arXiv:2303.15740v2 [cs.LG] 17 Sep 2024

压缩随机逼近的集中性：加性噪声与乘性噪声

Zaiwei Chen[1] Siva Theja Maguluri[2],和 Martin Zubeldia[3]1工业工程学院，普渡大学，*chen5252@purdue.edu*2工业与系统工程学院，佐治亚理工学院，*siva.theja@gatech.edu*3工业与系统工程系，明尼苏达大学，*zubeldia@umn.edu*

摘要

本文建立了随机逼近算法在某个任意范数（例如 ℓ_∞-范数）下的压缩算子作用下所生成迭代序列的最大集中界。我们考虑两种迭代序列可能无界的情形：具有有界乘性噪声的随机逼近，以及具有次高斯加性噪声的随机逼近。我们的最大集中不等式表明，在加性噪声情形下，收敛误差具有次高斯尾部；在乘性噪声情形下，收敛误差具有威布尔尾部（其衰减速度快于多项式衰减，但可能慢于指数衰减）。此外，我们还给出了一个不可能性结果，表明在乘性噪声下通常不可能具有次指数尾部。为建立最大集中界，我们提出了一种新颖的自举论证方法：对收敛误差的广义 Moreau 包络的一个修正版的矩母函数进行界定，并构造一个指数超鞅，从而可以使用 Ville 最大不等式。我们在线性随机逼近和强化学习的背景下，展示了理论结果的适用性。

目录

1 引言 **2**
1.1 挑战与我们的技术.41.2 相关文献.6

2 主要结果 **7**
2.1 乘性噪声随机逼近.82.1.1 收紧较小取值时的界 k .92.2 关于尾部衰减率的不可能性结果.10
2.3 次高斯加性噪声随机逼近.12

3 定理2.1 证明 **13**
3.1 时变最坏情况界.143.2 改进界的迭代框架.153.2.1 对数矩母函数的界.163.2.2 指数超鞅.21
3.3 完成自举论证.23

4 定理证明2.3 **23**
4.1 定理 2.3 第 (1) 部分的证明2.3 第 (1) 部分.244.2 定理 2.3 第 (2) 部分的证明2.3 第 (2) 部分.25

5 定理2.4 的证明 **25**
5.1 广义 Moreau 包络的对数矩母函数上界.255.2 最大集中不等式.29

6 应用 **30**
6.1 线性随机逼近.316.2 强化学习.326.2.1 用于预测的时序差分学习.336.2.2 Q-用于控制的学习.35

7 结论 **37**

定理2.1 的技术结果证明 43
A.1 引理的证明3.1 .43A.2 命题的证明3.3 .44A.3 命题的证明3.4 .45A.4 断言的证明3.3 .46A.5 消除对数因子的乘积.49

B 技术结果证明：支持定理 2.3 **51**
B.1 引理证明4.1 .51B.2 引理证明4.2 .51

C 技术 结果证明：支持定理 2.4 5 **1**
C.1 引理证明5.1 .51C.1.1 基于超鞅的方法 +Ville 最大不等式 52

D 技术证明 第 6 节5 **4**
D.1 引理证明6.1 .54D.2 引理证明6.2 .55D.3 引理证明6.3 .55D.4 定理证明6.2 .57

1 引言

随机逼近（SA）方法最早于 [65], 提出，如今已成为现代大规模优化和机器学习的基础工具，这些领域在解决不同领域的许多实际问题方面取得了巨大成功 [40, 43, 58, 68]。更正式地说，SA 是一种用于求解方程组的迭代方法，特别是求解形如 $\bar{F}(x) = x$ 的不动点方程，其中 $\bar{F}: \mathbb{R}^d \mapsto \mathbb{R}^d$ 是一个（可能为非线性的）算子。当算子 $F(\cdot)$ 的显式表达式可获得时，只要 $F(\cdot)$ 是 压缩映射 [1],，即存在 $\gamma_c \in (0, 1)$ 和一个范数 $\|\cdot\|_c$，使得对所有 x_1, $x_2 \in \mathbb{R}^d$ 都有 $\|F(x_1) - F(x_2)\|_c \leq \gamma_c \|x_1 - x_2\|_c$，就可以通过不动点迭代 $x_{k+1} = F(x_k)$ 高效地求解该方程。然而，在许多实际应用中，例如强化学习（RL）或大规模优化，我们没有足够的知识或计算能力来准确计算给定 x 的 $F(x)$，而这正是执行不动点迭代所需要的。为克服这一挑战，随机逼近被提出作为不动点迭代的一种小步长、数据驱动的随机变体，其形式如下：

$$x_{k+1} = x_k + \alpha_k (F(x_k, Y_k) - x_k), \tag{1}$$

其中 Y_k 是一个取值于概率空间 $\mathcal{Y}$中的随机变量， $F\colon \mathbb{R}^d \times \mathcal{Y} \mapsto \mathbb{R}^d$ 是另一个算子[-]（可视为 $F(\cdot)$ 的带噪版本），而 $\alpha_k > 0$ 是步长。步长通常选择为 $\alpha_k = \alpha/(k+h)^z$，其中 α， $h > 0$ 且 $z \in (0, 1]$。我们假设带噪估计 $F(x_k, Y_k)$ 对[-] $F(x_k)$ 是条件无偏的，即对于所有 $k \geq 0$，$\mathbb{E}[F(x_k, Y_k) \mid \mathcal{F}_k] = F(x_k)$，其中 $\mathcal{F}_k$是由 $\{x_0, Y_0, Y_1, \cdots, Y_{k-1}\}$ 生成的 σ-代数。

方程（1）中给出的随机逼近算法涵盖了许多现有的常用算法，并将它们作为特殊情形包括在内。例如，当 $\bar{F}(x) = -c\nabla J(x) + x$ 对于某个强凸且光滑的目标函数 $J\colon \mathbb{R}^d \mapsto \mathbb{R}$，其中 $c > 0$ 可以任意取值时，该算法便退化为用于最小化$J(\cdot)$ [11, 46], 的随机梯度下降（随机梯度下降）算法（SGD），后者也可以建模为压缩随机逼近 [66]。在强化学习的背景下，诸如 Q-学习 [80] 和时序差分（TD）学习 [70] 等常用算法都可以建模为方程（1）的形式[8], ，其中算子 $F(\cdot)$ 与贝尔曼算子 [5]密切相关。由于随机逼近具有广泛的应用，对方程（1）生成的序列 $\{x_k\}$ 的收敛行为进行理论理解具有基础性意义。

早期关于随机逼近的文献主要关注渐近收敛 [10, 44, 65, 74]。近年来，有限样本分析受到了广泛关注 [9, 17, 69]。在有限样本分析中，目标是根据迭代次数 k，界定随机迭代点 x_k 与其极限 x^* 之间的误差[-]（该极限满足 $F(x^*) = x^*$），并研究其衰减率。与渐近分析相比，有限样本分析不仅能更精细地刻画随机逼近迭代序列的行为，还可用于指导实际实现。

由于迭代序列具有随机性，因此有多种方式可以衡量迭代序列与极限点之间的距离。一种自然的方法是使用均方距离 $\mathbb{E}[\|x_k - x^*\|_c^2]$，这在文献中已得到广泛研究 [9, 17, 21, 69, 77]。另一种常用的方法是使用 $\|x_k - x^*\|_c \leq \epsilon$ 对某个 $\epsilon > 0$成立的概率。对该概率的界称为“高概率界”，有时相比均方界更受青睐，因为它不仅能提供收敛速率，还能提供置信水平。然而，总体而言，高概率界更难建立。例如，考虑大数定律的收敛速率 [1]。建立 $\mathcal{O}(1/k)$均方界要比建立指数尾界容易得多，例如霍夫丁不等式 [38], 、切尔诺夫界 [23],和伯恩斯坦不等式 [7], 等。

为了建立由方程（1）生成的 $\{x_k\}$ 的高概率界，随机过程$\{F(x_k, Y_k)\}$ 的性质（除条件无偏外）在分析中也发挥着重要作用。本文考虑两类可能导致迭代序列 $\{x_k\}$无界的噪声序列：有界乘性噪声（即存在 $\sigma > 0$，使得对所有 $k \geq 0$， $\|F(x_k, Y_k) - \mathbb{E}[F(x_k, Y_k) \mid \mathcal{F}_k]\|_c \leq \sigma(1 + \|x_k\|_c)$）或次高斯加性噪声（即随机向量 $F(x_k, Y_k) - \mathbb{E}[F(x_k, Y_k) \mid \mathcal{F}_k]$ 是范数次高斯的）。在现有文献中，大多数结果关注的是随机逼近算法中的噪声为加性且几乎必然（a.s.）有界的情形。举例而言，考虑如下形式的线性随机逼近：$x_{k+1} = x_k + \alpha_k(A(Y_k)x_k - b(Y_k))$，其中 $A\colon \mathcal{Y} \mapsto \mathbb{R}^{d\times d}$ 和 $b\colon \mathcal{Y} \mapsto \mathbb{R}^d$ 是确定性函数。这对应于方程（1），其中对任意 $c > 0$， $F(x, y) = c(A(y)x - b(y)) + x$。当 $A(Y_k)$不是随机变量，即 $A(Y_k) = \mathbb{E}[A(Y_k)]$, 时，噪声是纯加性的。在乘性噪声情形中，也就是 $A(Y_k)$为随机变量时，分析要困难得多。现有的高概率界要么其尾部衰减速度不超过多项式，要么需要较强的假设，例如 $A(Y_k)$几乎必然是赫尔维茨矩阵。更详细的文献综述请见第1.2节。

本文针对加性噪声和乘性噪声情形下的压缩随机逼近算法，推导了具有威布尔尾部的最大集中界。本文的主要贡献总结如下。具体结果概览另见表 1 。

• 乘性噪声情形。 当使用 dimin- 时，我们建立了具有威布尔尾部的高概率界。

[1]随机变量序列 $\frac{1}{k}\sum_{i=0}^{k-1} Y_k$ 的平均值可以通过以下方式迭代计算： $x_{k+1} = x_k + \frac{1}{k+1}(-x_k + Y_k)$ 其中 $x_0 = 0$，这是式（1）中随机逼近算法的一个特例。

	乘性噪声下的随机逼近		加性噪声下的随机逼近
	$D \le 0$	$D > 0$	
$z = 1$	$\tilde{\mathcal{O}}\left(\frac{\log(1/\delta)}{k}\right)$	界：$\tilde{\mathcal{O}}\left(\frac{\log^{\lceil 2\alpha D\rceil+1}(1/\delta)}{k}\right)$ 不可能性结果：$\tilde{\Omega}\left(\frac{\log^{2\alpha D+1}(1/\delta)}{k}\right)$	$\tilde{\mathcal{O}}\left(\frac{\log(1/\delta)}{k}\right)$
$z \in (0,1)$	不可能性结果：无法实现威布尔尾部。		$\tilde{\mathcal{O}}\left(\frac{\log(1/\delta)}{k^z}\right)$

表1：我们的结果总结：$\|x_k - x^*\|_c^2$ 的界，以至少 $1-\delta$ 的概率成立

注：参数 z 是步长 $\alpha_k = \alpha/(k+h)^z$ 中的指数。参数 $D = \sigma + \gamma_c - 1$ 取决于算子 $F(\cdot)$ 的压缩因子 γ_c，以及乘性噪声的参数 σ。$\tilde{}$ $\tilde{}$我们用 $\lceil x \rceil$ 表示大于 x 的最小整数。在 $\mathcal{O}(\cdot)$ 和 $\Omega(\cdot)$ 中，忽略对数因子。

采用形式为 $\alpha_k = \alpha/(k+h)$ 的消失步长，其中 $\alpha, h > 0$。重要的是，我们的结果给出了迭代序列整个尾部的界，因为我们的步长既不依赖于期望的精度水平，也不依赖于概率容差水平。此外，我们的界是“最大”的，因为它约束了迭代序列 $\{x_k\}$ 整条轨迹的集中行为。作为集中界的补充，我们给出了不可能性结果，表明使用 $\alpha_k = \alpha/(k+h)$ 时，通常无法实现具有次指数尾部的集中界。此外，当使用 $\alpha_k = \alpha/(k+h)^z$ 并以$z \in (0,1)$ 作为步长时，甚至无法实现威布尔尾。 据我们所知，这是针对具有乘性噪声的随机逼近给出的首个最大集中界（具有威布尔尾）。即使在线性随机逼近这一简单情形下（其随机 $A(Y_k)$ 并非几乎必然为赫尔维茨矩阵），文献中也尚无此类集中结果。

• 加性噪声情形。 我们还考虑纯加性噪声的情形。我们允许噪声无界，但要求其为次高斯。在这种情况下，对于采用 $\alpha_k = \alpha/(k+h)^z$ 形式步长的随机逼近算法，当任意 $z \in (0,1]$ 时，我们建立了具有次高斯尾部的最大集中界。据我们所知，除随机梯度下降这一特殊情形外，文献中尚无关于加性但无界噪声情形的此类集中结果。

• 方法论贡献。 为克服随机逼近算法中存在乘性噪声且迭代序列可能无界所带来的挑战，我们提出了一种新颖的自举论证，其中包括：（1）为收敛误差的广义 Moreau 包络的一个适当修正变体建立矩母函数（MGF）上界，该变体在我们的分析中充当李雅普诺夫函数；（2）构造指数超鞅，并使用 Ville 最大不等式。接下来，我们将进一步介绍这些挑战和我们的技术贡献。

1.1 挑战与我们的方法

我们以含乘性噪声的随机逼近为例，说明其中的挑战及我们采用的技术。含加性噪声的随机逼近分析也遵循类似的方法。

要获得尾部衰减速度快于多项式的高概率界，主要挑战来自无界迭代序列 与乘性噪声的结合。单独来看，这两个因素都不太棘手；但二者结合会形成恶性循环，使噪声的方差可能无界。在这种情况下，尽管期望算子 $F(x_k)$ 是收缩的，但由于"带噪"算子$F(x_k,\ Y_k)$ 以正概率呈扩张性，误差仍可能以相当大的概率极快增长。这带来了一个挑战，目前文献中的任何方法都无法普遍应对。

为克服这一挑战，我们提出一种自举论证，其思路类似于数学归纳法的证明技巧。具体而言，我们首先证明，随机逼近算法的迭代序列虽然不一致有界，但存在一个时变的几乎处处界，该界可能随时间多项式增长。这类似于归纳论证中的基础情形。为进行归纳步骤，假设某个非递减界以一定概率成立（这就是归纳假设）。接着，我们证明，更紧的界必定以稍大的概率成立。这为界的迭代改进提供了蓝图。最后，我们从最坏情况几乎处处界出发，反复使用归纳蓝图（有限次），从而完成证明。接下来，我们将更详细地阐述这一三步证明思路。

步骤1：时变最坏情况界。尽管含有乘性噪声的随机逼近算法的迭代序列并非由某个常数一致有界，但我们证明它们确实存在时变界。此类时变界的表现取决于期望算子中的收缩效应，以及乘性噪声中的扩张效应。一般而言，该界可能随时间增长为多项式。这个时变最坏情况界是我们自举论证的基础情形。

步骤 2：改进界的迭代框架。 我们的自举论证的关键在于通过归纳法建立以下命题。

命题 *1.1.* 给定容差水平 $\delta \in (0, 1)$，假设存在一个非递减序列$\{T_k(\delta)\}$ 使得 $\mathbb{P}(\|x_k - x^*\|_c^2 \leq T_k(\delta)\ \forall\ k \geq 0) \geq 1 - \delta$。那么，对于任意 $\delta' \in (0, 1-\delta)$，必然存在一个序列 $\{T_k(\delta, \delta')\}$，满足 $T_k(\delta, \delta') = \tilde{\mathcal{O}}(T_k(\delta)/k)$，使得 $\mathbb{P}(\|x_k - x^*\|_c^2 \leq T_k(\delta, \delta')\ \forall\ k \geq 0) \geq 1 - \delta - \delta'$。

这一结果使我们能够从误差的时变最坏情况界（该界可能呈多项式增长）出发，通过迭代改进，获得具有威布尔尾部的集中界以及所需的收敛速率。为建立命题 1.1，我们采用李雅普诺夫方法，该方法包括以下两个步骤。

• 步骤 *2.1*：对数矩母函数的递归上界。第一步是为范数平方函数的广义 Moreau 包络的一个变体建立对数矩母函数的递归上界$\|x_k - x^*\|_c^2$。展开这一递归后，我们还可以得到一个直接上界 $\|x_k - x^*\|_c^2$，它仅取决于 $\|x_0 - x^*\|_c^2$和其他模型参数。这些上界对所有 $k \geq 0$ 均成立，使我们能够准确把握噪声对误差的影响。

• 步骤 *2.2*：构造指数超鞅。我们构造一个形式为 $\{M_k\}_{k\geq 0} M_k = \exp(\|x_k - x^*\|_c^2 \alpha_k^{-1} T_k(\delta)^{-1} - C\sum_{i=0}^{k-1} \alpha_k)$的超鞅，其中 $C > 0$是经过适当选择的常数，然后使用 Ville 最大不等式 [76]，得到迭代序列的最大上界。具体而言，该最大上界表明，以高概率，对所有 $k \geq 0$，均有 $\|x_k - x^*\|_c^2 = \tilde{\mathcal{O}}(\alpha_k T_k(\delta))$。由于我们在乘性噪声情形下使用$\alpha_k = \mathcal{O}(1/k)$，因此建立了归纳蓝图。

步骤 3：完成自举论证。 证明我们的最大集中界的最后一步，是以收敛误差的最坏情况界为起点，并反复应用归纳步骤，逐步改进该界。请注意，与可以无限次应用的经典归纳论证不同，由于我们的归纳蓝图要求初始界 $\{T_k(\delta)\}$保持非递减，因此它只能应用有限次。

1.2 相关文献

在介绍问题设定和主要结果之前，我们首先总结一下关于以随机梯度下降、线性随机逼近和强化学习算法形式建立随机逼近算法集中界的相关工作。

随机梯度下降。 关于随机梯度下降（SGD）及其变体的指数型高概率界，已有大量研究。在 [37, 64], 中，作者在噪声条件无偏且迭代序列处于紧集内时，针对非光滑但强凸的目标函数，得到了随机梯度下降的指数型高概率界。随后， [36], 将这一结果推广到次高斯噪声和无界迭代序列的情形，使其成为少数几个在噪声无界时仍能得到指数型高概率界的案例之一。在 [27], 中，针对遍历镜像下降（噪声为马尔可夫型、条件偏置且方差一致有界），也得到了指数型高概率界，但需额外假设迭代序列处于紧集内。近年来， [52] 针对噪声具有重尾的线性模型随机梯度下降，得到了多项式型高概率界。最后，在 [72], 中，作者分析了采用常数步长的镜像下降，其中噪声独立同分布（i.i.d.），且方差一致有界，并且噪声几乎处处有界或为次高斯。通过适当选择常数步长，他们在这一设定下得到了指数型高概率界。

线性随机逼近。 对于线性随机逼近，文献 [45, 69] 给出了常数步长下误差的 ℓ_2-范数的一阶矩界。基于这些结果，可以得到高概率界，但其尾部为多项式型，而非指数型。据我们所知，关于线性随机逼近指数型高概率界的最强结果见于文献 [25]。其中，作者分析了采用递减步长的双时间尺度线性随机逼近，其噪声为乘性的、几乎处处有界的鞅差噪声。在此设置下，通过选择同时依赖于运行时间和置信水平的步长，他们获得了对所有充分大的迭代序列的最大指数型高概率界。另一方面，还有一系列工作主要关注随机矩阵的乘积，然后将这些结果应用于线性随机逼近。在文献 [29], 中，作者考虑了采用常数步长的线性随机逼近，其中噪声为马尔可夫型且几乎处处有界。在此设置下，他们针对随机矩阵乘积建立了高概率界：当随机矩阵几乎处处为赫尔维茨时，得到次指数型高概率界；当随机矩阵仅在期望意义下为赫尔维茨时，则得到多项式型高概率界。随后，文献 [28, 55] 将这一结果推广到了 Polyak-Ruppert 平均迭代的情形。

强化学习中的时序差分学习。 时序差分学习被提出作为一种用于解决强化学习中策略评估问题的随机逼近算法 [70, 71]。时序差分学习的均方界在 [9, 69], 中建立，高概率界则在 [24, 59]中建立。具体而言， [59] 的作者研究了最小二乘时序差分（LSTD）算法（其中包括向紧集的投影步骤），并获得了当步长为 $\mathcal{O}(k^{-1})$ 时，误差的 ℓ_2-范数的指数高概率界。 [24] 的作者分析了采用线性函数逼近和独立同分布采样的时序差分学习，并针对超过某个点的迭代序列，获得了最后一个迭代序列中误差的 ℓ_2-范数的最大指数高概率界，该点的阶为 $\log(1/\delta)$。在异策略设定下，时序差分学习的有限样本均方界在 [16, 18]中建立。据我们所知，目前尚无关于异策略时序差分学习的高概率界（尾部衰减速度快于多项式）的结果，无论是否采用函数逼近。

Q-强化学习中的学习。在关于强化学习中指数型高概率界的早期研究之一中， [31], 作者分析了同步的 Q-学习算法，其中步长为 $\mathcal{O}(k^{-z})$，且 $z \in (1/2,\ 1)$。在这种设定下，他们为所有足够大的迭代序列推导出了指数型高概率界。近期， [48–50], 作者分析了广受欢迎的 Q-学习算法，该算法采用常数步长，并具有一致有界的马尔可夫型、可能存在条件偏置的噪声。在这种设定下，给定运行时间和性能保证，只要运行时间足够长，他们便能在运行结束时获得指数型高概率界。

一般随机逼近。 对于任意范数和递减步长下的一般非线性随机逼近， [17, 21] 的作者获得了误差二阶矩的界。这些矩界可用于获得高概率界，但不具有指数尾。

在 [73], 中，作者研究了采用递减步长并具有鞅差次指数噪声的随机逼近。在此情形下，在迭代序列经过一段时间后足够接近不动点这一事件的条件下，他们获得了最大指数型高概率界。在后续工作 [13], 中，他们假设噪声是乘性的、有界的（几乎处处界），并且是马尔可夫型的。在此情形下，他们无需以某个未知事件为条件即可获得最大指数型高概率界。然而，他们的高概率界仅在经过一段时间后成立，并且界和概率都取决于经过一段时间后的迭代序列的未知范数（该范数是随机的，其分布未知）。

在另一项研究中， [63], 作者在无穷范数下研究了一般随机逼近，其中噪声包含一个几乎处处一致有界的鞅差部分，以及一个仅决定迭代序列中哪个坐标得到更新的马尔可夫型部分。由于这种结构化噪声，随机算子是原算子的条件有偏估计量。在这一设定下，假设迭代序列始终属于某个紧集，他们获得了指数型高概率界。最后，在 [54], 作者在任意巴拿赫空间中研究了一般随机逼近的方差缩减版本，其中采用常数步长以及独立同分布、乘性且几乎处处有界的噪声。通过适当选择步长和用于降低方差的平均化方法，他们获得了误差的指数型高概率界。

总而言之，文献中此前关于随机逼近的所有高概率界都存在以下一项或多项局限：（1）通过投影强制迭代序列属于某个紧集，或者对噪声引入严格的假设，使得迭代序列几乎处处属于某个紧集；（2）根据概率保证和总运行时间调整算法的参数；（3）不允许存在乘性噪声；（4）仅对某个特定迭代序列元素或有限范围内的迭代序列元素有效，而该范围可能取决于概率保证本身。

本文其余部分安排如下。第2节介绍我们的主要结果，包括有界乘性噪声和次高斯加性噪声下随机逼近的最大集中界，以及不可能性结果。第3、4和5节给出主要理论结果的证明。第6节展示主要结果在线性随机逼近和强化学习中的适用性。最后，我们在第7节总结本文。

2 主要结果

考虑使用式（ F）x中给出的随机逼近算法求解不动点方程 $= x$ 。对算子施加以下假设 F（·）。

假设2.1。 存在常数 $\gamma_c \in [0,1)$ 和范数 $\|\cdot\|_c$，使得 $\|F(x_1) - F(x_2)\|_c \leq \gamma_c\|x_1 - x_2\|_c$ 对所有 x_1， $x_2 \in \mathbb{R}^d$均成立。

应用巴拿赫不动点定理 [1], 假设 2.1 可知， $F(x) = x$ 具有唯一解 x^*。我们的结果在 $F(\cdot)$ 为伪压缩算子时同样成立，即对于所有 $\|F(x) - x^*\|_c \leq \gamma_c\|x - x^*\|_c$均有 $x \in \mathbb{R}^d$ [8]。然而，在这种情况下，必须假设 x^* 存在。压缩映射总是伪压缩映射。

我们的下一个假设表明，带噪估计 $F(x_k, Y_k)$ $\bar{F}(x_k)$ 具有无偏扰动 .

假设2.2。 对于所有 $k \geq 0$，均有 $\mathbb{E}[F(x_k, Y_k) \mid \mathcal{F}_k] = F(x_k)$ 几乎处处成立，其中 $\mathcal{F}_k$是由 $\{x_0, Y_0, Y_1, \cdots, Y_{k-1}\}$生成的 σ-代数。

满足假设 2.2 的一种特殊情形是 $\{Y_k\}$ 为一列独立同分布随机变量。假设 2.2 可放宽为 $\|\mathbb{E}[F(x_k, Y_k) \mid \mathcal{F}_k] - F(x_k)\|_c \leq L\|x_k - x^*\|_c$对所有 $k \geq 0$几乎必然成立，其中某个足够小的常数为 $L > 0$。对于具有一般有偏扰动的随机逼近（一个典型例子是 $\{Y_k\}$ 为马尔可夫链的情形），建立高概率界是未来的研究方向。尽管如此，现有允许有偏扰动的结果都要求 $\{x_k\}$ 几乎必然由某个确定性常数所界定，例如 Q-学习和遍历镜像下降。

2.1 乘性噪声下的随机逼近

满足 开始时，我们在以下假设中解释乘性噪声的含义 .

