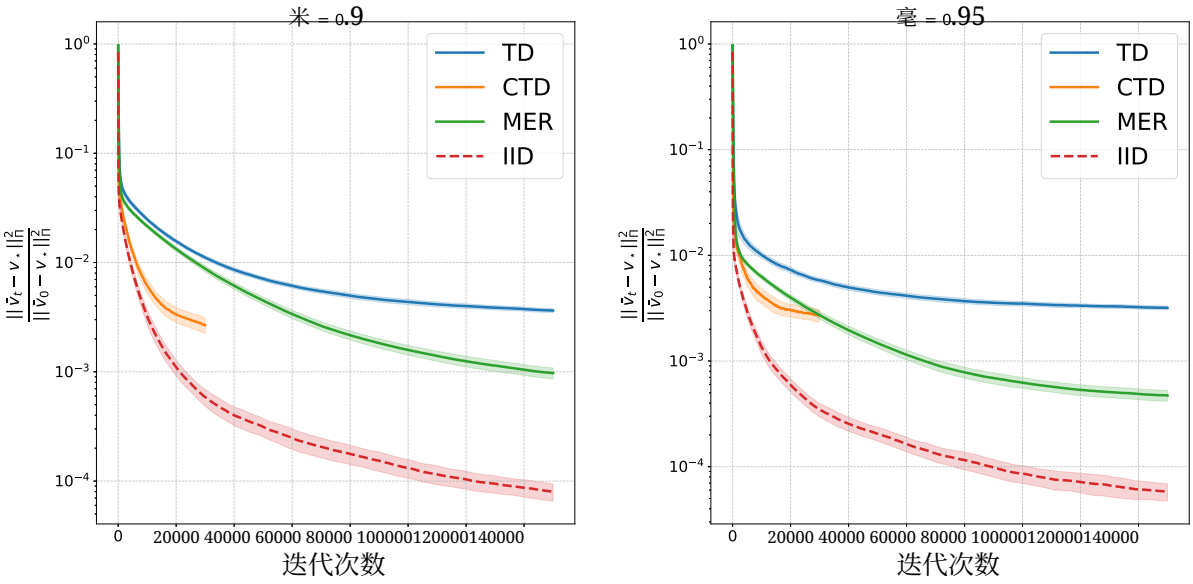
特征向量定义为

$$\phi(s) = \mathbf{e}_{s \wedge d} \in \mathbb{R}^d,$$

其中$\mathbf{e}_i$ 是第 i 个标准基向量。我们固定 $|\mathcal{S}| = 30$ 和 $d = 16$。对于每个 m 值，我们绘制归一化平方误差 $\frac{\|\theta_{T_k} - \theta^*\|^2}{\|\theta_0 - \theta^*\|^2}$，其中 θ_{T_k} 是~~平均输出~~，θ^* 是最优解，θ_0 是初始化。

为生成独立同分布曲线，我们在每个时间步从平稳分布中独立采样，并使用该样本应用时序差分算法。对于 MER、时序差分和 CTD 算法，结果均使用大小为 150000 的静态马尔可夫缓冲区获得。所有实验中，CTD 的跳过参数均设为 5。因此，使用我们的缓冲区大小时，CTD 算法在每种情况下都会因跳过而使用 30000 个样本。

对于 $m = 0.9$ （图2(a)），链的混合速度很快。由于采用了跳过策略，CTD 算法在运行的 30000 次迭代中，表现得与独立同分布数据流非常接近，但由于所有可用样本均已耗尽，误差无法进一步降低。TD 算法最终也能达到相近的误差，但其误差随迭代次数的下降明显较慢。MER 算法逐次迭代的表现显著优于 TD，同时最终误差的平稳值低于 CTD 和 TD（且更接近独立同分布误差）。当 m 增至 0.95（图2(b)）时，链的混合速度变慢，马尔可夫依赖变得显著。此时，CTD 使用的固定跳过参数 5 远小于有效混合时间，导致 CTD 的表现显著下降。MER 算法的表现仍与 $m = 0.9$时相近，说明其表现对混合参数不那么敏感。在这种情况下，TD 算法的效果仍然最差。



(a) (b)

图2. MDP 示例中各算法的比较。在 y 轴上，我们绘制了 50 个实例的欧几里得范数 $\|\cdot\|_\Pi$ 比值的平均值。

总之，CTD 的表现对跳过参数的选择高度敏感，而无论混合时间如何，MER 都能持续达到接近独立同分布情形的表现。正如预期，对于 m的两个取值，TD 的表现都是最差的。

6.2 带自回归协变量的逻辑回归

现在，我们在逻辑回归任务上测试这些算法。广义线性模型（GLM）的链接函数取为 S 形函数：

$$f : z \mapsto \frac{\exp(z)}{1+\exp(z)}.$$

协变量 $(a_t)_{t\geq 1}$ 由自回归过程生成

$$a_{t+1} = Aa_t + \epsilon_t, \text{ for } t = 1, 2 \ldots, \tag{31}$$

其中，$\epsilon_t \in \mathbb{R}^n$ 是均值为 0、方差为 10^{-2} 的独立同分布高斯随机变量序列。第一个协变量 a_1 通过将过程初始化为 0 并进行 10000 步的预热期得到。

矩阵 A 被构造为稳定对称矩阵。协变量中马尔可夫依赖的程度通过 A 中接近 1 的特征值数量来控制，我们将其记为 d_{L}。d_{L} 值越大，混合速度越慢。具体而言，给定 d_{L}，我们通过以下步骤生成A：

1. 特征值生成：我们生成一个特征值向量，其中前 d_{L} 个元素设为 0.995，其余特征值则从均值为 0 的高斯分布中随机选取

，方差为 10^{-4}。这确保其余 $d-d_{\mathrm{L}}$ 个特征值接近 0。所有特征值都汇集在一个 $d \times d$ 对角矩阵 E 中。

2. 矩阵构造：现在生成均匀随机的 $d \times d$ 正交矩阵 Q，并通过特征值分解得到A

$$A = QEQ^T.$$

生成矩阵后 A，根据式（a_t）$_{t\geq 1}$ 中描述的自回归过程生成协变量（31），并根据式（y_t）$_{t\geq 1}$ 中的广义线性模型生成相应的观测值（6），使用 S 形函数作为链接函数。

我们将协变量维度固定为 $d = 1000$，样本数量设为 $n = 10000$，并报告 $d_{\mathrm{L}} = 400$ 和500时的结果。所有实验中，CTD算法的跳过参数均设为10。因此，对于固定大小为10000个样本的缓冲区，CTD算法由于跳过样本，实际仅使用1000个样本。对于每个 d_{L}的取值，我们绘制归一化平方误差 $\frac{\|x_{T_k}-x_*\|^2}{\|x_0-x_*\|^2}$，其中 x_{T_k} 为输出，x_* 为最优解，x_0 为初始化。

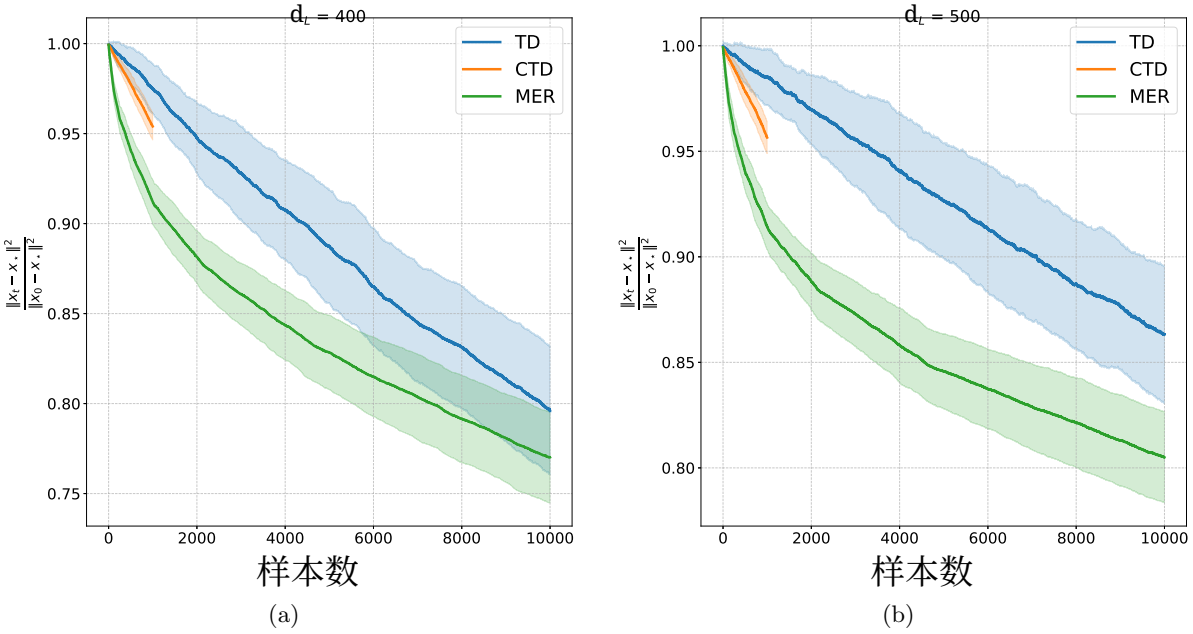


图3。对比比逻辑回归示例中的算法。在 y轴上，我们绘制出比率欧几里得范数 $\|\cdot\|_2$ 在50个实例上的平均值。

图3(a) 和3(b) 表明，即使链的混合速度较慢，MER算法的性能也始终优于TD和CTD。与之前一样，MER取得这一性能并不依赖任何混合信息。

7 证明

本节将给出所有主要结果的证明。

需要指出的是，尽管 MER 算法使用动态缓冲区，其中的样本会被删除和追加（如算法1所示），但针对静态缓冲区情形的理论收敛分析可直接扩展到这一设置。其关键依据在于对迭代变量进行逐轮重新初始化 $x_1^{(k)}$ 。在每一轮 k开始时，该变量的初始化独立于其先前的轨迹。这一机制确保在整个第 k轮中，算法实际上都在固定的记忆缓冲区快照上运行。在动态缓冲区中，所用样本之间的间隔甚至更大，因此，在静态缓冲区假设下推导出的性能保证对每一轮均成立。因此，本文针对静态情形建立的收敛保证，对于算法1的动态缓冲区实现仍然全局有效。

7.1 初步辅助引理

我们首先陈述并证明一些在分析的多个部分中都有用的引理。

引理 1. 假设 *4* 和 *5* 成立。对于每个 t， $t' \in \mathbb{Z}_+$ 和 $x \in X$，以概率 1，

$$\|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}]\| \leq C_M \cdot \rho^{t'} + C_B \cdot \rho^{t'} \|x - x^*\|. \tag{32}$$

因此，以概率 *1*，

$$\left\langle F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}], x - x^* \right\rangle \leq C_M \cdot \rho^{t'} \|x - x^*\| + C_B \cdot \rho^{t'} \|x - x^*\|^2. \tag{33}$$

证明。 令 $y = x^*$ ，代入不等式 (16)，得到

$$\begin{aligned} C_B \cdot \rho^{t'} \|x - x^*\| &\geq \|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}] - F(x^*) + \mathbb{E}[\widetilde{F}(x^*, \xi_{t+t'})|\mathcal{F}_{t-1}]\| \\ &\geq \|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}]\| - \|F(x^*) - \mathbb{E}[\widetilde{F}(x^*, \xi_{t+t'})|\mathcal{F}_{t-1}]\|. \end{aligned}$$

将不等式 (15)代入上面的不等式并整理各项，得到

$$\|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}]\| \leq C_M \rho^{t'} + C_B \rho^{t'} \|x - x^*\|,$$

由此完成引理1的证明。 □

Lem引理 2. 假设 *2–4* 成立。对于 t， $t' \in \mathbb{Z}_+$ 和 $x \in X$，我们有 e

$$\mathbb{E}\left[\|\widetilde{F}(x, \xi_{t+t'}) - F(x)\|^2\right] \leq \sigma^2 + 4C_M^2 \rho^{2t'} + (\zeta^2 + 8\bar{L}^2)\mathbb{E}\left[\|x - x^*\|^2\right],$$

$\bar{}$ $\tilde{}$其中 L：
$= L + L_1$。

证明。 首先，根据杨氏不等式，我们有

$$\|\widetilde{F}(x, \xi_{t+t'}) - F(x)\|^2 \leq 2\|\widetilde{F}(x, \xi_{t+t'}) - \mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}]\|^2 + 2\|\mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}] - F(x)\|^2. \tag{34}$$

为界定第二项，注意到：

$$\begin{aligned} \|\mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}] - F(x)\|^2 &\overset{\text{(i)}}{\leq} 2\|\mathbb{E}[\widetilde{F}(x, \xi_{t+t'})|\mathcal{F}_{t-1}] - F(x) - \mathbb{E}[\widetilde{F}(x^*, \xi_{t+t'})|\mathcal{F}_{t-1}] + F(x^*)\|^2 \\ &\quad + 2\|\mathbb{E}[\widetilde{F}(x^*, \xi_{t+t'})|\mathcal{F}_{t-1}] - F(x^*)\|^2 \end{aligned}$$

$$
\begin{aligned}
&\overset{\text{(ii)}}{\leq} 4\|F(x^*)-F(x)\|^2+4\|\mathbb{E}[\widetilde{F}(x,\xi_{t+t'})|\mathcal{F}_{t-1}]-\mathbb{E}[\widetilde{F}(x^*,\xi_{t+t'})|\mathcal{F}_{t-1}]\|^2\\
&\quad+2C_M^2\rho^{2t'}\\
&\overset{\text{(iii)}}{\leq} 4L^2\|x-x^*\|^2+4\mathbb{E}[\|\widetilde{F}(x,\xi_{t+t'})-\widetilde{F}(x^*,\xi_{t+t'})\|^2|\mathcal{F}_{t-1}]+2C_M^2\rho^{2t'}\\
&\overset{\text{(iv)}}{\leq} 4L^2\|x-x^*\|^2+4\mathbb{E}[\widetilde{L}_1^2\|x-x^*\|^2|\mathcal{F}_{t-1}]+2C_M^2\rho^{2t'},
\end{aligned}
$$

