arXiv:2601.05444v1 [math.ST] 9 Jan 2026

XGBoost 学习什么函数?

Dohyeong Ki*和 Adityanand Guntuboyina†

加州大学伯克利分校统计系

摘要

本文为 XGBoost 隐式学习的函数类建立了严谨的理论基础，弥合了其经验成功与理论理解之间的差距。我们引入了一个无限维函数类 $\mathcal{F}^{d,\infty}_{-}{}^{s}$ ST，它扩展了有限个有界深度回归树的集成；同时引入复杂度度量 $V^{d,\infty}_{-}{}^{s}$ XGB$(\cdot)$，将 XGBoost 中使用的 L^1 正则化惩罚项推广开来。我们证明，XGBoost 目标函数的每一个优化器，也是带有惩罚项 $V^{d,\infty}_{-}{}^{s}$ XGB$(\cdot)$的等价惩罚回归问题在 $\mathcal{F}^{d,\infty}_{-}{}^{s}$ ST 上的优化器，从而将 XGBoost 解释为隐式地针对一个更广泛的函数类。我们还根据 Hardy–Krause 变差，从平滑性的角度阐释了 $\mathcal{F}^{d,\infty}_{-}{}^{s}$ ST 和 $V^{d,\infty}_{-}{}^{s}$ XGB$(\cdot)$。我们证明，在 $\{f \in \mathcal{F}^{d,\infty}_{-}{}^{s}\text{ST} : V^{d,\infty}_{-}{}^{s}\text{XGB}(f) \leqslant V\}$ 上的最小二乘估计量达到了几乎极小极大最优的收敛速率 $n^{-2/3}(\log n)^{4(\min(s,d)-1)/3}$，，从而避免了维数灾难。我们的结果首次严谨刻画了 XGBoost 所依托的函数空间，澄清了其与经典变差概念之间的联系，并指出了一个重要的开放问题：XGBoost 算法本身是否能在该函数类上达到极小极大最优性。

MSC 2010 主题分类：主要分类 62G08。

关键词与短语：梯度提升、Hardy–Krause 变差、L1 正则化、混合偏导数、非参数回归、树集成方法。

1 引言

考虑具有数据 $(\mathbf{x}^{(1)},\ y_1),\ \ldots\ldots,\ (\mathbf{x}^{(n)},\ y_n)$ 的标准回归问题，其中每个 $\mathbf{x}^{(i)} \in \mathbb{R}^d$ 和 $y_i \in \mathbb{R}$。XGBoost 由 Chen 和 Guestrin [10], 提出，它通过（近似）最小化由最小二乘损失和正则化惩罚项组成的目标函数，拟合有限个回归树的和。下面我们将介绍 XGBoost 试图求解的优化问题；有关实现的更多细节，请参阅官方文档 [37] 。

XGBoost 使用右连续分裂构建单棵回归树，这意味着每次分裂的形式为 $x_j \geqslant t_j$ 与 $x_j < t_j$，其中 x_j 表示协变量向量$\mathbf{x}$的第 j^{th} 个坐标。每棵树还受到用户指定的最大深度限制，该限制由超参数 max depth 控制（默认值为 6）。请记住，树的深度指的是任意一条根到叶路径上的最大分裂次数。令 $\mathcal{F}^{d,\ s}{}_{\text{ST}}$ 表示所有深度至多为 s 的有限个右连续[1] 回归树之和构成的类（此处 ST 表示“树之和”）。更准确地说，$\mathcal{F}^{d,\ s}{}_{\text{ST}}$ 由所有形如以下形式的函数构成

*dohyeong_ki@berkeley.edu

†aditya@stat.berkeley.edu[1]本文中，“右连续”一词指逐坐标右连续性。

$\sum_{k=1}^{K} f_k$ 其中 $K \geqslant 1$，且每个 f_k 都是具有右连续分裂、深度为 $\leqslant s$的回归树。每个f_k 也可以是常值函数，此时我们称其为常值回归树，并称其深度为 0。对于 K，除了要求其有限之外，我们不作任何限制。在实现中，XGBoost 允许用户为 K指定一个上界，通常在几百到几千的量级。由于这个上界通常设得较大，为了理论上的便利，我们在定义 $\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$ 时不限制 K。

对于回归，XGBoost 在函数类 $\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$ 上最小化最小二乘损失，并加入下文所述的显式正则化惩罚项。这种显式正则化是一项关键创新，使 XGBoost 有别于早期的梯度提升方法，例如梯度提升机（见 Friedman [14]），以及随机森林等集成方法（见 Breiman [7]）。

对于函数 $f \in \mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$，假设 f 可表示为有限个树的和，并令 $\mathbf{w}_k$ 表示与第 k^{th} 棵树相关联的叶节点权重向量。XGBoost 的正则化惩罚项形式为 $\alpha \sum_k \|\mathbf{w}_k\|_1$，其中 $\alpha > 0$ 控制正则化强度，$\|\cdot\|_1$ 表示 L^1 范数。如果第 k^{th} 棵树是常数树，则设 $\|\mathbf{w}_k\|_1 = 0$。该惩罚项通过约束叶节点权重的大小，抑制过于复杂的树。由于每个函数 $f \in \mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$ 通常都可以表示为有限个树之和，且量 $\sum_k \|\mathbf{w}_k\|_1$ 取决于所选的具体表示方式，因此，我们可以通过对所有可能的树分解取下确界，得到与表示方式无关的复杂度度量。具体而言，对于 $f \in \mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$，定义

$$V_{\mathrm{XGB}}^{d,s}(f) = \inf\Big\{ \sum_k \|\mathbf{w}_k\|_1 \Big\} \tag{1}$$

其中，下确界取遍所有将 f 表示为具有右连续切分且深度至多为 s的回归树有限和的表示，$\mathbf{w}_k$ 表示该 k^{th} 树的叶节点权重向量。

在此记号下，XGBoost 是一种用于求解以下优化问题的贪心算法:

$$\operatorname*{argmin}_f \Big\{ \sum_{i=1}^{n} \big(y_i - f(\mathbf{x}^{(i)})\big)^2 + \alpha V_{\mathrm{XGB}}^{d,s}(f) : f \in \mathcal{F}_{\mathrm{ST}}^{d,s} \Big\}. \tag{2}$$

此外，通常还会加入形如 $\gamma \sum_k T_k$ 的叶节点数量惩罚项，其中 T_k 是 k^{th} 树中的叶节点数量。由于 γ 的默认值为 $\gamma = 0$ （见 [37]），因此全文省略该项。一些实现还会将 L^1 惩罚项替换为平方 L^2 惩罚项 $\|\mathbf{w}_k\|_2^2$。然而，对于树的和而言，这种惩罚项并未得到良好定义。为说明这一问题，考虑函数 $(x_1, \ldots, x_d) \mapsto \mathbf{1}(x_1 \geqslant 0)$，其平方 L^2 惩罚项为 1。相同的函数也可以表示为 $\sum_{k=1}^{K} (1/K) \cdot \mathbf{1}(x_1 \geqslant 0)$，，此时总平方 L^2 惩罚项等于 $1/K$，，并且当 $K \to \infty$时趋于零。与 L^1 惩罚项不同，平方 L^2 惩罚项可以通过适当增加分解中的树的数量而任意接近于零，这使得平方 L^2 惩罚项对于树集成而言定义不适定。因此，我们仅关注 L^1 惩罚项（1）。更广泛地说，在高维统计学中，L^1 相较于 L^2惩罚项的优势已得到充分确立（例如，参见 Donoho 和 Johnstone [11], Johnstone[20], Tibshirani [31], 以及 Tibshirani [32]）。第6.3节进一步讨论了树集成中 L^1 与 L^2惩罚项之间的差异。

XGBoost 已成为使用最广泛的机器学习算法之一，因其预测准确性和效率而广受赞誉。事实上，它在许多备受瞩目的机器学习竞赛中发挥了决定性作用，从业者也常指出，对于表格数据（包括我们的回归数据设定），XGBoost 的表现可以优于深度学习方法（例如，参见 Borisov 等 [5],Grinsztajn 等 [18], Shwartz-Ziv 和 Armon [28]）。然而，尽管 XGBoost 在实证上取得了成功，其理论性质仍缺乏充分理解。本文通过提供关于 XGBoost 目标函数（2）解的行为的理论洞见，朝着弥合这一差距迈出了一步。

我们的主要贡献是构造了一个函数类 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$ 以及与之相关的复杂度度量 $V^{d,\ \overset{s}{\infty-}\mathrm{XGB}}(\cdot)$，该函数类和复杂度度量具有以下性质：

1. 当 $\mathcal{F}^{d,s}\mathrm{ST} \subsetneq \mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$ 且 $V^{d,\overset{s}{\infty-}\mathrm{XGB}}(f) = V^{d,s}\mathrm{XGB}(f)$ 时，$f \in \mathcal{F}^{d,s}\mathrm{ST}$。换言之，$\mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$ 是一个严格大于 $\mathcal{F}^{d,s}\mathrm{ST}$ 的函数类，而 $V^{d,\overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 则是将 $V^{d,s}\mathrm{XGB}(\cdot)$ 扩展到这个更大的函数类。事实上，$\mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$ 包含许多连续函数，不同于仅包含分段常数函数的 $\mathcal{F}^{d,s}\mathrm{ST}$。

2. XGBoost 优化问题 (2) 的每个解也都是以下问题的解：

$$\operatorname*{argmin}_{f}\Big\{\sum_{i=1}^{n}\big(y_i - f(\mathbf{x}^{(i)})\big)^2 + \alpha V^{d,s}_{\infty-\mathrm{XGB}}(f) : f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\Big\}. \tag{3}$$

3. 在具有随机设计和平方误差损失的标准回归假设下，在函数类上的极小极大收敛速率

$$\big\{f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}} : V^{d,s}_{\infty-\mathrm{XGB}}(f) \leqslant V\big\} \tag{4}$$

对于固定的 $V > 0$ s，满足

$$\Omega(n^{-2/3}(\log n)^{2(\min(s,d)-1)/3}) \ \leqslant\ \text{minimax rate} \ \leqslant\ O(n^{-2/3}(\log n)^{4(\min(s,d)-1)/3}), \tag{5}$$

其中 $\Omega(\cdot)$ 和 $O(\cdot)$ 符号中的常数取决于 d，s 和 V。特别是，这两个界都随 V 增大而增大，说明随着 V 增大，极小极大收敛速率会变差，这符合直觉。

(5) 中的上界可由在 (4) 上的最小二乘估计量达到；这可以看作是求解 (3) 的一个受约束版本。根据标准的对偶论证，对于每个 $V > 0$，都存在一个 α（可能同时取决于 V 和数据），使得问题 (3) 的解能够达到 (5) 中的上界。

综合这些结果可以看出，尽管 XGBoost 拟合得到的解是由有限个回归树相加构成的，但仍可将其解释为隐式地针对更大、表达能力更强的函数类 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$。式 (5) 中的快速收敛速率不受通常维数灾难的影响，这表明，只要函数的复杂度 $V^{d,\ \overset{s}{\infty-}\mathrm{XGB}}(f^*)$ 不太高，XGBoost 就能准确估计函数 $f^* \in \mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$。这种观点至少在一定程度上为 XGBoost 在实践中的出色实证表现提供了理论解释：尽管现实世界中的回归函数很少是分段常数的，只要底层函数属于复杂度适中的 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$，XGBoost 就能表现良好。

与此同时，我们的结果也指出了 XGBoost 的潜在局限。如果真实回归函数 f^* 无法由 $\mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$ 中的元素在受控 $V^{d,\overset{s}{\infty-}\mathrm{XGB}}(f)$ 的条件下得到良好近似，那么就不应期待能够准确估计（更多详情请参见第 6.2 节）。最后，我们强调，本文不讨论 XGBoost 的算法方面。我们的结果刻画了 XGBoost 旨在优化的目标函数的解所具有的统计性质，而不保证该算法的某个特定实现能够达到这些速率（更多详情请参见 6.4）。

我们还给出了 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$ 一种基于光滑性的刻画，该刻画不依赖于与树的任何显式联系。具体而言，我们证明 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$ 与具有有限 Hardy–Krause（HK）变差的函数类密切相关。更确切地说，在命题 3 中，我们证明 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$ 与以下函数类一致：

具有有限 HK 变差且不存在阶数高于 s 的交互作用的右连续函数，具体定义见（11）。

HK 变差是多元全变差的一个经典概念（参见 Aistleitner和Dick [1], Leonov[23],Owen [24]），可解释为一种光滑度度量。事实上，对于足够光滑的函数f，HK 变差与 f 最高阶为一的混合偏导数的 L^1 范数密切相关，即

$$\frac{\partial^{r_1+\cdots+r_d} f}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}} \qquad \text{with } \max_j r_j = 1.$$

精确关系见第2.1节中的公式 (10)。我们还表明，复杂度度量$V^{d,\ \infty-^{s}\ \mathrm{XGB}}(\cdot)$ 与 HK 变差紧密相关：它的上界为 HK 变差，下界则为 HK 变差乘以一个常数（该常数取决于 s 和 d）（参见命题 5）。

