《机器学习研究论文集》第 XX 卷：1–21，2023

老虎机反馈下安全优化中约束集几何性质的影响

Spencer Hutchinson SHUTCHINSON@UCSB.EDU
加利福尼亚大学圣巴巴拉分校
Berkay Turan BTURAN@UCSB.EDU
加利福尼亚大学圣塔芭芭拉分校
Mahnoosh Alizadeh ALIZADEH@UCSB.EDU
加利福尼亚大学圣塔芭芭拉分校

arXiv:2305.00889v1 [cs.LG] 1 May 2023

摘要

我们考虑一个具有老虎机反馈的安全优化问题，其中智能体依次选择动作并观察来自环境的响应，目标是在遵守各阶段约束的同时，最大化响应的任意函数。我们为该问题提出了一种算法，并研究约束集的几何性质如何影响算法的遗憾。为此，我们引入特定约束集的*尖锐度*这一概念，用以刻画在不确定环境下于约束集内进行学习的难度。尖锐度这一概念使我们能够识别出所提算法能够保证实现次线性遗憾的约束集类别。该算法的仿真结果支持次线性遗憾界，并提供了经验证据，表明约束集的尖锐度会影响算法的性能。

关键词： 安全学习，多臂老虎机，优化

1. 引言

随着当代学习与控制范式不断拓展至具有严格安全要求的领域，对能够提供此类安全保证的控制机制的需求显著增长。这催生了文献中大量新颖的安全学习问题，涉及模型预测控制 Koller 等（2018年）；Hewing 等（2020年）；Chen 等（2022年）、强化学习 Junges 等（2016）；Garcıa 和 Fern´andez（2015年）、优化 Usmanova 等（2019年）；Fereydounian 等（2020年）、多臂老虎机 Sui 等（2015年）以及其他诸多领域。这类问题非常适用于控制算法需要与人类交互的应用场景，而这种交互会引入必须纳入考量的不确定性，以确保安全性（例如，临床试验 Sui 和 Burdick（2017年）、机器人系统 Berkenkamp 等（2021年），以及通过定价实现社会基础设施中的资源分配 Hutchinson 等（2022年））。

在这项工作中，我们研究一个序贯决策问题，其中决策必须处于任意且未知的紧致安全集内。我们考虑一个具有老虎机反馈的安全优化框架，其中奖励和约束集是动作与未知参数相乘后所得矩阵乘积的已知非线性函数。与现有文献相比，这个问题具有独特的挑战性，因为（1）决策和来自环境的反馈都是向量；（2）奖励是决策向量的任意函数；（3）决策向量上的安全约束是一个任意紧集。这些挑战

© 2023 S. Hutchinson、B. Turan 和 M. Alizadeh。

不过，考虑到这类问题出现在许多现实应用中，这样的研究是有必要的。例如，潮流约束通常是非线性且非凸的（Molzahn 等人 (2017年)），且通常通过（非线性）凸松弛求解（例如 Bai 等人 (2008)；Farivar 和 Low (2013)。

我们通过引入一种集合的几何性质来应对一般安全集带来的挑战，我们将其称为尖锐度。它与在特定安全集内进行学习的难度有关。这使我们能够根据安全集的尖锐度来刻画学习算法的性能，并以遗憾衡量。相应地，我们找出了能够建立次线性遗憾界的一类安全集（其中包括所有凸集）。

相关工作：在不确定性下带安全约束的序贯决策，日益成为学术界热门的研究领域。特别是，已有多项研究探讨约束函数不确定的优化问题。根据具体的问题设定，约束函数被假定为线性函数 Usmanova 等（2019年）；Chaudhary 和 Kalathil（2022年）；Fereydounian 等（2020年），具有高斯过程先验 Sui 等（2015年、2018年）；Berkenkamp 等（2021年），或满足一般的利普希茨条件 Usmanova 等（2020年）。在所有这些研究中，约束规定（未知）函数的输出必须位于非正正交象限（即 $g(x) \leq 0$）；而我们将约束建模为未知函数的输出必须位于某个任意集合中（即 $g(x) \in \mathcal{E}$，其中 $\mathcal{E}$为任意值）。这种模型需要采用独特的分析方法，研究这一任意约束集的几何性质如何影响算法的性能。

不确定约束也在随机线性多臂老虎机设定中得到了研究。在传统的随机线性多臂老虎机设定中（不含不确定约束），智能体在每个时间步选择一个动作向量，随后获得一个在期望上关于该动作呈线性的奖励，目标是最大化其累积奖励（参见例如 Dani 等人 (2008年); Abbasi-Yadkori 等人 (2011 年)）。带有不确定约束的一类随机线性多臂老虎机问题称为保守线性多臂老虎机 Kazerouni 等人 (2017年);Moradipari 等人 (2020年)，其中每一轮的期望奖励都需要接近一个基准奖励。另一些研究则考虑存在辅助约束函数的设定。具体而言，在 Amani 等人 (2019年) 中，约束函数取决于（经过线性变换的）奖励参数；而在 Pacchiano 等人 (2021 年); Moradipari 等人 (2021 年); Wang 等人 (2022 年) 中，约束函数与奖励参数无关，学习者会收到该函数的带噪反馈。与随机线性多臂老虎机类似，我们考虑环境的期望响应是动作的线性函数这一问题。不过，我们将环境响应视为向量而非标量，将奖励视为期望响应的任意函数，并要求环境的期望响应落在任意集合内。该问题的主要新意在于其任意约束集，这需要新的分析技术，而这些技术可能具有更广泛的适用性。

记号：对于向量 $v \in \mathbb{R}^d$ 和正定矩阵 $A \in \mathbb{R}^{d\times d}$，我们将加权二范数记为 $\|v\|_A = \sqrt{v^\top A v}$。方阵 M 的最小和最大特征值分别记为 $\lambda_{min}(M)$ 和 $\lambda_{max}(M)$。我们将半径为 r 且具有范数 $\|\cdot\|$ 的闭球和开球（均以原点为中心）分别记为 $\mathcal{B}_{\|\cdot\|}(r)$ 和 $\mathcal{B}_{\|\cdot\|}(r)$。矩阵$M$的条件数记为 $\kappa(M)$。对于集合 $\mathcal{D}$和矩阵 M，我们使用记号 $M\mathcal{D} = \{Mx\colon\ x \in \mathcal{D}\}$。集 $\{1, 2, ..., n\}$ 记为 $[n]$。我们用 $\mathcal{O}$ 表示忽略对数因子的大 O 记号。我们使用记号 $\mathbf{e}_i$ 表示一个向量，其第 i个位置为 1，其余位置均为 0。对于向量范数 $\|\cdot\|$，其对偶范数记为 $\|\cdot\|_\star$。诱导矩阵 p-范数记为 $\|M\|_p$，其中矩阵 M满足 $\|M\|_p = \sup_{\|x\|_p=1} \|Ax\|_p$。矩阵 M的 Frobenius 范数记为 $\|M\|_F$。

2. 问题设定

我们研究一个序贯决策问题：在每一轮中，智能体选择一个动作，随后环境根据所采取的动作选择一个响应。与随机线性多臂老虎机类似，我们假设环境的响应是动作的未知线性函数，且智能体观测到该线性函数的输出及一些噪声。不过，我们的问题与随机线性多臂老虎机不同：响应是多维的，而智能体的目标是累积奖励，奖励是响应的任意已知函数。在我们的问题中，智能体还需要确保每一轮环境的响应都落在安全集内。

问题的详细描述如下。每一轮 $t \in [T]$中，智能体从紧致动作集 $\mathcal{A} \subset \mathbb{R}^d$ 中选择一个动作 x_t，然后观察带噪声的响应 $y_t := \Theta_* x_t + \epsilon_t$。矩阵 $\Theta_* \in \mathbb{R}^{n \times d}$ 是一个满秩的未知参数，$\epsilon_t \in \mathbb{R}^n$ 是随机噪声，并且我们有 $n \leq d$。选择动作后，智能体获得奖励 $f(\Theta_* x_t)$，其中奖励函数 $f: \mathbb{R}^n \to \mathbb{R}$ 是已知的。智能体需要确保其选择动作时，响应 $\Theta_* x_t$ 位于已知的、具有非空内部的紧致安全集 $\mathcal{E} \subset \mathbb{R}^n$ 内；等价地，x_t 属于未知可行动作集 $\mathcal{X} := \{x \in \mathcal{A}: \Theta_* x \in \mathcal{E}\}$。

智能体通过选择行动，旨在最大化其获得的累积奖励，同时确保所有轮次均满足安全约束。因此，可以用累积遗憾衡量智能体的性能，$R_T := \sum_{t=1}^{T} (f(\Theta_* x_*) - f(\Theta_* x_t))$，其中 $x_* \in \arg\max_{x \in \mathcal{X}} f(\Theta_* x)$。

在以下两个假设中，我们假设未知参数和可行行动均有界，且噪声服从次高斯分布。这些假设在相关文献中很常见（例如 Abbasi-Yadkori 等人 （2011年）；Pacchiano 等人（2021年）。

假设 1 对于所有 x *in* $\mathcal{A}$，存在一个常数 L 使得 $\|x\|_2 \leq L$。此外，存在一个常数 S，使得对于所有 $i \in [n]$均有 $\|\theta_*^i\|_2 \leq S$，其中 θ_*^i 是 Θ_*的第 i行。智能体已知常数 S 和 L。

假设 2 对于所有 $t \in [T]$，噪声 ϵ_t 在逐元素意义上服从条件 R-次高斯分布，因此给定历史记录 $\mathcal{F}_t = \sigma(x_1, x_2, ..., x_{t+1}, \epsilon_1, \epsilon_2, ..., \epsilon_t)$，并将 ϵ_t 的第 i 个元素记为 ϵ_t^i，对于所有 $i \in [n]$，都有 $\mathbb{E}[\epsilon_t^i | \mathcal{F}_{t-1}] = 0$ 和 $\mathbb{E}[e^{\lambda \epsilon_t^i} | \mathcal{F}_{t-1}] \leq \exp(\frac{\lambda^2 R^2}{2})$ $\forall \lambda \in \mathbb{R}$。常数 R 为智能体所知。

此外，我们假设奖励函数满足 Lipschitz 条件。

假设 3 f在 $\mathcal{E}$上是 M-*Lipschitz* 连续的，使得 $|f(x) - f(y)| \leq M\|x - y\|_2$对所有 $x, y \in \mathcal{E}$均成立。

最后，我们提出一个假设，以确保假设1 所提供给智能体的知识足以让其选择初始安全行动。由于已知 Θ_* 属于$\mathcal{C}^0 = \{[\theta^1\ \theta^2\ ...\ \theta^n]^\top \in \mathbb{R}^{n \times d}: \|\theta^i\|_2 \leq S, \forall i \in [n]\}$ 根据假设 1，因此也可知 $\mathcal{G}^0 := \{x \in \mathcal{A}: \Theta x \in \mathcal{E}, \forall \Theta \in \mathcal{C}^0\}$ 是 $\mathcal{X}$的子集。因此，我们假设 $\mathcal{G}^0$ 的内部非空，以确保智能体能够在初始时选择安全行动。

