arXiv:2302.05430v2 [stat.ML] 19 Mar 2024

分段连续决策的预言机高效平滑在线学习

Adam Block、Alexander Rakhlin 和 Max Simchowitz

MIT

摘要

平滑在线学习已成为一种广受欢迎的框架，用于缓解从经典学习转向对抗式学习时在统计复杂度和计算复杂度方面产生的显著损失。遗憾的是，对于某些空间，研究表明，即使学习者可以访问该空间上的优化预言机，高效算法所遭受的遗憾仍比极小极大最优值差一个指数级数量级。为减轻这种指数依赖关系，本文提出了一种新的复杂度概念——广义括号数。该概念将对手所受的约束与空间的大小结合起来，并证明扰动跟随者的一种具体实现能够达到较低的遗憾，同时其优化预言机调用次数相对于平均遗憾呈最优扩展。随后，我们将这些界应用于若干值得关注的问题，包括分段连续函数的在线预测与规划；这类问题在计量经济学和机器人学等多种领域具有广泛应用。

目录

1 引言 2 2 形式化设定与符号 4 3 跟随扰动领袖与广义括号 6 4 指数扰动与分段连续函数 8 4.1 分段连续预测 .9 4.2 具有广义仿射边界的分段连续预测 . . . . . 10 4.3 具有多项式边界的分段连续预测 .11 **5 平滑多步规划 12**

A 相关工作 19
B 命题 3.1 的证明 19
C 命题 3.2 的证明 21
D 推论 3.1 的证明 22
25

E 定理 1 的证明 **27**E.1 界定稳定性项 .28E.2 完成证明.30**F 关于具有广义仿射边界的分段连续函数的证明** **32**F.1 定理 2 的证明 .32F.2 推论 4.1 的证明 .35F.3 将 ℓ 替换为 ℓ .35

G 第 4.3 节的证明 **36**
G.1 多项式平滑性 .37G.2 定理 3 的证明 .37

H 定理 4 的证明 **40**

1 引言

在线学习已成为研究具有依赖性且可能具有对抗性数据的序贯决策最受欢迎的范式。尽管这一范式因其高度的通用性和极少的假设而颇具吸引力 [Cesa-Bianchi 和 Lugosi， 2006], 对手的最坏情况性质带来了统计和计算方面的挑战 [Rakhlin 等人，2015年，Littlestone，1988，Hazan 和 Koren， 2016]。为缓解这些困难，Rakhlin 等人[2011] 提出了平滑设定，其中对手受到约束，必须从某个分布中采样数据；该分布相对于固定的控制测度，其似然比有一个上界 $1/\sigma$ ，这确保对手无法以高概率选择最坏情况输入。与其他在线学习设定一样，性能通过相对于事后最优比较器的遗憾来衡量 [Cesa-Bianchi 和 Lugosi， 2006]。

近期研究表明，平滑在线学习中存在显著的计算与统计权衡：尽管存在统计上高效的算法，能够实现遗憾 对数级增长 关于 $1/\sigma$，但预言机高效算法必然会遭受随 多项式级增长 的遗憾，其中 $1/\sigma$ [Haghtalab 等人，2022a，b，Block 等人， 2022], 假设学习者可以访问一个经验风险最小化（ERM）预言机，该预言机能够在参数空间上高效优化泛函。这一差距意义重大，因为在许多值得关注的应用中， σ 的自然增长规模是关于问题环境维度的 [Block 和 Simchowitz， 2022]。

一个自然的问题仍待解答：在何种平滑条件下，才能设计出预言机高效算法，使其遗憾随问题维度呈多项式级增长？Block 和 Simchowitz [2022], 给出了部分答案：他们提出了一种基于约翰椭球的高效算法，对于$\log(T/\sigma)\cdot\mathrm{poly}$(维度)遗憾界，该算法适用于无噪声线性分类，以及具有多项式特征的分类问题的一种适当推广。他们还证明，在另一种光滑性条件下—— σ_{dir}方向平滑性——感知机算法能够自动实现次线性于T、且关于 $1/\sigma_{\mathrm{dir}}$呈多项式级增长的遗憾。关键在于， σ_{dir} 对于许多我们关注的分布而言与维度无关，从而避开了先前$\mathrm{poly}(1/\sigma)$遗憾界 [Block 等人，2022，Haghtalab 等人，2022b]中遇到的维度灾难。

在本研究中，我们将预言机效率视为必要的前提，并拓展高效平滑在线学习所能解决的问题范围。一个值得牢记的核心例子是

分段仿射函数（PWA），其中，PWA 函数由欧几里得空间中的有限个区域定义，在每个区域内，该函数都是仿射的。这类模型自然出现在统计学和计量经济学中常见的分段回归应用里 [Feder，1975年，Bai 和 Perron，1998年，Yamamoto 和 Perron， 2013]，以及控制系统中的常用模型 [Borrelli，2003年，Henzinger 和 Sastry， 1998]。

遗憾的是，由于跨越不同区域时会出现不连续性，分段仿射回归器并非在对抗性设定下可学习，即使计算时间不受限制也是如此，因为它们包含线性阈值类，而众所周知，该类不具备在线可学习性[Littlestone， 1988]；然而，在这种设定下，平滑性假设很自然，因为实证研究者已经会注入噪声 [Posa 等人，2014，Suh 等人，2022a]。遗憾的是，注入噪声的性质决定了平滑性参数 σ 会像上文所述那样，随上下文空间的维度呈指数级增长，因此先前的保证不足以满足实际应用需求。于是，我们面临的问题是：如何设计在平滑在线学习设定下可证明具备（预言机）效率的实用算法。

下面，我们将提出一种基于经典括号数的复杂度度量 [Blum，1955年，Giné 和 Nickl， 2021] 若该度量有界，便可得到一种遗憾可证明较小的实用算法。具体而言，我们将考虑对著名的 Follow-the-Perturbed-Leader（FTPL）算法的实例化 [Kalai 和 Vempala， 2005]，在每个时刻 $1 \le t \le T$，我们对 Θ 上的随机路径 $\omega_t(\theta)$进行采样，并选择 $\theta_t \in \arg\min_\theta L_{t-1}(\theta) + \omega_t(\theta)$，其中 $L_{t-1}(\theta)$表示截至时刻 $t-1$的累积损失。FTPL 的标准分析 [Agarwal 等人，2019b，Suggala 和 Netrapalli，2020，Haghtalab 等人，2022a，Block 等人， 2022] 要求损失函数关于参数 θ满足利普希茨条件，而这显然不适用于分段仿射函数这一核心示例。不过，我们表明，光滑性能够保证许多损失函数在期望意义下满足利普希茨条件，至多相差一个加性常数；该常数取决于由我们提出的括号数推广形式所度量的类的复杂度。基于这一事实，我们给出一个用于证明 FTPL 惰性实例化的遗憾保证的框架。

尽管上述理论本身可能具有技术上的价值，我们仍通过若干示例具体展示了所得结果。我们以此前介绍的相关概念——方向平滑性——取代标准的光滑性概念 [Block 和 Simchowitz， 2022]。我们改编了Agarwal 等人 [2019b],Suggala 和 Netrapalli [2020] 关于采用指数分布扰动的 FTPL 的结果，并提出了一种实用且有可证明低遗憾保证的算法，适用于具有广义仿射边界的分段连续损失函数。随后，我们将这一结果推广至具有多项式边界的损失函数，并假设对手受到更严格的约束；最后，我们在一个受机器人规划启发的场景中具体展示了所得结果。具体而言：

- 在第 3 节中，我们引入一种衡量类大小的新指标——广义括号数。它结合了对手的假设与空间复杂度，因此在许多感兴趣的情形下都可能较小。我们利用广义括号数证明了命题3.2：如果对手受到适当约束，且相对于某个特定伪度量的广义括号数受到控制，那么惰性版本的 FTPL 所产生的遗憾较低。在此过程中，我们还在命题3.1中表明，对广义括号数的控制能够导出一个同时对参数和对手一致成立的集中不等式。

- 在定理 1中，我们将第 3 节所建立的一般理论应用于有限维 Θ这一特殊情形。具体而言，通过改编Agarwal 等人 [2019b]，Suggala 和

Netrapalli [2020], 我们证明，如果 Θ 的广义括号数受到 Θ 控制，那么算法2 可以实现至多为 ϵ 的平均遗憾，同时$\widetilde{\mathcal{O}}\left(\epsilon^{-2}\right)$ 对经验风险最小化预言机的调用次数达到最优。

- 在定理 2和推论4.1中，我们考虑一个更具体的设定，其中损失函数是分段连续的，且具有仿射边界。特别地，我们证明，如果对手是σ_{dir}-方向光滑的，那么算法2 可达到平均遗憾 ϵ ，且仅需$\widetilde{\mathcal{O}}\left(\sigma_{\text{dir}}^{-1}\epsilon^{-2}\right)$ 次经验风险最小化预言机调用，从而消除应用Block 等人 [2022] 所带来的维度指数依赖，并达到关于 ϵ的最优依赖。

- 在定理 3中，我们推广推论4.1 的结果，并证明，若对手进一步受到多项式光滑约束（见定义4.1），且损失函数是分段连续的，其边界由次数至多为 r的多项式定义，那么算法2 可达到平均遗憾 ϵ ，且至多需要 $\widetilde{\mathcal{O}}\left(\epsilon^{-2r}\right)$ 次经验风险最小化预言机调用。

- 在第 5 节中，我们考虑一类分段 Lipschitz“混合”动力系统 [Hen- zinger 和 Sastry， 1998]，其中区域内的边界是线性的或多项式的。这些系统可以对机器人学中常见的多种动力系统进行建模，尤其是分段仿射系统 [Borrelli， 2003年、Marcucci 和 Tedrake， 2019]以及分段多项式系统 [Posa 等人， 2015]。我们在定理 4 中证明，通过对输入和动力学进行平滑处理，我们提出的 FTPL 算法在在线规划场景下能够达到低遗憾。据我们所知，这是首个用于具有不连续性的混合系统规划的低遗憾算法。

我们首先对问题进行形式化设定，并介绍若干必要的基本概念，随后陈述并讨论我们的结果。出于篇幅考虑，对相关工作的扩展讨论留待附录 A 。

2 形式化设定与符号

形式化地，我们考虑受约束对手下的在线学习问题。给定某个决策空间 Θ 和上下文空间 $\mathcal{Z}$，以及损失函数 $\ell\colon \Theta\times\mathcal{Z}\to[0,1]$，在线学习按轮次 $1\le t\le T$进行。在每个时刻 t，对手选择某个 $z_t\in\mathcal{Z}$，学习者选择某个 $\theta_t\in\Theta$ 并承受损失 $\ell(\theta_t,\ z_t)$，目标是相对于事后最优的 $\theta\in\Theta$ 将遗憾降至最低，即 $\mathbb{E}\left[\mathrm{Reg}_T\right]=\mathbb{E}\left[\sum_{t=1}^{T}\ell(\theta_t,\ z_t)-\inf_{\theta\in\Theta}\sum_{t=1}^{T}\ell(\theta,\ z_t)\right]$。为衡量预言机复杂度，我们将尤其关注归一化遗憾 $T^{-1}\mathrm{Reg}_T$。在实际应用中，我们常会考虑在线监督学习这一特殊情形，其中$\mathcal{Z}=\mathcal{X}\times\mathcal{Y}$和 $z=(x,\ y)$由上下文 x 和标签 y组成；在这种情况下，我们区分适当 学习（学习者在看到 x_t之前选择 θ_t ）和非适当 学习（学习者能够根据已揭示的 x_t选择 θ_t ）。

由于完全对抗式在线学习在统计和计算方面面临挑战 [Rakhlin 等人、2015年、Hazan 和 Koren， 2016]，我们将限制对手，使其选择 $z_t\sim p_t$，其中$p_t\in\mathcal{M}\subset\Delta(\mathcal{Z})$ 是定义在 $\mathcal{Z}$ 上的分布，且可能依赖于截至时间 t 的历史，并且 $\mathcal{M}$ 是某个受限分布类。在本文中，我们主要关注这样一种情形： $\mathcal{M}$由某种意义上的光滑分布构成：

定义 2.1。给定空间 $\mathcal{X}$、测度 $\mu \in \Delta(\mathcal{X})$ 以及某个 $\sigma < 0$，若测度 p_t相对于 μ 是 σ-光滑的，我们称之为 p_t；这意味着相对于 μ 的似然比由 σ^{-1}一致有界，即 $\left\| \frac{dp_t}{d\mu} \right\|_\infty \leq \frac{1}{\sigma}$。若对于某个 d有 $\mathcal{Z} \subset \overline{\mathbb{R}^d}$，则称 p_t是 σ_{dir}-方向光滑的；也就是说，对于任意单位向量 $\mathbf{w} \in \mathcal{S}^{d-1}$，分布 $\langle \mathbf{w}, \mathbf{x} \rangle$ 相对于实数轴上的勒贝格测度是 σ_{dir}-光滑的，其中 $\mathbf{x} \sim p_t$。

如相关工作部分将进一步讨论的那样，光滑性最近已成为平滑在线学习中一项广受欢迎的假设。方向平滑性由 Block 和 Sim- chowitz [2022] 提出，并被 Block 等人 [2023], 采用，为缓解一些常用系统中标准光滑性对维度的依赖提供了一种自然的方法。

我们的算法将采用经验风险最小化（ERM）预言机这一计算原语：

定义 2.2. 给定空间 Θ 和泛函 $\ell_i : \Theta \to \mathbb{R}$ （其中 $1 \leq i \leq m$），将经验风险最小化（ERM）预言机定义为任何能够在 Θ 上进行优化的预言机，即 $\theta = \text{ERMOracle}(\sum_{i=1}^m \ell_i(\theta))$（当 $\theta \in \arg\min_{\theta \in \Theta} \sum_{i=1}^m \ell_i(\theta)$）时）。

定义2.2是计算高效在线学习研究中的一项常见假设[Hazan 和 Koren，2016年，Block 等人，2022，Haghtalab 等人，2022a], 许多适用于热门函数类的启发式方法均可用于实际应用 [LeCun 等人，2015年，Garulli 等人， 2012]。在下文中，我们始终假设学习者可以访问 ERM 预言机，并以算法的计算复杂度来衡量其预言机调用次数ERMOracle。我们尤其关注实现平均遗憾 ϵ 所需的预言机复杂度，即足以确保 $T^{-1} \cdot \mathbb{E}[\text{Reg}_T] \leq \epsilon$的预言机调用次数。虽然为使正文表述简洁，我们假设 ERMOracle 是精确的，但我们在附录中给出了只需近似预言机的陈述和证明，其误差可能依赖于扰动，并会以加法项的形式计入最终的遗憾保证。

在下一节中，我们将介绍一种新的复杂度概念，即空间 Θ 的广义括号数。这里，我们将回顾经典的括区熵概念，以便进行比较，并为后续讨论我们的一个结果作参考：

定义 2.3 （摘自Giné 和 Nickl [2021] 第 3.5.2 节）。对于函数类 $\mathcal{F} : \mathcal{Z} \to \mathbb{R}$ 和测度 $\mu \in \Delta(\mathcal{Z})$, 如果对于所有 $\mathcal{B}_i$，都有 $\mathbb{E}_\nu \left[\sup_{f,\ g \in \mathcal{B}_i} |f(z) - g(z)| \right] \leq \epsilon$，我们就称 $\mathcal{N} = \{\mathcal{B}_i\}$ 是 $\mathcal{F}$ 的一个关于 μ 的 ϵ-括号。括区数 $\mathcal{N}_{[]}(\mathcal{F}, \mu, \epsilon)$ 是此类划分的最小规模。

对函数类的括区数进行控制，经典上可得出一致大数定律和一致中心极限定理；许多常见函数类的这类数值性质良好。更多详情请参见 [Giné 和 Nickl， 2021]。

记号 下文中，我们用 z 表示上下文，用 θ 表示参数。我们始终用 T表示时域，用 ℓ表示损失函数，并将向量记为粗体。我们将使用 $\mathcal{O}(\cdot)$ 记号来省略通用常数，并使用 $\mathcal{O}(\cdot)$ 来省略多对数因子。我们用 $||\cdot||_1$ 表示欧几里得空间中的 ℓ_1 范数，用不加修饰的 $||\cdot||$ 表示欧几里得范数。

3 跟随扰动领导者与广义括号

本节中，我们提出算法，并定义确保期望遗憾较低的复杂度参数。下一节将给出示例。我们将考虑扰动跟随者（FTPL）类算法的一种实例化 [Kalai 和 Vempala，2005]，其中，在每个时刻 $1 \le t \le T$，我们构造一个样本路径 $\omega_t(\theta)$，该路径是在 t 上从某个随机过程独立同分布地抽取的，并选择 Θ

$$\theta_t = \arg\min_{\theta \in \Theta} L_{t-1}(\theta) + \omega_t(\theta), \tag{3.1}$$

其中 $L_{t-1}(\theta) = \sum_{s=1}^{t-1} \ell(\theta, z_s)$。FTPL 的经典分析采用所谓的“成为领导者”（Be-The-Leader）引理 [Kalai 和 Vempala，2005年，引理 3。1]将遗憾分解为扰动的大小和预测的稳定性。也就是说，如果学习者按照 θ_t （3.1）进行决策，则遗憾的上界如下：

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le 2 \cdot \mathbb{E}\left[\sup_{\theta \in \Theta} \omega_1(\theta)\right] + \sum_{t=1}^{T} \mathbb{E}\left[\ell(\theta_t, z_t) - \ell(\theta_{t+1}, z_t)\right]. \tag{3.2}$$

通常，分析 FTPL 所产生遗憾时的难点在于界定第二项（3.2）——稳定性项。此类分析中常见的一个假设是损失 ℓ 关于 θ [Agarwal 等人，2019年b，Suggala 和 Netrapalli，2020年，Block 等人，2022]是利普希茨连续的；遗憾的是，对于许多受关注的类，这一假设并不成立。

为说明我们方法的动机，考虑学习线性阈值这一简单情形，其中 $\theta \in [0,1]$ 和 $\ell(\theta, z) = \mathbb{I}[y \neq \mathrm{sign}(x - \theta)]$，对于 $z = (x, y) \in \mathcal{Z} = [0,1] \times \{\pm 1\}$。在这种情况下，显然$\theta \mapsto \ell(\theta, z)$不是利普希茨连续的（甚至不连续），因此 Agarwal 等人 [2019b]，Suggala 和 Netrapalli [2020]的结果并不适用；不过，简单计算可知，如果对手关于勒贝格测度是 σ-光滑的，那么 $\theta \mapsto \mathbb{E}_z[\ell(\theta, z)]$ 是利普希茨连续的。直观地，我们或许会希望稳定性项 $\mathbb{E}[\ell(\theta_t, z_t) - \ell(\theta_{t+1}, z_t)]$ 能够由 $|\theta_t - \theta_{t+1}|$ 控制，并据此应用与 Agarwal 等人 [2019b]，Suggala 和 Netrapalli [2020] 中类似的论证。但这一想法行不通，因为尽管对于任意固定的 $\theta \in \Theta$，在给定历史的条件下，z_t 的光滑性意味着 $\mathbb{E}[\ell(\theta_t, z_t) - \ell(\theta, z_t)] \lesssim |\theta_t - \theta|$，但实际上 θ_{t+1} 取决于 z_t，因此不能保证在给定 θ_{t+1} 的条件下 z_t 的分布必然是光滑的。不过，我们不会完全放弃这种方法；相反，我们将证明，如果函数类 $\theta \mapsto \ell(\theta, z)$ 关于某种特定的复杂度度量足够小，那么类似的论证仍然成立。为准确说明这一点，请看下面的定义：

定义 3.1。设 $\mathcal{M}$ 是某个空间 $\mathcal{Z}$ 上的一个分布类，并假设 $\rho: \Theta \times \Theta \times \mathcal{Z} \to \mathbb{R}$ 是空间 Θ 上由 $\mathcal{Z}$ 中的元素参数化的伪度量。若集合 $\{(\theta_i, \mathcal{B}_i)\} \subset \Theta \times 2^{\Theta}$ 满足 $\Theta \subset \bigcup_i \mathcal{B}_i$，且对所有 i 都有

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_{z \sim \nu}\left[\sup_{\theta \in \mathcal{B}_i} \rho(\theta, \theta_i, z)\right] \le \epsilon.$$

我们用 $\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon)$ 表示广义 ϵ-括号的最小规模。

请注意，定义 3.1 与定义 2.3中的经典概念相似：广义括号要求给定划分 $\mathcal{B}_i$ 的期望直径 在某个类中的测度上 $\mathcal{M}$一致地较小；事实上，如果 $\mathcal{M}$ 是单元素集合，我们就得到了经典概念。与其他复杂度概念（例如标准覆盖数）相比，广义ϵ括号的优点如下：

算法 1 惰性 FTPL

初始化 经验风险最小化预言机 ERMOracle，周期长度 n，扰动分布 Ω
对于 $\tau = 1, 2, \ldots, T/n$ **执行**
  采样 $\omega_\tau : \Theta \to \mathbb{R}$ 从 Ω (% 采样扰动)
  $\widetilde{\theta}_\tau \leftarrow \text{ERMOracle}\left(\widetilde{L}_\tau(\theta) + \omega_\tau(\theta)\right)$ (% 在扰动后的损失上调用 ERMOracle)
  对于 $t = (\tau-1)n+1, \ldots, \tau n$ **执行**
    观察 z_t，**预测** $\widetilde{\theta}_\tau$，**接收** $\ell(\widetilde{\theta}_\tau, z_t)$

命题 3.1. 设 $\mathcal{M}$ 和 ρ 如定义 3.1 所述，并假设 $z_1, \ldots, z_n$ 的生成方式满足：z_i 在由截至时间 i 的 z_j 生成的 σ-代数 $\mathcal{F}_i$ 条件下的分布 p_i 满足 $p_i \in \mathcal{M}$，对所有 $1 \le i \le n$ 均成立。进一步假设，对所有 $z \in \mathcal{Z}$，均有 $\sup_{\theta,\, \theta' \in \Theta} \rho(\theta, \theta', z) \le D$。那么对于任意 $\epsilon, \delta > 0$，以至少 $1-\delta$ 的概率，对所有 $\theta, \theta' \in \Theta$，以下结论同时成立：

$$\left|\sum_{i=1}^{n} \rho(\theta, \theta', z_i)\right| \le 4n \cdot \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', z)\right] + 8\epsilon n + 6D^2 \log\left(\frac{2\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon)}{\delta}\right). \tag{3.3}$$