这一视角表明，XGBoost 可被视为执行光滑度约束的非参数回归，其中光滑度通过控制最高阶为一的混合偏导数来量化。XGBoost 等基于树的方法通常被归类为“算法建模”传统，与统计建模不同（参见 Breiman [8]）。相比之下，我们的结果将 XGBoost 明确置于传统统计框架之中，即由可解释的光滑度惩罚项支配的正则化估计框架。

HK 变差此前已在 Fang 等人 [12]以及 Benkeser 和 van der Laan [2], Schuler等人 [25], van der Laan 等人 [34] 的非参数回归研究中被用作正则化惩罚项（后一组论文将该方法称为“高度自适应Lasso”）。然而，HK 变差缺乏对称性，因此作为正则化惩罚项显得不太自然。例如，当 $d = 2$时，指示函数 $\mathbf{1}(x_1 \geqslant t_1,\ x_2 \geqslant t_2)$和 $\mathbf{1}(x_1 < t_1,\ x_2 < t_2)$的 HK 变差值不同。这种不对称性源于 HK 变差需要指定一个锚点 $[1, 24],$，而任何特定的锚点选择都会打破对称性。以往研究 [2, 12] 通常将定义域的左下角作为锚点（在我们的设定中为 $(-\infty,\ \ldots,\ -\infty)$），但原则上，任何满足 $a_j \in \{-\infty,\ +\infty\}$的点 $(a_1,\ \ldots,\ a_d)$都可以用作锚点。所有这些选择都会使所得的 HK 变差呈现某种不对称性（见第 3.1 节）。

相比之下，复杂度度量 $V^{d,\ \infty-^{s}\ \mathrm{XGB}}(\cdot)$ 不存在这种对称性缺失的问题：上述两个指示函数在 $V^{d,\ \infty-^{s}\ \mathrm{XGB}}(\cdot)$下获得相同的值。由于这种对称性，$V^{d,\ \infty-^{s}\ \mathrm{XGB}}(\cdot)$比 HK 变差提供了更自然的正则化项。此外，由于对于任意锚点的选择，$V^{d,\ \infty-^{s}\ \mathrm{XGB}}(\cdot)$ 都一致小于 HK 变差，使用它可以避免过度收缩，同时仍能有效控制模型复杂度。

本文其余部分的组织如下。第2节和第3节介绍函数类 $\mathcal{F}^{d,\ \infty-^{s}\ \mathrm{ST}}$和复杂度度量 $V^{d,\ \infty-^{s}\ \mathrm{XGB}}(\cdot)$，并描述它们与 Hardy–Krause 变差之间的联系。第4节研究 XGBoost 优化问题 (2) 与在更广泛函数类 $\mathcal{F}^{d,\ \infty-^{s}\ \mathrm{ST}}$ 上的优化问题 (3) 之间的关系。第5节分析 (4) 上的极小极大收敛速率。讨论部分强调了若干与我们的主要结果相关的问题。所有结果的证明见附录。

2 函数类 $\mathcal{F}^{d,\ \infty-^{s}}$ ST

我们对 $\mathcal{F}^{d,\infty-^{s}\ \mathrm{ST}}$ 的定义，建立在与回归树相关的一类特定基函数之上。这些基函数的形式为

$$b_{\mathbf{l},\mathbf{u}}^{L,U}(x_1, \ldots, x_d) = \prod_{j \in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j \in U} \mathbf{1}(x_j < u_j), \tag{6}$$

其中， L 和 U 是 $[d]$ 的（不一定互不相交的）子集：$=\{1,\ldots,d\}$，且 $0<|L|+|U|\leqslant s$（其中 $|\cdot|$ 表示集合的基数）；$\mathbf{l}:=(l_j, j\in L)$ 和 $\mathbf{u}:=(u_j, j\in U)$ 是实值阈值向量。由于阈值可以取任意实数，基函数（6）的集合是不可数的。

任何具有右连续分裂且深度为 $\leqslant s$ 的非常数回归树，以及由此类树构成的任意有限和，都可以表示为这些基函数的有限线性组合。此表示是通过将树分解为与从根节点到各个叶节点的单条路径相对应的指示函数而得到的。对于每条从根节点到叶节点的路径，取

$$L:=\{j\in[d]:\ \text{该路径至少包含一个形式为 } x_j\geqslant t\},\ \text{的分裂}$$

$$U:=\{j\in[d]:\ \text{该路径至少包含一个形式为 } x_j< t\},\ \text{的分裂}$$

对于每个 $j\in L$（分别是 $j\in U$），令 l_j（分别是 u_j）取路径上这些分裂中出现的阈值 t 的最大值（分别是最小值）。请注意，如果某个坐标沿同一路径在两个方向上都发生分裂，那么它可能同时属于 L 和 U。因此，$|L|+|U|$ 不超过树的深度，而树的深度至多为 s。

由于阈值向量 $\mathbf{l}$ 和 $\mathbf{u}$有不可数多种选择，因此使用有符号测度来表示 $b^{L,U}{}_{\mathbf{l},\mathbf{u}}$ 的有限线性组合会更方便。更准确地说，可以用有符号测度对不同阈值向量 $\mathbf{l}$和 $\mathbf{u}$对应的 $b^{L,U}{}_{\mathbf{l},\mathbf{u}}$ 的系数进行编码。对于有限符号博雷尔测度 $\nu_{L,U}$（以 $L,U\subseteq[d]$ 为指标，且满足 $0<|L|+|U|\leqslant s$），定义在 $\mathbb{R}^{|L|+|U|}$和 $c\in\mathbb{R}$上，定义

$$f^{d,s}_{c,\{\nu_{L,U}\}}(x_1,\ldots,x_d)=c+\sum_{0<|L|+|U|\leqslant s}\int_{\mathbb{R}^{|L|+|U|}} b^{L,U}_{\mathbf{l},\mathbf{u}}(x_1,\ldots,x_d)\,d\nu_{L,U}(\mathbf{l},\mathbf{u}). \tag{7}$$

这个表达式提供了一种简单、统一的方式来表示基函数$b^{L,U}{}_{\mathbf{l},\mathbf{u}}$ 的有限线性组合。 $b^{L,U}{}_{\mathbf{l},\mathbf{u}}$ 的任何有限线性组合——因此 $\mathcal{F}^{d,s}{}_{\mathrm{ST}}$中的每个元素——都可以写成这种形式，其中离散有符号测度 $\nu_{L,U}$ 的支撑点有限。下一个结果（证明见附录 A.1.1）表明，反过来也成立：所有这类由具有有限支撑的离散有符号测度 $\nu_{L,U}$ 构成的函数 (7)，都属于 $\mathcal{F}^{d,s}{}_{\mathrm{ST}}$。

命题 1。 所有深度至多为 s 的右连续回归树的有限和所构成的类 $\mathcal{F}^{d,\ s}_{ST}$ 可刻画为

$$\mathcal{F}^{d,s}_{ST}=\big\{f^{d,s}_{c,\{\nu_{L,U}\}}:\nu_{L,U}\ \textit{are discrete signed measures with finitely many support points}\big\}.$$

根据命题 1，通过允许 (7) 中的 $\nu_{L,\ U}$ 为任意（即不一定是离散的）有限符号博雷尔测度，可以自然地扩展 $\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$ 。由此得到函数类 $\mathcal{F}^{d,\ s}_{\infty-}{}_{\mathrm{ST}}$。

定义 1。 对于固定的 $d\geqslant 1$ 和 $s\geqslant 1$，$\mathcal{F}^{d,\ s}_{\infty-ST}$由所有函数 $f^{d,\ s}_{c\ ,\ \{\nu_{L,\ U}\}}$（定义于(7)）构成，其中 $c\in\mathbb{R}$，且每个$\nu_{L,\ U}$都是定义在 $\mathbb{R}^{|L|+|U|}$ 上的有限符号博雷尔测度。

Th以下结果（证明见附录 A.1.6）记录了 $\mathcal{F}^{d,\ s}_{\infty-}{}_{\mathrm{ST}}$ 的一些基本性质 .

命题 2。 *(a)* $\mathcal{F}^{d,\ s}_{\infty-ST}$ 中的每个函数都是右连续的。

(b) 对于 $s_1\leqslant s_2$、$\mathcal{F}^{d,\ s_1}_{\infty-ST}\subseteq\mathcal{F}^{d,\ s_2}_{\infty-ST}$。
(c) 对于每个 $s\geqslant d$、$\mathcal{F}^{d,\ s}_{\infty-ST}=\mathcal{F}^{d,\ d}_{\infty-ST}$。 *(d)* 函数类 $\mathcal{F}^{d,\ s}_{\infty-ST}$ 是凸的。

接下来，我们将说明 $\mathcal{F}^{d,\ s}_{\infty-}{}_{\mathrm{ST}}$ 可以通过 Hardy–Krause（HK）变差来刻画。为此，我们先回顾 HK 变差的定义。

2.1 Hardy–Krause（HK）变差

Hardy–Krause（HK）变差通常用于定义在紧致定义域上的函数（参见 Aistleitner和Dick [1], Leonov [23], Owen [24]），但本文研究的是定义在整个空间 $\mathbb{R}^d$ 上的函数。因此，我们对标准定义稍作修改，以适应无界定义域 $\mathbb{R}^d$。在介绍我们版本的 HK 变差之前，我们先回顾 Vitali变差，它是 HK 变差的重要构成基础。

定义 2（拟体积）。设 g 是定义在 $\mathbb{R}^m$ 上的实值函数。对于 $(u_1,\dots,u_m),(v_1,\dots,v_m)\in\mathbb{R}^m$，其中对于所有 $j\in[m]$ 均有 $u_j<v_j$，g 在矩形 $\prod_{j=1}^m[u_j,v_j]$ 上的拟体积定义为

$$\Delta\Big(g;\prod_{j=1}^m[u_j,v_j]\Big)=\sum_{\boldsymbol{\delta}\in\{0,1\}^m}(-1)^{\delta_1+\cdots+\delta_m}\cdot g\big((1-\delta_1)v_1+\delta_1u_1,\dots,(1-\delta_m)v_m+\delta_mu_m\big).$$

定义 3（轴对齐划分）。设 $(a_1,\dots,a_m)$ 和 $(b_1,\dots,b_m)$ 是 $\mathbb{R}^m$ 中的向量，其中对于所有 j 均有 $a_j<b_j$。若集合 $\mathcal{P}$ 是 $\prod_{j=1}^m[a_j,b_j]$ 的子集，并且由如下形式的矩形构成，则称其为轴对齐划分：

$$\prod_{j=1}^m\big[u_{l_j}^{(j)},u_{l_j+1}^{(j)}\big]\quad for\ l_j\in[n_j]\ and\ j\in[m],$$

其中，对于每个 $j\in[m]$，$a_j=u_1^{(j)}<u_2^{(j)}<\cdots<u_{n_j+1}^{(j)}=b_j$ 都是 $[a_j,b_j]$ 的一个分割。

定义 4 (Vitali变差)。*(a)* g 在$\prod_{j=1}^m[a_j,b_j]$ 上的 ***Vitali***变差定义为

$$Vit\Big(g;\prod_{j=1}^m[a_j,b_j]\Big)=\sup_{\mathcal{P}}\sum_{R\in\mathcal{P}}|\Delta(g;R)|,$$

其中，上确界取遍$\prod_{j=1}^m[a_j,b_j]$的所有轴对齐划分 $\mathcal{P}$。

(b) g 在整个空间 $\mathbb{R}^m$ 上的 ***Vitali***变差定义为

$$Vit(g)=\sup_{\prod_{j=1}^m[a_j,b_j]\subseteq\mathbb{R}^m}Vit\Big(g;\prod_{j=1}^m[a_j,b_j]\Big).$$

如果 g 足够光滑，那么 g 在 $\mathbb{R}^m$ 上的 Vitali变差可表示为以下形式（例如参见 Owen [24,第 9]节）：

$$\mathrm{Vit}(g)=\int_{\mathbb{R}^m}\Big|\frac{\partial^m g(\mathbf{x})}{\partial x_1\cdots\partial x_m}\Big|\,d\mathbf{x}.\tag{8}$$

我们现在可以定义 $\mathbb{R}^d$上函数的 HK 变差。定义 HK 变差时需要指定一个锚点。当定义域是有界的轴对齐矩形时，锚点选为其中一个顶点。例如，当定义域为 $[0,1]^d$时，锚点通常选为左下角 $\mathbf{0}=(0,\dots,0)$。然而，在我们的情形中，定义域是整个空间 $\mathbb{R}^d$，锚点需要置于无穷远处（$-\infty$ 或 $+\infty$）。这要求函数在无穷远处具有适当的良好性质，具体含义如下。