假设4 初始可行集 $\mathcal{G}^0$ 具有非空内部。

我们将在下一节给出针对所述问题的算法。

算法1

$\mathcal{A}, \mathcal{E}, f, L, S$

// 纯探索

1 对于 $t = 1$，**执行** T'，**执行**

2 通过随机采样 $\mathcal{G}^0$选择 x_t，并观察响应 y_t。

3 结束

4 构造 $\mathcal{C}_{T'}$ 和 $\mathcal{G}_{T'}$ 分别满足（1）和（2）。

// 探索-利用

5 对 $t = T' + 1$ **到** T **执行**

6 选择一些（x_t， $\tilde{\Theta}_t$）$\in \arg\max_{(x,\Theta)\in\mathcal{G}_{t-1}\times\mathcal{C}_{t-1}} f(\Theta x)$，并观察响应 y_t。

7 更新 $\mathcal{C}_t$ 和 $\mathcal{G}_t$，分别用（1）和（2）表示。

8 结束

3. 所提算法

我们提出一种算法来解决所述问题：首先进行适当时长的纯探索 T'，具体时长将在分析中说明；随后在剩余轮次中进行探索与利用。该算法见算法1。

纯探索阶段通过从 $\mathcal{G}^0$ 中随机采样动作来进行，使得 $\lambda_- := \lambda_{\min}\left(\mathbb{E}\left[x_t x_t^\top\right]\right) > 0$ 对于 $t \in [T']$ 成立。由于 $\mathcal{G}^0$ 具有非空内部，因此这种方案是可行的，不过我们将具体的采样方案选择留作设计决策。[1]

在探索-利用阶段的每一轮，$t \in (T', T]$，首先根据当前知识确定能够确保安全的动作集，然后在该安全动作集中选择乐观动作。为了既识别哪些动作是安全的，又能乐观地选择动作，我们使用置信集，参数 Θ_* 以高概率位于其中。设 θ_*^i 为 i 的第 Θ_* 行，使得 $\Theta_* = [\theta_*^1\ \theta_*^2\ \ldots\ \theta_*^n]^\top$，并设 y_t^i 为 i 的第 y_t 个元素，使得 $y_t = [y_t^1\ y_t^2\ \ldots\ y_t^n]^\top$。则每个 θ_*^i 的正则化最小二乘估计量为 $\hat{}\theta_t^i = [V_t]^{-1}\sum_{s=1}^t x_s y_s^i$，其中 t 为第 $V_t = \nu I + \sum_{s=1}^t x_s [x_s]^\top$ 轮的 Gram矩阵。利用 Θ_* 每一行的正则化最小二乘估计量，我们根据 Abbasi-Yadkori 等人(2011) 的以下定理定义置信集。

定理1 *(Abbasi-Yadkori 等人 (2011)*中的定理2）若满足假设 *1* 和 *2* 。那么，如果 x_t属于 $\mathcal{A}$ 对所有 t，则以至少 $1-\delta$ 的概率，Θ_* 属于集合

$$\mathcal{C}_t = \left\{ [\theta^1\ \theta^2\ \ldots\ \theta^n]^\top \in \mathbb{R}^{n\times d} : \left\|\theta^i - \hat{\theta}_t^i\right\|_{V_t} \le \sqrt{\beta_t}, \forall i \in [n] \right\} \tag{1}$$

对所有 $t \ge 0$，其中

$$\sqrt{\beta_t} = R\sqrt{d\log\left(\frac{1+tL^2/\nu}{\delta/n}\right)} + \sqrt{\nu}S.$$

1. 我们给出一个具有 $\lambda_- > 0$ 的采样方案示例。由于 $\mathcal{G}^0$ 具有非空内部，因此存在 $v \in \mathbb{R}^d$ 和 $r > 0$，使得开球 $v + \mathcal{B}_2(r)$ 是 $\mathcal{G}^0$ 的子集。由此可知，闭球 $v + \bar{\mathcal{B}}_2(r/2)$ 是 $\mathcal{G}^0$ 的子集。因此，一种可行的采样方案是从单位球面上独立同分布地均匀采样 u_t，使得 $\mathbb{E}[u_t u_t^\top] = \frac{1}{d} I$，然后执行 $x_t = v + \frac{r}{2}u_t$。因此，在给定 v 固定的情况下，$\mathbb{E}[x_t x_t^\top] = \mathbb{E}[vv^\top] + \frac{r}{2}\mathbb{E}[vu_t^\top + u_t v^\top] + \frac{r^2}{4}\mathbb{E}[u_t u_t^\top] = v\bar{v}^\top + \frac{r^2}{4d} I$，进而可得 $\lambda_- = \frac{r^2}{4d} > 0$。

Using 该集合，我们定义可行动作集 ($\mathcal{X}$) 的 保守内近似 a s

$$\mathcal{G}_t := \{x \in \mathcal{A} : \Theta x \in \mathcal{E}, \forall \Theta \in \mathcal{C}_t\}. \tag{2}$$

注意，集合 $\mathcal{C}_t$ 和 $\mathcal{G}_t$ 会在每轮结束时更新，因此智能体在第$\mathcal{C}_{t-1}$ 轮可以访问 $\mathcal{G}_{t-1}$ 和 t。根据 $\mathcal{G}_t$的定义可知，对于任意 $\Theta \in \mathcal{C}_t$ 和 $x \in \mathcal{G}_t$，都可以保证 Θx 属于安全集 $\mathcal{E}$。定理1指出，以高概率， Θ_* 在所有轮次中都属于 $\mathcal{C}_t$ 。因此，如果算法从 x_t 中选择 $\mathcal{G}_{t-1}$，那么环境的响应$\{\Theta_* x_t\}_{\forall t \in (T', T]}$ 以高概率都将属于 $\mathcal{E}$，从而能够确保安全。

为了从这些保守动作集 $\{\mathcal{G}_{t-1}\}_{\forall t \in (T', T]}$中选择动作以获得较低的遗憾，算法采取乐观策略。也就是说，算法会在 $\mathcal{G}_{t-1}$ 中选择最佳动作，并假设根据现有信息，真实参数 Θ_* 尽可能有利。由于智能体知道，在第 t轮中， Θ_* 很可能位于置信集 $\mathcal{C}_{t-1}$ 内，因此算法采取乐观策略，在 $\mathcal{G}_{t-1}$ 中寻找一个动作，并在 $\mathcal{C}_{t-1}$中寻找一个参数，使可能获得的奖励最大化。因此，算法选择的动作如下:

$$(x_t, \tilde{\Theta}_t) \in \underset{(x,\Theta) \in \mathcal{G}_{t-1} \times \mathcal{C}_{t-1}}{\arg\max} f(\Theta x), \tag{3}$$

每一轮 $t \in (T', T]$。

需要注意的是，由于 $\mathcal{G}_{t-1}$ 是 $\mathcal{X}$的保守内近似，最优动作 x_* 可能不属于 $\mathcal{G}_{t-1}$。因此，$\mathcal{G}_{t-1}$ 对 $\mathcal{X}$ 的近似程度会影响 x_t 与最优动作 x_*之间的距离，进而影响 $f(\Theta_* x_*)$ 与 $f(\Theta_* x_t)$ 之间的差距（假设 f 满足利普希茨假设）。 $\mathcal{G}_t$ 对 $\mathcal{X}$ 的近似精度显然会受到 $\mathcal{C}_t$ 大小的影响，但正如我们将在下一节中展示的，安全约束 $\mathcal{E}$ 和动作集 $\mathcal{A}$ 的几何性质也起着重要作用。

4. 遗憾分析

本节中，我们将证明算法 1 的累积遗憾上界。影响遗憾的一个关键问题是安全集和动作集的几何性质。目前，我们尚未对这些集合做出任何假设。不过，我们将证明，这些集合的几何性质决定了我们能否证明该算法具有次线性遗憾。为便于分析，我们引入一种集合的几何性质，并将其称为 尖锐度，它在所提算法的遗憾界中起着关键作用。

4.1. 安全集的几何性质

智能体选择动作时，响应结果存在不确定性，因此有必要使用安全动作集的保守内近似。从该内近似中选择动作可以维持安全性，因为它能确保对所选动作所有合理可能的响应（即每个 Θx_t 对应于 $\Theta \in \mathcal{C}_{t-1}$）都满足安全约束。本质上，这确保了预期响应周围的某个区域位于安全约束之内。可以想象，这个区域无法很好地“嵌入”安全集可能具有的任何“尖锐”角落，因此，对于具有“尖锐”角落的安全集，其内近似在安全性方面会更宽松，从而导致较差的遗憾。我们通过以下一系列定义对这一 尖锐度 概念进行形式化。本节的证明见附录 A。

为了研究安全集的几何形状对保守内近似紧致性的影响，我们首先介绍一种更一般的内近似，称为集合的收缩版本。与保守内近似确保所有合理可能的响应都位于安全约束内类似，集合的收缩版本确保每个点处的一个闭球都位于原集合内。形式化定义如下[2]，其中使用闭球 $\mathcal{B}_{\|\cdot\|}(r) := \{x \in \mathbb{R}^m : \|x\| \leq r\}$， r 表示半径，特定的范数为 $\|\cdot\|$。

定义 2 对于紧集 $\mathcal{D} \subset \mathbb{R}^m$、范数 $\|\cdot\|$ 和非负标量 Δ，我们将 $\mathcal{D}$ 的收缩版本 定义为 $\mathcal{D}_\Delta^{\|\cdot\|} := \{x \in \mathcal{D} : \quad x + v \in \mathcal{D}, \forall v \in \bar{\mathcal{B}}_{\|\cdot\|}(\Delta)\}$.[2]

根据上述集合收缩版本的定义，可以考虑集合在仍然非空的情况下所能承受的最大收缩量。我们在以下定义中引入集合的最大收缩量。

定义 3 对于紧集 $\mathcal{D} \subset \mathbb{R}^m$ 和范数 $\|\cdot\|$，我们将 $\mathcal{D}$ 的最大收缩量 定义为 $H_\mathcal{D}^{\|\cdot\|} := \text{上确界}\{\Delta : \quad \mathcal{D}_\Delta^{\|\cdot\|} \neq \emptyset\}$.