其中，步骤（i）由杨氏不等式得出，步骤（ii）由假设 4 和詹森不等式得出，步骤（iii）由不等式（2）得出，步骤（iv）由假设2得出。取期望，并在不等式（34）中使用假设3，可得

$$
\mathbb{E}\left[\|\widetilde{F}(x,\xi_{t+t'})-F(x)\|^2\right]\leq\sigma^2+4C_M^2\rho^{2t'}+(\zeta^2+8\bar{L}^2)\mathbb{E}\left[\|x-x^*\|^2\right], \tag{35}
$$

其中我们进一步使用了 $L^2+\widetilde{L}_1^2\leq(L+\widetilde{L}_1)^2=\bar{L}^2$。至此，引理 2 的证明完成。 □

7.2 定理 1 的证明

我们先定义以下简写

$$
\Gamma_t\coloneqq(1+\mu\eta_k)^t. \tag{36}
$$

以下命题中的误差分解是证明的关键。

命题 1。 假设 *1-5* 成立，则对于算法 k第 *1* 轮

$$
\begin{aligned}
\mathbb{E}\|x_{T_k+1}^{(k)}-x^*\|^2\leq\underbrace{\frac{G_1}{A_{T_k}}\|x_1^{(k)}-x^*\|^2}_{R_1}+\underbrace{\frac{10\eta_kC_M}{A_{T_k}}\left(\sum_{t=1}^{\tau_M/\tau_k}\rho^{t\tau_k-1}\Gamma_t+\sum_{t=\tau_M/\tau_k+1}^{T_k}\rho^{\tau_M+\tau_k-1}\Gamma_t\right)}_{R_2}\\
+\underbrace{\frac{H}{A_{T_k}}\eta_k^2\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right)}_{R_3},
\end{aligned}
$$

其中

$$
\begin{aligned}
A_{T_k}&=\left(\frac{1}{2}-\eta_k^2\bar{L}^2+\eta_k\mu\right)\Gamma_{T_k},\\
G_1&=\left(\frac{1}{2}+2\eta_k^2E\right)(1+\eta_k\mu)+\eta_kC\sum_{t'=1}^{\alpha_k+1}\rho^{\tau_kt'-1}(1+\eta_k\mu)^{t'}+2\eta_k^2\bar{L}^2\sum_{t'=1}^{\alpha_k+1}(1+\eta_k\mu)^{t'}\\
&\quad+2\eta_k^2\bar{L}^2\sum_{t'=1}^{\alpha_k}(t'-1)(1+\eta_k\mu)^{t'},\\
H&=\sum_{t=1}^{\alpha_k}t\Gamma_t+\sum_{t=\alpha_k+1}^{T_k}(\alpha_k+1)\Gamma_t,\ E=\zeta^2+8\bar{L}^2,
\end{aligned}
$$

以及 α_k 定义见方程(19)。

在上文 G_1 的定义中，请回忆 $C = \left(\frac{C_M}{40} + C_B\right)$ 和 $\bar{L} = L + \widetilde{L}_1$ 来自方程 (20)。命题1 在第 7.3节中证明。与定理陈述相对应，第一项 R_1 是确定性误差，项 R_2 源于最优解 x^* 处的偏差，而项 R_3 是随机误差。请注意， $R_2 = 0$ 如果 $C_M = 0$。

暂且假设命题 1 的陈述成立，并继续证明定理 1。接下来，我们给出一个简单条件和一个命题，证明定理 1时将始终用到它们:

$$\frac{3\mu^2}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)} \leq \frac{3\mu^2}{128\bar{L}^2} \overset{\text{(i)}}{\leq} \frac{3}{128}, \tag{37}$$

其中，步骤 (i) 是由事实 $\mu \leq L$ 得出的。

命题 1。 在定理 *1*的条件下，步长序列 η_k 满足 s

$$\eta_k \leq \min\left\{\frac{3\mu}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)}, \frac{p_k \log T_k}{\mu T_k}\right\}. \tag{38}$$

我们暂且将此陈述视为已知，并在第 7.2.5 节中证明它。

为了继续证明，我们需要对 Γ_{α_k+1} 进行界定，其中 Γ_t 定义见式（36）。我们对这一项作如下界定:

$$\Gamma_{\alpha_k+1} = (1+\eta_k\mu)^{\alpha_k+1} \leq \left(1 + \frac{3\mu^2}{8(\zeta^2+16\bar{L}^2)\left(\alpha_k+1\right)}\right)^{\alpha_k+1} \leq e^{\frac{3\mu^2}{8(\zeta^2+16\bar{L}^2)}} < 2. \tag{39}$$

现在，我们可以对命题 1 中的各个误差项进行界定。

7.2.1 界定 R_1

首先，我们对选择 G_1 $\Gamma_t = (1+\eta_k\mu)^t$ 以及定理 η_k 1 中定义的 进行界定。根据命题1中 的定义，我们有

$$\begin{aligned}
G_1 &= \left(\frac{1}{2} + 2\eta_k^2 E\right)(1+\eta_k\mu) + \eta_k C \sum_{t'=1}^{\alpha_k+1} \rho^{\tau_k t'-1}(1+\eta_k\mu)^{t'} + 2\eta_k^2\bar{L}^2 \sum_{t'=1}^{\alpha_k+1}(1+\eta_k\mu)^{t'} \\
&\quad + 2\eta_k^2\bar{L}^2 \sum_{t'=1}^{\alpha_k}(t'-1)(1+\eta_k\mu)^{t'} \\
&\overset{\text{(i)}}{\leq} \left(\frac{1}{2} + \eta_k^2 E\right)(1+\eta_k\mu) + \eta_k C \sum_{t'=1}^{\alpha_k+1}(1+\eta_k\mu)^{t'} + 2\eta_k^2\bar{L}^2\frac{(1+\eta_k\mu)^{\alpha_k+2}}{\mu\eta_k} + 2\eta_k^2\bar{L}^2\alpha_k\frac{(1+\eta_k\mu)^{\alpha_k+1}}{\mu\eta_k} \\
&\leq \left(\frac{1}{2} + \eta_k^2 E\right)(1+\eta_k\mu) + C\frac{(1+\eta_k\mu)^{\alpha_k+1}}{\mu} + 2\eta_k\bar{L}^2\left(\alpha_k + (1+\eta_k\mu)\right)\frac{(1+\eta_k\mu)^{\alpha_k+1}}{\mu},
\end{aligned}$$

其中，步骤 (i) 源于 ρ 位于区间 (0，1) 内。将 E 的定义和 η_k 的条件代入上述不等式，可得

$$G_1 \overset{\text{(i)}}{\leq} \left(\frac{1}{2} + \frac{9\mu^2}{\left(\alpha_k+1\right)^2 64(\zeta^2+16\bar{L}^2)^2}(\zeta^2+8\bar{L}^2)\right)\left(1 + \frac{3\mu^2}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)}\right)$$

$$+\frac{2C}{\mu}+\frac{6\mu}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)}\bar{L}^2\left(\alpha_k+1+\frac{3\mu^2}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)}\right)\frac{2}{\mu}$$
$$\overset{\text{(ii)}}{\leq}\left(\frac{1}{2}+\frac{9}{1024}\right)\frac{131}{128}+\frac{2C}{\mu}+\frac{12}{128}\frac{\alpha_k+131/128}{\alpha_k+1}\leq 1+\frac{2C}{\mu},$$

其中，步骤 (i) 源于不等式 (38) 和不等式 (39，步骤 (ii) 源于不等式 (37).

接下来我们对 A_{T_k}进行界定。根据命题 A_{T_k} 1 中 的定义，我们有

$$\begin{aligned}A_{T_k}&=\left(\frac{1}{2}-\eta_k^2\bar{L}^2+\eta_k\mu\right)\Gamma_{T_k}=\left(\frac{1}{2}-\eta_k^2\bar{L}^2+\eta_k\mu\right)\left(1+\eta_k\mu\right)^{T_k}\\&\geq\left(\frac{1}{2}-\eta_k\mu\frac{3\bar{L}^2}{8\left(\alpha_k+1\right)\left(\zeta^2+16\bar{L}^2\right)}+\eta_k\mu\right)\left(1+\eta_k\mu\right)^{T_k}\\&\geq\left(\frac{1}{2}-\frac{3\eta_k\mu}{128}+\eta_k\mu\right)\left(1+\eta_k\mu\right)^{T_k}=\left(\frac{1}{2}+\frac{125\eta_k\mu}{128}\right)\left(1+\eta_k\mu\right)^{T_k}\geq\frac{1}{2}(1+\eta_k\mu)^{T_k}.\end{aligned}\tag{40}$$

最后结合对 G_1 和 A_{T_k}的界定，我们得到

$$\frac{G_1}{A_{T_k}}\leq 2\left(1+\frac{2C}{\mu}\right)\left(1+\eta_k\mu\right)^{-T_k}.$$

利用不等式 (38) 中的步长上界，并将两个正数的最大值替换为它们的和，我们得到

$$\begin{aligned}&2\left(1+\frac{2C}{\mu}\right)\left(1+\eta_k\mu\right)^{-T_k}\\&\leq 2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{p_k\log T_k}{T_k}\right)^{-T_k}\\&\leq 2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+2\left(1+\frac{2C}{\mu}\right)T_k^{-p_k}\\&\overset{\text{(i)}}{=}2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+\frac{6\left(1+\frac{2C}{\mu}\right)}{\mu^2T_kM\max(1,D^2)}(6\sigma^2+4\|F(x^*)\|^2),\end{aligned}$$

其中步骤 (i)　由将p_k 的定义代入式 (20) 可得。继续可得

$$\begin{aligned}2\left(1+\frac{2C}{\mu}\right)\left(1+\eta_k\mu\right)^{-T_k}&\overset{\text{(i)}}{\leq}2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}\\&\quad+\frac{3}{\mu^2T_kD^2}(6\sigma^2+4\|F(x^*)\|^2)\\&\leq 2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}\\&\quad+\frac{3}{\mu^2T_kD^2}(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2).\end{aligned}$$

其中，步骤 (i) 根据式 (20) 中 M 的定义以及利用 $^1{}_{\max(a,b)}\leq\frac{1}{b}$ 这一事实得出。因此我们得到

$$R_1=\frac{G_1}{A_{T_k}}\|x_1^{(k)}-x^*\|^2\leq 2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}\|x_1^{(k)}-x^*\|^2$$

$$+\frac{3}{\mu^2 T_k D^2}(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2)\|x_1^{(k)}-x^*\|^2.$$

代入 $\frac{\|x_1^{(k)}-x^*\|}{D}\leq 1$ 假设1 可得

$$R_1 \leq 2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k} D^2 + \frac{3}{\mu^2 T_k}(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2). \tag{41}$$

7.2.2 界定 R_2

现在考虑最优解 x^*处由偏差导致的误差。我们有

$$\begin{aligned}
R_2 &= \frac{10\eta_k C_M}{A_{T_k}}\left(\sum_{t=1}^{\tau_M/\tau_k}\rho^{t\tau_k-1}(1+\eta_k\mu)^t+\sum_{t=\tau_M/\tau_k+1}^{T_k}\rho^{\tau_M+\tau_k-1}(1+\eta_k\mu)^t\right)\\
&\overset{\text{(i)}}{\leq} \frac{10\eta_k C_M}{A_{T_k}}\left(\sum_{t=1}^{\tau_M/\tau_k}(1+\eta_k\mu)^t+\frac{\rho^{\tau_M+\tau_k-1}(1+\eta_k\mu)^{T_k+1}}{\eta_k\mu}\right)\\
&\overset{\text{(ii)}}{\leq} \frac{40C_M}{\mu}(1+\eta_k\mu)^{-T_k}+\frac{20C_M\rho^{\tau_M+\tau_k-1}}{\mu}(1+\eta_k\mu),
\end{aligned}$$

其中，步骤（i）由以下事实得到： $\rho\in(0,1)$，而步骤（ii）由不等式（39）以及不等式（ A_{T_k} 40）中的界得到。我们采用与界定 G_1/A_{T_k} 在 R_1中的比值类似的方法，对上述不等式中的第一项进行界定。具体而言，利用不等式（38），并将两个正量的最大值替换为它们的和，可得

$$\begin{aligned}
(1+\eta_k\mu)^{-T_k} &\leq \left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+\left(1+\frac{p_k\log T_k}{T_k}\right)^{-T_k}\\
&\leq \left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+T_k^{-p_k}\\
&= \left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}\\
&\quad+\frac{3}{\mu^2 M T_k\max(1,\|x_1^{(k)}-x^*\|^2)}(6\sigma^2+4\|F(x^*)\|^2)\\
&\leq \left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+\frac{3}{\mu^2\frac{40C_M}{\mu}T_k}(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2).
\end{aligned}$$

因此， x^* 处由偏差导致的误差有如下界：

$$R_2 \leq \frac{40C_M}{\mu}\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}+\frac{3}{\mu^2 T_k}(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2) + \frac{20C_M\rho^{\tau_M+\tau_k-1}}{\mu}(1+\eta_k\mu). \tag{42}$$

7.2.3 界定 R_3

现在我们对随机误差项进行界定。将 $\Gamma_t = (1+\eta_k\mu)^t$ 代入 D 命题1中的定义，可得

$$\begin{aligned} H &= \sum_{t=1}^{\tau_M/\tau_k} t(1+\eta_k\mu)^t + \sum_{t=\tau_M/\tau_k+1}^{T_k} (\tau_M/\tau_k+1)(1+\eta_k\mu)^t \\ &\le \sum_{t=1}^{T_k} (\tau_M/\tau_k+1)(1+\eta_k\mu)^t \\ &\le (\alpha_k+1)\frac{(1+\mu\eta_k)^{T_k+1}}{\mu\eta_k}. \end{aligned}$$

合并上界 H 由上文以及 A_{T_k} 由不等式（40）可得

$$\begin{aligned} \eta_k^2\frac{H}{A_{T_k}} &\le 2\eta_k(\alpha_k+1)\frac{(1+\mu\eta_k)}{\mu} \\ &\overset{\text{(i)}}{\le} 2\eta_k(\alpha_k+1)\frac{1+\frac{3\mu^2}{8(\alpha_k+1)(\zeta^2+16\bar{L}^2)}}{\mu} \\ &\overset{\text{(ii)}}{\le} \frac{3\eta_k}{\mu}(\alpha_k+1) \le \frac{3p_k\log T_k}{\mu^2 T_k}(\alpha_k+1), \end{aligned}$$