令$\mathbf{a}=(a_1,\dots,a_d)\in\{-\infty,+\infty\}^d$表示锚点。对于每个坐标，有两种可能的选择：$-\infty$ 或 $+\infty$。对于函数 f：$\mathbb{R}^d\to\mathbb{R}$、子集 $S\subseteq[d]$（其中 $S^c:=[d]\backslash S$），以及 $(x_j,j\in S)\in\mathbb{R}^{|S|}$，定义

$$f^S_{(a_j,j\in S^c)}(x_j,j\in S)=\lim_{(x_j,j\in S^c)\to(a_j,j\in S^c)}f(x_1,\dots,x_d).\tag{9}$$

对于每个 $S \subseteq [d]$，若上述极限对所有 $(x_j, j \in S) \in \mathbb{R}^{|S|}$都存在且有限，则称函数$f^S_{(a_j, j\in S^c)}$是良定义的。该函数可视为 f在定义域某一截面上的限制，该截面通过将 S^c 中的坐标固定为锚点值 a_j而得到。可以验证，只要 $f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$，则对所有 $S \subseteq [d]$ ，$f^S_{(a_j, j\in S^c)}$ 都是良定义的。

定义 5（HK 变差）。固定一个 $\in \{-\infty, +\infty\}^d$。令 $f\colon \mathbb{R}^d \to \mathbb{R}$ 为一个函数，使得对所有 $S \subseteq [d]$，$f^S_{(a_j,\ j\in S^c)}$ 均有定义。f 以$\mathbf{a}$ 为锚点的 *HK* 变差定义为

$$HK_{\mathbf{a}}(f) = \sum_{\varnothing \neq S \subseteq [d]} Vit(f^S_{(a_j, j\in S^c)}).$$

换言之，f 的 HK 变差是将 f 限制在定义域的各个截面上所得的 Vitali变差之和，其中部分坐标固定在 a_j。这解释了$\mathbf{a}$为何称为“锚点”。对于足够光滑的函数 f，由 (8) 可知，$\mathrm{HK}_{\mathbf{a}}(f)$ 也可表示为

$$\mathrm{HK}_{\mathbf{a}}(f) = \sum_{\varnothing \neq S \subseteq [d]} \int_{\mathbb{R}^{|S|}} \left| \frac{\partial^{|S|}}{\prod_{j\in S} \partial x_j} f^S_{(a_j, j\in S^c)}(x_j, j \in S) \right| d(x_j, j \in S). \tag{10}$$

2.2 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 与 HK 变差之间的联系

以下结果（证明见附录 A.1.7）表明，$\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 恰好由所有 HK 变差有限且满足交互限制条件的右连续函数构成：对于每个满足 $|S| > s$ 的子集 $S \subseteq [d]$,

$$\sum_{\boldsymbol{\delta} \in \{0,1\}^{|S|}} (-1)^{\sum_{j\in S} \delta_j} \cdot f^S_{(a_j, j\in S^c)}\big((1-\delta_j)w_j + \delta_j v_j, j \in S\big) = 0 \quad \text{for all } v_j < w_j, j \in S. \tag{11}$$

该条件排除了阶数大于 s 的变量间交互。对于任意锚点$\mathbf{a}$，此结果均成立，因为相对于不同锚点，HK 变差有限这一性质是等价的。

命题 3. 以下陈述等价：

(a) $f \in \mathcal{F}^{d,s}_{\infty-ST}$。*(b)* 对某个 $\mathbf{a} \in \{-\infty, +\infty\}^d$，$HK_{\mathbf{a}}(f) < \infty$，且 f右连续，并对所有满足 $|S| > s$ 的子集$S \subseteq [d]$ 满足(11)。*(c)* 对所有 $\mathbf{a} \in \{-\infty, +\infty\}^d$，$HK_{\mathbf{a}}(f) < \infty$，且 f右连续，并对所有满足 $|S| > s$ 的子集 $S \subseteq [d]$满足(11)。

注 1 $(d = 1)$。当 $d = 1$ 时，命题 *3* 可简化如下。对每个 $s \geqslant 1$,

$$\mathcal{F}^{1,s}_{\infty-ST} = \big\{f : TV(f) < \infty \text{ and } f \text{ is right-continuous}\big\}.$$

这里，$TV(f)$ 表示 f 在 $\mathbb{R}$ 上的通常全变差，定义为 $TV(f) = \sup_{a<b} TV(f;\ [a, b])$，其中

$$TV(f; [a, b]) = \sup \sum_{k=1}^{m} |f(z_{k+1}) - f(z_k)|,$$

其中 t 上确界取遍所有 $m \geqslant 1$ 以及 $[a$ 的所有分割 $a = z_1 < \cdots < z_{m+1} = b$， $b]$.

命题 3 确认，$\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 包含许多连续函数，与子类 $\mathcal{F}^{d,s}_{\mathrm{ST}}$形成对比，后者仅由分段常数函数构成。例如，任何混合偏导数的最高阶为一且足够光滑、其 L^1 范数有限的函数（回顾式 (10)）都属于 $\mathcal{F}^{d,d}_{\infty-\mathrm{ST}}$。

3 复杂度度量 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$

下面给出复杂度度量 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 在 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$上的定义。

定义 6 ($V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$)。 对于 $f\in\mathcal{F}^{d,s}_{\infty-ST}$，定义

$$V^{d,s}_{\infty-XGB}(f):=\inf\left\{\sum_{0<|L|+|U|\leqslant s}\|\nu_{L,U}\|_{TV}:f^{d,s}_{c,\{\nu_{L,U}\}}\equiv f\right\},\tag{12}$$

其中，下确界取遍 f 的所有表示形式 $f^{d,s}_{c,\{\nu_{L,U}\}}$。这里， $\|\nu\|_{TV}:=|\nu|(\mathbb{R}^{|L|+|U|})$ 表示符号测度 ν 的全变差。

基本性质这一复杂度度量的基本性质（证明见附录 A.1.2）总结如下 w.

命题 4。 *(a)* 对于 $s_1\leqslant s_2$, $V^{d,s_1}_{\infty-XGB}(f)\geqslant V^{d,s_2}_{\infty-XGB}(f)$，对所有 $f\in\mathcal{F}^{d,s_1}_{\infty-ST}$。

(b) 对于每个 $s\geqslant 2d$, $V^{d,s}_{\infty-XGB}(\cdot)\equiv V^{d,2d}_{\infty-XGB}(\cdot)$。

(c) $V^{d,s}_{\infty-XGB}(\cdot)$ 在 $\mathcal{F}^{d,s}_{\infty-ST}$ 上是凸的；也就是说，对于所有 $f,g\in\mathcal{F}^{d,s}_{\infty-ST}$ 以及 $\lambda\in[0,1]$,

$$V^{d,s}_{\infty-XGB}((1-\lambda)f+\lambda g)\leqslant(1-\lambda)\cdot V^{d,s}_{\infty-XGB}(f)+\lambda\cdot V^{d,s}_{\infty-XGB}(g).$$

下一个结果（证明见附录 A.1.3）表明，只要 $f\in\mathcal{F}^{d,s}_{\mathrm{ST}}$， $V^{d,s}_{\infty-\mathrm{XGB}}(f)$ 就与 XGBoost 惩罚项 $V^{d,s}_{\mathrm{XGB}}(f)$（定义见 (1)）一致。

定理 1. 对于每个 $f\in\mathcal{F}^{d,s}_{ST}$，我们有 $V^{d,s}_{\infty-XGB}(f)=V^{d,s}_{XGB}(f)$。

3.1 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 与 HK 变差之间的联系

当 $d=1$ 时，我们有以下关于 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 的显式公式——证明见附录 A.1.4——其用全变差（TV）表示（回顾一下，当 $d=1$ 时，HK 变差与全变差一致）。对于 $f\in\mathcal{F}^{1,s}_{\infty-\mathrm{ST}}$，我们有

$$V^{1,s}_{\infty-\mathrm{XGB}}(f)=\begin{cases}\mathrm{TV}(f) & \text{if } s=1,\\ \frac{1}{2}\left(\mathrm{TV}(f)+|\Delta(f)|\right) & \text{if } s=2,\end{cases}\tag{13}$$

其中 $\Delta(f):=\lim_{x\to+\infty}f(x)-\lim_{x\to-\infty}f(x)$。由于 $|\Delta(f)|\leqslant\mathrm{TV}(f)$，因此 t

$$\frac{1}{2}\mathrm{TV}(f)\leqslant V^{1,2}_{\infty-\mathrm{XGB}}(f)\leqslant\mathrm{TV}(f).\tag{14}$$

当 $d\geqslant 2$ 时，似乎无法给出关于 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 的 HK 变差直接公式，但类似于（14）的不等式仍然成立，如下一条结果所示（证明见附录 A.1.5）。

命题 5。 对于每个 $f\in\mathcal{F}^{d,s}_{\infty-ST}$，我们有

$$\frac{1}{\min(2^s-1,2^d)}\cdot\Big(\sup_{\mathbf{a}\in\{-\infty,+\infty\}^d}HK_{\mathbf{a}}(f)\Big)\leqslant V^{d,s}_{\infty-XGB}(f)\leqslant\inf_{\mathbf{a}\in\{-\infty,+\infty\}^d}HK_{\mathbf{a}}(f).\tag{15}$$

这个不等式的两边都是紧的，也就是说，存在非恒定函数 $\mathcal{F}^{d,s}_{\infty-ST}$，使得左侧和右侧的不等式分别取等号。

$V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 与 HK 变差的一个重要区别在于，HK 变差本质上是不对称的，而 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 是对称的。例如，当 $d = s = 2$ 且$\mathbf{a} = (-\infty, -\infty)$时，对于任意$t_1, t_2 \in \mathbb{R}$，我们有

$$\mathrm{HK}_{\mathbf{a}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 \geqslant t_1, x_2 \geqslant t_2)\big) = 1,$$

而

$$\mathrm{HK}_{\mathbf{a}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 < t_1, x_2 \geqslant t_2)\big) = 2 \quad \text{and} \quad \mathrm{HK}_{\mathbf{a}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 < t_1, x_2 < t_2)\big) = 3.$$

对于其他的$\mathbf{a}$取值，也会出现类似的不对称现象。相比之下，对于 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$，我们有

$$\begin{aligned} &V^{d,s}_{\infty-\mathrm{XGB}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 \geqslant t_1, x_2 \geqslant t_2)\big) = V^{d,s}_{\infty-\mathrm{XGB}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 < t_1, x_2 \geqslant t_2)\big) \\ &\quad = V^{d,s}_{\infty-\mathrm{XGB}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 \geqslant t_1, x_2 < t_2)\big) = V^{d,s}_{\infty-\mathrm{XGB}}\big((x_1, x_2) \mapsto \mathbf{1}(x_1 < t_1, x_2 < t_2)\big) = 1. \end{aligned}$$

HK 变差的不对称性源于锚点的存在。HK 变差将函数锚定在定义域的一个角上，由此导致不对称。因此，以 HK 变差作为正则化惩罚项得到的估计结果，可能会随着锚点移至另一个角，或等价地，随着定义域的某些坐标轴翻转而改变。相比之下，$V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 不受坐标轴翻转影响，正如下一个命题所形式化阐述的那样（证明见附录 A.1.8）；这表明$V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 是比 HK 变差 $\mathrm{HK}_{\mathbf{a}}(\cdot)$更自然的正则化项。

命题 6. 设 $f \in \mathcal{F}^{d,s}_{\infty-ST}$。固定 $j \in [d]$ 和 $t_j \in \mathbb{R}$，并定义 g： $\mathbb{R}^d \to \mathbb{R}$ 为

$$g(x_1, \ldots, x_d) = f(x_1, \ldots, x_{j-1}, t_j - x_j, x_{j+1}, \ldots, x_d) \quad \textit{for } (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

那么，$V^{d,s}_{\infty-XGB}(g) = V^{d,s}_{\infty-XGB}(f)$。

第 6.1 节进一步阐述了 HK 变差的不对称性，以及它与 $V^{d,s}_{\infty-\mathrm{XGB}}(\cdot)$ 的对称性之间的对比。

4 XGBoost 与(3)之间的优化等价性

本节分析优化问题 (2) 和 (3)。我们的首个结果证明这两个问题都存在解，并表明 (2) 的任意解同时也是 (3) 的解。因此，可以将 XGBoost 视为在更广泛的函数类 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$上进行隐式优化，该函数类既包含光滑函数，也包含分段常数函数。

定理 2。 存在一个(3) 的解，它同时也是(2)的解。此外，(2) 的每一个解也都是(3)的解。

在标准 XGBoost 实现中，回归树的分裂阈值通常限定为观测到的协变量值之间的中点。更准确地说，对于每个坐标 j，分裂阈值均从 j^{th} 协变量连续观测值之间的中点中选取。令 $v_1^{(j)} < \cdots < v_{n_j}^{(j)}$表示 j^{th} 协变量 x_j 中观测到的不同取值，并按递增顺序排列。注意，

$$\{v_1^{(j)}, \ldots, v_{n_j}^{(j)}\} = \{x_j^{(1)}, \ldots, x_j^{(n)}\}$$

其中，$x_j^{(i)}$ 表示数据点$\mathbf{x}^{(i)}$的 j^{th} 坐标。令 $\mathcal{F}^{d,s}_{\mathrm{ST}\bullet}$ 表示由深度至多为 s的（右连续）树的有限和构成的，s ST子类，其中每棵树都将分裂阈值限制在