假设 2.3。 存在 $\sigma > 0$，使得 $\|F(x_k,\ Y_k) - \mathbb{E}[F(x_k,\ Y_k) \mid \mathcal{F}_k]\|_c \le \sigma(1 + \|x_k\|_c)$ 几乎必然对所有 $k \ge 0$成立。

满足假设 2.3 的一种特殊情形是，算子 $F(x,\ y)$ 关于 x 一致地利普希茨连续，对所有 y均成立，下面将对此作出正式表述。

假设 2.3′。 存在 $L_c > 0$，使得 $\sup_{y\in\mathcal{Y}} \|F(x_1,\ y) - F(x_2,\ y)\|_c \le L_c\|x_1 - x_2\|_c$对所有$x_1$, $x_2 \in \mathbb{R}^d$均成立，且 $\sup_{y\in\mathcal{Y}} \|F(\mathbf{0},\ y)\|_c < \infty$。

为说明这一含义，在假设 2.3′下，由三角不等式可得

$$\begin{aligned}\|F(x_k, Y_k)\|_c &\le \|F(x_k, Y_k) - F(\mathbf{0}, Y_k)\|_c + \|F(\mathbf{0}, Y_k)\|_c \\ &\le L_c\|x_k\|_c + \sup_{y\in\mathcal{Y}}\|F(\mathbf{0}, y)\|_c \\ &\le \sigma(1 + \|x_k\|_c),\end{aligned} \tag{2}$$

wh其中 $\sigma := \max(L_c,\ \sup_{y\in\mathcal{Y}} \|F(\mathbf{0},\ y)\|_c) < \infty$。此外，詹森不等式意味着 at

$$\|\mathbb{E}[F(x_k, Y_k) \mid \mathcal{F}_k]\|_c \le \mathbb{E}[\|F(x_k, Y_k)\|_c \mid \mathcal{F}_k] \le \sigma(1 + \|x_k\|_c). \tag{3}$$

假设 2.3 可通过将方程（2）和（3）与三角不等式结合得到。注意，假设2.3′ 在线性随机逼近中自动满足，其更新方程为 $x_{k+1} = x_k + \alpha_k$（A（Y_k）$x_k - b$（Y_k））， 其中 A（·） 和 b（·） 是有界函数。事实上，“乘性噪声”这一术语的灵感来源于线性随机逼近。

接下来给出最大集中界。设 $D = \sigma + \gamma_c - 1$，其中 γ_c 是来自假设 2.1 的压缩因子，而 σ 是来自假设 2.3的参数。我们用于表述以下定理的其他参数 c_1、c_1'、c_1''、$\{c_i\}_{2\le i\le 4}$、以及 $D_0 \in$（0，1）都是（取决于问题的）常数，其表达式将在第 3节中给出；该节将给出定理2.1的完整证明。

定理 2.1。 考虑 $\{x_k\}$ 由式（*1*）生成。假设 *2.1 – 2.3* 均成立，且$\alpha_k = \alpha/(k + h)$。那么我们有以下结果。

(1) 当 $D > 0$ 时，选取 $\alpha > 2/D_0$，并令 h 足够大，对于任意 $\delta > 0$ 和 $K \ge 0$，以至少 $1 - \delta$ 的概率，对所有 $k \ge K$，都有[2]

$$\begin{aligned}\|x_k - x^*\|_c^2 \le \frac{c_1\alpha\|x_0 - x^*\|_c^2}{k + h}&\left[\log\left(\frac{m}{\delta}\right) + c_2 + c_3 + c_4\log\left(\frac{k - 1 + h}{h - 1}\right)\right]^{m-1} \\ &\times\left[\log\left(\frac{m}{\delta}\right) + c_2\left(\frac{h}{K + h}\right)^{\alpha D_0/2-1} + c_3 + c_4\log\left(\frac{k - 1 + h}{K - 1 + h}\right)\right],\end{aligned}$$

其中 $m = \lceil 2\alpha D\rceil + 1$。

(2) 当 $D = 0$ 时，选取 $\alpha > 2/D_0$，并令 h 足够大，对于任意 $\delta > 0$ 和 $K \ge 0$，以至少 $1 - \delta$ 的概率，对所有 $k \ge K$，都有

$$\begin{aligned}\|x_k - x^*\|_c^2 \le \frac{c_1'\alpha\|x_0 - x^*\|_c^2}{k + h}&\left[\log\left(\frac{k - 1 + h}{h - 1}\right)\right]^2\left[\log\left(\frac{1}{\delta}\right)\right. \\ &\left. + c_2\left(\frac{h}{K + h}\right)^{\alpha D_0/2-1} + c_3 + c_4\log\left(\frac{k - 1 + h}{K - 1 + h}\right)\right].\end{aligned}$$

[2]可以去掉对数项的乘积（即 $\mathcal{O}(\log(k)^{m-1})$），代价是尾部略有变差。结果见附录A.5。

(3) 当 $D<0$ 时，通过将 $\alpha>2/D_0$ 和 h 取得足够大，对于任意 $\delta>0$ 和 $K\geq 0$，以至少 $1-\delta$ 的概率，对所有 $k\geq K$ 均有

$$\|x_k-x^*\|_c^2\leq \frac{c_1''\alpha\|x_0-x^*\|_c^2}{k+h}\left[\log\left(\frac{1}{\delta}\right)+c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}+c_3+c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right].$$

下面给出几点说明。首先讨论由参数 D决定的尾部。在定理 2.1 (2) 和 (3) 中，由于 $D\leq 0$，且 δ 在平方误差 $\log(1/\delta)$ 中以 $\|x_k-x^*\|_c^2$的形式出现，因此收敛误差 $\|x_k-x^*\|_c$ 具有次高斯尾部。当 $D>0$ 时（参见定理 2.1 (1)），情况则更为微妙。在这种情况下，尾部取决于参数 m。由于 $m=\lceil 2\alpha D\rceil+1$ 和 $D>0$，一般而言，我们只能得到威布尔尾部。 m 关于 D 是 D 的仿射函数（忽略上取整函数）这一事实，直观上是合理的，因为较大的 D 意味着噪声更大的更新，进而意味着更重的尾部[3]。

接下来，我们讨论关于 k 和 K的收敛速率。我们只考虑 $D>0$的情况，这是最有意思的情形。定理 2.1 (1) 指出，以至少 $1-\delta$的概率，所有迭代序列都位于一个锥内，该锥的初始半径为 $\Theta((1+\log^{m/2}(1/\delta))K^{-1/2})$，对应于 $k=K$。这与 [19]中得到的均方误差速率一致。接着，对于任意 $k>K$，该锥的半径为 $\Theta((\log^{m/2}(1/\delta)+\log^{m/2}(k))k^{-1/2})$ 阶。

定理2.1 有若干推论。具体而言，最大集中界可立即推出固定迭代次数下的集中界，进而得到完整尾界。这里我们只给出 $D>0$时的结果。对于 $D\leq 0$ 的情形，可以采用类似的方法。

推论 2.2. 假设定理 *2.1 (1)* 的相同假设成立。

(1) 对任意 $\delta>0$ 和 $k\geq 0$，以至少 $1-\delta$ 的概率，有

$$\|x_k-x^*\|_c^2\leq \frac{c_1\alpha\|x_0-x^*\|_c^2}{k+h}\left[\log\left(\frac{m}{\delta}\right)+c_2+c_3+c_4\log\left(\frac{k-1+h}{h-1}\right)\right]^m.$$

因此，要达到 $\|x_k-x^*\|_c\leq\epsilon$ 所需的样本复杂度为 $\tilde{\mathcal{O}}(\epsilon^{-2}$对数$^m(1/\delta))$

(2) 存在 $C_1>0$，使得对所有 $\epsilon>0$ 和 $k\geq 0$，以下不等式均成立：

$$\mathbb{P}\left(\frac{\sqrt{k+h}\,\|x_k-x^*\|_c}{\log(k)^{m/2}}>\epsilon\right)<m\exp\left(-C_1\epsilon^{2/m}\right).$$

推论 2.2 (1) 可通过在定理 $K=k$ 中令 2.1 (1) 得到；推论 2.2 (2) 则可将推论 δ 2.2 (1) 中的容差水平表示为精度水平 ϵ的函数得到。注意，推论 2.2 (2) 给出了误差 $\|x_k-x^*\|_c$ 在任意 $k\geq 0$下的整个互补累积分布函数（CDF）的上界。因此，我们可以使用公式 $\mathbb{E}[\|x_k-x^*\|_c^r]=\int_0^\infty\mathbb{P}(\|x_k-x^*\|_c^r>x)dx$ (其中 r 为任意正整数)对该界进行积分，从而得到任意时刻误差任意阶矩的上界。

2.1.1 收紧小 k取值下的界

注意，对于 $k=0$以及其他较小的 k值，定理 2.1 中的界可能与真实的初始误差 $\|x_0-x^*\|_c$相差甚远。为了在 k较小时获得更紧的界，我们采用时变最坏情况界（参见命题 3.1）。当 k 时，该界关于 $D>0$呈多项式增长，而且，

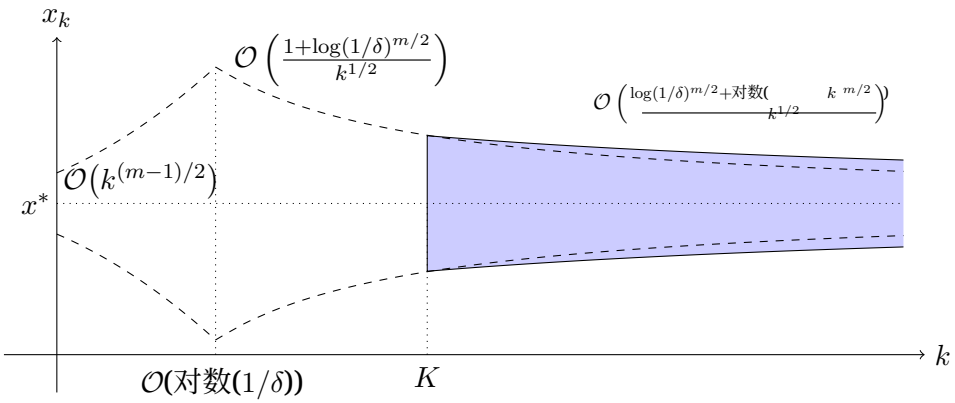

图1　　：当 $D > 0$ 时，所有迭代序列以至少 $1-\delta$ 的概率落在蓝色阴影区域内。

事实上，这是我们证明定理 2.1的第一步。结合这两个界，我们得到一个如图 1所示的非单调界。

从图 1 可以看出，定理 2.1的高概率界与命题 3.1 的最坏情况界相结合后呈现出一种令人惊讶的行为：组合界先增大到某一点，然后开始减小。特别是，该界会一直增大到数量级为 $\mathcal{O}$(对数($1/\delta$))的时刻，此时平方误差的界为 $\mathcal{O}$(对数$(1/\delta)^{m-1}$)的数量级。为了说明这并非我们的证明所造成的假象，接下来我们将基于以下例子进行数值模拟。

例2.1。 考虑如下给定的一维随机逼近

$$x_{k+1} = x_k + \alpha_k(Y_k x_k - x_k), \tag{4}$$

其中， $\{Y_k\}$ 是一个独立同分布随机变量序列，满足 $\mathbb{P}(Y_k = 0.65) = \mathbb{P}(Y_k = 1.15) = 1/2$，且$\alpha_k = 60/(k+120)$。

请注意，公式（4）中的线性随机逼近是公式（1）中的压缩随机逼近的一个特例，其中 $F(x, y) = xy$。此外，容易验证，假设2.1成立，且最小 γ_c 为0.9；假设2.2成立（因为 $\{Y_k\}$ 是独立同分布的）；假设 2.3成立，且最小 σ 为0.25。因此，我们有 $D = \sigma + \gamma_c - 1 = 0.15 > 0$，这与定理 2.1（1）以及图1所示的图相对应。在数值模拟中，我们从初始条件 $x_0 = 10$ 开始，运行方程（4）20,000次，并在图2中绘制若干百分位数。显然，百分位数先增大后减小，这与图1一致。

2.2 关于尾部衰减率的不可能性结果

˜定理 2.1 表明，一般而言，具有乘性噪声的随机逼近能够达到具有威布尔尾部的 $\mathcal{O}(1/k)$ 收敛速率。我们可能会问，是否可以达到次高斯尾部（或次指数尾部）。在本节中，我们通过以下示例表明，只要我们只能获得威布尔尾部，就不可能得到一般性的次指数尾部界。

例 2.2。 考虑1维线性随机逼近，其形式见方程（4）。在此情况下，令 $\{Y_k\}$ 为一个实值随机变量的独立同分布序列，使得 $\mathbb{P}(Y_k = a + N) = 1/(N+1)$ 且 $\mathbb{P}(Y_k = a - 1) =$

[3]请注意，若假设 2.1 和 2.3 对某些 γ_c 和 σ成立，那么对任意 $\gamma_c' \in (\gamma_c, 1)$ 以及 $\sigma' > \sigma$也必然成立。因此，我们总能使 D 为正。然而，根据定理 2.1，随着 D 增大，尾部会变得更重（从次高斯尾部变为威布尔尾部）。因此，为了获得具有最佳尾部衰减率的集中界，应将参数 γ_c 和 σ 视为使假设 2.1 和 2.3 成立的最小值。

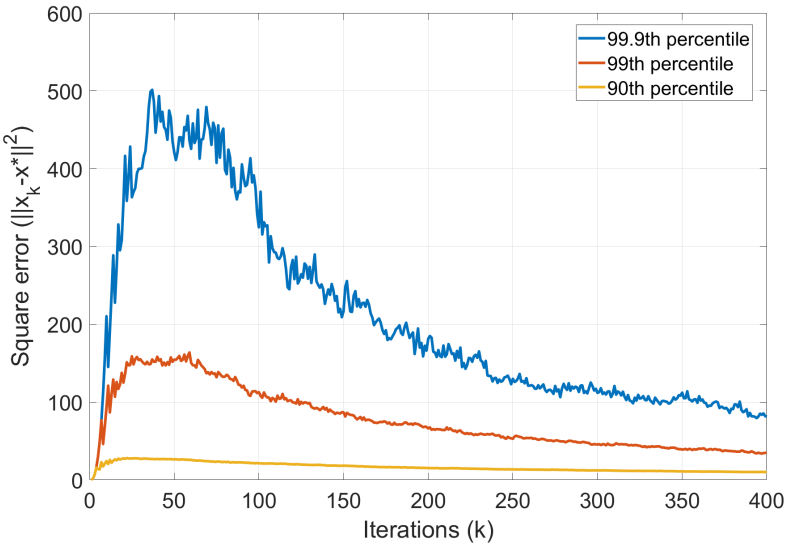

图2：带有 $D > 0$的一维随机逼近的平方误差百分位数。

$N/(N+1)$，其中 $a \in (0, 1)$， $N \geq 1$ 为可调参数。请注意，更新方程也可以等价地写成

$$x_{k+1} = (1 + (Y_k - 1)\alpha_k)x_k. \tag{5}$$

在上述示例中，可以很容易地验证，假设2.1成立，且最小值 γ_c为 a；由于 $\{Y_k\}$ 是一个独立同分布序列，假设2.2成立；最小值 σ 为 N时，假设 2.3成立。因此，我们有 $D = a + N - 1$。接下来，我们应用定理 2.1，为式（5）中的随机逼近算法推导高概率界。假设 $x_0 > 0$, $\alpha_k = \alpha/(k + h)$，其中 $\alpha > 1/(1 - a)$，且 h 足够大，使得 $\alpha_0 < 1/2$。那么，存在 $K_1, K_2 > 0$，使得对所有 $\epsilon > 0$和 $k \geq 0$，以下不等式均成立：

$$\mathbb{P}\left(\frac{\sqrt{k+h}\, x_k}{\log(k)^{m_e/2}} > \epsilon\right) < K_1 \exp\left(-K_2 \epsilon^{\frac{2}{m_e}}\right), \tag{6}$$

其中 $m_e = \lceil 2\alpha D \rceil + 1$。

接下来，我们通过以下定理研究式（5）中随机逼近算法的下界，其证明见第4节。

定理 2.3。考虑 $\{x_k\}$，由式（5）生成。假设 $\alpha_k = \alpha/(k + h)^z$，其中 $z \in (0, 1], \alpha > 0$，且 h 的选取满足 $\alpha_0 < 1/2$。

(1) 当 $z = 1$ 时，对于任意 $\tilde{\beta} > 2/(1 + 2\alpha D)$，都有 $\liminf_{k\to\infty} \mathbb{E}[\exp(\lambda[(k + h)^{1/2} x_k]^{\tilde{\beta}})] = \infty$ 对所有 $\lambda > 0$ 均成立。因此，不存在 K_1'、$K_2' > 0$ 使得 $\mathbb{P}\left((k + h)^{1/2} x_k \geq \epsilon\right) \leq K_1' \exp(-K_2' \epsilon^{\beta})$ 对任意$\epsilon > 0$ 和 $k \geq 0$均成立。

(2) 当 $z \in (0, 1)$ 时，对任意 $\tilde{\beta}, \tilde{\beta}' > 0$，都有 $\liminf_{k\to\infty} \mathbb{E}[\exp(\lambda(k + h)^{\tilde{\beta}'} x_k^{\beta})] = \infty$，对所有 $\lambda > 0$ 均成立。因此，不存在 $\bar{K}_1', \bar{K}_2' > 0$ 使得对任意$\epsilon > 0$ 和 $k \geq 0$，都有 $\mathbb{P}\left((k + h)^{\tilde{\beta}'/\tilde{\beta}} x_k \geq \epsilon\right) \leq \bar{K}_1' \exp\left(-\bar{K}_2' \epsilon^{\tilde{\beta}}\right)$。

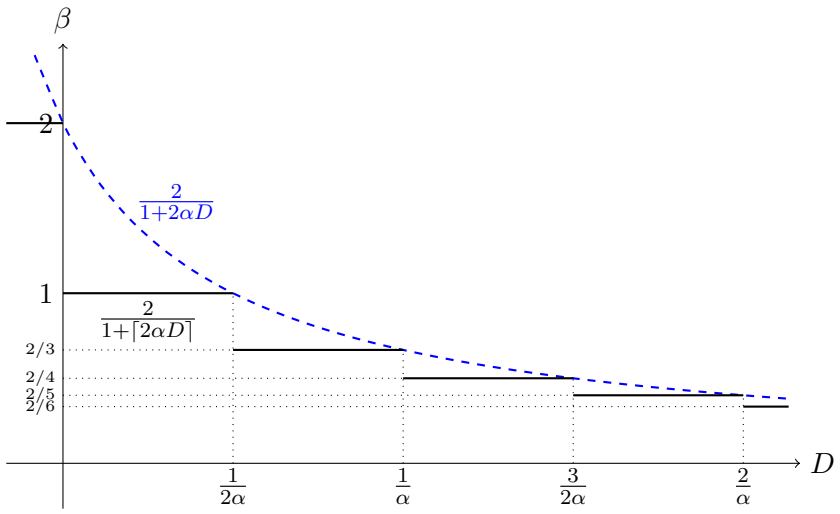

图 3：式（6）中最佳尾部指数（黑色）与定理 2.3 给出的最佳可能尾部指数上界（蓝色虚线）（1）

˜在定理 2.3 **(1)** 中，由于 $\beta > 2/(1+2\alpha D) \geq 2/(1+\lceil 2\alpha D\rceil) = 2/m_e$，我们的集中界几乎是紧的：它具有最佳尾部衰减率（至少在 $2\alpha D$ 为整数时），但在 k中的衰减率稍差。特别是，这意味着只要次指数尾部上界可实现，我们就能得到这样的上界。图3对此作了说明。注意，定理 2.3 **(2)** 表明，当 $\alpha_k = \alpha/(k+h)^z$ （其中 $z \in (0,1)$）时，任何多项式收敛速率下，甚至威布尔尾部界都不可能成立。

2.3 具有次高斯加性噪声的随机逼近

本节还考虑 $\{x_k\}$由式（1）中给出的随机逼近算法生成，但具有加性次高斯噪声的情形，具体说明见以下假设。

假设 2.4。 存在 $\bar{\sigma} > 0$ 和一个（可能依赖于维度的）常数 $c_d > 0$，使得对于任意 $k \geq 0$以及任意 $\mathcal{F}_k$-可测随机向量 v，以下两个不等式成立：

$$\mathbb{E}\left[\exp\left(\lambda\langle F(x_k, Y_k) - \bar{F}(x_k), v\rangle\right)\middle|\mathcal{F}_k\right] \leq \exp\left(\lambda^2\bar{\sigma}^2\|v\|_{c,*}^2/2\right) \tag{7}$$

对所有 $\lambda > 0$成立，且

$$\mathbb{E}\left[\exp\left(\lambda\left\|F(x_k, Y_k) - \bar{F}(x_k)\right\|_c^2\right)\middle|\mathcal{F}_k\right] \leq \left(1 - 2\lambda\bar{\sigma}^2\right)^{-\frac{c_d}{2}} \tag{8}$$

对所有 $\lambda \in \left(0,\ 1/2\bar{\sigma}^2\right)$成立，其中 $\|\cdot\|_c^*$是压缩范数 $\|\cdot\|_c$的对偶范数 [2]。

假设 2.4 可以视为对随机向量范数次高斯这一标准定义的推广，适用于使用任意范数 [39] 而非 $\|\cdot\|_c$的情形 $\|\cdot\|_2$。事实上，当$\|\cdot\|_c = \|\cdot\|_2$ 且 $c_d = d$时，方程（7）和（8正是次高斯随机向量的等价定义 [39, 77]。由于我们使用任意范数，因此允许存在一个（可能）随维度变化的不同常数 c_d。满足假设 2.1、2.2和 2.4 的一种特殊情形是，噪声 Y_k 完全是加性的，并且是鞅差序列，或具有次高斯尾部的独立同分布零均值序列。

接下来，我们给出集中界。在下述定理中，参数 $\{\bar{c}_i\}_{1\le i\le 5}$ 和 $D_1 \in (0,1)$ 是（依赖于问题的）常数，其显式表达式以及该定理的完整证明见第 5 节。

定理 2.4。考虑 $\{x_k\}$ 由公式（1）生成。假设 2.1、2.2和 2.4均成立，并且 $\alpha_k = \alpha/(k+h)^z$，其中 $z \in (0,1]$ 和 α，$h > 0$ 均经过适当选择。则有以下结果。

（1）当 $z = 1$ 时，通过选择 $\alpha > 2/D_1$，对于任意 $\delta > 0$ 和 $K \ge 0$，以至少 $1-\delta$ 的概率，对于所有 $k \ge K$，有

$$\|x_k - x^*\|_c^2 \le \frac{\bar{c}_1 \log(1/\delta)}{k+h} + \bar{c}_2 \|x_0 - x^*\|_c^2 \left(\frac{h}{k+h}\right)^{\bar{D}_1\alpha/2} + \frac{\bar{c}_3 + \bar{c}_4 \log((k+1)/K^{1/2})}{k+h}.$$

（2）当 $z \in (0,1)$ 时，通过选择 $\alpha > 0$ 和 $h \ge (4z/(\bar{D}_1\alpha))^{1/(1-z)}$，对于任意 $\delta > 0$ 和 $K \ge 0$，以至少 $1-\delta$ 的概率，对于所有 $k \ge K$，有

$$\|x_k - x^*\|_c^2 \le \frac{\bar{c}_1 \log(1/\delta)}{(k+h)^z} + \bar{c}_2 \|x_0 - x^*\|_c^2 \exp\left(-\frac{\bar{D}_1\alpha((k+h)^{1-z} - h^{1-z})}{2(1-z)}\right) + \frac{\bar{c}_5 + \bar{c}_4 \log((k+1)/K^{1/2})}{(k+h)^z}.$$

我们将从其对 δ、K和 k 的依赖关系方面讨论定理 2.4 的含义。由于容差水平 δ 在范数平方界中以 $\log(1/\delta)$ 的形式出现，范数误差 $\|x_k - x^*\|_c$ 具有次高斯尾部。关于对 K 和 k的依赖关系，定理 2.4 表明，以至少 $1-\delta$ 的概率，所有迭代序列均位于一个半径为 $\tilde{\Theta}\left((\log(1/\delta)^{1/2} + \log(k/K^{1/2}))\, k^{-z/2}\right)$ 的锥内，对于所有 $k \ge K$。

顺便指出，当 $z = 1$时，为了达到 $\mathcal{O}(1/k)$ 收敛速率，步长中的参数 α必须与零保持一定距离（$\alpha > 2/D_1$ 严格来说）。然而，当 $z < 1$时，参数 α 只需为正，这表明使用 $\alpha_k = \alpha/(k+h)^z$ 步长时的收敛速率虽非最优，却更加稳健。这与研究均方误差 [9, 17]的文献中的观察结果一致。此外，回顾乘性噪声情形，如果我们使用$\alpha_k = \alpha/(k+h)^z$ 作为步长，则不可能得到具有多项式收敛速率和威布尔尾部的收敛界（参见定理 2.3 (2)），这表明，乘性噪声与加性噪声下由随机逼近生成的随机迭代点在行为上存在根本差异。

与第 2.2 节中的推论 2.1类似，在加性噪声情形下，可以利用最大集中界得到阶为 $k^{-z/2}$的固定时间集中界，进而得到完整尾界。这里省略相关结果。

3 定理2.1的证明

我们将遵循第 1.1 节介绍的高层思路来证明该结果。具体而言，在第3.1节中，我们证明由公式（$\{x_k\}$ 1）生成的，尽管不是一致有界的，但满足时变最坏情况界。在第 3.2节中，我们证明命题 1.1，该命题为我们的自举论证提供了蓝图。最后，在第 3.3节中，我们实施自举论证以完成证明：首先将最坏情况界用作初始化，然后利用命题 1.1 逐步改进该界。

3.1 时变最坏情况界

以下命题给出了由公式（ $\{x_k\}$ 中所述的随机逼近算法生成的 1）所满足的时变最坏情况界。回顾 $D = \sigma + \gamma_c - 1$。

命题 3.1. 假设 *2.1* 和 *2.3* 均成立，且 $\alpha_k = \alpha/(k+h)$，其中 α， $h > 0$。那么，我们有 $\|x_k - x^*\|_c \leq B_k(D)$ 对所有 $k \geq 0$几乎必然成立，其中

$$B_k(D) = \begin{cases} \left(\dfrac{k-1+h}{h-1}\right)^{\alpha D}\left(\|x_0 - x^*\|_c + \dfrac{\sigma(1+\|x^*\|_c)}{D}\right), & \text{if } D > 0, \\ \|x_0 - x^*\|_c + \sigma(1+\|x^*\|_c)\alpha \log\left(\dfrac{k-1+h}{h-1}\right), & \text{if } D = 0, \\ \|x_0 - x^*\|_c - \dfrac{\sigma(1+\|x^*\|_c)}{D}, & \text{if } D < 0. \end{cases}$$

注。 直观而言，参数 γ_c 刻画了期望算子的收缩效应，而参数 σ 刻画了噪声的扩张效应。两者的综合影响由参数 $D = \sigma + \gamma_c - 1$刻画。命题 3.1 指出，当 $\|x_k\|_c$ 时，$D < 0$一致有界于一个确定性常数；当 $D = 0$时，增长至多为对数级；当$D > 0$时，则可以按 $\mathcal{O}(k^{\alpha D})$ 的多项式速率增长。

¯命题 *3.1* 的证明。 利用更新方程（1）以及 $F(x^*) = x^*$ 这一事实，对于所有 $k \geq 0$，我们有

$$x_{k+1} - x^* = x_k - x^* + \alpha_k(F(x_k, Y_k) - \bar{F}(x_k) + \bar{F}(x_k) - \bar{F}(x^*) + x^* - x_k).$$

由此可得

$$\begin{aligned} \|x_{k+1} - x^*\|_c &\leq (1-\alpha_k)\|x_k - x^*\|_c + \alpha_k(\|F(x_k, Y_k) - \bar{F}(x_k)\|_c + \|\bar{F}(x_k) - \bar{F}(x^*)\|_c) \\ &\leq (1-\alpha_k)\|x_k - x^*\|_c + \alpha_k(\sigma(1+\|x_k\|_c) + \gamma_c\|x_k - x^*\|_c) \\ &\leq (1-\alpha_k)\|x_k - x^*\|_c + \alpha_k(\sigma\|x_k - x^*\|_c + \sigma(1+\|x^*\|_c) + \gamma_c\|x_k - x^*\|_c) \\ &= (1 + (\sigma + \gamma_c - 1)\alpha_k)\|x_k - x^*\|_c + \alpha_k\sigma(1+\|x^*\|_c) \\ &= (1 + D\alpha_k)\|x_k - x^*\|_c + \alpha_k\sigma(1+\|x^*\|_c), \end{aligned} \tag{9}$$

其中第二个不等式由假设 2.1和假设 2.3得出。接下来，我们需要以下引理，这是求解递归不等式的一般结果。

引理 3.1。 考虑满足以下条件的标量值序列 $\{w_k\}$

$$w_{k+1} \leq (1 + \beta_1\alpha_k)w_k + \beta_2\alpha_k, \quad \forall\, k \geq 0,$$

其中 $w_0 \geq 0$ 和 $\{\alpha_k\}$ 是一个正实数序列。假设 $\beta_1 > -1/\alpha_0$ 和 $\beta_2 > 0$。那么，对于所有 $k \geq 0$ ，我们有

$$w_k \leq \begin{cases} e^{\beta_1 \sum_{i=0}^{k-1}\alpha_i} w_0 + \dfrac{\beta_2}{\beta_1}(e^{\beta_1 \sum_{i=0}^{k-1}\alpha_i} - 1), & \text{if } \beta_1 > 0, \\ w_0 + \beta_2 \displaystyle\sum_{i=0}^{k-1}\alpha_i, & \text{if } \beta_1 = 0, \\ w_0 - \dfrac{\beta_2}{\beta_1}, & \text{if } \beta_1 < 0. \end{cases}$$

引理3.1 的证明见附录A.1。将引理3.1 应用于公式（9），可得

$$\|x_k - x^*\|_c \leq \begin{cases} e^{D\sum_{i=0}^{k-1}\alpha_i}\|x_0 - x^*\|_c + \dfrac{\sigma(1+\|x^*\|_c)}{D}(e^{D\sum_{i=0}^{k-1}\alpha_i} - 1), & \text{if } D > 0, \\ \|x_0 - x^*\|_c + \sigma(1+\|x^*\|_c)\displaystyle\sum_{i=0}^{k-1}\alpha_i, & \text{if } D = 0, \\ \|x_0 - x^*\|_c - \dfrac{\sigma(1+\|x^*\|_c)}{D} & \text{if } D < 0. \end{cases} \tag{10}$$