其中，步骤（i）由不等式（38）得出，步骤（ii）由不等式（37）得出。因此，

$$\begin{aligned} R_3 &= \frac{H}{A_{T_k}}\eta_k^2(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2) \\ &\le \frac{3p_k\log T_k}{\mu^2 T_k}(\alpha_k+1)(6\sigma^2+25C_M^2\rho^{2\tau_k}+4\|F(x^*)\|^2). \end{aligned} \tag{43}$$

7.2.4 合并上界

将不等式（41）、（42）和（43）中的界代入命题 1，可得 $\mathbb{E}\|x_{T_k+1}^{(k)}-x^*\|^2$的上界为

$$\begin{aligned} &\frac{G_1}{A_{T_k}}\|x_1^{(k)}-x^*\|^2 + \frac{H}{A_{T_k}}\eta_k^2(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2) + \frac{1}{A_{T_k}}\sum_{t=1}^{T_k}\Gamma_t\frac{10\eta_k C_M^2\rho^{(\kappa+1)\tau_k-1}}{\mu} \\ &\le 2\left(1+\frac{2C}{\mu}\right)\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}D^2 \\ &\quad + \left(\frac{3}{\mu^2 T_k}+\frac{3p_k\log T_k}{\mu^2 T_k}(\alpha_k+1)+\frac{3}{\mu^2 T_k}\right)(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2) \\ &\quad + \frac{40C_M}{\mu}\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k} + \frac{20C_M\rho^{\tau_M+\tau_k-1}}{\mu}(1+\eta_k\mu) \\ &\le M\left(1+\frac{3\mu^2}{(\alpha_k+1)\,8(\zeta^2+16\bar{L}^2)}\right)^{-T_k}(D^2+1) \end{aligned}$$

$$+\frac{3(2+p_k\log T_k)}{\mu^2 T_k}(\alpha_k+1)\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right)+\frac{20C_M\rho^{\tau_M+\tau_k-1}}{\mu}(1+\eta_k\mu).$$

至此，定理1中的论证得证。接下来需要证明断言 1。

7.2.5 证明断言1

根据定理 1，我们选择的步长为

$$\eta_k=\min\left\{\frac{3\mu}{16(\zeta^2+16\bar{L}^2)},\frac{p_k\log T_k}{\mu T_k}\right\}.$$

现在我们将该断言的证明分为两种情况。

Ca情况 I： 对于 $\tau_k\geq\tau_M$，即 $\alpha_k\leq 1$和 $\frac{1}{1+\alpha_k}\geq\frac{1}{2}$的情况，我们有如下结论：

$$\frac{3\mu}{16(\zeta^2+16\bar{L}^2)}\leq\frac{3\mu}{8\left(1+\alpha_k\right)(\zeta^2+16\bar{L}^2)}.$$

因此，当$\eta_k=\frac{3\mu}{16(\zeta^2+16\bar{L}^2)}$时，我们有

$$\eta_k=\frac{3\mu}{16(\zeta^2+16\bar{L}^2)}\leq\frac{3\mu}{8\left(1+\alpha_k\right)(\zeta^2+16\bar{L}^2)}.$$

每当 $\eta_k=\frac{p_k\log T_k}{\mu T_k}$时，我们都有以下不等式组：

$$\eta_k=\frac{p_k\log T_k}{\mu T_k}<\frac{3\mu}{16(\zeta^2+16\bar{L}^2)}\leq\frac{3\mu}{8\left(1+\alpha_k\right)(\zeta^2+16\bar{L}^2)}.$$

因此我们得到

$$\eta_k\leq\min\left\{\frac{3\mu}{8(1+\alpha_k)(\zeta^2+16\bar{L}^2)},\frac{p_k\log T_k}{\mu T_k}\right\}$$

至此，我们完成了引理 1 在情况 I 中的证明。

情况 II： 对于情况 $\tau_k\leq\tau_M$，即 $\alpha_k\geq 1$和 $\frac{1}{1+\alpha_k}\leq\frac{1}{2}$，我们有

$$\frac{3\mu}{16(\zeta^2+16\bar{L}^2)}\geq\frac{3\mu}{8\left(1+\alpha_k\right)(\zeta^2+16\bar{L}^2)}.\tag{44}$$

由条件 $B\geq\frac{32p_1\tau_M(\zeta^2+16\bar{L}^2)}{3\mu^2}\log\left(\frac{32p_1\tau_M(\zeta^2+16\bar{L}^2)}{3\mu^2}\right)$，我们有如下结论：

$$\frac{\log B}{B}\leq\frac{3\mu^2}{16p_1\tau_M(\zeta^2+16\bar{L}^2)}.$$

注意，总轮数的上界为 $\log_2 B$，即对于所有 $k=1,\ldots,K$，均有 $k\leq\log_2 B$。因此我们有

$$\frac{p_1 k\log 2}{\mu}\leq\frac{3\mu B}{16\tau_M(\zeta^2+16\bar{L}^2)}.$$

代入定义 $T_k = 2^k$ 和 $\tau_k = \frac{B}{T_k}$，我们得到 —

$$\frac{p_1 \log T_k}{\mu T_k} \leq \frac{3\mu\tau_k}{16\tau_M(\zeta^2 + 16\bar{L}^2)}.$$

注意，$p_k \leq p_1$ 由于 p_k 在各轮次之间是单调的，并且在情况 II 中 $(1+\alpha_k) \leq \frac{2\tau_M}{\tau_k}$。代入这些界，我们得到

$$\frac{p_k \log T_k}{\mu T_k} \leq \frac{3\mu}{8\,(1+\alpha_k)\,(\zeta^2 + 16\bar{L}^2)}.$$

结合不等式（44），我们得到情况 II 下的如下结果：

$$\eta_k = \frac{p_k \log T_k}{\mu T_k} \leq \frac{3\mu}{8\,(1+\alpha_k)\,(\zeta^2 + 16\bar{L}^2)} \leq \frac{3\mu}{16(\zeta^2 + 16\bar{L}^2)}.$$

结合情况 I 和情况 II 中关于步长参数的不等式，我们可以看到，以下条件始终成立：

$$\eta_k \leq \min\left\{\frac{3\mu}{8\,(\alpha_k+1)\,(\zeta^2 + 16\bar{L}^2)}, \frac{p_k \log T_k}{\mu T_k}\right\}. \tag{45}$$

至此，命题 1 的证明完成，从而定理 1的证明也告完成。

7.3 命题 1 的证明

我们首先给出算法 1的“三点引理”，以刻画式（12）中随机逼近更新的最优性条件。

引理 3。令 $x_{t+1}^{(k)}$ 按照方程 (12)中定义的更新规则生成。则对于任意 $x \in X$，我们有

$$\eta_k\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}), x_{t+1}^{(k)} - x\rangle + \frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 \leq \frac{1}{2}\|x_t^{(k)} - x\|^2 - \frac{1}{2}\|x_{t+1}^{(k)} - x\|^2. \tag{46}$$

我们将在第 7.3.2 节稍后证明此引理。

此引理构成了我们分析的基础。我们首先将不等式 (46) 中的内积分解如下：

$$\begin{aligned}
\left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}), x_{t+1}^{(k)} - x \right\rangle &= \left\langle F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x \right\rangle + \left\langle F(x_t^{(k)}) - F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x \right\rangle \\
&\quad + \left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)}), x_{t+1}^{(k)} - x \right\rangle \\
&\overset{\text{(i)}}{\geq} \left\langle F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x \right\rangle - L\|x_{t+1}^{(k)} - x_t^{(k)}\|\|x_{t+1}^{(k)} - x\| \\
&\quad + \left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)}), x_{t+1}^{(k)} - x \right\rangle,
\end{aligned} \tag{47}$$

其中步骤（i）由柯西–施瓦茨不等式以及不等式（2）中的利普希茨性质得出。我们给出以下命题，以界定不等式（47）中的最后一项。

命题 2。 设 $x_{t-\kappa}^{(k)}$ 是某个参数 $\kappa \in \{0,1,2,\ldots\}$ 对应的过去迭代点。那么，

$$\begin{aligned}\left\langle \widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)}), x_{t+1}^{(k)} - x \right\rangle \geq & -\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|\|x_{t+1}^{(k)} - x_t^{(k)}\| \\ & + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)},\xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\ & - \sum_{t'=t-\kappa}^{t-1} \eta_k \Big(2\bar{L}^2\|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x,\xi_{t'\tau_k})\|^2 + 2\bar{L}^2\|x_{t-\kappa}^{(k)} - x\|^2 \\ & \qquad + 2\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big).\end{aligned}$$

命题 2 将在第 7.3.3节中证明。

将命题2 代入不等式（47），可得

$$\begin{aligned}&\langle \widetilde{F}(x_t^{(k)},\xi_{t\tau_k}), x_{t+1}^{(k)} - x \rangle \\ &\geq \left\langle F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x \right\rangle - L\|x_{t+1}^{(k)} - x_t^{(k)}\|\|x_{t+1}^{(k)} - x\| \\ &\quad - \|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|\|x_{t+1}^{(k)} - x_t^{(k)}\| + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)},\xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\ &\quad - \sum_{t'=t-\kappa}^{t-1} \eta_k \Big(2\bar{L}^2\|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x,\xi_{t'\tau_k})\|^2 + 2\bar{L}^2\|x_{t-\kappa}^{(k)} - x\|^2 + 2\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big).\end{aligned} \tag{48}$$

将不等式（48）中的界代入不等式（46）的三点引理，可得

$$\begin{aligned}&\frac{1}{2}\|x_t^{(k)} - x\|^2 - \frac{1}{2}\|x_{t+1}^{(k)} - x\|^2 \\ &\geq \frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 + \eta_k \left\langle F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x \right\rangle - \eta_k L\|x_{t+1}^{(k)} - x_t^{(k)}\|\|x_{t+1}^{(k)} - x\| \\ &\quad - \eta_k\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|\|x_{t+1}^{(k)} - x_t^{(k)}\| + \eta_k \left\langle \widetilde{F}(x_{t-\kappa}^{(k)},\xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\ &\quad - \sum_{t'=t-\kappa}^{t-1} \eta_k^2 \Big(2\bar{L}^2\|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x,\xi_{t'\tau_k})\|^2 + 2\bar{L}^2\|x_{t-\kappa}^{(k)} - x\|^2 + 2\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big).\end{aligned} \tag{49}$$

根据杨氏不等式，我们有

$$\begin{aligned}&\frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 - \eta_k L\|x_{t+1}^{(k)} - x_t^{(k)}\|\|x_{t+1}^{(k)} - x\| - \eta_k\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|\|x_{t+1}^{(k)} - x_t^{(k)}\| \\ &\geq \frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 - \frac{1}{4}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 - \eta_k^2 L^2\|x_{t+1}^{(k)} - x\|^2 - \frac{1}{4}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 \\ &\qquad - \eta_k^2\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \\ &\overset{\text{(i)}}{\geq} -\eta_k^2\bar{L}^2\|x_{t+1}^{(k)} - x\|^2 - \eta_k^2\|\widetilde{F}(x_t^{(k)},\xi_{t\tau_k}) - F(x_t^{(k)})\|^2,\end{aligned}$$

其中，步骤（i）由以下事实可得 $-L \geq -\bar{L}$。将此界代入不等式（49），可得

$$\frac{1}{2}\|x_t^{(k)} - x\|^2 t \geq \left(\frac{1}{2} - \eta_k^2\bar{L}^2\right)\|x_{t+1}^{(k)} - x\|^2 + \eta_k \left\langle F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x \right\rangle + \eta_k \left\langle \widetilde{F}(x_{t-\kappa}^{(k)},\xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle$$

$$- \eta_k^2 \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 - \sum_{t'=t-\kappa}^{t-1} \eta_k^2 \Big(2\bar{L}^2 \|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x, \xi_{t'\tau_k})\|^2 + 2\bar{L}^2 \|x_{t-\kappa}^{(k)} - x\|^2 + 2\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big). \tag{50}$$

利用不等式（3）中的单调性，我们有

$$\eta_k \left\langle F(x_{t+1}^{(k)}), x_{t+1}^{(k)} - x^* \right\rangle \geq \eta_k \mu \|x_{t+1}^{(k)} - x^*\|^2.$$

取 $x = x^*$ 于不等式（50）中，并利用上述界，我们得到

$$\begin{aligned}
&\frac{1}{2}\|x_t^{(k)} - x^*\|^2 \\
&\geq \left(\frac{1}{2} - \eta_k^2 \bar{L}^2 + \eta_k \mu\right) \|x_{t+1}^{(k)} - x^*\|^2 + \eta_k \left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x^* \right\rangle \\
&\quad - \eta_k^2 \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \\
&\quad - \sum_{t'=t-\kappa}^{t-1} \eta_k^2 \Big(2\bar{L}^2 \|x_{t'}^{(k)} - x^*\|^2 + 2\|\widetilde{F}(x^*, \xi_{t'\tau_k})\|^2 + 2\bar{L}^2 \|x_{t-\kappa}^{(k)} - x^*\|^2 + 2\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big).
\end{aligned} \tag{51}$$

对不等式（51）两边取期望，并利用杨氏不等式，我们得到

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\|x_t^{(k)} - x^*\|^2 \\
&\geq \left(\frac{1}{2} - \eta_k^2 \bar{L}^2 + \eta_k \mu\right) \mathbb{E}\|x_{t+1}^{(k)} - x^*\|^2 + \eta_k \mathbb{E}\left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x^* \right\rangle \\
&\quad - \eta_k^2 \mathbb{E}\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \\
&\quad - \sum_{t'=t-\kappa}^{t-1} \eta_k^2 \Big(2\bar{L}^2 \mathbb{E}\|x_{t'}^{(k)} - x^*\|^2 + 4\mathbb{E}\|\widetilde{F}(x^*, \xi_{t'\tau_k}) - F(x^*)\|^2 \\
&\qquad + 4\|F(x^*)\|^2 + 2\bar{L}^2 \mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2 + 2\mathbb{E}\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big).
\end{aligned} \tag{52}$$