命题3.1 的证明见附录 B，证明通过应用 Freedman 不等式，并用上确界之和控制和的上确界。尽管这个界看起来相当宽松，但令人有些意外的是，我们仍能在多种设置下获得优于预期的 $\widetilde{\mathcal{O}}\left(\epsilon^{-2}\right)$ 预言机复杂度保证。

关键在于，由于 (3.3) 对 $\theta' \in \Theta$ 一致成立，我们可以将命题 3.1 应用于 $\theta' = \theta_{t+1}$，从而避开因在以 θ_{t+1} 为条件时 z_t 不光滑而带来的难题。在介绍算法之前，还有两个问题需要解决。首先，由于 (3.3) 中存在加性统计误差，如果 n 太小，命题 3.1 就会变得空洞。为缓解这一问题，我们将按训练轮次运行 FTPL。对于某个固定的 $n \in \mathbb{N}$，以及所有 $\tau \ge 1$，令 $L_\tau(\theta) = L_{\tau n}(\theta)$，并定义 $\mathcal{I}_\tau = \{i \mid (\tau-1)n + 1 \le i \le \tau n\}$ 以及 $\ell_\tau(\theta) = \sum_{t \in \mathcal{I}_\tau} \ell(\theta, z_t)$。我们将运行一个惰性版本的 FTPL，在每个训练轮次开始时更新 $\theta_t = \theta_\tau$，并让 $\theta_t = \theta_\tau$ 保持不变，直到训练轮次发生变化。采用惰性更新可以使我们应用命题 3.1 时，(3.3) 中的第一项占主导地位。完整算法总结见算法 1。

第二个挑战是将（3.2）中的稳定性项与伪度量联系起来 ρ 在连续的 θ_τ 上的取值。因此，我们要求损失函数满足以下结构条件：

定义 3.2。 设 Θ 是某个赋范空间的子集，该空间配备了范数 $||\cdot||$。如果对于分布类 $\mathcal{M}$ 和范数 $||\cdot||$，伪度量 $\rho : \Theta \times \Theta \times \mathcal{Z} \to \mathbb{R}$ 满足具有参数（α，β）的伪等距性质，即对所有 θ，$\theta' \in \Theta$，均有

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_{z \sim \nu}\left[\rho(\theta, \theta', z)\right] \le \alpha \cdot \left|\left|\theta - \theta'\right|\right|^\beta.$$

现在，我们已经可以陈述第一个结果，该结果给出了算法1的一个实例的遗憾上界：

命题 3.2. 假设我们处于受约束在线学习设定中，对手受到约束，只能从类 $\mathcal{M}$ 中的某个分布采样 z_t。进一步假设，在 Θ 上存在一个由 $\mathcal{Z}$ 参数化的伪度量 ρ，满足定义

3.2 中的伪等距性质，并且对于所有 θ、$\theta' \in \Theta$，都有 $\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu [\ell(\theta, z) - \ell(\theta', z)] \leq \sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu [\rho(\theta, \theta', z)]$。如果学习者采用算法 1，且 $\sup_{\theta,\theta'\in\Theta} \rho(\theta, \theta', z) \leq D$，那么对于任意 $\epsilon > 0$，期望遗憾的上界为：

$$\mathcal{O}\left(\mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] + \epsilon T + \frac{TD^2}{n} \cdot \log\left(T \cdot \mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon)\right) + 2n\alpha \cdot \sum_{\tau=1}^{T/n} \mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right|\right|^\beta\right]\right).$$

我们在附录 C中给出完整证明。首先，我们证明 Kalai 和 Vempala [2005]提出的 Be-the-Leader 引理的一个变体，该变体允许进行惰性更新，然后应用命题 3.1 以及伪等距性质，通过学习者的预测相对于相关范数的稳定性，控制惰性更新相对于评估损失函数的稳定性项。将上述各部分结合起来即可完成证明。我们注意到，按照目前的表述，似乎将 n 设置得尽可能大并无不利之处；事实上，最终求和中的 n 依赖关系似乎会相互抵消，而增大 n 会减小第三项。然而，不出所料，事实并非如此，因为增大 n 会降低学习者预测的稳定性，从而隐式地增大最后一项，这一点在该结果的应用中显而易见。

命题 3.2 为证明算法 1的不同实例的遗憾界提供了一个模板。具体而言，对于给定的损失函数 $\ell(\cdot, \cdot)$，只需找到一个伪度量ρ、范数 $||\cdot||$和噪声分布Ω，使得（a） ρ 相对于范数 $||\cdot||$是伪等距的；（b） Θ 相对于 ρ的广义括号数较小；以及（c）扰动使惰性更新具有稳定性，即 $\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right|\right|\right]$ 较小。作为下一节结果的简单预热，我们展示：当相关函数类具有经典意义下较小的括区熵时，可以通过一种明显更简单的证明，恢复 Block 等人 [2022], 关于带高斯过程扰动的适当平滑在线学习的预言机复杂度上界的一个较弱版本。

在这个启发性示例中，我们假设 $\Theta = \mathcal{F}$ 表示一个函数类，并且我们处于~在线监督学习 $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$设定中，即 $\ell(\theta, z) = \ell(f(x)\ y)$。我们进一步假设，对手相对于已知的基准测度 μ 是 σ-平滑的（回顾定义 2.1）。与 Block 等人 [2022,定理 10], 中一样，我们考虑高斯过程扰动，其中独立抽取x_1，……，$x_m \sim \mu$，即 γ_1，……，γ_m标准高斯随机变量，并令 $\omega(f) = \eta \cdot \sum_{i=1}^m \gamma_i f(x_i)$。

推论 3.1。 假设我们处于平滑在线学习设定中，且函数类为~$\mathcal{F}$： $\mathcal{X} \to \{\pm 1\}$ 并且 ℓ 取值在单位区间内，且对于第二个参数的所有取值，它关于第一个参数都是利普希茨连续的。如果学习者采用算法 1 以及上述高斯扰动，并选择正确的超参数（见 附录 D），那么学习者可以达到平均遗憾 ϵ，其中 $\widetilde{\mathcal{O}}\left(\frac{\epsilon^{-4}L^{3/5}}{\sigma^{2/5}} \cdot \text{对数}^{3/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)\right)$次调用经验风险最小化预言机。

请注意，预言机复杂度保证弱于Block 等人 [2022]；我们将推论 3.1及其证明收录于附录 D，仅作为对我们所使用技术及其与函数类复杂度这一更经典概念之间关系的简单演示。现在，我们继续讨论一些示例，在这些基本设置中，我们的方法能够提供全新的遗憾界。

4 指数　　性扰动与分段连续函数　　s

在上一节中，我们观察到命题3.2 为证明 FTPL 不同实例的遗憾界提供了一个模板，并运用这一技巧重新得到了早先关于

算法 2 带指数噪声的惰性 FTPL

初始化 经验风险最小化预言机 ERMOracle，周期长度 n，扰动幅度 η
对于 $\tau = 1, 2, \ldots, T/n$ **执行**
  采样 $\xi = (\xi_1, \ldots, \xi_d) \overset{iid}{\sim}$ 参数为 1 的指数分布 (% 采样扰动)
  $\widetilde{\theta}_\tau \leftarrow \text{ERMOracle}\left(\widetilde{L}_\tau(\theta) - \eta \langle \xi, \theta \rangle\right)$ (% 在扰动后的损失上调用 ERMOracle)
  对于 $t = (\tau - 1)n + 1, \ldots, \tau n$ **执行**
    观察 z_t，**预测** $\widetilde{\theta}_\tau$，**接收** $\ell(\widetilde{\theta}_\tau, z_t)$

平滑在线学习。本节针对一个重要设定——分段连续函数——给出新的结果。在正式定义分段连续函数之前，我们考虑集合 $\Theta \subset \mathbb{R}^d$ 对于某个维度 d的一般情形。命题3.2 所提供的模板要求我们指定一个扰动分布；此前我们使用的是高斯过程，而在这里我们采用Agarwal 等人 [2019b], Suggala 和 Netrapalli [2020] 的方法，并使用指数扰动。如算法 2所概述的，我们保留算法1 中的惰性更新，但指定 $\omega(\theta) = -\eta \cdot \langle \xi, \theta \rangle$ 其中某个尺度参数为 $\eta > 0$，并且 $\xi = (\xi_1, \ldots, \xi_d)$ 中的 $\xi_i \sim$ 参数为 1 的指数分布彼此独立。利用指数扰动，我们得到以下遗憾界：

定理 *1.* 假设我们处于命题 *3.2* 的受约束在线学习设定中，其中 $\Theta \subset \mathbb{R}^d$ 满足 $\sup_{\theta, \theta' \in \Theta} ||\theta - \theta'||_1 = D < \infty$。进一步假设，$\mathcal{Z}$-参数化的伪度量 ρ 在 $\mathbb{R}^d$ 上相对于 ℓ_1 满足定义 *3.2* 的伪等距性质，并且 $\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu [\ell(\theta, z) - \ell(\theta', z)] \leq \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu [\rho(\theta, \theta', z)]$。如果学习者执行算法 *2*，且 $\eta = \Omega(n^2)$，则期望遗憾有界：

$$\mathbb{E}[\mathrm{Reg}_T] \leq \mathcal{O}\left(\eta + \frac{T}{n} \cdot \log\left(\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, 1/T)\right) + T\alpha \left(\frac{\log \mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, 1/T)}{\eta}\right)^{\frac{\beta}{4-2\beta}}\right).$$

调整 η和 n后，遗憾按 $\widetilde{\mathcal{O}}\left(T^{\frac{4-2\beta}{4-\beta}}\right)$ 的规模增长，需调用优化预言机 $\widetilde{\mathcal{O}}\left(T^{\frac{2}{4-\beta}}\right)$ 次，因此调用 ERMOracle $\widetilde{\mathcal{O}}\left(\epsilon^{-2/\beta}\right)$ 次即可达到 ϵ的平均遗憾。特别是，在最佳情况下，当 $\beta = 1$时，我们恢复了达到平均遗憾不超过 ϵ 所需的最优预言机复杂度 $\widetilde{\mathcal{O}}\left(\epsilon^{-2}\right)$；若每轮调用一次预言机并将遗憾控制在 $\widetilde{\mathcal{O}}\left(\sqrt{T}\right)$，也会得到这一结果。 —

定理 定理 1 的完整证明可见于附录 E，这里我们简要概述。虽然我们遵循命题3.2的一般框架，但为获得略有改进的速率，并未直接应用该结果。与更一般命题的证明一样，我们借助 Be-the-Leader 引理，将分析简化为界定学习者的预测相对于评估损失函数的稳定性。随后，我们运用Agarwal 等人 [2019b],Suggala 和 Netrapalli [2020] 的技术 $||\cdot||_1$ 在 $\mathbb{R}^d$中证明，若稳定性项较小，则学习者的预测相对于 $||\cdot||_1$ 在 $\mathbb{R}^d$是稳定的。最后，我们结合伪等距假设和对广义括号数的控制，并利用命题 3.1 通过自界定论证得出结论。

4.1 分段连续预测

现在，我们将前述结果应用于若干值得关注的问题。在本节余下部分，我们将展示，边界性质良好的分段连续函数能够同时具有较小的

括区间数，并且在对手仅满足方向光滑性的假设下，对于适当选取的 ρ，还具有伪等距性质。形式上，我们假设 $\Theta = \Theta_{\mathrm{c}} \times \Theta_{\mathrm{d}}$ 可以分解为连续部分和离散部分，且对于某个维度 m，有 $\Theta \subset \mathbb{R}^m$。我们按如下方式构造函数 g。首先，考虑类 g_k：$\Theta_{\mathrm{c}} \times \mathcal{Z} \to \mathbb{R}$，其中 $1 \le k \le K$ 满足：对于所有 $z \in \mathcal{Z}$，$g_k(\cdot, z)$ 作为 θ_{c} 的函数，相对于 Θ 上的 ℓ_1 范数是利普希茨的。现在，对于固定的 ϕ：$\Theta_{\mathrm{d}} \times [K] \times \mathcal{Z} \to \mathbb{R}$，我们定义

$$k_\phi(\theta_{\mathrm{d}}, z) = \arg\max_{k \in [K]} \phi(\theta_{\mathrm{d}}, k, z), \qquad \ell(\theta, z) = g_{k_\phi(\theta_{\mathrm{d}}, z)}(\theta_{\mathrm{c}}, z). \tag{4.1}$$

(4.1) 的表述兼具灵活性与简洁性，但另一种相关构造更易于分析：令 ϕ：$\Theta_{\mathrm{d}} \times [K]^{\times 2} \times \mathcal{Z} \to \mathbb{R}$ 满足 $\phi(\theta_{\mathrm{d}}, k, k', z) = -\phi(\theta_{\mathrm{d}}, k', k, z)$对所有 $\theta_{\mathrm{d}} \in \Theta_{\mathrm{d}}$，$k, k' \in [K]$以及 $z \in \mathcal{Z}$均成立。此外，令

$$\overline{k}_\phi(\theta_{\mathrm{d}}, z) = \arg\max_{k \in [K]} \sum_{k' \neq k} \mathbb{I}\left[\overline{\phi}(\theta_{\mathrm{d}}, k, k', z) \ge \phi(\theta_{\mathrm{d}}, k', k, z)\right],$$

若出现平局，则按字典序打破平局，即，$\overline{k}_\phi$ 是赢得锦标赛中最多场比赛的最小索引 k，其中胜负由 $\phi(\theta_{\mathrm{d}}, k, k', z)$的符号决定。接下来定义

$$\overline{\ell}(\theta, z) = g_{\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, z)}(\theta_{\mathrm{c}}, z). \tag{4.2}$$

为简洁起见，本节将重点讨论(4.2)的锦标赛表述。在附录 F.1中，我们将把结果扩展到带有额外间隔假设的(4.1)情形。我们还指出，(4.2)可视为(4.1)中自然函数类的一种非适当松弛，因此足以用于非适当在线学习[1]。最后需要指出，尽管为简洁起见我们描述的是锦标赛式聚合系统，但正如证明所示，对这些 $\binom{K}{2}$ 事件 $\phi(\theta_{\mathrm{d}}, k, k', z) \ge 0$ 进行任意聚合也会得到类似结论，且适用范围更广。例如，这种推广使我们能够高效地表示面数与其成正比的多面体区域。K 其与面的数量成正比。

4.2 具有广义仿射边界的分段连续预测

我们首先研究仿射决策边界这一重要的特殊情形。并注意到，式（4.1）所描述的设定涵盖了分段仿射函数这一核心示例：令 $\Theta_{\mathrm{c}} = \left(\mathbb{R}^{m \times d}\right)^{\times K}$, $\Theta_{\mathrm{d}} = (\mathbb{R}^{d+1})^{\times K}$, $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}^m$，并令 $\phi(\theta_{\mathrm{d}}, k, z) = \langle \mathbf{w}_k, (\mathbf{x}, 1) \rangle$，即可取

$$\ell(\theta, z) = ||\mathbf{y} - \mathbf{W}_{k^\star}\mathbf{x}||^2, \qquad k^\star = \arg\max_{k \in [K]} \langle \mathbf{w}_k, (\mathbf{x}, 1) \rangle,$$

其中我们在末尾额外添加一个1 维坐标，以考虑可能存在的仿射常数。我们证明，若 ℓ 如（4.2）所示具有仿射边界且分段连续，则广义括号数较小，并且相对于 ℓ_1 范数满足伪等距，只要对手是 σ_{dir}方向平滑的。

[1]参见Block 等人 [2022]，了解适当在线学习与非适当在线学习之间差异的讨论。

定理 2. 假设 $\mathcal{Z} \subset \mathbb{R}^d$ 且 Θ 是欧几里得空间的子集，其 ℓ_1 直径有界于 D，且 $\Theta_{\mathrm{d}} \subset (\mathcal{S}^d)^{\times\binom{K}{2}}$；令 $w_{kk'}$ 表示给定 $\theta_{\mathrm{d}} \in \Theta_{\mathrm{d}}$ 的坐标。进一步假设 $\phi(\theta_{\mathrm{d}}, k, k', \mathbf{z}) = \psi(\langle w_{kk'}, (\mathbf{z}, 1)\rangle)$，其中存在某个可微的奇链接函数 $\psi\colon \mathbb{R} \to \mathbb{R}$，满足 $a \le |\psi'(x)| \le A$ 对所有 x 均成立，并令 $\mathcal{M}$ 由满足 $\|\mathbf{z}\|_\infty \le B$ 的 σ_{dir}-方向平滑分布类构成。令

$$\rho(\theta, \theta', \mathbf{z}) = 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right] + \max_{1 \le k \le K} \left\|\theta_{\mathrm{c}}^{(k)} - \theta_{\mathrm{c}}^{\prime(k)}\right\|_1 . \tag{4.3}$$

那么，ρ 是一个满足伪等距性质的伪度量，其中 $\alpha = \frac{2A}{(B \vee 1)\, a\sigma_{\mathrm{dir}}}$ 和 $\beta = 1$。此外，对于所有 $\epsilon > 0$、$\mathcal{N}_{\mathcal{M},\,[\,]}(\Theta, \rho, \epsilon) \le \left(\frac{9AK^2BD}{a\sigma_{\mathrm{dir}}\epsilon}\right)^{K^2(d+1)}$。

我们在附录 F.1 中详细证明定理 2。这两个结论的证明都依赖于同一个关键步骤，即引理 F.1。该引理表明，对于固定的 θ_{d}^0，尽管事件 $\mathbb{I}\left[k_\phi(\theta_{\mathrm{d}}, \mathbf{z}) \neq k_\phi(\theta_{\mathrm{d}}^0)\right]$ 不是关于 θ_{d} 的连续函数，但如果 $\mathbf{z}$ 是 σ_{dir}-方向光滑的，其期望就是利普希茨函数。因此，引理 F.1 极大地推广了第 3 节中涉及一维阈值的启发性论证。该关键引理通过利用仿射函数作用于方向光滑随机变量时的反集中性质得以证明；这也是 ℓ 的分析与式 (4.1) 中函数 ℓ 的分析唯一不同之处。随后，我们利用这一结果推出伪等距性质，并证明，关于 $\overline{\ell}_1$ 覆盖 Θ，可得到关于 $\mathcal{M}$ 和 ρ 的广义 ϵ-括号。

根据 定理 2，我们可以为算法 2 在具有仿射边界的分段连续函数类上证明一个具体的遗憾界：

推论 4.1. 设 ℓ 如 (4.2) 所示，其中 ϕ 和 Θ 如定理 2 所示，且 $B \ge 1$ 和 ℓ 的绝对值均一致有界于 1。若我们设定 $\eta = \widetilde{\Theta}\left(\left(TK^2dDBA(a\sigma_{\mathrm{dir}})^{-1}\right)^{2/3}\right)$ 和 $n = \sqrt{\eta}$，那么算法 2 的遗憾为 $\mathbb{E}[\mathrm{Reg}_T] \le \widetilde{\mathcal{O}}\left(\left(TAK^2dBD(a\sigma_{\mathrm{dir}})^{-1}\right)^{2/3}\right)$。特别地，要达到平均遗憾 ϵ，只需调用 ERMOracle $\widetilde{\mathcal{O}}\left(\frac{AK^2dDB}{a\sigma_{\mathrm{dir}}\epsilon^2}\right)$ 次。

推论 4.1 的证明见 附录 F.2，几乎可以直接由定理 1 和 2推出。链接函数最简单的例子是令 $\psi(x) = x$ 为恒等函数，此时我们可以得到具有仿射边界的分段连续函数的遗憾界。

4.3 具有多项式边界的分段连续预测

为了拓展应用范围，我们现在考虑更一般的区域间边界。如上所述，证明定理 2的类似结论的关键，在于将仿射函数应用于方向光滑随机变量时的反集中性质。尽管对更一般函数的反集中性质仍是一个活跃的研究领域，但多项式的某些子类（例如独立变量的多重线性函数）已知在非常一般的情况下具有反集中性质 [Mossel 等人， 2010]并且足以利用我们的技术，将结果扩展到具有这些决策边界的损失函数；不过，我们转而关注一般的多项式边界，并进一步限制 $\mathcal{M}$:

定义 4.1. 对于多项式 $f\colon \mathbb{R}^d \to \mathbb{R}$ 满足 $f(x) = \sum_{\mathcal{I} \subset [n]} \alpha_{\mathcal{I}} x^{\mathcal{I}}$ 的情形，令 $r = \deg(f) = \max\{|\mathcal{I}|\,|\alpha_{\mathcal{I}} \neq 0\}$ 表示其次数，并令 $\mathrm{coeff}_r(f) = \sqrt{\sum_{|\mathcal{I}|=r} \alpha_{\mathcal{I}}^2}$ 为其最高次项系数向量的欧几里得范数

。对于多项式 f 的最高次项系数向量，我们称分布 ν 是$\sigma_{\text{poly},r}$-多项式光滑的，若对于所有 $a \in \mathbb{R}$，以及所有次数为 f 且满足 $\text{coeff}_r(f) = 1$ 的多项式，均有 $\mathbb{P}_{x\sim\nu}(|f(x) - a| \leq \epsilon) \leq \sigma_{\text{poly},r}^{\epsilon^r}$。

继续之前，我们先作几点说明。首先，方向平滑性并不足以保证多项式平滑性，正如 Glazer 和 Mikulincer [2022, 示例 3]所示，因此，确实需要更受约束的对手才能应用我们的方法。其次，我们注意到，定义 4.1扩展了方向平滑性的概念；后者对应于$\sigma_{\text{poly},1}$-平滑性。最后，我们注意到，若干常见的分布族显然具有与维度无关的多项式平滑性$\sigma_{\text{poly},r}$，例如高斯分布，以及更一般地，由对数凹边缘分布构成的乘积测度 [Glazer 和 Mikulincer，2022，推论 4]；我们将在附录 G.1中进一步讨论这一点。假设对手是多项式平滑的，我们证明了一个类似于定理 2的结果，由此得到以下遗憾界：