坐标上的以下集合中：

$$\Big\{(v_1^{(j)}+v_2^{(j)})/2,\ldots,(v_{n_j-1}^{(j)}+v_{n_j}^{(j)})/2\Big\}.$$

因此，XGBoost 算法也可以看作是一种用于贪心求解的过程 ：

$$\operatorname*{argmin}_f\Big\{\sum_{i=1}^{n}\big(y_i-f(\mathbf{x}^{(i)})\big)^2+\alpha V_{\mathrm{XGB}}^{d,s}(f):f\in\mathcal{F}_{\mathrm{ST},\bullet}^{d,s}\Big\}. \tag{16}$$

以下结果（证明见附录 A.2.1）表明，问题 (3) 也与 (16) 密切相关。

定理 3. 存在一个 (3)的解，同时也是 (16)的解。此外，每个解也都是 (3)的解。 (16)

上述两个定理是以下引理（证明见附录 A.2.2）的直接推论。该引理断言，对于每个 $f\in\mathcal{F}^{d,\ \infty-\mathrm{ST}}$，都存在一个属于 $\mathcal{F}^{d,\ s\ \mathrm{ST}\bullet}$ 的函数，它在每个数据点 $f\ \mathbf{x}^{(i)}$ 处都与之相同，且复杂度不更高。

引理 1. 对于每个 $f\in\mathcal{F}^{d,\ s}_{\infty-ST}$，都存在 $f^{d,\ c\ ,\ \{\nu_{L,\ U}\}}\in\mathcal{F}^{d\ ,\ ST,\ \bullet}$，使得

(a) $\nu_{L,\ U}$ 是支撑在格点上的离散有符号博雷尔测度

$$\prod_{j\in L}\Big\{(v_1^{(j)}+v_2^{(j)})/2,\ldots,(v_{n_j-1}^{(j)}+v_{n_j}^{(j)})/2\Big\}\times\prod_{j\in U}\Big\{(v_1^{(j)}+v_2^{(j)})/2,\ldots,(v_{n_j-1}^{(j)}+v_{n_j}^{(j)})/2\Big\} \tag{17}$$

(b) $f^{d,\ c\ ,\ \{\nu_{L,\ U}\}}(\mathbf{x}^{(i)})=f(\mathbf{x}^{(i)})$ 对于 $i=1,\ \ldots\ldots,\ n$

(c)

$$V_{\infty-XGB}^{d,s}(f_{c,\{\nu_{L,U}\}}^{d,s})=\sum_{0<|L|+|U|\leqslant s}\|\nu_{L,U}\|_{TV}\leqslant V_{\infty-XGB}^{d,s}(f).$$

即使将中点 $(v_{m_j}^{(j)}+v_{m_j+1}^{(j)})/2$ 替换为区间 $(v_{m_j}^{(j)},\ v_{m_j+1}^{(j)})$内的任意其他点，引理 1 仍然成立。我们采用中点，是因为这种选择符合标准的 XGBoost 实现。默认情况下，当数据集较小时，XGBoost 使用中点；而数据集较大时，由于计算限制，它会改用基于分位数的划分（见 [37]）。

条件 (c) 第一部分中的等式值得特别关注。由于 $V^{d,\ \infty-\ \mathrm{XGB}}(\cdot)$ 被定义为对所有可容许积分表示 (7) 取下确界，因此一般来说，$V^{d,\ \infty-\mathrm{XGB}}(f^{d,\ c\ ,\ \{\nu_{L,\ U}\}})$与有符号测度 $\nu_{L,\ U}$的全变差之和之间通常只有不等式成立。然而，对于引理 1 中构造的函数，等式可以取到。这样便无需取下确界，并且惩罚项可以明确表示为相应有符号测度的全变差之和。

5 极小极大风险

在本节中，我们研究函数类 (4) 上的极小极大收敛速率。全文假设$(\mathbf{x}^{(1)},y_1),\ldots,(\mathbf{x}^{(n)},y_n)$ 按照以下模型生成

$$y_i=f^*(\mathbf{x}^{(i)})+\xi_i$$

其中 f^* 是真实回归函数。我们采用随机设计设定，假设协变量 $\mathbf{x}^{(i)}$独立同分布，其密度为 p_0 ，支撑位于一个紧矩形上，且有上界：

$$p_0(\mathbf{x}) = 0 \ \text{ when } \mathbf{x} \notin \prod_{j=1}^{d} \Big[-\frac{M_j}{2}, \frac{M_j}{2} \Big] \quad \text{and} \quad B := M_1 \cdots M_d \cdot \sup_{\mathbf{x}} p_0(\mathbf{x}) < \infty. \tag{18}$$

注意，当 p_0 是 $\prod_{j=1}^{d}[-M_j/2, M_j/2]$, 上的均匀分布的密度时，我们有 $B = 1$。我们进一步假设误差项 ξ_i 是独立同分布的、均值为零的，并且与协变量 $\mathbf{x}^{(i)}$相互独立。

类 (4) 上的极小极大风险定义为

$$\mathfrak{M}_{n,V}^{d,s} := \inf_{\hat{f}_n} \sup_{\substack{f^* \in \mathcal{F}_{\infty-\mathrm{ST}}^{d,s} \\ V_{\infty-\mathrm{XGB}}^{d,s}(f^*) \leqslant V}} \mathbb{E} \| \hat{f}_n - f^* \|_{p_0,2}^2, \tag{19}$$

其中，下确界是在所有基于数据 $(\mathbf{x}^{(1)}, y_1), \ldots, (\mathbf{x}^{(n)}, y_n)$的估计量 $\hat{} f_n$ 上取的。这里， $\| \hat{} f_n - f^* \|_{p_0,2}$表示 $\hat{} f_n$与 f^*之间的 $L^2(p_0)$ 损失：

$$\| \hat{f}_n - f^* \|_{p_0,2}^2 := \int_{\mathbb{R}^d} (\hat{f}_n - f^*)^2(\mathbf{x}) \cdot p_0(\mathbf{x}) \, d\mathbf{x}.$$

本节的第一个主要结果给出了极小极大风险 (19) 的一个上界。我们通过分析类 (4) 上的一个特定最小二乘估计量，得到这个上界。具体而言，我们考虑满足以下附加限制的 (4) 上的最小二乘估计量：与之相关的有符号测度$\nu_{L,U}$ 满足引理 1 第4节中的条件 (a) 以及条件 (c) 中的等式：

$$\begin{aligned} \hat{f}_{n,V}^{d,s} \in \operatorname*{argmin}_{f} \Bigg\{ \sum_{i=1}^{n} \big(y_i - f(\mathbf{x}^{(i)}) \big)^2 : f \equiv f_{c,\{\nu_{L,U}\}} \in \mathcal{F}_{\infty-\mathrm{ST}}^{d,s}, \\ \sum_{0<|L|+|U| \leqslant s} \| \nu_{L,U} \|_{\mathrm{TV}} \leqslant V, \text{ and } \nu_{L,U} \text{ satisfy condition (a) of Lemma 1} \Bigg\}. \end{aligned} \tag{20}$$

引理 1 保证，$\hat{} f^{d,s}_{n\,V}$ 也在原始类别 (4) 上最小化最小二乘准则。换言之，它是类别 (4) 上的最小二乘估计量：

$$\hat{f}_{n,V}^{d,s} \in \operatorname*{argmin}_{f} \Bigg\{ \sum_{i=1}^{n} \big(y_i - f(\mathbf{x}^{(i)}) \big)^2 : f \in \mathcal{F}_{\infty-\mathrm{ST}}^{d,s} \text{ and } V_{\infty-\mathrm{XGB}}^{d,s}(f) \leqslant V \Bigg\}.$$

还可以进一步验证，对于每个 V，都存在一个 α，它可能取决于 V 和数据，使得 $\hat{} f^{d,s}_{n\,V}$ 也是原始惩罚形式 *(3)* 的一个解。更确切地说，如果将 选为 *(20)* 的拉格朗日对偶问题的解，那么 $\hat{},^{s}_{n\,V}$ 也是 (20) 的惩罚版本的一个解，因此也是 (3) 的一个解（回顾引理 1）。

下面的定理（证明见附录 A.3.1）给出了 $\hat{} f^{d,s}_{n\,V}$ 的风险上界。对于这一结果，我们对误差项 施加一个额外假设。具体而言，我们假设它们具有有限的,1 范数：

$$\| \xi_i \|_{3,1} := \int_0^{\infty} \mathbb{P}(|\xi_i| > t)^{1/3} \, dt < \infty. \tag{21}$$

该范数条件较为宽松：它比要求 L^3 范数有限更强，但比要求任意 $\epsilon > 0$ 的有限$L^{3+\epsilon}$ 范数更弱（例如参见 Grafakos [17, 第 1 章4]）。

定理 4。 固定一个真实回归函数 $f^*:\ \mathbb{R}^d \to \mathbb{R}$，它不一定属于 $\mathcal{F}^{d,\ s}_{\infty-ST}$。假设密度 p_0 满足 （18），且误差项 ξ_i 满足 （21）。那么，对于每个 $f_0 \in \mathcal{F}^{d,\ s}_{\infty-ST}$，若$V^{d,\ s}_{\infty-XGB}(f_0) < V$，我们有

$$\mathbb{E}\big[\|\hat{f}^{d,s}_{n,V} - f^*\|^2_{p_0,2}\big] \leqslant C\|f_0 - f^*\|^2_{p_0,2} + O\big(d^{4\bar{s}}(1+\log d)^{4(\bar{s}-1)}(V+1)^2 \cdot n^{-2/3}(\log n)^{4(\bar{s}-1)/3}\big), \tag{22}$$

其中， $C > 0$ 是一个普适常数， $\bar{s}:= \min(s、\ d)$，而 $O(\cdot)$ 所对应的常数因子取决于B、 s、 ξ_i的矩，以及

$$\sup_{\mathbf{x}\in\prod_{j=1}^d [-M_j/2, M_j/2]} |f_0(\mathbf{x}) - f^*(\mathbf{x})|.$$

定理 4是在模型设定错误情形下表述的，允许真实函数 f^*为任意函数。如果 $f^* \in \mathcal{F}^{d,\ s}_{\infty-\mathrm{ST}}$且 $V^{d,\ s}_{\infty-\mathrm{XGB}}(f^*) < V$，那么我们可以在（22）中取 $f_0 = f^*$，从而推出

$$\mathbb{E}\big[\|\hat{f}^{d,s}_{n,V} - f^*\|^2_{p_0,2}\big] \leqslant O\big(d^{4\bar{s}}(1+\log d)^{4(\bar{s}-1)}(V+1)^2 \cdot n^{-2/3}(\log n)^{4(\bar{s}-1)/3}\big),$$

其中， $O(\cdot)$ 所对应的常数因子取决于 B， s以及 ξ_i的矩。这表明，当$f^* \in \mathcal{F}^{d,\ s}_{\infty-\mathrm{ST}}$时，在类（4）上，最小二乘估计量$\hat{f}^{d,\ s}_{n,\ V,}$ 以 $n^{-2/3}$,的速率收敛到 f^*，差一个乘法对数因子。该上界通过因子 $(V+1)^2$取决于复杂度上界 V以及 $V^{d,\ s}_{\infty-\mathrm{XGB}}(f^*)$，这表明，随着目标函数复杂度的增加，$\hat{f}^{d,\ s}_{n,\ V,}$ 的准确性会降低。还值得注意的是，上界对 d 的依赖是多项式的。

备注 2。鉴于第 *3.1* 节讨论的 $V^{d、\ s}_{\infty-XGB}(\cdot)$ 与 *HK* 变差之间的关系，自然会将定理 *4* 与现有关于 *HK* 变差去噪的结果进行比较，例如 *Fang* 等人*[12]*的研究。具体而言， *[12]* 中的定理 *4.5* 表明，在 *HK* 变差约束下的最小二乘估计量也能达到 $n^{-2/3}$ 收敛速率（仅相差一个略有不同的乘性对数因子）。

考虑到 *HK* 变差与 $V^{d、\ s}_{\infty-XGB}(\cdot)$之间的紧密联系，尤其是命题 *5*，这种相似性并不令人意外。然而，两者存在重要差异。 *[12]* 中的定理 *4.5* 是在固定设计情形下建立的，其中设计点$\mathbf{x}^{(i)}$ 构成一个格点；而我们的结果假设设计是随机的，这在许多应用中更为相关。此外， *[12]* 的分析仅限于 $s = d$ 的情形（在他们的框架中，这意味着允许协变量之间具有所有交互阶数），而我们的结果适用于所有 $1 \leqslant s \leqslant d$。此外， *[12]* 中的界没有明确说明对 d的依赖关系。