将引理1和引理2 按 τ_k 缩放后代入不等式（52），得到

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\|x_t^{(k)} - x^*\|^2 \\
&\geq \left(\frac{1}{2} - \eta_k^2 \bar{L}^2 + \eta_k \mu\right) \mathbb{E}\|x_{t+1}^{(k)} - x^*\|^2 - \eta_k C_B \rho^{(\kappa+1)\tau_k - 1} \mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2 - \eta_k C_M \rho^{(\kappa+1)\tau_k - 1} \mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\| \\
&\quad - \eta_k^2 \left(\sigma^2 + 4C_M^2 \rho^{2(\tau_k - 1)} + \left(\zeta^2 + 8\bar{L}^2\right) \mathbb{E}\|x_t^{(k)} - x^*\|^2\right) \\
&\quad - \sum_{t'=t-\kappa}^{t-1} \eta_k^2 \Big(2\bar{L}^2 \mathbb{E}\|x_{t'}^{(k)} - x^*\|^2 + 4\sigma^2 + 16C_M^2 \rho^{2(\tau_k - 1)} + 4\|F(x^*)\|^2 + 2\bar{L}^2 \mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2
\end{aligned}$$

$$+ 2\sigma^2 + 8C_M^2\rho^{2(\tau_k-1)} + 2\left(\zeta^2 + 8\bar{L}^2\right)\mathbb{E}\|x_t^{(k)} - x^*\|^2\Big). \tag{53}$$

重新整理各项后，得到

$$\begin{aligned}
&\left(\frac{1}{2} - \eta_k^2\bar{L}^2 + \eta_k\mu\right)\mathbb{E}\|x_{t+1}^{(k)} - x^*\|^2 \\
&\leq \left(\frac{1}{2} + \sum_{t'=t-\kappa}^{t} 2\eta_k^2(\zeta^2 + 8\bar{L}^2)\right)\mathbb{E}\|x_t^{(k)} - x^*\|^2 + \eta_k C_B\rho^{(\kappa+1)\tau_k-1}\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2 \\
&\quad + \eta_k C_M\rho^{(\kappa+1)\tau_k-1}\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\| \ + \sum_{t'=t-\kappa}^{t-1}\eta_k^2(2\bar{L}^2\mathbb{E}\|x_{t'}^{(k)} - x^*\|^2 + 2\bar{L}^2\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2) \\
&\quad + \sum_{t'=t-\kappa}^{t}\eta_k^2(6\sigma^2 + 24C_M^2\rho^{2(\tau_k-1)} + 4\|F(x^*)\|^2).
\end{aligned}$$

对上述不等式应用杨氏不等式和詹森不等式，得到

$$\begin{aligned}
&\left(\frac{1}{2} - \eta_k^2\bar{L}^2 + \eta_k\mu\right)\mathbb{E}\|x_{t+1}^{(k)} - x^*\|^2 \\
&\leq \left(\frac{1}{2} + (\kappa+1)\,2\eta_k^2\left(\zeta^2 + 8\bar{L}^2\right)\right)\mathbb{E}\|x_t^{(k)} - x^*\|^2 + \eta_k C_B\rho^{(\kappa+1)\tau_k-1}\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2 \\
&\quad + 10\eta_k C_M\rho^{(\kappa+1)\tau_k-1} + \frac{\eta_k C_M\rho^{(\kappa+1)\tau_k-1}\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2}{40} \\
&\quad + \sum_{t'=t-\kappa}^{t-1}\eta_k^2\left(2\bar{L}^2\mathbb{E}\|x_{t'}^{(k)} - x^*\|^2 + 2\bar{L}^2\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2\right) \\
&\quad + (\kappa+1)\,\eta_k^2\left(6\sigma^2 + 24C_M^2\rho^{2(\tau_k-1)} + 4\|F(x^*)\|^2\right).
\end{aligned} \tag{54}$$

现在我们设定自由参数 κ的值，它决定了我们条件化时回溯到多远的过去。若 $\kappa = \tau_M/\tau_k$，则设 $t \geq \tau_M/\tau_k + 1$；若 $\kappa = t-1$，则设 $t \leq \tau_M/\tau_k$。两边乘以 $\Gamma_t = (1+\eta_k\mu)^t$（式（36）），利用 $C = \left(\frac{C_M}{40} + C_B\right)$ 的定义（式（17））， 并从 $t = 1$ 求和至 T_k， 可得

$$\begin{aligned}
&\sum_{t=1}^{T_k}\left(\frac{1}{2} - \eta_k^2\bar{L}^2 + \eta_k\mu\right)\Gamma_t\mathbb{E}\|x_{t+1}^{(k)} - x^*\|^2 \\
&\leq \sum_{t=1}^{T_k}\left(\frac{1}{2} + (\kappa+1)\,2\eta_k^2\left(\zeta^2 + 8\bar{L}^2\right)\right)\Gamma_t\mathbb{E}\|x_t^{(k)} - x^*\|^2 \\
&\quad + 10\sum_{t=1}^{T_k}\Gamma_t\eta_k C_M\rho^{(\kappa+1)\tau_k-1} + \sum_{t=1}^{T_k}\eta_k C\rho^{(\kappa+1)\tau_k-1}\Gamma_t\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2 \\
&\quad + \sum_{t=1}^{T_k}\sum_{t'=t-\kappa}^{t-1}\eta_k^2\Gamma_t(2\bar{L}^2\mathbb{E}\|x_{t'}^{(k)} - x^*\|^2 + 2\bar{L}^2\mathbb{E}\|x_{t-\kappa}^{(k)} - x^*\|^2)
\end{aligned}$$

$$+\sum_{t=1}^{T_k}\Gamma_t\left(\kappa+1\right)\eta_k^2\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right).$$

现在我们将参数 A_t 和 G_t 定义如下:

$$A_t=\left(\frac{1}{2}-\eta_k^2\bar{L}^2+\eta_k\mu\right)\Gamma_t, \tag{55}$$

$$G_t=\begin{cases}\Gamma_1\left(\frac{1}{2}+2\eta_k^2E\right)+\eta_kC\sum\limits_{t'=1}^{\alpha_k+1}\rho^{\tau_kt'-1}\Gamma_{t'}+2\eta_k^2\bar{L}^2\sum\limits_{t'=1}^{\alpha_k+1}\Gamma_{t'}+2\eta_k^2\bar{L}^2\sum\limits_{t'=1}^{\alpha_k}(t'-1)\Gamma_{t'}, & t=1\\ \Gamma_t\left(\frac{1}{2}+2\eta_k^2Et\right)+\eta_kC\rho^{\tau_M+\tau_k-1}\Gamma_{t+\alpha_k}+2\eta_k^2\bar{L}^2\sum\limits_{t'=t+1}^{t+\alpha_k}\Gamma_{t'}+2\eta_k^2\bar{L}^2\alpha_k\Gamma_{t+\alpha_k}, & t\in[2,\alpha_k]\\ \Gamma_t\left(\frac{1}{2}+2\eta_k^2E(\alpha_k+1)\right)+\eta_kC\rho^{\tau_M+\tau_k-1}\Gamma_{t+\alpha_k}+2\eta_k^2\bar{L}^2\sum\limits_{t'=t+1}^{t+\alpha_k}\Gamma_{t'}+2\eta_k^2\bar{L}^2\alpha_k\Gamma_{t+\alpha_k}, & t\in(\alpha_k,T_k-\alpha_k]\\ \Gamma_t\left(\frac{1}{2}+2\eta_k^2E(\alpha_k+1)\right)+2\eta_k^2\bar{L}^2\sum\limits_{t'=t+1}^{T_k}\Gamma_{t'}, & t\in(T_k-\alpha_k,T_k]\end{cases}$$

回顾命题 1 可知 $H=\sum_{t=1}^{\alpha_k}t\Gamma_t+\sum_{t=\alpha_k+1}^{T_k}(\alpha_k+1)\Gamma_t$。

将 A_t 和 G_t 在公式（55）中的定义以及 H 代入，可得

$$\sum_{t=1}^{T_k}A_t\mathbb{E}\|x_{t+1}^{(k)}-x^*\|^2\le\sum_{t=1}^{T_k}G_t\mathbb{E}\|x_t^{(k)}-x^*\|^2+10\sum_{t=1}^{T_k}\Gamma_t\eta_kC_M\rho^{(\kappa+1)\tau_k-1}$$
$$+H\eta_k^2\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right).$$

的定义 G_t 可通过简单的代数运算，并计算不同时间步 t范围内的系数数量得到。上述公式直接给出了 A_t和 H 的定义。我们断言 $G_{t+1}\le A_t$，并将在第 7.3.1 节中验证这一点。利用该条件可得

$$A_{T_k}\mathbb{E}\|x_{T_k+1}^{(k)}-x^*\|^2$$
$$\le G_1\|x_1^{(k)}-x^*\|^2+10\sum_{t=1}^{T_k}\Gamma_t\eta_kC_M\rho^{(\kappa+1)\tau_k-1}+H\eta_k^2\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right).$$

展开上述界中因最优解 x^* 处的偏差而出现的求和项，可得

$$10\sum_{t=1}^{T_k}\Gamma_t\eta_kC_M\rho^{(\kappa+1)\tau_k-1}=10\eta_kC_M\sum_{t=1}^{\tau_M/\tau_k}\rho^{t\tau_k-1}\Gamma_t+10\eta_kC_M\sum_{t=\tau_M/\tau_k+1}^{T_k}\rho^{\tau_M+\tau_k-1}\Gamma_t.$$

最后，我们得到

$$\mathbb{E}\|x_{T_k+1}^{(k)}-x^*\|^2\le\frac{G_1}{A_{T_k}}\|x_1^{(k)}-x^*\|^2+\frac{10\eta_kC_M}{A_{T_k}}\left(\sum_{t=1}^{\tau_M/\tau_k}\rho^{t\tau_k-1}\Gamma_t+\sum_{t=\tau_M/\tau_k+1}^{T_k}\rho^{\tau_M+\tau_k-1}\Gamma_t\right)$$
$$+\frac{H}{A_{T_k}}\eta_k^2\left(6\sigma^2+25C_M^2\rho^{2(\tau_k-1)}+4\|F(x^*)\|^2\right),$$

如命题 1 中所述。

接下来需要验证条件 $A_t\ge G_{t+1}$ 并证明引理 3 和断言 2。

7.3.1 验证 $A_t \geq G_{t+1}$

现在验证，对于给定的参数选择 $A_t \geq G_{t+1}$ η_k 和 Γ_t，条件成立。根据命题 1的定义，我们有

$$
\begin{aligned}
A_t &= \left(\frac{1}{2} - \eta_k^2 \bar{L}^2 + \eta_k \mu\right) \Gamma_t \text{ and for all } t \in [1, T_k - 1] \\
G_{t+1} &\leq \Gamma_{t+1}\left(\frac{1}{2} + 2\eta_k^2 E\left(\alpha_k + 1\right)\right) + \eta_k C \rho^{\tau_M + \tau_k - 1} \Gamma_{t+1+\alpha_k} + 2\eta_k^2 \bar{L}^2 \sum_{t'=t+2}^{t+1+\alpha_k} \Gamma_{t'} + 2\eta_k^2 \bar{L}^2 \alpha_k \Gamma_{t+1+\alpha_k} \\
&\overset{\text{(i)}}{\leq} \Gamma_{t+1}\left(\frac{1}{2} + \eta_k^2 E\left(\alpha_k + 1\right)\right) + \eta_k C \rho^{\tau_M + \tau_k - 1} \Gamma_{t+1+\alpha_k} + 2\eta_k^2 \bar{L}^2 \alpha_k \Gamma_{t+1+\alpha_k} + 2\eta_k^2 \bar{L}^2 \alpha_k \Gamma_{t+1+\alpha_k} \\
&\overset{\text{(ii)}}{\leq} \Gamma_{t+1}\left(\frac{1}{2} + \eta_k^2 (\zeta^2 + 8\bar{L}^2)\left(\alpha_k + 1\right)\right) + \eta_k C \rho^{\tau_M + \tau_k - 1} \Gamma_{t+1+\alpha_k} + 4\eta_k^2 \bar{L}^2 \alpha_k \Gamma_{t+1+\alpha_k},
\end{aligned}
$$

其中，步骤（i）由参数 Γ_t 在式（20）中的递增性质得出，而步骤（ii）通过代入 $E = \zeta^2 + 8\bar{L}^2$的定义得出。为了证明 $A_t \geq G_{t+1}$，我们按如下步骤进行:

$$
\begin{aligned}
A_t - G_{t+1} &\geq \left(\frac{1}{2} - \eta_k^2 \bar{L}^2 + \eta_k \mu\right) \Gamma_t - \Gamma_{t+1}\left(\frac{1}{2} + \eta_k^2 (\zeta^2 + 8\bar{L}^2)\left(\alpha_k + 1\right)\right) \\
&\quad - \eta_k C \rho^{\tau_M + \tau_k - 1} \Gamma_{t+1+\alpha_k} - 4\eta_k^2 \bar{L}^2 \alpha_k \Gamma_{t+1+\alpha_k} \\
&\overset{\text{(i)}}{\geq} \Gamma_t \left(\frac{\eta_k \mu}{2} - \eta_k^2 \bar{L}^2 - \eta_k^2 (\zeta^2 + 8\bar{L}^2)\left(\alpha_k + 1\right)\Gamma - \eta_k C \rho^{\tau_M + \tau_k - 1} \Gamma_{\alpha_k + 1} - 4\eta_k^2 \bar{L}^2 \alpha_k \Gamma_{\alpha_k + 1}\right) \\
&\geq \Gamma_t \left(\frac{\eta_k \mu}{2} - \eta_k^2 (\zeta^2 + 13\bar{L}^2)\left(\alpha_k + 1\right)\Gamma_{\alpha_k + 1} - \eta_k C \rho^{\tau_M + \tau_k - 1} \Gamma_{\alpha_k + 1}\right) \\
&\overset{\text{(ii)}}{\geq} \Gamma_t \left(\frac{\eta_k \mu}{2} - \frac{3\eta_k \mu}{8} \Gamma_{\alpha_k + 1} - \frac{\eta_k \mu}{18} \Gamma_{\alpha_k + 1}\right),
\end{aligned} \tag{56}
$$

其中，步骤（i）由 Γ_t 的定义得出，而步骤（ii）由关于 η_k的条件得出，该条件在不等式（38）中讨论过，以及条件 $\tau_M = \log 18C/\mu \log 1/\rho$ 。将不等式（ Γ_{α_k+1} 中的界（39）代入不等式（56），得到

$$
A_t - G_{t+1} \geq 0.
$$

因此，我们已验证了对于给定的参数选择，条件 $A_t \geq G_{t+1}$ 得以满足。

7.3.2 引理3的证明

定义函数 $\Phi(x)$: $= \eta_k \langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) x \rangle + \frac{1}{2}\|x_t^{(k)} - x\|^2$。利用$\Phi(x)$的强凸性，我们有

$$
\Phi(x) \geq \Phi(x_{t+1}^{(k)}) + \langle \nabla \Phi(x_{t+1}^{(k)}), x - x_{t+1}^{(k)} \rangle + \frac{1}{2}\|x_{t+1}^{(k)} - x\|^2. \tag{57}
$$

利用 $\Phi(x)$ 的最优性条件，我们有 $\langle \nabla \Phi(x_{t+1}^{(k)}) x - x_{t+1}^{(k)} \rangle \geq 0$。因此

$$
\Phi(x) \geq \Phi(x_{t+1}^{(k)}) + \frac{1}{2}\|x_{t+1}^{(k)} - x\|^2. \tag{58}
$$

代入 $\Phi(x)$ 的定义，即得引理3.