定理 *3.* 假设 $\mathcal{Z} \subset \mathbb{R}^d$ 和 Θ 是欧几里得空间的一个子集，其 ℓ_1 直径有界于 D。令 Θ_{d} 参数化如下元组集合：在 $\mathbb{R}^d$ 上的 $\binom{K}{2}$ 次数为 r 的多项式 $(f_{\mathbf{w}_{kk'}})$，满足 $\mathit{coeff}_r(f_{\mathbf{w}_{kk'}}) = 1$，对所有 $k \in [K]$均成立。假设 ℓ 按 *(4.2)* 中的定义，且 $\phi(\theta_{\text{d}}, k, k', z) = f_{\mathbf{w}_{kk'}}(z)$，并且 ℓ 取值于单位区间。若 $\mathcal{M}$ 是由 $\sigma_{\text{poly},r}$-多项式平滑分布构成的类，且 $||z||_\infty \leq B$ 几乎必然成立，那么采用附录 *G.2* 中给出的适当的 η, n 取值后，算法 *2* 的 $\mathbb{E}[\mathit{Reg}_T] \leq \widetilde{\mathcal{O}}\left(\left(TK^2r^2d^rDB\sigma^{-1}\mathit{poly}_r\right)^{\frac{4r}{4r-1}}\right)$。因此，实现平均遗憾至多为 ϵ 所需的预言机复杂度由 $\widetilde{\mathcal{O}}\left(\left(\frac{K^2r^2d^rDB}{\sigma_{\text{poly},r}} \cdot \epsilon^{-2r}\right)\right)$控制。

我们以与证明定理 3类似的方式证明4.1，即，我们证明定理 2 的一个类似结论，适用于多项式光滑分布，以控制关于 ρ （4.3）的广义括号数和伪等距常数，然后再应用定理 1。完整细节见附录 G.2。我们指出，定理 3和推论 4.1 的证明有一个共同之处：来自 $\mathcal{M}$ 中分布的随机变量样本的函数具有足够强的反集中性，能够平滑损失函数的非连续部分。最后，请注意，我们可以将 ℓ 替换为 ℓ （4.1），并采用与附录 F.3中类似的间隔假设。

5 平滑的多步规划

在前面的章节中，我们关注的是在线预测；这里我们聚焦于相关的多步决策问题。具体而言，我们研究多步规划这一情形：学习者选择一系列动力学输入（用控制领域的术语来说，即开环规划），以便在有限规划时域内最小化累积控制损失。我们关注“混合动力学” [Borrelli，2003年，Henzinger 和 Sastry， 1998]，，其中每个状态空间都被划分为若干区域（称为模式），在每个区域内，动力学均满足利普希茨条件。这里我们考虑模式之间具有仿射决策边界的情形，并将多项式边界的讨论留待附录 H。我们指出，由于模式之间可能出现不连续性，这个问题颇具挑战性，这同样限制了以往技术的适用性。这一类足以对机器人规划中常见的分段仿射动力学建模 [Hogan 和 Rodriguez，2016年，Anitescu 和 Potra，1997年，Aydinoglu 等人， 2021]；在附录中，我们进一步将其推广到多项式决策边界[Posa 等人， 2015]。另请参阅附录 A中的相关工作。

形式化地，我们固定一个规划时域 $H \in \mathbb{N}$，并考虑一族状态为 $\mathbf{x}_h \in \mathcal{X} \subset \mathbb{R}^m$、输入为 $\mathbf{u}_h \in \mathcal{U} \subset \mathbb{R}^d$的动力系统。我们的决策变量是计划 $\theta = \bar{\mathbf{u}}_{1:H} \in \mathcal{K} \subset \mathcal{U}^{\times H}$，上下文则是元组 $z_t = (\mathbf{x}_{t,1}, \boldsymbol{\eta}_{t,1:H}, \boldsymbol{\xi}_{t,1:H}, g_{t;1:H,1:K}, \ell_t^v, \mathbf{W}_{t,1:H})$，由初始状态 $\mathbf{x}_{t,1} \in \mathcal{X}$、噪声 $\boldsymbol{\eta}_{t,h} \in \mathcal{X}$和 $\boldsymbol{\xi}_{t,h} \in \mathcal{U}$、定义模式-$k$ 在步骤 h时动力学的连续函数 $g_{t,h,k}$、随时间变化的连续损失函数 ℓ_t^v，以及决定模式边界的矩阵 $\mathbf{W}_{t,h} \in \mathbb{R}^{K(m+d+1)}$构成，其中 $\mathbf{W}_{t,h}$ 的各行为 $\mathbf{w}_{t,h,k} \in \mathcal{S}^{m+d}$。我们用 $\mathbf{v} \in \mathcal{V} = \mathcal{X} \times \mathcal{U}$ 表示状态与输入的拼接。我们假设动力学是分段连续的，其中

$$\begin{aligned} \mathbf{x}_{t,h+1}(\theta) &= g_{t,h,k_{t,h}(\mathbf{v}_{t,h}(\theta))}(\mathbf{v}_{t,h}(\theta)) + \boldsymbol{\eta}_{t,h}, \quad \text{and} \\ \mathbf{u}_{t,h}(\theta) &= \bar{\mathbf{u}}_{t,h} + \boldsymbol{\xi}_{t,h}, \quad \mathbf{v}_{t,h}(\theta) = (\mathbf{x}_{t,h}(\theta), \quad \mathbf{u}_{t,h}(\theta)), \\ k_{t,h}(\mathbf{v}) &= \arg\max_{k \in [K]} \phi_{t,h}(k, \mathbf{v}), \quad \text{and} \quad \phi_{t,h}(k, \mathbf{v}) = \langle \mathbf{w}_{t,h,k}, (\mathbf{v}, 1) \rangle . \end{aligned} \tag{5.1}$$

换言之，对于每个时刻 t，都有长度为 H 的轨迹，它们按照分段连续动力学演化，其中每一段（模式）由前一状态和输入的仿射函数决定。我们的目标是最小化损失 $\ell(\theta, z_t) := \ell_t^v(\mathbf{v}_{t,1:H}(\theta))$ 下的遗憾，其中 $\ell_t^v : \mathcal{V}^H \to \mathbb{R}$ 是关于状态和输入的 1- 利普希茨函数。我们假设，对于固定的模式序列 $k_{1:h} \in [K]^h$，利普希茨动力学映射$g_{t,h,k_h} \circ g_{t,h-1,k_{h-1}} \circ \cdots \circ g_{t,1,k_1}$ 的 $h \in [H]$重组合成关于 $\theta \in \mathcal{K}$ 的 L-利普希茨函数，且满足某种 $\ell_1 \to \ell_1$ 意义上的条件（精确定义见附录）。尽管在最坏情况下，L 可能随 H 呈指数增长，但常见的稳定性条件可确保 L 有更合理的界；更多说明请参见注 H.1。最后，为了纳入光滑性，令 $\mathcal{F}_t$ 表示由 $(z_{1:t-1}, \ell_t^v, g_{t,1:H,1:K}, \mathbf{W}_{t,1:H})$ 生成的滤过；对于 $h \geq 0$，令 $\mathcal{F}_{t,h}$表示由 $\mathcal{F}_t$ 和$\boldsymbol{\xi}_{t,1:h}, \boldsymbol{\eta}_{t,1:h}, \mathbf{x}_{t,1}$ 生成的滤过；我们假设动力学与输入噪声构成的元组 $(\boldsymbol{\xi}_{t,h}, \boldsymbol{\eta}_{t,h})$ 在给定$\mathcal{F}_{t,h}$的条件下是 σ_{dir}-方向光滑的。

虽然限制为开环计划似乎会有所局限，但我们注意到，对我们的 $g_{t,\,h,\,k}$ 定义所赋予的灵活性，使我们只需稍作修改，就能纳入各种各样的状态依赖策略。例如，我们的框架涵盖了广受欢迎的线性控制设定，其中学习者采用一个将状态映射为输入的仿射函数；令 $g_{t,\,h,\,k}$ 关于输入矩阵和状态为多线性，并令损失为二次函数，由于有界性假设，两者均保持利普希茨连续，因此我们自然可以得到著名的线性二次调节器（LQR）的一种分段推广。我们的主要结果如下。

定理 *4.* 假设我们处于 *(5.1)* 所述的情形，$(\boldsymbol{\eta}_{t,h}, \boldsymbol{\xi}_{t,h}) | \mathcal{F}_{t,h-1}$ σ_{dir}-方向光滑，$\sup_{\mathbf{v} \in \mathcal{V}} ||\mathbf{v}||_1 \leq D$，$\ell_t^v$ 均为利普希茨连续且有界，并且 $g_{t,h,k}$ 满足定理 *10* 中给出的技术性连续性假设。若存在某个裕度参数 $\gamma > 0$，使得对所有 $t \in [T]$ 和 $h \in [H]$，均有 $\min_{k \neq k' \in [K]} ||\mathbf{w}_{t,h,k} - \mathbf{w}_{t,h,k'}|| \geq \gamma$，且规划器根据算法 *2* 执行 $\bar{\mathbf{u}}_{t,h}$，那么实现平均遗憾 ϵ 所需的预言机复杂度至多为 $\widetilde{\mathcal{O}}\left((dH^5 K^4 (DL/(\gamma \sigma_{\text{dir}}))^2)^{\frac{1}{3}} \epsilon^{-2} \right)$。

证明、假设的详细说明以及向多项式决策边界的扩展见 附录 H。证明遵循上一节的模板；为处理多步设置，我们论证平滑动力噪声足以保证：当 θ、$\theta' \in \mathcal{K}$ 足够接近时，光滑性以高概率确保模式序列 $k_{t,\,h}(\mathbf{v}_{t,\,h}(\theta))k_{t,\,h}(\mathbf{v}_{t,\,h}(\theta))$ 对所有 $h \in [H]$ 都一致；这需要一个在思想上类似于强化学习中 Kakade 的性能差异引理的望远镜求和论证 [Kakade, 2003]。

致谢

AB 感谢美国国家科学基金会研究生研究奖学金（资助编号 1122374）的支持。我们还感谢美国海军研究办公室（资助编号 N00014-20-1-2336）、美国能源部（资助编号 DE-SC0022199）以及美国国家科学基金会（奖项编号 DMS-2031883）提供的支持。MS 感谢 Amazon.com Services LLC 的资助；采购订单号为 2D-06310236。我们还感谢 Russ Tedrake、Terry H.J. Suh 和 Tao Pang 提供的宝贵意见。

参考文献

Yasin Abbasi-Yadkori 和 Csaba Szepesvári。线性二次系统自适应控制的遗憾界。载于第*24*届年度学习理论会议论文集，第1–26页。JMLR Workshop and Conference Proceedings，2011年。

Jacob Abernethy、Chansoo Lee、Abhinav Sinha 和 Ambuj Tewari。通过平滑实现在线线性优化。载于学习理论会议，第807–823页。PMLR，2014年。

Jacob D Abernethy、Chansoo Lee 和 Ambuj Tewari。用一种新型平滑性应对老虎机问题。《神经信息处理系统进展》，第28卷，2015年。

Alekh Agarwal、Daniel Hsu、Satyen Kale、John Langford、Lihong Li 和 Robert Schapire。驯服怪兽：一种快速而简单的上下文老虎机问题算法。载于国际机器学习会议，第1638–1646页。PMLR，2014年。

Naman Agarwal、Brian Bullins、Elad Hazan、Sham Kakade 和 Karan Singh。《带有对抗性扰动的在线控制》。载于国际机器学习会议，第 111–119 页。PMLR，2019a年。

Naman Agarwal、Alon Gonen 和 Elad Hazan。在优化预言机下的非凸博弈学习。载于 学习理论会议，第18–29页。PMLR，2019年b。David Angeli。李雅普诺夫方法与增量稳定性。载于 第*39*届*IEEE*决策与控制会议论文集（目录号*00CH37187*），第3卷，第2947–2952页。电气电子工程师学会，2000年。Mihai Anitescu 和 Florian A Potra。将带摩擦的动态多刚体接触问题表述为可解的线性互补问题。《非线性动力学》，14：231–247，1997年。Alp Aydinoglu、Philip Sieg、Victor M Preciado 和 Michael Posa。通过接触感知控制器实现互补系统的稳定化。《*IEEE*机器人学汇刊》，38(3)：1735–1754，2021年。Jushan Bai 和 Pierre Perron。具有多个结构变化的线性模型的估计与检验。《计量经济学》，第47–78页，1998年。Maria-Florina Balcan、Travis Dick 和 Ellen Vitercik。面向数据驱动算法设计、在线学习和隐私优化的离散度。载于*2018*年*IEEE*第*59*届计算机科学基础年度研讨会（*FOCS*），第603–614页。电气电子工程师学会，2018年。

Maria-Florina Balcan、Dan DeBlasio、Travis Dick、Carl Kingsford、Tuomas Sandholm 和 Ellen Vitercik。学习高性能算法需要多少数据？数据驱动算法设计的泛化保证。载于 第 *53* 届 *ACM SIGACT*计算理论年度研讨会论文集，第919–932页，2021年。Adam Block 和 Max Simchowitz。广义线性函数的高效且近乎最优的平滑在线学习。 *arXiv*预印本 *arXiv:2205.13056*，2022。Adam Block、Yuval Dagan、Noah Golowich 和 Alexander Rakhlin。平滑在线学习与统计学习同样容易。载于 学习理论会议，第1716–1786页。PMLR，2022。Adam Block、Max Simchowitz 和 Russ Tedrake。分段仿射系统中的预测平滑在线学习。 *arXiv* 预印本 *arXiv:2301.11187*，2023 年。JR Blum。关于经验分布函数的收敛性。 《数学统计年鉴》，第527–529页，1955年。Francesco Borrelli。 线性与混合系统的约束最优控制，第290卷。施普林格，2003年。Anthony Carbery 和 James Wright。凸体上rn中多项式的分布与lq范数不等式。 《数学研究快报》，8(3)：233–248，2001年。Nicolo Cesa-Bianchi 和 Gábor Lugosi。 预测、学习与博弈。剑桥大学出版社，2006年。Alon Cohen 和 Tamir Hazan。在线结构化学习中的扰动跟随者。载于 Francis Bach 和 David Blei 编， 第*32*届国际机器学习会议论文集，第37卷， 《机器学习研究论文集》，第1034–1042页，法国里尔，2015年7月7–9日。PMLR。网址https://proceedings.mlr.press/v37/cohena15.html。Alon Cohen、Avinatan Hasidim、Tomer Koren、Nevena Lazic、Yishay Mansour 和 Kunal Talwar。在线线性二次控制。载于 国际机器学习会议，第1029–1038页。PMLR，2018年。Yan Dai、Haipeng Luo 和 Liyu Chen。带赌博机反馈的对抗性马尔可夫决策过程中的扰动跟随者。载于 《神经信息处理系统进展》，2022。Sarah Dean 和 Benjamin Recht。确定性等价的基于感知的控制。载于 动力学与控制学习，第399–411页。PMLR，2021年。Sarah Dean、Horia Mania、Nikolai Matni、Benjamin Recht 和 Stephen Tu。线性二次调节器的样本复杂度。 《计算数学基础》，20(4)：633–679，2020年。John DeHardt。格利文科–坎泰利定理的推广。 《数学统计年鉴》，42(6)：2050–2055，1971年。Manfred Deistler、K Peternell 和 Wolfgang Scherrer。子空间方法的一致性与相对效率。 《自动化》，31(12)：1865–1875，1995年。

Richard M Dudley。经验测度的中心极限定理。《概率年刊》，第899–929页，1978年。Paul I Feder。分段回归问题的渐近分布理论——可辨识情形。《统计年刊》，3(1):49–83，1975年。Dylan Foster、Tuhin Sarkar 和 Alexander Rakhlin。基于单条轨迹学习非线性动力系统。载于动力学与控制学习，第851–861页。PMLR，2020年。Andrea Garulli、Simone Paoletti 和 Antonio Vicino。切换系统与分段仿射系统辨识综述。*IFAC*会议论文集，45(16):344–355，2012年。Evarist Giné 和 Richard Nickl。无限维统计模型的数学基础。剑桥大学出版社，2021年。Itay Glazer 和 Dan Mikulincer。多项式的反集中：无维数依赖的协方差界与傅里叶系数衰减。《泛函分析杂志》，283(9):109639，2022。Nika Haghtalab、Tim Roughgarden 和 Abhishek Shetty。在线学习与差分隐私学习的平滑分析。《神经信息处理系统进展》，33:9203–9215，2020年。Nika Haghtalab、Yanjun Han、Abhishek Shetty 和 Kunhe Yang。面向超越最坏情况对手的预言机高效在线学习。*arXiv*预印本 *arXiv:2202.08549*，2022a。Nika Haghtalab、Tim Roughgarden 和 Abhishek Shetty。针对自适应对手的平滑分析。载于*2021 IEEE*第*62*届计算机科学基础年会（*FOCS*），第942–953页。电气电子工程师学会，2022b。Elad Hazan 和 Tomer Koren。在线学习中优化的计算能力。载于第四十八届美国计算机协会计算理论年会论文集，第128–141页，2016年。Elad Hazan 和 Karan Singh。在线非随机控制导论。*arXiv*预印本 *arXiv:2211.09619*，2022年。Elad Hazan、Karan Singh 和 Cyril Zhang。通过谱滤波学习线性动力系统。《神经信息处理系统进展》，30，2017年。Thomas A Henzinger 和 Shankar Sastry。《混合系统：计算与控制：第一届国际研讨会，*HSCC'98*，伯克利，加利福尼亚州，美国，*1998*年*4*月*13*日至*15*日：论文集》。施普林格，1998年。François Robert Hogan 和 Alberto Rodriguez。推杆-滑块系统的反馈控制：混合与欠驱动接触动力学的故事。*arXiv*预印本 *arXiv:1611.08268*，2016年。Ali Jadbabaie 和 John Hauser。一般终端代价下无约束滚动时域控制的稳定性。载于第*40*届*IEEE*决策与控制会议论文集（目录号*01CH37228*），第5卷，第4826–4831页。电气电子工程师学会，2001年。

Sham Kakade、Akshay Krishnamurthy、Kendall Lowrey、Motoya Ohnishi 和 Wen Sun。在线非线性控制的信息论遗憾界。《神经信息处理系统进展》，33:15312–15325，2020年。Sham Machandranath Kakade。《强化学习的样本复杂度》。伦敦大学学院（英国），2003年。Adam Kalai 和 Santosh Vempala。在线决策问题的高效算法。《计算机与系统科学杂志》，71(3):291–307，2005年。Yann LeCun、Yoshua Bengio 和 Geoffrey Hinton。深度学习。《自然》，521(7553):436–444，2015年。Zifan Li 和 Ambuj Tewari。超越风险率：用于对抗性多臂老虎机的更多扰动算法。《机器学习研究杂志》，18:183–1，2017年。Nick Littlestone。当无关属性大量存在时快速学习：一种新的线性阈值算法。《机器学习》，2:285–318，1988年。Lennart Ljung。系统辨识：用户理论。普伦蒂斯·霍尔，新泽西州，1999年。Lennart Ljung 和 Bo Wahlberg。用于估计传递函数和扰动谱的最小二乘法的渐近性质。《应用概率进展》，24(2):412–440，1992年。Horia Mania、Michael I Jordan 和 Benjamin Recht。具有保证的非线性系统辨识主动学习。*arXiv*预印本 *arXiv:2006.10277*，2020年。Tobia Marcucci 和 Russ Tedrake。分段仿射系统最优控制的混合整数建模。载于第*22*届美国计算机协会混合系统：计算与控制国际会议论文集，第230–239页，2019年。Zakaria Mhammedi、Dylan J Foster、Max Simchowitz、Dipendra Misra、Wen Sun、Akshay Krishnamurthy、Alexander Rakhlin 和 John Langford。基于非线性观测学习线性二次调节器。《神经信息处理系统进展》，33:14532–14543，2020年。Elchanan Mossel、Ryan O’Donnell 和 Krzysztof Oleszkiewicz。低影响函数的噪声稳定性：不变性与最优性。《数学年刊》，第295–341页，2010年。Richard Nickl 和 Benedikt M Pötscher。Besov型和Sobolev型函数类的括号度量熵速率与经验中心极限定理。《理论概率杂志》，20:177–199，2007年。Samet Oymak 和 Necmiye Ozay。基于单条轨迹对LTI系统进行非渐近辨识。载于*2019*年美国控制会议（*ACC*），第5655–5661页。电气电子工程师学会，2019年。Daniel Pfrommer、Thomas TCK Zhang、Stephen Tu 和 Nikolai Matni。Tasil：泰勒级数模仿学习。*arXiv*预印本 *arXiv:2205.14812*，2022年。Michael Posa、Cecilia Cantu 和 Russ Tedrake。刚体接触运动轨迹优化的直接方法。《国际机器人研究杂志》，33(1):69–81，2014年。

Michael Posa、Mark Tobenkin 和 Russ Tedrake。冲击与摩擦作用下刚体系统的稳定性分析与控制。*IEEE* 自动控制汇刊，61(6):1423–1437，2015年。Alexander Rakhlin、Karthik Sridharan 和 Ambuj Tewari。在线学习：随机、受约束和平滑对手。《神经信息处理系统进展》，24，2011年。Alexander Rakhlin、Karthik Sridharan 和 Ambuj Tewari。通过序列复杂度进行在线学习。机器学习研究期刊，16(1):155–186，2015年。Yahya Sattar 和 Samet Oymak。非线性动力系统的非渐近精确学习。机器学习研究期刊，23(140):1–49，2022。Yahya Sattar、Zhe Du、Davoud Ataee Tarzanagh、Laura Balzano、Necmiye Ozay 和 Samet Oymak。马尔可夫跳变系统的辨识与自适应控制：样本复杂度与遗憾界。*arXiv* 预印本 *arXiv:2111.07018*，2021年。Max Simchowitz 和 Dylan Foster。朴素探索对于在线线性二次调节是最优的。载于国际机器学习会议，第8937–8948页。PMLR，2020年。Max Simchowitz、Horia Mania、Stephen Tu、Michael I Jordan 和 Benjamin Recht。无需混合的学习：迈向线性系统辨识的精确分析。载于学习理论会议，第439–473页。PMLR，2018年。Max Simchowitz、Karan Singh 和 Elad Hazan。非随机控制的非适当学习。载于学习理论会议，第3320–3436页。PMLR，2020年。Arun Sai Suggala 和 Praneeth Netrapalli。在线非凸学习：跟随扰动领导者是最优的。载于算法学习理论，第845–861页。PMLR，2020年。Hyung Ju Suh、Max Simchowitz、Kaiqing Zhang 和 Russ Tedrake。可微模拟器能否带来更好的策略梯度？载于国际机器学习会议，第20668–20696页。PMLR，2022a。Hyung Ju Terry Suh、Tao Pang 和 Russ Tedrake。通过随机平滑计算接触过程中的捆绑梯度。*IEEE* 机器人与自动化快报，7(2):4000–4007，2022b。Anastasios Tsiamis 和 George J Pappas。随机系统辨识的有限样本分析。载于*2019*年 *IEEE* 第*58*届决策与控制会议 *(CDC)*，第3648–3654页。IEEE，2019年。Anastasios Tsiamis、Nikolai Matni 和 George Pappas。未知系统卡尔曼滤波的样本复杂度。载于动力学与控制学习，第435–444页。PMLR，2020年。Anastasios Tsiamis、Ingvar Ziemann、Nikolai Matni 和 George J Pappas。控制的统计学习理论：有限样本视角。*arXiv* 预印本 *arXiv:2209.05423*，2022。Roman Vershynin。高维概率：导论及其在数据科学中的应用，第47卷。剑桥大学出版社，2018年。Yohei Yamamoto 和 Pierre Perron。利用频带谱回归估计和检验线性模型中的多重结构变化。计量经济学期刊，16(3):400–429，2013年。

相关工作

在本节中 本节将继续讨论引言中提及的相关研究 n.