证明的其余部分是计算当 $\alpha_k = \alpha/(k+h)$ 时的 $\sum_{i=0}^{k-1}\alpha_i$ 和 $e^{D\sum_{i=0}^{k-1}\alpha_i}$。注意到

$$\sum_{i=0}^{k-1}\alpha_i = \sum_{i=0}^{k-1}\frac{\alpha}{i+h} \leq \alpha\int_{-1}^{k-1}\frac{1}{x+h}dx = \alpha\log\left(\frac{k-1+h}{h-1}\right),$$

这进一步意味着

$$\exp\left(D\sum_{i=0}^{k-1}\alpha_i\right) \leq \left(\frac{k-1+h}{h-1}\right)^{\alpha D}.$$

将我们得到的关于 $\sum_{i=0}^{k-1}\alpha_i$ 和 $e^{D\sum_{i=0}^{k-1}\alpha_i}$ 的上界与公式（10）结合起来，即可得到最终结果。

□

3.2 改进该界的迭代框架

既然我们已经建立了随机逼近迭代序列的时变最坏情况界，下一步便是建立命题 1.1。该命题为界的迭代改进提供了蓝图，现将其重述如下，以便于阐述。

命题 3.2。给定一个概率容差水平 $\delta \in (0, 1)$，假设存在一个非递减序列 $\{T_k(\delta)\}$，使得

$$\mathbb{P}(\|x_k - x^*\|_c^2 \leq T_k(\delta), \forall\, k \geq 0) \geq 1 - \delta. \tag{11}$$

那么，对于任意 $\delta' \in (0, 1-\delta)$，存在 $T_k(\delta, \delta') = \mathcal{O}(T_k(\delta)\alpha_k)$ 使得

$$\mathbb{P}(\|x_k - x^*\|_c^2 \leq T_k(\delta, \delta'), \forall\, k \geq 0) \geq 1 - \delta - \delta'. \tag{12}$$

由于 x_0 是确定性地初始化的，因此必有 $T_0(\delta) \geq \|x_0 - x^*\|_c^2$ 几乎处处成立。一旦命题 3.2 得到证明，我们就可以将命题 3.1作为初始化，并利用命题3.2 迭代地改进界，从而证明定理 2.1。为证明命题 3.2，我们采用李雅普诺夫论证。李雅普诺夫函数的构造需要以下定义。

定义 3.1。设 $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ 是一个凸可微函数。当且仅当 $f(\cdot)$ L-光滑（相对于某个范数 $\|\cdot\|$）时，称其为 $f(y) \leq f(x) + \nabla f(x)^\top(y-x) + \frac{L}{2}\|x-y\|^2$ 对所有 x，$y \in \mathbb{R}^d$ 均成立。

受 [17],启发，我们将使用定义如下的广义 Moreau 包络

$$M(x) = \min_{u\in\mathbb{R}^d}\left\{\frac{1}{2}\|u\|_c^2 + \frac{1}{2\mu}\|x-u\|_s^2\right\}$$

作为我们的李雅普诺夫函数，其中 $\|\cdot\|_s$ 是经过选择的平滑范数，使得 $\frac{1}{2}\|\cdot\|_s^2$ 是一个关于 L-光滑函数，且 $\|\cdot\|_s$，而 $\mu > 0$ 是一个可调常数。直观而言，广义 Moreau 包络是对范数平方函数 $\|x\|_c^2$的光滑近似，而该函数通常本身并不光滑（例如，$\|\cdot\|_\infty^2$）。这一点已在 [17, 引理 21] 中得到形式化证明，并在下面的引理中给出，以便完整起见。设 ℓ_{cs} 和 u_{cs} 是满足以下条件的两个正常数：$\ell_{cs}\|x\|_s \le \|x\|_c \le u_{cs}\|x\|_s$ 对所有 $x \in \mathbb{R}^d$均成立。由于有限维空间中范数之间的等价性，这总是可以做到的。不失一般性，我们假设 $\ell_{cs} \in (0,1]$ 和 $u_{cs} \in [1,+\infty)$。

引理 3.2 ([17]的引理 2.1）。广义 *Moreau* 包络 $M(\cdot)$ 具有以下性质：（*1*）函数 $M(\cdot)$ 是凸的，并且关于 $\|\cdot\|_s$ 是 L/μ –光滑的。（*2*）存在一个范数，记为 $\|\cdot\|_M$，使得对所有 $x \in \mathbb{R}^d$，$M(x) = \frac{1}{2}\|x\|_M^2$。（*3*）对所有 $x \in \mathbb{R}^d$，成立 $(1+\mu\ell_{cs}^2)^{1/2}\|x\|_M \le \|x\|_c \le (1+\mu u_{cs}^2)^{1/2}\|x\|_M$。

为简化记号，记 $\ell_{cM} = (1+\mu\ell_{cs}^2)^{1/2}$、$u_{cM} = (1+\mu u_{cs}^2)^{1/2}$ 和 $\tilde{\gamma}_c = \gamma_c u_{cM}/\ell_{cM}$。可调参数 $\mu > 0$ 的选取满足 $\tilde{\gamma}_c < 1$，这总是可行的，因为 $\gamma_c \in (0,1)$ 且 $\lim_{\mu\to 0} u_{cM}/\ell_{cM} = 1$。

接下来回顾一下，在推导 Hoeffding 不等式和切尔诺夫界等经典集中不等式时，一个重要步骤是对所关注随机变量的矩母函数进行界定。完成这一步后，即可结合马尔可夫不等式和矩母函数的界，推导出集中界。受此启发，我们将对广义 Moreau 包络的一个修改版本的矩母函数进行界定。引入这种修改是为了应对迭代序列可能无界的问题$\{x_k\}$。

令 $E_k(\delta) = \{\|x_t - x^*\|_c^2 \le T_t(\delta)\ \forall\, t = 0, 1, \cdots, k\}$。注意，根据定义，$\{E_k(\delta)\}_{k\ge 0}$ 是一个递减事件序列，即对于所有 $k \ge 0$，$E_{k+1}(\delta) \subseteq E_k(\delta)$。此外，根据式 (11)，对于任意 $k \ge 0$，我们有$\mathbb{P}(E_k(\delta)) \ge 1-\delta$。令 $\lambda_k = \theta\alpha_k^{-1}T_k(\delta)^{-1}$，其中 θ 是一个尚待选择的可调常数，α_k 是步长。对于任意 $k \ge 0$，令

$$Z_k = \log\left(\mathbb{E}\left[\exp\left(\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k - x^*)\right)\right]\right), \tag{13}$$

它是广义 Moreau 包络某个修正版的对数矩母函数，后续分析中会频繁用到。为了理解 Z_k 的定义背后的直觉，假设 $\|x_k - x^*\|_c$ 被某个确定性常数一致有界，即 Proposition 3.1 中 $D < 0$ 的情形。此时，我们可以令 $T_k(\delta)$ 等于 Proposition 3.1 (3) 给出的最坏情形常数上界，这意味着 $\mathbb{1}_{E_k(\delta)} = 1$ 几乎必然成立。在这种情况下，Z_k 就是标准的对数矩母函数。由于我们没有如此强的有界性性质，因此引入额外参数 $T_k(\delta)$ 和 $\mathbb{1}_{E_k(\delta)}$；这些参数对于构造自举论证至关重要，而该论证用于克服迭代序列无界这一挑战。

3.2.1 对数矩母函数的界

为了界定 Z_k，我们首先推导一个连接 Z_k 与 Z_{k+1} 的递归上界，进而通过求解递推关系得到 Z_k 的直接上界。为陈述这一结果，我们需要以下记号。令 $D_0 = 2(1-\tilde{\gamma}_c) \in (0,1)$、$D_1 = 4\sigma^2/\ell_{cM}^2$ 和 $D_2 = 2L(2+\sigma)^2 u_{cM}^2/(\mu\ell_{cs}^2)$。参数 θ 取为 $\theta = D_0\|x_0 - x^*\|_c^2/[8D_1((1+\|x^*\|_c)^2 + \|x_0 - x^*\|_c^2)]$。步长按以下条件选取。

条件 3.1。对所有 $k \ge 0$，令 $\alpha_k = \alpha/(k+h)$，其中选择 $\alpha > 2/D_0$ 和 $h > 1$ 使得 $\alpha_0 \le \min(1, D_0, D_0/(4D_2))$。

现在，我们可以在以下命题中给出 Z_k 的递归上界。

命题 3.3。结论成立：

$$Z_{k+1} \leq \exp\left(-\frac{\alpha D_0/2-1}{\alpha}\alpha_k\right) Z_k + 2\alpha_k^2\lambda_k D_2(1+\|x^*\|_c)^2, \quad \forall\, k \geq 0. \tag{14}$$

命题 *3.3* 的证明。利用广义 Moreau 包络的光滑性 $M(\cdot)$（参见引理 3.2 (1)）以及更新方程（1），对于所有 $k \geq 0$，我们有

$$\begin{aligned}
M(x_{k+1}-x^*) &\leq M(x_k-x^*)+\alpha_k\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-x_k)+\frac{L\alpha_k^2}{2\mu}\|F(x_k,Y_k)-x_k\|_s^2 \\
&= M(x_k-x^*)+\alpha_k\nabla M(x_k-x^*)^\top(\bar{F}(x_k)-x_k)+\frac{L\alpha_k^2}{2\mu}\|F(x_k,Y_k)-x_k\|_s^2 \\
&\quad +\alpha_k\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k)) \\
&\leq (1-2\alpha_k(1-\tilde{\gamma}_c))M(x_k-x^*)+\alpha_k\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k)) \\
&\quad +\frac{L\alpha_k^2}{2\mu}\|F(x_k,Y_k)-x_k\|_s^2,
\end{aligned}$$

其中，最后一个不等式由 [19, 引理 A.1]推出。因此，先将$\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}$ 相乘，再对前一个不等式两边取指数，得到

$$\begin{aligned}
&\exp\left(\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}M(x_{k+1}-x^*)\right) \\
\leq\ &\exp\left(\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}(1-2\alpha_k(1-\tilde{\gamma}_c))M(x_k-x^*)\right) \\
&\times\exp\left(\alpha_k\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right) \\
&\times\exp\left(\frac{L\alpha_k^2}{2\mu}\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}\|F(x_k,Y_k)-x_k\|_s^2\right) \\
\leq\ &\exp\left(\lambda_{k+1}\mathbb{1}_{E_k(\delta)}(1-2\alpha_k(1-\tilde{\gamma}_c))M(x_k-x^*)\right) \\
&\times\exp\left(\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right) \\
&\times\exp\left(\frac{L\alpha_k^2}{2\mu}\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\|F(x_k,Y_k)-x_k\|_s^2\right),
\end{aligned}$$

其中，在最后一个不等式中，我们用到了 $\{E_k(\delta)\}$ 是递减事件序列这一事实（这意味着 $\mathbb{1}_{E_{k+1}(\delta)} \leq \mathbb{1}_{E_k(\delta)}$）。对前一个不等式两边关于 $\mathcal{F}_k$ 取条件期望，我们得到

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}M(x_{k+1}-x^*)\right)\mid\mathcal{F}_k\right] \\
\leq\ &\exp\left(\lambda_{k+1}\mathbb{1}_{E_k(\delta)}(1-2\alpha_k(1-\tilde{\gamma}_c))M(x_k-x^*)\right) \\
&\times\mathbb{E}\left[\exp\left(\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right)\right. \\
&\left.\times\exp\left(\frac{L\alpha_k^2}{2\mu}\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\|F(x_k,Y_k)-x_k\|_s^2\right)\ \middle|\ \mathcal{F}_k\right] \\
\leq\ &\exp\left(\lambda_{k+1}\mathbb{1}_{E_k(\delta)}(1-2\alpha_k(1-\tilde{\gamma}_c))M(x_k-x^*)\right) \\
&\times\underbrace{\mathbb{E}\left[\exp\left(2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right)\mid\mathcal{F}_k\right]^{1/2}}_{T_1}
\end{aligned}$$

$$\times \underbrace{\mathbb{E}\left[\exp\left(\frac{L\alpha_k^2}{\mu}\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\|F(x_k,Y_k)-x_k\|_s^2\right)\,\middle|\,\mathcal{F}_k\right]^{1/2}}_{T_2}, \tag{15}$$

其中，最后一行由条件柯西–施瓦茨不等式得出。接下来，我们对 T_1 和 T_2 进行界定。为了界定 T_1，我们将使用条件霍夫丁引理。注意，假设2.2 蕴含

$$\begin{aligned}
&\mathbb{E}\left[2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\nabla M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\mid\mathcal{F}_k\right]\\
&=2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\nabla M(x_k-x^*)^\top(\mathbb{E}\left[F(x_k,Y_k)\mid\mathcal{F}_k\right]-\bar{F}(x_k))\\
&=0.
\end{aligned} \tag{16}$$

此外，我们有

$$\begin{aligned}
&\left|2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right|\\
&\le 2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\|\nabla M(x_k-x^*)\|_M^*\|F(x_k,Y_k)-\bar{F}(x_k)\|_M &(17)\\
&\le 2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\|x_k-x^*\|_M\|\nabla\|x_k-x^*\|_M\|_M^*\|F(x_k,Y_k)-\bar{F}(x_k)\|_M &(18)\\
&\le 2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}\|x_k-x^*\|_M\|F(x_k,Y_k)-\bar{F}(x_k)\|_M &(19)\\
&\le \frac{2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}}{\ell_{cM}}\|x_k-x^*\|_M\|F(x_k,Y_k)-\bar{F}(x_k)\|_c &(20)\\
&\le \frac{2\sigma\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}}{\ell_{cM}}\|x_k-x^*\|_M(1+\|x_k\|_c) &(21)\\
&\le \frac{2\sigma\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}}{\ell_{cM}}\|x_k-x^*\|_M(1+\|x_k-x^*\|_c+\|x^*\|_c)\\
&\le \frac{2\sigma\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}}{\ell_{cM}}\|x_k-x^*\|_M(1+\|x^*\|_c+T_k^{1/2}(\delta)), &(22)
\end{aligned}$$

其中，方程（17）由霍尔德不等式得出，其中 $\|\cdot\|_M^*$是 $\|\cdot\|_M$的对偶范数；方程（18）由 $M(x)=\frac{1}{2}\|x\|_M^2$（参见引理 3.2）以及微积分链式法则得出；对于所有 $x\in\mathbb{R}^d$ [67, , 引理 2.6], 中的 $\|\nabla\|x\|_M\|_M^*\le 1$ 导出方程（19）；方程（20）由引理 3.2得出；方程（21）由假设2.3得出；方程（22）由 $\|x_k-x^*\|_c^2\le T_k(\delta)$ 在 $E_k(\delta)$ 上得出。方程（16）和（22）结合后，使我们能够对随机变量 $2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))$ 使用条件霍夫丁引理，从而得到

$$\begin{aligned}
T_1&=\mathbb{E}\left[\exp\left(2\alpha_k\lambda_{k+1}\mathbb{1}_{E_k(\delta)}M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right)\,\middle|\,\mathcal{F}_k\right]^{1/2}\\
&\le\exp\left(\frac{\sigma^2\alpha_k^2\lambda_{k+1}^2\mathbb{1}_{E_k(\delta)}}{\ell_{cM}^2}\|x_k-x^*\|_M^2(1+\|x^*\|_c+T_k^{1/2}(\delta))^2\right)\\
&\le\exp\left(\frac{4\sigma^2\alpha_k^2\lambda_{k+1}^2\mathbb{1}_{E_k(\delta)}}{\ell_{cM}^2}M(x_k-x^*)[(1+\|x^*\|_c)^2+T_k(\delta))]\right),
\end{aligned} \tag{23}$$

其中，最后一个不等式由 $M(x)=\frac{1}{2}\|x\|_M^2$（参见引理3.2）以及 $(a+b)^2\le 2(a^2+b^2)$ 对所有 $a,\ b\in\mathbb{R}$成立。

接下来，我们对下式中的项进行界定：T_2 见式（15）。注意到

$$\|F(x_k,Y_k)-x_k\|_s\le\frac{1}{\ell_{cs}}\|F(x_k,Y_k)-x_k\|_c$$

$$
\begin{aligned}
&= \frac{1}{\ell_{cs}} \left\| F(x_k, Y_k) - \bar{F}(x_k) + \bar{F}(x_k) - \bar{F}(x^*) + x^* - x_k \right\|_c \\
&\leq \frac{1}{\ell_{cs}} \left(\|F(x_k, Y_k) - \bar{F}(x_k)\|_c + \|\bar{F}(x_k) - \bar{F}(x^*)\|_c + \|x^* - x_k\|_c \right) \\
&\leq \frac{1}{\ell_{cs}} \left(\sigma(1 + \|x_k\|_c) + (\gamma_c + 1)\|x_k - x^*\|_c \right) \qquad (24) \\
&\leq \frac{1}{\ell_{cs}} \left(\sigma(1 + \|x_k - x^*\|_c + \|x^*\|_c) + 2\|x_k - x^*\|_c \right) \\
&\leq \frac{1}{\ell_{cs}} \left((2 + \sigma)\|x_k - x^*\|_c + \sigma(1 + \|x^*\|_c) \right),
\end{aligned}
$$

其中，式（24）由假设2.1 和2.3可得。因此，我们有

$$
\begin{aligned}
T_2 &= \mathbb{E}\left[\exp\left(\frac{L\alpha_k^2}{\mu} \lambda_{k+1} \mathbb{1}_{E_k(\delta)} \|F(x_k, Y_k) - x_k\|_s^2 \right) \,\middle|\, \mathcal{F}_k \right]^{1/2} \\
&\leq \exp\left(\frac{L\alpha_k^2 \lambda_{k+1} \mathbb{1}_{E_k(\delta)}}{2\mu\ell_{cs}^2} \left((2 + \sigma)\|x_k - x^*\|_c + \sigma(1 + \|x^*\|_c) \right)^2 \right) \\
&\leq \exp\left(\frac{L\alpha_k^2 \lambda_{k+1} \mathbb{1}_{E_k(\delta)}}{\mu\ell_{cs}^2} \left((2 + \sigma)^2 \|x_k - x^*\|_c^2 + \sigma^2 (1 + \|x^*\|_c)^2 \right) \right) \\
&\leq \exp\left(\frac{L\alpha_k^2 \lambda_{k+1} \mathbb{1}_{E_k(\delta)}}{\mu\ell_{cs}^2} \left((2 + \sigma)^2 u_{cM}^2 \|x_k - x^*\|_M^2 + \sigma^2 (1 + \|x^*\|_c)^2 \right) \right) \\
&\leq \exp\left(\frac{2L(2 + \sigma)^2 u_{cM}^2 \alpha_k^2 \lambda_{k+1} \mathbb{1}_{E_k(\delta)}}{\mu\ell_{cs}^2} M(x_k - x^*) + \frac{L\sigma^2 \alpha_k^2 \lambda_{k+1}}{\mu\ell_{cs}^2} (1 + \|x^*\|_c)^2 \right).
\end{aligned}
$$

利用我们对 T_1 （参见公式（23））和 T_2 （参见前一个不等式）在公式（15）中得到的上界，我们有

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_{k+1} \mathbb{1}_{E_{k+1}(\delta)} M(x_{k+1} - x^*) \right) \mid \mathcal{F}_k \right] \\
&\leq \exp\left(\lambda_{k+1} \mathbb{1}_{E_k(\delta)} (1 - 2\alpha_k(1 - \tilde{\gamma}_c)) M(x_k - x^*) \right) \\
&\quad \times \exp\left(\frac{4\sigma^2 \alpha_k^2 \lambda_{k+1}^2 \mathbb{1}_{E_k(\delta)}}{\ell_{cM}^2} M(x_k - x^*)[(1 + \|x^*\|_c)^2 + T_k(\delta))] \right) \\
&\quad \times \exp\left(\frac{2L(2+\sigma)^2 u_{cM}^2 \alpha_k^2 \lambda_{k+1} \mathbb{1}_{E_k(\delta)}}{\mu\ell_{cs}^2} M(x_k - x^*) + \frac{L\sigma^2 \alpha_k^2 \lambda_{k+1}}{\mu\ell_{cs}^2} (1 + \|x^*\|_c)^2 \right) \\
&= \exp\left(\lambda_{k+1} \mathbb{1}_{E_k(\delta)} M(x_k - x^*) \left(1 - 2\alpha_k(1 - \tilde{\gamma}_c) \right. \right. \\
&\quad \left. \left. + \frac{4\sigma^2 \alpha_k^2 \lambda_{k+1} [(1 + \|x^*\|_c)^2 + T_k(\delta))]}{\ell_{cM}^2} + \frac{2L(2 + \sigma)^2 u_{cM}^2 \alpha_k^2}{\mu\ell_{cs}^2} \right) \right) \\
&\quad \times \exp\left(\frac{L\sigma^2 \alpha_k^2 \lambda_{k+1}}{\mu\ell_{cs}^2} (1 + \|x^*\|_c)^2 \right).
\end{aligned}
$$

回顾一下，我们将 $D_0 = 2(1 - \tilde{\gamma}_c)$、 $D_1 = 4\sigma^2/\ell_{cM}^2$ 和 $D_2 = 2L(2 + \sigma)^2 u_{cM}^2/(\mu\ell_{cs}^2)$记作上述形式。此外，为简化记号，令

$$
T_3 = \frac{\lambda_{k+1}}{\lambda_k} \left(1 - \alpha_k D_0 + \alpha_k^2 \lambda_{k+1} D_1 ((1 + \|x^*\|_c)^2 + T_k(\delta)) + \alpha_k^2 D_2 \right), \qquad (25)
$$

$$T_4 = \alpha_k^2 \frac{\lambda_{k+1}}{\lambda_k}. \tag{26}$$

那么，前一个不等式可写为

$$\begin{aligned} &\mathbb{E}\left[\exp\left(\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)} M(x_{k+1} - x^*)\right) \mid \mathcal{F}_k\right] \\ &\leq \exp\left(T_3 \lambda_k \mathbb{1}_{E_k(\delta)} M(x_k - x^*)\right) \exp\left(T_4 \lambda_k D_2 (1 + \|x^*\|_c)^2\right). \end{aligned} \tag{27}$$

接下来，我们对方程（ T_3 和 T_4 （25）以及（26）中分别定义的项进行界定。一方面，显然有 $T_3 \geq 0$ ，因为 $\alpha_0 \leq D_0$。另一方面，由于 $\lambda_k = \theta \alpha_k^{-1} T_k$（$\delta$）$^{-1}$，我们有

$$\begin{aligned} T_3 &= \frac{\lambda_{k+1}}{\lambda_k}\left(1 - \alpha_k D_0 + \alpha_k^2 \lambda_{k+1} D_1((1 + \|x^*\|_c)^2 + T_k(\delta)) + \alpha_k^2 D_2\right) \\ &= \frac{\alpha_k T_k(\delta)}{\alpha_{k+1} T_{k+1}(\delta)}\left(1 - \alpha_k D_0 + \frac{\theta \alpha_k^2 D_1((1 + \|x^*\|_c)^2 + T_k(\delta))}{\alpha_{k+1} T_{k+1}(\delta)} + \alpha_k^2 D_2\right) \\ &\leq \frac{\alpha_k}{\alpha_{k+1}}\left(1 - \alpha_k D_0 + \frac{\theta \alpha_k^2 D_1((1 + \|x^*\|_c)^2 + T_k(\delta))}{\alpha_{k+1} T_{k+1}(\delta)} + \alpha_k^2 D_2\right), \end{aligned} \tag{28}$$

其中，最后一行源于 $\{T_k(\delta)\}$ 是一个非递减序列。接下来，注意到

$$\frac{\theta \alpha_k^2 D_1((1 + \|x^*\|_c)^2 + T_k(\delta))}{\alpha_{k+1} T_{k+1}(\delta)} \leq \frac{\theta \alpha_k^2 D_1}{\alpha_{k+1}} \frac{(1 + \|x^*\|_c)^2 + T_{k+1}(\delta)}{T_{k+1}(\delta)} \tag{29}$$

$$\leq \frac{\theta \alpha_k^2 D_1}{\alpha_{k+1}} \frac{(1 + \|x^*\|_c)^2 + \|x_0 - x^*\|_c^2}{\|x_0 - x^*\|_c^2} \tag{30}$$

$$= \frac{\alpha_k^2 D_0}{8\alpha_{k+1}} \tag{31}$$

$$\leq \frac{\alpha_k D_0}{4}, \tag{32}$$

其中，式 (29) 由 $\{T_k(\delta)\}$ 是非递减序列得出，式 (30) 由 $\|x_0 - x^*\|_c^2 \leq T_0(\delta) \leq T_{k+1}(\delta)$ 以及数值不等式 $(a + c)/(b + c) \leq a/b$ （对任意 $a, b > 0$ ，满足 $a \geq b$ 和 $c > 0$）得出，式 (31) 由选择 $\theta = D_0 \|x_0 - x^*\|_c^2 / [8D_1((1 + \|x^*\|_c)^2 + \|x_0 - x^*\|_c^2)]$，得出，式 (32) 则由 $\alpha_k/\alpha_{k+1} \leq (h + 1)/h \leq 2$ 得出（参见条件 3.1）。因此，将式 (28) 中的前述不等式代入，可得

$$\begin{aligned} T_3 &\leq \frac{\alpha_k}{\alpha_{k+1}}\left(1 - \alpha_k D_0 + \frac{\alpha_k D_0}{4} + \alpha_k^2 D_2\right) \\ &\leq \frac{\alpha_k}{\alpha_{k+1}}\left(1 - \frac{\alpha_k D_0}{2}\right) \qquad (33) \\ &= \frac{k + 1 + h}{k + h}\left(1 - \frac{\alpha D_0}{2(k + h)}\right) \\ &\leq \left(\frac{k + h + 1}{k + h}\right) \exp\left(-\frac{D_0 \alpha}{2(k + h)}\right) \qquad (34) \\ &= \left[\left(1 + \frac{1}{k + h}\right)^{k+h}\right]^{1/(k+h)} \exp\left(-\frac{D_0 \alpha}{2(k + h)}\right) \\ &\leq \exp\left(\frac{1}{k + h} - \frac{D_0 \alpha}{2(k + h)}\right) \qquad (35) \end{aligned}$$

$$= \exp\left(-\frac{\alpha D_0/2-1}{k+h}\right)$$
$$= \exp\left(-\frac{\alpha D_0/2-1}{\alpha}\alpha_k\right), \tag{36}$$

其中，式 (33) 由 $\alpha_k \le D_0/(4D_2)$ 得出 （参见条件 3.1），式 (34) 由 $1+x \le e^x$ 对所有$x \in \mathbb{R}$，而式 (35) 由 $(1+1/x)^x \le e$ 对所有 $x>0$ 得出。注意， $T_3 < 1$ 因为 $\alpha > 2/D_0$ （参见条件 3.1）。

现在，考虑式 T_4 （26）中的项。根据 λ_k 的定义，我们有

$$T_4 = \alpha_k^2 \frac{\lambda_{k+1}}{\lambda_k} = \alpha_k^2 \frac{\alpha_k T_k(\delta)}{\alpha_{k+1} T_{k+1}(\delta)} \le \frac{\alpha_k^3}{\alpha_{k+1}} \le 2\alpha_k^2,$$

其中我们利用了 $\{T_k（\delta）\}$ 是一个非递减序列，以及 $\alpha_k/\alpha_{k+1} \le （h+1）/h \le 2$ （参见条件3.1）。将我们得到的项 T_3 （参见式（36））和 T_4 （参见前一个不等式）的上界代入式（27），可得

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}M(x_{k+1}-x^*)\right) \mid \mathcal{F}_k\right]\\ \le\ &\exp\left(\exp\left(-\frac{\alpha D_0/2-1}{\alpha}\alpha_k\right)\lambda_k \mathbb{1}_{E_k(\delta)}M(x_k-x^*)\right)\exp\left(2\alpha_k^2\lambda_k D_2(1+\|x^*\|_c)^2\right).\end{aligned} \tag{37}$$

鉴于式（13）中Z_k 的定义，可通过先取全期望，再取对数，最后对式（37）两边应用詹森不等式，得到 Z_k 的递归上界。这证明了命题3.3。

□

反复使用命题3.3 中的式（14），可得到 Z_k的总体界。结果陈述如下引理，其证明只涉及标准的代数运算，故推迟至附录A.2。

引理 3.3。 对所有 $k \ge 0$，均有

$$Z_k \le Z_0\left(\frac{h}{k+h}\right)^{\alpha D_0/2-1} + \frac{4\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2}, \quad \forall\, k \ge 0. \tag{38}$$