7.3.3 命题 2 的证明

现在我们着重对不等式（47）中的最后一项进行界定。

我们先利用某个过去的迭代点 $x_{t-\kappa}^{(k)}$，对参数 $\kappa \in \{0,1,2,\ldots\}$ 的情形下的 $\left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)}) \right.$ $\left. x_{t+1}^{(k)} - x \right\rangle$ 进行分解。这是一种分析工具，旨在处理数据中的马尔可夫相关性。因此，我们有

$$
\begin{aligned}
&\left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)}), x_{t+1}^{(k)} - x \right\rangle \\
&= \left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)}), x_{t+1}^{(k)} - x_{t-\kappa}^{(k)} \right\rangle + \left\langle \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}), x_{t-\kappa}^{(k)} - x \right\rangle \\
&\quad + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle + \left\langle F(x_{t-\kappa}^{(k)}) - F(x_t^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\
&\overset{\text{(i)}}{\geq} -\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_{t-\kappa}^{(k)}\| - (\widetilde{L}_1 + L)\|x_{t-\kappa}^{(k)} - x_t^{(k)}\| \|x_{t-\kappa}^{(k)} - x\| \\
&\quad + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\
&\overset{\text{(ii)}}{=} -\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_{t-\kappa}^{(k)}\| - \bar{L}\|x_{t-\kappa}^{(k)} - x_t^{(k)}\| \|x_{t-\kappa}^{(k)} - x\| \\
&\quad + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle,
\end{aligned} \tag{59}
$$

其中，步骤 (i) 根据柯西–施瓦茨不等式和不等式 (2) 中的利普希茨假设 $\bar{}$ $\widetilde{}$以及不等式 (13) 得出。在步骤 (ii) 中，我们使用参数 $L = L_1 + L$，其定义见公式 (20)。

现在，我们分别对不等式 (59) 中的各项进行界定。我们提出以下不等式，并将在本节末尾给出证明：

$$
\|x_{t+1}^{(k)} - x_t^{(k)}\| \leq 2\eta_k \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k})\|. \tag{60}
$$

利用式（60）和三角不等式，我们得到 $\|x_t^{(k)} - x_{t-\kappa}^{(k)}\|$ 如下界 s:

$$
\|x_t^{(k)} - x_{t-\kappa}^{(k)}\| \leq \sum_{t'=t-\kappa}^{t-1} 2\eta_k \|\widetilde{F}(x_{t'}^{(k)}, \xi_{t'\tau_k})\| \overset{\text{(i)}}{\leq} \sum_{t'=t-\kappa}^{t-1} 2\eta_k \left(\bar{L}\|x_{t'}^{(k)} - x\| + \|\widetilde{F}(x, \xi_{t'\tau_k})\| \right), \tag{61}
$$

其中步骤（i） $\widetilde{}$ $\bar{}$由利普希茨假设以及 $L_1 \leq L$这一事实可得。因此 y,

$$
\begin{aligned}
\|x_t^{(k)} - x_{t-\kappa}^{(k)}\| \|x_{t-\kappa}^{(k)} - x\| &\leq \sum_{t'=t-\kappa}^{t-1} 2\eta_k \left(\bar{L}\|x_{t'}^{(k)} - x\| \|x_{t-\kappa}^{(k)} - x\| + \|\widetilde{F}(x, \xi_{t'\tau_k})\| \|x_{t-\kappa}^{(k)} - x\| \right) \\
&\overset{\text{(i)}}{\leq} \sum_{t'=t-\kappa}^{t-1} \eta_k \left(\frac{1}{\bar{L}} \|\widetilde{F}(x, \xi_{t'\tau_k})\|^2 + \bar{L}\|x_{t'}^{(k)} - x\|^2 + 2\bar{L}\|x_{t-\kappa}^{(k)} - x\|^2 \right)
\end{aligned} \tag{62}
$$

其中步骤（i）由杨氏不等式得出。

我们还有

$$
\begin{aligned}
&\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_{t-\kappa}^{(k)}\| \\
&\leq \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \left(\|x_{t+1}^{(k)} - x_t^{(k)}\| + \|x_t^{(k)} - x_{t-\kappa}^{(k)}\| \right) \\
&\overset{\text{(i)}}{\leq} \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_t^{(k)}\|
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{t'=t-\kappa}^{t-1} 2\eta_k \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \left(\bar{L}\|x_{t'}^{(k)} - x\| + \|\widetilde{F}(x, \xi_{t'\tau_k})\| \right) \\
&\overset{\text{(ii)}}{\leq} \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_t^{(k)}\| \\
&\quad + \sum_{t'=t-\kappa}^{t-1} \eta_k \left(\bar{L}^2 \|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 + \|\widetilde{F}(x, \xi_{t'\tau_k})\|^2 \right). \qquad (63)
\end{aligned}
$$

其中，步骤（i）由三角不等式和不等式（61）得出，而步骤（ii）由杨氏不等式得出。

将不等式 (62) 和不等式 (63) 中的界代入不等式 (59)，得到

$$
\begin{aligned}
U_3 \geq & -\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_t^{(k)}\| \\
& - \sum_{t'=t-\kappa}^{t-1} \eta_k \left(\bar{L}^2 \|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 + \|\widetilde{F}(x, \xi_{t'\tau_k})\|^2 \right) \\
& - \sum_{t'=t-\kappa}^{t-1} \eta_k \bar{L} \left(\frac{1}{\bar{L}} \|\widetilde{F}(x, \xi_{t'\tau_k})\|^2 + \bar{L}\|x_{t'}^{(k)} - x\|^2 + 2\bar{L}\|x_{t-\kappa}^{(k)} - x\|^2 \right) \\
& + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\
= & -\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\| \|x_{t+1}^{(k)} - x_t^{(k)}\| + \left\langle \widetilde{F}(x_{t-\kappa}^{(k)}, \xi_{t\tau_k}) - F(x_{t-\kappa}^{(k)}), x_{t-\kappa}^{(k)} - x \right\rangle \\
& - \sum_{t'=t-\kappa}^{t-1} \eta_k \Big(2\bar{L}^2 \|x_{t'}^{(k)} - x\|^2 + 2\|\widetilde{F}(x, \xi_{t'\tau_k})\|^2 + 2\bar{L}^2 \|x_{t-\kappa}^{(k)} - x\|^2 \\
& \qquad + 2\|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}) - F(x_t^{(k)})\|^2 \Big),
\end{aligned}
$$

至此，命题2的证明完成。接下来只需证明命题 (60)。

命题 (60)**的证明。** 由公式 (12) 可得

$$
\left\langle \eta_k \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}), x_{t+1}^{(k)} \right\rangle + \frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 \leq \left\langle \eta_k \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}), x \right\rangle + \frac{1}{2}\|x - x_t^{(k)}\|^2 \text{ for all } x \in X.
$$

代入 $x = x_t^{(k)}$，得到

$$
\left\langle \eta_k \widetilde{F}(x_t^{(k)}, \xi_{t\tau_k}), x_{t+1}^{(k)} - x_t^{(k)} \right\rangle + \frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2 \leq 0.
$$

应用柯西–施瓦茨不等式可得

$$
-\eta_k \|\widetilde{F}(x_t^{(k)}, \xi_{t\tau_k})\| \|x_{t+1}^{(k)} - x_t^{(k)}\| \leq -\frac{1}{2}\|x_{t+1}^{(k)} - x_t^{(k)}\|^2.
$$

重新排列各项，得到压缩引理。

至此，命题1 的证明结束，包括其中的辅助引理和断言。

7.4 定理 2 的证明

为了证明定理2，需要正式定义串行跳跃经验回放（SSER）算法。该算法只需跟踪 $\tau_k = \beta\tau_M$所在的 MER 算法轮次。回顾第 4.3 节中的记号，令 x_{T+1}表示从某个初始化出发，执行T步式 (64) 中的 SSER 更新后得到的迭代序列。在给定初始化的条件下， x_T 中的随机性来自用于更新迭代序列的样本 $\{\xi_{t\beta\tau_M}\}_{t=1}^{T}$ 。类似地，定义 $\widetilde{x}_{T+1}$为从相同的初始化出发，在样本 $\{\widetilde{\xi}_{t\beta\tau_M}\}_{t=1}^{T}$构成的独立同分布序列上执行 T 步随机逼近后得到的迭代序列。回顾误差项 $\Delta_{T+1} = x_{T+1} - x^*$和 $\widetilde{\Delta}_{T+1} = \widetilde{x}_{T+1} - x^*$。两个算法均初始化于 x_1，因此 $\widetilde{x}_1 = x_1$。令 $T = {}^{B}_{\beta\tau_M}$。

算法 2 串行跳跃经验回放

输入：记忆缓冲区 $\{\xi_1, \xi_1, \ldots, \xi_B\}$，马尔可夫链的有效混合时间 τ_M，初始化x_1。

对于 $t = 1, \ldots, T$**执行**

使用样本 $\xi_{t\beta\tau_M}$，执行一步随机逼近：

$$x_{t+1} = \arg\min_{x\in X} \eta\langle \widetilde{F}(x_t, \xi_{t\beta\tau_M}), x\rangle + \frac{1}{2}\|x_t - x\|^2. \tag{64}$$

结束循环

完成符号设定后，我们开始证明。我们定义具有无约束更新的辅助序列如下：

$$\begin{aligned} x'_{t+1} &= x_t - \eta\widetilde{F}(x_t, \xi_{t\beta\tau_M}), \text{ for } t = 1, \ldots, T \\ \widetilde{x}'_{t+1} &= \widetilde{x}_t - \eta\widetilde{F}(\widetilde{x}_t, \xi_{t\beta\tau_M}), \text{ for } t = 1, \ldots, T. \end{aligned} \tag{65}$$

因此，在时间步 t，对应于 x_t 的迭代值在投影之前为 x'_t。由于 X 是一个闭凸集，且对这类集合的投影具有压缩性，因此我们有

$$\|x_{t+1} - \widetilde{x}_{t+1}\| \le \|x'_{t+1} - \widetilde{x}'_{t+1}\|, \text{ for all } t = 1, \ldots, T. \tag{66}$$

接下来，我们证明将随机变量 $(\xi_{j\beta\tau_M})_{j=1}^{T}$映射到 $\|\Delta_{T+1}\|$的函数 h是一个利普希茨连续函数，该函数将马尔可夫链的状态映射为误差。我们有

$$\begin{aligned} \|x_{T+1} - \widetilde{x}_{T+1}\| &\overset{\text{(i)}}{\le} \|x'_{T+1} - \widetilde{x}'_{T+1}\| \\ &\overset{\text{(ii)}}{=} \|x_T - \eta\widetilde{F}(x_T, \xi_{T\beta\tau_M}) - \widetilde{x}_T + \eta\widetilde{F}(\widetilde{x}_T, \widetilde{\xi}_{T\beta\tau_M})\| \\ &\le \|x_T - \widetilde{x}_T\| + \eta\|\widetilde{F}(x_T, \xi_{T\beta\tau_M}) - \widetilde{F}(\widetilde{x}_T, \widetilde{\xi}_{T\beta\tau_M})\| \\ &\overset{\text{(iii)}}{\le} (1 + \eta\widetilde{L}_1)\|x_T - \widetilde{x}_T\| + \eta\widetilde{L}_2\|\xi_{T\beta\tau_M} - \widetilde{\xi}_{T\beta\tau_M}\|, \end{aligned}$$

其中，步骤（i）由不等式（66）得出，步骤（ii）根据公式（65）中的定义得出，步骤（iii）则由假设2和假设7得出。递归应用上述关系可得

$$\|x_{T+1} - \widetilde{x}_{T+1}\| \overset{\text{(i)}}{\le} \eta\widetilde{L}_2 \sum_{i=1}^{T} \left((1 + \eta\widetilde{L}_1)^{T-i} \|\xi_{i\beta\tau_M} - \widetilde{\xi}_{i\beta\tau_M}\| \right)$$

$$\leq \eta\widetilde{L}_2(1+\eta\widetilde{L}_1)^{T-1}\sum_{i=1}^{T}\left(\|\xi_{i\beta\tau_M}-\widetilde{\xi}_{i\beta\tau_M}\|\right),$$

其中，步骤（i）由以下事实得出： $x_1=\widetilde{x}_1$。因此，我们有

$$\left|\|\Delta_{T+1}\|-\|\widetilde{\Delta}_{T+1}\|\right| \overset{\text{(i)}}{\leq} \|\Delta_{T+1}-\widetilde{\Delta}_{T+1}\| = \|x_{T+1}-\widetilde{x}_{T+1}\| \leq \eta\widetilde{L}_2(1+\eta\widetilde{L}_1)^{T-1}\sum_{i=1}^{T}\left(\|\xi_{i\beta\tau_M}-\widetilde{\xi}_{i\beta\tau_M}\|\right),$$