平滑在线学习平滑在线学习最初由Rakhlin 等人[2011], 提出，近期的相关工作包括Haghtalab 等人 [2020, 2022a、b], Block 等人 [2022]。具体而言，Haghtalab 等人 [2022b] 刻画了平滑在线分类的统计速率，而Block 等人 [2022] 则对更一般的实值函数设定进行了同样的分析。关于平滑在线学习预言机高效算法的首次分析由Haghtalab 等人[2022a], Block 等人 [2022], 开展，两项工作都提出了适当和不适当算法。两项工作还给出了下界，表明低效算法与预言机高效算法所产生的遗憾在平滑参数 σ 上的依赖存在指数级差距。

为了解决预言机高效平滑在线学习中遗憾保证呈指数级恶化的问题，Block 和 Simchowitz [2022] 研究了一个特殊情形：损失函数是线性阈值函数，其参数取自 Θ。在限制更严格的可实现设定中，存在某个 $\theta \in \Theta$ 使累积损失为零，他们的工作成功将遗憾界降至关于 T/σ的对数级。不幸的是，无噪声假设并不现实，而且由此得到的算法在取消该假设后不具备鲁棒性；因此，Block 和 Simchowitz [2022] 提出了方向平滑性这一新概念，将平滑性假设放宽为一种更适合线性结构的假设。在此基础上，Block 等人 [2023] 证明了在具有挑战性的分段仿射（PWA）设定下，可以实现预言机高效平滑在线学习，其遗憾仅以多项式形式依赖于所有相关问题参数；不过，该算法不太实用，而且遗憾对时域的依赖显著更差。然而，我们指出，Block 等人 [2023] 的结果无法与我们的结果比较，因为我们的算法要求的经验风险最小化预言机概念比Block 等人 [2023], 所用的更强，下面将对此进一步说明。

跟随扰动领袖与预言机高效在线学习。跟随扰动领袖（FTPL）最早由 Kalai 和 Vempala [2005] 针对线性损失提出并进行了分析。在这项工作中，作者引入了 Be-the-Leader 引理，将遗憾分解为扰动项和稳定性项；这仍是证明此类算法遗憾界最常用的方法。此后，这一算法框架广受欢迎，并应用于多臂老虎机问题 [Abernethy 等人、 2015], 强化学习 [Dai 等人、2022], 以及在线结构化学习 [Cohen 和 Hazan 等领域。与本文更为相关的是，Agarwal 等人 [2019b], Suggala 和 Netrapalli [2020], 的研究表明，在对抗性在线学习设定中，如果损失函数满足利普希茨条件，那么采用指数扰动的 FTPL 可以达到最优遗憾。在定理 1 中，我们运用新的复杂度概念，将这两项工作的思路拓展到了利普希茨条件不成立的情形。由于指数分布具有无记忆性，它成为分析 FTPL 实例时最常用的扰动之一 [Kalai 和 Vempala，2005，Suggala 和 Netrapalli，2020，Agarwal 等人，2019b], ；不过，也有研究采用不同技术考察了其他分布[Abernethy 等人，2014，2015，Li 和 Tewari，2017，Block 等人，2022，Haghtalab 等人，2022a]。我们的分析高度依赖这一无记忆性，因此将研究范围限定于这一实例。未来一个值得探讨的问题是，类似结果能否推广到更一般的扰动分布。

预言机高效在线学习的下界证明起来远比上界困难。在Hazan 和 Koren [2016], 中，作者证明，在实现平均遗憾至多为 ϵ时，统计复杂度与计算复杂度之间存在指数级差距；类似地，Block 等人 [2022], Haghtalab 等人 [2022a] 的下界也是基于归约到这一结果得出的。另一方面，这一下界似乎有些脆弱，因为它只适用于适当学习，且对经验风险最小化预言机的概念有所限制。尽管Block 等人 [2023]使用的预言机符合这一模型，但我们假设的预言机，以及Hazan 和 Koren [2016], Suggala 和 Netrapalli [2020],使用的预言机则不符合。关于这一点的更多讨论，请参见Hazan 和 Koren [2016]。

分段仿射系统中的预测与规划 我们的示例部分源于对分段仿射系统中的规划与预测的研究，更广泛地说，也源于对利普希茨区域之间具有多项式边界的系统的研究。分段仿射动力学在受约束模型预测控制和混合系统文献中十分常见， [Henzinger 和 Sastry，1998年，Borrelli， 2003], 部分原因在于它们能够对机器人系统中的接触动力学进行建模 [Marcucci 和 Tedrake，2019年，Anitescu 和 Potra，1997年，Suh 等人，2022b]；多项式边界则见 [Posa 等人， 2015]。Suh 等人 [2022b，b]研究了随机噪声注入在穿越具有这些形式的不连续性的系统进行轨迹规划时的优势，并展示了诸多优点。对于典型的噪声分布（例如高斯分布），这些随机噪声注入会引入与本文所利用的相同类型的光滑性。

控制与动力学预测的统计学习和在线学习。基于数十年来的系统辨识文献 [Ljung，1999年；Ljung 和 Wahlberg，1992年；Deistler 等， 1995], 近期研究已针对多种感兴趣的情形，为线性动力系统中的参数恢复提供了有限样本统计保证 [Simchowitz 等，2018年；Dean 等，2020年；Oymak 和 Ozay，2019年；Tsiamis 和 Pappas，2019年；Tsiamis 等， 2022]。进一步的研究探讨了光滑非线性动力学 [Mania 等，2020年；Sattar 和 Oymak，2022；Foster 等， 2020],以及仅观测模型为非线性的情形 [Dean 和 Recht，2021年；Mhammedi 等，2020]。与本文相关的是，Sattar 等人 [2021]研究了马尔可夫跳变系统，其中系统动力学会在有限个线性系统（“模式”）之间切换，模式间的切换由（离散）马尔可夫链控制。相比之下，本文研究的具有分段仿射边界的动力学，其模式取决于状态。

除了有限样本系统辨识方面的最新进展之外，大量研究还从遗憾的角度探讨了线性控制任务 [Abbasi-Yadkori 和 Szepesvári，2011年，Agarwal 等人，2019a年，Simchowitz 等人，2020年，Simchowitz 和 Foster，2020年，Cohen 等人， 2018]。Kakade 等人 [2020]研究了具有特定线性参数结构的固定非线性系统的非线性在线控制；与本文类似，尽管其适用范围较窄，但他们通过高斯平滑保证低遗憾，这是其中的关键步骤。相比之下，第 5 节允许时变动力学，且不依赖于低维参数的恢复。

我们的预测保证与 [Hazan 等人，2017年，Tsiamis 等人， 2020]在线性控制设置中实现的在线预测在精神上相似；不过，它们当然适用于范围广得多的动力系统类。

括区熵 括区间数的概念最早可追溯至 Blum [1955],De-Hardt [1971] ，并被用于证明一致大数定律。随后，Dudley[1978] 利用它们证明了一致中心极限定理。对于

值得关注的具体函数类，已有许多相关界，其中最著名的可能是 Besov 类和 Sobolev类[Nickl and Pötscher， 2007]。我们的定义 3.1 通过将绝对值改为一般的伪度量，并且更重要的是，要求期望在一族测度上保持一致，从这两个方面推广了这一概念。

PAC学习分段利普希茨函数与"弥散"条件我们强调，我们研究的是分段利普希茨函数的在线学习。在*PAC*学习设定中，可以通过界定分段仿射函数的伪维数，推导出分布上的泛化界 [Balcan et al.， 2021]。此外，我们的函数类不满足"弥散条件" [Balcan et al.， 2018], 而满足该条件便可使此函数类无需平滑即可进行在线学习；此外，如何基于弥散条件开发预言机高效的在线学习算法，仍是一个未解决的问题。

B 命题3.1的证明

本节将通过 Freedman 不等式证明命题3.1 。回顾如下：

引理 B.1 (Freedman 不等式，Agarwal 等人 [2014])。设 Z_t 对于 $1 \le t \le T$ 是一个实值鞅差序列，且在给定 $Z_{1:t-1}$的条件下，几乎必然 $|Z_t| \le R$。那么对于任意$0 < \eta < \frac{1}{R}$，以至少 $1-\delta$的概率，有

$$\sum_{t=1}^{T} Z_t \le \eta \cdot \sum_{t=1}^{T} \mathbb{E}_{t-1}\left[Z_t^2\right] + \frac{R\log\left(\frac{1}{\delta}\right)}{\eta}.$$

现在我们可以证明这一关键结果：

命题 *3.1* 的证明。 令

$$Z_i(\theta,\theta') = \rho(\theta,\theta',x_i) - \mathbb{E}_i\left[\rho(\theta,\theta',x_i)\right],$$

其中，我们使用简便记法 $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{i-1}]$， $\mathcal{F}_i$ 的定义与命题陈述中相同。首先注意，根据假设，$|Z_i| \le 2D$ ，并且

$$\mathbb{E}_i\left[Z_i^2\right] \le \mathbb{E}_i\left[\rho(\theta,\theta',x_i)^2\right] \le D\cdot\mathbb{E}_i\left[\rho(\theta,\theta',x_i)\right].$$

应用引理 B.1，可知对于任意固定的 $0 < \eta < \frac{1}{R}$ 和 θ， $\theta' \in \Theta$，以至少$1-\frac{\delta}{2}$ 的概率，有

$$\begin{aligned}\sum_{i=1}^{n}\rho(\theta,\theta',x_i) &\le (1+\eta D)\sum_{i=1}^{n}\mathbb{E}_i\left[\rho(\theta,\theta',x_i)\right] + \frac{2D\log\left(\frac{2}{\delta}\right)}{\eta} \\ &\le (1+\eta D)n\cdot\sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\rho(\theta,\theta',x)\right] + \frac{2D\log\left(\frac{2}{\delta}\right)}{\eta}.\end{aligned}$$

令 $\mathcal{N} = \{(\theta_j,\mathcal{B}_j)\}$ 表示一个最小广义 ϵ-括号。根据并集界并令 $\eta = \frac{1}{3D}$，可知以至少 $1-\frac{\delta}{2}$ 的概率，对所有 $\theta_j,\theta_k \in \mathcal{N}$，均有

$$\sum_{i=1}^{n}\rho(\theta_j,\theta_k,x_i) \le 4n\cdot\sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\rho(\theta_j,\theta_k,x)\right] + 2D^2\cdot\log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right).$$

类似地，我们定义

$$\widetilde{Z}_i^j = \sup_{\theta\in\mathcal{B}_j} \rho(\theta,\theta_j,x_i) - \mathbb{E}_i\left[\sup_{\theta\in\mathcal{B}_j} \rho(\theta,\theta_j,x_i)\right]$$

并注意到，根据广义括号的定义，

$$\left|\widetilde{Z}_i^j\right| \le 2D \qquad\qquad \mathbb{E}_i\left[(\widetilde{Z}_i^j)^2\right] \le D\epsilon.$$

因此，再次应用引理 B.1 以及并集界可知，以至少 $1-\frac{\delta}{2}$的概率，对所有 $\theta_j \in \mathcal{N}$，

$$\sup_{\theta\in\mathcal{B}_j} \sum_{i=1}^n \rho(\theta,\theta_j,x_i) \le 4n\epsilon + 2D^2\log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right).$$

根据三角不等式和并集界可知，以至少 $1-\delta$ 的概率，若对所有 θ，$\theta' \in \Theta$，令 θ_j 为 θ 在 $\mathcal{N}$ 上的投影，并令 θ_k 为 θ' 在 $\mathcal{N}$ 上的投影，则有

$$\begin{aligned}
\sum_{i=1}^n \rho(\theta,\theta',x_i) &\le \inf_{\theta_j\in\mathcal{N}}\left\{\sum_{i=1}^n \rho(\theta,\theta_j,x_i)\right\} + \sum_{i=1}^n \rho(\theta_j,\theta_k,x_i) + \inf_{\theta_k\in\mathcal{N}}\left\{\sum_{i=1}^n \rho(\theta_k,\theta',x_i)\right\} \\
&\le 4n\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta_j,\theta_k,x)\right] + 4n\epsilon + 6D^2\log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right) \\
&\le 4n\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta,\theta',x)\right] + \left(2n\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta,\theta_j,x)\right] + 2n\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta_k,\theta',x)\right]\right) \\
&\quad + 4n\epsilon + 6D^2\log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right) \\
&\le 4n\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta,\theta',x)\right] + 8n\epsilon + 6D^2\log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right).
\end{aligned}$$

结论得证。 □

C 命题3.2的证明

本节通过结合 Kalai 和 Vempala [2005] 的经典 Be-the-Leader 引理与命题 3.1，证明命题 3.2 的一个更一般版本。我们首先陈述并证明 Be-the-Leader 引理的一个惰性版本。我们沿用 Block 等人 [2022,引理 31]的证明：

引理 C.1 (Be-the-Leader 引理)。令 $n \in \mathbb{N}$，并假设对于每个 $1 \le \tau \le T/n$，学习者都会选择扰动累积损失的某个近似极小化点。更具体地说，对于随机过程上的某个~实值函数 γ，学习者选择某个 θ_τ 满足

$$L_{(\tau-1)\cdot n}(\widetilde{\theta}_\tau) + \omega_{(\tau-1)\cdot n+1}(\widetilde{\theta}_\tau) \le \gamma(\omega_{(\tau-1)\cdot n+1}) + \inf_{\theta\in\Theta} L_{(\tau-1)\cdot n}(\theta) + \omega_{(\tau-1)\cdot n+1}(\theta)$$

~并对所有 $(\tau-1)\cdot n+1 \le t \le \tau\cdot n$ 执行 θ_τ。假设这些 ω_t 在 t 上相互独立且同分布，并且是满足 $\mathbb{E}[\text{上确界}_{\theta\in\Theta}\,\omega_t(\theta)] \ge 0$ 的随机过程。那么，学习者的遗憾如下：

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \frac{T}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right] + \mathbb{E}\left[\sup_\theta \omega_1(\theta)\right] + \sum_{\tau=1}^{T/n}\mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \ell(\widetilde{\theta}_\tau, z_t) - \ell(\widetilde{\theta}_{\tau+1}, z_t)\right].$$

证明。 我们将 Block 等人 [2022, 引理 31] 的证明应用于 n 步的累积损失。因此，对于每个 $1 \le \tau \le T' = \lfloor T/n \rfloor$，回忆

$$\widetilde{\ell}_\tau(\theta) = \sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \ell(\theta, z_t).$$

我们将对 T' 应用不等式的归纳证明

$$\mathbb{E}\left[\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta_{\tau+1})\right] \le \mathbb{E}\left[\sum_{\tau=1}^{T'} \ell_\tau(\theta_{T'+1}) + \omega_{T'\cdot n+1}(\theta_{T'+1})\right] + \mathbb{E}\left[\sup_{\theta\in\Theta}\omega_1(\theta)\right] + \frac{T'}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right]. \quad \text{(C.1)}$$

对于 $T'=0$的基础情形，该陈述显然成立。假设对于某个固定的 $T'-1$，(C.1)成立。由构造可知

$$\begin{aligned}\mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau(\theta_{T'}) + \omega_{T'\cdot n+1}(\theta_{T'})\right] &\le \mathbb{E}\left[\inf_{\theta\in\Theta}\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau(\theta) + \omega_{T'\cdot n+1}(\theta)\right] + \mathbb{E}\left[\gamma(\omega_{T'\cdot n+1})\right] \\ &\le \mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}\left(\theta_{T'+1}\right) + \omega_{(T'+1)\cdot n+1}\left(\theta_{T'+1}\right)\right] + \mathbb{E}\left[\gamma(\omega_{(T'+1)\cdot n+1})\right],\end{aligned}$$

其中，第二个不等式源于以下事实：这些 ω_t 相互独立且同分布，并且根据 $\theta_{T'+1}$的构造。将归纳假设 (C.1)与上述不等式结合可得

$$\mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau(\theta_{\tau+1})\right] \le \mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau\left(\theta_{T'+1}\right) + \omega_{(T'+1)\cdot n+1}\left(\theta_{T'+1}\right)\right] + \mathbb{E}\left[\sup_{\theta\in\Theta}\omega_1(\theta)\right] + T'\cdot\mathbb{E}\left[\gamma(\omega_1)\right].$$

在两边加上 $\mathbb{E}\left[\widetilde{\ell}_{T'}(\theta_{T'+1})\right]$即可完成归纳证明。

继续进行，我们计算：

$$\begin{aligned}\mathbb{E}\left[\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta_{T'+1}) + \omega_{T'+1}(\theta_{T'+1})\right] &\le \mathbb{E}\left[\inf_{\theta\in\Theta}\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau\left(\theta\right) + \sup_{\theta'\in\Theta}\omega_1(\theta)\right] + \frac{T'}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right] \\ &\le \mathbb{E}\left[\inf_{\theta\in\Theta}\sum_{t=1}^{T} \ell(\theta, z_t)\right] + \mathbb{E}\left[\sup_{\theta\in\Theta}\omega_1(\theta)\right]\frac{T'}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right],\end{aligned}$$

~其中，第一个不等式使用了 $\theta_{T'+1}$ 的构造，第二个不等式使用了 ℓ_τ 的定义。最后，我们应用（C.1）并注意到：

$$\begin{aligned}\mathbb{E}\left[\mathrm{Reg}_T\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\ell(\theta_{\tau(t)}) - \inf_{\theta\in\Theta}\sum_{t=1}^{T}\ell(\theta, z_t)\right] \\ &\leq \mathbb{E}\left[\sum_{\tau=1}^{T'}\widetilde{\ell}_\tau(\theta_\tau)\right] - \mathbb{E}\left[\sum_{\tau=1}^{T'}\widetilde{\ell}_\tau(\theta_{\tau+1})\right] + 2\mathbb{E}\left[\sup_{\theta\in\Theta}\omega_1(\theta)\right] + \frac{T'}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right],\end{aligned}$$

whe ~re 我们用 $\tau(t) = \lfloor t/n \rfloor$ 表示。该结论可由 ℓ_τ 的构造得出。 □

现在我们可以证明本节的主要结果了。

命题 *C.2.* 假设我们处于受约束在线学习设定中，其中对手只能从类 $\mathcal{M}$ 中的某个分布采样 z_t 。进一步假设存在一个由 $\mathcal{Z}$ 参数化的 Θ 上的伪度量 ρ，满足定义 *3.2* 中的伪等距性质，使得对于所有 $\theta, \theta' \in \Theta$，都有 $\sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\ell(\theta, z) - \ell(\theta', z)\right] \leq \sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\rho(\theta, \theta', z)\right]$。若~学习者按照引理 *C.1* 中的方式进行操作，且 $\sup_{\theta,\theta'\in\Theta}\rho(\theta, \theta', z) \leq D$，那么对于任意 $\epsilon > 0$，期望遗憾有如下界：

$$\begin{aligned}\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \frac{T}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right] + \mathbb{E}\left[\sup_{\theta\in\Theta}\omega_1(\theta)\right] + 8\epsilon T + 1 \\ + \frac{6TD^2}{n}\cdot\log\left(T\cdot\mathcal{N}_{\mathcal{M},[]}\left(\Theta, \rho, \epsilon\right)\right) + 4n\alpha\cdot\sum_{\tau=1}^{T/n}\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right|\right|^\beta\right].\end{aligned}$$

证明。 应用引理 C.1，可得

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \frac{T}{n}\cdot\mathbb{E}\left[\gamma(\omega_1)\right] + \mathbb{E}\left[\sup_{\theta}\omega_1(\theta)\right] + \sum_{\tau=1}^{T/n}\mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n}\ell(\widetilde{\theta}_\tau, z_t) - \ell(\widetilde{\theta}_{\tau+1}, z_t)\right].$$

因此，我们只需对上面的最后一项求和进行界定。由于损失函数关于该伪度量满足利普希茨条件，我们有：

$$\sum_{\tau=1}^{T/n}\mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n}\ell(\widetilde{\theta}_\tau, z_t) - \ell(\widetilde{\theta}_{\tau+1}, z_t)\right] \leq \sum_{\tau=1}^{T/n}\mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n}\rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, z_t)\right].$$

因此，对于任意固定的 τ，以及所有 ϵ，$\delta > 0$,

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n}\rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, z_t)\right] \\ &\leq \mathbb{E}\left[4n\sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, z)\right] + 8\epsilon n + 6D^2\log\left(\frac{2\mathcal{N}_{\mathcal{M},[]}\left(\Theta, \rho, \epsilon\right)}{\delta}\right) + \delta n\right] \\ &\leq 4n\alpha\cdot\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right|\right|^\beta\right] + (8\epsilon + \delta)n + 6D^2\cdot\log\left(\frac{2\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon)}{\delta}\right),\end{aligned}$$

其中，第一个不等式由命题 3.1 得出，第二个不等式则基于以下假设：ρ 满足定义 3.2的条件。对 τ 求和并令 $\delta = T^{-1}$ 即可完成证明。

□

最后，我们注意到，命题 3.2 是前述分析的一个特例：

命题 *3.2* 的证明。 取 $\gamma = 0$ 在命题C.2 中为均匀分布，即可立即得到该结果。

□

D 推论 3.1的证明

本节首先证明，在此情形下，广义括号可以直接取为经典括号；然后应用 Block 等人[2022]给出的稳定性界，从而证明推论 3.1 。为遵循记号惯例，我们将用函数类 Θ 替换 $\mathcal{F}$，并考虑函数 $f \in \mathcal{F}$ 而非参数 $\theta \in \Theta$。我们令

$$\rho(f, f', z) = L \cdot \left| f(x) - f'(x) \right|, \qquad z = (x, y) \in \mathcal{Z},$$

并证明，在此 ρ下，经典括号经过缩放后即成为广义括号：

引理 D.1。 令 $\mathcal{M}$ 表示由关于某个分布 μ 在 $\mathcal{X}$上 σ-光滑的分布构成的类。若 ρ 如上所述，则

$$\mathcal{N}_{\mathcal{M},[]}(\mathcal{F}, \rho, \epsilon) \leq \mathcal{N}_{[]}\left(\mathcal{F}, \mu, \sigma\epsilon/L\right).$$

此外 r， ρ 对 $\alpha = L \cdot \sigma^{-1}$ 和 $\beta = 1$ 满足关于范数 $L^1(\mu$的伪等距).