现在，我们可以将集合的尖锐度定义为：集合中任意点到该集合收缩版本中最近点的最大距离。

定义 4 对于紧集 $\mathcal{D} \subset \mathbb{R}^m$ 和范数 $\|\cdot\|$，我们将 $\mathcal{D}$ 的尖锐度定义为

$$\text{Sharp}_\mathcal{D}^{\|\cdot\|}(\Delta) := \sup_{x \in \mathcal{D}} \inf_{y \in \mathcal{D}_\Delta^{\|\cdot\|}} \|y - x\|_2 ,$$

对于所有使 $\mathcal{D}_\Delta^{\|\cdot\|}$ 非空的非负 Δ。[3]

尖锐度适用于安全学习算法的分析，因为它给出了安全集内某个最优点（例如某个集合 $\mathcal{D}$）与该安全集的保守内近似（例如 $\mathcal{D}_\Delta^{\|\cdot\|}$ 或 $\mathcal{D}_\Delta^{\|\cdot\|}$ 的超集）之间距离的上界。为了展示集合的几何形状如何影响其尖锐度，图 $\mathbb{R}^2$ 1 中绘制了 1中若干不同集合的尖锐度。可以看出，对于相同的 Δ值，具有更“尖锐”角的集合拥有更大的尖锐度。还需注意，我们使用 $\mathcal{D}_\Delta^p$、$H_\mathcal{D}^p$ 和$\text{Sharp}_\mathcal{D}^p(\Delta)$ 分别表示某个集合的收缩集、最大收缩量和尖锐度 $\mathcal{D}$（相对于 p-范数）。

现在，我们来说明一些简单性质，这些性质涉及集合的收缩版本何时非空，以及因此尖锐度何时有定义。首先，对于紧集，存在某个正的收缩量使其收缩版本非空，当且仅当该集合具有非空内部。

命题 5 对于紧集 $\mathcal{D} \subset \mathbb{R}^m$，存在一个 $\Delta > 0$，使得 $\mathcal{D}_\Delta^{\|\cdot\|}$ 非空，当且仅当 $\mathcal{D}$ 具有非空内部。

2. 我们也可以用闵可夫斯基差来定义 $\mathcal{D}_\Delta^{\|\cdot\|}$。集合 $A, B \subseteq \mathbb{R}^m$ 的闵可夫斯基差定义为 $A \ominus B := \{a - b : a \in A, b \in B\}$，或者等价地， $A \ominus B = \{x \in \mathbb{R}^m : x + B \subseteq A\}$ （Schneider (2014)）。因此，对于 $\Delta \geq 0$，我们可以写成 $\mathcal{D}_\Delta^{\|\cdot\|} = \mathcal{D} \ominus \mathcal{B}_{\|\cdot\|}(\Delta)$。3. 也可以用集合之间的豪斯多夫度量来定义尖锐度（参见 Schneider (2014) 第 1.8 节），使得 $\text{Sharp}_\mathcal{D}^{\|\cdot\|}(\Delta) = d_H(\mathcal{D}, \mathcal{D}_\Delta^{\|\cdot\|})$。

接下来，我们证明，具有非空内部的紧集的收缩版本，在收缩量小于或等于该集合的最大收缩量时始终非空。这表明，尖锐度定义在从零到最大收缩量的闭区间上。

命题 6 对于具有非空内部的紧集 $\mathcal{D} \subset \mathbb{R}^m$，对所有 $\Delta \in [0, H_{\mathcal{D}}^{\|\cdot\|}]$，$\mathcal{D}_{\Delta}^{\|\cdot\|}$ 均非空。

在本节余下部分，我们将研究具有非空内部的不同类型紧集的尖锐度。我们研究的第一类集合是多面体，它是有限个点的凸包，或者等价地说，是有限个闭半空间的有界交。多面体涵盖了现实世界中范围广泛的约束，并常用于安全学习中的安全集（例如 Chaudhary 和 Kalathil（2022年）；Fereydounian 等（2020年）；Usmanova 等（2019年））。我们使用多面体的多面体表示，$\mathcal{D} = \{x \in \mathbb{R}^m : \ Ax \leq b\}$，其中有 $A \in \mathbb{R}^{p \times m}$ 和$b \in \mathbb{R}^p$，且不存在冗余约束。我们将 $a_j \in \mathbb{R}^m$ 定义为 A 的第 j 行，使得 $A = [a_1 \ a_2 \ ... \ a_p]^\top$；将 $b_j \in \mathbb{R}$ 定义为 b 的第 j 个元素，使得 $b = [b_1 \ b_2 \ ... \ b_p]^\top$。我们还用 $\mathcal{I}_A$ 表示由所有满足以下条件的 m 指数组成的集合：对于每个 $\{i_1, i_2, ..., i_m\} \in \mathcal{I}_A$，向量

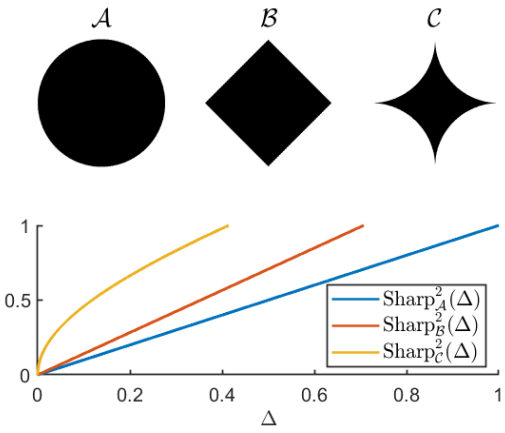


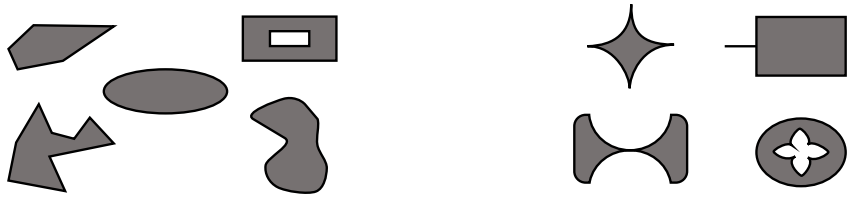
图 1：$\mathbb{R}^2$中三个不同集合的 2-范数尖锐度。

a_{i_1}，a_{i_2}，..... a_{i_m} 线性无关。对于每个 $\ell \in \mathcal{I}_A$ 满足 $\ell = \{i_1, i_2, ..., i_m\}$ 的情形，我们记 $A^\ell = [a_{i_1} \ a_{i_2} \ ... \ a_{i_m}]^\top$，并用 $\kappa(A^\ell)$ 表示其条件数。使用这一记号，下面的命题表明，多面体的尖锐度有一个关于收缩呈线性变化的上界。

命题 7 对于具有非空内部的多面体 $\mathcal{D} = \{x \in \mathbb{R}^m : \ Ax \leq b\}$，有 尖锐度$_{\mathcal{D}}^{\|\cdot\|}$（$\Delta$）$\leq \sqrt{m} C_{\|\cdot\|} K_{\mathcal{D}} \Delta$，其中 $K_{\mathcal{D}} :=$ 最大值$_{\ell \in \mathcal{I}_A}$ κ（A^ℓ），且 $C_{\|\cdot\|} :=$ 最大值$_{\|y\|=1}$ $\|y\|_2$。

命题 7 中的尖锐度界与常数成正比 $K_{\mathcal{D}}$，该常数是所有由 m 个线性无关约束构成的集合的最大条件数。由于每个顶点处都有 m 个活跃的线性无关约束，$K_{\mathcal{D}}$ 为每个顶点处活跃约束的条件数提供了上界。鉴于约束的条件数反映了它们彼此接近平行的程度，因此这是衡量多面体尖锐度的一种直观方式。另外请注意，命题 $C_{\|\cdot\|}$ 中的项 7 可能取决于维度。例如，使用无穷范数时，$C_{\|\cdot\|_\infty} = \sqrt{m}$；使用 1-范数时，$C_{\|\cdot\|_1} = 1$。

利用我们为多面体推导的尖锐度界，我们可以研究更一般的集合。研究更一般集合时所采用的关键直觉是，我们可以定义原集合的子集，并通过适当构造这些子集，以尖锐度为依据界定原集合。具体而言，我们构造原集合的多面体子集，从而得到关于收缩呈线性的尖锐度界。能够建立尖锐度的线性界十分重要，因为这使我们能够为所提算法建立次线性遗憾界，我们将在.....中讨论这一点。



(*a*) (*b*)

图 2：集合示例 $\mathbb{R}^2$，它们具有多面体尖锐性，因此可通过命题10 (a) 根据收缩量得到线性界；以及不具有多面体尖锐性的集合 (b)。

下一节。为了建立一个利用多面体子集的界，我们定义了可用于界定给定集合尖锐度的多面体族。

定义 8对于紧集 $\mathcal{D} \subset \mathbb{R}^m$中的点 x，我们将 $F_{\mathcal{D}}(x)$ 定义为包含 x 且为 $\mathcal{D}$子集的、具有非空内部的多面体族。

由此，我们定义这样一类集合：对于它们，可以利用多面体子集来界定尖锐度。

定义 9若紧集 $\mathcal{D} \subset \mathbb{R}^m$ 满足对于所有$x \in \mathcal{D}$， $F_{\mathcal{D}}(x)$ 均非空，则称其具有多面体尖锐性。

可以看出，具有多面体尖锐性的集合是这样一类集合：对于集合中的每个点，都可以构造一个多面体族，使其包含该点，同时仍为原集合的子集。图2展示了满足和不满足这一条件的集合示例。根据这个定义，我们可以给出以下命题，它为具有多面体尖锐性的集合提供了线性尖锐度界。

命题 10 设 $\mathcal{D} \subset \mathbb{R}^m$ 是一个多面体尖锐的紧集，并为每个 $x \in \mathcal{D}$ 任意选取某个 $\mathcal{F}_x \in F_{\mathcal{D}}(x)$。那么，我们有

$$\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) \leq \sqrt{m}\bar{C}_{\|\cdot\|}\Gamma_{\mathcal{D}}\Delta,$$

其中， $\bar{C}_{\|\cdot\|} = \max(C_{\|\cdot\|}, 1)$ 且

$$\Gamma_{\mathcal{D}} := \max\left\{\bar{K}_{\mathcal{F}}, \frac{r_{\mathcal{D}}}{\bar{H}_{\mathcal{F}}^{\|\cdot\|}}\right\},$$

其中 $\bar{K}_{\mathcal{F}} := \max_{x\in\mathcal{D}} K_{\mathcal{F}_x}$, $\bar{H}_{\mathcal{F}}^{\|\cdot\|} := \min_{x\in\mathcal{D}} H_{\mathcal{F}_x}^{\|\cdot\|}$ 且 $r_{\mathcal{D}} := \max_{x,\ y\in\mathcal{D}} \|x-y\|_2$。

除了为多面体尖锐集提供线性尖锐度界外，命题 10还表明，当多面体较小时，即 $\bar{H}_{\mathcal{F}}^{\|\cdot\|}$ 较小，或多面体较尖锐时，即 $\bar{K}_{\mathcal{F}}$ 较大，尖锐度界会更大，因此更不理想。根据命题 10，我们还可以立即证明，每个具有非空内部的紧凸集都是线性尖锐的。

推论 11 如果紧集 $\mathcal{D} \subset \mathbb{R}^m$ 具有非空内部且为凸集，则它是多面体尖锐的，并且有 $\text{尖锐}_{\mathcal{D}}^{\|\cdot\|}(\Delta) \leq \sqrt{m}\bar{C}_{\|\cdot\|}\Gamma_{\mathcal{D}}\Delta$。

需要注意的是，尽管所有紧凸集都是多面体尖锐的，但也存在多面体尖锐的非凸集。

4.2. 遗憾界

现在，我们将利用上一节的结果，为算法1的累积遗憾建立一个次线性界。为此，我们将可行响应集定义为 $\mathcal{Y} := \Theta_* \mathcal{A} \cap \mathcal{E}$，其中采用记号 $\Theta_* \mathcal{A} = \{\Theta_* x : x \in \mathcal{A}\}$。集合 $\mathcal{Y}$ 反映了在给定动作集 $\mathcal{A}$ 和安全集 $\mathcal{E}$ 的情况下可能出现的响应集合。定理 12 所示，算法1的遗憾界使用了 $\mathcal{Y}$ 的尖锐度。尽管人们可能会认为遗憾界中应包含 $\mathcal{E}$ 的尖锐度（而非 $\mathcal{Y}$ 的尖锐度），但集合 $\mathcal{A}$ 会影响最优动作 x^* 到集合 $\mathcal{G}_{t-1}$ 的距离，因此，仅使用 $\mathcal{E}$ 的尖锐度进行遗憾分析是不够的。因此，我们使用 $\mathcal{Y}$ 的尖锐度，来同时体现 $\mathcal{E}$ 的尖锐度以及特定问题设定中的具体 $\mathcal{A}$ 可能造成的任何不利影响。