其中，步骤 (i) 由三角不等式得出。两边平方可得

$$\left|\|\Delta_{T+1}\|-\|\widetilde{\Delta}_{T+1}\|\right|^2 \overset{\text{(i)}}{\leq} T\eta^2\widetilde{L}_2^2\left(1+\eta\widetilde{L}_1\right)^{2(T-1)}\sum_{i=1}^{T}\|\xi_{i\beta\tau_M}-\widetilde{\xi}_{i\beta\tau_M}\|^2,$$

其中，步骤 (i) 由杨氏不等式得出。这证明函数 h 是常数为 $\sqrt{T}\eta\widetilde{L}_2\left(1+\eta\widetilde{L}_1\right)^{T-1}$ 的利普希茨连续函数。

现在注意到，根据康托罗维奇对偶性，有

$$\mathcal{W}_1(\mu,\nu)=\sup_{g\in\text{Lip}_1}\left|\mathbb{E}_{Y\sim\mu}[g(Y)]-\mathbb{E}_{Y\sim\nu}[g(Y)]\right|, \tag{67}$$

其中，Lip_1 表示所有 1-利普希茨连续函数 $g:\mathcal{X}\mapsto\mathbb{R}$构成的集合。由于 h 是 $\sqrt{T}\eta\widetilde{L}_2\left(1+\eta\widetilde{L}_1\right)^{T-1}$-利普希茨连续的，应用上述不等式可得

$$\left|\mathbb{E}\|\Delta_{T+1}\|-\mathbb{E}\|\widetilde{\Delta}_{T+1}\|\right| \leq \sqrt{T}\eta\widetilde{L}_2\left(1+\eta\widetilde{L}_1\right)^{T-1}\cdot\mathcal{W}_1\left(\{\xi_{i\beta\tau_M}\}_{i=1}^{T},\{\widetilde{\xi}_{i\beta\tau_M}\}_{i=1}^{T}\right) \tag{68}$$

其中，$\mathcal{W}_1\left(\{\xi_{i\beta\tau_M}\}_{i=1}^{T},\{\widetilde{\xi}_{i\beta\tau_M}\}_{i=1}^{T}\right)$ 表示 SSER 用于更新迭代序列的马尔可夫链与独立同分布样本序列之间的 1-瓦瑟斯坦距离。下面的引理将给出该 1-瓦瑟斯坦距离的上界。

引理 4。 设 τ_M 为马尔可夫链的有效混合时间。定义随机过程

$$Z=\{\xi_1,\xi_{\beta\tau_M},\xi_{2\beta\tau_M},...,\xi_{T\beta\tau_M}\},$$
$$\widetilde{Z}=\{\widetilde{\xi}_1,\widetilde{\xi}_{\beta\tau_M},\widetilde{\xi}_{2\beta\tau_M},...,\widetilde{\xi}_{T\beta\tau_M}\},$$

其中，$\xi_i\sim\pi$ 独立于其他所有变量，从平稳分布中抽取。于是，对某个通用正常数 c_0，以下结论成立：

$$\mathcal{W}_1(Z,\widetilde{Z})\leq c_0Te^{-\beta}.$$

我们将此引理的证明推迟到本节末尾，并指出，针对全变差混合，已有类似引理得到证明（例如，参见 Pananjady 等人，2024；Nakul 等人，2025）。将引理 4 代入不等式（68），得到

$$\left|\mathbb{E}\|\Delta_{T+1}\|-\mathbb{E}\|\widetilde{\Delta}_{T+1}\|\right| \leq c_0\eta T^{3/2}\widetilde{L}_2\left(1+\eta\widetilde{L}_1\right)^{T-1}e^{-\beta}.$$

最后，利用步长的定义$\eta\leq\frac{1}{T\widetilde{L}_1}$即可完成定理2的证明。

7.4.1 引理 4 的证明4

In 为了证明该引理，我们先定义一个中间随机过程 :

$$\widetilde{Z}_i = \{\xi_1, ..., \xi_{(i-1)\beta\tau_M}, \widetilde{\xi}_{i\beta\tau_M}, \xi_{(i+1)\beta\tau_M}, ..., \xi_{T\beta\tau_M}\}.$$

该随机过程将 i-th 个样本从 Z 替换为来自平稳分布的一个样本，且该样本与其他所有因素相互独立。在假设 6下， $\mathcal{W}_1(\xi_{i\beta\tau_M}, \widetilde{\xi}_{i\beta\tau_M}) \leq c_0 e^{-\beta}$对所有 $i = 1, ...,$ T均成立，这是 Mou 等人（2024）中的引理 4 所述结论。由 Wasserstein 度量的三角不等式可得

$$\mathcal{W}_1(Z, \widetilde{Z}) \leq \sum_{i=1}^{T} \mathcal{W}_1(Z, \widetilde{Z}_i) \leq c_0 T e^{-\beta}. \tag{69}$$

7.5 推论 1 的证明

为证明推论 1 ，我们首先验证必要的假设，然后调用定理 1 的陈述。

首先，我们证明式（7）中定义的随机算子关于其第一个自变量是 $L_f \cdot D_a^2$-利普希茨连续的:

$$\begin{aligned}
\|\widetilde{F}(x_1, \xi) - \widetilde{F}(x_2, \xi)\| &= \|af(a^T x_1) - ay - af(a^T x_2) + ay\| \\
&= \|af(a^T x_1) - af(a^T x_2)\| \\
&\overset{\text{(i)}}{\leq} L_f \cdot \|a\|^2 \cdot \|x_1 - x_2\| \\
&\overset{\text{(ii)}}{\leq} L_f \cdot D_a^2 \cdot \|x_1 - x_2\|,
\end{aligned}$$

其中，步骤（i）由链接函数的 Lipschitz 连续性 f 得到，步骤（ii）则由回归量轨迹的有界性得到，即 $\|a_t\| \leq D_a$ 几乎处处对所有 t均成立。因此，对于广义线性模型，参数为 $\bar{L} = L + \widetilde{L}_1 = 2 \cdot L_f \cdot D_a^2$，从而验证了假设 2。

现在，我们验证以下广义线性模型的假设 3 ：

$$\begin{aligned}
&\mathbb{E}\left[\left\|\widetilde{F}(x, \xi_{t'}) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}]\right\|^2 \middle| \mathcal{F}_{t-1}\right] \\
&\overset{\text{(i)}}{=} \mathbb{E}\left[\left\|a_{t'}(f(a_{t'}^T x) - y_{t'}) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}] + \mathbb{E}[\widetilde{F}(x^*, \xi_{t'})|\mathcal{F}_{t-1}]\right\|^2 \middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\left\|a_{t'}(f(a_{t'}^T x) - f(a_{t'}^T x^*)) + a_{t'}(f(a_{t'}^T x^*) - y_{t'}) - \mathbb{E}[a_{t'}(f(a_{t'}^T x) - f(a_{t'}^T x^*))|\mathcal{F}_{t-1}]\right\|^2 \middle| \mathcal{F}_{t-1}\right] \\
&\overset{\text{(ii)}}{\leq} 3\mathbb{E}\left[\left\|a_{t'}(f(a_{t'}^T x) - f(a_{t'}^T x^*))\right\|^2 \middle| \mathcal{F}_{t-1}\right] + 3\mathbb{E}\left[\left\|a_{t'}(f(a_{t'}^T x^*) - y_{t'})\right\|^2 \middle| \mathcal{F}_{t-1}\right] \\
&\quad + 3\mathbb{E}\left[\left\|\mathbb{E}[a_{t'}(f(a_{t'}^T x) - f(a_{t'}^T x^*))|\mathcal{F}_{t-1}]\right\|^2 \middle| \mathcal{F}_{t-1}\right],
\end{aligned}$$

~其中，步骤 (i) 是由广义线性模型 $\mathbb{E}[F(x^*, \xi_{t'})|\mathcal{F}_{t-1}] = \mathbb{E}[a_{t'} v_{t'}|\mathcal{F}_{t-1}] = 0$ 满足这一事实得出的，因为 $\zeta_{t'}$是零均值独立同分布噪声；步骤 (ii) 则由杨氏不等式得出。现在，利用回归量轨迹的有界性以及 f 的利普希茨连续性，我们有:

$$\mathbb{E}\left[\left\|\widetilde{F}(x, \xi_{t'}) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}]\right\|^2 \middle| \mathcal{F}_{t-1}\right]$$

$$\leq 6 \cdot L_f^2 \cdot D_a^4 \cdot \|x - x^*\|^2 + 3\mathbb{E}\left[\left\|a_{t'} v_{t'}\right\|^2 \middle| \mathcal{F}_{t-1}\right]$$
$$\leq 6 \cdot L_f^2 \cdot D_a^4 \cdot \|x - x^*\|^2 + 3 \cdot D_a^2 \cdot \sigma_v^2.$$

因此，假设3 在 $\sigma^2 = 6D_a^2\sigma_v^2$ 和 $\zeta^2 = 12L_f^2 D_a^4$下成立。

~对于广义线性模型， $F(x^*) = 0$ 并根据公式 (6可知，我们有 $\mathbb{E}[F(x^*, \xi_{t'})|\mathcal{F}_{t-1}] = \mathbb{E}[a_{t'} v_{t'}|\mathcal{F}_{t-1}] =$0。因此，假设 4 在 $C_M = 0$下成立。

令 π 表示 $\xi_t = (a_t, y_t)$ 的平稳分布，并回顾一下， $P_{t-1}^{t'}$ 表示在给定 $\xi_{t'}$ 的条件下的条件分布 $\mathcal{F}_{t-1}$。我们通过以下步骤验证假设 5 ：

$$\|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}] - F(y) + \mathbb{E}[\widetilde{F}(y, \xi_{t'})|\mathcal{F}_{t-1}]\|$$
$$= \|\mathbb{E}_{\xi\sim\pi}[\widetilde{F}(x, \xi) - \widetilde{F}(y, \xi)] - \mathbb{E}[\widetilde{F}(x, \xi_{t'}) - \widetilde{F}(y, \xi_{t'})|\mathcal{F}_{t-1}]\|$$
$$= \left\| \int_{\xi\in\Xi} (\widetilde{F}(x, \xi) - \widetilde{F}(y, \xi))(d\pi(\xi) - dP_{t-1}^{t'}(\xi)) \right\|.$$

利用 Lipsch ~itz 性质以及全变差距离的定义，可得

$$\|F(x) - \mathbb{E}[\widetilde{F}(x, \xi_{t'})|\mathcal{F}_{t-1}] - F(y) + \mathbb{E}[\widetilde{F}(y, \xi_{t'})|\mathcal{F}_{t-1}]\|$$
$$\leq 2\widetilde{L}_1 \|x - y\| d_{\mathsf{TV}}(\Pi, P_{t-1}^{t'}) \overset{\text{(i)}}{\leq} 2\widetilde{L}_1 C_{TV} \cdot \rho_{TV}^{t'-t} \cdot \|x - y\|,$$

其中，步骤 (i) 同样由假设 8得出。因此，对于广义线性模型，假设 5 在 $_\sim\rho_2 = \rho_{TV}$和 $C_B = 2L_1 C_{TV}$下成立。

由于所有这些假设均成立，应用定理 1即可证明推论 1。

7.6 推论2 的证明

~定理2 要求随机算子 F 关于其第二个自变量满足利普希茨条件，即对于广义线性模型，协变量和观测值（a_t， y_t）。

回顾一下， D_f 是由不等式（ f 30）给出的链接函数 $a_t \leq D_a$的界。我们还回顾，几乎必然地，对于所有 t 和 X ，其直径至多为 D_x。

我们有

$$\begin{aligned}
\|\widetilde{F}(x, \xi_1) - \widetilde{F}(x, \xi_2)\| &= \|a_1 f(a_1^T x) - a_1 y_1 - a_2 f(a_2^T x) + a_2 y_2\| \\
&\leq \|a_1 f(a_1^T x) - a_2 f(a_2^T x)\| + \|a_1 y_1 - a_2 y_2\| \\
&= \|a_1 f(a_1^T x) - a_1 f(a_2^T x) + a_1 f(a_2^T x) - a_2 f(a_2^T x)\| \\
&\quad + \|a_1 y_1 - a_1 y_2 + a_1 y_2 - a_2 y_2\| \\
&\leq L_f \|a_1\| \|x\| \|a_1 - a_2\| + |f(a_2^T x)| \|a_1 - a_2\| \\
&\quad + \|a_1\| |y_1 - y_2| + |y_2| \|a_1 - a_2\|. \qquad (70)
\end{aligned}$$

将 $|f(a^T x)| \leq D_f$代入不等式（70），并结合假设 $\|x_t\| \leq D_x$ 且 $|y_t| \leq D_y$对所有 t 成立，可得

$$\|\widetilde{F}(x, \xi_1) - \widetilde{F}(x, \xi_2)\| \leq (L_f D_a D_x + D_f + D_y) \|a_1 - a_2\| + D_a |y_1 - y_2|. \qquad (71)$$

~这证明随机算子 F 关于其第二个自变量满足利普希茨条件，其 Lipschitz 常数为 $\widetilde{L}_2 = \sqrt{2}$ max$(L_f D_a D_x + \bar{D}_f + D_y, D_a)$。

H在验证了必要的假设后，调用定理2即可证明推论 2.