证明。 令 $\mathcal{N} = \{\mathcal{B}_i\}$ 表示 $\mathcal{F}$ 相对于 μ的一个 $\widetilde{\epsilon}$-括号（经典意义下），其中$\widetilde{\epsilon} = \frac{\sigma\epsilon}{L}$，并令 f_i 表示一$\mathcal{B}_i$中的任意元素。那么对所有 $\nu \in \mathcal{M}$，我们有：

$$\begin{aligned}
\mathbb{E}_\nu \left[\sup_{f \in \mathcal{B}_i} \rho(f, f_i, z) \right] &= \mathbb{E}_\nu \left[\sup_{f \in \mathcal{B}_i} |f(x) - f_i(x)| \right] \\
&= \mathbb{E}_\mu \left[\frac{d\nu}{d\mu} \sup_{f \in \mathcal{B}_i} |f(x) - f_i(x)| \right] \\
&\leq \frac{1}{\sigma} \cdot \mathbb{E}_\mu \left[\sup_{f \in \mathcal{B}_i} |f(x) - f_i(x)| \right] \\
&\leq \frac{\epsilon}{\sigma},
\end{aligned}$$

根据经典括号的定义。注意到再次根据定义， $\mathcal{B}_i$ 覆盖 $\mathcal{F}$，即可完成第一个结论的证明。

根据光滑分布的定义，第二个结论显然成立。 □

我们现在回顾一个稳定性结果：

引理 D.2 （Block 等人 [2022]中的引理 34）。 假设我们处于推论*3.1* 的设定中，并令 $\widehat{\mu}$ 表示在抽样得到的 x_i上的经验测度。如果函数 ℓ 在 L第一个参数上是 -利普希茨的，并且

$$\sup_{f,f'\in\mathcal{F}}\left|||f-f'||^2_{L^2(\mu)}-||f-f'||^2_{L^2(\widehat{\mu})}\right|\leq\Delta, \tag{D.1}$$

那么对于任意固定的 y,

$$\mathbb{E}\left[||\ell(f_t(\cdot),y)-\ell(f_{t+1}(\cdot),y)||_{L^1(p_t)}\right]\leq\frac{30L^3\log(\eta)}{\sigma\eta}\cdot\mathbb{E}\left[1+\sup_{f\in\mathcal{F}}\omega(f)\right]+\frac{2L\Delta}{\sigma}.$$

引理 D.2 中关于经验范数与总体范数彼此接近的假设，是经典学习理论中的标准推论。因此，我们可以给出主要证明:

推论 *3.1* 的证明。 根据命题 3.2，算法1 所经历 es

$$\mathbb{E}\left[\mathrm{Reg}_T\right]\leq\mathcal{O}\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}\omega(f)\right]+\epsilon T+\frac{T}{n}\cdot\log\left(T\cdot\mathcal{N}_{\mathcal{M},[]}(\mathcal{F},\rho,\epsilon)\right)+2T\alpha\max_{\tau\leq T/n}\mathbb{E}\left[||f_\tau-f_{\tau+1}||^\beta\right]\right).$$

根据引理 D.1 的结果，我们可以取 $\alpha=\frac{1}{\sigma}$ 和 $\beta=1$ 如上所述，以恢复

$$\mathbb{E}\left[\mathrm{Reg}_T\right]\leq\mathcal{O}\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}\omega(f)\right]+\epsilon T+\frac{T}{n}\cdot\log\left(T\cdot\mathcal{N}_{[]}\left(\mathcal{F},\mu,\frac{\sigma\epsilon}{L}\right)\right)+\frac{T}{\sigma}\cdot\max_{\tau}\mathbb{E}\left[||f_\tau-f_{\tau+1}||\right]\right).$$

~现在注意到，如果 ℓ 是 L-利普希茨的，那么 ℓ（长度为 n 的一个训练轮次内的累积损失）根据三角不等式是Ln-利普希茨的。因此，我们得到

$$\max_{\tau}\mathbb{E}\left[||f_\tau-f_{\tau+1}||\right]\leq\mathcal{O}\left(\frac{L^3n^3\log(\eta)}{\sigma\eta}\cdot\mathbb{E}\left[1+\sup_{f\in\mathcal{F}}\omega(f)\right]+\frac{2Ln\Delta}{\sigma}\right),$$

其中我们在事件 (D.1) 上应用了引理 D.1。应用 Block 等人 [2022, 引理 36], , 我们看到以至少 $1-\delta$ 的概率，可以取

$$\Delta\leq\mathcal{O}\left(\frac{\mathbb{E}\left[\sup_{f\in\mathcal{F}}\omega(f)\right]}{\eta m^{3/2}}+\frac{\sqrt{\log\left(\frac{1}{\delta}\right)}}{m}\right).$$

应用Giné 和 Nickl [2021, 定理 3.5。13], 我们可以看到

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\omega(f)\right]\leq\mathcal{O}\left(\eta\sqrt{m\cdot\log\left(\mathcal{N}_{[]}(\mathcal{F},\mu,\epsilon)\right)}+\eta m\epsilon\right)$$

对所有 $\epsilon>0$ 均成立。因此，令

$$m=\sqrt{T},\qquad\eta=\frac{1}{\sigma}\cdot\sqrt{\frac{TL^3n^3}{m}},\qquad n=\frac{T^{1/5}\sigma^{2/5}}{L^{3/5}}\cdot\log^{2/5}\left(\mathcal{N}_{[]}\left(\mathcal{F},\mu,\frac{\sigma}{LT}\right)\right)$$

可得

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\frac{T^{4/5}L^{3/5}}{\sigma^{2/5}} \cdot \log^{3/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)\right)$$

使用相同数量的预言机调用。因此，特别地，为了使平均遗憾至多为ϵ，只需调用 ERMOracle

$$\widetilde{\mathcal{O}}\left(\frac{\epsilon^{-4}L^{3/5}}{\sigma^{2/5}} \cdot \log^{3/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)\right)$$

次。 □

E 定理 1 的证明

本节中，我们证明一个更一般的定理 1版本。回顾一下，对于固定的 n 和 $1 \leq \tau \leq T/n$，我们定义

$$\mathcal{I}_\tau = \{t|(\tau-1)n \leq t \leq \tau n\} \qquad \widetilde{\ell}_\tau(\theta) = \sum_{t\in\mathcal{I}_\tau} \ell(\theta, x_t)$$

$$L_t(\theta) = \sum_{s=1}^{t} \ell(\theta, x_s) \qquad \widetilde{L}_\tau(\theta) = \sum_{\tau'=1}^{\tau} \widetilde{\ell}_{\tau'}(\theta).$$

此外，我们假设 $\widetilde{\theta}_\tau$ 的选取满足：对于某个实值函数 $\gamma : \mathbb{R}^d \to \mathbb{R}$，有

$$\widetilde{L}_{\tau-1}(\widetilde{\theta}_\tau) - \eta\left\langle \xi, \widetilde{\theta}_\tau \right\rangle \leq \gamma(\eta\xi) + \inf_{\theta\in\Theta} \widetilde{L}_{\tau-1}(\theta) - \eta\langle \xi, \theta\rangle . \tag{E.1}$$

我们假设 $\xi \sim$ 参数为 1 的指数分布是 $\mathbb{R}^d$ 中的一个随机向量，其各坐标相互独立，并且 $\widetilde{\ }$ $\widetilde{\ }$按照标准指数分布抽取。对于固定的 ξ，令 $\theta_\tau(\xi)$ 表示某个 θ_τ，满足（E.1）。我们证明以下结果：

定理 *5.* 假设我们处于命题 *3.2* 的受约束在线学习设定中，其中 $\Theta \subset \mathbb{R}^d$ 满足 $\sup_{\theta, \theta'\in\Theta} ||\theta - \theta'||_1 = D < \infty$。进一步假设，$\mathcal{Z}$-参数化伪度量 ρ 相对于 ℓ^1 在 $\mathbb{R}^d$上满足定义 *3.2* 的伪等距性质，并且 $\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\ell(\theta, z) - \ell(\theta', z)\right] \leq \sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', z)\right]$。如果学习者采用算法 *2*，且 $\eta = \Omega(n^2)$（精确关系见 *(E.3)*），则期望遗憾有如下界：

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \eta D d + 2 + + \frac{3DT}{n}\log\left(T\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, 1/T)\right)$$
$$+ 8T\alpha \cdot d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right] + 8\epsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}}, .$$

注意，定理 1 通过考虑特殊情况 $\gamma(\xi) = 0$，可立即得出。

证明 定理 5 首先通过引理C.1，然后对稳定性项进行界定。对稳定性项的控制分为两部分：第一部分中，我们采用

Agarwal 等人的技术Agarwal 等人 [2019b], Suggala 和 Netrapalli [2020], 证明如果稳定性项较小，那么 $\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1$ 其期望也较小；第二部分中，我们结合伪等距假设与广义括号的控制，采用自界方法完成证明。

E.1 界定稳定性项

在本节中，我们采用 Suggala 和 Netrapalli [2020] 的技术来控制 $\widetilde{}$在 θ_τ 中的期望 $||\cdot||_1$稳定性。我们有以下关键引理：

引理 E.1。 假设 $\theta_\tau \in \mathbb{R}^d$ 按照 (E.1) 选取。进一步假设 ℓ^∞的直径 Θ 的上界为 D。那么有

$$\mathbb{E}\left[||\theta_\tau - \theta_{\tau+1}||_1\right] \leq d \cdot \sqrt{\frac{D}{\eta} \cdot \mathbb{E}\left[4\gamma(\xi) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|\right]}.$$

为了证明该结果，我们需要对 Suggala 和 Netrapalli [2020], 的关键单调性引理做少许修改，在运用他们的技术时不对损失函数施加 Lipschitz 假设。首先，有：

引理 E.2。 假设 ξ， $\xi' \in \mathbb{R}^d$满足 $\theta = \widetilde{\theta}_\tau(\xi)$ 且 $\theta' = \widetilde{\theta}_\tau(\xi')$ 对某个固定的 τ成立，如 (E.1)所示。那么以下不等式成立：

$$\eta \cdot \langle \xi' - \xi, \theta' - \theta \rangle \geq -\left(\gamma(\xi) + \gamma(\xi')\right).$$

证明。 我们计算：

$$\begin{aligned}
\widetilde{L}_\tau(\theta) - \eta \langle \xi, \theta \rangle &\leq \widetilde{L}_\tau(\theta') - \eta \langle \xi, \theta' \rangle + \gamma(\xi) \\
&= \widetilde{L}_\tau(\theta') - \eta \langle \xi', \theta' \rangle + \eta \langle \xi' - \xi, \theta' \rangle + \gamma(\xi) \\
&\leq \widetilde{L}_\tau(\theta) - \eta \langle \xi', \theta \rangle + \eta \langle \xi' - \xi, \theta' \rangle + \gamma(\xi) + \gamma(\xi').
\end{aligned}$$

结论得证。 □

第二个必要结果是Suggala 和 Netrapalli [2020, 引理 6]的类似结果：

引理 E.3。 设 ξ、$\xi' \in \mathbb{R}^d$满足 $\widetilde{\theta}_\tau = \widetilde{\theta}_\tau(\xi)$、$\widetilde{\theta}_{\tau'} = \widetilde{\theta}_\tau(\xi')$，并且 $\widetilde{\theta}_{\tau+1}$、$\widetilde{\theta}'_{\tau+1}$ 针对某个固定的 τ 以类似方式定义。则有以下不等式成立：

$$\begin{aligned}
\min\left(\left\langle \widetilde{\theta}'_\tau, \xi' - \xi \right\rangle, \left\langle \widetilde{\theta}'_{\tau+1}, \xi' - \xi \right\rangle\right) \geq \max\left(\left\langle \widetilde{\theta}_\tau, \xi' - \xi \right\rangle, \left\langle \widetilde{\theta}_{\tau+1}, \xi' - \xi \right\rangle\right) \\
- \frac{2\left(\gamma(\xi) + \gamma(\xi')\right) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta}.
\end{aligned}$$

证明。 根据构造，我们计算得：

$$\begin{aligned}
\widetilde{L}_\tau(\widetilde{\theta}_\tau) - \eta \left\langle \xi, \widetilde{\theta}_\tau \right\rangle &= \widetilde{L}_{\tau-1}(\widetilde{\theta}_\tau) - \eta \left\langle \xi, \widetilde{\theta}_\tau \right\rangle + \widetilde{\ell}_\tau(\widetilde{\theta}_\tau) \\
&\leq \widetilde{L}_{\tau-1}(\widetilde{\theta}_{\tau+1}) - \eta \left\langle \xi, \widetilde{\theta}_{\tau+1} \right\rangle + \widetilde{\ell}_\tau(\widetilde{\theta}_\tau) + \gamma(\xi) \\
&= \widetilde{L}_\tau(\widetilde{\theta}_{\tau+1}) - \eta \left\langle \xi, \widetilde{\theta}_{\tau+1} \right\rangle + \gamma(\xi) + \widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1}).
\end{aligned}$$

同样根据构造，我们有：

$$\begin{aligned}\widetilde{L}_\tau(\widetilde{\theta}_\tau) - \eta\left\langle \xi, \widetilde{\theta}_\tau\right\rangle &= \widetilde{L}_\tau(\widetilde{\theta}_\tau) - \eta\left\langle \xi', \widetilde{\theta}_\tau\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau\right\rangle \\ &\geq \widetilde{L}_\tau(\widetilde{\theta}'_{\tau+1}) - \eta\left\langle \xi', \widetilde{\theta}'_{\tau+1}\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau\right\rangle - \gamma(\xi') \\ &= \widetilde{L}_\tau(\widetilde{\theta}'_{\tau+1}) - \eta\left\langle \xi, \widetilde{\theta}'_{\tau+1}\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau - \widetilde{\theta}'_{\tau+1}\right\rangle - \gamma(\xi') \\ &\geq \widetilde{L}_\tau(\widetilde{\theta}_{\tau+1}) - \eta\left\langle \xi, \widetilde{\theta}_{\tau+1}\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau - \widetilde{\theta}'_{\tau+1}\right\rangle - \gamma(\xi').\end{aligned}$$

结合前面的两个展示式可得以下不等式：

$$\eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau - \widetilde{\theta}'_{\tau+1}\right\rangle \geq -2\left(\gamma(\xi) + \gamma(\xi')\right) - \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|.$$

同样的论证可得

$$\eta\left\langle \xi' - \xi, \widetilde{\theta}'_\tau - \widetilde{\theta}_{\tau+1}\right\rangle \geq -2\left(\gamma(\xi) + \gamma(\xi')\right) - \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|.$$

应用引理 E.2可得

$$\begin{aligned}\eta\left\langle \xi' - \xi, \widetilde{\theta}'_\tau - \widetilde{\theta}_\tau\right\rangle &\geq -\left(\gamma(\xi) + \gamma(\xi')\right) \\ \eta\left\langle \xi' - \xi, \widetilde{\theta}'_{\tau+1} - \widetilde{\theta}_{\tau+1}\right\rangle &\geq -\left(\gamma(\xi) + \gamma(\xi')\right).\end{aligned}$$

结合这些不等式即可完成证明。 □

现在我们可以证明稳定性界：

引理 *E.1*的证明。 对于某个固定的 τ 和 ξ，对所有 $1 \leq i \leq d$，令

$$\theta_{\max,i} = \max(\widetilde{\theta}_{\tau,i}, \widetilde{\theta}_{\tau+1,i}), \qquad \theta_{\min,i} = \min(\widetilde{\theta}_{\tau,i}, \widetilde{\theta}_{\tau+1,i}),$$

其中 $\widetilde{\theta}_{\tau,\ i}$ 表示 $\widetilde{\theta}_\tau$ 的 i^{th} 坐标。注意，$\left|\widetilde{\theta}_{\tau,\ i} - \widetilde{\theta}_{\tau+1,\ i}\right| = \theta_{\max,\ i} - \theta_{\min,\ i}$。假设 $\xi \sim$ 服从参数为 1 的指数分布，并令 $\xi' = \xi + k\mathbf{e}_i$。然后，利用指数分布的无记忆性，并记

$$\mathbb{E}_{-i}[\cdot] = \mathbb{E}\left[\cdot|\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_d\right],$$

我们有

$$\begin{aligned}\mathbb{E}_{-i}\left[\theta_{\min,i}\right] &= \mathbb{P}\left(\xi_i \leq k\right) \cdot \mathbb{E}_{-i}\left[\theta_{\min,i}|\xi_i \leq k\right] + \mathbb{P}\left(\xi_i > k\right) \cdot \mathbb{E}_{-i}\left[\theta_{\min,i}|\xi_i > k\right] \\ &\geq \left(1 - e^{-k}\right)\left(\mathbb{E}_{-i}\left[\theta_{\max,i}\right] - D\right) + e^{-k} \cdot \mathbb{E}_{-i}\left[\theta_{\min,i}|\xi_i > k\right] \\ &= \left(1 - e^{-k}\right)\left(\mathbb{E}_{-i}\left[\theta_{\max,i}\right] - D\right) + e^{-k} \cdot \mathbb{E}\left[\theta'_{\min,i}\right]\end{aligned}$$

其中 $\theta'_{\min,i} = \theta_{\min,i}(\xi')$。该不等式由关于 Θ 直径的假设得出，而第二个等式则由指数分布的无记忆性得出。应用引理 E.3 并注意到 $\left\langle \widetilde{\theta}_\tau, \xi' - \xi\right\rangle = k\widetilde{\theta}_{\tau,i}$，我们得到

$$\mathbb{E}_{-i}\left[\theta'_{\min,i}\right] \geq \mathbb{E}_{-i}\left[\theta_{\max,i}\right] - \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi) + \gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right].$$

因此，结合前面的各式可得

$$\mathbb{E}_{-i}\left[\theta_{\min,i}\right] \geq \mathbb{E}_{-i}[\theta_{\max,i}] - \left(1-e^{-k}\right) D - \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right].$$

那么，

$$\begin{aligned}
\mathbb{E}_{-i}\left[\left|\widetilde{\theta}_{\tau,i} - \widetilde{\theta}_{\tau+1,i}\right|\right] &= \mathbb{E}_{-i}\left[\theta_{\max,i} - \theta_{\min,i}\right] \\
&\leq \left(1-e^{-k}\right) D + \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right] \\
&\leq kD + \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right].
\end{aligned}$$

对 $1 \leq i \leq d$ 求和并对 k 取最小值，即可完成证明。 □

E.2 证明的收尾

我们将应用 Be-the-Leader 引理；为此，需要对扰动项和稳定性项进行界定。对于前者，我们有如下结果：

引理 E.4. 假设 $\Theta \subset \mathbb{R}^d$ 满足 $\sup_\theta ||\theta||_\infty \leq D$。那么

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \langle\theta, \xi\rangle\right] \leq Dd,$$

其中 ξ ~服从参数为 1 的指数分布。

证明。 注意到

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \langle\theta, \xi\rangle\right] \leq \mathbb{E}\left[\sup_{\theta\in\Theta} ||\theta||_\infty ||\xi||_1\right] \leq D\mathbb{E}\left[||\xi||_1\right] = Dd.$$

□

现在我们已经准备好证明主要结果：

定理 *5* 的证明。根据引理 C.1，只需分别界定扰动项和稳定性项。为了界定稳定性项，注意到根据关于 ρ 的 Lipschitz 性假设，我们有

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right] &= \mathbb{E}\left[\sum_{t\in\mathcal{I}_\tau} \ell(\widetilde{\theta}_\tau, x_t) - \ell(\widetilde{\theta}_{\tau+1}, x_t)\right] \\
&\leq \mathbb{E}\left[\sum_{t\in\mathcal{I}_\tau} \rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, x_t)\right].
\end{aligned}$$

现在应用命题 3.1 ，并注意到对于所有 ϵ， $\delta > 0$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t\in\mathcal{I}_\tau}\rho(\widetilde{\theta}_\tau,\widetilde{\theta}_{\tau+1},x_t)\right] &\le \mathbb{E}\left[4n\sup_{\nu\in\mathcal{M}}\left[\rho(\widetilde{\theta}_\tau,\widetilde{\theta}_{\tau+1},x)\right]+8\epsilon n+\delta+3D\log\left(\frac{4\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)\right] \\
&\le 4n\left(\alpha\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau-\widetilde{\theta}_{\tau+1}\right|\right|_1^\gamma\right]+\beta\right)+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right) \\
&\le 4n\left(\alpha\left(\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau-\widetilde{\theta}_{\tau+1}\right|\right|_1\right]\right)^\beta\right)+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right),
\end{aligned} \tag{E.2}$$