定理 12 假设 *1–4* 成立。以至少 $1 - 2\delta$ 的概率，算法*1*的遗憾有如下界：

$$R_T \leq 2M\sqrt{n}LST' + M(T - T')\mathrm{Sharp}_{\mathcal{Y}}^{\infty}\left(\frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu + \lambda_- T'}}\right) + M\max(H_{\mathcal{Y}}^{\infty}, 1)\sqrt{n8\beta_T(T - T')d\log\left(\frac{1 + TL^2}{d\nu}\right)}.$$

对任意 $T' \geq \max(t_\delta, t_h)$，其中 $t_\delta := \frac{8L^2}{\lambda_-}\log(\frac{d}{\delta})$，且 $t_h := \frac{8\beta_T L^2 (H_{\mathcal{Y}}^{\infty})_2}{\lambda_-} - \frac{2\nu}{\lambda_-}$。

推论13 假设与定理 *12*相同。若 $\mathcal{Y}$ 是多面体尖锐的，则算法 *1* 满足

$$R_T \leq 2M\sqrt{n}LST' + \frac{2n\sqrt{2\beta_T}\Gamma_{\mathcal{Y}} LM(T - T')}{\sqrt{2\nu + \lambda_- T'}} + M\max(H_{\mathcal{Y}}^{\infty}, 1)\sqrt{n8\beta_T(T - T')d\log\left(\frac{1 + TL^2}{d\nu}\right)}$$

以至少 $1-2\delta$ 的概率，当 $T' \geq \max(t_\delta, t_h)$ 时成立。特别地，选择 $T' = \max(T^{2/3}, t_\delta, t_h)$ 可确保 $R_T = \tilde{\mathcal{O}}\left(T^{2/3}\right)$。

我们可以看出，遗憾界取决于 $\mathcal{Y}$的尖锐度，并且如推论13所示，当$\tilde{\mathcal{O}}\left(T^{2/3}\right)$ 为 $\mathcal{Y}$多面体尖锐时，遗憾界为 $\mathcal{Y}$。请注意，智能体需要知道 T'的最大收缩量，或其下界，以便适当地选择$\mathcal{Y}$ 。如果已知 $\mathcal{E}$ 的某个子集，或者已知 $\Theta_* \mathcal{A}$ 是 $\mathcal{Y}$的子集，那么智能体可以计算出 $\mathcal{Y}$的最大收缩量下界。

定理的完整证明12 见附录B。该证明利用了如下瞬时遗憾分解

$$r_t := f(\Theta_* x_*) - f(\Theta_* x_t) = \underbrace{f(\Theta_* x_*) - f(\tilde{\Theta}_t x_t)}_{\text{Term I}} + \underbrace{f(\tilde{\Theta}_t x_t) - f(\Theta_* x_t)}_{\text{Term II}}. \tag{4}$$

<escape>~</escape>项 I 刻画了乐观对(<escape>x_t</escape>，<escape> Θ_t</escape>) 相对于（3）的次优性，而项 II 则刻画了<escape>~</escape>置信集的收缩<escape> $\mathcal{C}_t$</escape>。由于(<escape>x_t</escape>，<escape> Θ_t</escape>) 可能是次优的，这是因为<escape>$\mathcal{G}_t$ </escape>是<escape> $\mathcal{X}$</escape>的真子集，而这对于确保安全性是必要的。尽管项 II 的分析可以采用传统的老虎机分析方法处理，但项 I 的分析需要运用新的技术，包括尖锐度，正如我们将在下一段中讨论的那样。

项 I 的界由以下引理给出。

引理14 假设 *1–4* 成立。对于 $t \in (T', T]$, 项 I 有如下界

$$\text{Term I} := f(\Theta_* x_*) - f(\tilde{\Theta}_t x_t) \leq M\text{Sharp}_{\mathcal{Y}}^{\infty}\left(\frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu + \lambda_- T'}}\right)$$

当 $T' \geq$ 最大值为（t_δ、 t_h）时，其发生的概率至少为 $1 - 2\delta$。

引理14的证明见附录 B.1，其中考虑了 $\mathcal{Y}$ 的收缩版本，使得收缩版本 $\mathcal{Y}$ 中的每个可能的 y 都可以表示为 Θx ，其中存在某个 $\Theta \in \mathcal{C}_t$ 和某个 $x \in \mathcal{G}_t$。这意味着，对于收缩版本 $\mathcal{Y}$ 中的每个 y， $f(\Theta x_t)$ 大于或等于 $f(y)$，因此我们可以用最优奖励（$f(y_*)$，其中 $y_* = \Theta_* x_*$）与收缩版本 $\mathcal{Y}$ 中某个 y 所获得的奖励之差来界定项 I。在对 f 作出利普希茨假设的情况下，该差值可以进一步用 y_* 与收缩版本 $\mathcal{Y}$ 中某个 y 之间的差来界定。令 y 为收缩版本 $\mathcal{Y}$ 中距离 y_* 最近的点，最终即可利用 $\mathcal{Y}$ 的尖锐度界定遗憾，如引理14所述。

5. 数值实验

我们在一个 $n = 3$且 $d = 3$ 的问题设定中，模拟了算法在三个尖锐度各异的多面体安全集上的结果。图 3 展示了算法在每个多面体安全集上的探索-利用阶段的瞬时遗憾累积和。实线表示六次试验的平均值，阴影区域表示试验结果的 95% 置信区间。对于每个安全集，图中展示了其 K常数，定义见命题 7。每次模拟中的噪声$\{\epsilon_t\}_{t\in[T]}$均有不同的实现。除此之外，每次模拟的问题参数和算法参数都相同—

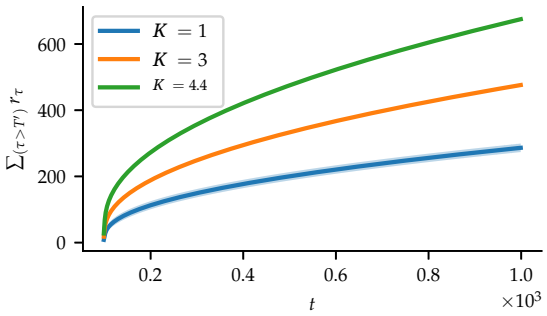


图 3：算法在探索-利用阶段的瞬时遗憾累积和，其中多面体约束集具有不同的 K 常数（定义见命题 7）。

ulation，以及所有多面体安全集都具有相同的最大收缩量。另外，注意动作集 $\mathcal{A}$ 被选择为非限制性的，使得 $\mathcal{Y} = \mathcal{E}$。图3 为定理12 中的次线性遗憾界提供了一些实证支持，同时也表明安全集的尖锐度会影响算法的遗憾。仿真的详细信息见附录C。

致谢

本研究得到美国国家科学基金会资助项目#1847096的支持。

参考文献

Yasin Abbasi-Yadkori、D´avid P´al 和 Csaba Szepesv´ari。线性随机老虎机的改进算法。神经信息处理系统进展，24，2011年。Sanae Amani、Mahnoosh Alizadeh 和 Christos Thrampoulidis。安全约束下的线性随机老虎机。神经信息处理系统进展，32，2019年。Xiaoqing Bai、Hua Wei、Katsuki Fujisawa 和 Yong Wang。用于最优潮流问题的半定规划。《国际电力与能源系统杂志》，30(6-7)：383–392，2008年。Felix Berkenkamp、Andreas Krause 和 Angela P Schoellig。带安全约束的贝叶斯优化：机器人学中的安全自动参数调优。《机器学习》，第1–35页，2021年。Stephen Boyd、Stephen P Boyd 和 Lieven Vandenberghe。凸优化。剑桥大学出版社，2004年。Sapana Chaudhary 和 Dileep Kalathil。具有未知线性安全约束的安全在线凸优化。载于*AAAI*人工智能大会论文集，第36卷，第6175–6182页，2022年。Shaoru Chen、Ning-Yuan Li、Victor M Preciado 和 Nikolai Matni。通过系统级综合实现时滞系统的鲁棒模型预测控制。*arXiv*预印本 *arXiv:2209.11841*，2022年。Varsha Dani、Thomas P Hayes 和 Sham M Kakade。老虎机反馈下的随机线性优化。2008年。Masoud Farivar 和 Steven H Low。支路潮流模型：松弛与凸化——第一部分。《*IEEE*电力系统汇刊》，28(3)：2554–2564，2013年。Mohammad Fereydounian、Zebang Shen、Aryan Mokhtari、Amin Karbasi 和 Hamed Hassani。不确定目标和约束下的安全学习。*arXiv*预印本 *arXiv:2006.13326*，2020年。Javier Garcıa 和 Fernando Fern´andez。安全强化学习综合综述。《机器学习研究杂志》，16(1)：1437–1480，2015年。Lukas Hewing、Kim P Wabersich、Marcel Menner 和 Melanie N Zeilinger。基于学习的模型预测控制：迈向控制中的安全学习。《控制、机器人与自主系统年度评论》，3：269–296，2020年。Spencer Hutchinson、Berkay Turan 和 Mahnoosh Alizadeh。用于带老虎机反馈的分布式资源分配的安全定价机制。载于*2022 IEEE*第*61*届决策与控制会议（*CDC*），第5092–5098页。IEEE，2022年。Sebastian Junges、Nils Jansen、Christian Dehnert、Ufuk Topcu 和 Joost-Pieter Katoen。马尔可夫决策过程的安全约束强化学习。载于系统构造与分析工具和算法国际会议，第130–146页。Springer，2016年。