7.7 推论 3的证明

利用特征向量有界和奖励有界这一假设，我们可以证明，对于强化学习中的策略评估问题，式（10）中定义的随机算子在其第一个参数上是$(1+\gamma)\cdot D_\psi^2$-利普希茨连续的：

$$\begin{aligned}\left\|\widetilde{F}(\theta,\xi_k)-\widetilde{F}(\theta',\xi_k)\right\| &= \left\|\left(\langle\psi(s_k),\theta-\theta'\rangle-\gamma\langle\psi(s_{k+1}),\theta-\theta'\rangle\right)\psi(s_k)\right\| \\ &= \left\|\langle\psi(s_k)-\gamma\psi(s_{k+1}),\theta-\theta'\rangle\psi(s_k)\right\| \\ &\overset{\text{(i)}}{\leq} \|\psi(s_k)\|\,\|\psi(s_k)-\gamma\psi(s_{k+1})\|\,\left\|\theta-\theta'\right\| \\ &\overset{\text{(ii)}}{\leq} (1+\gamma)\cdot D_\psi^2\cdot\|\theta-\theta'\|,\end{aligned}$$

其中，步骤 (i) 根据柯西–施瓦茨不等式得出，步骤 (ii) 则根据有界性得出。因此，我们有$\bar{L}=L+\widetilde{L}_1=(1+\gamma)(D_\psi^2+\lambda_{max}(Q))$。

假设 3 验证如下：

$$\begin{aligned}&\mathbb{E}\left[\left\|\widetilde{F}(\theta,\xi_{t'})-\mathbb{E}[\widetilde{F}(\theta,\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right] \\ &=\mathbb{E}\left[\left\|\widetilde{F}(\theta,\xi_{t'})-\widetilde{F}(\bar{\theta},\xi_{t'})+\widetilde{F}(\bar{\theta},\xi_{t'})-\mathbb{E}[\widetilde{F}(\bar{\theta},\xi_{t'})|\mathcal{F}_{t-1}]+\mathbb{E}[\widetilde{F}(\bar{\theta},\xi_{t'})|\mathcal{F}_{t-1}]-\mathbb{E}[\widetilde{F}(\theta,\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right] \\ &\overset{\text{(i)}}{\leq} 3\mathbb{E}\left[\left\|\widetilde{F}(\theta,\xi_{t'})-\widetilde{F}(\bar{\theta},\xi_{t'})\right\|^2\Big|\mathcal{F}_{t-1}\right]+3\mathbb{E}\left[\left\|\widetilde{F}(\bar{\theta},\xi_{t'})-\mathbb{E}[\widetilde{F}(\bar{\theta},\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right]+ \\ &\quad 3\mathbb{E}\left[\left\|\mathbb{E}[\widetilde{F}(\bar{\theta},\xi_{t'})|\mathcal{F}_{t-1}]-\mathbb{E}[\widetilde{F}(\theta,\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right] \\ &\overset{\text{(ii)}}{\leq} 6(1+\gamma)^2D_\psi^4\|\theta-\bar{\theta}\|^2+3\mathbb{E}\left[\left\|\widetilde{F}(\bar{\theta},\xi_{t'})-\mathbb{E}[\widetilde{F}(\bar{\theta},\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right] \\ &\overset{\text{(iii)}}{\leq} 6(1+\gamma)^2D_\psi^4\|\theta-\bar{\theta}\|^2+6\mathbb{E}\left[\left\|\widetilde{F}(\bar{\theta},\xi_{t'})\right\|^2\Big|\mathcal{F}_{t-1}\right]+6\mathbb{E}\left[\left\|\mathbb{E}[\widetilde{F}(\bar{\theta},\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right],\end{aligned}$$

其中，步骤 (i) 和步骤 (iii) 根据杨氏不等式得出，步骤 (ii) 则根据 F 的利普希茨 ~性质得出。现在，我们对上述不等式中的最后一项作如下界定：

$$\begin{aligned}\|\widetilde{F}(\bar{\theta},\xi_{t'})\|^2 &= \left\|(\langle\psi(s_{t'})-\gamma\psi(s_{t'+1}),\bar{\theta}\rangle-R(s_{t'},s_{t'+1}))\psi(s_{t'})\right\|^2 \\ &\overset{\text{(i)}}{\leq} 2\|\psi(s_{t'})-\gamma\psi(s_{t'+1})\|^2\|\bar{\theta}\|^2\|\psi(s_{t'})\|^2+2|R(s_{t'},s_{t'+1}))|^2\|\psi(s_{t'})\|^2 \\ &\overset{\text{(ii)}}{\leq} 2(1+\gamma)^2D_\psi^4D^2+2\bar{R}^2D_\psi^2,\end{aligned}$$

其中，步骤 (i) 根据杨氏不等式和柯西–施瓦茨不等式得出，步骤 (ii) 则根据有界性和假设 1得出。利用上述界，我们有

$$\begin{aligned}&\mathbb{E}\left[\left\|\widetilde{F}(\theta,\xi_{t'})-\mathbb{E}[\widetilde{F}(\theta,\xi_{t'})|\mathcal{F}_{t-1}]\right\|^2\Big|\mathcal{F}_{t-1}\right] \\ &\leq 6(1+\gamma)^2D_\psi^4\|\theta-\bar{\theta}\|^2+24(1+\gamma)^2D_\psi^4D^2+24\bar{R}^2D_\psi^2.\end{aligned}$$

因此，假设 3 在 $\sigma^2=48(1+\gamma)^2D_\psi^4D^2+48D_\psi^2\bar{R}^2$ 和 $\zeta^2=12(1+\gamma)^2D_\psi^4$下得到满足。假设 4 和假设 5 分别由 Li 等人 (2023) 的引理 1 和引理 2 验证。因此，我们调用定理 1 以完成推论 3的证明。

7.8 推论 4 的证明

现在，我们开始验证与策略评估相关的定理 2 的假设。为了验证假设 7，我们需要证明策略评估中的迭代序列保持有界。我们陈述以下引理，其证明将在第 7.8.1节给出。

引理 5. 假设特征向量有界，即对于所有 $s \in \mathcal{S}, \|\psi(s)\| \leq D_\psi$ ，奖励函数有界，即对于所有 $s, s' \in \mathcal{S}$ $R(s,s') \leq R$，且步长满足 $\eta \leq \frac{1}{(1+\gamma)\cdot D_\psi^2}$。那么我们有

$$\|\theta_T\| \leq \left(1+\sqrt{d}\right)\left(\|\theta_0\| + \frac{\bar{R}}{(1+\gamma)\cdot D_\psi}\right).$$

为方便记号，我们定义 $D_\theta := \left(1+\sqrt{d}\right)\left(\|\theta_0\| + \frac{\bar{R}}{1+\gamma}()\cdot D_\psi\right)$。现在我们陈述用于验证强化学习中策略评估的假设7 的引理，证明将在第 7.8.2 节给出。

引理 6. 在引理 5的前提下，随机算子 F 关于其第二个参数是利普希茨连续的，*Lipschitz* 常数为 $\sqrt{3}\left((2+\gamma)\cdot D_\psi \cdot D_\theta + \bar{R}\right)$。

在验证了这些假设后，我们调用定理 2以完成推论 4的证明。现在我们继续证明引理 5和引理6。

7.8.1 引理5的证明

我们有

$$\begin{aligned}\widetilde{F}(\theta_t,\xi_t) &= \left(\langle\psi(s_t),\theta_t\rangle - R(s_t,s_{t+1}) - \gamma\langle\psi(s_{t+1}),\theta_t\rangle\right)\psi(s_t)\\ &= \psi(s_t)\psi(s_t)^T\theta_t - \gamma\psi(s_t)\psi(s_{t+1})^T\theta_t - \psi(s_t)R(s_t,s_{t+1}).\end{aligned} \tag{72}$$

我们的目标是找到投影贝尔曼方程（ $\bar{\theta} \in \mathbb{R}^d$ 的固定点9）。这是一个无界问题，式（64）中的更新会简化为式（65）的无约束更新。将式（72）中的随机算子表达式代入式（65），可得

$$\theta_{t+1} = \theta_t - \eta\left(\psi(s_t)\psi(s_t)^T\theta_t - \gamma\psi(s_t)\psi(s_{t+1})^T\theta_t - \psi(s_t)R(s_t,s_{t+1})\right).$$

定义 $\Delta_t := \theta_t - \theta_0$。因此我们有

$$\begin{aligned}\Delta_{t+1} &= \Delta_t - \eta\left(\psi(s_t)\psi(s_t)^T - \gamma\psi(s_t)\psi(s_t)^T\right)\Delta_t + \eta\psi(s_t)R(s_t,s_{t+1})\\ &\quad - \eta\left(\psi(s_t)\psi(s_t)^T - \gamma\psi(s_t)\psi(s_{t+1})^T\right)\theta_0\\ &= \left(I - \eta\psi(s_t)\psi(s_t)^T + \eta\gamma\psi(s_t)\psi(s_t)^T\right)\Delta_t + \eta\psi(s_t)R(s_t,s_{t+1})\\ &\quad - \eta\left(\psi(s_t)\psi(s_t)^T - \gamma\psi(s_t)\psi(s_{t+1})^T\right)\theta_0.\end{aligned}$$

取 l_∞ 范数，并利用 $\left(I - \eta\psi(s_t)\psi(s_t)^T + \eta\gamma\psi(s_t)\psi(s_t)^T\right)$ 在 l_∞ 范数下的非扩张性质，可得

$$\|\Delta_{t+1}\|_\infty \leq \|\Delta_t\|_\infty + \eta D_\psi\bar{R} + \eta(1+\gamma)D_\psi^2\|\theta_0\|_\infty.$$

从 $t=0$ 到 $T-1$ 求和，可得

$$\|\Delta_T\|_\infty \le T\eta D_\psi \bar{R} + T\eta(1+\gamma)D_\psi^2 \|\theta_0\|_\infty .$$

将 l_∞ 转换为 ℓ_2 范数，并使用步长 $\eta \le \frac{1}{T\widetilde{L}_1}$ 的定义，我们得到

$$\|\Delta_T\| \le \frac{\sqrt{d}\cdot D_\psi \cdot \big(\bar{R} + (1+\gamma)D_\psi \|\theta_0\|\big)}{\widetilde{L}_1}.$$

这就完成了引理5的证明。

7.8.2 引理6的证明

我们有

$$\begin{aligned}\left\|\widetilde{F}(\theta,\xi_t) - \widetilde{F}(\theta,\xi_{t'})\right\| \le{}& \underbrace{\|\langle\psi(s_t),\theta\rangle\psi(s_t) - \langle\psi(s_{t'}),\theta\rangle\psi(s_{t'})\|}_{E_1} \\ &+ \underbrace{\|R(s_t,s_{t+1})\psi(s_t) - R(s_{t'},s_{t'+1})\psi(s_{t'})\|}_{E_2} \\ &+ \gamma\underbrace{\|\langle\psi(s_{t+1}),\theta\rangle\psi(s_t) - \langle\psi(s_{t'+1}),\theta\rangle\psi(s_{t'})\|}_{E_3}. \end{aligned} \tag{73}$$

现在我们分别对项 E_1、 E_2和 E_3 进行界定。

$$\begin{aligned} E_1 &= \|\langle\psi(s_t),\theta\rangle\psi(s_t) - \langle\psi(s_{t'}),\theta\rangle\psi(s_{t'})\| \\ &\le \|\langle\psi(s_t),\theta\rangle\psi(s_t) - \langle\psi(s_{t'}),\theta\rangle\psi(s_t)\| + \|\langle\psi(s_{t'}),\theta\rangle\psi(s_t) - \langle\psi(s_{t'}),\theta\rangle\psi(s_{t'})\| \\ &= \|\langle\psi(s_t) - \psi(s_{t'}),\theta\rangle\psi(s_t)\| + \|\langle\psi(s_{t'}),\theta\rangle(\psi(s_t) - \psi(s_{t'}))\| \\ &\le 2\cdot D_\psi \cdot \|\theta\| \cdot \|\psi(s_t) - \psi(s_{t'})\|. \end{aligned} \tag{74}$$

对于 E_2 ，我们有

$$\begin{aligned} E_2 &= \|R(s_t,s_{t+1})\psi(s_t) - R(s_{t'},s_{t'+1})\psi(s_{t'})\| \\ &\le \|R(s_t,s_{t+1})\psi(s_t) - R(s_t,s_{t+1})\psi(s_{t'})\| + \|R(s_t,s_{t+1})\psi(s_{t'}) - R(s_{t'},s_{t'+1})\psi(s_{t'})\| \\ &= \|R(s_t,s_{t+1})(\psi(s_t) - \psi(s_{t'}))\| + \|(R(s_t,s_{t+1}) - R(s_{t'},s_{t'+1}))\psi(s_{t'})\| \\ &\le \bar{R}\cdot\|\psi(s_t) - \psi(s_{t'}\| + D_\psi \cdot |(R(s_t,s_{t+1}) - R(s_{t'},s_{t'+1})|. \end{aligned} \tag{75}$$

对于 E_3 ，我们有

$$\begin{aligned} E_3 &= \gamma\,\|\langle\psi(s_{t+1}),\theta\rangle\psi(s_t) - \langle\psi(s_{t'+1}),\theta\rangle\psi(s_{t'})\| \\ &\le \gamma\|\langle\psi(s_{t+1}),\theta\rangle\psi(s_t) - \langle\psi(s_{t'+1}),\theta\rangle\psi(s_t)\| + \gamma\|\langle\psi(s_{t'+1}),\theta\rangle\psi(s_t) - \langle\psi(s_{t'+1}),\theta\rangle\psi(s_{t'})\| \\ &\le \gamma\|\langle\psi(s_{t+1}) - \psi(s_{t'+1}),\theta\rangle\psi(s_t)\| + \gamma\|\langle\psi(s_{t'+1}),\theta\rangle(\psi(s_t) - \psi(s_{t'}))\| \\ &\le \gamma D_\psi \cdot \|\theta\| \cdot \|\psi(s_{t+1}) - \psi(s_{t'+1}\| + \gamma D_\psi \cdot \|\theta\| \cdot \|\psi(s_t) - \psi(s_{t'})\|. \end{aligned} \tag{76}$$

将上述界代入不等式（74）、（75）和（76），并使用引理5，可得

$$\left\|\widetilde{F}(\theta,\xi_t) - \widetilde{F}(\theta,\xi_{t'})\right\| \le \big((2+\gamma)\cdot D_\psi \cdot D_\theta + \bar{R}\big)\cdot\|\psi(s_t) - \psi(s_{t'})\|$$

$$+\gamma D_\psi \cdot D_\theta \cdot \|\psi(s_{t+1}) - \psi(s_{t'+1}\| + D_\psi \cdot |(R(s_t, s_{t+1}) - R(s_{t'}, s_{t'+1})|.$$