现在，我们已成功建立修正对数矩母函数的递归上界和总体界 Z_k，下一步是利用 Z_k 构造一个超鞅，并应用 Ville 最大不等式，以完成对命题3.2的证明。

3.2.2 指数超鞅

对于任意 $k \ge 0$，令 M_k定义为 $M_k = \exp(\lambda_k \mathbb{1}_{E_k(\delta)}M(x_k - x^*) - D_3\sum_{i=0}^{k-1}\alpha_i)$，其中 $D_3 = D_0D_2/(4D_1)$。接下来我们证明，$\{M_k\}$ 关于滤子 $\{\mathcal{F}_k\}$是一个超鞅。显然， $\{M_k\}$ 适应于 $\{\mathcal{F}_k\}$，且其期望有限（参见引理3.3）。此外，对于任意 $k \ge 0$，由式（37）可得

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\lambda_{k+1}\mathbb{1}_{E_{k+1}(\delta)}M(x_{k+1}-x^*)\right) \mid \mathcal{F}_k\right]\\ \le\ &\exp\left(\lambda_k \mathbb{1}_{E_k(\delta)}M(x_k-x^*)\right)\exp\left(2\alpha_k^2\lambda_k D_2(1+\|x^*\|_c)^2\right)\\ =\ &\exp\left(\lambda_k \mathbb{1}_{E_k(\delta)}M(x_k-x^*)\right)\exp\left(\frac{2\alpha_k\theta D_2(1+\|x^*\|_c)^2}{T_k(\delta)}\right)\end{aligned} \tag{39}$$

$$\leq \exp\left(\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k - x^*)\right) \exp\left(\frac{2\alpha_k \theta D_2 (1 + \|x^*\|_c)^2}{\|x_0 - x^*\|_c^2}\right) \tag{40}$$

$$\leq \exp\left(\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k - x^*)\right) \exp\left(\frac{\alpha_k D_0 D_2}{4D_1}\right) \tag{41}$$

$$= \exp\left(\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k - x^*)\right) \exp\left(D_3 \alpha_k\right), \tag{42}$$

其中，式（39）由 $\lambda_k = \theta \alpha_k^{-1} T_k(\delta)^{-1}$ 得出，式（40）由 $\|x_0 - x_k\|_c^2 \leq T_0(\delta) \leq T_k(\delta)$ 得出，式（41）由选取 $\theta = D_0 \|x_0 - x^*\|_c^2 / [8D_1((1 + \|x^*\|_c)^2 + \|x_0 - x^*\|_c^2)]$, 得出，而式（42）由我们的记号 $D_3 = D_0 D_2/(4D_1)$ 得出。前述不等式意味着 $\mathbb{E}\left[M_{k+1} \mid \mathcal{F}_k\right] \leq M_k$。因此，随机过程 $\{M_k\}$ 是适应于 $\{\mathcal{F}_k\}$ 的超鞅。

为此，我们已证明 $\{M_k\}$ 是一个超鞅，并给出了 M_k 期望的一个界（参见引理 3.3）。下一步，我们将使用 Ville 最大不等式来建立第一个最大浓缩界。结果陈述如下命题，其证明是 Ville 最大不等式的标准应用，详见附录 A.3。

命题 3.4。 对于任意 $\delta' \in (0,\ 1)$ 和 $K \geq 0$，以下不等式以至少$1 - \delta'$的概率成立：

$$\sup_{k \geq K} \{\lambda_k \mathbb{1}_{E_k(\delta)} \|x_k - x^*\|_c^2\} \leq 2u_{cM}^2 \log(1/\delta') + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2} \left(\frac{h}{K+h}\right)^{\alpha D_0/2 - 1} + \frac{16 u_{cM}^2 \alpha e D_2 \theta}{\alpha D_0 - 2} \frac{(1 + \|x^*\|_c)^2}{\|x_0 - x^*\|_c^2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k - 1 + h}{K - 1 + h}\right). \tag{43}$$

注。假设迭代序列 $\{x_k\}$ 被某个确定性常数一致有界。那么，取$T_k(\delta)$为一致范数平方界，可得对于某个常数 c， $\lambda_k = c/\alpha_k$ 且 $\mathbb{1}_{\{E_k(\delta)\}} = 1$ 几乎必然成立。在这种情况下，将式（43）两边乘以 α_k，即可得到所需的集中界，其收敛速率为 $\mathcal{O}(1/k)$，尾部 $\mathcal{O}(\log(1/\delta'))$呈指数级小。然而，在乘性噪声的情况下，迭代序列 $\{x_k\}$ 通常没有一致界。由于 $\lambda_k = \theta \alpha_k^{-1} T_k(\delta)^{-1}$且 $T_k(\delta)$可能是递增函数，式（43）无法提供所需的收敛速率。

接下来，我们完成证明命题 3.2 的最后一步。为简化记号，将公式（43）的右端项记为 $\epsilon(k,\ K,\ \delta')$。对于任意 $K \geq 0$，注意到

$$\begin{aligned}
&\mathbb{P}(\lambda_k \|x_k - x^*\|_c^2 \leq \epsilon(k, K, \delta'),\ \forall\, k \geq K) \\
&= \mathbb{P}\left(\bigcap_{k=K}^{\infty} \{\lambda_k \|x_k - x^*\|_c^2 \leq \epsilon(k, K, \delta')\}\right) \\
&\geq \mathbb{P}\left(\bigcap_{k=K}^{\infty} \{\lambda_k \mathbb{1}_{E_k(\delta)} \|x_k - x^*\|_c^2 \leq \epsilon(k, K, \delta')\} \cap E_k(\delta)\right).
\end{aligned}$$

继续之前，注意对于任意两个事件， A 和 B，我们有

$$\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c) = \mathbb{P}(A) + \mathbb{P}(B) - 1.$$

因此，我们有

$$\begin{aligned}
&\mathbb{P}(\|x_k - x^*\|_c^2 \leq \epsilon(k, K, \delta')/\lambda_k,\ \forall\, k \geq K) \\
&\geq \mathbb{P}\left(\bigcap_{k=K}^{\infty} \{\lambda_k \mathbb{1}_{E_k(\delta)} \|x_k - x^*\|_c^2 \leq \epsilon(k, K, \delta')\} \cap E_k(\delta)\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \mathbb{P}\left(\bigcap_{k=K}^{\infty}\{\lambda_k \mathbb{1}_{E_k(\delta)}\|x_k - x^*\|_c^2 \leq \epsilon(k,K,\delta')\}\right) + \mathbb{P}\left(\bigcap_{k=0}^{\infty} E_k(\delta)\right) - 1 \\
&= \mathbb{P}\left(\lambda_k \mathbb{1}_{E_k(\delta)}\|x_k - x^*\|_c^2 \leq \epsilon(k,K,\delta'),\ \forall\, k \geq K\right) + \lim_{k\to\infty}\mathbb{P}\left(E_k(\delta)\right) - 1 \\
&\geq (1-\delta') + (1-\delta) - 1 \\
&= 1 - \delta - \delta'
\end{aligned}
$$

利用 $\epsilon(k,\ K,\ \delta')$ 和 λ_k 的定义，可得以下结果。

命题 3.5 对于任意 $\delta' \in (0, 1-\delta)$ 和 $K \geq 0$，其概率至少为 $1-\delta-\delta'$，我们有

$$
\begin{aligned}
\|x_k - x^*\|_c^2 \leq \frac{\alpha_k T_k(\delta)}{\theta}\Bigg[&2u_{cM}^2 \log(1/\delta') + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} \\
&+ \frac{16u_{cM}^2 \alpha e D_2 \theta}{\alpha D_0 - 2}\frac{(1+\|x^*\|_c)^2}{\|x_0 - x^*\|_c^2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{h-1}\right)\Bigg],\ \forall\, k \geq K.
\end{aligned}
$$

在本命题中令 $K=0$，即可得到命题 3.2，这就是我们的自举蓝图。注意，命题 3.5 是命题 3.2 的更强版本，因为该结果对所有 $K \geq 0$都成立。

3.3 完成自举论证

我们从命题 3.1 推导出的最坏情况界出发，然后使用命题 3.2 迭代地改进该界；但在最后一步中，我们使用命题 3.5，以获得任意$K \geq 0$对应的最大界。每次应用命题 3.2 时，该界都会大约缩小 $1/k$ 倍，因此最终会得到一个递减的界。此后，命题 3.2 不再适用，因为初始化自举论证要求该界非递减。为给出详细推导，在 $D > 0$的情况下，我们假设 $2\alpha D$ 是正整数；这确实不失一般性，因为 $D = \gamma_c + \sigma - 1$，并且如果假设 2.3 对某个 $\sigma > 0$成立，那么对所有 $\sigma' > \sigma$也成立。根据命题 3.1，$\|x_k - x^*\|_c^2$ 可以按 $\mathcal{O}(k^{2\alpha D})$的速率多项式增长。因此，我们需要自举 $m := 2\alpha D + 1$ 次。当 $D \leq 0$时，由于 $\|x_k - x^*\|_c^2$ 几乎必然地受某个确定性常数所界，或者至多按对数增长，我们只需自举一次。

断言。 定理2.1 经由自举后，证明即告完成。

上述断言的证明只涉及标准的代数运算。更多细节请参见附录A.4 。

4 定理2.3 的证明

证明该定理需要以下两个引理。

引理 4.1。 给定 c_1、$c_2 > 0$，假设存在 C_1、$C_2 > 0$ 使得

$$
\mathbb{P}\left((k+h)^{c_2} x_k^{c_1} \geq \epsilon^{c_1}\right) \leq C_1 \exp\left(-C_2 \epsilon^{c_1}\right), \quad \forall\, \epsilon > 0, k \geq 0.
$$

那么，对于任意 $\lambda \in (0, C_2)$，我们有 上极限$_{k\to\infty}\,\mathbb{E}\left[\text{指数}\ \left(\lambda(k+h)^{c_2} x_k^{c_1}\right)\right] < \infty$。

引理的证明4.1（见附录B.1）可通过使用公式$\mathbb{E}[X] = \int_0^\infty \mathbb{P}\left(X > x\right) dx$对任意正的随机变量 X。

引理 4.2. 考虑定义在 $[0,\infty)$ 上的函数 $\ell(x) = e^x - (1+cx)$，其中 $c > 1$。存在$x_c > 0$，使得对所有 $x \in [0, x_c]$，$\ell(x) \leq 0$；并且对所有 $x \in [x_c, \infty)$，$\ell(x) \geq 0$。

引理 4.2 的证明可通过考察函数 $\ell(\cdot)$ 的单调性并利用其导数得出。更多详情请参见附录B.2。

接下来，我们证明定理2.3。

4.1 定理 2.3 第（1）部分的证明

由于 $\beta > 2/(1+2\alpha(a+N-1))$，存在 $\epsilon > 0$ 使得

$$\tilde{\beta} > \frac{2}{1+2\alpha(a+N-1)/(1+\epsilon)}.$$

应用引理 4.2，存在 $k = k(\epsilon) > 0$ 使得

$$\exp(\alpha_k(a+N-1)/(1+\epsilon)) \le 1+\alpha_k(a+N-1), \quad \forall\, k \ge k(\epsilon). \tag{44}$$

现在，对于任意 $\lambda > 0$，对任意 $k \ge k(\epsilon)$，有

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda\left[(k+h)^{1/2}x_k\right]^{\tilde{\beta}}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\lambda x_0^{\tilde{\beta}}(k+h)^{\tilde{\beta}/2}\prod_{i=0}^{k-1}(1+\alpha_i(Y_i-1))^{\tilde{\beta}}\right)\right] &(45)\\
&\ge \frac{1}{(N+1)^k}\exp\left(\lambda x_0^{\tilde{\beta}}(k+h)^{\tilde{\beta}/2}\prod_{i=0}^{k-1}(1+\alpha_i(a+N-1))^{\tilde{\beta}}\right) &(46)\\
&\ge \frac{1}{(N+1)^k}\exp\left(\lambda x_0^{\tilde{\beta}}(k+h)^{\tilde{\beta}/2}\prod_{i=k(\epsilon)}^{k-1}(1+\alpha_i(a+N-1))^{\tilde{\beta}}\right) \\
&\ge \frac{1}{(N+1)^k}\exp\left(\lambda x_0^{\tilde{\beta}}(k+h)^{\tilde{\beta}/2}\exp\left(\tilde{\beta}\sum_{i=k(\epsilon)}^{k-1}\frac{\alpha_i(a+N-1)}{(1+\epsilon)}\right)\right), &(47)
\end{aligned}$$

其中，公式（45）由更新方程（5）得到，公式（46）由 Y_k 的分布（参见例2.2）得到，公式（47）由公式（44）得到。注意到

$$\sum_{i=k(\epsilon)}^{k-1}\alpha_i = \sum_{i=k(\epsilon)}^{k-1}\frac{\alpha}{i+h} \ge \int_{k(\epsilon)}^{k}\frac{\alpha}{x+h}dx = \alpha\ln\left(\frac{k+h}{k(\epsilon)+h}\right).$$

因此，根据前面的不等式和公式（47）可得

$$\mathbb{E}\left[\exp\left(\lambda\left[(k+h)^{1/2}\,x_k\right]^{\tilde{\beta}}\right)\right] \ge \exp\left(\lambda x_0^{\tilde{\beta}}\left(\frac{k+h}{k(\epsilon)+h}\right)^{\frac{\tilde{\beta}}{2}\left(1+\frac{2\alpha(a+N-1)}{1+\epsilon}\right)} - k\ln(N+1)\right).$$

由于 ϵ 的选取满足 $\beta(1+2\alpha(a+N-1)/(1+\epsilon)) > 2$，由前面的不等式可得

$$\liminf_{k\to\infty}\mathbb{E}\left[\exp\left(\lambda\left[(k+h)^{1/2}\,x_k\right]^{\tilde{\beta}}\right)\right] = \infty, \quad \forall\,\lambda > 0.$$

因此，根据引理 4.1，不存在 K_1'，$K_2' > 0$ 使得

$$\mathbb{P}\left((k+h)^{1/2}\,x_k \ge \epsilon\right) \le K_1'\exp\left(-K_2'\epsilon^{\tilde{\beta}}\right), \quad \forall\,\epsilon > 0, k \ge 0.$$

4.2 定理2.3 （2）部分的证明

首先，根据引理 4.2，存在 $k_0 > 0$ 使得

$$\exp(\alpha_k(a+N-1)) \le 1+\alpha_k(a+N-1), \quad \forall\, k \ge k_0.$$

因此，采用与本定理第（1）部分证明相同的分析，对于任意 $\lambda > 0$ 和$k \ge k_0$，我们有

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda(k+h)^{\tilde\beta'} x_k^{\tilde\beta}\right)\right] \\
\ge\ & \frac{\exp\left(\lambda x_0^{\tilde\beta}(k+h)^{\tilde\beta'} \exp\left(\tilde\beta \sum_{i=k_0}^{k-1} \alpha_i(a+N-1)\right)\right)}{(N+1)^k} \\
\ge\ & \exp\left(\lambda x_0^{\tilde\beta}(k+h)^{\tilde\beta'} \exp\left(\frac{\alpha\tilde\beta(a+N-1)}{(1-z)}((k+h)^{1-z}-(k_0+h)^{1-z})\right) - k\ln(N+1)\right),
\end{aligned}$$

其中，最后一行由下式得出

$$\sum_{i=k_0}^{k-1} \alpha_i \ge \int_{k_0}^{k} \frac{\alpha}{(x+h)^z} dx = \frac{\alpha}{1-z}((k+h)^{1-z} - (k_0+h)^{1-z}).$$

因此，我们总是有

$$\liminf_{k\to\infty} \mathbb{E}\left[\exp\left(\lambda(k+h)^{\tilde\beta'} x_k^{\tilde\beta}\right)\right] = \infty, \quad \forall\, \lambda > 0.$$

因此，根据引理4.1，不存在 K_1'，$K_2' > 0$ 使得

$$\mathbb{P}\left((k+h)^{\tilde\beta'/\tilde\beta}\, x_k \ge \epsilon\right) \le \bar K_1' \exp\left(-\bar K_2' \epsilon^{\tilde\beta}\right), \quad \forall\, \epsilon > 0, k \ge 0.$$

5 定理2.4的证明

该证明的总体思路与定理 2.1的思路相似。具体而言，我们首先建立广义 Moreau 包络的矩母函数上界，然后利用 Ville 最大不等式推导最大浓缩界。不同之处在于，由于噪声具有加性，我们不需要使用自举论证。

5.1 广义 Moreau 包络的对数矩母函数界

回顾一下，对于所有 $x \in \mathbb{R}^d$，我们将广义 Moreau 包络定义为 $M(x) = \min_{u\in\mathbb{R}^d}\{\frac{1}{2}\|u\|_c^2 + \frac{1}{2\mu}\|x-u\|_s^2\}$。常数 ℓ_{cs}、u_{cs}、ℓ_{cM} 和 u_c、$_M$ 的定义，以及选择 μ 的要求，与第 3 节中的定义和要求相同。$M(\cdot)$ 的性质已在引理 3.2 中总结。具体而言，$M(\cdot)$ 关于 $\|\cdot\|_s$ 是一个 L/μ–光滑函数，并且可以写成 $M(x) = \frac{1}{2}\|x\|_M^2$，其中范数为$\|\cdot\|_M$。以下常数将在我们的推导中频繁使用。它们与 ℓ_{cs} 和 u_{cs} 之间的关系也已列出。

$$\begin{aligned}
&\ell_{M2}\|\cdot\|_2 \le \|\cdot\|_M \le u_{M2}\|\cdot\|_2, \quad && \ell_{M2} = \ell_{c2}(1+\mu u_{cs}^2)^{-1/2}, \quad && u_{M2} = u_{c2}(1+\mu\ell_{cs}^2)^{-1/2}, \\
&\ell_{sM}\|\cdot\|_M \le \|\cdot\|_s \le u_{sM}\|\cdot\|_M, && \ell_{sM} = (1+\mu\ell_{cs}^2)^{1/2} u_{cs}^{-1}, && u_{sM} = (1+\mu u_{cs}^2)^{1/2}\ell_{cs}^{-1}.
\end{aligned}$$

我们还定义 $u_{cM,\,*}$使得$\|x\|_M \le u_{cM,\,*}\|x\|_{c,\,*}$对所有 $x \in \mathbb{R}^d$均成立，其中$\|\cdot\|_c^*$是$\|\cdot\|_c$的对偶范数。令 $\bar D_0 = \mu\ell_{cs}^2/(8\bar\sigma^2 L)$、$\bar D_1 = 2(1-\tilde\gamma_c)$、$\bar D_2 = 8Lu_{cM}^2/(\mu\ell_{cs}^2)$、$\bar D_3 = 2\bar\sigma^2 u_{cM,\,*}^2$以及 $\bar D_4 = 2c_d\bar\sigma^2 L/(\mu\ell_{cs}^2)$。令 $\lambda_k = \theta/\alpha_k$ 满足 $\theta = D_1/(8D_3)$。步长按照以下条件选取。

条件 5.1。 我们使用 $\alpha_k = \alpha/(k+h)^z$，其中 $z \in (0,1]$,和 $\alpha > 0$， $h \geq 1$的选取满足 $\alpha_0 \leq \min(4D_0D_3/D_1,\ 1/D_1,\ D/(4D_2))$。此外，当 $z = 1$时，我们选取 $\alpha > 2/D_1$；当$z \in (0,\ 1)$时，我们选取 $h \geq (2z/(\bar{D}_1\alpha))^{1/(1-z)}$。

下面的命题给出了广义 Moreau 包络的对数矩母函数所满足的递归不等式。

命题 5.1。 对所有 $k \geq 0$，均有

$$Z_{k+1} \leq \frac{\alpha_k}{\alpha_{k+1}}(1 - \bar{D}_1\alpha_k/2)Z_k + \frac{\bar{D}_1\bar{D}_4}{4\bar{D}_3}\alpha_k, \tag{48}$$

其中 $Z_k = \log(\mathbb{E}[\exp(\lambda_k M(x_k - x^*))])$。

命题 *5.1* 的证明。 由于 $M(\cdot)$关于 L/μ 是 $\|\cdot\|_s$– 光滑的，根据更新方程（1），我们有

$$\begin{aligned} M(x_{k+1} - x^*) \leq\ & M(x_k - x^*) + \alpha_k \nabla M(x_k - x^*)^\top (\bar{F}(x_k) - x_k) \\ & + \alpha_k \nabla M(x_k - x^*)^\top (F(x_k, Y_k) - \bar{F}(x_k)) + \frac{L\alpha_k^2}{2\mu}\|F(x_k, Y_k) - x_k\|_s^2. \end{aligned} \tag{49}$$

接下来，我们对前一个不等式右端项中的所有项进行界定。首先，由 [19,引理 A.1]可得

$$\nabla M(x_k - x^*)^\top (\bar{F}(x_k) - x_k) \leq -2\,(1 - \tilde{\gamma}_c)\, M(x_k - x^*). \tag{50}$$

接下来，对于式（49）右端项中的二次项，我们有

$$\|F(x_k, Y_k) - x_k\|_s^2 \leq \left(\left\|F(x_k, Y_k) - \bar{F}(x_k)\right\|_s + \left\|\bar{F}(x_k) - \bar{F}(x^*)\right\|_s + \|x_k - x^*\|_s\right)^2.$$

为继续推导，注意到

$$\begin{aligned} &\|F(x_k, Y_k) - \bar{F}(x_k)\|_s \leq \frac{1}{\ell_{cs}}\|F(x_k, Y_k) - \bar{F}(x_k)\|_c, \\ &\|\bar{F}(x_k) - \bar{F}(x^*)\|_s \leq \frac{1}{\ell_{cs}}\|\bar{F}(x_k) - \bar{F}(x^*)\|_c \leq \frac{\gamma_c}{\ell_{cs}}\|x_k - x^*\|_c \leq \frac{\gamma_c u_{cM}}{\ell_{cs}}\|x_k - x^*\|_M, \\ &\|x_k - x^*\|_s \leq u_{sM}\|x_k - x^*\|_M \leq \frac{u_{cM}}{\ell_{cs}}\|x_k - x^*\|_M. \end{aligned}$$

因此，我们有

$$\begin{aligned} \|F(x_k, Y_k) - x_k\|_s^2 &\leq \left(\frac{1}{\ell_{cs}}\|F(x_k, Y_k) - \bar{F}(x_k)\|_c + \frac{2u_{cM}}{\ell_{cs}}\|x_k - x^*\|_M\right)^2 \\ &\leq \frac{2}{\ell_{cs}^2}\|F(x_k, Y_k) - \bar{F}(x_k)\|_c^2 + \frac{16u_{cM}^2}{\ell_{cs}^2}M(x_k - x^*). \end{aligned} \tag{51}$$

最后一行由$(a+b)^2 \leq 2(a^2+b^2)$对所有 a， $b \in \mathbb{R}$以及 $M(x) = \frac{1}{2}\|x\|_M^2$得出（参见引理3.2）。利用方程（50）和（51）中得到的上界，并代入方程（49），可得

$$\begin{aligned} M(x_{k+1} - x^*) \leq\ & \left(1 - 2\alpha_k(1 - \tilde{\gamma}_c) + \frac{8Lu_{cM}^2\alpha_k^2}{\mu\ell_{cs}^2}\right) M(x_k - x^*) \\ & + \frac{L\alpha_k^2}{\mu\ell_{cs}^2}\left\|F(x_k, Y_k) - \bar{F}(x_k)\right\|_c^2 + \alpha_k M(x_k - x^*)^\top (F(x_k, Y_k) - \bar{F}(x_k)). \end{aligned} \tag{52}$$

对于任意 $\lambda > 0$，方程（52）意味着

$$\begin{aligned}
\mathbb{E}[\exp(\lambda M(x_{k+1}-x^*)) \mid \mathcal{F}_k] \leq\ & \exp\left[\lambda\left(1-2\alpha_k(1-\tilde{\gamma}_c)+\frac{8Lu_{cM}^2\alpha_k^2}{\mu\ell_{cs}^2}\right)M(x_k-x^*)\right] \\
& \times \mathbb{E}\Bigg[\exp\left(\frac{\lambda L\alpha_k^2}{\mu\ell_{cs}^2}\left\|F(x_k,Y_k)-\bar{F}(x_k)\right\|_c^2\right) \\
& \times \exp\left(\alpha_k\lambda M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right) \Bigg| \mathcal{F}_k\Bigg] \\
\leq\ & \exp\left[\lambda\left(1-2\alpha_k(1-\tilde{\gamma}_c)+\frac{8Lu_{cM}^2\alpha_k^2}{\mu\ell_{cs}^2}\right)M(x_k-x^*)\right] \\
& \times \underbrace{\mathbb{E}\left[\exp\left(2\alpha_k\lambda M(x_k-x^*)^\top(F(x_k,Y_k)-\bar{F}(x_k))\right)\Big|\mathcal{F}_k\right]^{1/2}}_{N_1} \\
& \times \underbrace{\mathbb{E}\left[\exp\left(\frac{2\lambda L\alpha_k^2}{\mu\ell_{cs}^2}\left\|F(x_k,Y_k)-\bar{F}(x_k)\right\|_c^2\right)\Big|\mathcal{F}_k\right]^{1/2}}_{N_2}, \qquad (53)
\end{aligned}$$

最后一行由条件柯西–施瓦茨不等式得出。接下来，我们对式 (53) 中的两个条件期望给出上界。

对于项 N_1，根据假设 2.4 （或式 (7)），我们有

$$\begin{aligned}
N_1 &\leq \exp\left(\lambda^2\bar{\sigma}^2\alpha_k^2\left\|\nabla M(x_k-x^*)\right\|_{c,*}^2\right) \\
&= \exp\left(\lambda^2\bar{\sigma}^2\alpha_k^2\left\|\nabla\|x_k-x^*\|_M\right\|_{c,*}^2\|x_k-x^*\|_M^2\right) \qquad (54) \\
&= \exp\left(2\lambda^2\bar{\sigma}^2\alpha_k^2\left\|\nabla\|x_k-x^*\|_M\right\|_{c,*}^2 M(x_k-x^*)\right), \\
&\leq \exp\left(2\lambda^2\bar{\sigma}^2\alpha_k^2u_{cM,*}^2M(x_k-x^*)\right) \qquad (55)
\end{aligned}$$

其中，式 (54) 由 $M(x)=\frac{1}{2}\|x\|_M^2$（参见引理 3.2）得出，而式 (55) 由 [67,引理 2。6]得出。对于项 N_2，由假设 2.4 可知

$$N_2 \leq \left(1-\frac{4\lambda\bar{\sigma}^2L\alpha_k^2}{\mu\ell_{cs}^2}\right)^{-c_d/4} \leq \exp\left(\frac{2c_d\lambda\bar{\sigma}^2L\alpha_k^2}{\mu\ell_{cs}^2}\right)., \qquad (56)$$

其中，第二个不等式是通过选择 $\lambda \leq \bar{D}_0/\alpha_k^2$ 使得 $4\lambda\bar{\sigma}^2L\alpha_k^2/(\mu\ell_{cs}^2) \leq 1/2$，并利用数值不等式 $1/(1-x) \leq e^{2x}$（对任意 $x\in[0,1/2]$）得出的。

将我们对项 N_1（参见式 (55)）和 N_2（参见式 (56)）得到的上界代入式 (53)，可得

$$\begin{aligned}
\mathbb{E}[\exp(\lambda M(x_{k+1}-x^*)) \mid \mathcal{F}_k] \leq\ & \exp\left[\lambda\left(1-2\alpha_k(1-\tilde{\gamma}_c)+\frac{8Lu_{cM}^2\alpha_k^2}{\mu\ell_{cs}^2}\right)M(x_k-x^*)\right] \\
& \times \exp\left(2\lambda^2\bar{\sigma}^2\alpha_k^2u_{cM,*}^2M(x_k-x^*)\right)\exp\left(\frac{2c_d\lambda\bar{\sigma}^2L\alpha_k^2}{\mu\ell_{cs}^2}\right) \\
\leq\ & \exp\left[\lambda M(x_k-x^*)\left(1-\alpha_k\bar{D}_1+\bar{D}_2\alpha_k^2+\bar{D}_3\alpha_k^2\lambda\right)+\bar{D}_4\alpha_k^2\lambda\right], \qquad (57)
\end{aligned}$$