Abbas Kazerouni、Mohammad Ghavamzadeh、Yasin Abbasi Yadkori 和 Benjamin Van Roy。保守型上下文线性老虎机。神经信息处理系统进展，30，2017年。Torsten Koller、Felix Berkenkamp、Matteo Turchetta 和 Andreas Krause。用于安全探索的基于学习的模型预测控制。载于*2018*年*IEEE*决策与控制会议，第6059–6066页。IEEE，2018。Daniel K Molzahn、Florian D¨orfler、Henrik Sandberg、Steven H Low、Sambuddha Chakrabarti、Ross Baldick 和 Javad Lavaei。电力系统分布式优化与控制算法综述。*IEEE*智能电网汇刊，8(6):2941–2962，2017年。Ahmadreza Moradipari、Christos Thrampoulidis 和 Mahnoosh Alizadeh。分阶段保守线性老虎机。神经信息处理系统进展，33:11191–11201，2020。Ahmadreza Moradipari、Sanae Amani、Mahnoosh Alizadeh 和 Christos Thrampoulidis。利用辅助信息的安全线性汤普森采样。*IEEE*信号处理汇刊，69: 3755–3767，2021。Aldo Pacchiano、Mohammad Ghavamzadeh、Peter Bartlett 和 Heinrich Jiang。带线性约束的随机老虎机。载于人工智能与统计国际会议，第2827–2835页。PMLR，2021。Rolf Sc hneider。凸体：布伦–闵可夫斯基理论。第151卷。剑桥大学出版社，2014。Yanan S ui 和 Joel W Burdick。应用于大型决策空间临床治疗的相关性对决老虎机。载于第*26*届国际人工智能联合会议论文集，第2793–2799页，2017年。Yanan Sui、Alkis Gotovos、Joel Burdick 和 Andreas Krause。利用高斯过程进行优化的安全探索。载于国际机器学习会议，第997–1005页。PMLR，2015年。Yanan Sui、Vincent Zhuang、Joel Burdick 和 Yisong Yue。基于高斯过程的分阶段安全贝叶斯优化。载于国际机器学习会议，第4781–4789页。PMLR，2018年。Ilnura Usmanova、Andreas Krause 和 Maryam Kamgarpour。不确定约束下的安全凸学习。载于第*22*届人工智能与统计国际会议，第2106–2114页。PMLR，2019年。Ilnura Usmanova、Andreas Krause 和 Maryam Kamgarpour。应用于策略搜索的安全非光滑黑盒优化。载于动力学与控制学习，第980–989页。PMLR，2020年。Zhenlin Wang、Andrew J Wagenmaker 和 Kevin Jamieson。带安全约束的最佳臂识别。载于人工智能与统计国际会议，第9114–9146页。PMLR，2022年。

附录 A. 第 4.1 节的证明 4.1

A.1. 命题的证明5

证明 首先证明“若”方向。根据定义， $\mathcal{D}$ 具有非空内部，当且仅当存在一点 $x \in \mathcal{D}$，使得对于某个 $\epsilon > 0$， $x + \mathcal{B}_2(\epsilon) \subseteq \mathcal{D}$，其中 $\mathcal{B}_2(\epsilon) = \{x \in \mathbb{R}^m : \|x\|_2 < \epsilon\}$ 是半径为 ϵ 的开球。由于 $\mathcal{D}$ 是闭集，这等价于条件 $x + \mathcal{B}_2(\epsilon) \subseteq \mathcal{D}$，其中 $\mathcal{B}_2(\epsilon) = \{x \in \mathbb{R}^m : \|x\|_2 \leq \epsilon\}$。有限维向量空间上的范数彼此等价，因此对于某个范数 $\|\cdot\|$，存在正常数 C，使得对所有 $y \in \mathcal{B}_{\|\cdot\|}(\tilde{\epsilon})$，都有 $C\|y\|_2 \leq \|y\| \leq \tilde{\epsilon}$。我们可以选取 $\tilde{\epsilon} = C\epsilon$，使得 $\mathcal{B}_{\|\cdot\|}(\tilde{\epsilon}) \subseteq \mathcal{B}_2(\epsilon)$，从而确保 $x + \mathcal{B}_{\|\cdot\|}(\tilde{\epsilon}) \subseteq \mathcal{D}$。因此，我们知道对于 $\Delta = \tilde{\epsilon} > 0$， $x + v \in \mathcal{D}, \forall v \in \mathcal{B}_{\|\cdot\|}(\Delta)$，进而 $\mathcal{D}_\Delta^{\|\cdot\|}$ 非空。

接下来，我们证明“仅当”方向。存在一个 $\Delta > 0$，使得 $\mathcal{D}_\Delta^{\|\cdot\|}$ 非空，当且仅当存在一点 $x \in \mathcal{D}_\Delta^{\|\cdot\|}$。点 x 必须满足 $x + v \in \mathcal{D}, \forall v \in \bar{\mathcal{B}}_{\|\cdot\|}(\Delta)$。由此可得，即 $x + \mathcal{B}_{\|\cdot\|}(\Delta) \subseteq \mathcal{D}$。根据前面讨论的范数等价，可知存在某个，使得 $\epsilon > 0$ ($\mathcal{B}_2$) $\subseteq \mathcal{B}_{\|\cdot\|}(\Delta)$，因此 $x + \mathcal{B}_2(\epsilon) \subseteq \mathcal{D}$。这意味着 $x + \mathcal{B}_2(\epsilon) \subseteq \mathcal{D}$，因为 $\mathcal{B}_2(\epsilon) \subset \mathcal{B}_2(\epsilon)$。因此， $\mathcal{D}$ 具有非空内部。

■

A.2. 命题的证明6

证明 令 $E := \{\Delta \geq 0 : \mathcal{D}_\Delta^{\|\cdot\|} \neq \emptyset\}$，并注意到根据定义， $H_{\mathcal{D}}^{\|\cdot\|} = \sup E$。首先，我们通过证明 E 非空且有上界，说明 $\sup E$ 确实存在。我们知道 E 非空，因为 $\mathcal{D}$ 具有非空内部，而命题 5 告诉我们，对于这样的集合，存在一个 $\Delta > 0$，使得 $\mathcal{D}_\Delta^{\|\cdot\|}$ 非空，因此 E 非空。此外，注意到 E 有上界，因为 $\mathcal{D}$ 有界（即存在有限的 r，使得对所有 $x, y \in \mathcal{D}$，都有 $\|x - y\|_2 \leq r$）。为此，先注意，对于任意 $\Delta \in E$ （即满足 $\mathcal{D}_\Delta^{\|\cdot\|} \neq \emptyset$ 的 Δ ），存在 $x \in \mathcal{D}$，使得对所有 $v \in \mathcal{B}_{\|\cdot\|}$ Δ $x + \tilde{v} \subseteq \mathcal{D}$ （ ），都有 $x + v \subseteq \mathcal{D}$；并且由于范数等价，对于某个有限的 $C > 0$，对所有 $\tilde{v} \in \mathcal{B}_2$（$C\Delta$）都有此性质。取 $\tilde{v} = \mathrm{e}_1 C\Delta$，我们有 x，且 $x + \mathrm{e}_1 C\Delta$ 属于 $\mathcal{D}$，因此 $\|x - (x + \mathrm{e}_1 C\Delta)\|_2 = C\Delta \leq r$。因此， E 有上界。由于 E 非空且有上界，$\sup E$ 存在。

接下来，我们证明 $H := H_{\mathcal{D}}^{\|\cdot\|} = \sup E$ 属于 E。根据上确界的性质，必然存在一个 $\Delta_\epsilon \in E$，使得 $\Delta_\epsilon > H - \epsilon$ 对每个 $\epsilon > 0$都成立。因此，对每个 $\epsilon \in (0, H]$ ，都有 $\mathcal{B}_{\|\cdot\|}(\Delta_\epsilon) \supset \mathcal{B}_{\|\cdot\|}(H - \epsilon)$，并且存在某个 x 使得 $x + \mathcal{B}_{\|\cdot\|}(\Delta_\epsilon) \subseteq \mathcal{D}$。由此可得， $x + \mathcal{B}_{\|\cdot\|}(H - \epsilon) \subset \mathcal{D}$ 对每个 $\epsilon \in (0, H]$都成立。我们声称这意味着 $x + \mathcal{B}_{\|\cdot\|}(H) \subset \mathcal{D}$。反设不然，即 $x + \mathcal{B}_{\|\cdot\|}(H) \not\subset \mathcal{D}$。这意味着存在 y $\|y\| < H$ 使得 and $x + y \notin \mathcal{D}$。由于 $x + y \notin \mathcal{D}$ $x + \mathcal{B}_{\|\cdot\|}$ $H - \epsilon \subset \mathcal{D}$ $\epsilon \in (0, H]$ y 且对每个 (, , 都有 ( )，可知 不可能属于 $\mathcal{B}_{\|\cdot\|}(H - \epsilon)$，因此 $\|y\| \not\leq H - \epsilon$，即 $\|y\| > H - \epsilon$。因此，对每个 $\epsilon \in (0, H]$ 都必须有 $H - \epsilon < \|y\| < H$。重新整理后，对每个 $\epsilon \in (0, H]$都必须有 $\epsilon > H - \|y\| > 0$，所以我们可以选择 $\epsilon = H - \|y\| > 0$ 来推翻前述结论。因此， $x + \mathcal{B}_{\|\cdot\|}(H) \subset \mathcal{D}$成立。由于 $\mathcal{D}$ 是闭集，它也包含 $x + \mathcal{B}_{\|\cdot\|}(H)$，因此，$H$ 属于 E。

最后，我们证明所有 $\Delta \in [0, H]$ E $x \in \mathcal{D}$ $x + \mathcal{B}_{\|\cdot\|}$ H $\subseteq$都属于。由于存在这样的，使得（ ） $\mathcal{D}$ 且 $\mathcal{B}_{\|\cdot\|}$（Δ） $\subseteq \mathcal{B}_{\|\cdot\|}$（$H$）对所有 $\Delta \in [0, H]$均成立，因此 $x + \mathcal{B}_{\|\cdot\|}$（$\Delta$） $\subseteq \mathcal{D}$也成立。

■

A.3. 命题7的证明

证明 首先，我们证明 $\mathcal{D}_{\Delta}^{\|\cdot\|}$ 是 $\mathcal{D}$中的一个多面体。为此，我们将第j个约束对应的半空间定义为 $\mathcal{D}_j = \{x \in \mathbb{R}^m : \ a_j^\top x \le b_j\}$ 使得 $\mathcal{D} = \bigcap_{j\in[p]} \mathcal{D}_j$。稍微滥用一下记号，令

$$\begin{aligned}
\mathcal{D}_{j,\Delta}^{\|\cdot\|} &:= \{x \in \mathbb{R}^m : a_j^\top (x+v) \le b_j, \forall v \in \bar{\mathcal{B}}_{\|\cdot\|}(\Delta)\} \\
&= \{x \in \mathbb{R}^m : \max_{v\in\bar{\mathcal{B}}_{\|\cdot\|}(\Delta)} a_j^\top (x+v) \le b_j\} \\
&= \{x \in \mathbb{R}^m : a_j^\top x + \Delta\|a_j\|_\star \le b_j\} \\
&= \{x \in \mathbb{R}^m : a_j^\top x \le b_j - \Delta\|a_j\|_\star\},
\end{aligned}$$

其中，$\|\cdot\|_\star$是 $\|\cdot\|$的对偶范数。因此，$\mathcal{D}_{\Delta}^{\|\cdot\|} = \bigcap_{j\in[p]} \mathcal{D}_{j\ \Delta}^{\|\cdot\|}$，是一个多面体，其每个组成约束都与 $\mathcal{D}$的相应约束平行，即对于所有 $j \in [p]$，$\mathcal{D}_{j\ \Delta}^{\|\cdot\|}$，都与 $\mathcal{D}_j$ 平行。然而，需要注意的是，某些约束 $\mathcal{D}_{j\ \Delta}^{\|\cdot\|}$，可能是冗余的。令 E 为这些非冗余约束构成的集合，即满足$\mathcal{D}_{\Delta}^{\|\cdot\|} = \bigcap_{j\in E} \mathcal{D}_{j\ \Delta}^{\|\cdot\|}$，的最小集合 $E \subseteq [p]$。为了研究 $\mathcal{D}$的尖锐度，我们将使用多面体 $\tilde{\mathcal{D}} := \bigcap_{j\in E} \mathcal{D}_j$。