其中，第二个不等式由伪等距性质得出，最后一个不等式由詹森不等式以及 $\gamma \le 1$这一事实得出。根据引理E.1，我们有

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau-\widetilde{\theta}_{\tau+1}\right|\right|_1\right] &\le d\sqrt{\frac{D}{\eta}\cdot\mathbb{E}\left[4\gamma(\xi)+\left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau)-\widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|\right]} \\
&\le d\sqrt{\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)\right)} \\
&\quad + d\cdot\sqrt{\frac{2nD\alpha}{\eta}}\cdot\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau-\widetilde{\theta}_{\tau+1}\right|\right|_1\right]^{\frac{\beta}{2}},
\end{aligned}$$

将前面的公式代入引理E.1 并应用平方根的次可加性，即可得到第二个不等式。重新整理可得

$$\begin{aligned}
&\mathbb{E}\left[\left|\left|\widetilde{\theta}_\tau-\widetilde{\theta}_{\tau+1}\right|\right|_1\right] \\
&\le \max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}},\left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).
\end{aligned}$$

将其代入式（E.2），可得

$$\begin{aligned}
&\mathbb{E}\left[\widetilde{\ell}_\tau(\widetilde{\theta}_\tau)-\widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right] \le 8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right) \\
&+8n\alpha\cdot\max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}},\left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).
\end{aligned}$$

Su对 τ求和后，我们可以看到，引理 C.1 中的稳定性项上界为 y

$$\begin{aligned}
&8\epsilon T+\delta\frac{T}{n}+\frac{3DT}{n}\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right) \\
&+8T\alpha\cdot\max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}},\left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).
\end{aligned}$$

应用引理 C.1和 E.4 可知，期望遗憾的上界为

$$\eta Dd + 8\epsilon T + \delta\frac{T}{n} + \frac{3DT}{n}\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)$$
$$+ 8T\alpha\cdot\max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\epsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\epsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}},\left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).$$

若我们设定 $\delta=\epsilon=\frac{1}{T}$，注意只要

$$\eta \geq \frac{d^{4-2\beta}\cdot D^{2-\beta}\cdot d^{4-2\beta}\cdot\alpha^2}{\left(4\mathbb{E}\left[\gamma(\xi)\right]+3D\log\left(T\cdot\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,1/T)\right)\right)^{2\beta}}\cdot n^2, \tag{E.3}$$

第一个参数就不小于第二个参数，因此证明得证。 □

F 与分段连续函数及广义仿射边界相关的证明

在本节中，我们给出 定理 2 的详细证明。随后，我们陈述并证明一个类似的结果，将 ℓ 替换为 $\bar{\ell}$（见4.1），并假设边界满足额外的间隔条件。这样做既是为了提高结果的普适性，也为了将其应用于第 5 节的多步规划问题。

F.1 定理2 的证明

本节中，我们证明定理 2。首先介绍关键步骤，证明当 $\mathbb{P}\left(k_\phi(\theta_{\mathrm{d}},\ \mathbf{z})\neq k_\phi(\theta'_{\mathrm{d}},\ \mathbf{z})\right)\lesssim||\theta_{\mathrm{d}}-\theta'_{\mathrm{d}}||_1$ 且 $\mathbf{z}$ 来自一个 σ_{dir}方向光滑分布时，该结论成立。随后，我们将此结果用于控制伪等距常数，并界定广义括号数。

我们先给出以下引理:

引理 F.1。 假设 Θ_{d}， ϕ和 k_ϕ 如定理 2中所定义，并假设 $\mathbf{z}$ 是从某个 σ_{dir}-方向光滑分布中选取的，且满足 $||\mathbf{z}||_\infty\leq B$ 几乎必然成立。于是，

$$\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right)\leq\frac{AB}{a\sigma_{\mathrm{dir}}}\cdot\left|\left|\theta_{\mathrm{d}}-\theta'_{\mathrm{d}}\right|\right|_1.$$

证明。 我们首先注意到

$$\begin{aligned}
\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right)&=\mathbb{P}\left(\arg\max_{k\in[K]}\sum_{k'\neq k}\mathbb{I}\left[\overline{\phi}(\theta_{\mathrm{d}},k,k',z)\geq 0\right]\neq\arg\max_{k\in[K]}\sum_{k'\neq k}\mathbb{I}\left[\overline{\phi}(\theta'_{\mathrm{d}},k,k',z)\geq 0\right]\right)\\
&\leq\mathbb{P}\left(\bigcup_{k,k'\in[K]}\left\{\overline{\phi}(\theta_{\mathrm{d}},k,k',z)\geq 0>\overline{\phi}(\theta'_{\mathrm{d}},k,k',z)\right\}\right)\\
&=\mathbb{P}\left(\bigcup_{k,k'\in[K]}\left\{\psi\left(\langle\mathbf{w}_{kk'},\mathbf{z}\rangle\right)\geq 0>\psi\left(\left\langle\mathbf{w}'_{kk'},(\mathbf{z},1)\right\rangle\right)\right\}\right)\\
&\leq\sum_{k,k'\in[K]}\mathbb{P}\left(\psi\left(\langle\mathbf{w}_{kk'},(\mathbf{z},1)\rangle\right)\geq 0>\psi\left(\left\langle\mathbf{w}'_{kk'},(\mathbf{z},1)\right\rangle\right)\right),
\end{aligned}\tag{F.1}$$

其中，第一个不等式源于 ϕ 在 (k, k') 中是反对称的，第二个等式源于 ϕ 的构造，最后一个不等式则源于并集界。现在我们注意到，对于固定的 $k, k' \in [K]$,

$$\begin{aligned}
&\mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \geq 0 > \psi(\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\rangle)\right) \\
&\quad \leq \mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \leq \left|\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) - \psi(\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\rangle)\right|\right) \\
&\quad \leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle|) \leq A\left|\langle \mathbf{w}_{kk'} - \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\rangle\right|\right) \\
&\quad \leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle|) \leq AB\left|\left|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right|\right|_1\right) \\
&\quad \leq \frac{AB}{a\sigma_{\text{dir}}} \cdot \left|\left|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right|\right|_1, \qquad \text{(F.2)}
\end{aligned}$$

其中，第一个不等式由三角不等式得出，第二个不等式由 ψ 是 A-利普希茨函数这一假设得出，第三个不等式由 Hölder不等式以及 $||\mathbf{z}||_\infty \leq B$ 几乎必然成立这一事实得出，最后一个不等式则由 $\mathbf{w}_{kk'} \in \mathcal{S}^d$ 以及$\mathbf{z}$的方向平滑性得出，同时还用到了Block 和 Simchowitz [2022, 引理 36]。代入第一个公式并求和可得

$$\begin{aligned}
\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right) &\leq \sum_{k,k' \in [K]} \frac{AB}{a\sigma_{\text{dir}}} \cdot \left|\left|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right|\right|_1 \\
&= \frac{AB}{a\sigma_{\text{dir}}} \left|\left|\theta_{\text{d}} - \theta'_{\text{d}}\right|\right|_1,
\end{aligned}$$

由此证明得证。 □

现在我们使用引理F.1来证明伪等距性质成立：

引理 F.2。假设我们处于定理 2 的情形，且 $\mathcal{M}$ 是由 σ_{dir}-方向光滑分布构成的类，其 $||\cdot||_\infty$ 几乎必然有界于 $B > 0$。那么

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta, \theta', \mathbf{z})\right] \leq \frac{2AB}{a\sigma_{\text{dir}}} \cdot \left|\left|\theta - \theta'\right|\right|_1.$$

证明。我们计算：

$$\begin{aligned}
\mathbb{E}_\nu \left[\rho(\theta, \theta', \mathbf{z})\right] &= \mathbb{E}_\nu \left[2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right] + \max_{k \in [K]} \left|\left|\theta_{\text{c}}^{(k)} - \theta_{\text{c}}^{\prime(k)}\right|\right|_1\right] \\
&\leq \frac{2AB}{a\sigma_{\text{dir}}} \left|\left|\theta_{\text{d}} - \theta'_{\text{d}}\right|\right|_1 + \max_{k \in [K]} \left|\left|\theta_{\text{c}}^{(k)} - \theta_{\text{c}}^{\prime(k)}\right|\right|_1 \\
&\leq \frac{2A(B \vee 1)}{a\sigma_{\text{dir}}} \cdot \left(\left|\left|\theta_{\text{d}} - \theta'_{\text{d}}\right|\right|_1 + \max_{k \in [K]} \left|\left|\theta_{\text{c}}^{(k)} - \theta_{\text{c}}^{\prime(k)}\right|\right|_1\right) \\
&\leq \frac{2A(B \vee 1)}{a\sigma_{\text{dir}}} \cdot \left|\left|\theta - \theta'\right|\right|_1,
\end{aligned}$$

其中，第一个不等式由期望的线性性和引理F.1得出。结论成立。

□

现在我们来控制 Θ 的广义括号数：

引理 F.3. 如果我们处于 定理 2 所述情形， $\epsilon > 0$,

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon) \leq \left(\frac{9AK^2B}{a\sigma_{\text{dir}}\epsilon}\right)^{K^2d}.$$

证明。 令 $\mathcal{N} = \{\theta_i = (\theta_{\text{c}}^i, \theta_{\text{d}}^i)\}$ 表示相对于 ℓ_1的一个 $\widetilde{\epsilon}$-网，其中 $\widetilde{\epsilon} = \frac{a\sigma_{\text{dir}}}{3K^2AB} \cdot \epsilon$，并令

$$\mathcal{B}_i = \{\theta \in \Theta | \, ||\theta - \theta_i||_1 \leq \widetilde{\epsilon}\}.$$

我们声称， $\{(\theta_i, \mathcal{B}_i)\}$ 构成相对于 $\mathcal{M}$的一个广义 ϵ-括号； $\mathcal{M}$是由 σ_{dir}-方向光滑分布构成的类，其 ℓ_∞ 范数有界于 B。为验证这一点，首先注意，根据 $\widetilde{\epsilon}$-网的定义， $\mathcal{B}_i$ 的并集覆盖了 Θ。现在，固定 θ_i,$\mathbf{z}$ ，并注意到对于 $\theta \in \mathcal{B}_i$，有

$$\begin{aligned}
\rho(\theta, \theta_i, \mathbf{z}) &= 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right] + \max_{k \in [K]} \left|\left|\theta_{\text{c}} - \theta_{\text{c}}^i\right|\right|_1 \\
&\leq 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right] + \left|\left|\theta_{\text{c}} - \theta_{\text{c}}^i\right|\right|_1.
\end{aligned}$$

现在我们来计算：

$$\begin{aligned}
&\mathbb{P}\left(\exists \theta \in \mathcal{B}_i \text{ s.t. } \overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right) \\
&\leq \mathbb{P}\left(\exists \theta \in \mathcal{B}_i, k, k' \in [K] \text{ s.t. } \psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \geq 0 > \psi(\langle \mathbf{w}_{kk'}^i, (\mathbf{z}, 1)\rangle)\right) \\
&\leq \mathbb{P}\left(\exists \theta \in \mathcal{B}_i, k, k' \in [K] \text{ s.t. } \left|\psi(\langle \mathbf{w}_{kk'}^i, (\mathbf{z}, 1)\rangle)\right| \leq AB\left|\left|\mathbf{w}_{kk'}^i - \mathbf{w}_{kk'}\right|\right|_1\right) \\
&\leq \mathbb{P}\left(\exists \theta \in \mathcal{B}_i, k, k' \in [K] \text{ s.t. } \left|\psi(\langle \mathbf{w}_{kk'}^i, (\mathbf{z}, 1)\rangle)\right| \leq AB\widetilde{\epsilon}\right) \\
&\leq \sum_{k,k' \in [K]} \mathbb{P}\left(\left|\psi(\langle \mathbf{w}_{kk'}^i, (\mathbf{z}, 1)\rangle)\right| \leq B\widetilde{\epsilon}\right) \\
&\leq \frac{K^2AB}{a\sigma_{\text{dir}}}\widetilde{\epsilon},
\end{aligned}$$

其中，第一个不等式由与 (F.1) 中相同的推理得出，第二个不等式由与 (F.2) 中相同的推理得出，第三个不等式由 $\mathcal{B}_i$的构造得出，第四个不等式由并集界得出，最后一个不等式则由方向平滑性的假设得出。因此，我们注意到，

$$\begin{aligned}
\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} \rho(\theta, \theta_i, \mathbf{z})\right] &= \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right] + \left|\left|\theta_{\text{c}} - \theta_{\text{c}}^i\right|\right|_1\right] \\
&\leq 2 \cdot \sup_{\nu \in \mathcal{M}}\left\{\mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right]\right]\right\} + \sup_{\theta \in \mathcal{B}_i}\left|\left|\theta_{\text{c}} - \theta_{\text{c}}^i\right|\right|_1 \\
&\leq \left(\frac{2K^2AB}{a\sigma_{\text{dir}}} + 1\right)\widetilde{\epsilon} \\
&\leq \epsilon.
\end{aligned}$$

因此，我们证明了 $\{(\theta_i, \mathcal{B}_i)\}$ 是关于 $\mathcal{M}$ 的广义 ϵ-括号。接下来需要界定其大小。为此，注意到根据构造，只需界定一个 $\widetilde{\epsilon}$-覆盖的大小，其

关于 ℓ_1 在 Θ。但注意， Θ 包含在一个 ℓ_1 半径为 D 的球中，因此通过一个简单的体积论证（例如参见 Vershynin [2018, 第 4.2 节。1] 例如），可知由于 $\Theta \subset \mathbb{R}^{Kd+K^2(d-1)}$，我们可以取

$$|\mathcal{N}| \leq \left(\frac{3D}{\widetilde{\epsilon}}\right)^{Kd+K^2(d-1)} \leq \left(\frac{9K^2DAB}{a\sigma_{\text{dir}}\epsilon}\right)^{K^2d}.$$

结论随即得出。 □

定理的证明2 可通过结合引理 F.3和 F.2得出。

F.2 推论的证明4.1

通过应用 定理 1 和 2，只需证明 ℓ 关于 ρ （4.3）中定义的对象是利普希茨的。然而，我们注意到

$$\begin{aligned}
&\overline{\ell}(\theta, \mathbf{z}) - \overline{\ell}(\theta', \mathbf{z}) \\
&= (\overline{\ell}(\theta, \mathbf{z}) - \overline{\ell}(\theta', \mathbf{z})) \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) = \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right] + (\overline{\ell}(\theta, \mathbf{z}) - \overline{\ell}(\theta', \mathbf{z})) \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right] \\
&\leq 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right] + \max_{k \in [K]} g_k(\theta_{\text{c}}, \mathbf{z}) - g_k(\theta'_{\text{c}}, \mathbf{z}) \\
&\leq 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right] + \max_{k \in [K]} \left|\left|\theta_{\text{c}}^{(k)} - \theta_{\text{c}}^{\prime(k)}\right|\right|_1.
\end{aligned}$$

根据 ρ 的定义，结论随即得出。

F.3 将ℓ 替换为 $\overline{\ell}$

尽管 附录 F.1 足以证明 定理 2，但为了规划，我们可能希望将损失函数 ℓ 替换为 $\overline{\ell}$ （4.1）中简单得多的形式。然而，为了应用我们的技术，我们需要 $\theta_{\text{d}} = (\mathbf{w}_1, \ldots, \mathbf{w}_K)$ 满足某个间隔条件。因此，附录 F.1 的对应结果如下：

定理 6. 假设 $\mathcal{Z} \subset \mathbb{R}^d$ 和 Θ 是欧几里得空间的子集，其 ℓ_1 直径不超过 D，并且

$$\Theta_{\text{d}} \subset \left\{(\mathbf{w}_1, \ldots, \mathbf{w}_K) \in (\mathcal{S}^d)^{\times K} \middle| \min_{k \neq k' \in [K]} \left|\left|\mathbf{w}_{k,\widehat{d}} - \mathbf{w}_{k',\widehat{d}}\right|\right|_2 \geq \gamma\right\},$$

其中，我们用 w_k 表示给定 $\theta_{\text{d}} \in \Theta_{\text{d}}$ 的坐标，并令 $w_{k\widehat{d}}$ 表示 w_k的前 d 个坐标。进一步假设，对于某个链接函数 ψ，如定理 2 中所述，有 $\phi(\theta_{\text{d}}, k, z) = \psi(\langle w_k, (z, 1)\rangle)$。如果 $\mathcal{M}$ 由 σ_{dir}-方向平滑分布构成，且对于某个 $B \geq 1$， $||z||_\infty \leq B$ 几乎必然成立，那么令 ρ 如 (4.3) 中所定义，则有： ρ 是一个满足伪等距性质的伪度量，具有 $\alpha = \frac{4AB}{a\gamma\sigma_{\text{dir}}}$ 和 $\beta = 1$。此外，对于所有 $\epsilon > 0$，

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon) \leq \left(\frac{18AK^2BD}{a\gamma\sigma_{\text{dir}}\epsilon}\right)^{2K(d+1)}$$

证明。 证明 定理 2 中给出的 附录 F.1基本相同。事实上，我们只需证明引理 F.1 的一个版本，证明的其余部分经过 相应的调整即可适用。为此，请注意，我们可以通过设定 $\mathbf{w}_{kk'} = \mathbf{w}_k - \mathbf{w}_{k'}$来仿照上述证明；这与之前的情形几乎相同，唯一的区别在于现在我们可以取 $||\mathbf{w}_{kk'}|| \neq 1$。这只会在应用方向平滑性时产生问题；因此，假设 $\mathbf{z}$ 是 σ_{dir}方向平滑的，并注意到（F.1）中的不等式链仍然成立。继续推导可知，对于固定的 $k,\ k' \in [K]$,

$$
\begin{aligned}
&\mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \geq 0 > \psi(\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\rangle)\right) \\
&\quad \leq \mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \leq \left|\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) - \psi(\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\rangle)\right|\right) \\
&\quad \leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle|) \leq A\left|\langle \mathbf{w}_{kk'} - \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\rangle\right|\right) \\
&\quad \leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle|) \leq AB\left|\left|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right|\right|_1\right) \\
&\quad \leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_k - \mathbf{w}_{k'}, (\mathbf{z}, 1)\rangle|) \leq AB(\left|\left|\mathbf{w}_k - \mathbf{w}'_k\right|\right|_1 + \left|\left|\mathbf{w}_{k'} - \mathbf{w}'_{k'}\right|\right|_1)\right) \\
&\quad \leq \frac{AB}{a\sigma_{\text{dir}} \cdot \left|\left|\mathbf{w}_{k,\widehat{d}} - \mathbf{w}_{k',\widehat{d}}\right|\right|_2} \cdot \left(\left|\left|\mathbf{w}_k - \mathbf{w}'_k\right|\right|_1 + \left|\left|\mathbf{w}_{k'} - \mathbf{w}'_{k'}\right|\right|_1\right) \\
&\quad \leq \frac{AB}{a\sigma_{\text{dir}}\gamma} \cdot \left(\left|\left|\mathbf{w}_k - \mathbf{w}'_k\right|\right|_1 + \left|\left|\mathbf{w}_{k'} - \mathbf{w}'_{k'}\right|\right|_1\right),
\end{aligned}
$$

其中，前四个不等式的推导与（F.2）中的相同，第五个不等式由 $\mathbf{w}_{kk'}$ 的定义得出，倒数第二个不等式的推导与前一个证明中的相同，最后一个不等式则由间隔假设得出。接下来，我们可以沿用定理 Theorem 2 证明中的相同思路，结果仍然成立，只需更改 Θ_{d} 的维度，并乘以因子 2，以计入对上文 $||\mathbf{w}_k - \mathbf{w}'_k||_1$ 差值求和两次的影响。 □

沿用推论 4.1 中的相同论证，可得定理 2 所述情形下算法的遗憾界，此处对应定理 Theorem 6:

推论 F.1。假设 ℓ 如 (4.1) 中所述，且 ϕ 和 Θ 如 定理 6 所述，其中 $B \geq 1$ 和 ℓ 的绝对值均一致有界于 1。若设 $\eta = \widetilde{\mathcal{O}}\left(\left(TKAdDB(\gamma a\sigma_{\text{dir}})^{-1}\right)^{2/3}\right)$ 且 $n = \sqrt{\eta}$，则算法2 的遗憾为

$$
\mathbb{E}\left[\text{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\left(\frac{TAKdBD}{a\gamma\sigma_{\text{dir}}}\right)^{2/3}\right).
$$

特别地，实现平均遗憾 ϵ 的预言机复杂度为 $\widetilde{\mathcal{O}}\left(\frac{AKdDB}{\gamma a\sigma_{\text{dir}}\epsilon^2}\right)$。

G 第4.3 节的证明

在本附录中，我们讨论多项式光滑假设，并给出满足这一要求的常见分布示例。我们还给出一个反例，说明方向平滑性并不足以确保多项式平滑性。随后，我们证明定理 3。

G.1 多项式平滑性

本节讨论定义4.1中的多项式平滑性概念。我们首先回顾 Glazer 和 Mikulincer 的简单 [2022, 示例 3], ，该示例表明 σ_{dir}- 平滑性严格弱于多项式平滑性：

例 G.1。我们证明，存在一个分布类，它是 σ_{dir}-方向光滑的，且 σ_{dir} 随维度按多项式衰减，但只有当σ_{poly},2随维度呈指数衰减时，才是σ_{poly},2-多项式平滑的。令 ν_d 表示单位欧氏球 $\mathcal{B}^d \subset \mathbb{R}^d$上的均匀测度。根据 Block 和 Simchowitz [2022, 例 1]，我们注意到 ν_d 是 σ_{dir}-方向光滑的，且 $\sigma_{\text{dir}} = \Omega\left(\frac{1}{d}\right)$。另一方面，对于多项式 $f(x) = \frac{1}{d} \cdot ||x||_2^2$，我们看到 $\text{coeff}_2(f) = 1$，但测度集中（例如参见 Vershynin [2018] ）告诉我们$\mathbb{P}(|f(x) - 1| \leq \epsilon) \geq 1 - e^{-\Omega(d)\epsilon}$。