¯其中我们使用了 $\{D_i\}_{1\leq i\leq 4}$ 来简化最后一行中的记号。

由于前面的不等式对所有 $\lambda \in (0,\bar{D}_0/\alpha_k^2]$均成立，接下来我们将 λ 替换为随时间变化的 $\lambda_{k+1}=\theta/\alpha_{k+1}$。继续之前，我们先验证这一 λ_{k+1} 选择对所有 $k\geq 0$都满足 $\lambda_{k+1}\in(0,\bar{D}_0/\alpha_k^2]$。

由于对所有 $a \geq b > 0$ 和 $c \geq 0$都有 $a/b \geq (a+c)/(b+c)$，因此对于任意 $z \in (0,1]$ 和 $h \geq 1$，都有

$$\frac{\alpha_k}{\alpha_{k+1}} = \left(\frac{k+h+1}{k+h}\right)^z \leq \left(\frac{h+1}{h}\right)^z \leq 1 + \frac{1}{h} \leq 2.$$

由此可得

$$\lambda_{k+1} = \frac{\theta}{\alpha_{k+1}} = \frac{\bar{D}_0}{\alpha_k^2}\frac{\theta\alpha_k^2}{\bar{D}_0\alpha_{k+1}} \leq \frac{\bar{D}_0}{\alpha_k^2}\frac{2\theta}{\bar{D}_0}\alpha_k = \frac{\bar{D}_0}{\alpha_k^2}\frac{\bar{D}_1\alpha_k}{4\bar{D}_0\bar{D}_3}.$$

因此，当 $\alpha_0 \leq 4\bar{D}_0\bar{D}_3/\bar{D}_1$ （参见条件 5.1）时，有 $\lambda_{k+1} \leq \bar{D}_0/\alpha_k^2$。将 $\lambda = \lambda_{k+1}$ 应用于式（57），对所有 $k \geq 0$，均有

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda_{k+1}M(x_{k+1}-x^*)) \mid \mathcal{F}_k\right] \\
\leq\ &\exp\left(\lambda_k M(x_k - x^*)\frac{\lambda_{k+1}}{\lambda_k}\left(1 - \bar{D}_1\alpha_k + \bar{D}_2\alpha_k^2 + \bar{D}_3\lambda_{k+1}\alpha_k^2\right)\right)\exp(\bar{D}_4\lambda_{k+1}\alpha_k^2) \\
=\ &\exp\left(\lambda_k N_{1,k} M(x_k - x^*)\right)\exp(N_{2,k}),
\end{aligned} \tag{58}$$

为简化记号，我们记作 $N_{1,\,k} = (\lambda_{k+1}/\lambda_k)(1 - \bar{D}_1\alpha_k + \bar{D}_2\alpha_k^2 + \bar{D}_3\lambda_{k+1}\alpha_k^2)$ 以及 $N_{2,\,k} = \bar{D}_4\lambda_{k+1}\alpha_k^2$。利用 λ_{k+1}的显式表达式，可得

$$N_{2,k} = \bar{D}_4\lambda_{k+1}\alpha_k^2 = \bar{D}_4\theta\frac{\alpha_k^2}{\alpha_{k+1}} \leq 2\bar{D}_4\theta\alpha_k = \frac{\bar{D}_1\bar{D}_4}{4\bar{D}_3}\alpha_k. \tag{59}$$

接下来对$N_{1,\,k}$进行如下界定：

$$\begin{aligned}
N_{1,k} &= \frac{\lambda_{k+1}}{\lambda_k}(1 - \bar{D}_1\alpha_k + \bar{D}_2\alpha_k^2 + \bar{D}_3\lambda_{k+1}\alpha_k^2) \\
&= \frac{\alpha_k}{\alpha_{k+1}}(1 - \bar{D}_1\alpha_k + \bar{D}_2\alpha_k^2 + \bar{D}_3\theta\alpha_k^2/\alpha_{k+1}) \\
&\leq \frac{\alpha_k}{\alpha_{k+1}}\left(1 - \bar{D}_1\alpha_k + \bar{D}_2\alpha_k^2 + \bar{D}_1\alpha_k/4\right) &(60)\\
&\leq \frac{\alpha_k}{\alpha_{k+1}}\left(1 - \bar{D}_1\alpha_k + \bar{D}_1\alpha_k/4 + \bar{D}_1\alpha_k/4\right) &(61)\\
&= \frac{\alpha_k}{\alpha_{k+1}}\left(1 - \bar{D}_1\alpha_k/2\right). &(62)
\end{aligned}$$

其中，式（60）由 $\alpha_k/\alpha_{k+1} \leq 2$ 和 $\theta = D_1/(8D_3)$ 得出，式（61）由 $\alpha_k \leq \alpha_0 \leq D_1/(4D_2)$ 得出（参见条件 5.1）。将我们对 $N_{1,\,k}$ （参见式（62））和 $N_{2,\,k}$ （参见式（59））所得的上界代入式（58），对所有 $k \geq 0$，均有

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda_{k+1}M(x_{k+1}-x^*)) \mid \mathcal{F}_k\right] \\
\leq\ &\exp\left(\frac{\alpha_k}{\alpha_{k+1}}\left(1 - \frac{\bar{D}_1\alpha_k}{2}\right)\lambda_k M(x_k - x^*) + \frac{\bar{D}_1\bar{D}_4}{4\bar{D}_3}\alpha_k\right).
\end{aligned} \tag{63}$$

下一步，我们首先对前述不等式两边取全期望，然后取对数，接着应用詹森不等式。为此，我们需要验证 $(\alpha_k/\alpha_{k+1})\left(1 - \bar{D}_1\alpha_k/2\right)$ 属于区间 $(0,\ 1)$。

这一事实源于我们对 $\alpha_0 \leq 1/\bar{D}_1$ 的选择（参见条件 5.1）。为证明 $(\alpha_k/\alpha_{k+1})\left(1 - \bar{D}_1\alpha_k/2\right) < 1$，注意到

$$\frac{\alpha_k}{\alpha_{k+1}}\left(1 - \bar{D}_1\alpha_k/2\right) = \left(\frac{k+h+1}{k+h}\right)^z\left(1 - \frac{\bar{D}_1\alpha}{2(k+h)^z}\right)$$

$$
\begin{aligned}
&\leq \left(\frac{k+h+1}{k+h}\right)^{z} \exp\left(-\frac{\bar{D}_1\alpha}{2(k+h)^z}\right) \\
&= \left[\left(1+\frac{1}{k+h}\right)^{k+h}\right]^{z/(k+h)} \exp\left(-\frac{\bar{D}_1\alpha}{2(k+h)^z}\right) \\
&\leq \exp\left(\frac{z}{k+h}-\frac{\bar{D}_1\alpha}{2(k+h)^z}\right).
\end{aligned}
$$

¯当 $z=1$ 时，由于 $\alpha > 2/D_1$ （参见条件 5.1），我们有

$$
\frac{\alpha_k}{\alpha_{k+1}}\left(1-\bar{D}_1\alpha_k/2\right) \leq \exp\left(\frac{z}{k+h}-\frac{\bar{D}_1\alpha}{2(k+h)^z}\right) \leq 1.
$$

当 $z \in (0,1)$ 时，由于 $h \geq (4z/(\bar{D}_1\alpha))^{1/(1-z)}$ (参见条件 5.1)，我们还有

$$
\begin{aligned}
\frac{\alpha_k}{\alpha_{k+1}}\left(1-\bar{D}_1\alpha_k/2\right) &\leq \exp\left(\frac{z}{k+h}-\frac{\bar{D}_1\alpha}{2(k+h)^z}\right) \\
&= \exp\left(\frac{2z-(k+h)^{1-z}\bar{D}_1\alpha}{2(k+h)}\right) \\
&\leq \exp\left(\frac{2z-h^{1-z}\bar{D}_1\alpha}{2(k+h)}\right) \\
&\leq 1.
\end{aligned}
$$

现在我们已经验证了 $(\alpha_k/\alpha_{k+1})\left(1-\bar{D}_1\alpha_k/2\right) \in (0,1)$，先对式（63）两边取全期望，再取对数，然后应用詹森不等式，我们得到对所有 $k \geq 0$ 都有

$$
Z_{k+1} \leq \frac{\alpha_k}{\alpha_{k+1}}(1-\bar{D}_1\alpha_k/2)Z_k + \frac{\bar{D}_1\bar{D}_4}{4\bar{D}_3}\alpha_k.
$$

这就证明了命题 5.1。 □

反复使用命题 48）中的式子可得以下引理。引理 5.1的证明只涉及标准代数运算，故推迟至附录 C.1。

引理 5.1。 对所有 $k \geq 0$，均有

$$
Z_k \leq \begin{cases} Z_0\left(\frac{h}{k+h}\right)^{\bar{D}_1\alpha/2-1} + \frac{e\bar{D}_1\bar{D}_4\alpha}{\bar{D}_3(\bar{D}_1\alpha/2-1)}, & z=1, \\ Z_0\left(\frac{k+h}{h}\right)^{z}\exp\left(-\frac{\bar{D}_1\alpha}{2(1-z)}((k+h)^{1-z}-h^{1-z})\right)+\frac{\bar{D}_4}{\bar{D}_3}, & z\in(0,1). \end{cases} \tag{64}
$$

5.2 最大集中不等式

为了利用命题 5.1 和引理 5.1建立2种最大集中不等式。方法 *1* 基于构造超鞅并使用 Ville 最大不等式（我们在定理2.1的证明中也使用了该不等式），而方法 *2* 则基于将马尔可夫不等式与望远镜技巧结合使用。在 $z<1$ 的情况下，方法 *2* 给出的界更优。在 $z=1$ 的情况下，方法 *2* 对涉及初始条件的项给出了更优的衰减率，但在衰减锥上的初始半径较差。定理 2.4基于方法 *2*给出，其证明如下。关于方法 *1* 的细节见附录 C.1.1。

对于任意 $\epsilon > 0$ 和 $k \geq 0$，有

$$\begin{aligned}\mathbb{P}\left(\lambda_k M(x_k - x^*) > \epsilon\right) &= \mathbb{P}\left(\exp\left(\lambda_k M(x_k - x^*)\right) > e^{\epsilon}\right) \\ &\leq \mathbb{E}\left[\exp\left(\lambda_k M(x_k - x^*) - \epsilon\right)\right] \\ &= \exp\left(Z_k - \epsilon\right).\end{aligned}$$

令 $\delta = \exp(Z_k - \epsilon)$。于是，前面的不等式表示：以至少 $1 - \delta$ 的概率，有$M(x_k - x^*) \leq (Z_k + \log(1/\delta))/\lambda_k$。接下来，给定 $K > 0$，定义 $\delta_k = K\delta/(k(k+1))$。根据并集界，对于任意 $\delta > 0$，有

$$\begin{aligned}\mathbb{P}\left(M(x_k - x^*) \leq \frac{Z_k + \log(1/\delta_k)}{\lambda_k},\ \forall\, k \geq K\right) &\leq 1 - \sum_{k=K}^{\infty} \delta_k \\ &= 1 - \sum_{k=K}^{\infty} \frac{K\delta}{k(k+1)} \\ &= 1 - K\delta \sum_{k=K}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1}\right) \\ &= 1 - \delta.\end{aligned}$$

首先使用引理 3.2 将 $M(\cdot)$ 转换回 $\|\cdot\|_c^2$，然后使用引理 Z_k 5.1 中推导出的 Z_k 上界5.1，可得以下结果。

¯(1) 当 $z = 1$ 且 $\alpha > 2/D_1$ 时，对于任意 $K \geq 0$，以至少 $1 - \delta$ 的概率，以下不等式对所有 $k \geq K$成立：

$$\begin{aligned}\|x_k - x^*\|_c^2 \leq \frac{16\bar{D}_3 u_{cM}^2 \alpha}{\bar{D}_1(k+h)} \log(1/\delta) + \frac{u_{cM}^2}{\ell_{cM}^2}\|x_0 - x^*\|_c^2 \left(\frac{h}{k+h}\right)^{\bar{D}_1\alpha/2} \\ + \frac{16e u_{cM}^2 \bar{D}_4 \alpha^2}{(\bar{D}_1\alpha/2 - 1)} \frac{1}{k+h} + \frac{32 u_{cM}^2 \bar{D}_3 \alpha}{\bar{D}_1} \frac{\log((k+1)/K^{1/2})}{k+h}.\end{aligned}$$

(2) 当 $z \in (0,\ 1)$、 $\alpha > 0$ 和 $h \geq (\frac{4z}{\bar{D}_1\alpha})^{1/(1-z)}$ 时，对于任意 $\delta > 0$ 和 $K \geq 0$，以下不等式对所有 $k \geq K$ 成立：

$$\begin{aligned}\|x_k - x^*\|_c^2 \leq \frac{16\bar{D}_3 u_{cM}^2 \alpha}{\bar{D}_1(k+h)^z} \log(1/\delta) + \frac{16e u_{cM}^2 \bar{D}_4 \alpha}{\bar{D}_1} \frac{1}{(k+h)^z} \\ + \frac{u_{cM}^2}{\ell_{cM}^2}\|x_0 - x^*\|_c^2 \exp\left(-\frac{\bar{D}_1\alpha}{2(1-z)}((k+h)^{1-z} - h^{1-z})\right) \\ + \frac{32 u_{cM}^2 \bar{D}_3 \alpha}{\bar{D}_1} \frac{\log((k+1)/K^{1/2})}{(k+h)^z}.\end{aligned}$$

定理 2.4 可通过使用我们的记号 $\bar{c}_1 = 16\bar{D}_3 u_{cM}^2 \alpha/\bar{D}_1$， $\bar{c}_2 = u_{cM}^2/\ell_{cM}^2$， $\bar{c}_3 = 16e u_{cM}^2 \bar{D}_4 \alpha^2/(\bar{D}_1\alpha/2 - 1)$， $\bar{c}_4 = 32 u_{cM}^2 \bar{D}_3 \alpha/\bar{D}_1$以及 $\bar{c}_5 = 16e u_{cM}^2 \bar{D}_4 \alpha/\bar{D}_1$ 来简化上述表达式。

6 应用

在本节中，w 我们讨论理论结果在线性随机逼近和强化学习背景下的适用性。

6.1 线性随机逼近

考虑求解线性方程组 $\bar{A}x = \bar{b}$的问题，其中 $\bar{A} \in \mathbb{R}^{d\times d}$ 和 $\bar{b} \in \mathbb{R}^d$。这类问题出现在许多实际应用中 [6],，典型例子包括：（1）求解最小二乘问题；（2）在强化学习中使用时序差分（TD）学习方法解决策略评估问题（将在下一小节讨论）；（3）控制理论中的线性动力系统等。

假设我们无法获得问题中参数 A 和 b 的完整信息，但可以得到它们的带噪估计。在这种情况下，通常使用下文介绍的随机逼近来求解问题：

$$x_{k+1} = x_k + \alpha_k(A(Y_k)x_k - b(Y_k)), \tag{65}$$

其中 $\{Y_k\}$（取值于 $\mathcal{Y}$）是一列独立同分布随机变量，且 $A: \mathcal{Y} \mapsto \mathbb{R}^{d\times d}$ 和 $b: \mathcal{Y} \mapsto \mathbb{R}^d$是确定性函数。为研究由方程（65）生成的 $\{x_k\}$，我们采用以下标准假设。

假设 6.1。 满足 $\mathbb{E}[A(Y_k)] = A$ 且 $\mathbb{E}[b(Y_k)] = b$ 对所有 $k \geq 0$均成立。

只需假设 $\{Y_k\}$ 具有第 2.1 节中所述的条件无偏扰动。这里为便于说明，采用独立同分布假设。

假设 6.2。 有 $\sup_{y\in\mathcal{Y}} \|A(y)\|_2 < \infty$ 以及 $\sup_{y\in\mathcal{Y}} \|b(y)\|_2 < \infty$。

记 $A_{\max} = \sup_{y\in\mathcal{Y}} \|A(y)\|_2$ 和 $b_{\max} = \sup_{y\in\mathcal{Y}} \|b(y)\|_2$。假设 6.2 在研究线性随机逼近的渐近收敛 [8] 和有限样本收敛 [69]时被广泛采用。注意，当 $\mathcal{Y}$ 为有限集时，该假设会自动满足。

假设 6.3。 矩阵 A 是赫尔维茨矩阵，即 A 的所有特征值的实部都严格为负。

假设 6.3 通常用于确保式（65）中所述线性随机逼近算法的稳定性 [69]。事实上，考虑与该随机逼近算法相关的常微分方程（ODE） [10]:

$$\dot{x}(t) = \bar{A}x(t) - \bar{b}. \tag{66}$$

当 $\bar{A}$为赫尔维茨矩阵时，李雅普诺夫方程 $\bar{A}^\top P + P\bar{A} + I_d = 0$存在唯一的正定解 [41],，记为 $\bar{P}$。由李雅普诺夫定理 [41] 可知，常微分方程（66）的唯一平衡点 $x^* = \bar{A}^{-1}\bar{b}$具有指数稳定性， [34], 进而通过常微分方程方法保证随机逼近算法渐近收敛 [10]。

接下来，我们证明，式（65）中的线性随机逼近可以等价地建模为具有乘性噪声、形式如式（1）的压缩随机逼近。因此，定理 2.1 使我们能够为线性随机逼近建立最大集中界（具有威布尔尾）。令 $\beta = \lambda_{\max}^{-1}(\bar{A}^\top\bar{P}\bar{A})/2$。本文中， $\lambda_{\max}(\cdot)$（分别是， $\lambda_{\min}(\cdot)$）返回对称矩阵的最大（分别是最小）特征值。令$F_\beta: \mathcal{Y}\times\mathbb{R}^d \mapsto \mathbb{R}^d$ 为定义如下的算子

$$F_\beta(x, y) = \beta A(y)x - \beta b(y) + x, \quad \forall\, x \in \mathbb{R}^d, y \in \mathcal{Y}.$$

于是，式（65）可等价地写为

$$x_{k+1} = x_k + \frac{\alpha_k}{\beta}(F_\beta(x_k, Y_k) - x_k),$$

其形式与算法（1）相同，因为我们可以将常数 β 吸收到步长中。接下来，我们证明在线性随机逼近的情形下，假设 2.1 – 2.3 均成立。令 $\|\cdot\|_{\bar{P}}$ 为如下定义的范数：$\|x\|_{\bar{P}} = (x^\top\bar{P}x)^{1/2}$对所有 $x \in \mathbb{R}^d$，其中我们回忆， $\bar{P}$是李雅普诺夫方程 $\bar{A}^\top P + P\bar{A} + I_d = 0$的唯一正定解。以下引理的证明见附录 D。

引理 6.1。 假设 *6.2* 和 *6.3* 成立，且 $\beta = \lambda_{\max}^{-1}\left(\bar{A}^\top \bar{P}\bar{A}\right)/2$。那么，我们有以下结果。

(1) 存在 $\bar{\gamma} \in (0,1)$ 使得算子 $F_\beta(\cdot) = \mathbb{E}[F_\beta(\cdot, Y_0)]$ 关于 $\|\cdot\|_{\bar{P}}$ 是一个 $\bar{\gamma}$–压缩映射。

(2) 对所有 $k \geq 0$ 均成立： $\mathbb{E}[F_\beta(x_k, Y_k) \mid \mathcal{F}_k] = F_\beta(x_k)$，其中 $\mathcal{F}_k$ 是由 $\{x_0, Y_0, Y_1, \cdots, Y_{k-1}\}$ 生成的 σ-代数。

(3) 存在 $\hat{\sigma} > 0$，使得对所有 $k \geq 0$， $\|F_\beta(x_k, Y_k) - F_\beta(x_k)\|_{\bar{P}} \leq \hat{\sigma}(\|x\|_{\bar{P}} + 1)$。

引理 6.1使我们能够应用定理 2.1 ，从而建立式（65）中所述线性随机逼近算法的最大浓缩界。结果如下。

定理 *6.1*。假设假设 *6.2* 和 *6.3* 成立，且 $\alpha_k = \alpha\beta/(k+h)$，其中 α 和 h 经过适当选择。那么，此处成立的界与定理 *2.1* 中的相同。因此，存在一个整数 $m_\ell > 0$，使得对于任意 $\epsilon > 0$ 和 $\delta \in (0, 1)$，若要以至少 $1-\delta$ 的概率达到 $\|x_k - x^*\|_{\bar{P}} \leq \epsilon$，迭代复杂度为 $\tilde{\mathcal{O}}((1 + \log^{m_\ell}(1/\delta))\epsilon^{-2})$。

由于线性随机逼近有广泛的应用，已有许多结果分析其集中行为 [28–30]。然而，只有当矩阵 $A(Y_k)$ 几乎必然是赫尔维茨矩阵时，才建立了集中界（尾部衰减快于多项式）；否则，得到的仅是多项式尾界 [29]。定理 6.1 指出，由线性随机逼近生成的迭代序列（其中随机矩阵 $\{A(Y_k)\}$ 在期望意义下是赫尔维茨矩阵，但并非几乎必然是赫尔维茨矩阵）具有 $\mathcal{O}(1/k)$ 威布尔尾部的收敛速率，相比文献中的结果有显著改进。

6.2 强化学习

本节将讨论我们的结果在强化学习背景下的适用性。近年来，强化学习已被用于解决许多涉及序贯决策的实际问题。强化学习问题通常建模为马尔可夫决策过程（MDP） [71]。然而，环境模型（包括转移动态和奖励函数）对智能体而言是未知的。因此，强化学习智能体必须通过主动与环境交互来收集信息，并学习如何做出决策。

数学上，考虑一个无限时域折扣马尔可夫决策过程，它由有限状态空间 $\mathcal{S}$、有限动作空间 $\mathcal{A}$、一组转移概率矩阵 $\{P_a \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|} \mid a \in \mathcal{A}\}$、奖励函数$\mathcal{R}$：$\mathcal{S}\times\mathcal{A} \mapsto [0,1]$[4]以及折扣因子 $\gamma \in (0,1)$定义。注意，在强化学习中，转移概率和奖励函数对智能体而言是未知的。给定一个平稳策略 $\pi : \mathcal{S} \mapsto \Delta^{|\mathcal{A}|}$，其中 $\Delta^{|\mathcal{A}|}$表示 $|\mathcal{A}|$ 维概率单纯形，其价值函数 V^π：$\mathcal{S} \mapsto \mathbb{R}$ 和 Q函数Q^π：$\mathcal{S}\times\mathcal{A} \mapsto \mathbb{R}$ 定义为 $V^\pi(s) = \mathbb{E}_\pi\left[\sum_{k=0}^\infty \gamma^k \mathcal{R}(S_k, A_k) \mid S_0 = s\right]$ 对所有s以及 $Q^\pi(s,a) = \mathbb{E}_\pi[\cdot]$ (s,a)，均成立；对所有(s,a)均成立。这里我们用记号表示动作是根据策略 π选取的。由于我们研究的是有限马尔可夫决策过程，价值函数也可以视为一个处于 $|\mathcal{S}|$维欧几里得空间中的向量，其中 $|\mathcal{S}|$ 表示状态空间 $\mathcal{S}$的基数， Q函数也类似。

在强化学习中，预测问题（也称为策略评估问题）是指针对给定策略 π 估计价值函数 V^π（或 Q 函数 Q^π）的问题；而控制问题是指寻找一个最优策略 π^*，使其价值函数 V^*（或等价地，其 Q 函数 Q^π）在所有情况下均达到最大，即对于所有 $s \in \mathcal{S}$ 和策略 π， $V^*(s) \geq V^\pi(s)$（或等价地，对于所有 $(s,\ a)$ 和策略 π， $Q^*(s,\ a) \geq Q^\pi(s,\ a)$）。从现在起，我们将从 Q 函数的角度研究预测问题和控制问题。按照类似的方法，也可以得到关于价值函数的相同结果。

[4]由于我们处理的是有限马尔可夫决策过程，假设奖励函数落在区间 $[0,1]$ 内并不失一般性

6.2.1 用于预测的时序差分学习

我们首先考虑预测问题，通常使用时序差分学习 [70] 及其变体来求解。为介绍该算法，我们先陈述马尔可夫决策过程理论中的一个重要结果——贝尔曼方程，它指出 $Q^\pi \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ 是以下不动点方程的唯一解 $Q = \mathcal{H}^\pi(Q)$，其中 $\mathcal{H}^\pi : \mathbb{R}^{|\mathcal{S}||\mathcal{A}|} \mapsto \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ 是定义如下的贝尔曼算子

$$[\mathcal{H}^\pi(Q)](s,a) = \mathbb{E}_\pi \left[\mathcal{R}(S_0, A_0) + \gamma Q(S_1, A_1) \mid S_0 = s, A_0 = a\right] \tag{67}$$

对所有 $(s,\ a) \in \mathcal{S} \times \mathcal{A}$ 和 $Q \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$均成立。利用 Q-函数的定义和马尔可夫性质，可以推导出 $Q^\pi = \mathcal{H}^\pi(Q^\pi)$；此外，由于 $\mathcal{H}^\pi(\cdot)$ 是压缩映射 [8]，因此该解的唯一性得以保证。而且，可以很容易地验证，Q^π 是多步贝尔曼方程的唯一解

$$Q = \underbrace{\mathcal{H}^\pi \cdot \mathcal{H}^\pi \cdots \mathcal{H}^\pi}_{n \text{ times}}(Q) := \mathcal{H}_n^\pi(Q) \tag{68}$$

对于任意 $n \geq 1$ 均成立；只要 $\mathcal{H}^\pi(\cdot)$ 是压缩映射，这一点就可由 $\mathcal{H}_n^\pi(\cdot)$ 是压缩映射推出。

鉴于贝尔曼方程，解决预测问题的一种自然方法是执行不动点迭代来求解 $Q = \mathcal{H}^\pi(Q)$；根据巴拿赫不动点定理，其几何收敛性得到了保证 [1]。然而，执行这种不动点迭代需要完全了解模型参数，以计算公式（67）中的条件期望。为克服这一挑战，时序差分学习被提出，作为一种数据驱动的随机逼近算法，用于求解（多步）贝尔曼方程。

令 π_b 为智能体用于与环境交互并收集样本的策略，通常称为行为策略。注意，π_b 不一定与目标策略 π 一致；我们旨在估计该目标策略的 Q-函数。假设对于所有 $(s,\ a)$，$\pi_b(a \mid s) > 0$，且由此诱导的状态马尔可夫链具有唯一的平稳分布，记为 $\kappa_b \in \Delta(\mathcal{S})$，并满足对于所有 $s \in \mathcal{S}$，$\kappa_b(s) > 0$。我们对样本轨迹作出以下假设。

假设6.4。样本轨迹 $\{(S_k^0,\ A_k^0,\ S_k^1,\ A_k^1,\ \cdots,\ S_k^n,\ A_k^n)\}_{k\geq 0}$（其中样本是一个 n 元组）是一个独立同分布序列，满足对于所有 $i \in \{0, 1, \cdots, n-1\}$ 和 $k \geq 0$，$S_k^0 \sim \kappa_b(\cdot)$、$A_k^i \sim \pi_b(\cdot \mid S_k^i)$，以及 $S_k^{i+1} \sim P_{A_k^i}(S_k^i,\ \cdot)$。

为便于解析处理，现有文献中广泛采用独立同分布采样 [9, 24,26, 83]。在实践中，假设由 π_b 诱导的马尔可夫链 $\{S_k\}_{k\geq 0}$ 不可约且非周期[15, 22, 42, 62, 81, 82, 84]（这意味着具有一致遍历性）。通过先让马尔可夫链 $\{S_k\}$ 演化一小段时间（由于几何混合，$\{S_k\}$ 的分布会接近 κ_b），然后隔一段时间收集一个样本（从而使样本近似独立），即可近似实现独立同分布采样。与更自然的马尔可夫采样相比（在马尔可夫采样中，智能体使用行为策略 π_b 与环境交互，并使用由此生成的一条状态-动作对轨迹中的每个样本），独立同分布采样假设确实存在局限。未来一项直接的研究方向是，探究对于有偏更新的随机逼近，能否得到尾部衰减快于多项式的最大集中界，从而使我们能够研究采用马尔可夫采样的强化学习。

接下来，我们将介绍一种用于解决策略评估问题的通用基于学习的算法，它涵盖了许多现有时序差分学习变体作为特例。在此之前，我们需要先介绍带函数逼近的强化学习，它用于克服维数灾难。