~~多面体 $\mathcal{D}$ 很有用，因为对于 $\mathcal{D}$的每个非冗余约束，都存在 $\mathcal{D}_{\Delta}^{\|\cdot\|}$ 的一个平行的非冗余约束。我们还知道 $\mathcal{D} \subseteq \tilde{\mathcal{D}}$，因为

$$\mathcal{D} = \bigcap_{j\in[p]} \mathcal{D}_j = \left(\bigcap_{j\in E} \mathcal{D}_j\right) \cap \left(\bigcap_{j\in[p]\setminus E} \mathcal{D}_j\right) \subseteq \bigcap_{j\in E} \mathcal{D}_j = \tilde{\mathcal{D}}.$$

因此，有

$$\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) = \max_{x\in\mathcal{D}} \min_{y\in\mathcal{D}_{\Delta}^{\|\cdot\|}} \|x-y\|_2 \le \max_{x\in\tilde{\mathcal{D}}} \min_{y\in\mathcal{D}_{\Delta}^{\|\cdot\|}} \|x-y\|_2.$$

~~因此，我们研究 $\mathcal{D}$ 中投影距离最大的点，即

$$u \in \arg\max_{x\in\tilde{\mathcal{D}}} \underbrace{\min_{y\in\mathcal{D}_{\Delta}^{\|\cdot\|}} \|y-x\|_2}_{g(x)}.$$

由于 $\|y-x\|_2$ 在(x,y)上是凸的，我们知道 $g(x)$ 是凸函数（Boyd 等人 (2004) 第 3.2.5 节）。凸函数在多面体上的最大值出现在某个顶点处，[4] 因此 u 位于某个顶点~~处，该顶点属于 $\mathcal{D}$。

现在，我们对 u 与 $\mathcal{D}_{\Delta}^{\|\cdot\|}$ 中和 u 具有相同活跃约束的顶点之间的距离进行界定。为此，我们先回顾正文中的符号，并引入一些新符号。符号 $\mathcal{I}_A$ 指一组长度为 m 的集合，其中对于每个 $\ell = \{i_1, i_2, ..., i_m\} \in \mathcal{I}_A$，向量 a_{i_1}、a_{i_2}、.....、a_{i_m} 均线性无关。我们还使用符号 $A^\ell = [a_{i_1}\ a_{i_2}\ ...\ a_{i_m}]^\top$

~~4. 多面体中的每个点都可以表示为 $x = \sum_{i\in[n]} \lambda_i v_i$，$v_i$ 为顶点，且满足 $\lambda_i \ge 0$，$\sum_{i\in[n]} \lambda_i =$，其中 1 $f\ x\ = f\ \sum_{i\in[n]} \lambda_i v_i\ \le \sum_{i=[n]} \lambda_i f\ v_i\ \le\ _{i\in[n]} f\ v_i$ v_i 。由詹森不等式可得 ( ) ( ) ( ) max ( )。由于这些点都位于多面体中，最大值集合必定至少包含其中一个点。~~

以及 $b^\ell = [b_{i_1}\ b_{i_2}\ ...\ b_{i_m}]^\top$。注意，对于没有冗余约束的多面体的每个顶点，该顶点处都有 m 个线性无关约束处于活跃状态，这些约束构成 $\mathcal{I}_A$ 的一个元素。因此，对于顶点 v，我们使用符号 $A^v = [a_{i_1}\ a_{i_2}\ ...\ a_{i_m}]^\top$ 和 $b^v = [b_{i_1}\ b_{i_2}\ ...\ b_{i_m}]^\top$，其中 $\ell_v = \{i_1, i_2, ..., i_m\}$ 是该顶点处 m 个线性无关活跃约束构成的集合。现在，我们继续对 u 与顶点 $v \in \mathcal{D}_\Delta^{\|\cdot\|}$ 之间的距离进行界定，该顶点与 u 具有相同的活跃约束。由于在 u 和 v 处，相同的约束均取等号，因此有 $A^v u = b^v$ 和 $A^v v = b^v - \Delta\alpha_v$，其中 $\alpha_v := [\|a_{i_1}\|_\star\ ...\ \|a_{i_m}\|_\star]^\top$。因此，我们有 $u = (A^v)^{-1}b^v$ 和 $v = (A^v)^{-1}(b^v - \Delta\alpha_v)$。由此可得

$$\begin{aligned}\|u - v\|_2 &= \|(A^v)^{-1}b^v - (A^v)^{-1}\left(b^v - \Delta\alpha_v\right)\|_2 \\ &= \Delta\|(A^v)^{-1}\alpha_v\|_2 \\ &\le \Delta\|(A^v)^{-1}\|_2\|\alpha_v\|_2 \\ &= \Delta\frac{\sqrt{\sum_{i\in\ell_v}\|a_i\|_\star^2}}{\sigma_{min}(A^v)}\end{aligned}$$

现在，考虑上述分子：

$$\sqrt{\sum_{i\in\ell_v}\|a_i\|_\star^2} \le C_{\|\cdot\|}\sqrt{\sum_{i\in\ell_v}\|a_i\|_2^2} = C_{\|\cdot\|}\|A^v\|_F \le \sqrt{m}C_{\|\cdot\|}\|A^v\|_2 = \sqrt{m}C_{\|\cdot\|}\sigma_{max}(A^v)$$

其中 $C_{\|\cdot\|} = \max_{\|x\|_2=1}\|x\|_\star$ 使得 $\|x\|_\star \le C_{\|\cdot\|}\|x\|_2$ 对所有 $x \in \mathbb{R}^m$ 均成立。因此，我们有

$$\|u - v\|_2 \le \Delta\frac{\sqrt{\sum_{i\in\ell_v}\|a_i\|_\star^2}}{\sigma_{min}(A^v)} \le \sqrt{m}C_{\|\cdot\|}\frac{\sigma_{max}(A^v)}{\sigma_{min}(A^v)}\Delta = \sqrt{m}C_{\|\cdot\|}\kappa(A^v)\Delta,$$

其中 $\kappa(A^v)$ 是 A^v 的条件数。由于 v 属于 $\mathcal{D}_\Delta^{\|\cdot\|}$，我们有

$$\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) \le \min_{y\in\mathcal{D}_\Delta^{\|\cdot\|}}\|y - u\|_2 \le \|v - u\|_2 \le \sqrt{m}C_{\|\cdot\|}K\Delta$$

其中 $K = \max_{\ell\in\mathcal{I}_A}\kappa(A^\ell)$。我们进一步将 $C_{\|\cdot\|}$ 刻画为

$$C_{\|\cdot\|} = \max_{\|x\|_2=1}\|x\|_\star = \max_{\|x\|_2=1}\max_{\|y\|=1} y^\top x = \max_{\|y\|=1}\max_{\|x\|_2=1} y^\top x = \max_{\|y\|=1}\|y\|_2.$$

注意，对所有 $x \in \mathbb{R}^m$，$\|x\|_2 \le C_{\|\cdot\|}\|x\|$ 均成立。 ■

A.4. 命题10

首先，我们给出一个证明中需要用到的引理。

引理15 对于具有非空内部且满足 $\mathcal{A} \subseteq \mathcal{D}$ 的紧集 $\mathcal{A}$、$\mathcal{D} \subset \mathbb{R}^m$，对于所有 $x \in \mathcal{D}$，均有 $\min_{y\in\mathcal{D}_\Delta^{\|\cdot\|}}\|y - x\|_2 \le \min_{y\in\mathcal{A}_\Delta^{\|\cdot\|}}\|y - x\|_2$。

证明 首先，我们证明 $\mathcal{A}_\Delta^{\|\cdot\|} \subseteq \mathcal{D}_\Delta^{\|\cdot\|}$。对于任意 $a \in \mathcal{A}_\Delta$，有 $a + \bar{\mathcal{B}}_{\|\cdot\|}(\Delta) \subseteq \mathcal{A}$。由于 $\mathcal{A} \subseteq \mathcal{D}$，可得 $a + \bar{\mathcal{B}}_{\|\cdot\|}(\Delta) \subseteq \mathcal{D}$，从而 $a \in \mathcal{D}_\Delta^{\|\cdot\|}$。因此，$\mathcal{A}_\Delta^{\|\cdot\|} \subseteq \mathcal{D}_\Delta^{\|\cdot\|}$，从而对于所有 $x \in \mathcal{D}$，有 $\min_{y \in \mathcal{D}_\Delta^{\|\cdot\|}} \|y - x\|_2 \leq \min_{y \in \mathcal{A}_\Delta^{\|\cdot\|}} \|y - x\|_2$。 ■

于是，我们得到命题10的证明。

证明 为界定尖锐度，对于每个 $\mathcal{F}_x \in F_{\mathcal{D}}(x)$，我们任意选取一些，并使用引理15得到

$$\begin{aligned}
\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) &= \max_{x \in \mathcal{D}} \min_{y \in \mathcal{D}_\Delta^{\|\cdot\|}} \|y - x\|_2 \\
&\leq \max_{x \in \mathcal{D}} \min_{y \in \mathcal{F}_{x,\Delta}^{\|\cdot\|}} \|y - x\|_2 \\
&\leq \max_{x \in \mathcal{D}} \max_{z \in \mathcal{F}_x} \min_{y \in \mathcal{F}_{x,\Delta}^{\|\cdot\|}} \|y - z\|_2 \\
&= \max_{x \in \mathcal{D}} \mathrm{Sharp}_{\mathcal{F}_x}^{\|\cdot\|}(\Delta) \\
&\leq \Delta \sqrt{m} C_{\|\cdot\|} \max_{x \in \mathcal{D}} K_{\mathcal{F}_x}
\end{aligned}$$

这对于 $\Delta \in [0, \bar{H}_{\mathcal{F}}^{\|\cdot\|}]$均成立，其中 $\bar{H}_{\mathcal{F}}^{\|\cdot\|} := \text{最小值}_{x \in \mathcal{D}} H_{\mathcal{F}_x}^{\|\cdot\|} > 0$，因为根据定义和命题 5，每个 $\mathcal{F}_x$ 都具有¯非空内部。我们使用记号 $K_{\mathcal{F}} = \text{最大值}_{x \in \mathcal{D}} K_{\mathcal{F}_x}$。

随后，我们可以利用 $\mathcal{D}$ 的有界性，为尖锐度定义域的其余部分建立线性界，即对于 $\Delta \in (\bar{H}_{\mathcal{F}}^{\|\cdot\|}, H_{\mathcal{D}}^{\|\cdot\|}]$。由 $\mathcal{D}$的有界性可知，直径 $r_{\mathcal{D}} = \text{上确界}_{x,\, y \in \mathcal{D}} \|x - y\|_2$ 是有限的，因此我们得到平凡的尖锐度界 $\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) \leq r_{\mathcal{D}}$。这为所有 $\Delta \in (\bar{H}_{\mathcal{F}}^{\|\cdot\|}, H_{\mathcal{D}}^{\|\cdot\|}]$给出了线性界 $\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) \leq \frac{r_{\mathcal{D}}}{\bar{H}_{\mathcal{F}}^{\|\cdot\|}} \Delta$。因此，我们有

$$\mathrm{Sharp}_{\mathcal{D}}^{\|\cdot\|}(\Delta) \leq \max\left\{\sqrt{m} C_{\|\cdot\|} \bar{K}_{\mathcal{F}}, \frac{r_{\mathcal{D}}}{\bar{H}_{\mathcal{F}}^{\|\cdot\|}}\right\} \Delta \leq \sqrt{m} \bar{C}_{\|\cdot\|} \max\left\{\bar{K}_{\mathcal{F}}, \frac{r_{\mathcal{D}}}{\bar{H}_{\mathcal{F}}^{\|\cdot\|}}\right\} \Delta,$$