现在我们来考察哪些类型的分布是多项式光滑的。我们所掌握的关键工具是著名的 Carbery-Wright 不等式（的一个特例），它指出：

定理 7 (Carbery 和 Wright [2001]的定理 8。如果 ν 是定义在 $\mathbb{R}^d$上的对数凹测度，且 $f: \mathbb{R}^d \to \mathbb{R}$ 是一个次数为 r 的多项式，那么对于所有 $\epsilon > 0$，若 $X \sim \nu$,

$$\mathbb{P}(|f(X)| \leq \epsilon) \leq Cr \frac{\epsilon^{\frac{1}{r}}}{\mathbb{E}[f(X)^2]^{\frac{1}{2r}}}.$$

因此，定理 7 告诉我们，如果 ν 是对数凹的，并且可以确定所有系数$r(f) \geq 1$ 较大的多项式 f 都具有较大的二阶矩，那么 ν 是 $\sigma_{\text{poly},\, r}$-多项式光滑的，其中 $\sigma_{\text{poly},\, r}$随维度良好地变化。证明系数较大的多项式确实具有较大的二阶矩，仍是一个活跃的研究领域；下面我们提供一个例子，并将其改写成我们的表述：

定理 8 (Glazer 和 Mikulincer [2022]的推论 4。假设 $\nu = \mu^{\otimes n}$ 是一个对数凹的各向同性乘积测度。那么 ν是 $\sigma_{\text{poly},\, r}$-多项式光滑的，其中 $\sigma_{\text{poly},\, r} \geq \Omega\left(\frac{1}{r}\right)$。

请注意，上述结果涵盖高斯测度，并可根据需要进行缩放。沿着Glazer 和 Mikulincer [2022] 的思路取得的进一步结果，可直接推广到一类更广泛的已知为 $\sigma_{\text{poly},r}$-多项式光滑的测度。

G.2 定理3证明

本节将按照推论 4.1的思路给出定理 3的证明。尽管我们不再重复论证，但顺带指出，将锦标赛式 ℓ 替换为更简单的函数 ℓ （见式4.1），并加入间隔假设，即可在此情形下得到定理 6 的类似结果。为了证明定理 3，我们先证明定理 2的类似结果：

定理 *9*。假设 $\mathcal{Z} \subset \mathbb{R}^d$，且 Θ 是欧几里得空间的子集，其 ℓ_1 直径有界于 D。令 Θ_{d} 对定义在 $\mathbb{R}^d$ 上、次数为 r 的多项式组 $(f_{\mathbf{w}_{kk'}})$ 进行参数化，使得对所有 $k \in [K]$, $\textit{coeff}_r(f_{\mathbf{w}_{kk'}}) = 1$。若 $\phi(\theta_{\text{d}}, k, k', z) = f_{\mathbf{w}_{kk'}}(z)$，且 $\mathcal{M}$ 是满足 $||z||_\infty \leq B$ 几乎必然成立的 $\sigma_{\text{poly},r}$- 多项式光滑分布类，则 *(4.3)* 中定义的 ρ 是

一个满足伪等距性质的伪度量，其中 $\alpha = \frac{2B^r D}{\sigma_{\text{poly}\,r}}$，且 $\beta = \frac{1}{r}$。此外，对所有 $\epsilon > 0$，

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon) \leq \left(\frac{9K^2 B}{\sigma_{\text{poly},r}\epsilon}\right)^{K^2 r^2 d^r}.$$

为证明该结论，我们需要引理 F.1、F.2 和 F.3 的类似结论。我们首先证明 ρ 中第一项在期望意义下的 Lipschitz 性：

引理 G.1。假设 Θ_{d}、ϕ 和 k_ϕ 与定理 9 中的条件相同，并假设$\mathbf{z}$ 取自一个 $\sigma_{\text{poly},r}$-多项式平滑分布，且 $||\mathbf{z}||_\infty \leq B$ 几乎必然成立，其中 $B \geq 1$。那么，

$$\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right) \leq \frac{B^r K^{2-\frac{4}{r}}}{\sigma_{\text{poly},r}} \cdot ||\theta_{\text{d}} - \theta'_{\text{d}}||_1^{\frac{1}{r}}.$$

证明。根据式（F.1）中的相同论证，我们有

$$\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right) \leq \sum_{k,k' \in [K]} \mathbb{P}\left(f_{\mathbf{w}_{kk'}}(\mathbf{z}) \geq 0 > f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right).$$

注意，由三角不等式，

$$\left|f_{\mathbf{w}_{kk'}}(\mathbf{z}) - f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right| \leq B^r \cdot ||\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}||_1.$$

因此，应用式（F.2）中的论证，可得

$$\begin{aligned}
\mathbb{P}\left(f_{\mathbf{w}_{kk'}}(\mathbf{z}) \geq 0 > f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right) &\leq \mathbb{P}\left(\left|f_{\mathbf{w}_{kk'}}(\mathbf{z})\right| \leq \left|\left|f_{\mathbf{w}_{kk'}}(\mathbf{z}) - f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right|\right|\right) \\
&\leq \mathbb{P}\left(\left|f_{\mathbf{w}_{kk'}}(\mathbf{z})\right| \leq B^r \cdot ||\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}||_1\right) \\
&\leq \frac{B}{\sigma_{\text{poly},r}} \cdot ||\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}||_1^{\frac{1}{r}},
\end{aligned}$$

其中，最后一个不等式由多项式平滑性的定义得出。应用 Hölder不等式并求和，即可完成证明。□

利用这一结果，我们可以证明引理F.2的一个类似结论：

引理 G.2。假设我们处于定理 9的情形，并且 $\mathcal{M}$ 是由 $\sigma_{\text{poly}、r}$-多项式光滑分布构成的类，这些分布的样本的无穷范数几乎必然一致有界于某个 $B \geq 1$。如果 ℓ_1 Θ 的直径被 D 所界定，那么

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', \mathbf{z})\right] \leq \frac{2BD}{\sigma_{\text{poly},r}} \cdot ||\theta - \theta'||^{\frac{1}{r}}.$$

证明。我们计算：

$$\begin{aligned}
\mathbb{E}_\nu\left[\rho(\theta, \theta', \mathbf{z})\right] &= \mathbb{E}_\nu\left[2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\text{d}}, \mathbf{z})\right] + \max_{k \in [K]} \left|\left|\theta_{\text{c}}^{(k)} - \theta_{\text{c}}^{\prime(k)}\right|\right|_1\right] \\
&\leq \frac{2B}{\sigma_{\text{poly},r}} \cdot ||\theta_{\text{d}} - \theta'_{\text{d}}||_1^{\frac{1}{r}} + ||\theta_{\text{c}} - \theta'_{\text{c}}||_1 \\
&\leq \frac{2BD}{\sigma_{\text{poly},r}} \cdot ||\theta - \theta'||^{\frac{1}{r}},
\end{aligned}$$

其中，第二个不等式由引理 G.1 得出，最后一个不等式则由关于直径的假设得出。□

最后，我们需要引理 F.3的一个类似版本：

引理 G.3. 若处于定理 9 的情形，则对任意 $\epsilon > 0$，均有

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \epsilon) \leq \left(\frac{9K^2B}{\sigma_{\text{poly},r}\epsilon}\right)^{K^2r^2d^r}.$$

证明。 我们仿照引理F.3 的证明，但改用引理G.1 而非引理F.1。具体而言，我们假设 $\mathcal{N} = \left\{\theta_i = \left(\theta_{\text{c}}^i, \theta_{\text{d}}^i\right)\right\}$ 是一个 $\widetilde{\epsilon}$-网，覆盖 Θ，相对于 ℓ_1，其中 $\widetilde{\epsilon} = \left(\frac{\sigma_{\text{poly}}}{3K^2B^r} \cdot \epsilon\right)^r$ 并类似地令 $\mathcal{B}_i \subset \Theta$ 表示 $\widetilde{\epsilon}$ 中距 θ_i 不超过 ℓ_1 的参数集合，按F.3 引理的证明中的方法计算可得

$$\begin{aligned}
\mathbb{P}\left(\exists\theta \in \mathcal{B}_i \text{ s.t. } \overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right) &\leq \mathbb{P}\left(\exists\theta \in \mathcal{B}_i \text{ s.t. } \left|f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right| \leq \left|f_{\mathbf{w}_{kk'}}(\mathbf{z}) - f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right|\right) \\
&\leq \mathbb{P}\left(\exists\theta \in \mathcal{B}_i \text{ s.t. } \left|f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right| \leq B^r \cdot \left|\left|\mathbf{w}_{kk'} - \mathbf{w}_{kk'}^i\right|\right|_1\right) \\
&\leq \mathbb{P}\left(\exists\theta \in \mathcal{B}_i \text{ s.t. } \left|f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right| \leq B^r \cdot \widetilde{\epsilon}\right) \\
&\leq \frac{K^2B}{\sigma_{\text{poly},r}} \cdot \widetilde{\epsilon}^{\frac{1}{r}}.
\end{aligned}$$

Thus

$$\begin{aligned}
\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta_i, \mathbf{z})\right] &\leq 2 \cdot \sup_{\nu\in\mathcal{M}}\left\{\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i} \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\text{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\text{d}}^i, \mathbf{z})\right]\right]\right\} + \sup_{\theta\in\mathcal{B}_i}\left|\left|\theta_{\text{c}} - \theta_{\text{c}}^i\right|\right|_1 \\
&\leq \frac{2K^2B}{\sigma_{\text{poly},r}} \cdot \widetilde{\epsilon}^{\frac{1}{r}} + \widetilde{\epsilon} \\
&\leq \frac{3K^2B}{\sigma_{\text{poly},r}} \cdot \widetilde{\epsilon}^{\frac{1}{r}} \\
&\leq \epsilon.
\end{aligned}$$

因此， $\mathcal{N}$ 是相对于 ϵ的广义 $\mathcal{M}$ -括号，且 ρ。按照引理 $\mathcal{N}$ 的证明中的相同方法，我们可以给F.3，注意到$\mathbf{w}_{kk'}$ 所在空间的维度为

$$\sum_{i\leq r}\binom{d}{i} \leq \left(\frac{ed}{r}\right)^r.$$

□

结合引理G.2 和G.3 即可完成定理 9的证明。现在我们可以证明定理 3：

定理 *3* 的证明。 按照推论 4.1 的证明中相同的推理，我们观察到 ℓ 关于 ρ 是利普希茨连续的。因此，我们可以应用 定理 1 和 定理 9 ，得到如果运行算法 2 ，那么

$$\mathbb{E}\left[\text{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\eta + \frac{T}{n}K^2r^2d^r\log\left(\frac{1}{\sigma_{\text{poly},r}}\right) + \frac{TB^rD}{\sigma_{\text{poly},r}} \cdot \left(\frac{K^2r^2d^r}{\eta}\right)^{\frac{1}{4r-2}}\right).$$

设定

$$\eta = \widetilde{\Theta}\left(\left(\frac{TK^2r^2d^rDB}{\sigma_{\text{poly},r}}\right)^{\frac{4r-2}{4r-1}}\right) \qquad n = \widetilde{\Theta}\left(\left(\frac{TK^2r^2d^rDB}{\sigma_{\text{poly},r}}\right)^{\frac{2r-1}{4r-1}}\right)$$

证明到此结束。 □

H 定理4 的证明

本节将陈述并证明 定理 4 的一个正式版本。回顾一下，我们处于（5.1）的情形，目标是相对于最佳计划 $\bar{\mathbf{u}}_{1:H}$ 最小化遗憾。以下我们令 $\|\cdot\|_1$ 表示 ℓ_1 范数，并按拼接向量的自然方式解释；例如 $\|\bar{\mathbf{u}}_{1:H}\|_1 = \sum_{h=1}^{H}\|\mathbf{u}_h\|_1$。我们首先引入一个记号，以大幅简化后续表述：

定义 H.1. 给定模式序列 $k_{1:H} \in [K]^H$，根据（5.1）回顾，状态的演化方式如下

$$\bar{\mathbf{x}}_{t,\,h+1}(\theta;\,k_{1:H}) = g_{t,h,k_h}(\tilde{\mathbf{x}}_{t,h}(\theta;k_{1:H}),\mathbf{u}_{t,h}(\theta)) + \boldsymbol{\eta}_{t,h},\quad \tilde{\mathbf{x}}_{t,1}(\theta;k_{1:H}) = \mathbf{x}_{t,1},\quad \mathbf{u}_{t,h}(\theta) = \bar{\mathbf{u}}_h + \boldsymbol{\xi}_{t,h},$$

这里与（5.1）中的情形不同之处在于，此处模式序列是给定的，而在（5.1）中，模式序列取决于状态和输入。我们定义函数

$$G_t(\theta;k_{1:H}) := (\tilde{\mathbf{x}}_{t,1}(\theta;k_{1:H}),\ldots,\tilde{\mathbf{x}}_{t,H}(\theta;k_{1:H})) \in \mathcal{X}^H \subset \mathbb{R}^{mH},$$

它将一个计划和给定的模式序列映射到相应的轨迹。

现在，我们可以正式陈述定理 4：

定理 10。对于固定的规划时域 H，假设轨迹$\mathbf{x}_{t,\,1:H}$ 按照(5.1)演化，其中学习者在每个时刻 $\theta \in \mathcal{K} \subset \mathcal{U}^{\times H}$ 选择一个计划 t，而对手给出元组$\mathbf{z}_t$，如第 5 节所述。假设对于所有 $t \in [T]$，以下性质在对手策略下几乎必然成立 p_t：

1. $\mathbf{x}_{t,\,1} \mid \mathcal{F}_t$ 以及 $(\boldsymbol{\eta}_{t,\,h},\boldsymbol{\xi}_{t,\,h}) \mid \mathcal{F}_{t,\,h-1}$ 均为 σ_{dir}-方向光滑。
2. 对于所有模式序列 $k_{1:H} \in [K]^H$以及 $\theta,\theta' \in \mathcal{U}^{\times H}$, $||G_t(\theta,k_{1:H}) - G_t(\theta',k_{1:H})|| \le L\,||\theta-\theta'||_1$, 也就是说，函数 G_t 关于 ℓ_1 范数是 L-利普希茨的。
3. 对所有 $h \in [H]$，$\sup_{\theta\in\mathcal{K}}||\mathbf{u}_{t,\,h}(\theta)||_1 \vee ||\mathbf{x}_{t,\,h}(\theta)||_1 \le D$。
4. 存在某个 $\gamma > 0$，使得对所有 $h \in [H]$，$\min_{k\neq k'}\left|\left|\mathbf{w}_{t,\,h,\,k,\,\widehat{d}} - \mathbf{w}_{t,\,h,\,k',\,\widehat{d}}\right|\right|_2 \ge \gamma$ 成立，其中我们令$\mathbf{w}_{\widehat{d}}$表示向量$\mathbf{w} \in \mathbb{R}^{d+1}$的前 d 个坐标。
5. 对于所有 $\mathbf{v}_{1:H},\mathbf{v}'_{1:H} \in \mathcal{V}^H$且满足 $\|\mathbf{v}_{1:H}\|_1 \vee \|\mathbf{v}'_{1:H}\|_1 \le 2D$ 的情况，损失函数 ℓ_t^v关于 ℓ_1范数是利普希茨连续且有界的，即 $|\ell_t^v(\mathbf{v}_{1:H}) - \ell_t^v(\mathbf{v}'_{1:H})| \le \|\mathbf{v}_{1:H} - \mathbf{v}_{1:H}\|_1$且 $|\ell_t^v(\mathbf{v}_{1:H})| \le 1$。

如果规划器按照算法 θ_t 2并使用 $\eta = d^{1/3}H^{5/3}K^{4/3}\left(\frac{TLD}{\gamma\sigma_{\text{dir}}}\right)^{2/3}$ 和 $n = \sqrt{\eta}$进行决策，则

$$\mathbb{E}\left[\text{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(d^{1/3}H^{5/3}K^{4/3}\left(\frac{TLd}{\gamma\sigma_{\text{dir}}}\right)^{2/3}\right).$$

因此，达到平均遗憾 ϵ 所需的预言机复杂度为 $\widetilde{\mathcal{O}}\left(d^{1/3}D^{2/3}H^{5/3}K^{4/3}L^{2/3}(\gamma\sigma_{\text{dir}})^{-2/3}\epsilon^{-2}\right)$。

注 H.1（ L 和 D的缩放）。请注意，参数 L 和 D 的缩放取决于动态映射 g_t、$_k$、$_h$的 H 重复复合。因此，在上述定理的第二个假设中，要求映射 G_t 为 L-利普希茨映射，在最坏情况下，如果我们仅假设函数 g_t、$_k$、$_h$ 是 Λ-利普希茨映射，那么朴素地说， L 可能按 $\text{poly}(H)\cdot(\Lambda)^H$的规模增长。同样，如果我们仅假设 g_t、$_k$、$_h(\mathbf{v}) \leq c_2\|\mathbf{v}\| + c_1$，那么上述第三个也是最后一个假设中的界 D 可能按 $(c_2)^H$的规模增长。尽管一般而言这些界是紧的，但通过常见的稳定性假设可以减轻对指数的依赖；这类假设常用于控制领域。例如，在复合动力学 [满足增量稳定性的条件下，Pfrommer 等人，2022，Angeli， 2000]、 L的规模将仅呈多项式增长于 H。Jadbabaie 和 Hauser [2001]中提出的输入到状态稳定性等其他概念，则可得到多项式有界的 D。因此，在许多值得关注的实际场景中，参数 L、 D 相对于所有相关问题参数都仅呈多项式增长。请注意，在流行的线性二次调节器框架中（我们的设定是对其的广泛推广），这些稳定性假设是标准的 [Hazan 和 Singh， 2022]。

与之前对定理1的应用一样，我们将证明一个类似于定理 2的结果，在其中引入一个伪度量 ρ 并证明它满足伪等距性质，且能够控制广义括号数。随后，我们将证明我们的 ℓ 关于 ρ 是利普希茨的，并借助 定理 1得出结论。为简单起见，我们暂时省略指标 t，并比较给定的计划 $\bar{\mathbf{u}}_{1:H}$ 及其对应的动力学 $\mathbf{x}_{1:H}$ 与另一个计划$\bar{\mathbf{u}}'_{1:H}$及其对应的动力学 $\mathbf{x}'_{1:H}$，其中这些动力学共享噪声序列 $\eta_{1:H}$， $\xi_{1:H}$。我们将简写为

$$\rho(\theta,\theta'') = \rho(\theta,\theta',\boldsymbol{\eta}_{1:H},\boldsymbol{\xi}_{1:H},\theta) = \left|\left|\theta-\theta'\right|\right|_1 + \sum_{h=1}^{H}\left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|. \tag{H.1}$$

我们还将简写 为 $k_h = k_h(\mathbf{v}_h)$ 和 $k'_h = k_h(\mathbf{v}'_h)$。我们首先证明以下引理：

引理 H.1. 考虑事件

$$\mathcal{A}_h = \left\{k_{h'} = k'_{h'} \text{ for all } h' < h \text{ and } k_h \neq k'_h\right\}.$$

于是，对所有 $h \in [H]$,

$$\left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|_1 \leq L\cdot\left|\left|\theta-\theta'\right|\right|_1 + 2D\cdot\sum_{h'=1}^{h-1}\mathbb{I}\left[\mathcal{A}_{h'}\right].$$

证明。 由（5.1），可得

$$\begin{aligned}
\left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|_1 &= \left|\left|g_{k_{h-1}}(\mathbf{v}_{h-1}) - g_{k'_{h-1}}(\mathbf{v}'_{h-1})\right|\right|_1 \\
&\leq \left|\left|g_{k_{h-1}}(\mathbf{v}_{h-1}) - g_{k_{h-1}}(\mathbf{v}'_{h-1})\right|\right|_1 \cdot \mathbb{I}\left[\bigcup_{h'\geq h} \mathcal{A}_{h'}\right] + 2D \cdot \mathbb{I}\left[\bigcup_{h'<h} \mathcal{A}_{h'}\right] \\
&\leq \left|\left|G(\theta, k_{1:H}) - G\left(\theta', k'_{1:H}\right)\right|\right| + 2D \cdot \mathbb{I}\left[\bigcup_{h'<h} \mathcal{A}_{h'}\right] \\
&\leq L \cdot \left|\left|\bar{\mathbf{u}}_{1:H} - \bar{\mathbf{u}}'_{1:H}\right|\right|_1 + 2D \sum_{h'=1}^{h-1} \mathbb{I}\left[\mathcal{A}_{h'}\right],
\end{aligned}$$

其中等号由构造可知，第一个不等式由 $\mathcal{Z}$的有界性得出，第二个不等式由以下事实得出：对于 $\mathcal{A}_{h'}$， $h' \geq h$ 的定义意味着$k_{h''}$ 对应的 $h'' < h$ 模式与$k'_{h''}$相同；最后一个不等式则由并集界以及定理 10中的第二个条件得出。结论成立。 □

现在我们可以证明伪等距性质了：

引理 H.2。 设 $\mathcal{M}$ 表示由定理 *10* 中的设定诱导的分布类，并令 ρ 如(H.1)所示。则 ρ 相对于 $||\cdot||_1$ 满足伪等距性质，其中 $\alpha = \frac{6DH^2K^2L}{\gamma\sigma_{\text{dir}}}$ 且 $\beta = 1$，即

$$\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta,\theta')\right] \leq \frac{6DH^2K^2L}{\gamma\sigma_{\text{dir}}} \cdot \left|\left|\theta - \theta'\right|\right|_1.$$

证明。 由 (H.1) 和引理 H.1，可得

$$\begin{aligned}
\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta,\theta')\right] &= \sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\left|\left|\theta - \theta'\right|\right|_1 + \sum_{h=1}^{H} \left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|\right] \\
&\leq 2LH \cdot \left|\left|\theta - \theta'\right|\right|_1 + 2DH \cdot \sup_{\nu\in\mathcal{M}} \sum_{h=1}^{H} \mathbb{P}_\nu(\mathcal{A}_h).
\end{aligned}$$

现在计算：

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_h) &= \mathbb{P}\left(k_{h'} = k_{h'} \text{ for } h' < h \text{ and } k_h \neq k_h\right) \\
&= \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \arg\max_{k\in[K]} \left\langle \mathbf{w}^\star_{k,h}, (\mathbf{v}_h, 1)\right\rangle \neq \arg\max_{k'\in[K]} \left\langle \mathbf{w}^\star_{k',h}, (\mathbf{v}'_h, 1)\right\rangle\right) \\
&\leq \sum_{k\neq k'\in[K]} \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle \geq 0 > \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}'_h, 1)\right\rangle\right),
\end{aligned} \tag{H.2}$$