以下关于括区熵的上界（在附录 A.3.2 中证明）是定理 4 证明的关键要素。令 $\mathcal{F}_{\mathbf{M}}(V)$表示所有满足以下条件的函数 $f^{d,c\ \ s}_{,\{\nu_{L,U}\}} \in \mathcal{F}^{d\ \ s}_{,\infty-\mathrm{ST}}$组成的类，其形式如 (7):

(a) $\nu_{L,U}$的支撑包含于 $\prod_{j\in L}(-M_j/2, M_j/2] \times \prod_{j\in U}(-M_j/2, M_j/2]$(
b) $\sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} \leqslant V$。

类 $\mathcal{F}_{\mathbf{M}}(V)$ 不是全有界的，因为它包含所有常值函数。因此，我们将关注范围限制在子类

$$B(V,t) = \{f \in \mathcal{F}_{\mathbf{M}}(V) : \|f\|_{p_0,2} \leqslant t\}.$$

引理 2. 存在常数 $C_s > 0$，仅依赖于 s；以及 $C_{B,s} > 0$，仅依赖于 B 和 s，使得对于任意 $\epsilon, t, V > 0$,

对数 $N_{[\ ]}(\epsilon, B(V,t), \|\cdot\|_{p_0,\ 2}) \leqslant$对数 $\left(2 + \dfrac{C_s(V+t)}{\epsilon}\right) + C_{B,\ s} d^{2\bar{s}}(1+$对数 $d)^{2(\bar{s}-1)}\left(2+\dfrac{V}{\epsilon}\right)\Big[$对数 $\left(2+\dfrac{V}{\epsilon}\right)\Big]^{2(\bar{s}-1)}$.这里， $N_{[\ ]}(\epsilon, \mathcal{F}, \|\cdot\|_{p_0,\ 2})$ 表示类 $\mathcal{F}$相对于范数$\|\cdot\|_{p_0,\ 2}$ 的 ϵ-括区数。

这一结果建立在 Gao [15] 关于多元累积分布函数的括号熵界之上，这些函数对应于支撑在固定紧矩形上的概率测度。类 $\mathcal{F}_{\mathbf{M}}(V)$ 与多元累积分布函数类之间的联系，源于以下观察：式（7）中的每一项 $\int b^{L,\ U\mathbf{l},\ \mathbf{u}} d\nu_{L,\ U}(\mathbf{l},\ \mathbf{u})$ 都类似于累积分布函数，因为基函数 $b^{L,\ U\mathbf{l},\ \mathbf{u}}$ 由指示函数构造而成。

Blei 等人较早的工作针对相同的累积分布函数类建立了度量熵界（而非括号熵界），并给出了更尖锐的对数因子。然而，对于定理 4 的证明而言，括区熵是必不可少的，因此 Blei 等人的结果 [4] 无法直接应用。

定理 4 立即推出以下推论（证明见附录 A.3.3）。

推论 1。 极小极大风险 $\mathfrak{M}^{d,\ n^{\ s},\ V}$ 满足

$$\mathfrak{M}_{n,V}^{d,s} \leqslant O\big(d^{4\bar{s}}(1+\log d)^{4(\bar{s}-1)}(V+1)^2 \cdot n^{-2/3}(\log n)^{4(\bar{s}-1)/3}\big),$$

其中， $O(\cdot)$ 背后的常数因子取决于 B、 s以及 ξ_i的矩。

现在我们转向本节的第二个主要结果（证明见附录 A.3.4），该结果建立了极小极大风险的下界。对于这一关于下界的结果，除（18）之外，我们进一步假设密度 p_0 在其支撑上与零有正距离，即

$$b := M_1 \cdots M_d \cdot \inf_{\mathbf{x} \in \prod_{j=1}^d [-M_j/2, M_j/2]} p_0(\mathbf{x}) > 0. \tag{23}$$

注意，当 p_0 是$\prod_{j=1}^d [-M_j/2, M_j/2]$上的均匀密度时， $b = 1$ 。对于误差项 ξ_i，我们不采用（21），而是假设它们服从高斯分布：

$$\xi_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2). \tag{24}$$

定理 5. 假设密度 p_0 满足 (18) 和 (23)，且误差项 ξ_i 满足 (24)。那么，存在常数 $C_{b,B,\bar{s}} > 0$，它们仅取决于 b、 B 和 $\bar{s} = \min(s$、 $d)$；还存在 $C_{B,\bar{s}} > 0$，它们仅取决于 B 和 $\bar{s}$，使得

$$\mathfrak{M}_{n,V}^{d,s} \geqslant C_{b,B,\bar{s}} \Big(\frac{\sigma^2 V}{n}\Big)^{2/3} \left[\log\Big(\frac{nV^2}{\sigma^2}\Big)\right]^{2(\bar{s}-1)/3},$$

前提是 $n \geqslant C_{B,\bar{s}}(\sigma^2/V^2)$。

结合推论 1 和定理 5，我们得出，在类 (4) 上的极小极大收敛速率为 $n^{-2/3},$，至多相差乘法对数因子，其指数介于 $2(\bar{s}-1)/3$ 和 $4(\bar{s}-1)/3$ 之间。这种近乎与维度无关的速率表明，即使在高维情形下，类 (4) 也得到了充分正则化。换言之，复杂度度量 $V^{d,\infty-\ s\ \mathrm{XGB}}(\cdot)$（因此也包括 $V^{d,s\ \mathrm{XGB}}(\cdot)$）随着维度 d 增大，能够有效地发挥正则化作用，充分控制高维场景下的模型复杂度。

这一观察为 XGBoost 强大的实证表现提供了一种可能的解释，同时也补充了这样一个事实：XGBoost 优化问题 (2) 的每个解，也都是函数类 $\mathcal{F}^{d,\infty-\ s\ \mathrm{ST}}$ 上惩罚最小二乘问题 (3) 的解（定理 2）；该函数类包含许多分段常数函数之外的函数（命题 3）。

6 讨论

6.1 与对称化 HK 变差的联系

如第 3.1 节所述，由于需要指定一个锚点，HK 变差缺乏对称性。可以尝试将所有对应于 $\mathbf{a} = (a_1, \dots, a_d)$ 且具有 $a_j \in \{-\infty, +\infty\}$ 的 2^dversions $\mathrm{HK}_{\mathbf{a}}(\cdot)$ 组合起来，以恢复对称性。在数学图像处理文献中，将多个全变差概念组合成一个量的自然方法是下确界卷积（例如，Bergounioux [3], Bredies 和 Holler [6], Chambolle、Lions [9], Setzer 和 Steidl [26], Setzer 等 [27]）。遵循这一思路，可以考虑对所有 $\mathbf{a} \in \{-\infty, +\infty\}^d$ 对应的 Hardy–Krause 变差 $\mathrm{HK}_{\mathbf{a}}(\cdot)$ 取下确界卷积:

$$\inf\Big\{ \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} \mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) : \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} f_{\mathbf{a}} \equiv f,\ f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\ \forall \mathbf{a} \Big\}. \tag{25}$$

那么，一个自然的问题是，这个同样满足命题 6 中所述对称条件的量，与 $V^{d,\ \overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 有何关系。

可以证明，若修改 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$ 和 $V^{d,\ \overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 的定义，禁止在每棵树的分裂中重复使用同一个变量，那么所得的复杂度度量就与 (25) 一致。为明确说明这一点，考虑函数类 $\widetilde{\mathcal{F}}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$，它由所有形如 (7) 的函数 $f : \mathbb{R}^d \to \mathbb{R}$ 组成，但求和仅限于 $[d]$ 的不相交子集 L 和 U，并满足 $0 < |L| + |U| \leqslant s$。对每个 $f \in \widetilde{\mathcal{F}}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$，将复杂度 $\widetilde{V}^{d,\ \overset{s}{\infty-}\mathrm{XGB}}(f)$ 定义为 f 的复杂度，方式与 (12) 类似，同样将求和限制在满足 $0 < |L| + |U| \leqslant s$ 的不相交子集 L 和 U 上。

可以验证，对于所有 $f \in \mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$ 以及所有 $\mathbf{a} \in \{-\infty, +\infty\}^d$，$\widetilde{\mathcal{F}}^{d,\overset{s}{\infty-}\mathrm{ST}} = \mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$，并且 $V^{d,\overset{s}{\infty-}\mathrm{XGB}}(f) \leqslant \widetilde{V}^{d,\overset{s}{\infty-}\mathrm{XGB}}(f) \leqslant \mathrm{HK}_{\mathbf{a}}(f)$。更重要的是，$\widetilde{V}^{d,\overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 与式（25）中的下确界卷积相一致，如以下命题所示（证明见附录 A.4.1）。

命题 7。对于每个 $f \in \mathcal{F}^{d,\overset{s}{\infty-ST}}$，我们有

$$\widetilde{V}^{d,s}_{\infty-XGB}(f) = \inf\Big\{ \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} HK_{\mathbf{a}}(f_{\mathbf{a}}) : \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} f_{\mathbf{a}} \equiv f,\ f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-ST}\ \forall \mathbf{a} \Big\}.$$

尽管 $\widetilde{V}^{d,\overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 具有对称性，并且可以通过不同锚点上的 HK 变差的下确界卷积得到简洁的刻画，但它并不能充分反映实践中回归树的行为。回归树的一个重要特性是能够在单棵树中多次按同一个变量进行分裂，从而沿某个坐标进行局部细化。在实践中，禁止此类重复分裂可能会降低估计精度。复杂度 $\widetilde{V}^{d,\ \overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 对应于这种受限情形，而 $V^{d,\overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 则允许单棵树内对同一个变量进行重复分裂。因此，尽管这两种概念都满足对称性，$V^{d,\overset{s}{\infty-}\mathrm{XGB}}(\cdot)$ 更贴合回归树固有的结构灵活性，因此是此处更合适的全变差概念。

6.2 超越 $\mathcal{F}^{d,\ \overset{s}{\infty-}\mathrm{ST}}$

我们已经论证，XGBoost 有望有效估计类 $\mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$ 中的函数。特别是，如果 $f^* \in \mathcal{F}^{d,\overset{s}{\infty-}\mathrm{ST}}$，且复杂度度量 $V^{d,\overset{s}{\infty-}\mathrm{XGB}}(f^*)$ 可视为常数，那么理想化 XGBoost 估计量——即 XGBoost 优化问题的一个解——能够达到避免维数灾难的收敛速率 $n^{-2/3}$，且仅相差对数因子。

一个自然的后续问题是，当 f^* 位于 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 之外时会发生什么。此类函数的一个简单例子是

$$f^*(\mathbf{x}) := \mathbf{1}(x_1 + \cdots + x_d \geqslant 0, \mathbf{x} \in [-1,1]^d). \tag{26}$$

可以证明，这个函数不属于 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$。一种证明方法是注意到 f^* 具有无限的 Hardy–Krause 变差；例如，参见 Owen [24, 命题 17]。

对于 f^* 不属于 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 的函数，根据第5节的定理 4 和定理 5，可以自然地猜想，理想化 XGBoost 估计量的风险具有以下形式

$$\inf_{V}\left((V+1)^{\beta}\cdot n^{-2/3}(\log n)^{\gamma} + \inf_{\substack{f_0\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\\ V^{d,s}_{\infty-\mathrm{XGB}}(f_0)\leqslant V}} \|f_0 - f^*\|^2_{p_0,2}\right), \tag{27}$$

其中常数为 $\beta \in [2/3, 2]$ 和 $\gamma \in [2(\bar{s}-1)/3, 4(\bar{s}-1)/3],$，且 $\bar{s} = \min(s,d)$。β 和 γ 的上界由定理 4 中的风险上界得出，而下界则可依据定理 5 中的极小极大下界预期得到。为简洁起见，我们省略对其他参数的依赖，例如 d、s，以及协变量和误差项的分布。

I若我们假设 $\beta = 2/3$ 并忽略对数因子 $(\log n)^{\gamma}$，则（27）化简为

$$\inf_{V}\left((V+1)^{2/3}\cdot n^{-2/3} + \inf_{\substack{f_0\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\\ V^{d,s}_{\infty-\mathrm{XGB}}(f_0)\leqslant V}} \|f_0 - f^*\|^2_{p_0,2}\right).$$

为进一步简化，假设 p_0 是定义在 $[-1,1]^d$上的均匀密度。那么，对于函数（26），可以证明，当 V足够大时，

$$\inf_{\substack{f_0\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\\ V^{d,s}_{\infty-\mathrm{XGB}}(f_0)\leqslant V}} \|f_0 - f^*\|^2_{p_0,2} = \Omega(V^{-1/(d-1)}).$$

因此，即使在这一最有利的情形下（其中 $\beta = 2/3$），该 f^* 的理想化 XGBoost 估计量的收敛速率也不快于

$$\inf_{V}\left((V+1)^{2/3}\cdot n^{-2/3} + V^{-1/(d-1)}\right) \asymp n^{-2/(2d+1)}.$$