¯其中 $C_{\|\cdot\|} = \max(C_{\|\cdot\|}, 1)$。 ■

A.5. 推论的证明11

证明 根据命题 10，只需证明可以构造一个具有非空内部的多面体，使其包含 $\mathcal{D}$ 中的每个点，同时是 $\mathcal{D}$的子集。由于 $\mathcal{D}$ 具有非空内部，我们可以在 $\mathcal{D}$ 中构造一个具有非空内部的多面体 $\mathcal{A}$（例如超立方体）。然后，根据¯的凸性 $\mathcal{D}$，可以看出，多面体 $\mathcal{A} = \mathrm{conv}(\mathcal{A}, x)$ 包含 x，且对每个 $x \in \mathcal{D}$都包含于 $\mathcal{D}$ 中。因此，凸集 $\mathcal{D}$ 是多面体尖锐的，命题 10 的结论适用。 ■

A.6. 附加性质

命题 16 对于具有非空内部的紧集 $\mathcal{A}$、$\mathcal{D} \subset \mathbb{R}^m$，其中 $\mathcal{A} \subseteq \mathcal{D}$，我们有 $H_{\mathcal{A}}^{\|\cdot\|} \leq H_{\mathcal{D}}^{\|\cdot\|}$。

证明 任取 $\Delta \in [0, H_{\mathcal{A}}^{\|\cdot\|}]$。由命题 6 以及收缩集的定义，存在 $x \in \mathcal{A}_{\Delta}^{\|\cdot\|}$ 使得 $x + \mathcal{B}_{\|\cdot\|}(\Delta) \subseteq \mathcal{A}$。由于 $\mathcal{A} \subseteq \mathcal{D}$，可知 $x + \mathcal{B}_{\|\cdot\|}(\Delta) \subseteq \mathcal{D}$，因此 $\mathcal{D}_{\Delta}^{\|\cdot\|}$ 非空且 $\Delta \leq H_{\mathcal{D}}^{\|\cdot\|}$。由于这对所有 $\Delta \in [0, H_{\mathcal{A}}^{\|\cdot\|}]$均成立，因此 $H_{\mathcal{A}}^{\|\cdot\|} \leq H_{\mathcal{D}}^{\|\cdot\|}$。 ■

附录 B. 定理的证明 12

我们采用式（4）中给出的瞬时遗憾分解。因此，我们在第 B.1 节讨论项 I，在第 B.2 节讨论项 II，并在第 B.3 节讨论完整遗憾界。

B.1. 项 I（引理的证明14）

在开始引理 14的证明之前，我们需要 Amani 等人 （2019年）中的一个引理。

引理 17 *(引理 1，出自 Amani 等人 (2019年))* 以至少 $1-\delta$ 的概率，我们有

$$\lambda_{min}(V_{T'}) \geq \nu + \frac{\lambda_{-} T'}{2}, \tag{5}$$

对于 $T' \geq t_{\delta} := \frac{8L^2}{\lambda_{-}} \log(\frac{d}{\delta})$。

我们还需要证明 $\mathcal{Y}$ 的尖锐度定义良好，为此需要以下引理。

引理 18 对于具有非空内部的集合 $\mathcal{D} \subset \mathbb{R}^d$ 和满足 $n \leq d$ 的满秩矩阵 $N \in \mathbb{R}^{n \times d}$，集合 $\mathcal{H} = N\mathcal{D}$ 具有非空内部。

证明：考虑一个点 x_0，它位于 $\mathcal{D}$ 的内部。根据定义，存在一个半径为$\epsilon > 0$、以 x_0 为中心且位于 $\mathcal{D}$ 内的开球。我们可以在该开球内选取 $n+1$ 个点，x_0、x_1、...、x_n，由此定义多面体 $\text{conv}(x_0, x_1, ..., x_n) \subset \mathcal{D}$。这些点在 N 下的像定义为：对所有 $i \in \{0, 1, ..., n\}$，$v_i = Nx_i$，使得 $\text{conv}(v_0, v_1, ..., v_n) \subset \mathcal{H}$。如果向量 $v_0 - v_1$、$v_0 - v_2$、...、$v_0 - v_n$ 可以选为线性无关，那么 $\text{conv}(v_0, v_1, ..., v_n)$ 具有非空内部，因此 $\mathcal{H}$ 也具有非空内部。为说明这是可行的，首先注意，对于所有 $i \in \{1, 2, ..., n\}$，我们都可以选取 x_i，使得对于任意维度为d的单位向量 w_i，$x_0 - x_i = \frac{\epsilon}{2} w_i$。由此可得，对于所有 $i \in \{1, 2, ..., n\}$，$z_i = v_0 - v_i = N(x_0 - x_i) = \frac{\epsilon}{2} N w_i$。由于 N 的秩为 n，N 的像空间维数为 n，因此存在 w_1、w_2、...、w_n，使得z_1、z_2、...、z_n 线性无关（即像空间的一组基）。 ■

根据假设4，我们还可得 $\mathcal{G}^0$ 具有非空内部，这意味着 $\mathcal{X}$ 具有非空内部，因为 $\mathcal{X} \supset \mathcal{G}^0$。由于我们可以写成 $\mathcal{Y} = \Theta_* \mathcal{X}$ 且 $\mathcal{X}$ 具有非空内部，根据引理18，我们知道 $\mathcal{Y}$ 具有非空内部。我们还知道 $\mathcal{Y}$ 是紧致的，因为 $\mathcal{A}$ 和 $\mathcal{E}$ 都是紧致的，且有限维线性算子保持紧致性。由于 $\mathcal{Y}$ 是紧致的且具有非空内部，因此 $\mathcal{Y}$ 的尖锐度定义良好。

接下来我们给出引理14的证明。

证明 下文不再特别提及，我们令 Θ_* 属于 $\mathcal{C}_t$ 对所有 $t \in [T]$均成立，其成立概率至少为 $1-\delta$，由定理 1可知。同理，我们令 $\lambda_{\min}(V_{T'}) \geq \nu + \frac{\lambda_{-} T'}{2}$ 成立，其成立概率

至少为 $1-\delta$，根据引理 17，并且 $T' \geq t_\delta$。由并集界可知，以下各项$\Theta_* \in \mathcal{C}_t$, $\forall t \in [T]$ 以及 $\lambda_{\min}(V_{T'}) \geq \nu + \frac{\lambda_- T'}{2}$同时成立的概率至少为 $1-2\delta$。

F首先，我们定义未知参数的置信集的扩展版本，

$$\tilde{\mathcal{C}}_t := \left\{ [\theta^1\ \theta^2\ ...\ \theta^n]^\top \in \mathbb{R}^{n\times d} : \left\|\theta^i - \theta_*^i\right\|_{V_t} \leq 2\sqrt{\beta_t}, \forall i \in [n] \right\} \supseteq \mathcal{C}_t.$$

这是因为对于所有 Θ 属于 $\mathcal{C}_t$的情况，我们都有对于所有 $i \in [n]$ 而言

$$\left\|\theta^i - \theta_*^i\right\|_{V_t} = \left\|\theta^i - \hat{\theta}_t^i + \hat{\theta}_t^i - \theta_*^i\right\|_{V_t} \leq \left\|\theta^i - \hat{\theta}_t^i\right\|_{V_t} + \left\|\hat{\theta}_t^i - \theta_*^i\right\|_{V_t} \leq 2\sqrt{\beta_t}.$$

我们利用这一点定义一个收缩后的安全动作集:

$$\tilde{\mathcal{G}}_t := \{x \in \mathcal{A} : \Theta x \in \mathcal{E}, \forall \Theta \in \tilde{\mathcal{C}}_t\} \subseteq \mathcal{G}_t \tag{6}$$

然后我们使用 e 引理 17 来定义一个进一步收缩的安全价格集，使得对于 $t \geq T'$，我们有 th at

$$\bar{\mathcal{G}} = \{x \in \mathcal{A} : \Theta_* x + v \in \mathcal{E}, \forall v \in \bar{\mathcal{B}}_\infty(\ell)\} \subseteq \tilde{\mathcal{G}}_t, \tag{7}$$

其中 $\ell := \frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu+\lambda_- T'}}$ 要证明（7）成立，请注意，对于任意 x和 $\Theta \in \tilde{\mathcal{C}}_t$，对所有 $i \in [n]$，都有

$$\begin{aligned}
x^\top \theta^i &\in \left[x^\top \theta_*^i - 2\sqrt{\beta_t}\|x\|_{V_t^{-1}}, x^\top \theta_*^i + 2\sqrt{\beta_t}\|x\|_{V_t^{-1}} \right] \\
&\subseteq \left[x^\top \theta_*^i - \frac{2\sqrt{\beta_T}L}{\sqrt{\lambda_{min}(V_t)}}, x^\top \theta_*^i + \frac{2\sqrt{\beta_T}L}{\sqrt{\lambda_{min}(V_t)}} \right] \\
&\subseteq \left[x^\top \theta_*^i - \frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu+\lambda_- T'}}, x^\top \theta_*^i + \frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu+\lambda_- T'}} \right] \\
&= \left[x^\top \theta_*^i - \ell, x^\top \theta_*^i + \ell \right].
\end{aligned}$$

因此，对于所有 $v \in \mathcal{B}_\infty(\ell)$，任何满足 $\Theta_* x + v$ 属于 $\mathcal{E}$ 的 x，也都能确保 Θx 对所有 $\Theta \in \mathcal{C}_t$ 属于 $\mathcal{E}$，进而 $\mathcal{G} \subseteq \mathcal{G}_t$。于是我们有，

$$\begin{aligned}
\bar{\mathcal{Y}} := \{\Theta_* x : x \in \bar{\mathcal{G}}\} &= \{\Theta_* x : x \in \mathcal{A}\} \cap \{y : y + v \in \mathcal{E}, \forall v \in \bar{\mathcal{B}}_\infty(\ell)\} \\
&= \{\Theta_* x : x \in \mathcal{A}\} \cap \mathcal{E}_\ell^\infty \\
&\supseteq \mathcal{Y}_\ell^\infty,
\end{aligned}$$

其中 $\mathcal{Y} := \{\Theta_* x : \ x \in \mathcal{A}\} \cap \mathcal{E}$。我们还需要定义 $\tilde{\mathcal{Y}}^t := \{\Theta x : \ x \in \mathcal{G}_t, \ \Theta \in \mathcal{C}_t\}$。注意，对于 $t \geq T'$，有 $\tilde{\mathcal{Y}}^t \supseteq \bar{\mathcal{Y}} \supseteq \mathcal{Y}_\ell^\infty$。