函数逼近的核心思想，是从预先指定的函数类中逼近所需的高维目标（在策略评估中为Q^π）。在这项工作中，为了便于分析，我们考虑线性函数逼近。设 $\{\phi_i\}_{1\le i\le d}$ 为一组线性无关的基向量，其中对于每个 $i \in \{1,2,\cdots,d\}$，ϕ_i 都是一个 $|\mathcal{S}||\mathcal{A}|$维向量。令$\phi(s,\ a) = (\phi_1(s,\ a)\ \phi_2(s,\ a)\ \cdots,\ \phi_d(s,\ a)) \in \mathbb{R}^d$，它通常称为与状态-动作对（$s,\ a$）相关联的特征。我们不失一般性地假设基向量已归一化，因此对于所有（$s,\ a$），$\|\phi(s,\ a)\|_2 \le 1$。引入基向量后，在线性函数逼近中，目标是找到基向量的一个线性组合 $\sum_{i=1}^d \phi_i w_i$（其中 $w \in \mathbb{R}^d$），以尽可能准确地逼近 Q^π。顺带一提，如果 $\{\phi_i\}_{1\le i\le d}$ 选为标准基，那么该设定就退化为表格设定。

现在，我们可以给出通用算法了。给定独立同分布的样本轨迹$\{(S_k^0,\ A_k^0,\ S_k^1,\ A_k^1,\ \cdots,\ S_k^n,\ A_k^n)\}_{k\ge 0}$后，智能体根据以下公式迭代更新权重 w_k：

$$w_{k+1} = w_k + \alpha_k \phi(S_k, A_k) \sum_{i=k}^{k+n-1} \gamma^{i-k} \prod_{j=k+1}^{i} c(S_k^j, A_k^j) \times \left(\mathcal{R}(S_k^i, A_k^i) + \gamma \rho(S_k^j, A_k^j)\phi(S_k^{i+1}, A_k^{i+1})^\top w_k - \phi(S_k^i, A_k^i)^\top w_k\right), \tag{69}$$

其中，$c,\ \rho:\ \mathcal{S}\times\mathcal{A}\mapsto\mathbb{R}$ 称为广义重要性采样因子。为了进一步说明该算法并提供直观理解，我们来看以下几个特殊情况。

情况 1。假设我们使用标准基向量，且 $\pi_b = \pi$（即行为策略与目标策略相同）。此时，若对所有 $c(s,\ a) = \rho(s,\ a) = 1$ 选择 $(s,\ a)$，式（69）便化为标准同策略 n步时序差分学习的更新方程。该随机逼近算法用于求解n步 Bellman 方程（68）[71]。

情况 2：假设我们使用标准基向量和 $\pi_b \ne \pi$，即处于异策略设定。此时，若对所有（$s,\ a$）选择 $c(s,a) = \rho(s,a) = \pi(a\mid s)/\pi_b(a\mid s)$，式（69）便化为标准异策略 n步时序差分学习的更新方程。尽管选择 $c(s,a) = \rho(s,a) = \pi(a\mid s)/\pi_b(a\mid s)$可得到 Q^π的无偏估计量，但众所周知，该算法的方差很大（这是由重要性抽样导致的）[32]。为解决这一问题，人们提出了各种广义重要性采样因子，由此产生了 $Q^\pi(\lambda)$ [35], Tree-Backup(λ) [60], Retrace(λ) [56], 和Q-trace [19], 等算法，它们的形式都符合式（69）。

情况 3： 假设我们确实处于函数逼近设定中，其中 $\{\phi_i\}_{1\le i\le d}$构成一个 d维线性子空间 $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$。那么，公式（69）涵盖了流行的 n步时序差分学习与线性函数逼近在同策略和异策略设定下的特例 [14, 16, 75]。

为了建立公式（69）中所述通用时序差分学习算法的集中界，我们证明公式（69）可以等价地表述为线性随机逼近。对于任意 $y = (s^0、\ a^0、\ s^1、\ a^1、\ \cdots、\ s^n、\ a^n) \in \mathcal{Y} := \mathcal{S}^n \times \mathcal{A}^n$，令 $A:\ \mathcal{Y}\mapsto\mathbb{R}^{d\times d}$ 定义为

$$A(y) = \phi(s^0,a^0)\sum_{i=0}^{n-1}\gamma^i\prod_{j=1}^{i} c(s^j,a^j)\left(\gamma\rho(s^j,a^j)\phi(s^{i+1},a^{i+1}) - \phi(s^i,a^i)\right)^\top,$$

并令 $b:\ \mathcal{Y}\mapsto\mathbb{R}^d$ 定义为

$$b(y) = -\phi(s^0,a^0)\sum_{i=0}^{n-1}\gamma^i\prod_{j=1}^{i} c(s^j,a^j)\mathcal{R}(s^i,a^i).$$

于是，更新方程（69）可等价地写为

$$w_{k+1} = w_k + \alpha_k(A(Y_k)w_k - b(Y_k)),$$

其中 $\{Y_k = (S_k^0,\ A_k^0,\ \cdots,\ S_k^n,\ A_k^n)\}_{k\geq 0}$ 是一个独立同分布随机变量序列。接下来，我们证明应用定理 6.1 所需的所有假设在时序差分学习的情形下均成立。由于集合 $\mathcal{Y}$ 是有限的，且采用独立同分布采样，因此假设6.1 和 6.2 自动成立，我们只需验证假设 6.3。以下引理的证明见附录 D.2。

引理 6.2。 以下结果成立。

(1) 在同策略设定下（即，$\pi_b = \pi$），矩阵 A 始终是赫尔维茨矩阵。

(2) 在异策略设定下（即，$\pi_b \neq \pi$），采用适当选择的广义重要性抽样比率 $c(\cdot,\cdot)$ $\rho(\cdot,\cdot)$ 以及参数 n[5]时，矩阵 A 是赫尔维茨矩阵。

现在，将定理 6.1 应用于式（69）中给出的通用时序差分学习算法，我们得到一个渐近最大集中界，其具有 $\mathcal{O}(1/k)$ 收敛速率和威布尔尾部。

鉴于强化学习的广泛应用，文献中有许多结果研究了在独立同分布采样或马尔可夫采样 [9, 12, 13, 16, 18, 24, 69]下，同策略和异策略时序差分学习的均方界与集中界。与我们最接近的工作是 [12, 13, 24]。具体而言，在 [24] 中，他们针对采用线性函数逼近的 TD(0) 得到了最大集中界，其噪声为线性乘性噪声。然而，他们的界仅对足够大的迭代序列有效（阈值取决于概率 δ），且其收敛速率可能比 $\mathcal{O}(1/k)$更差。 [12, 13] 中的作者为采用线性函数逼近和马尔可夫采样的同策略时序差分学习给出了最大集中界。然而，他们要求 $\|F(x_k, Y_k)\|_2 \leq c_1\|x\|_2 + c_2$ 满足 $c_1 \in (0, 1)$，并且只有当K 足够大时，其界才开始成立（阈值是随机的）。据我们所知，在异策略设定下，文献中尚无关于高概率界的结果。

6.2.2 Q-学习与控制

转向控制问题，目标是找到一个最优策略 π^*，使其对应的 Q 函数Q^* 满足：对于所有 (s, a) 和 π，$Q^*(s, a) \geq Q^\pi(s, a)$。求解控制问题最受欢迎的算法之一是 Q-学习，这也是我们在此关注的内容。

令 $\mathcal{H}:\ \mathbb{R}^{|\mathcal{S}||\mathcal{A}|} \mapsto \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ 为定义如下的贝尔曼最优算子

$$[\mathcal{H}(Q)](s, a) = \mathcal{R}(s, a) + \gamma \mathbb{E}\left[\max_{a' \in \mathcal{A}} Q(S_{k+1}, a') \,\middle|\, S_k = s, A_k = a\right]$$

对于所有 $Q \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ 和 (s, a) [8]。众所周知，Q^* 是不动点方程$Q^* = \mathcal{H}(Q^*)$ 的唯一解。此外，一旦得到 Q^*，就可以通过基于 Q^* 贪心地选择动作来获得最优策略；也就是说，任何满足 $\{a \mid \pi(a \mid s) > 0\} \subseteq \arg\max_{a' \in \mathcal{A}} Q^*(s, a')$ 的策略 π 都是最优策略。更多详情请参见 [8, 61] 。

著名的 Q学习算法可视为一种数据驱动的随机逼近算法，旨在求解$Q = \mathcal{H}(Q)$。令 π_b 为用于收集样本的行为策略。假设由 π_b 诱导的马尔可夫链 $\{S_k\}$具有唯一的平稳分布 $\kappa_b \in \Delta(\mathcal{S})$，且对所有 $s \in \mathcal{S}$，均满足 $\kappa_b(s) > 0$ 。与时序差分学习类似，我们对样本轨迹作如下独立同分布假设。

假设 6.5。 样本轨迹 $\{(S_k,\ A_k,\ S_k')\}_{k\geq 0}$ （其中每个样本都是一个三元组）是一个独立同分布序列，满足 $S_k \sim \kappa_b(\cdot)$、$A_k \sim \pi_b(\cdot|S_k)$，且对所有 $k \geq 0$，均有 $S_k' \sim P_{A_k}(S_k,\ \cdot)$。

如上一节所示，文献中通常采用独立同分布采样假设 [47, 57, 79]。此外，当由 π_b 诱导的马尔可夫链 $\{S_k\}$ 不可约且非周期时，

[5]明确要求见附录D.2。

因此具有一致遍历性，并且存在一种近似实现独立同分布采样的方法。尽管如此，从单条马尔可夫样本轨迹中学习是一种更实用、更自然的设定 [31, 48, 49]，这要求研究有偏更新下的随机逼近，也是本工作的一个未来研究方向。

给定样本轨迹后， Q-学习算法会迭代地更新对 Q^*的估计值 Q_k ，具体公式如下：

$$Q_{k+1}(S_k, A_k) = Q_k(S_k, A_k) + \alpha_k(\mathcal{R}(S_k, A_k) + \gamma \max_{a' \in \mathcal{A}} Q_k(S'_k, a') - Q_k(S_k, A_k))$$

对所有 $k \geq 0$，其中 Q_0 的初始化值可以任意选取，但须满足 $\|Q_0\|_\infty \leq 1/(1-\gamma)$。

接下来，我们将 Q-学习重新表述为式（1）中给出的随机逼近算法的形式。令 $F: \mathbb{R}^{|\mathcal{S}||\mathcal{A}|} \times \mathcal{S} \times \mathcal{A} \times \mathcal{S} \mapsto \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ 为定义如下的一个算子

$$[F(Q, s_0, a_0, s_1)](s, a) = \mathbb{1}_{\{(s_0,a_0)=(s,a)\}}\big(\mathcal{R}(s_0, a_0) + \gamma \max_{a' \in \mathcal{A}} Q(s_1, a') - Q(s_0, a_0)\big) + Q(s, a)$$

对所有（s， a以及（Q， s_0， a_0， s_1）。那么， Q-学习的更新方程可以等价地写成 as

$$Q_{k+1} = Q_k + \alpha_k(F(Q_k, S_k, A_k, S'_k) - Q_k)$$

对所有 $k \geq 0$，其形式与式（1）中的随机逼近算法相同，其中 x_k 为 Q_k ，而 Y_k 为三元组 (S_k, A_k, S'_k)。

接下来，我们将利用关于次高斯加性噪声下随机逼近的结果，建立 Q-学习的最大集中界。我们首先在以下引理中验证，假设2.1、2.2和2.4在 Q-学习的背景下成立。令 D_b 为一个 $|\mathcal{S}||\mathcal{A}| \times |\mathcal{S}||\mathcal{A}|$ 的对角矩阵，其对角分量为 $\{\kappa_b(s)\pi_b(a|s)\}_{(s,a)\in\mathcal{S}\times\mathcal{A}}$。记 D_b 的最小对角元素为 $D_{b,\min}$。令 $\mathcal{F}_k$ 为由 $\{S_i, A_i, S'_i\}_{0\leq i\leq k-1}$生成的$\sigma$-代数。注意， Q_k 相对于 $\mathcal{F}_k$可测。以下引理的证明见附录D.3。

引理 6.3。 算子 $F(\cdot) := \mathbb{E}[F(\cdot, S_k, A_k, S'_k)]$ 可明确表示为 $F(Q) = D_b\mathcal{H}(Q) + (I_{|\mathcal{S}||\mathcal{A}|} - D_b)Q$ 对所有 $Q \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$均成立。此外，我们还有以下结果。

(1) $F(\cdot)$ 是关于 $\|\cdot\|_\infty$ 的 $\hat{\gamma}_c$-压缩映射，其中 $\hat{\gamma}_c = 1 - D_{b,\min}(1-\gamma)$。

(2) $\bar{F}(Q_k) = \mathbb{E}[F(Q_k, S_k, A_k, S'_k) \mid \mathcal{F}_k]$ *for all* $k \geq 0$.

(3) 假设 2.4 在 $\bar{\sigma} = 4/(1-\gamma)$ 和 $c_d = 1$下成立。

引理 6.3 使我们能够应用定理2.4 ，从而得到 Q-学习的最大浓缩界。结果如下述定理所示，其证明见附录 D.4。

定理 6.2。 假设 6.5 成立，并且 $\alpha_k = \alpha/(k+h)$，其中 $\alpha > 2/(1-\hat{\gamma}_c)$ 以及 h均作了适当选择。那么，对于任意 $K \geq 0$ 和 $\delta \in (0,1)$，以至少 $1-\delta$的概率，对于所有$k \geq K$ 均有

$$\|Q_k - Q^*\|_\infty^2 \leq c_q \left[\frac{\log(1/\delta)}{k+h} + \left(\frac{h}{k+h}\right)^{(1-\hat{\gamma}_c)\alpha/2} + \frac{1+\log((k+1)/K^{1/2})}{k+h}\right],$$

其中 $c_q = \frac{\log(|\mathcal{S}||\mathcal{A}|)\, D_{b,\min}^3}{(1-\gamma)^5}$。

根据定理 6.2，达到 $\|Q_k - Q^*\|_\infty \leq \epsilon$所需的样本复杂度为 $\tilde{\mathcal{O}}(D_{b,\text{最小值}}^{-3}(1-\gamma)^{-5}\epsilon^{-2})$。由于采用异步更新，收敛速率自然取决于马尔可夫链 $\{(S_k, A_k)\}$平稳分布的最小分量$D_{b,\min}$。该量至多与状态-动作空间的大小成反比。样本复杂度取决于 $1/(1-\gamma)$ 这一点是直观的，因为随着 γ 增大，智能体在做决策时应当考虑更远的未来，这会使问题更具挑战性。现有研究中也观察到了类似的依赖关系，例如Q-学习 [3, 4, 20, 31, 47, 49, 78, 79]以及策略梯度等其他算法 [51, 53]。

7 结论

本文为具有加性噪声与乘性噪声的一般压缩随机逼近建立了最大集中界。具体而言，我们证明，样本路径以高概率保持在一个锥内（其半径随时间衰减）。此外，我们展示了如何将这些一般界应用于线性随机逼近和各种强化学习算法。在方法上，为了克服迭代序列无界这一难题，我们提出了一种新颖的自举论证：从一个可能较松的界出发，逐步改进该界，从而得到更紧的界。关键步骤包括对收敛误差的广义 Moreau 包络某个修正版的对数矩母函数进行界定，并仔细构造超鞅以获得最大界。

未来工作。 我们的主要结果要求噪声序列是条件无偏的。然而，在实际应用中，样本轨迹往往可能包含有偏扰动。例如，在强化学习中，假设智能体与环境交互，以收集一条样本轨迹。那么，该样本轨迹会形成一个马尔可夫链，其中包含有偏扰动。将我们的结果扩展到算子可能具有有偏扰动的情形，是本工作的一个直接未来方向。另一个方向是将我们的结果扩展到更具挑战性但同样具有实际意义的多时间尺度随机逼近设置，即随机逼近算法维护一组迭代序列，并使用可能在阶数量级上不同的步长分别更新它们。在应用方面，我们希望了解我们的结果是否能够应用于线性随机逼近和强化学习之外的其他算法。

致谢

我们感谢来自伊利诺伊大学厄巴纳-香槟分校的 R. Srikant 教授，感谢他就使用望远镜技巧建立最大集中界提出的深刻见解。

参考文献

[1] Banach, S. (1922). 关于抽象集合中的运算及其在积分方程中的应用。*Fund.math*，3(1)：133–181。[2] Beck, A. (2017). 优化中的一阶方法，第25卷。SIAM。[3] Beck, C. L. and Srikant, R. (2012). 常数步长 Q学习的误差界。《系统与控制快报》，61(12)：1203–1208。[4] Beck, C. L. and Srikant, R. (2013). 改进常数步长Q学习中期望误差的上界。在*2013*年美国控制会议，第1926–1931页。IEEE。[5] Bellman, R., Bellman, R., and Corporation, R. (1957). 动态规划。兰德公司研究报告。普林斯顿大学出版社。[6] Benveniste, A., Métivier, M., and Priouret, P. (2012). 自适应算法与随机逼近，第22卷。施普林格科学与商业媒体。[7] Bernstein, S. (1924). 关于切比雪夫不等式及拉普拉斯误差公式的一种修正。乌克兰萨武科学研究所数学分部年刊，1(4)：38–49。[8] Bertsekas, D. P. and Tsitsiklis, J. N. (1996). 神经动态规划。Athena Scientific。

[9] Bhandari, J., Russo, D., and Singal, R. (2021). 采用线性函数逼近的时序差分学习的有限时间分析。《运筹学》，69(3)：950–973。[10] Borkar, V. S. (2009). 随机逼近：动力系统视角，第48卷。施普林格。[11] Bottou, L., Curtis, F. E., and Nocedal, J. (2018). 大规模机器学习的优化方法。《*SIAM*评论》，60(2)：223–311。[12] Chandak, S. and Borkar, V. S. (2023). 采用函数逼近的td (0)的集中界。*arXiv*预印本 *arXiv:2312.10424*。[13] Chandak, S., Borkar, V. S., and Dodhia, P. (2022). 压缩随机逼近与强化学习的集中性。《随机系统》。[14] Chen, Z., Khodadadian, S., and Maguluri, S. T. (2022a). 采用线性函数逼近的离策略自然演员-评论家的有限样本分析。《*IEEE*控制系统快报》，6：2611–2616。[15] Chen, Z., Ma, S., and Zhou, Y. (2021a). 用于零和马尔可夫博弈的样本高效随机策略外梯度算法。在国际学习表征会议。[16] Chen, Z. and Maguluri, S. T. (2022). 离策略采样和线性函数逼近下基于策略的方法的样本复杂度。在国际人工智能与统计会议，第11195–11214页。PMLR。[17] Chen, Z., Maguluri, S. T., Shakkottai, S., and Shanmugam, K. (2020). 使用光滑凸包络的压缩随机逼近有限样本分析。神经信息处理系统进展，33。[18] Chen, Z., Maguluri, S. T., Shakkottai, S., and Shanmugam, K. (2021b). 通过广义贝尔曼算子进行离策略时序差分学习的有限样本分析。神经信息处理系统进展，34：21440–21452。[19] Chen, Z., Maguluri, S. T., Shakkottai, S., and Shanmugam, K. (2023a). 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》。[20] Chen, Z., Maguluri, S. T., Shakkottai, S., and Shanmugam, K. (2023b). 马尔可夫型随机逼近有限样本保证的李雅普诺夫理论。《运筹学》。[21] Chen, Z., Zhang, S., Doan, T. T., Clarke, J.-P., and Maguluri, S. T. (2022b). 非线性随机逼近的有限样本分析及其在强化学习中的应用。《自动化》，146：110623。[22] Chen, Z., Zhou, Y., Chen, R.-R., and Zou, S. (2022c). 具有有限时间分析的样本与通信高效去中心化演员-评论家算法。在国际机器学习会议，第3794–3834页。PMLR。[23] Chernoff, H. (1952). 基于观测值总和的假设检验的渐近效率度量。《数理统计年刊》，第493–507页。[24] Dalal, G., Szörényi, B., Thoppe, G., and Mannor, S. (2018a). 采用函数逼近的TD (0)有限样本分析。在*AAAI*人工智能会议论文集，第32卷。[25] Dalal, G., Thoppe, G., Szörényi, B., and Mannor, S. (2018b). 双时间尺度随机逼近的有限样本分析及其在强化学习中的应用。在学习理论会议，第1199–1233页。PMLR。

[26] Doan, T., Maguluri, S. 和 Romberg, J.（2019）。多智能体强化学习中采用线性函数逼近的分布式 TD(0) 有限时间分析。载于国际机器学习会议，第 1626–1635 页。[27] Duchi, J. C., Agarwal, A., Johansson, M. 和 Jordan, M. I.（2012）。遍历镜像下降。*SIAM* 优化期刊，22(4)：1549–1578。[28] Durmus, A., Moulines, E., Naumov, A. 和 Samsonov, S.（2024）。线性随机逼近的 Polyak–Ruppert 平均迭代序列的有限时间高概率界。运筹学数学。[29] Durmus, A., Moulines, E., Naumov, A., Samsonov, S., Scaman, K. 和 Wai, H.-T.（2021a）。固定步长线性随机逼近的精确高概率界。神经信息处理系统进展，34：30063–30074。[30] Durmus, A., Moulines, E., Naumov, A., Samsonov, S. 和 Wai, H.-T.（2021b）。马尔可夫噪声下随机矩阵乘积的稳定性：在线性随机逼近与时序差分学习中的应用。预印本 *arXiv:2102.00185v1*。[31] Even-Dar, E. 和 Mansour, Y.（2003）。 Q学习的学习速率。《机器学习研究杂志》，5（12）：1–25。[32] Glynn, P. W. 和 Iglehart, D. L.（1989）。随机模拟中的重要性抽样。管理科学，35(11)：1367–1392。[33] Gosavi, A.（2006）。 Q学习中迭代序列的有界性。《系统与控制快报》，55(4)：347–349。[34] Haddad, W. M. 和 Chellaboina, V.（2011）。非线性动力系统与控制：一种基于 *Lyapunov* 的方法。普林斯顿大学出版社。[35] Harutyunyan, A., Bellemare, M. G., Stepleton, T. 和 Munos, R.（2016）。 $q(\lambda)$及其离策略校正。载于国际算法学习理论会议，第 305–320 页。施普林格。[36] Harvey, N. J. A., Liaw, C. 和 Randhawa, S.（2019）。强凸随机梯度下降的简洁且最优的高概率界。预印本 *arXiv:1909.00843v1*。[37] Hazan, E. 和 Kale, S.（2014）。超越遗憾最小化障碍：随机强凸优化的最优算法。《机器学习研究杂志》，15：2489–2512。[38] Hoeffding, W.（1994）。有界随机变量之和的概率不等式。*Wassily Hoeffding* 文集，第 409–426 页。[39] Jin, C., Netrapalli, P., Ge, R., Kakade, S. M. 和 Jordan, M. I.（2019）。关于次高斯范数随机向量集中不等式的简短说明。预印本 *arXiv:1902.03736*。[40] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. 等（2021）。利用 AlphaFold 实现高精度蛋白质结构预测。《自然》，596(7873)：583–589。[41] Khalil, H. K. 和 Grizzle, J. W.（2002）。非线性系统，第 3 卷。普伦蒂斯·霍尔，上萨德尔河，新泽西州。[42] Khodadadian, S., Doan, T. T., Romberg, J. 和 Maguluri, S. T.（2022）。双时间尺度自然演员–评论家算法的有限样本分析。*IEEE* 自动控制汇刊，68(6)：3273–3284。

[43] Kober, J., Bagnell, J. A. 和 Peters, J.（2013）。机器人学中的强化学习：综述。《国际机器人研究杂志》，32(11)：1238–1274。[44] Kushner, H. J. 和 Clark, D. S.（2012）。约束与非约束系统的随机逼近方法，第 26 卷。施普林格科学与商业媒体。[45] Lakshminarayanan, C. 和 Szepesvari, C.（2018）。线性随机逼近：常数步长与迭代平均能达到什么效果？载于国际人工智能与统计会议，第 1347–1355 页。[46] Lan, G.（2020）。机器学习的一阶与随机优化方法。施普林格。[47] Lee, D.（2024）。 Q学习的最终迭代收敛界：切换系统方法。*IEEE* 自动控制汇刊。[48] Li, G., Cai, C., Chen, Y., Gu, Y., Wei, Y. 和 Chi, Y.（2021a）。收紧 q学习样本复杂度对时域的依赖。《机器学习研究论文集》。[49] Li, G., Cai, C., Chen, Y., Wei, Y. 和 Chi, Y.（2024）。 Q学习是否具有极小极大最优性？精确样本复杂度分析。《运筹学》，72(1)：222–236。[50] Li, G., Wei, Y., Chi, Y., Gu, Y. 和 Chen, Y.（2021b）。异步 Q学习的样本复杂度：更精细的分析与方差缩减。《*IEEE*信息论汇刊》，68(1)：448–473。[51] Li, G., Wei, Y., Chi, Y., Gu, Y. 和 Chen, Y.（2021c）。Softmax策略梯度方法可能需要指数时间才能收敛。载于学习理论会议，第 3107–3110 页。PMLR。[52] Lou, Z., Zhu, W. 和 Wu, W. B.（2022）。超越次高斯噪声：随机梯度下降的精细集中分析。《机器学习研究杂志》。[53] Mei, J., Xiao, C., Szepesvari, C. 和 Schuurmans, D.（2020）。关于 Softmax策略梯度方法的全局收敛速率。载于国际机器学习会议，第 6820–6829 页。PMLR。[54] Mou, W., Khamaru, K., Wainwright, M. J., Bartlett, P. L. 和 Jordan, M. I.（2022）。巴拿赫空间中的最优方差缩减随机逼近。预印本 *arXiv:2201.08518v1*。[55] Mou, W., Li, C. J., Wainwright, M. J., Bartlett, P. L. 和 Jordan, M. I.（2020）。关于线性随机逼近：细粒度的 Polyak-Ruppert 分析与非渐近集中。载于学习理论会议，第 2947–2997 页。PMLR。[56] Munos, R., Stepleton, T., Harutyunyan, A. 和 Belle mare, M. G.（2016）。安全高效的离策略强化学习。载于第*30*届国际神经信息处理系统大会论文集，第 1054–1062 页。[57] Na, H. 和 Lee, D.（2024）。同步双 q学习的有限时间分析。*arXiv* 预印本 *arXiv:2406.09946*。[58] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. 等（2022）。训练语言模型，使其能够遵循指令并利用人类反馈。神经信息处理系统进展，35：27730–27744。

[59] Prashanth, L. A., Korda, N. 和 Munos, R.（2021）。线性函数逼近下时序差分学习的集中界：批量数据和均匀采样的情形。机器学习，110(3):559–618。[60] Precup, D.（2000）。异策略策略评估的资格迹。计算机科学系教职工出版物系列，第 80 页。[61] Puterman, M. L.（1995）。马尔可夫决策过程：离散随机动态规划。运筹学会杂志，46(6):792–792。[62] Qiu, S.、Yang, Z.、Ye, J. 和 Wang, Z.（2019）。关于演员-评论家算法的有限时间收敛性。载于神经信息处理系统进展强化学习优化基础研讨会（*NeurIPS*）。[63] Qu, G. 和 Wierman, A.（2020）。异步随机逼近和q-学习的有限时间分析。载于学习理论会议，第 3185–3205 页。PMLR。[64] Rakhlin, A.、Shamir, O. 和 Sridharan, K.（2012）。使梯度下降对强凸随机优化达到最优。载于国际机器学习会议。[65] Robbins, H. 和 Monro, S.（1951）。一种随机逼近方法。《数理统计年刊》，第 400–407 页。[66] Ryu, E. K. 和 Boyd, S.（2016）。单调算子方法入门。*Appl.Comput. Math*，15(1):3–43。[67] Shalev-Shwartz, S. 等（2012）。在线学习与在线凸优化。*Foundations andTrends®*机器学习，4(2):107–194。[68] Silver, D.、Schrittwieser, J.、Simonyan, K.、Antonoglou, I.、Huang, A.、Guez, A.、Hubert, T.、Baker, L.、Lai, M.、Bolton, A. 等（2017）。在没有人类知识的情况下掌握围棋。《自然》，550(7676):354。[69] Srikant, R. 和 Ying, L.（2019）。线性随机逼近和时序差分学习的有限时间误差界。载于学习理论会议，第 2803–2830 页。[70] Sutton, R. S.（1988）。通过时序差分方法学习预测。机器学习，3(1):9–44。[71] Sutton, R. S. 和 Barto, A. G.（2018）。强化学习：导论。麻省理工学院出版社。[72] Telgarsky, M.（2022）。具有二次有界损失的随机线性优化在一般数据上从不发生过拟合。载于学习理论会议，第 5453–5488 页。PMLR。[73] Thoppe, G. 和 Borkar, V.（2019）。通过阿列克谢耶夫公式得到的随机逼近集中界。《随机系统》，9(1):1–26。[74] Tsitsiklis, J. N.（1994）。异步随机逼近和 Q-学习。机器学习，16(3):185–202。[75] Tsitsiklis, J. N. 和 Van Roy, B.（1997）。函数逼近下时序差分学习的分析。载于神经信息处理系统进展，第 1075–1081 页。[76] Ville, J.（1939）。关于集体概念的批判性研究。*Bull.Amer.Math.Soc*，45(11):824。