不同于当 $f^* \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 且 $V^{d,s}_{\infty-\mathrm{XGB}}(f^*)$有界时所达到的避免维数灾难的收敛速率，上述速率会随着维数 d 的增加而迅速恶化。这表明，尽管 XGBoost 在模型设定错误的情况下仍可能实现一致性，但它并不适合估计远远超出函数类 $\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$的函数。

6.3 L^1 与 L^2 正则化

我们重点讨论了式 (1) 中定义的 XGBoost 的 L^1 惩罚项。如引言中所述，XGBoost 实现通常也会使用平方 L^2 惩罚项，其中 $\|\mathbf{w}_k\|_1$ 被替换为$\|\mathbf{w}_k\|_2^2$。更一般地，对于任意 $p \geqslant 1$ 和 $f \in \mathcal{F}^{d,s}_{\mathrm{ST}}$，我们可以定义

$$V^{d,s}_{\mathrm{XGB}}(f;p) := \inf\Big\{\sum_k \|\mathbf{w}_k\|_p^p\Big\},$$

其中 $\|\cdot\|_p$ 表示通常的 L^p 范数，并且与 (1) 中一样，下确界是在将 f表示为深度至多为 s的右连续回归树的有限和的所有表示中取得的。然而，这个变差泛函

只有当 $p = 1$时，才会产生有意义的正则化惩罚项。具体而言，当 $p > 1$时，该惩罚项会退化，正如下列结果所示（证明见附录 A.4.2）。这种退化解释了为何本文仅关注 L^1 惩罚项。

引理 3. 假设 $p > 1$。那么，对于每个 $f \in \mathcal{F}^d$， $V^{d,}{}_{XGB}^{s}(f;\ p) = 0_{ST}^{s}$。

在实践中，XGBoost 使用函数类 $\mathcal{F}^{d,}{}_{s\,\mathrm{ST}}(K)$，该函数类由形如 $\sum_{k=1}^{K} f_k$ 的函数组成，其中每个 f_k 都是具有右连续分裂且深度为 $\leqslant s$ 的回归树。这里， K 是一个固定的有限数值，通常通过交叉验证来选择。 $\mathcal{F}^{d,}{}_{s\,\mathrm{ST}}$ 与 $\mathcal{F}^{d,}{}_{s\,\mathrm{ST}}(K)$ 的区别在于，前者允许任意数量的树，而后者仅考虑由至多 K 棵树组成的集成。

Wi在这个受限的函数类 $\mathcal{F}^{d,}{}_{s\,\mathrm{ST}}(K)$ 中，我们可以将惩罚项截断为

$$V_{\mathrm{XGB}}^{d,s}(f;p,K) := \inf\Big\{ \sum_{k=1}^{K} \|\mathbf{w}_k\|_p^p \Big\},$$

其中，下确界现在取遍所有将 $f \in \mathcal{F}^{d,}{}_{s\,\mathrm{ST}}(K)$ 表示为至多 K 棵深度至多为 s 的回归树之和的方式。对于所有 $p \geqslant 1$，包括 $p = 2$，这个修正后的惩罚项很可能都有良好定义。然而，由于它对超参数 K 有严格的依赖关系，因此在理论上较难处理。具体而言，当 $K \to \infty$ 时，这种表述无法给出有意义的极限。此外，由于实践中的 K 依赖于数据，将其视为固定数值并不自然。

基于这些原因，我们仅关注 $p = 1$ 的情形，因为这一选择能够提供稳定的正则化惩罚项，并可自然推广到连续树集成，同时避免在树的数量未固定时，含有 $p > 1$ 的范数所固有的惩罚项趋于零的问题。

6.4 XGBoost 所用迭代算法的分析

我们的分析聚焦于 XGBoost 旨在逼近的正则化优化问题 (2) 的解的统计行为。我们不研究实际采用的贪心树提升算法能否在类 $\mathcal{F}^{d,}{}_{\infty-\,\mathrm{ST}}^{s}$ 上实现相同的收敛速率；如何建立此类保证仍是一个重要的开放问题。近期对贪心树构建算法的分析取得了一些进展；例如，参见 Tan 等人 [30]。

尽管存在这一局限，我们的结果仍与 XGBoost 的实践直接相关。通过刻画目标优化问题的行为，我们的理论为 XGBoost 在优化效果良好的情况下能够达到的表现提供了有依据的基准。具体而言，结果阐明了何时能够实现无维数依赖的速率，以及何时会因模型设定错误而出现内在的逼近障碍。这一视角有助于区分统计局限——源于树集成的表达能力及其相关正则化——与贪心提升过程本身的算法局限，从而为理解 XGBoost 在实践中的成功与失效提供了统一框架。

致谢

作者感谢 Ryan Tibshirani、Erez Buchweitz 和 Arnab Mitra 提供的宝贵讨论。

资金支持

作者衷心感谢美国国家科学基金会通过 DMS-2210504 和 DMS-2515470 号资助提供的支撑。

参考资料

[1] Aistleitner, C. 和 J. Dick（2015）。有界变差函数、有符号测度与一般 Koksma–Hlawka不等式。《*Acta Arithmetica*》*167*(2), 143–171.[2] Benkeser, D. 和 M. van der Laan（2016）。高度自适应套索估计量。载于*IEEE*数据科学与高级分析国际会议（*DSAA*），第 689–696 页。[3] Bergounioux, M.（2016）。图像处理下确界卷积模型的数学分析。《优化理论与应用杂志》*168*(1), 1–21.[4] Blei, R., F. Gao 和 W. Li（2007）。高维分布的度量熵。*Proc. Amer. Math.Soc. 135*(12), 4009–4018.[5] Borisov, V., T. Leemann, K. Seßler, J. Haug, M. Pawelczyk 和 G. Kasneci（2022）。深度神经网络与表格数据：综述。《*IEEE*神经网络与学习系统汇刊》*35*(6), 7499– 7519.[6] Bredies, K. 和 M. Holler（2020）。高阶全变差方法及其推广。《反问题》*36*(12), 123001.[7] Breiman, L.（2001a）。随机森林。《机器学习》*45*(1), 5–32.[8] Breiman, L.（2001b）。统计建模：两种文化（附评论及作者回应）。*Statist. Sci. 16*(3), 199–231.[9] Chambolle, A. 和 P.-L. Lions（1997）。通过全变差最小化及相关问题实现图像复原。《数值数学》*76*(2), 167–188.[10] Chen, T. 和 C. Guestrin（2016）。XGBoost：一种可扩展的树提升系统。载于第*22*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集，第 785–794 页。[11]Donoho, D. L. 和 J. M. Johnstone（1994）。通过小波收缩实现理想的空间自适应。《生物统计学》*81*(3), 425–455.[12] Fang, B., A. Guntuboyina 和 B. Sen（2021）。通过全单调性和 Hardy–Krause 变差对保序回归与全变差去噪进行多元扩展。统计年刊 *49*(2), 769–792.[13] Folland, G. B.（1999）。《实分析：现代技术及其应用》（第二版）。约翰·威立父子出版公司。[14]Friedman, J. H.（2001）。贪心函数逼近：梯度提升机。统计年刊 *29*(5), 1189–1232.[15] Gao, F.（2013）。高维分布的括区熵。载于《高维概率*VI*》，第 66 卷《概率论进展》，第 3–17 页。巴塞尔：Birkh¨auser。

[16] Gin´e, E., R. Lata-la 和 J. Zinn（2000）。U统计量的指数不等式与矩不等式。载于《高维概率 II》，《概率论进展》第 47 卷，第 13–38 页。马萨诸塞州波士顿：Birkh¨auser。[17] Grafakos, L.（2014）。《经典傅里叶分析》（第三版）。数学研究生教材。纽约州纽约市：施普林格。[18] Grinsztajn, L., E. Oyallon 和 G. Varoquaux（2022）。为什么基于树的模型在典型表格数据上仍优于深度学习？神经信息处理系统进展 35, 507–520。[19] Han, Q. 和 J. A. Wellner（2019）。具有重尾误差的最小二乘回归估计量的收敛速率。统计年刊 47 (4), 2286–2319。[20] Johnstone, I. M.（2019）。高斯估计：序列与小波模型。手稿，2019年9月。可访问 https://imjohnstone.su.domains/GE_09_16_19.pdf。[21] Ki, D., B. Fang 和 A. Guntuboyina（2024）。通过 LASSO 实现 MARS。统计年刊 52 (3), 1102–1126。[22] Ledoux, M. 和 M. Talagrand（1991）。《巴拿赫空间中的概率：等周不等式与随机过程》，《现代数学综述丛书》第 23 卷。柏林、海德堡：施普林格。[23] Leonov, A. S.（1996）。关于多变量函数的全变差及 Helly选择原理的多维类比。数学札记 63 (1), 61–71。[24] Owen, A. B.（2005）。拟蒙特卡罗的多维变差。载于《当代多元分析与实验设计》，《生物统计学丛书》第 2 卷，第 49–74 页。新泽西州哈肯萨克：世界科学出版社。[25] Schuler, A., Y. Li 和 M. van der Laan（2022）。套索树提升。arXiv 预印本 arXiv:2205.10697。[26] Setzer, S. 和 G. Steidl（2008）。图像处理中采用高阶导数的变分方法。Approximation 12, 360–386。[27] Setzer, S., G. Steidl 和 T. Teuber（2011）。采用离散 ℓ_1 型泛函的下确界卷积正则化。数学科学通讯 9 (3), 797–827。[28] Shwartz-Ziv, R. 和 A. Armon（2022）。表格数据：深度学习并非所需的一切。信息融合 81, 84–90。[29] Talagrand, M.（2021）。《随机过程的上界与下界：分解定理》（第二版），《现代数学综述丛书》第 60 卷。沙姆：施普林格。[30] Tan, Y. S., J. M. Klusowski 和 K. Balasubramanian（2024）。递归自适应划分估计量的统计-计算权衡。arXiv 预印本 arXiv:2411.04394。[31] Tibshirani, R.（1996）。通过套索实现回归收缩与选择。英国皇家统计学会学报 B 辑 58 (1), 267–288。[32] Tibshirani, R. J.（2014）。通过趋势滤波进行自适应分段多项式估计。统计年刊 42 (1), 285–323。[33] van de Geer, S.（2016）。《稀疏性下的估计与检验： ´Ecole d’ ´Et ´e de Probabilit ´es de Saint-Flour XLV-2015》，数学讲义第 2159 卷。沙姆：施普林格。

[34] van der Laan, M. J.、D. Benkeser 和 W. Cai（2023）。采用欠平滑高适应性套索对路径可微目标参数进行高效估计。《国际生物统计学杂志》*19*(1)，261–289。[35] van der Vaart, A. W.（2000）。《渐近统计学》，剑桥统计与概率数学丛书第 3 卷。剑桥：剑桥大学出版社。[36] van der Vaart, A. W. 和 J. A. Wellner（1996）。《弱收敛与经验过程：及其在统计学中的应用》。施普林格统计学丛书。纽约州纽约市：施普林格。[37] XGBoost 开发者（2025）。XGBoost 3.0.5 文档。
https://xgboost.readthedocs.io/en/ stable/parameter.html。访问日期：2025-09-14。

A 证明

A.1 第2节和第3节中命题及定理的证明

A.1.1 命题 1 的证明

命题 *1* 的证明。我们已经看到，$\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$ 中的任意元素都可以表示为形式（7），其中$\nu_{L,\ U}$ 为具有有限支撑的有符号测度。因此，只需证明反向包含关系即可。

注意，每个基函数 $b^{L,\ U\,\mathbf{l},\ \mathbf{u}}$ 都可以看作一个采用右连续分裂、深度至多为 s的回归树，其叶节点权重除一个权重为 1 的叶节点外均为 0。因此，每个基函数 $b^{L,\ U\,\mathbf{l},\ \mathbf{u}}$ 都属于 $\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$。由于 $\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$ 对加法和数乘封闭，可知这些基函数的任意有限线性组合——因此，任何由具有有限支撑的离散有符号测度 $\nu_{L,\ U}$ 构成的形如 (7) 的函数——也属于 $\mathcal{F}^{d,\ s}{}_{\mathrm{ST}}$。证明完毕。

□

A.1.2 命题 4 的证明

命题 *4* 的证明。回顾一下，(7) 中的求和遍历所有满足 $0 < |L| + |U| \leqslant s$的 $L,\ U \subseteq [d]$。因此，对于每个函数 f，可容许表示集合 $f^{d,\ c}{}^{\ s}{}_{,\ \{\nu_{L,\ U}\}} \equiv f$会随着 s增大而扩大。由于$V^{d,\ \infty}{}^{s}_{-\ \mathrm{XGB}}(\cdot)$ 定义为这些表示的下确界，因此可得 (a)。

由于 $|L| + |U|$ 始终不超过 $2d$，将 s 增加到 $2d$ 以上不会扩大可容许表示的集合。因此，$V^{d,\ \infty}{}^{s}_{-\ \mathrm{XGB}}(\cdot)$ 在 $s \geqslant 2d$之后便趋于稳定，这就证明了 (b)。

最后，(c) 源于 有限符号博雷尔测度空间上全变差的凸性 $\|\cdot\|_{\mathrm{TV}}$.