$\widetilde{\ }$这证明了策略评估中的随机算子 $\underline{\ }F$ 关于其第二个参数满足利普希茨连续性，其 Lipschitz 常数为 $\widetilde{L}_2 = \sqrt{3}\left((2+\gamma) \cdot D_\psi \cdot D_\theta + \bar{R}\right)$。

至此，引理6 的证明得以完成，推论4的证明也随之完成。

8 讨论

我们介绍了 MER 算法，这是一种新颖的方法，可在能够访问记忆缓冲区的情况下，利用马尔可夫数据求解随机变分不等式。我们给出了该算法随机误差的上界，表明其误差衰减速度快于基于朴素随机逼近的串行方法。此外，我们还给出了相对于独立同分布设置的双侧期望误差界，以说明 MER 能有效缓解数据依赖性带来的挑战，使该算法能够模拟并紧密追踪专为独立同分布样本设计的随机逼近方法的性能。我们方法的一项关键优势在于，无需预先了解马尔可夫链的混合时间即可实现上述特性；而在相关研究中，这是一项常见但往往限制颇多的要求。我们还提供了数值结果，以证实 MER 算法在解决实际问题方面的有效性，例如广义线性模型中的信号估计以及强化学习中的策略评估。从更广泛的角度来看，MER 算法可视为一种去相关装置，为经验回放启发式方法提供理论依据。我们认为，这种装置可用于解决涉及相关数据的更广泛问题。

我们的工作提出了若干值得关注的未来研究方向，下面列出一些具体的技术问题。首先，当前的 MER 算法只能达到次优的确定性误差。若能将变分不等式的加速技术（例如 Kotsalis 等人 (2022a,b) 纳入其中，从而达到确定性误差下界（Nemirovsky，1992），将会很有意义。其次，尽管 MER 方法达到了阶最优，其随机误差仍有可能在最坏情况保证之外得到进一步锐化。例如，针对策略评估和更一般的不动点方程求解，已有实例依赖下界（Li 等人，2023；Mou 等人，2023），并且随机逼近方法结合方差缩减（Li 等人，2023）以及样本均值近似（Pananjady 和 Wainwright，2020）已经达到了这些下界。未来一个有趣的研究方向，是针对带有记忆缓冲区的马尔可夫设定，建立实例依赖的信息论下界，并设计能够达到这些下界的高效 MER 类算法。

致谢

本研究部分得到美国国家科学基金会项目 CCF-2107455 和 DMS-2210734、谷歌研究学者奖，以及 Adobe、Amazon 和 Mathworks 提供的研究奖项与捐赠资助。

参考文献

A. Auslender 和 M. Teboulle。单调变分不等式的内点投影类方法。数学 -
《数学规划》, 104(1):39–68, 2005. F. Bach、J. Mairal 和 J. Ponce. 凸稀疏矩阵分解. 技术报告, INRIA，2008.

A. Ben-Tal、A. Nemirovski 和 L. El Ghaoui. 鲁棒优化. 普林斯顿大学出版社，2009. A. Benveniste、M. M´etivier 和 P. Priouret. 自适应算法与随机逼近，第 22 卷. 施普林格科学与商业媒体，2012. D. Bertsekas. 强化学习与最优控制，第 1 卷. Athena Scientific，2019. D. P. Bertsekas. 非线性规划. 运筹学会志，48(3):334–334，1997. J. Bhandari、D. Russo 和 R. Singal. 线性函数逼近下时序差分学习的有限时间分析. 收录于 学习理论会议，第 1691–1692 页. PMLR，2018. V. S. Borkar. 随机逼近：动力系统视角，第 9 卷. Springer，2008. V. S. Borkar 和 S. P. Meyn. 用于随机逼近和强化学习收敛的常微分方程方法. *SIAM*控制与优化杂志，38(2):447–469，2000. F. E. Browder. 非线性变分不等式解的存在性与逼近. 美国国家科学院院刊，56(4):1080–1086，1966. A. Chambolle 和 T. Pock. 一种用于凸问题并应用于成像的一阶原始-对偶算法. 数学成像与视觉杂志，40:120–145，2011. Y. Chen、G. Lan 和 Y. Ouyang. 一类变分不等式的加速方案.《数学规划》，165:113–149，2017. Z. Chen、S. T. Maguluri、S. Shakkottai 和 K. Shanmugam. 马尔可夫随机逼近有限样本保证的李雅普诺夫理论.《运筹学》，72(4):1352–1367，2024. S. Cui 和 U. V. Shanbhag. 单投影方案在单调随机变分不等式问题中的方差缩减与随机投影变体分析. 集值与变分分析，29:453–499，2021. C. D. Dang 和 G. Lan. 广义单调算子变分不等式非欧氏外梯度方法的收敛性质. 计算优化与应用，60:277–310，2015. S. Di-Castro、S. Mannor 和 D. Di Castro. 通过回放缓冲区分析随机过程. 收录于国际机器学习会议，第 5039–5060 页. PMLR，2022. R. Dorfman 和 K. Y. Levy. 在具有马尔可夫数据的随机优化中适应混合时间. 收录于国际机器学习会议，第 5429–5446 页. PMLR，2022. J. C. Duchi、A. Agarwal、M. Johansson 和 M. I. Jordan. 遍历镜像下降. *SIAM* 优化杂志，22(4):1549–1578，2012. A. Durmus、E. Moulines、A. Naumov、S. Samsonov 和 H.-T. Wai. 带马尔可夫噪声的随机矩阵乘积的稳定性：在线性随机逼近和 TD 学习中的应用. 收录于学习理论会议，第 1711–1752 页. PMLR，2021. W. Fedus、P. Ramachandran、R. Agarwal、Y. Bengio、H. Larochelle、M. Rowland 和 W. Dabney. 重新审视经验回放的基础. 收录于国际机器学习会议，第 3061–3071 页. PMLR，2020. I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio. 生成对抗网络. *ACM*通信，63(11):139–144，2020.

S. U. Haque 和 S. T. Maguluri。具有无界马尔可夫噪声的随机逼近：一个通用定理。*arXiv* 预印本 *arXiv:2410.21704*，2024。J. Harold、G. Kushner 和 G. Yin。随机逼近、递归算法及其应用。《数学应用》，35(10)，1997。S. Ilandarideva、A. Juditsky、G. Lan 和 T. Li。具有状态依赖噪声的加速随机逼近。《数学规划》，第 1–42 页，2024。A. N. Iusem、A. Jofr´e、R. I. Oliveira 和 P. Thompson。用于随机变分不等式的方差缩减超梯度法。*SIAM* 优化杂志，27(2):686–724，2017。C. Ju 和 G. Lan。在线强化学习的自动探索。*arXiv* 预印本 *arXiv:2512.06244*，2025。A. Juditsky 和 A. Nemirovski。通过凸优化进行统计推断。普林斯顿大学出版社，2020。A. Juditsky、A. Nemirovski 和 C. Tauvel。使用随机镜像近端算法求解变分不等式。《随机系统》，1(1):17–58，2011。K. Khamaru、A. Pananjady、F. Ruan、M. J. Wainwright 和 M. I. Jordan。时序差分学习是否最优？实例依赖分析。《*SIAM* 数据科学数学期刊》，3(4):1013–1040，2021。G. M. Korpelevich。用于求解鞍点及其他问题的超梯度法。$Matecon$，12:747–756，1976。G. Kotsalis、G. Lan 和 T. Li。随机变分不等式的简单最优方法，I：算子外推。*SIAM* 优化杂志，32(3):2041–2073，2022a。G. Kotsalis、G. Lan 和 T. Li。随机变分不等式的简单最优方法，II：强化学习中的马尔可夫噪声与策略评估。*SIAM* 优化杂志，32(2):1120–1155，2022b。S. Kowshik、D. Nagaraj、P. Jain 和 P. Netrapalli。使用反向经验回放进行流式线性系统辨识。《神经信息处理系统进展》，34:30140–30152，2021。R. Kumar 和 D. M. Nagaraj。内省式经验回放：受到意外时回顾过去。《机器学习研究汇刊》，2024。N. Lazic、D. Yin、Y. Abbasi-Yadkori 和 C. Szepesvari。正则化策略迭代中改进的遗憾界与经验回放。载于国际机器学习会议，第 6032–6042 页。PMLR，2021。G. Li、C. Cai、Y. Chen、Y. Wei 和 Y. Chi。Q 学习是否达到极小极大最优？紧确的样本复杂度分析。《运筹学》，72(1):222–236，2024。T. Li、G. Lan 和 A. Pananjady。利用线性函数逼近进行加速且实例最优的策略评估。《*SIAM* 数据科学数学期刊》，5(1):174–200，2023。L.-J. Lin。基于强化学习、规划和教学的自我改进反应式智能体。《机器学习》，8:293–321，1992。M. Lou、K. A. Verchand、S. Fridovich-Keil 和 A. Pananjady。准确、可证明且快速的非线性断层重建：一种变分不等式方法。*arXiv* 预印本 *arXiv:2503.19925*，2025。

Y. Malitsky。用于单调变分不等式的投影反射梯度方法。*SIAM* 优化杂志，25(1):502–520，2015。S. P. Meyn 和 R. L. Tweedie。马尔可夫链与随机稳定性。施普林格科学与商业媒体，2012。Y. Min、T. Wang、D. Zhou 和 Q. Gu。利用线性函数逼近进行方差感知离策略评估。《神经信息处理系统进展》，34:7598–7610，2021。V. Mnih、K. Kavukcuoglu、D. Silver、A. A. Rusu、J. Veness、M. G. Bellemare、A. Graves、M. Riedmiller、A. K. Fidjeland、G. Ostrovski 等。通过深度强化学习实现人类水平的控制。《自然》，518(7540):529–533，2015。R. D. Monteiro 和 B. F. Svaiter。关于混合近端超梯度方法在迭代序列和遍历平均方面的复杂度。*SIAM* 优化杂志，20(6):2755–2787，2010。W. Mou、A. Pananjady 和 M. J. Wainwright。投影不动点方程的最优预言机不等式及其在策略评估中的应用。运筹研究数学，48(4):2308–2336，2023。W. Mou、A. Pananjady、M. J. Wainwright 和 P. L. Bartlett。马尔可夫线性随机逼近的最优保证与实例依赖保证。数理统计与学习，7(1):41–153，2024。D. Nagaraj、X. Wu、G. Bresler、P. Jain 和 P. Netrapalli。基于马尔可夫数据的最小二乘回归：基本极限与算法。《神经信息处理系统进展》，33:16666–16676，2020。D. Nagaraj、N. Agarwal、P. Netrapalli、P. Jain 和 S. Chaudhari。采用反向经验回放的在线目标 Q 学习：高效求解线性马尔可夫决策过程中的最优策略。载于国际学习表征会议（*ICLR*），2022。M. Nakul、V. Muthukumar 和 A. Pananjady。逐频率估计平稳质量。载于第三十八届学习理论会议论文集，第 291 卷，机器学习研究论文集，第 4359–4359 页。PMLR，2025。A. Nemirovski。近端方法以 $O(1/t)$ 的收敛速率求解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。*SIAM* 优化杂志，15(1):229–251，2004。A. Nemirovski、A. Juditsky、G. Lan 和 A. Shapiro。用于随机规划的鲁棒随机逼近方法。*SIAM* 优化杂志，19(4):1574–1609，2009。A. S. Nemirovsky。线性算子方程的基于信息的复杂性。《复杂性杂志》，8(2):153–175，1992。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》，109(2):319–344，2007。Y. Nesterov 和 L. Scrimali。求解强单调变分不等式与拟变分不等式。技术报告，CORE，2006。A. Pananjady 和 D. P. Foster。高信号区间中的单指标模型。*IEEE* 信息论汇刊，67(6):4092–4124，2021。A. Pananjady 和 M. J. Wainwright。表格型强化学习中策略评估的实例依赖 ℓ_∞ 界。*IEEE* 信息论汇刊，67(1):566–585，2020。

A. Pananjady、V. Muthukumar 和 A. Thangaraj。Just Wing It：马尔可夫序列中缺失质量的近最优估计。《机器学习研究杂志》，25(312):1–43，2024。M. L. Puterman。《style id='3'>马尔可夫决策过程：离散随机动态规划。约翰·威立父子公司，2014。R. T. Rockafellar。凸函数、单调算子与变分不等式。载于单调算子的理论与应用，第 35–65 页。Citeseer，1969。R. T. Rockafellar。单调算子与近端点算法。*SIAM*控制与优化杂志，14(5):877–898，1976。M. Sibony。M´ethodes it´eratives pour les ´ équations et in´équations aux d´eriv´ees partielles non lin´eaires de type monotone。*Calcolo*，7:65–183，1970。J. C. Spall。随机搜索与优化导论：估计、模拟与控制。约翰·威立父子公司，2005。R. Srikant 和 L. Ying。线性随机逼近和 TD 学习的有限时间误差界。载于学习理论会议，第 2803–2830 页。PMLR，2019。R. S. Sutton。通过时间差分方法学习预测。《机器学习》，3:9–44，1988。R. S. Sutton 和 A. G. Barto。强化学习：导论。麻省理工学院出版社，2018。J. Tsitsiklis 和 B. Van Roy。利用函数逼近分析时间差分学习。《神经信息处理系统进展》，9，1996。Z. Wang、V. Bapst、N. Heess、V. Mnih、R. Munos、K. Kavukcuoglu 和 N. de Freitas。具有经验回放的样本高效演员-评论家。载于国际学习表征会议，2017。W. Wu、Y. Wei 和 A. Rinaldo。马尔可夫链的不确定性量化及其在时序差分学习中的应用。*arXiv* 预印本 *arXiv:2502.13822*，2025。F. Yousefian、A. Nedi´c 和 U. V. Shanbhag。随机变分不等式问题随机逼近方法中的平滑、正则化与平均。《数学规划》，165: 391–431，2017。S. Zhang 和 R. S. Sutton。深入探讨经验回放。*arXiv* 预印本 *arXiv:1712.01275*，2017。L. Zhu、J. Lee 和 Y. Xie。超越最大似然：广义线性模型的变分不等式估计。*arXiv* 预印本 *arXiv:2511.03087*，2025。