[77] Wainwright, M. J.（2019a）。高维统计：一种非渐近视角。剑桥大学出版社。[78] Wainwright, M. J.（2019b）。锥收缩算子下的随机逼近：针对 ℓ_∞-界的尖锐 Q-学习界。预印本 *arXiv:1905.06265*。[79]Wainwright, M. J.（2019c）。方差缩减 Q-学习达到极小极大最优。预印本 *arXiv:1906.04697*。[80] Watkins, C. J. 和 Dayan, P.（1992）。 Q-学习。机器学习，8(3-4):279–292。[81] Wu, Y. F.、Zhang, W.、Xu, P. 和 Gu, Q.（2020）。双时间尺度演员-评论家方法的有限时间分析。神经信息处理系统进展，33:17617–17628。[82] Xu, T. 和 Liang, Y.（2021）。基于价值的强化学习算法的样本复杂度界。载于国际人工智能与统计会议，第 811–819 页。PMLR。[83] Zeng, S. 和 Doan, T.（2024）。快速双时间尺度随机梯度方法及其在强化学习中的应用。载于第三十七届年度学习理论会议，第 5166–5212 页。PMLR。[84] Zou, S.、Xu, T. 和 Liang, Y.（2019）。线性函数逼近下 SARSA 的有限样本分析。载于神经信息处理系统进展，第 8668–8678 页。

附录

支持定理的技术结果证明2.1

A.1 引理证明3.1

由于 $\beta_1 = 0$ 情形的证明是平凡的，我们将只关注 $\beta_1 \neq 0$ 的情形。通过归纳法可以很容易地证明，对于所有 $k \geq 0$，都有 $w_k > 0$。利用数值不等式 $1 + x \leq e^x$，对于所有 $x \in \mathbb{R}$，我们有，对于所有 $k \geq 0$，

$$w_{k+1} \leq e^{\beta_1 \alpha_k} w_k + \beta_2 \alpha_k.$$

将前一个不等式的两边同乘 $e^{-\beta_1 \sum_{i=0}^{k} \alpha_i}$，我们得到

$$\begin{aligned}
e^{-\beta_1 \sum_{i=0}^{k} \alpha_i} w_{k+1} &\leq e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_k + \beta_2 \alpha_k e^{-\beta_1 \sum_{i=0}^{k} \alpha_i} \\
&= e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_k + \frac{\beta_2}{\beta_1} (1 + \beta_1 \alpha_k - 1) e^{-\beta_1 \sum_{i=0}^{k} \alpha_i} \\
&\leq e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_k + \frac{\beta_2}{\beta_1} (e^{\beta_1 \alpha_k} - 1) e^{-\beta_1 \sum_{i=0}^{k} \alpha_i} \\
&\leq e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_k + \frac{\beta_2}{\beta_1} (e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i} - e^{-\beta_1 \sum_{i=0}^{k} \alpha_i}).
\end{aligned}$$

通过望远镜求和，我们有

$$e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_k \leq w_0 + \frac{\beta_2}{\beta_1} (1 - e^{-\beta_1 \sum_{i=0}^{k-1} \alpha_i}),$$

这意味着

$$w_k \leq e^{\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_0 + \frac{\beta_2}{\beta_1} (e^{\beta_1 \sum_{i=0}^{k-1} \alpha_i} - 1).$$

这证明了 $\beta_1 > 0$ 的情形。当 $\beta_1 < 0$ 时，注意在这种情形下，我们有 $e^{\beta_1 \sum_{i=0}^{k-1} \alpha_i} \in (0, 1)$。因此，根据前面的不等式，我们得到

$$\begin{aligned}
w_k &\leq e^{\beta_1 \sum_{i=0}^{k-1} \alpha_i} w_0 + \frac{\beta_2}{\beta_1} (e^{\beta_1 \sum_{i=0}^{k-1} \alpha_i} - 1) \\
&= e^{\beta_1 \sum_{i=0}^{k-1} \alpha_i} \left(w_0 + \frac{\beta_2}{\beta_1} \right) - \frac{\beta_2}{\beta_1} \\
&\leq \begin{cases} w_0, & w_0 + \frac{\beta_2}{\beta_1} \geq 0, \\ -\frac{\beta_1}{\beta_1}, & w_0 + \frac{\beta_2}{\beta_1} < 0, \end{cases} \\
&\leq w_0 - \frac{\beta_2}{\beta_1}.
\end{aligned}$$

证明现已完成。

A.2 命题3.3 的证明

反复使用公式（14），对于所有 $k \geq 0$ 都有

$$Z_k \leq \underbrace{\exp\left(-\frac{\alpha D_0/2-1}{\alpha}\sum_{i=0}^{k-1}\alpha_i\right)}_{T_5} Z_0 + 2D_2(1+\|x^*\|_c)^2 \underbrace{\sum_{i=0}^{k-1}\alpha_i^2\lambda_i \exp\left(-\frac{\alpha D_0/2-1}{\alpha}\sum_{j=i+1}^{k-1}\alpha_j\right)}_{T_6}. \tag{70}$$

接下来需要对 T_5 和 T_6进行界定。我们先考虑项 T_5。注意，对于任意非增函数 f： $\mathbb{R} \mapsto \mathbb{R}$ 和 $n_1 \leq n_2$，都有

$$\int_{n_1}^{n_2+1} f(x)dx \leq \sum_{i=n_1}^{n_2} f(i) \leq \int_{n_1-1}^{n_2} f(x)dx.$$

因此，我们有

$$\sum_{i=0}^{k-1}\alpha_i = \sum_{i=0}^{k-1}\frac{\alpha}{i+h} \geq \int_0^k \frac{\alpha}{x+h}dx = \alpha\log\left(\frac{k+h}{h}\right).$$

由此可得

$$T_5 \leq \exp\left(-(\alpha D_0/2-1)\log\left(\frac{k+h}{h}\right)\right) = \left(\frac{h}{k+h}\right)^{\alpha D_0/2-1}.$$

至于项 T_6，我们有

$$\begin{aligned} T_6 &= \sum_{i=0}^{k-1}\alpha_i^2\lambda_i \exp\left(-\frac{\alpha D_0/2-1}{\alpha}\sum_{j=i+1}^{k-1}\alpha_j\right) \\ &\leq \sum_{i=0}^{k-1}\alpha_i^2\lambda_i \exp\left(-\frac{\alpha D_0/2-1}{\alpha}\alpha\log\left(\frac{k+h}{i+1+h}\right)\right) \\ &= \sum_{i=0}^{k-1}\alpha_i^2\lambda_i\left(\frac{i+1+h}{k+h}\right)^{\alpha D_0/2-1} \\ &= \left(\frac{1}{k+h}\right)^{\alpha D_0/2-1}\sum_{i=0}^{k-1}\alpha_i^2\lambda_i(i+1+h)^{\alpha D_0/2-1}. \end{aligned}$$

接下来，由于 $\lambda_k = \theta\alpha_k^{-1}T_k(\delta)^{-1}$ 和 $\|x_0-x^*\|_c^2 \leq T_0(\delta) \leq T_k(\delta)$，我们有

$$\alpha_k^2\lambda_k = \frac{\theta\alpha_k}{T_k(\delta)} \leq \frac{\theta\alpha_k}{T_0(\delta)} \leq \frac{\theta\alpha_k}{\|x_0-x^*\|_c^2}.$$

由此可得

$$T_6 \leq \left(\frac{1}{k+h}\right)^{\alpha D_0/2-1}\frac{\theta}{\|x_0-x^*\|_c^2}\sum_{i=0}^{k-1}\frac{\alpha}{i+h}(i+1+h)^{\alpha D_0/2-1}$$

$$= \left(\frac{1}{k+h}\right)^{\alpha D_0/2-1} \frac{\theta}{\|x_0-x^*\|_c^2} \sum_{i=0}^{k-1} \frac{\alpha(i+1+h)}{i+h}(i+1+h)^{\alpha D_0/2-2}$$
$$\leq 2\alpha \left(\frac{1}{k+h}\right)^{\alpha D_0/2-1} \frac{\theta}{\|x_0-x^*\|_c^2} \sum_{i=0}^{k-1}(i+1+h)^{\alpha D_0/2-2}.$$

类似地，利用积分对求和进行界定，可得

$$\sum_{i=0}^{k-1} \frac{1}{(i+1+h)^{2-\alpha D_0/2}} \leq \begin{cases} \dfrac{1}{\alpha D_0/2-1}(k+h)^{\alpha D_0/2-1}, & \alpha \in (2/D_0, 4/D_0), \\ k, & \alpha = 4/D_0, \\ \dfrac{e}{\alpha D_0/2-1}(k+h)^{\alpha D_0/2-1}, & \alpha > 4/D_0. \end{cases}$$

由此可得

$$T_6 \leq \frac{2\alpha e\theta}{(\alpha D_0/2-1)\|x_0-x^*\|_c^2}.$$

利用我们对各项得到的上界 T_5 和 T_6 在式（70）中，可得

$$Z_k \leq Z_0 \left(\frac{h}{k+h}\right)^{\alpha D_0/2-1} + \frac{4\alpha e\theta D_2}{\alpha D_0/2-1} \frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2}.$$

这就证明了命题3.3的式（38）。

A.3 命题3.4的证明

对于任意 $\epsilon > 0$ 和 $K \geq 0$，有

$$\mathbb{P}\left(\sup_{k\geq K}\left\{\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k-x^*) - D_3\sum_{i=0}^{k-1}\alpha_i\right\} > \epsilon\right)$$
$$= \mathbb{P}\left(\sup_{k\geq K}\left\{\exp\left(\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k-x^*) - D_3\sum_{i=0}^{k-1}\alpha_i\right)\right\} > e^{\epsilon}\right)$$
$$\leq \mathbb{E}\left[\exp\left(\lambda_K \mathbb{1}_{E_K(\delta)} M(x_K-x^*) - D_3\sum_{i=0}^{K-1}\alpha_i - \epsilon\right)\right] \tag{71}$$
$$\leq \exp\left(Z_0\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + \frac{4\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2} - D_3\sum_{i=0}^{K-1}\alpha_i - \epsilon\right), \tag{72}$$

其中，式（71）由 Ville 最大不等式得出，而式（72）由命题 3.3 得出。令

$$\delta' = \exp\left(Z_0\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + \frac{4\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2} - D_3\sum_{i=0}^{K-1}\alpha_i - \epsilon\right),$$

这意味着

$$\epsilon = \log(1/\delta') + Z_0\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + \frac{4\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2} - D_3\sum_{i=0}^{K-1}\alpha_i.$$

因此，以至少 $1-\delta'$ 的概率，我们有

$$\sup_{k\geq K}\left\{\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k - x^*) - D_3 \sum_{i=0}^{k-1}\alpha_i\right\}$$
$$\leq \log(1/\delta') + Z_0\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + \frac{4\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2} - D_3\sum_{i=0}^{K-1}\alpha_i. \tag{73}$$

最后一步是对 Z_0进行界定。注意到

$$Z_0 = \lambda_0 M(x_k - x^*) = \frac{\theta\|x_k-x^*\|_M^2}{2\alpha_0 T_0(\delta)} \leq \frac{\theta\|x_0-x^*\|_M^2}{2\alpha_0\|x_0-x^*\|_c^2} \leq \frac{\theta}{2\alpha_0\ell_{cM}^2} \leq \frac{D_0}{16\alpha_0 D_1\ell_{cM}^2},$$

其中我们使用了 $M(x)=\frac{1}{2}\|x\|_M^2$(参见引理 3.2), $\|x_0-x^*\|_c^2 \leq T_0(\delta)$，以及 $\theta = D_0\|x_0-x^*\|_c^2/[8D_1((1+\|x^*\|_c)^2+\|x_0-x^*\|_c^2)]$。利用我们在式 (73) 中对 Z_0 得到的界，即可得出结果。

我们以至少 $1-\delta'$ 的概率有:

$$\sup_{k\geq K}\left\{\lambda_k \mathbb{1}_{E_k(\delta)} M(x_k-x^*)\right\} \leq \log(1/\delta') + \frac{D_0}{16\alpha_0 D_1\ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}$$
$$+ \frac{4\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2} + D_3\sum_{i=K}^{k-1}\alpha_i.$$

最后一步是计算 $\sum_{i=K}^{k-1}\alpha_i$，并将 $M(x_k-x^*)$与 $\|x_k-x^*\|_c^2$联系起来。由于

$$\sum_{i=K}^{k-1}\alpha_i = \sum_{i=K}^{k-1}\frac{\alpha}{i+h} \leq \int_{K-1}^{k-1}\frac{\alpha}{x+h}dx = \alpha\log\left(\frac{k-1+h}{K-1+h}\right)$$

and

$$M(x_k-x^*) = \frac{1}{2}\|x_k-x^*\|_M^2 \geq \frac{1}{2u_{cM}^2}\|x_k-x^*\|_c^2, \qquad \text{(Lemma 3.2 (3))}$$

我们以至少 $1-\delta'$ 的概率有

$$\sup_{k\geq K}\left\{\lambda_k \mathbb{1}_{E_k(\delta)}\|x_k-x^*\|_c^2\right\} \leq 2u_{cM}^2\log(1/\delta') + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1\ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}$$
$$+ \frac{8u_{cM}^2\alpha e D_2\theta}{\alpha D_0/2-1}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2} + 2\alpha D_3 u_{cM}^2\log\left(\frac{k-1+h}{K-1+h}\right).$$

A.4 断言3.3 的证明

Let

$$D_4 = \frac{D_0((1+\|x^*\|_c)^2}{8D_1((1+\|x^*\|_c)^2+\|x_0-x^*\|_c^2)},\ c_1 = \frac{(32D_1(1+D_4)u_{cM}^2)^m(D^2+\sigma^2 D_4)}{D_0^m D^2},$$
$$c_2 = \frac{D_0}{16\alpha_0 D_1\ell_{cM}^2}, c_3 = \frac{8\alpha e D_2 D_4}{\alpha D_0-2},\ c_4 = \alpha D_3,\ \text{ and } c_1' = \frac{32u_{cM}^2 D_1(1+\sigma D_4\alpha^2)(D_4+1)}{D_0}.$$

我们先从命题3.1开始，它指出

$$\mathbb{P}(\|x_k - x^*\|_c^2 \le B_k(D)^2, \forall\, k \ge 0) = 1,$$

其中

$$B_k(D) = \begin{cases} \left(\dfrac{k-1+h}{h-1}\right)^{\alpha D}\left(\|x_0 - x^*\|_c + \dfrac{\sigma(1+\|x^*\|_c)}{D}\right), & D > 0, \\ \|x_0 - x^*\|_c + \sigma(1+\|x^*\|_c)\alpha \log\left(\dfrac{k-1+h}{h-1}\right), & D = 0. \end{cases}$$

(1) 我们首先考虑 $D > 0$的情形。回顾一下，我们已假设（不失一般性）$2\alpha D$ 是一个整数，并定义了 $m = 2\alpha D + 1$。令 $\delta_1,\ \delta_2,\ \cdots,\ \delta_m > 0$ 满足 $\sum_{i=1}^m \delta_i \le 1$。对 m 重复使用命题 3.5 次，我们以至少 $1 - \sum_{i=1}^m \delta_i$ 的概率得到

$$\begin{aligned} \|x_k - x^*\|_c^2 \le{}& \frac{\alpha_k^m B_k^2(D)}{\theta^m} \prod_{i=1}^{m-1} \left[2u_{cM}^2 \log(1/\delta_i) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\right. \\ &\left. + \frac{16u_{cM}^2 \alpha e D_2 \theta}{\alpha D_0 - 2}\frac{(1+\|x^*\|_c)^2}{\|x_0 - x^*\|_c^2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{h-1}\right)\right] \\ &\times \left[2u_{cM}^2 \log(1/\delta_m) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right. \\ &\left. + \frac{16u_{cM}^2 \alpha e D_2 \theta}{\alpha D_0 - 2}\frac{(1+\|x^*\|_c)^2}{\|x_0 - x^*\|_c^2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{K-1+h}\right)\right] \\ ={}& \frac{2^m u_{cM}^{2m} \alpha_k^m B_k^2(D)}{\theta^m} \prod_{i=1}^{m-1}\left[\log(1/\delta_i) + c_2 + c_3 + c_4 \log\left(\frac{k-1+h}{h-1}\right)\right] \\ &\times \left[\log(1/\delta_m) + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4 \log\left(\frac{k-1+h}{K-1+h}\right)\right]. \end{aligned}$$

对所有 $i \in \{1, 2, \cdots, m\}$ 取 $\delta_i = \delta/m$ 后，前一个不等式可写为

$$\begin{aligned} \|x_k - x^*\|_c^2 \le{}& \frac{2^m u_{cM}^{2m} \alpha_k^m B_k^2(D)}{\theta^m}\left[\log(m/\delta) + c_2 + c_3 + c_4 \log\left(\frac{k-1+h}{h-1}\right)\right]^{m-1} \\ &\times \left[\log(m/\delta) + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4 \log\left(\frac{k-1+h}{K-1+h}\right)\right]. \end{aligned}$$

接下来，注意到

$$\begin{aligned} \frac{2^m u_{cM}^{2m} \alpha_k^m B_k^2(D)}{\theta^m} &= \frac{2^{4m} u_{cM}^{2m} D_1^m (1+D_4)^m \alpha_k \alpha^{m-1}}{D_0^m (k+h)^{m-1}}\left(\frac{k-1+h}{h-1}\right)^{m-1} \\ &\quad \times \left(1 + \sigma D_4^{1/2}/D\right)^2 \|x_0 - x^*\|_c^2 \\ &\le \frac{32^m u_{cM}^{2m} D_1^m (1+D_4)^m \left(1 + \sigma^2 D_4/D^2\right)}{D_0^m} \alpha_k \|x_0 - x^*\|_c^2 \\ &= c_1 \alpha_k \|x_0 - x^*\|_c^2. \end{aligned}$$

因此，对于任意 $\delta > 0$ 和 $K \ge 0$，我们以至少 $1 - \delta$ 的概率得到

$$\|x_k - x^*\|_c^2 \le c_1 \alpha_k \|x_0 - x^*\|_c^2 \left[\log(m/\delta) + c_2 + c_3 + c_4 \log\left(\frac{k-1+h}{h-1}\right)\right]^{m-1}$$

$$\times \left[\log(m/\delta) + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4 \log\left(\frac{k-1+h}{K-1+h}\right)\right]$$

对所有 $k \geq K$。定理 2.1 (1) 的证明完毕。

（2）当 $D = 0$时，使用命题3.5 一次，我们以至少 $1-\delta$ 的概率有

$$\begin{aligned}
\|x_k - x^*\|_c^2 &\leq \frac{\alpha_k B_k(D)^2}{\theta}\left[2u_{cM}^2 \log(1/\delta) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right. \\
&\quad \left. + \frac{16u_{cM}^2 \alpha e D_2 D_4}{\alpha D_0 - 2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{K-1+h}\right)\right] \\
&\leq \frac{16D_1(1+\sigma D_4\alpha^2)(D_4+1)}{D_0}\|x_0 - x^*\|_c^2 \alpha_k \log^2\left(\frac{k-1+h}{h-1}\right) \\
&\quad \times \left[2u_{cM}^2 \log(1/\delta) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right. \\
&\quad \left. + \frac{16u_{cM}^2 \alpha e D_2 D_4}{\alpha D_0 - 2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{K-1+h}\right)\right]
\end{aligned}$$

对所有 $k \geq K$均成立。因此，根据 $\{c_i\}_{2\leq i\leq 4}$ 和 c_1'的定义，前一个不等式可写为

$$\begin{aligned}
\|x_k - x^*\|_c^2 &\leq c_1' \|x_0 - x^*\|_c^2 \alpha_k \log^2\left(\frac{k-1+h}{h-1}\right) \\
&\quad \times \left[\log(1/\delta) + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4 \log\left(\frac{k-1+h}{K-1+h}\right)\right].
\end{aligned}$$

定理2.1的（2）部分证明完毕。

（3）当 $D < 0$时，使用命题 3.5一次，我们以至少 $1-\delta$ 的概率有

$$\begin{aligned}
\|x_k - x^*\|_c^2 &\leq \frac{\alpha_k B_k(D)^2}{\theta}\left[2u_{cM}^2 \log(1/\delta) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right. \\
&\quad \left. + \frac{16u_{cM}^2 \alpha e D_2 D_4}{\alpha D_0 - 2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{K-1+h}\right)\right] \\
&\leq \frac{8\alpha_k D_1((1+\|x^*\|_c)^2 + \|x_0 - x^*\|_c^2)}{D_0 \|x_0 - x^*\|_c^2}\left(\|x_0 - x^*\|_c - \frac{\sigma(1+\|x^*\|_c)}{D}\right)^2 \\
&\quad \times \left[2u_{cM}^2 \log(1/\delta) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right. \\
&\quad \left. + \frac{16u_{cM}^2 \alpha e D_2 D_4}{\alpha D_0 - 2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{K-1+h}\right)\right] \\
&\leq \frac{16(D^2+\sigma^2)D_1}{D_0 D^2}\frac{(\|x_0 - x^*\|_c^2 + (1+\|x^*\|_c)^2)^2}{\|x_0 - x^*\|_c^4}\|x_0 - x^*\|_c^2 \alpha_k \\
&\quad \times \left[2u_{cM}^2 \log(1/\delta) + \frac{2u_{cM}^2 D_0}{16\alpha_0 D_1 \ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right. \\
&\quad \left. + \frac{16u_{cM}^2 \alpha e D_2 D_4}{\alpha D_0 - 2} + 2\alpha D_3 u_{cM}^2 \log\left(\frac{k-1+h}{K-1+h}\right)\right]
\end{aligned}$$

$$\leq c_1''\|x_0-x^*\|_c^2\alpha_k\left[\log(1/\delta)+c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right.$$
$$\left.+c_3+c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right],$$

其中

$$c_1''=\frac{16(D^2+\sigma^2)D_1}{D_0D^2}\frac{(\|x_0-x^*\|_c^2+(1+\|x^*\|_c)^2)^2}{\|x_0-x^*\|_c^4}.$$

定理 2.1(3) 的证明已完成。

A.5 去除对数因子的乘积

定理 2.1′。 在定理 *2.1 (1)* 的相同假设下，对于任意 $\delta\in(0,1)$ 以及 $K\geq 0$，以至少 $1-\delta$ 的概率，对于所有 $k\geq K$，有

$$\|x_k-x^*\|_c^2\leq\frac{c_1''\alpha\|x_0-x^*\|_c^2}{k+h}\left[\left(\log\left(\frac{m+1}{\delta}\right)\right)^m+1\right]\left[\log\left(\frac{m+1}{\delta}\right)\right.$$
$$\left.+c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}+c_3+c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right].$$

定理*2.1′*的证明。 回顾命题 3.2，我们要求初始界 $\{T_k(\delta)\}$ 是非递减的，以便初始化自举论证。然而，定理 2.1(2) 中不等式的右端项最终会递减，因为当 k 足够大时，多项式项将主导对数项。为克服这一困难，定义

$$\tilde{T}_k(\delta)=c_1\|x_0-x^*\|_c^2\sup_{0\leq k'\leq k}\left\{\alpha_{k'}\left[\log\left(\frac{m}{\delta}\right)+c_2+c_3+c_4\log\left(\frac{k'-1+h}{h-1}\right)\right]^m\right\}.$$

对于任意 $\delta>0$ 和 $k\geq 0$。注意， $T_k(\delta)$ 按定义是一个非递减序列，而定理 2.1 (2) 意味着

$$\mathbb{P}\left(\|x_k-x^*\|_c^2\leq\tilde{T}_k(\delta),\ \forall\, k\geq 0\right)\geq 1-\delta. \tag{74}$$

˜既然 $\{T_k(\delta)\}$ 是一个非递减序列，再进行一步自举以消除对数项的乘积，即可证明定理 2。1′下面我们给出详细过程。

我们从式（74）开始，并使用命题 3.5进行一步自举。对于任意 δ， $\delta'\in(0,1)$且 $K\geq 0$，以至少 $1-\delta-\delta'$的概率，对所有 $k\geq K$，均有

$$\|x_k-x^*\|_c^2\leq\frac{\alpha_k\tilde{T}_k(\delta)}{\theta}\left[2u_{cM}^2\log(1/\delta')+\frac{2u_{cM}^2D_0}{16\alpha_0D_1\ell_{cM}^2}\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1}\right.$$
$$\left.+\frac{16u_{cM}^2\alpha eD_2\theta}{\alpha D_0-2}\frac{(1+\|x^*\|_c)^2}{\|x_0-x^*\|_c^2}+2\alpha D_3u_{cM}^2\log\left(\frac{k-1+h}{K-1+h}\right)\right]$$
$$=\frac{16c_1D_1(D_4+1)\alpha_ku_{cM}^2}{D_0}\|x_0-x^*\|_c^2$$
$$\times\sup_{0\leq k'\leq k}\left\{\alpha_{k'}\left[\log\left(\frac{m}{\delta}\right)+c_2+c_3+c_4\log\left(\frac{k'-1+h}{h-1}\right)\right]^m\right\}$$

$$\times \left[\log(1/\delta') + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right].$$

对所有 $n \geq 0$ 和 $a,\ b > 0$，利用数值不等式 $(a+b)^{n+1} \leq 2^n(a^{n+1}+b^{n+1})$，可得

$$\begin{aligned}
&\sup_{0\leq k'\leq k}\left\{\alpha_{k'}\left[\log\left(\frac{m}{\delta}\right) + c_2 + c_3 + c_4\log\left(\frac{k'-1+h}{h-1}\right)\right]^m\right\} \\
&\leq 2^{m-1}\sup_{0\leq k'\leq k}\left\{\alpha_{k'}\left[\left(\log\left(\frac{m}{\delta}\right)\right)^m + \left(c_2 + c_3 + c_4\log\left(\frac{k'-1+h}{h-1}\right)\right)^m\right]\right\} \\
&\leq 2^{m-1}\left(\log\left(\frac{m}{\delta}\right)\right)^m + 2^{m-1}\sup_{k'\geq 0}\left\{\alpha_{k'}\left(c_2 + c_3 + c_4\log\left(\frac{k'-1+h}{h-1}\right)\right)^m\right\} \\
&= 2^{m-1}(1+c_5)\left[\left(\log\left(\frac{m}{\delta}\right)\right)^m + 1\right],
\end{aligned}$$

其中

$$c_5 := \sup_{k'\geq 0}\left\{\alpha_{k'}\left(c_2 + c_3 + c_4\log\left(\frac{k'-1+h}{h-1}\right)\right)^m\right\}.$$

因此，对于任意 $\delta,\ \delta' > 0$ 和 $K \geq 0$，以下不等式以至少 $1-\delta-\delta'$ 的概率成立：

$$\begin{aligned}
\|x_k - x^*\|_c^2 &\leq \frac{16c_1D_1(D_4+1)\alpha_k u_{cM}^2}{D_0}\|x_0 - x^*\|_c^2 \\
&\quad \times 2^{m-1}(1+c_5)\left[\left(\log\left(\frac{m}{\delta}\right)\right)^m + 1\right] \\
&\quad \times \left[\log(1/\delta') + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right] \\
&\leq \frac{8^{2m+1}D_1^{m+1}(1+D_4)^{m+1}u_{cM}^{2m+2}(1+c_5)}{D_0^{m+1}}\left(1 + \frac{\sigma^2 D_4}{D^2}\right) \\
&\quad \times \alpha_k\|x_0 - x^*\|_c^2\left[\left(\log\left(\frac{m}{\delta}\right)\right)^m + 1\right] \\
&\quad \times \left[\log(1/\delta') + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right] \\
&\leq c_1''\alpha_k\|x_0 - x^*\|_c^2\left[\left(\log\left(\frac{m}{\delta}\right)\right)^m + 1\right] \\
&\quad \times \left[\log(1/\delta') + c_2\left(\frac{h}{K+h}\right)^{\alpha D_0/2-1} + c_3 + c_4\log\left(\frac{k-1+h}{K-1+h}\right)\right],
\end{aligned}$$