其中论证与（F.1）中的类似。对于固定的 $k \neq k' \in [K]$，接下来我们计算

$$\begin{aligned}
&\mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle \geq 0 > \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}'_h, 1)\right\rangle\right) \\
&\leq \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle\right| \leq \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h - \mathbf{v}'_h, 0)\right\rangle\right|\right) \\
&\leq \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle\right| \leq 2 \cdot \left|\left|\mathbf{v}_h - \mathbf{v}'_h\right|\right|_1\right),
\end{aligned}$$

其中，第一个不等式由三角不等式得出，第二个不等式由 Hölder不等式得出。现在我们注意到

$$\left|\left|\mathbf{v}_h - \mathbf{v}'_h\right|\right|_1 \leq \left|\left|\mathbf{u}_h - \mathbf{u}'_h\right|\right|_1 + \left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|_1,$$

并且，此外，由于模式序列 $k_{1:h-1} = k'_{1:h-1}$，在这种情况下我们有

$$\left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|_1 \leq \left|\left|G(\theta, k_{1:H}) - G(\theta', k_{1:H})\right|\right|_1 \leq L \cdot \left|\left|\theta - \theta'\right|\right|_1. \tag{H.3}$$

因此，我们有

$$\begin{aligned}
&\mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle\right| \leq 2 \cdot \left|\left|\mathbf{v}_h - \mathbf{v}'_h\right|\right|_1\right) \\
&\leq \mathbb{P}\left(\left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, \mathbf{v}_h\right\rangle\right| \leq 3L \cdot \left|\left|\theta - \theta'\right|\right|_1\right) \\
&= \mathbb{P}\left(\left|\left\langle \frac{\mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}}{\left|\left|\mathbf{w}^\star_{k,h,\widehat{d}} - \mathbf{w}^\star_{k',h,\widehat{d}}\right|\right|_2}, (\mathbf{v}_h, 1)\right\rangle\right| \leq \frac{3L}{\left|\left|\mathbf{w}^\star_{k,h,\widehat{d}} - \mathbf{w}^\star_{k',h,\widehat{d}}\right|\right|_2} \cdot \left|\left|\theta - \theta'\right|\right|_1\right) \\
&\leq \frac{3L}{\sigma_{\text{dir}} \cdot \left|\left|\mathbf{w}^\star_{k,h,\widehat{d}} - \mathbf{w}^\star_{k',h,\widehat{d}}\right|\right|_2} \cdot \left|\left|\theta - \theta'\right|\right|_1 \\
&\leq \frac{3L}{\sigma_{\text{dir}}\gamma} \cdot \left|\left|\theta - \theta'\right|\right|_1,
\end{aligned} \tag{H.4}$$

其中，第一个不等式由前面的计算得出，等式显然成立，第二个不等式由方向平滑性假设得出，最后一个不等式则由间隔假设得出。代回后即可得到结论。

□

最后，我们证明括号数可以得到控制：

引理 H.3。 设 $\mathcal{M}$ 和 ρ 分别如引理 *H.2* 和 $\mathcal{K}$ 如 *定理 10*中所述。则对于任意 $\epsilon > 0$，均有

$$\mathcal{N}_{\mathcal{M},[]}(\mathcal{K}, \rho, \epsilon) \leq \left(\frac{36K^2DH^2L}{\gamma\sigma_{\text{dir}}\epsilon}\right)^{H(d+1)}.$$

证明。令 $\mathcal{N} = \{\theta^i\} \subset \mathcal{U}^{\times H}$ 表示相对于 ℓ_1的一个 $\widetilde{\epsilon}$-网，其中 $\widetilde{\epsilon} = \frac{\gamma\sigma_{\text{dir}}}{12DH^2L} \cdot \epsilon$。与引理 F.3 和 G.3 中一样，我们将证明：若 $\mathcal{B}_i$ 表示与 θ^i的距离至多为 $\widetilde{\epsilon}$ 的 θ 集合，则 $\{(\theta^i, \mathcal{B}_i)\}$ 构成相对于 $\mathcal{M}$ 的广义 ϵ- 括号。将引理 F.1 和 G.1 替换为引理 H.1 以及引理 H.2 中的论证后，证明过程基本相同。具体而言，对于固定的 i 和任意的 $\nu \in \mathcal{M}$，我们有

$$\begin{aligned}
\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i} \rho(\theta, \theta^i)\right] &= \mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i} \left|\left|\theta - \theta^i\right|\right|_1 + \sum_{h=1}^H \left|\left|\mathbf{x}_h - \mathbf{x}^i_h\right|\right|_1\right] \\
&\leq \widetilde{\epsilon} + \mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i} \sum_{h=1}^H \left|\left|\mathbf{x}_h - \mathbf{x}^i_h\right|\right|_1\right],
\end{aligned}$$

其中我们令 $\mathbf{x}^i_{1:H}$ 表示在 θ^i下演化得到的动力学。对于第二项，我们应用引理H.1并计算：

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\sum_{h=1}^{H}\left|\left|\mathbf{x}_h-\mathbf{x}^i_h\right|\right|_1\right] &\le \sum_{h=1}^{H}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\left|\left|\mathbf{x}_h-\mathbf{x}^i_h\right|\right|_1\right]\\ &\le \sum_{h=1}^{H}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}L\cdot\left|\left|\theta-\theta^i\right|\right|+2D\cdot\sum_{h'=1}^{h-1}\mathbb{I}\left[\mathcal{A}_{h'}\right]\right]\\ &\le LH\widetilde{\epsilon}+2DH^2\cdot\max_{h\in[H]}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\mathbb{I}\left[\mathcal{A}_h\right]\right].\end{aligned}$$

现在，我们以与引理 H.2类似的方式进行推理：

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\mathbb{I}\left[\mathcal{A}_h\right]\right] &= \mathbb{P}\left(\bigcup_{\bar{\mathbf{u}}_{1:H}\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'}\text{ for }h'<h\text{ and }k_h\neq k^i_h\right\}\right)\\ &\le \sum_{k\neq k'\in[K]}\mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'}\text{ for }h'<h\text{ and }\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 2\cdot\left|\left|\mathbf{v}_h-\mathbf{v}^i_h\right|\right|_1\right\}\right),\end{aligned}$$

其中，不等式是沿用（H.2）和（H.2）中的同一套推理得到的。与（H.3）一样，我们注意到，如果模式序列 $k_{h'}=k^i_{h'}$ 对于 $h'<h$，那么

$$\left|\left|\mathbf{v}_h-\mathbf{v}^i_h\right|\right|_1\le 2L\cdot\left|\left|\theta-\theta^i\right|\right|_1.$$

因此，根据 $\mathcal{B}_i$ 的构造，我们看到

$$\begin{aligned}&\mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'}\text{ for }h'<h\text{ and }\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 2\cdot\left|\left|\mathbf{v}_h-\mathbf{v}^i_h\right|\right|_1\right\}\right)\\ &\le \mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 6L\cdot\left|\left|\theta-\theta^i\right|\right|_1\right\}\right)\\ &\le \mathbb{P}\left(\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 6L\widetilde{\epsilon}\right)\\ &\le \frac{6L\widetilde{\epsilon}}{\gamma\sigma_{\text{dir}}},\end{aligned}$$

其中，最后一个不等式由方向平滑性和间隔假设得出，与（H.4）中的论证相同。代回 $\widetilde{\epsilon}$ 的定义可知，$\{(\theta^i,\ \mathcal{B}_i)\}$ 构成了所需的广义 ϵ-括号。注意，$\mathcal{K}$ 位于 ℓ_1 球内，其半径为 D，且该球包含于 $\mathbb{R}^{dH}$；因此，应用与 定理 2 中相同的论证即可完成证明。 □

最后，我们可以证明主要结果了：

定理 *10* 的证明。 根据 定理 1，只需证明损失关于 ρ是利普希茨连续的。

为此，注意到

$$\begin{aligned}\ell(\mathbf{z}_t) - \ell(\mathbf{z}'_t) &\le \left|\left|\mathbf{v}_{t,1:H} - \mathbf{v}'_{t,1:H}\right|\right|_1 \\ &\le \left|\left|\theta - \theta'\right|\right|_1 + \sum_{h=1}^{H} \left|\left|\mathbf{x}_{t,h} - \mathbf{x}'_{t,h}\right|\right|_1 \\ &= \rho(\theta, \theta').\end{aligned}$$

因此，根据 定理 1，若学习者使用算法 2 且 $n = \sqrt{\eta}$，则其会经历

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \widetilde{\mathcal{O}}\left(\eta + \frac{TDH^2K^2L}{\gamma\sigma_{\mathrm{dir}}} \cdot \sqrt{\frac{Hd}{\eta}}\right)$$

借助引理H.2 和H.3。按照定理陈述中的设置取 η，即可完成证明。

□

最后，我们证明一个与 定理 10 类似的结果，但这里我们假设模式之间的决策边界是多项式。该陈述与 定理 10 几乎等价，区别在于现在模式边界由多项式参数化，因此预言机复杂度也相应增加，具体方式与 定理 3 类似。陈述如下：

定理 11。 假设我们处于 (5.1) 所述情形，但区域由次数至多为 r 的多项式定义。更准确地说，我们假设

$$\begin{aligned}\mathbf{x}_{t,h+1}(\theta) &= g_{t,h,k_{t,h}(\mathbf{v}_{t,h}(\theta))}(\mathbf{v}_{t,h}(\theta)) + \boldsymbol{\eta}_{t,h}, \quad \textit{and} \\ \mathbf{u}_{t,h}(\theta) &= \bar{\mathbf{u}}_{t,h} + \boldsymbol{\xi}_{t,h}, \quad \mathbf{v}_{t,h}(\theta) = (\mathbf{x}_{t,h}(\theta), \quad \mathbf{u}_{t,h}(\theta)), \\ k_{t,h}(\mathbf{v}) &= \arg\max_{k\in[K]} \phi_{t,h}(k, \mathbf{v}), \quad \textit{and} \quad \phi_{t,h}(k, \mathbf{v}) = f_{\mathbf{w}_{t,k,h}}(\mathbf{v}),\end{aligned}$$

其中，$f_{\mathbf{w}_{t,\,k,\,h}}$ 是次数为 r 的多项式，$\mathbf{w}_{t,\,k,\,h}$ 对其系数进行参数化。假设对于所有 $t \in [T]$，在对手策略 p_t下，以下性质几乎必然成立：

1. $\mathbf{x}_{t,\,1} \mid \mathcal{F}_t$ 和 $(\boldsymbol{\eta}_{t,\,h}, \boldsymbol{\xi}_{t,\,h}) \mid \mathcal{F}_{t,\,h-1}$ 均为 $\sigma_{\mathrm{poly},\,r}$-多项式光滑。

2. 对于所有模式序列 $k_{1:H} \in [K]^H$和 θ、$\theta' \in \mathcal{U}^{\times H}$、$||G_t(\theta,\ k_{1:H}) - G_t(\theta',\ k_{1:H})|| \le L\,||\theta - \theta'||_1$，即，函数 G_t 关于 L范数均为 -利普希茨，其中 ℓ_1，G_t 是定义 *H.1* 中定义的映射。

3. 对于所有 $h \in [H]$，$\sup_{\theta\in\mathcal{K}} ||\mathbf{u}_{t,\,h}(\theta)||_1 \vee ||\mathbf{x}_{t,\,h}(\theta)||_1 \le D$。

4. 对于某个 $\gamma > 0$，对所有 $h \in [H]$ 均有 $\min^{k\neq k'\in[K]}\mathrm{coeff}_r\left(f_{\mathbf{w}_{t,k,h}} - f_{\mathbf{w}_{t,k',h}}\right) \ge \gamma$，其中$\mathrm{coeff}_r(\cdot)$ 的定义见定义 *4.1*。

5. 对于所有 $\mathbf{v}_{1:H}$、$\mathbf{v}'_{1:H} \in \mathcal{V}^H$，满足 $\|\mathbf{v}_{1:H}\|_1 \vee \|\mathbf{v}'_{1:H}\|_1 \le 2D$，损失函数 ℓ^v_t关于 ℓ_1范数是利普希茨连续的且有界，即 $|\ell^v_t(\mathbf{v}_{1:H}) - \ell^v_t(\mathbf{v}'_{1:H})| \le \|\mathbf{v}_{1:H} - \mathbf{v}_{1:H}\|_1$，并且 $|\ell^v_t(\mathbf{v}_{1:H})| \le 1$。

6. 多项式 $f_{\mathbf{w}_{t,\,k和h}}$ 的系数 $\mathbf{w}_{t,\,k和h}$ 具有单位欧几里得范数。

如果规划器根据算法 2并结合 $\eta = \left(\frac{LTK^2r^2H^{2+r}d^rDB}{\gamma\sigma_{\text{poly},r}}\right)^{\frac{4r-2}{4r-1}}$和 $n = \sqrt{\eta}$进行决策，则

$$\mathbb{E}\left[\text{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\left(\frac{LTK^2r^2H^{2+r}d^rDB}{\gamma\sigma_{\text{poly},r}}\right)^{\frac{4r-2}{4r-1}}\right).$$

因此，实现平均遗憾 ϵ 所需的预言机复杂度为 $\widetilde{\mathcal{O}}\left(\left(\frac{LK^2r^2H^{2+r}d^rDB}{\gamma\sigma_{\text{poly},r}}\right)^{\frac{4r-2}{4r-1}}\right)\cdot\epsilon^{-\frac{2}{r}}$。

证明。证明过程与定理 10的证明类似。具体而言，我们仍将使用相同的ρ ，如（H.1）所示，并沿用该证明中的记号。根据定理 1，只需控制伪等距和广义括号数。我们首先断言，若 $\mathcal{M}$ 是由当前设定所诱导的对手分布类，则

$$\sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\rho(\theta,\theta')\right] \leq \frac{4D^2H^2K^2L}{\gamma^{\frac{1}{r}}\cdot\sigma_{\text{poly},r}}\cdot\left|\left|\theta-\theta'\right|\right|_1^{\frac{1}{r}}. \tag{H.5}$$

为此，我们注意到，若 $\mathcal{A}_h$ 如引理H.1中所述，则根据同一结论

$$\begin{aligned}\sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\rho(\theta,\theta')\right] &= \sup_{\nu\in\mathcal{M}}\mathbb{E}_\nu\left[\left|\left|\theta-\theta'\right|\right| + \sum_{h=1}^{H}\left|\left|\mathbf{x}_h-\mathbf{x}'_h\right|\right|\right]\\ &\leq 2LH\cdot\left|\left|\theta-\theta'\right|\right|_1 + 2DH\cdot\sup_{\nu\in\mathcal{M}}\sum_{h=1}^{H}\mathbb{P}_\nu\left(\mathcal{A}_h\right).\end{aligned}$$

现在，我们按照引理H.2计算，

$$\begin{aligned}\mathbb{P}\left(\mathcal{A}_h\right) &= \mathbb{P}\left(k_{h'}=k'_{h'}\text{ for }h'<h\text{ and }k_h\neq k'_h\right)\\ &\leq \sum_{k\neq k'\in[K]}\mathbb{P}\left(k_{h'}=k'_{h'}\text{ for }h'<h\text{ and }f_{\mathbf{w}_{k,h}}(\mathbf{v}_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\geq 0> f_{\mathbf{w}_{k,h}}(\mathbf{v}'_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}'_h)\right).\end{aligned}$$

对于固定的 $k\neq k'$，我们接着计算

$$\begin{aligned}&\mathbb{P}\left(k_{h'}=k'_{h'}\text{ for }h'<h\text{ and }f_{\mathbf{w}_{k,h}}(\mathbf{v}_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\geq 0> f_{\mathbf{w}_{k,h}}(\mathbf{v}'_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}'_h)\right)\\ &\leq \mathbb{P}\left(k_{h'}=k'_{h'}\text{ for }h'<h\text{ and }\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right|\leq\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}'_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}'_h)\right|+\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right|\right)\\ &\leq \mathbb{P}\left(k_{h'}=k'_{h'}\text{ for }h'<h\text{ and }\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right|\leq 2D^r\cdot\left|\left|\mathbf{v}_h-\mathbf{v}'_h\right|\right|_1\right),\end{aligned}$$

其中，第二个不等式源于以下假设：$\mathbf{w}_{k,\ h}$ 具有单位范数；幂函数是局部利普希茨的；以及$\mathbf{v}_h$ 的范数不超过 D。应用多项式平滑性的假设，可知最后一个表达式有如下上界：

$$\begin{aligned}&\mathbb{P}\left(k_{h'}=k'_{h'}\text{ for }h'<h\text{ and }\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right|\leq 2D^r\cdot\left|\left|\mathbf{v}_h-\mathbf{v}'_h\right|\right|_1\right)\\ &\leq \frac{2D}{\gamma^{\frac{1}{r}}\cdot\sigma_{\text{poly},r}}\left|\left|\mathbf{v}_h-\mathbf{v}_{h'}\right|\right|_1^{\frac{1}{r}}.\end{aligned}$$

现在，我们以与引理 H.2类似的方式得出结论，并注意到

$$\left|\left|\mathbf{v}_h - \mathbf{v}'_h\right|\right|_1^{\frac{1}{r}} \leq 2\left|\left|\mathbf{u}_h - \mathbf{u}'_h\right|\right|_1^{\frac{1}{r}} + 2\left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|_1^{\frac{1}{r}}$$

并且由于在该事件中模式序列 $k_{1:h-1} = k'_{1:h-1}$，

$$\left|\left|\mathbf{x}_h - \mathbf{x}'_h\right|\right|_1^{\frac{1}{r}} \leq L^{\frac{1}{r}} \cdot \left|\left|\theta - \theta'\right|\right|_1^{\frac{1}{r}}.$$

将所有内容综合起来，即足以证明（H.5）成立。

现在我们声称，对于 $0 < \epsilon < 1$,

$$\mathcal{N}_{\mathcal{M},[]}\left(\Theta, \mathcal{M}, \epsilon\right) \leq \left(\frac{12D^2H^2K^2L}{\sigma_{\text{poly},r}\gamma^{\frac{1}{r}}}\epsilon\right)^{K^2r^2(Hd)^r}. \tag{H.6}$$

为此，令 $\mathcal{N} = \left\{\theta^i\right\} \subset \mathcal{U}^{\times H}$ 表示关于 $\widetilde{\epsilon}$的一个 ℓ_1-网，其中 $\widetilde{\epsilon} = \gamma\left(\frac{\sigma_{\text{poly}}}{4D^2H^2K^2L} \cdot \epsilon\right)^r$。我们的证明与引理 H.3 中的证明类似，并且我们将证明：如果 $\mathcal{B}_i$ 表示以 $\widetilde{\epsilon}$为中心的 θ^i -球的集合，那么相应的 $\left(\theta^i,\ \mathcal{B}_i\right)$ 关于 ϵ构成一个广义 $\mathcal{M}$-括号。事实上，对于固定的 i 和 $\nu \in \mathcal{M}$，我们有：

$$\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\rho(\theta,\theta^i)\right] \leq \widetilde{\epsilon} + \mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\sum_{h=1}^{H}\left|\left|\mathbf{x}_h - \mathbf{x}_h^i\right|\right|_1\right].$$

对于第二项，我们再次调用引理 H.1 并注意到

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\sum_{h=1}^{H}\left|\left|\mathbf{x}_h - \mathbf{x}_h^i\right|\right|_1\right] &\leq \sum_{h=1}^{H}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i} L\cdot\left|\left|\theta-\theta^i\right|\right|_1 + 2D\cdot\sum_{h'=1}^{h-1}\mathbb{I}\left[\mathcal{A}_{h'}\right]\right] \\ &\leq LH\widetilde{\epsilon} + 2DH^2\cdot\max_{h\in[H]}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\mathbb{I}\left[\mathcal{A}_h\right]\right].\end{aligned}$$

采用与引理 H.3 中相同的推理，并以与引理 G.3 相同的方式使用多项式平滑性假设，我们可知

$$\mathbb{P}_\nu\left(\mathcal{A}_h\right) \leq \sum_{k\neq k'\in[K]}\mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'} = k_{h'}^i \text{ for } h' < h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h^i) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h^i)\right| \leq 2D^r\cdot\left|\left|\mathbf{v}_h - \mathbf{v}_h^i\right|\right|_1\right\}\right).$$

再次注意到，

$$\left|\left|\mathbf{v}_h - \mathbf{v}_h^i\right|\right|_1 \leq 2L\cdot\left|\left|\theta-\theta^i\right|\right|$$

根据 G_t 的 Lipschitz 性，我们可知

$$\begin{aligned}\mathbb{P}_\nu\left(\mathcal{A}_h\right) &\leq \sum_{k\neq k'\in[K]}\mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'} = k_{h'}^i \text{ for } h' < h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h^i) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h^i)\right| \leq 4D^rL\widetilde{\epsilon}\right\}\right) \\ &\leq \sum_{k\neq k'\in[K]}\mathbb{P}\left(\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h^i) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h^i)\right| \leq 4D^rL\widetilde{\epsilon}\right) \\ &\leq K^2\frac{4DL^{\frac{1}{r}}}{\gamma^{\frac{1}{r}}\cdot\sigma_{\text{poly},r}}\cdot\widetilde{\epsilon}^{\frac{1}{r}}.\end{aligned}$$

将所有内容代回上述推导中，我们确实可以看到 $\mathcal{N}$ 是一个广义 ϵ-网。使用引理 F.3、G.3 和 H.3证明中的同一体积论证，我们可知（H.6）成立。

最后，我们注意到 ℓ^v 关于 ρ 显然是利普希茨连续的，因此我们可以应用 定理 1，该定理指出，如果我们运行算法 2，那么对于 $n = \sqrt{\eta}$,

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\eta + \frac{T}{n} K^2 r^2 H^r d^r \log\left(\frac{1}{\sigma_{\mathrm{poly},r}}\right) + \frac{TD}{\gamma^{\frac{1}{r}} \sigma_{\mathrm{poly},r}} \cdot \left(\frac{K^2 r^2 H^r d^r}{\eta}\right)^{\frac{1}{4r-2}}\right).$$

按陈述中的方式设置 η，即可完成证明。 □