为了投影到 $\mathcal{Y}_\ell^\infty$ 上，我们需要它非空。当安全探索阶段足够长，使得 $\ell \leq H_{\mathcal{Y}}^\infty$ 时，它便非空。因此，我们可以通过以下方式确保 T' 足够大。

$$\begin{aligned}
\ell = \frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu+\lambda_- T'}} &\leq H_{\mathcal{Y}}^\infty \\
T' \geq \frac{8\beta_T L^2}{\lambda_- \left(H_{\mathcal{Y}}^\infty\right)^2} - \frac{2\nu}{\lambda_-} &=: t_h
\end{aligned}$$

结合引理17，我们需要 $T' \geq \max(t_\delta,\ t_h)$。

现在我们可以直接证明该引理的结论。由于 (3) 对所有 y在 $\tilde{\mathcal{Y}}^t$和 $\tilde{\mathcal{Y}}^t \supseteq \mathcal{Y}_\ell^\infty$上都是乐观的，因此有 $f(\tilde{y}_t) \geq f(\bar{y})$，其中 $\tilde{y}_t :=\tilde{\Theta}_t x_t$，$y_* = \Theta_* x_*$以及 $\bar{y} \in \operatorname{argmin}_{y\in\mathcal{Y}_\ell^\infty} \|y_* - y\|_2$。由此可得

$$\begin{aligned}
\text{Term I} &:= f(y_*) - f(\tilde{y}_t) \\
&\leq f(y_*) - f(\bar{y}) \\
&\leq |f(y_*) - f(\bar{y})| \\
&\leq M\|y_* - \bar{y}\| \\
&\leq M\,\text{Sharp}_{\mathcal{Y}}^{\infty}(\ell) \\
&= M\,\text{Sharp}_{\mathcal{Y}}^{\infty}\left(\frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu + \lambda_- T'}}\right)
\end{aligned}$$

该引理的结论立即成立。 ■

B.2. 项 II

首先，我们需要一个引理，出自 Abbasi-Yadkori 等人(2011)。

引理 19 *11 2011* $\{x_t\}_{t=1}^\infty$ $\nu > 0$ $V_t = \nu I + \sum_{s=1}^t x_s [x_s]^\top$ *(*引理，出自 *Abbasi-Yadkori* 等人 *())* 对于 *with* 和，我们有

$$\sum_{t=1}^{T} \min(\|x_t\|^2_{[V_{t-1}]^{-1}}, 1) \leq 2(d\log((\mathit{trace}(\nu I) + TL^2)/d) - \log \mathit{det}(\nu I))$$

当 $\|x_t\|_2 \leq L$ 对所有 $t \in [T]$ 均成立时。

接下来，我们用以下引理对项 II 进行界定。

Lem引理 20 假设 *1–4* 成立。对于 $T > T' \geq \max(t_h, t_\delta)$，项 II 有如下界

$$\sum_{t=T'+1}^{T} \text{Term II} \leq M\max(LS, 1)\sqrt{n8\beta_T(T-T')d\log(1 + TL^2/(d\nu))}.$$

其成立的概率至少为 $1-\delta$。

证明 与引理 14 的证明类似，我们取 $\Theta_* \in \mathcal{C}_t$，$\forall t \in [T]$和 $\lambda_{\min}(V_{T'}) \geq \nu + \frac{\lambda_- T'}{2}$，它们同时成立的概率至少为 $1-2\delta$。

根据假设 3，对于 $t > T'$，我们有

$$\begin{aligned}
\text{Term II} :=& f(\tilde{\Theta}_t x_t) - f(\Theta_* x_t) \\
\leq& |f(\tilde{\Theta}_t x_t) - f(\Theta_* x_t)| \\
\leq& M\|\tilde{\Theta}_t x_t - \Theta_* x_t\|_2 \\
\leq& M\sqrt{n}\max_{i\in[n]}\left(x_t^\top \tilde{\theta}_t^i - x_t^\top \theta_*^i\right)
\end{aligned}$$

令 $\bar{r}_t^{II} := x_t^\top \tilde{\theta}_t^i - x_t^\top \theta_*^i$，使得

$$\begin{aligned}\bar{r}_t^{II} :=& x_t^\top \tilde{\theta}_t^i - x_t^\top \theta_*^i \\ \leq& \|x_t\|_{[V_{t-1}]^{-1}} \left\|\tilde{\theta}_t^i - \theta_*^i\right\|_{V_{t-1}} \\ \leq& \|x_t\|_{[V_{t-1}]^{-1}} \left\|\tilde{\theta}_t^i - \hat{\theta}_t^i + \hat{\theta}_t^i - \theta_*^i\right\|_{V_{t-1}} \\ \leq& 2\sqrt{\beta_t}\, \|x_t\|_{[V_{t-1}]^{-1}}.\end{aligned}$$

由于纯探索阶段，我们有

$$2\sqrt{\beta_t}\, \|x_t\|_{[V_{t-1}]^{-1}} \leq \frac{2\sqrt{\beta_T} L}{\sqrt{\lambda_{min}(V_{t-1})}} \leq \frac{2\sqrt{2\beta_T} L}{\sqrt{2\nu + \lambda_- T'}} \leq H_{\mathcal{Y}}^\infty.$$

因此，当 T 足够大，使得 $\beta_T \geq 1$ 时，以下界成立。

$$\begin{aligned}\bar{r}_t^{II} \leq& \min\left(H_{\mathcal{Y}}^\infty, 2\sqrt{\beta_t}\, \|x_t\|_{[V_{t-1}]^{-1}}\right) \\ \leq& 2\sqrt{\beta_T} \max(H_{\mathcal{Y}}^\infty, 1) \min\left(1, \|x_t\|_{[V_{t-1}]^{-1}}\right)\end{aligned}$$

接下来，我们可以使用引理 19，因为

$$\begin{aligned}\sum_{t=T'+1}^{T} [\bar{r}_t^{II}]^2 =& 4\beta_T \max((H_{\mathcal{Y}}^\infty)^2, 1) \sum_{t=T'+1}^{T} \min\left(1, \|x_t\|^2_{[V_{t-1}]^{-1}}\right) \\ \leq& 4\beta_T \max((H_{\mathcal{Y}}^\infty)^2, 1) \sum_{t=1}^{T} \min\left(1, \|x_t\|^2_{[V_{t-1}]^{-1}}\right) \\ \leq& 8\beta_T \max((H_{\mathcal{Y}}^\infty)^2, 1)(d \log((\mathsf{trace}(\nu I) + TL^2)/d) - \log \det(\nu I)) \\ =& 8\beta_T \max((H_{\mathcal{Y}}^\infty)^2, 1)(d \log((d\nu + TL^2)/d) - d\log(\nu)) \\ =& 8\beta_T \max((H_{\mathcal{Y}}^\infty)^2, 1) d \log(1 + TL^2/(d\nu)).\end{aligned}$$

因此，应用柯西-施瓦茨不等式可得

$$\begin{aligned}\sum_{t=T'+1}^{T} \bar{r}_t^{II} \leq& \sqrt{(T - T') \sum_{t=T'+1}^{T} [r_{t,i}^{II}]^2} \\ \leq& \max(H_{\mathcal{Y}}^\infty, 1) \sqrt{8\beta_T (T - T') d \log(1 + TL^2/(d\nu))}.\end{aligned}$$

我们由此得到，

$$\sum_{t=T'+1}^{T} \text{Term II} \leq M \max(H_{\mathcal{Y}}^\infty, 1) \sqrt{nd 8\beta_T (T - T') \log(1 + TL^2/(d\nu))}.$$

■

B.3. 完整遗憾界

最后，我们可以证明定理12。

证明 首先，瞬时遗憾的平凡上界为

$$r_t := f(\Theta_* x_*) - f(\Theta_* x_t) \le M\|\Theta_* x_* - \Theta_* x_t\|_2 \le M\sqrt{n}\max_{i\in[n]}\left(x_*^\top \theta_*^i - x_t^\top \theta_*^i\right) \le 2M\sqrt{n}LS$$

我们在最初的 T' 个时间步中使用这个平凡上界。因此，我们可以按时间分解遗憾，并利用引理 14 和引理20 得到

$$\begin{aligned} R_T &= \sum_{t=1}^{T'} r_t + \sum_{t=T'+1}^{T} (\text{Term I}) + \sum_{t=T'+1}^{T} (\text{Term II}) \\ &= 2M\sqrt{n}LST' + M\max(H_{\mathcal{Y}}^{\infty}, 1)\sqrt{n8\beta_T(T-T')d\log(1+TL^2/(d\nu))} \\ &\quad + M(T-T')\text{Sharp}_{\mathcal{Y}}^{\infty}\left(\frac{2\sqrt{2\beta_T}L}{\sqrt{2\nu+\lambda_- T'}}\right) \end{aligned}$$

■

附录 C. 数值实验

数值实验在动作维度为 $d = 3$ 、响应维度为 $n = 3$ 的设置下进行。奖励函数为线性函数，即 $f(y) = a^\top y$，其中 $a \in \mathbb{R}^n$。为系统地研究尖锐度的影响，我们采用形式为 $\mathcal{E}_b = \{y \in \mathbb{R}^n : \|\text{diag}(b)y\|_1 \le 1\}$ 的安全集，其中 $b \in \mathbb{R}^n$。通过适当选择 b，我们可以在不改变最大收缩量的情况下调整集合的尖锐度。动作集 $\mathcal{A}$ 的选择不受限制，使得 $\mathcal{Y} = \mathcal{X}$。

我们研究的三个安全集分别为 $\mathcal{E}_{b_1}$、 $\mathcal{E}_{b_2}$ 和 $\mathcal{E}_{b_3}$，其中 $b_1 = [0.1, 0.1, 0.1]$、 $b_2 = [0.1/2, 0.1, 0.1]$ 和 $b_3 = [0.1/3, 0.1, 0.1]$。请注意， $\mathcal{E}_{b_1}$、 $\mathcal{E}_{b_2}$ 和 $\mathcal{E}_{b_3}$ 的 2-范数最大收缩量均为 10。为单独考察尖锐度的影响，我们通过选择 $a = \mathbf{e}_1$，确保最优响应 $y_* = \theta_* x_*$ 位于多面体中“最尖锐”的角上。参数 Θ 通过随机选取各元素 $\Theta^{i,j} \sim U[-1, 1]$ 进行采样，其中所有 $i \in [n]$ 和 $j \in [d]$ 均如此。对于所有 $i \in [n]$，噪声 ϵ_t 的每个元素均从 $\epsilon_t^i \sim N(\sigma^2)$ 中采样，其中 $\sigma = 10^{-3}$。此外，我们选择 $T = 10^3$。另外， $S = \|\Theta_*\|_2 + 0.1$ 和 $L = \max^{b\in\{b_1,b_2,b_3\}} \max_{x\in\mathcal{E}_b} \|x\|_2 + 0.1$。算法参数为 $\nu = 0.1$、 $T' = T^{2/3} = 10^2$ 和 $\delta = 0.01$。

每个时间步中，式（3）中的乐观动作通过以下方式计算：使用 ℓ_1 置信集，该置信集是 $\mathcal{C}_t$ 的一个外逼近，这与Dani 等人(2008）和Amani 等人(2019年）采用的方法相同。具体而言，我们使用Amani 等人(2019年）中详述的方法，概述如下。由于 f 是线性的，式（3）可以通过枚举 ℓ_1 置信集的顶点，并根据每个顶点处的参数优化奖励来求解。最优奖励是各顶点处奖励的最大值，而乐观动作则是使该奖励达到最大的动作。