其中

$$c_1'' = \frac{8^{2m+1}D_1^{m+1}(1+D_4)^{m+1}u_{cM}^{2m+2}(1+c_5)}{D_0^{m+1}}\left(1 + \frac{\sigma^2 D_4}{D^2}\right).$$

重新指定 $\delta \leftarrow \delta m/(m+1)$ 和 $\delta' \leftarrow \delta/(m+1)$ 后，即可得到该结果。 □

注意，定理 2.1′ 意味着存在 $C_1' > 0$ 使得

$$\mathbb{P}(\sqrt{k+h}\,\|x_k - x^*\|_c > \epsilon) < (m+1)\exp\left(-C_1'\epsilon^{2/(m+1)}\right), \quad \forall\, k \geq 0, \epsilon > 0.$$

与推论2.2 （定理 2.1的（2）相比，我们可以看到，收敛速率提升了一个对数因子，但尾部更重。

B 技术结果的证明（用于支持定理2.3

B.1 引理的证明4.1

对于任意 $k \geq 0$，我们有

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda(k+h)^{c_2}x_k^{c_1}\right)\right] &= \int_0^\infty \mathbb{P}\left(\exp\left(\lambda(k+h)^{c_2}x_k^{c_1}\right) > x\right) dx \\
&\leq e + \int_e^\infty \mathbb{P}\left(\exp\left(\lambda(k+h)^{c_2}x_k^{c_1}\right) > x\right) dx \\
&= e + \int_e^\infty \mathbb{P}\left((k+h)^{c_2}x_k^{c_1} > \log(x)/\lambda\right) dx \\
&\leq e + \int_e^\infty C_1 \exp\left(-\frac{C_2\log(x)}{\lambda}\right) dx \\
&= e + \int_e^\infty C_1 x^{-\frac{C_2}{\lambda}} dx \\
&= e + \frac{C_1\lambda}{\lambda - C_2}\, x^{1-\frac{C_2}{\lambda}}\Big|_e^\infty \\
&= e + \frac{C_1\lambda}{C_2 - \lambda} e^{1-\frac{C_2}{\lambda}}
\end{aligned}$$

其中，最后一行由 $\lambda < C_2$得出。

B.2 引理的证明4.2

$\ell(\cdot)$ 的导数由 $\ell'(x) = e^x - c$ 给出；当 $x > \ln(c)$ 时，该导数严格为正，而当 $x < \ln(c)$ 时，严格为负。因此，$\ell(\cdot)$ 在 $[0,\ln(c))$ 上是严格递减函数，在 $(\ln(c)\ \infty)$ 上是严格递增函数。此外，由于 $\ell(0) = 0$、$\ell(\ln(c)) < 0$，且 $\lim_{x\to\infty}\ell(x) = \infty$，方程 $\ell(x) = 0$ 在 $(0,\infty)$ 上有唯一解（记为 x_c）。因此，$\ell(x) \leq 0$ 在 $[0, x_c]$ 上，而$\ell(x) \geq 0$ 在 $[x_c, \infty)$ 上。

C 定理的技术结果证明2.4

C.1 引理证明5.1

反复使用公式 (48)，可得对所有 $k \geq 0$ 均有

$$Z_k \leq \prod_{j=0}^{k-1}\frac{\alpha_j}{\alpha_{j+1}}\left(1 - \frac{\bar{D}_1\alpha_j}{2}\right) Z_0 + \frac{\bar{D}_1\bar{D}_4}{4\bar{D}_3}\sum_{i=0}^{k-1}\alpha_i \prod_{j=i+1}^{k-1}\frac{\alpha_j}{\alpha_{j+1}}\left(1 - \frac{\bar{D}_1\alpha_j}{2}\right). \tag{75}$$

接下来我们对 th　前一个不等式右端项中的两项。对于第一项，我们有　e

$$\begin{aligned}
\prod_{j=0}^{k-1}\frac{\alpha_j}{\alpha_{j+1}}\left(1 - \frac{\bar{D}_1\alpha_j}{2}\right) &= \frac{\alpha_0}{\alpha_k}\prod_{j=0}^{k-1}\left(1 - \frac{\bar{D}_1\alpha_j}{2}\right) \\
&= \left(\frac{k+h}{h}\right)^z \prod_{j=0}^{k-1}\left(1 - \frac{\bar{D}_1\alpha}{2(k+h)^z}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \left(\frac{k+h}{h}\right)^{z} \exp\left(-\sum_{j=0}^{k-1} \frac{\bar{D}_1 \alpha}{2(k+h)^z}\right) \\
&\leq \left(\frac{k+h}{h}\right)^{z} \exp\left(-\frac{\bar{D}_1 \alpha}{2} \int_0^k \frac{1}{(x+h)^z} dx\right) \\
&= \begin{cases} \left(\frac{h}{k+h}\right)^{\bar{D}_1 \alpha/2-1}, & z=1, \\ \left(\frac{k+h}{h}\right)^{z} \exp\left(-\frac{\bar{D}_1 \alpha}{2(1-z)}((k+h)^{1-z}-h^{1-z})\right), & z \in (0,1). \end{cases}
\end{aligned}
$$

对于公式（75）右端项中的第二项，我们有

$$
\begin{aligned}
\frac{\bar{D}_1 \bar{D}_4}{4\bar{D}_3} \sum_{i=0}^{k-1} \alpha_i \prod_{j=i+1}^{k-1} \frac{\alpha_j}{\alpha_{j+1}} \left(1-\frac{\bar{D}_1 \alpha_j}{2}\right) &= \frac{\bar{D}_1 \bar{D}_4}{4\bar{D}_3} \sum_{i=0}^{k-1} \alpha_i \frac{\alpha_{i+1}}{\alpha_k} \prod_{j=i+1}^{k-1} \left(1-\frac{\bar{D}_1 \alpha_j}{2}\right) \\
&\leq \frac{\bar{D}_1 \bar{D}_4}{4\bar{D}_3 \alpha_k} \sum_{i=0}^{k-1} \alpha_i^2 \prod_{j=i+1}^{k-1} \left(1-\frac{\bar{D}_1 \alpha_j}{2}\right), \qquad (76)
\end{aligned}
$$

其中最后一行由 $\{\alpha_k\}$ 是递减序列得出。式（76）右端项中的这一项在研究迭代算法的现有文献中经常出现，并已得到详细分析。例如， [19,附录 A.3.7。] 中证明了

$$
\sum_{i=0}^{k-1} \alpha_i^2 \prod_{j=i+1}^{k-1} (1-c'\alpha_j) \leq \begin{cases} \frac{4e\alpha}{c'\alpha-1}\alpha_k, & z=1, \alpha > 1/c', \\ \frac{2\alpha_k}{c'}, & z \in (0,1), \alpha > 0, h \geq \left(\frac{4z}{\bar{D}_1 \alpha}\right)^{1/(1-z)} \end{cases}
$$

对于所有 $k \geq 0$ 均成立，其中 $c' \in (0, 1/\alpha_0)$ 是任意常数。因此，我们有

$$
\frac{\bar{D}_1 \bar{D}_4}{4\bar{D}_3 \alpha_k} \sum_{i=0}^{k-1} \alpha_i^2 \prod_{j=i+1}^{k-1} \left(1-\frac{\bar{D}_1 \alpha_j}{2}\right) \leq \begin{cases} \frac{e\bar{D}_1 \bar{D}_4 \alpha}{\bar{D}_3(\bar{D}_1 \alpha/2-1)}, & z=1, \alpha > 2/\bar{D}_1, \\ \frac{\bar{D}_4}{\bar{D}_3}, & z \in (0,1), \alpha > 0, \end{cases}.
$$

利用 u　　我们得到的式（75）右端项中两项的上界，可得

$$
Z_k \leq \begin{cases} Z_0 \left(\frac{h}{k+h}\right)^{\bar{D}_1 \alpha/2-1} + \frac{e\bar{D}_1 \bar{D}_4 \alpha}{\bar{D}_3(\bar{D}_1 \alpha/2-1)}, & z=1, \alpha > 2/\bar{D}_1, \\ Z_0 \left(\frac{k+h}{h}\right)^{z} \exp\left(-\frac{\bar{D}_1 \alpha((k+h)^{1-z}-h^{1-z})}{2(1-z)}\right) + \frac{\bar{D}_4}{\bar{D}_3}, & z \in (0,1), \alpha > 0. \end{cases}
$$

证明完毕。

C.1.1 基于超鞅 + Ville 最大不等式的方法

对所有 $k \geq 0$，令 $M(k) = \exp(\lambda_k M(x_k - x^*) - \bar{D}_5 \sum_{i=0}^{k-1} \alpha_i)$，其中 $\bar{D}_5 = \bar{D}_1 \bar{D}_4/(4\bar{D}_3)$。接下来，我们证明 $\{M(k)\}$ 是关于滤子 $\mathcal{F}_k$ 的超鞅。显然， $M(k)$ 关于

关于 $\mathcal{F}_k$可测，且期望有限（参见命题 5.1）。此外，对于任意 $k \geq 0$，由等式（63）可得

$$\begin{aligned}
\mathbb{E}[\overline{M}(k+1) \mid \mathcal{F}_k] &= \mathbb{E}\left[\exp\left(\lambda_{k+1} M(x_{k+1} - x^*) - \bar{D}_5 \sum_{i=0}^{k} \alpha_i\right) \,\middle|\, \mathcal{F}_k\right] \\
&\leq \exp\left(\lambda_k M(x_k - x^*) + \frac{\bar{D}_1 \bar{D}_4}{4\bar{D}_3}\alpha_k - \bar{D}_5 \sum_{i=0}^{k} \alpha_i\right) \\
&= \exp\left(\lambda_k M(x_k - x^*) - \bar{D}_5 \sum_{i=0}^{k-1} \alpha_i\right) \\
&= \overline{M}(k).
\end{aligned}$$

因此， $\{M(\overline{k})\}$ 是一个超鞅。

N现在我们可以使用 Ville 最大不等式了。对于任意 $K \geq 0$ 和 $\epsilon > 0$，有

$$\begin{aligned}
&\mathbb{P}\left(\sup_{k \geq K}\left\{\lambda_k M(x_k - x^*) - \bar{D}_5 \sum_{i=0}^{k-1} \alpha_i\right\} > \epsilon\right) \\
&= \mathbb{P}\left(\sup_{k \geq K}\left\{\exp\left(\lambda_k M(x_k - x^*) - \bar{D}_5 \sum_{i=0}^{k-1} \alpha_i\right)\right\} > e^{\epsilon}\right) \\
&\leq \mathbb{E}\left[\exp\left(\lambda_K M(x_K - x^*) - \bar{D}_5 \sum_{i=0}^{K-1} \alpha_i - \epsilon\right)\right] \\
&= \exp\left(Z_K - \bar{D}_5 \sum_{i=0}^{K-1} \alpha_i - \epsilon\right).
\end{aligned}$$

令 $\delta = \exp(Z_k - \bar{D}_5 \sum_{i=0}^{K-1} \alpha_i - \epsilon)$，这意味着 $\epsilon = \log(1/\delta) + Z_K - \bar{D}_5 \sum_{i=0}^{K-1} \alpha_i$。于是，前面的不等式可写为

$$M(x_k - x^*) \leq \frac{1}{\lambda_k}\left(\log(1/\delta) + Z_K + \bar{D}_5 \sum_{i=K}^{k-1} \alpha_i\right), \ \forall\, k \geq K.$$

以至少为 $1 - \delta$ 的概率。由于

$$\frac{1}{2\ell_{cM}^2}\|x_k - x^*\|_c^2 \geq M(x_k - x^*) = \frac{1}{2}\|x_k - x^*\|_M^2 \geq \frac{1}{2u_{cM}^2}\|x_k - x^*\|_c^2$$

and

$$\begin{aligned}
\sum_{i=K}^{k-1} \alpha_i &\leq \int_{K-1}^{k-1} \frac{\alpha}{(x+h)^z} dx \\
&= \begin{cases} \alpha \log\left(\dfrac{k-1+h}{K-1+h}\right), & z = 1, \\ \dfrac{\alpha}{1-z}\left((k-1+h)^{1-z} - (K-1+h)^{1-z}\right), & z \in (0,1), \end{cases}
\end{aligned}$$

通过使用 e 的显式上界 Z_K 在命题 5.1中确立了以下结果 .

¯(1) 当 $z=1$ 和 $\alpha>2/D_1$时，对于任意 $K\geq 0$，以至少 $1-\delta$ 的概率，有

$$\|x_k-x^*\|_c^2\leq \frac{16\bar{D}_3u_{cM}^2\alpha\log(1/\delta)}{\bar{D}_1(k+h)}+\frac{u_{cM}^2}{\ell_{cM}^2}\|x_0-x^*\|_c^2\frac{h^{\bar{D}_1\alpha/2}}{(k+h)(K+h)^{\bar{D}_1\alpha/2-1}}$$
$$+\frac{16e\bar{D}_4u_{cM}^2\alpha^2}{(\bar{D}_1\alpha/2-1)(k+h)}+\frac{16\bar{D}_3\bar{D}_5u_{cM}^2\alpha^2}{\bar{D}_1(k+h)}\log\left(\frac{k-1+h}{K-1+h}\right).$$

（2）当 $z\in(0,1)$ 时，对于任意 $K\geq 0$，我们以至少 $1-\delta$ 的概率有

$$\|x_k-x^*\|_c^2\leq \frac{16\bar{D}_3u_{cM}^2\alpha\log(1/\delta)}{\bar{D}_1(k+h)^z}+\frac{16\bar{D}_4u_{cM}^2\alpha}{\bar{D}_1(k+h)^z}$$
$$+\frac{u_{cM}^2}{\ell_{cM}^2}\|x_0-x^*\|_c^2\left(\frac{K+h}{k+h}\right)^z\exp\left(-\frac{\bar{D}_1\alpha}{2(1-z)}((K+h)^{1-z}-h^{1-z})\right)$$
$$+\frac{16\bar{D}_3\bar{D}_5u_{cM}^2\alpha^2}{\bar{D}_1(1-z)}\frac{(k-1+h)^{1-z}-(K-1+h)^{1-z}}{(k+h)^z}.$$

第 6 节技术结果证明

D.1 引理 6.1 的证明

（1）显然，$\bar{F}_\beta(x)=\beta(\bar{A}x-\bar{b})+x$。因此，对于任意 $x_1,\ x_2\in\mathbb{R}^d$，我们有

$$\|\bar{F}_\beta(x_1)-\bar{F}_\beta(x_2)\|_{\bar{P}}=\|(\beta\bar{A}+I)(x_1-x_2)\|_{\bar{P}}\leq\|\beta\bar{A}+I\|_{\bar{P}}\|x_1-x_2\|_{\bar{P}}.$$

¯接下来需要界定诱导矩阵范数$\|\beta A+I\|_{\bar{P}}$。注意到

$$\begin{aligned}\|\beta\bar{A}+I_d\|_{\bar{P}}^2&=\max_{x:\|x\|_{\bar{P}}=1}x^\top(\beta\bar{A}+I_d)^\top\bar{P}(\beta\bar{A}+I_d)x\\&=\max_{x:\|x\|_{\bar{P}}=1}x^\top(\beta^2\bar{A}^\top\bar{P}\bar{A}+\beta\bar{A}^\top\bar{P}+\beta\bar{P}\bar{A}+\bar{P})x\\&=1+\max_{x:\|x\|_{\bar{P}}=1}x^\top(\beta^2\bar{A}^\top\bar{P}\bar{A}-\beta I_d)x\qquad(77)\\&\leq 1+\max_{x:\|x\|_{\bar{P}}=1}\|x\|_2^2(\beta^2\lambda_{\max}(\bar{A}^\top\bar{P}\bar{A})-\beta)\\&=1-\frac{\max_{x:\|x\|_{\bar{P}}=1}\|x\|_2^2}{4\lambda_{\max}(\bar{A}^\top\bar{P}\bar{A})}\qquad(78)\\&\leq 1-\frac{1}{4\lambda_{\max}(\bar{P})\lambda_{\max}(\bar{A}^\top\bar{P}\bar{A})},\qquad(79)\end{aligned}$$

其中，式（77）由李雅普诺夫方程 $\bar{A}^\top\bar{P}+\bar{P}\bar{A}+I_d=0$ 以及式（78）由选取 $\beta=\lambda_{\max}^{-1}(\bar{A}^\top\bar{P}\bar{A})/2$得到。因此，$\bar{F}_\beta(\cdot)$ 关于 $\|\cdot\|_{\bar{P}}$ 是一个压缩映射，其压缩因子为

$$\bar{\gamma}=\left(1-\frac{1}{4\lambda_{\max}(\bar{P})\lambda_{\max}(\bar{A}^\top\bar{P}\bar{A})}\right)^{1/2}.$$

(2) 由于 $\{Y_k\}$ 是独立同分布序列，我们有

$$\mathbb{E}[F_\beta(x_k,Y_k)\mid\mathcal{F}_k]=\beta\bar{A}x_k-\beta\bar{b}+x_k=\bar{F}_\beta(x_k).$$

(3) 在假设 6.2下，对于所有 $k \geq 0$都有

$$\begin{aligned}\|F_\beta(x_k, Y_k) - \bar{F}_\beta(x_k)\|_2 &= \beta\|(A(Y_k) - \bar{A})x_k - (b(Y_k) - \bar{b})\|_2 \\ &\leq \beta(\|A(Y_k)\|_2 + \|\bar{A}\|_2)\|x_k\|_2 + \beta(\|b(Y_k)\|_2 + \|\bar{b}\|_2) \\ &\leq 2\beta A_{\max}\|x_k\|_2 + 2\beta b_{\max}.\end{aligned}$$

因此，对所有 $k \geq 0$，都有

$$\begin{aligned}\|F_\beta(x_k, Y_k) - \bar{F}_\beta(x_k)\|_{\bar{P}} &\leq \lambda_{\max}(\bar{P})\|F_\beta(x_k, Y_k) - \bar{F}_\beta(x_k)\|_2 \\ &\leq 2\beta\lambda_{\max}(\bar{P})(A_{\max}\|x_k\|_2 + b_{\max}) \\ &\leq 2\beta\lambda_{\max}(\bar{P})(A_{\max}\|x_k\|_{\bar{P}}/\lambda_{\min}(\bar{P}) + b_{\max}) \\ &\leq 2\beta\lambda_{\max}(\bar{P})(A_{\max}/\lambda_{\min}(\bar{P}) + b_{\max})(\|x_k\|_{\bar{P}} + 1).\end{aligned}$$

令 $\sigma = 2\beta\lambda_{\max}(P)(A_{\max}/\lambda_{\min}(P) + b_{\max})$，即可得到该结果。

D.2 引理证明6.2

在同策略设定下，对于更具挑战性的 TD(λ) 情形，A 的赫尔维茨性已在 [75] 中得到证明。对于 n 步时序差分学习，同样的结论也成立。

在异策略设定下，令

$$\begin{aligned}D_{\rho,\max} &= \max_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} \pi_b(a \mid s)\rho(s,a), & D_{\rho,\min} &= \min_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} \pi_b(a \mid s)\rho(s,a), \\ D_{c,\max} &= \max_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} \pi_b(a \mid s)c(s,a), & D_{c,\min} &= \min_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} \pi_b(a \mid s)c(s,a).\end{aligned}$$

已在 [16, 命题 4.1] 中证明，只要对于所有 (s,a)，$0 < c(s,a) \leq \rho(s,a)$，$D_{\rho,\max} < 1/\gamma$，$\gamma(D_{\rho,\max} - D_{c,\min})(1-\gamma D_{c,\min})\sqrt{\min_{s,a}\kappa_b(s)\pi_b(a|s)} < 1$，且参数 n 的选取满足 $1 - (1-\gamma D_{\rho,\max})(1-(\gamma D_{c,\min})^n)\frac{1-\gamma D_{c,\min}}{\leq} \min_{s,a}\sqrt{\kappa_b(s)\pi_b(a \mid s)}$，则矩阵 $\bar{A}$ 是赫尔维茨矩阵。

D.3 引理的证明 6.3

我们首先计算 $F(\cdot) = \mathbb{E}[F(\cdot, S, A, S')]$，的显式表达式，其中 $S \sim \kappa_b(\cdot)$、$A \sim \pi_b(\cdot|S)$ 和 $S' \sim P_A(S, \cdot)$。利用 $F(\cdot)$ 的显式表达式，我们有

$$\begin{aligned}&[\bar{F}(Q)](s,a) \\ &= \mathbb{E}\left[\mathbb{1}_{\{(S,A)=(s,a)\}}(\mathcal{R}(S,A) + \gamma\max_{a'\in\mathcal{A}} Q(S',a') - Q(S,A)) + Q(s,a)\right] \\ &= \kappa_b(s)\pi_b(a|s)\left(\mathcal{R}(s,a) + \gamma\sum_{s'\in\mathcal{S}} P_a(s,s')\max_{a'\in\mathcal{A}} Q(s',a')\right) \\ &\quad + (1 - \kappa_b(s)\pi_b(a|s))Q(s,a) \\ &= (1 - D_b((s,a),(s,a)))[\mathcal{H}(Q)](s,a) + (1 - D_b((s,a),(s,a)))Q(s,a),\end{aligned}$$

其中，$\mathcal{H}(\cdot)$ 是 Q-函数的贝尔曼最优算子。因此，

$$\bar{F}(Q) = D_b\mathcal{H}(Q) + (I - D_b)Q.$$

¯（1）由于 ℓ_∞-范数压缩性质， $\mathcal{H}(\cdot)$ 的压缩性质 $F(\cdot)$ 已在¯现有文献 [19]中得到确立。[19, 命题 31] 具体表明， $F(\cdot)$ 是关于 ℓ_∞-范数的压缩算子，其压缩因子为

$$\hat{\gamma}_c = 1 - D_{b,\min}(1-\gamma).$$

¯(2) 根据我们对 $F(\cdot)$ 的定义以及 $\{(S_k, A_k, S_k')\}$ 是独立同分布序列这一事实，可知其无偏性。

(3) 首先，文献中通过归纳论证表明， Q-学习的迭代序列满足确定性一致界： $\|Q_k\|_\infty \le 1/(1-\gamma)$。现在，对于任意 (s,a) 和 $k \ge 0$，我们有

$$F(Q_k, S_k, A_k, S_k') - \mathbb{E}[F(Q_k, S_k, A_k, S_k') \mid \mathcal{F}_k] = F(Q_k, S_k, A_k, S_k') - \bar{F}(Q_k).$$

如果 $(S_k, A_k) \neq (s,a)$，则有

$$\begin{aligned}
&|[F(Q_k, S_k, A_k, S_k')](s,a) - [\bar{F}(Q_k)](s,a)| \\
&= |D_b((s,a),(s,a))(Q_k(s,a) - [\mathcal{H}(Q_k)](s,a))| \\
&\le \|Q_k - \mathcal{H}(Q_k)\|_\infty \\
&\le \|Q_k - Q^*\|_\infty + \|\mathcal{H}(Q^*) - \mathcal{H}(Q_k)\|_\infty \\
&\le 2\|Q_k - Q^*\|_\infty \qquad \text{(the contraction property of } \mathcal{H}(\cdot)) \\
&\le \frac{2}{1-\gamma}.
\end{aligned}$$

如果 $(S_k, A_k) = (s,a)$，则有

$$\begin{aligned}
&|[F(Q_k, S_k, A_k, S_k')](s,a) - [\bar{F}(Q_k)](s,a)| \\
&= |\mathcal{R}(s,a) + \gamma \max_{a' \in \mathcal{A}} Q_k(S_k', a') - D_b((s,a),(s,a))[\mathcal{H}(Q_k)](s,a) \\
&\quad + (1 - D_b((s,a),(s,a)))Q_k(s,a)| \\
&\le \frac{4}{1-\gamma}.
\end{aligned}$$

结合这两种情况，我们有

$$\|F(Q_k, S_k, A_k, S_k') - \mathbb{E}[F(Q_k, S_k, A_k, S_k') \mid \mathcal{F}_k]\|_\infty \le \frac{4}{1-\gamma}.$$

A因此，对于相对于 $\mathcal{F}_k$ 可测的任意随机向量 $v \in \mathbb{R}^d$，我们有

$$\mathbb{E}[\langle F(Q_k, S_k, A_k, S_k') - \mathbb{E}[F(Q_k, S_k, A_k, S_k') \mid \mathcal{F}_k], v\rangle \mid \mathcal{F}_k] = 0 \quad \text{w.p. } 1$$

and

$$\langle F(Q_k, S_k, A_k, S_k') - \mathbb{E}[F(Q_k, S_k, A_k, S_k') \mid \mathcal{F}_k], v\rangle \le \frac{4\|v\|_1}{1-\gamma} \quad \text{w.p. } 1$$

由条件霍夫丁引理可得

$$\mathbb{E}\left[\exp\left(\lambda\langle F(x_k, Y_k) - \mathbb{E}\left[F(x_k, Y_k)|\mathcal{F}_k\right], v\rangle\right)|\mathcal{F}_k\right] \le \exp\left(\frac{8\lambda^2\|v\|_1^2}{(1-\gamma)^2}\right)$$

对所有 $\lambda > 0$。此外，我们有

$$\mathbb{E}\left[\exp\left(\lambda\|F(x_k,Y_k)-\mathbb{E}\left[F(x_k,Y_k)|\mathcal{F}_k\right]\|_\infty^2\right)\middle|\mathcal{F}_k\right] \le \exp\left(\frac{16\lambda}{(1-\gamma)^2}\right)$$
$$\le \left(1-\frac{32\lambda}{(1-\gamma)^2}\right)^{-1/2},$$

其中最后一行由 $(1-2x)^{1/2} \le e^{-x}$ 对所有 $x \ge 0$成立。因此，假设 2.4 在 $\bar{\sigma} = 4/(1-\gamma)$ 和 $c_d = 1$下成立。

D.4 定理6.2 的证明

我们只需确定常数 $\{\bar{c}_i\}_{1\le i\le 4}$ 对状态-动作空间大小和有效时域的依赖关系。沿用 [17] 最初提出使用广义 Moreau 包络研究任意范数压缩下随机逼近的方法，当 $\|\cdot\|_c = \|\cdot\|_\infty$ 时，我们选择 $\|\cdot\|_s = \|\cdot\|_p$，其中 $p = 2\log(|\mathcal{S}||\mathcal{A}|)$。由此可得 $\ell_{cs} = 1/\sqrt{e}$、$u_{cs} = 1$、$L = p-1 \le 2\log(d)$、$\mu = ((1+\hat{\gamma}_c)/2\hat{\gamma}_c)^2 - 1$，以及 $u_{cM,*} = \sqrt{e}$。

为简化记号，我们用 $a \lesssim b$ 表示存在一个数值常数 C 使得 $a \le C b$。根据 $\{\bar{c}_i\}_{1\le i\le 5}$ 的定义，我们有

$$\bar{D}_0 \lesssim \frac{\mu(1-\gamma)^2}{\log(|\mathcal{S}||\mathcal{A}|)},\ \bar{D}_1 \ge 1-\hat{\gamma}_c,\ \bar{D}_2 \lesssim \frac{\log(|\mathcal{S}||\mathcal{A}|)}{1-\hat{\gamma}_c},\ \bar{D}_3 \lesssim \frac{1}{(1-\gamma)^2},$$
$$\bar{D}_4 \lesssim \frac{\log(|\mathcal{S}||\mathcal{A}|)}{(1-\gamma)^2\mu},$$

进而可得

$$\bar{c}_1 \lesssim \frac{(1+\mu)}{(1-\gamma)^2},\ \bar{c}_2 \lesssim 1,\ \bar{c}_3 \lesssim \frac{\log(|\mathcal{S}||\mathcal{A}|)}{(1-\gamma)^2(1-\hat{\gamma}_c)^3},$$
$$\bar{c}_4 \lesssim \frac{\log(|\mathcal{S}||\mathcal{A}|)}{(1-\gamma)^2(1-\hat{\gamma}_c)^3},\ \bar{c}_5 \lesssim \frac{\log(|\mathcal{S}||\mathcal{A}|)}{(1-\gamma)^2(1-\hat{\gamma}_c)^3}.$$

通过将定理2.4 （1）应用于 Q学习。