□

A.1.3 定理1的证明

在证明定理之前，我们先观察并证明 $V^{d,\ s}{}_{\mathrm{XGB}}(\cdot)$的另一种刻画；它最初是通过 (1) 定义的。

引理 4。 复杂度度量 $V^{d,\ }{}^{s}_{XGB}(\cdot)$ 还可以刻画为

$$V^{d,s}_{XGB}(f) = \inf\left\{ \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{TV} : f^{d,s}_{c,\{\nu_{L,U}\}} \equiv f \ \textit{and} \right.$$
$$\left. \nu_{L,U} \ \textit{are discrete signed measures with finitely many support points} \right\}.$$

请注意，与 $V^{d,\infty}{}^{s}_{-\ \mathrm{XGB}}(\cdot)$ 的定义 (12) 相比，唯一的区别在于有符号测度 $\nu_{L,U}$ 被限制为具有有限支撑的离散测度。我们将在定理 1 的证明中说明，对于 $\mathcal{F}^{d,s}{}_{\mathrm{ST}}$ 中的函数，这一额外限制不会影响下确界的取值。

引理 *4* 的证明。 首先假设所有 $\nu_{L,U}$ 都是只有有限个支撑点的离散有符号测度。那么，$f^{d,c}{}^{\ s}{}_{,\{\nu_{L,U}\}}$是基函数 $b^{L,U\mathbf{l},\mathbf{u}}$的有限线性组合，其系数为 $\nu_{L,U}(\{(\mathbf{l},\mathbf{u})\})$（再加上一个常数）。回想一下，每个基函数 $b^{L,U\,\mathbf{l},\mathbf{u}}$ 都可以看作一棵具有右连续分裂且深度至多为 s 的回归树，其叶节点权重除一个权重为 1 的叶节点外均为 0。因此，$f^{d,c}{}^{\ s}{}_{,\{\nu_{L,U}\}}$可以表示为同类回归树的有限和，这些树的叶权重向量

各自都只有一个非零元素，且该元素等于 $\nu_{L,U}(\{(\mathbf{l},\mathbf{u})\})$。对于这种表示，叶权重向量的总 ℓ^1 范数恰好等于 $\nu_{L,U}$的全变差之和。这就证明了引理中的下确界大于或等于式（1）中的下确界。

现在，假设 $f \in \mathcal{F}^{d,s}_{\mathrm{ST}}$，并且它可以表示为有限个具有右连续分裂且深度至多为 s 的回归树之和。令 $\mathbf{w}_k$ 表示第 k^{th} 棵树的叶权重向量。将每棵树分解为与从根节点到各叶节点的路径相对应的基函数 $b^{L,U,\mathbf{l},\mathbf{u}}$，即可得到 f 由这些基函数构成的有限线性组合表示，其系数向量的 ℓ^1范数不超过 $\sum_k \|\mathbf{w}_k\|_1$。等价地，存在一个表示 $f^{d,c,s}_{,\{\nu_{L,U}\}} \equiv f$，其中有有限支撑的离散有符号测度 $\nu_{L,U}$ 满足，其全变差之和 $\nu_{L,U}$ 不超过$\sum_k \|\mathbf{w}_k\|_1$。这表明引理中的下确界不大于 (1) 中的下确界。

□

定理 *1* 的证明。 我们先引入证明中使用的一些记号。对于函数 $g\colon [0,1]^d \to \mathbb{R}$ 和非空子集 $S \subseteq [d]$，定义

$$g^S(x_j, j \in S) := \lim_{(x_j, j\in S^c)\to(-\infty, j\in S^c)} g(x_1, \ldots, x_d) \qquad \text{for } (x_j, j \in S) \in \mathbb{R}^{|S|}$$

只要极限存在，其中 $S^c = [d]\backslash S$。

固定 $f \in \mathcal{F}^{d,s}_{\mathrm{ST}}$。由于 f是回归树的有限和，因此对每个 $j \in [d]$，都存在 $\mathbb{R}$ 的一个分割 $-\infty = v_0^{(j)} < v_1^{(j)} < \cdots < v_{n_j}^{(j)} < v_{n_j+1}^{(j)} = +\infty$，使得 f 在

$$\prod_{j\in S}(v_{m_j}^{(j)}, v_{m_j+1}^{(j)}) \times \prod_{j\in S^c}\{v_{m_j}^{(j)}\} \tag{28}$$

对于每个非空的 $S \subseteq [d]$，$m_j \in \{0, \ldots\ldots, n_j\}$（当 $j \in S$ 时），以及 $m_j \in \{1, \ldots\ldots, n_j\}$（当 $j \in S^c$ 时）。

对于每个满足 $|S| \leqslant s$ 且 $\mathbf{m} = (m_j,\ j \in S) \in \prod_{j\in S}[n_j]$ 的非空 $S \subseteq [d]$，定义交替和泛函

$$\Delta^S_{\mathbf{m}}(g) := \lim_{\epsilon\to 0+} \sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}} (-1)^{\sum_{j\in S}\delta_j} \cdot g^S\big(v_{m_j}^{(j)} - \delta_j\epsilon, j \in S\big)$$

对于式 (28) 中的分段常数函数 g。

假设 $f^{d,c,s}_{,\{\nu_{L,U}\}} \equiv f$。那么，显然有

$$\Delta^S_{\mathbf{m}}(f^{d,s}_{c,\{\nu_{L,U}\}}) = \Delta^S_{\mathbf{m}}(f) \tag{29}$$

对所有非空的 $S \subseteq [d]$，满足 $|S| \leqslant s$且 $\mathbf{m}\in \prod_{j\in S}[n_j]$。事实上，条件（29）几乎涵盖了恒等式 $f^{d,c,s}_{,\{\nu_{L,U}\}} \equiv f$ 中的所有信息。下面的引理将对此作出精确说明，其证明将在当前证明之后给出。该引理将在后文发挥重要作用。

引理 5. 若 $g, h \in \mathcal{F}^{d,s}_{,ST}$ 如 (28) 所示为分段常数，并且满足

$$\Delta^S_{\mathbf{m}}(g) = \Delta^S_{\mathbf{m}}(h)$$

对所有非空的 $S \subseteq [d]$，满足 $|S| \leqslant s$ 且 $\mathbf{m} \in \prod_{j\in S}[n_j]$，则 g 与 h 仅相差一个加法常数；也就是说，存在 $b \in \mathbb{R}$，使得对所有 $\mathbf{x} \in \mathbb{R}^d$，$g(\mathbf{x}) = h(\mathbf{x}) + b$。

我们将（29）化简，并更明确地用 $\nu_{L,U}$ 表示。固定一个非空的 $S \subseteq [d]$，满足 $|S| \leqslant s$ 且 $\mathbf{m} = (m_j, j \in S) \in \prod_{j\in S}[n_j]$。展开（29）的左侧可得

$$\Delta^S_{\mathbf{m}}(f^{d,s}_{c,\{\nu_{L,U}\}}) = \sum_{\substack{L,U: L\subseteq S\subseteq L\cup U \\ |L|+|U|\leqslant s}} \lim_{\epsilon\to 0+} \sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}} (-1)^{\sum_{j\in S}\delta_j} \tag{30}$$

$$\cdot \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}\big(v_{m_j}^{(j)} - \delta_j\epsilon \geqslant l_j\big) \cdot \prod_{j\in U\cap S} \mathbf{1}\big(v_{m_j}^{(j)} - \delta_j\epsilon < u_j\big)\, d\nu_{L,U}(\mathbf{l},\mathbf{u}).$$

通过交换求和与积分的顺序，并分析每个指示项，内极限可以化简。具体而言，

$$\begin{aligned}
&\lim_{\epsilon\to 0+} \sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}} (-1)^{\sum_{j\in S}\delta_j} \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}\big(v_{m_j}^{(j)} - \delta_j\epsilon \geqslant l_j\big) \cdot \prod_{j\in U\cap S} \mathbf{1}\big(v_{m_j}^{(j)} - \delta_j\epsilon < u_j\big)\, d\nu_{L,U}(\mathbf{l},\mathbf{u}) \\
&= \lim_{\epsilon\to 0+} \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L\setminus U} \big\{\mathbf{1}\big(v_{m_j}^{(j)} \geqslant l_j\big) - \mathbf{1}\big(v_{m_j}^{(j)} - \epsilon \geqslant l_j\big)\big\} \cdot \prod_{j\in L\cap U} \big\{\mathbf{1}\big(l_j \leqslant v_{m_j}^{(j)} < u_j\big) - \mathbf{1}\big(l_j \leqslant v_{m_j}^{(j)} - \epsilon < u_j\big)\big\} \\
&\qquad \cdot \prod_{j\in (U\setminus L)\cap S} \big\{\mathbf{1}\big(v_{m_j}^{(j)} < u_j\big) - \mathbf{1}\big(v_{m_j}^{(j)} - \epsilon < u_j\big)\big\}\, d\nu_{L,U}(\mathbf{l},\mathbf{u}) \\
&= (-1)^{|(U\setminus L)\cap S|} \sum_{K\subseteq L\cap U} (-1)^{|(L\cap U)\setminus K|} \cdot \nu_{L,U}\Big(\prod_{j\in (L\setminus U)\cup K} \{v_{m_j}^{(j)}\} \times \prod_{j\in (L\cap U)\setminus K} (-\infty, v_{m_j}^{(j)}) \\
&\qquad\qquad \times \prod_{j\in K} (v_{m_j}^{(j)}, +\infty) \times \prod_{\substack{j\in ((L\cap U)\setminus K)\\ \cup ((U\setminus L)\cap S)}} \{v_{m_j}^{(j)}\} \times \mathbb{R}^{|U\setminus S|}\Big).
\end{aligned}$$

因此，条件（29）对于所有非空的 $S\subseteq[d]$（满足 $|S|\leqslant s$ 且 $\mathbf{m}=(m_j, j\in S)\in\prod_{j\in S}[n_j]$，前提是 $f^{d,s}_{c,\{\nu_{L,U}\}}\equiv f$）均成立，可写为

$$\begin{aligned}
&\sum_{\substack{L,U: L\subseteq S\subseteq L\cup U\\ |L|+|U|\leqslant s}} (-1)^{|(U\setminus L)\cap S|} \sum_{K\subseteq L\cap U} (-1)^{|(L\cap U)\setminus K|} \\
&\qquad \cdot \nu_{L,U}\Big(\prod_{j\in (L\setminus U)\cup K} \{v_{m_j}^{(j)}\} \times \prod_{j\in (L\cap U)\setminus K} (-\infty, v_{m_j}^{(j)}) \times \prod_{j\in K} (v_{m_j}^{(j)}, +\infty) \times \prod_{\substack{j\in ((L\cap U)\setminus K)\\ \cup ((U\setminus L)\cap S)}} \{v_{m_j}^{(j)}\} \times \mathbb{R}^{|U\setminus S|}\Big) \\
&\qquad = \Delta^S_{\mathbf{m}}(f). \qquad (31)
\end{aligned}$$

因此，我们有

$$V^{d,s}_{\infty-\mathrm{XGB}}(f) \geqslant \inf\Big\{ \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} : \nu_{L,U} \text{ satisfy (31)} \Big\}. \qquad (32)$$

Now，我们证明（32）中的下确界由支撑在.....上的离散有符号测度取得

$$\prod_{j\in L} \{v_1^{(j)}, \ldots, v_{n_j}^{(j)}\} \times \prod_{j\in U} \{v_1^{(j)}, \ldots, v_{n_j}^{(j)}\}. \qquad (33)$$

假设 $\nu_{L,U}$ 是满足 (31) 的有符号博雷尔测度。对于每个 $j\in[d]$，令 $V^{(j)} = \{v_1^{(j)}, \ldots, v_{n_j}^{(j)}\}$、$V^{(j)}_{m_j} = \{v_{m_j+1}^{(j)}, \ldots, v_{n_j}^{(j)}\}$ 和 $V^{(j)}_{m_j} = \{v_1^{(j)}, \ldots, v_{m_j-1}^{(j)}\}$。定义支撑在格点（33）上的离散有符号测度 $\mu_{L,U}$，如下：

$$\begin{aligned}
\mu_{L,U}\big(\{(v_{p_j}^{(j)}, j\in L; v_{q_j}^{(j)}, j\in U)\}\big) = &\sum_{\substack{\widetilde{U}\supseteq U\\ |\widetilde{U}|\leqslant s-|L|}} \sum_{T\subseteq (L\cap\widetilde{U})\setminus U} (-1)^{|((L\cap\widetilde{U})\setminus U)\setminus T|} \\
&\cdot \nu_{L,\widetilde{U}}\Big(\prod_{j\in L\setminus(((L\cap\widetilde{U})\setminus U)\setminus T)} \{v_{p_j}^{(j)}\} \times \prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T} \big((-\infty, v_{p_j}^{(j)})\setminus \underline{V}^{(j)}_{p_j}\big) \\
&\times \prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T} \{v_{p_j}^{(j)}\} \times \prod_{j\in T} \big((v_{p_j}^{(j)}, +\infty)\setminus \overline{V}^{(j)}_{p_j}\big) \times \prod_{j\in U} \{v_{q_j}^{(j)}\} \times \prod_{j\in(\widetilde{U}\setminus L)\setminus U} (\mathbb{R}\setminus V^{(j)})\Big)
\end{aligned}$$

对于 $(p_j, j\in L)\in\prod_{j\in L}[n_j]$ 和 $(q_j, j\in U)\in\prod_{j\in U}[n_j]$。注意，对于满足 $|L|+|U|\leqslant s$ 的 $L\subseteq S\subseteq L\cup U$，

$$
\begin{aligned}
&\mu_{L,U}\Big(\prod_{j\in(L\backslash U)\cup K}\{v_{m_j}^{(j)}\}\times\prod_{j\in(L\cap U)\backslash K}(-\infty,v_{m_j}^{(j)})\times\prod_{j\in K}(v_{m_j}^{(j)},+\infty)\times\prod_{\substack{j\in((L\cap U)\backslash K)\\ \cup((U\backslash L)\cap S)}}\{v_{m_j}^{(j)}\}\times\mathbb{R}^{|U\backslash S|}\Big)\\
&=\sum_{\mathbf{r}\in\prod_{j\in(L\cap U)\backslash K}\underline{V}_{m_j}^{(j)}\times\prod_{j\in K}\overline{V}_{m_j}^{(j)}\times\prod_{U\backslash S}V^{(j)}}\mu_{L,U}\Big(\prod_{j\in(L\backslash U)\cup K}\{v_{m_j}^{(j)}\}\times\prod_{j\in(L\cap U)\backslash K}\{v_{r_j}^{(j)}\}\\
&\qquad\times\prod_{j\in K}\{v_{r_j}^{(j)}\}\times\prod_{\substack{j\in((L\cap U)\backslash K)\\ \cup((U\backslash L)\cap S)}}\{v_{m_j}^{(j)}\}\times\prod_{j\in U\backslash S}\{v_{r_j}^{(j)}\}\Big)\\
&=\sum_{\mathbf{r}\in\prod_{j\in(L\cap U)\backslash K}\underline{V}_{m_j}^{(j)}\times\prod_{j\in K}\overline{V}_{m_j}^{(j)}\times\prod_{U\backslash S}V^{(j)}}\ \sum_{\substack{\tilde{U}\supseteq U\\ |\tilde{U}|\leqslant s-|L|}}\ \sum_{T\subseteq(L\cap\tilde{U})\backslash U}(-1)^{|((L\cap\tilde{U})\backslash U)\backslash T|}\\
&\qquad\cdot\nu_{L,\tilde{U}}\Big(\prod_{j\in(L\backslash\tilde{U})\cup K\cup T}\{v_{m_j}^{(j)}\}\times\prod_{j\in(L\cap U)\backslash K}\{v_{r_j}^{(j)}\}\times\prod_{j\in((L\cap\tilde{U})\backslash U)\backslash T}\big((-\infty,v_{m_j}^{(j)})\backslash\underline{V}_{m_j}^{(j)}\big)\\
&\qquad\times\prod_{j\in((L\cap\tilde{U})\backslash U)\backslash T}\{v_{m_j}^{(j)}\}\times\prod_{j\in T}\big((v_{m_j}^{(j)},+\infty)\backslash\overline{V}_{m_j}^{(j)}\big)\\
&\qquad\times\prod_{j\in K}\{v_{r_j}^{(j)}\}\times\prod_{\substack{j\in((L\cap U)\backslash K)\\ \cup((U\backslash L)\cap S)}}\{v_{m_j}^{(j)}\}\times\prod_{j\in U\backslash S}\{v_{r_j}^{(j)}\}\times\prod_{j\in(\tilde{U}\backslash L)\backslash U}(\mathbb{R}\backslash V^{(j)})\Big)\\
&=\sum_{\substack{\tilde{U}\supseteq U\\ |\tilde{U}|\leqslant s-|L|}}\ \sum_{T\subseteq(L\cap\tilde{U})\backslash U}(-1)^{|((L\cap\tilde{U})\backslash U)\backslash T|}\\
&\qquad\cdot\nu_{L,\tilde{U}}\Big(\prod_{j\in(L\backslash\tilde{U})\cup K\cup T}\{v_{m_j}^{(j)}\}\times\prod_{j\in(L\cap U)\backslash K}\underline{V}_{m_j}^{(j)}\times\prod_{j\in((L\cap\tilde{U})\backslash U)\backslash T}\big((-\infty,v_{m_j}^{(j)})\backslash\underline{V}_{m_j}^{(j)}\big)\\
&\qquad\times\prod_{j\in((L\cap\tilde{U})\backslash U)\backslash T}\{v_{m_j}^{(j)}\}\times\prod_{j\in T}\big((v_{m_j}^{(j)},+\infty)\backslash\overline{V}_{m_j}^{(j)}\big)\\
&\qquad\times\prod_{j\in K}\overline{V}_{m_j}^{(j)}\times\prod_{\substack{j\in((L\cap U)\backslash K)\\ \cup((U\backslash L)\cap S)}}\{v_{m_j}^{(j)}\}\times\prod_{j\in U\backslash S}V^{(j)}\times\prod_{j\in(\tilde{U}\backslash L)\backslash U}(\mathbb{R}\backslash V^{(j)})\Big).
\end{aligned}
$$

因此，式 (31) 的左侧对于 $\mu_{L,\ U}$ 变为

$$
\begin{aligned}
&\sum_{\substack{L,U:L\subseteq S\subseteq L\cup U\\ |L|+|U|\leqslant s}}(-1)^{|(U\backslash L)\cap S|}\sum_{K\subseteq L\cap U}(-1)^{|(L\cap U)\backslash K|}\\
&\qquad\cdot\mu_{L,U}\Big(\prod_{j\in(L\backslash U)\cup K}\{v_{m_j}^{(j)}\}\times\prod_{j\in(L\cap U)\backslash K}(-\infty,v_{m_j}^{(j)})\times\prod_{j\in K}(v_{m_j}^{(j)},+\infty)\times\prod_{\substack{j\in((L\cap U)\backslash K)\\ \cup((U\backslash L)\cap S)}}\{v_{m_j}^{(j)}\}\times\mathbb{R}^{|U\backslash S|}\Big)\\
&=\sum_{\substack{L,U:L\subseteq S\subseteq L\cup U\\ |L|+|U|\leqslant s}}(-1)^{|(U\backslash L)\cap S|}\sum_{K\subseteq L\cap U}(-1)^{|(L\cap U)\backslash K|}\sum_{\substack{\tilde{U}\supseteq U\\ |\tilde{U}|\leqslant s-|L|}}\ \sum_{T\subseteq(L\cap\tilde{U})\backslash U}(-1)^{|((L\cap\tilde{U})\backslash U)\backslash T|}\\
&\qquad\cdot\nu_{L,\tilde{U}}\Big(\prod_{j\in(L\backslash\tilde{U})\cup K\cup T}\{v_{m_j}^{(j)}\}\times\prod_{j\in(L\cap U)\backslash K}\underline{V}_{m_j}^{(j)}\times\prod_{j\in((L\cap\tilde{U})\backslash U)\backslash T}\big((-\infty,v_{m_j}^{(j)})\backslash\underline{V}_{m_j}^{(j)}\big)\\
&\qquad\times\prod_{j\in((L\cap\tilde{U})\backslash U)\backslash T}\{v_{m_j}^{(j)}\}\times\prod_{j\in T}\big((v_{m_j}^{(j)},+\infty)\backslash\overline{V}_{m_j}^{(j)}\big)\\
&\qquad\times\prod_{j\in K}\overline{V}_{m_j}^{(j)}\times\prod_{\substack{j\in((L\cap U)\backslash K)\\ \cup((U\backslash L)\cap S)}}\{v_{m_j}^{(j)}\}\times\prod_{j\in U\backslash S}V^{(j)}\times\prod_{j\in(\tilde{U}\backslash L)\backslash U}(\mathbb{R}\backslash V^{(j)})\Big).
\end{aligned}
$$

~ ~通过改变 U 和 U 中求和的顺序，并利用 $K = K \cup T$ 合并 K 和 T 中的求和，我们可以将右侧简化为

$$\begin{aligned}
&\sum_{\substack{L,\tilde{U}:L\subseteq S\subseteq L\cup\tilde{U}\\ |L|+|\tilde{U}|\leqslant s}} (-1)^{|(\tilde{U}\setminus L)\cap S|} \sum_{U:S\setminus L\subseteq U\subseteq\tilde{U}} \sum_{K\subseteq L\cap U} \sum_{T\subseteq(L\cap\tilde{U})\setminus U} (-1)^{|(L\cap U)\setminus K|}\cdot(-1)^{|((L\cap\tilde{U})\setminus U)\setminus T|} \\
&\qquad \cdot \nu_{L,\tilde{U}}\Big(\prod_{j\in(L\setminus\tilde{U})\cup K\cup T} \{v_{m_j}^{(j)}\} \times \prod_{j\in(L\cap U)\setminus K} \underline{V}_{m_j}^{(j)} \times \prod_{j\in((L\cap\tilde{U})\setminus U)\setminus T} \big((-\infty, v_{m_j}^{(j)})\setminus \underline{V}_{m_j}^{(j)}\big) \\
&\qquad\quad \times \prod_{j\in((L\cap\tilde{U})\setminus U)\setminus T} \{v_{m_j}^{(j)}\} \times \prod_{j\in T} \big((v_{m_j}^{(j)}, +\infty)\setminus \overline{V}_{m_j}^{(j)}\big) \\
&\qquad\quad \times \prod_{j\in K} \overline{V}_{m_j}^{(j)} \times \prod_{\substack{j\in((L\cap U)\setminus K)\\ \cup((U\setminus L)\cap S)}} \{v_{m_j}^{(j)}\} \times \prod_{j\in U\setminus S} V^{(j)} \times \prod_{j\in(\tilde{U}\setminus L)\setminus U} (\mathbb{R}\setminus V^{(j)})\Big) \\
=\ &\sum_{\substack{L,\tilde{U}:L\subseteq S\subseteq L\cup\tilde{U}\\ |L|+|\tilde{U}|\leqslant s}} (-1)^{|(\tilde{U}\setminus L)\cap S|} \sum_{\widetilde{K}\subseteq L\cap\tilde{U}} (-1)^{|(L\cap\tilde{U})\setminus\widetilde{K}|} \sum_{U:S\setminus L\subseteq U\subseteq\tilde{U}} \\
&\qquad \cdot \nu_{L,\tilde{U}}\Big(\prod_{j\in(L\setminus\tilde{U})\cup\widetilde{K}} \{v_{m_j}^{(j)}\} \times \prod_{j\in((L\cap\tilde{U})\setminus\widetilde{K})\cap U} \underline{V}_{m_j}^{(j)} \times \prod_{j\in((L\cap\tilde{U})\setminus\widetilde{K})\setminus U} \big((-\infty, v_{m_j}^{(j)})\setminus \underline{V}_{m_j}^{(j)}\big) \\
&\qquad\quad \times \prod_{j\in\widetilde{K}\setminus U} \big((v_{m_j}^{(j)}, +\infty)\setminus \overline{V}_{m_j}^{(j)}\big) \times \prod_{j\in\widetilde{K}\cap U} \overline{V}_{m_j}^{(j)} \times \prod_{\substack{j\in((L\cap\tilde{U})\setminus\widetilde{K})\\ \cup((\tilde{U}\setminus L)\cap S)}} \{v_{m_j}^{(j)}\} \\
&\qquad\quad \times \prod_{j\in U\setminus S} V^{(j)} \times \prod_{j\in(\tilde{U}\setminus L)\setminus U} (\mathbb{R}\setminus V^{(j)})\Big).
\end{aligned}$$

计算 U 上的内层求和，得到

$$\begin{aligned}
&\sum_{\substack{L,U:L\subseteq S\subseteq L\cup U\\ |L|+|U|\leqslant s}} (-1)^{|(U\setminus L)\cap S|} \sum_{K\subseteq L\cap U} (-1)^{|(L\cap U)\setminus K|} \\
&\qquad \cdot \mu_{L,U}\Big(\prod_{j\in(L\setminus U)\cup K} \{v_{m_j}^{(j)}\} \times \prod_{j\in(L\cap U)\setminus K} (-\infty, v_{m_j}^{(j)}) \times \prod_{j\in K} (v_{m_j}^{(j)}, +\infty) \times \prod_{\substack{j\in((L\cap U)\setminus K)\\ \cup((U\setminus L)\cap S)}} \{v_{m_j}^{(j)}\} \times \mathbb{R}^{|U\setminus S|}\Big) \\
=\ &\sum_{\substack{L,\tilde{U}:L\subseteq S\subseteq L\cup\tilde{U}\\ |L|+|\tilde{U}|\leqslant s}} (-1)^{|(\tilde{U}\setminus L)\cap S|} \sum_{\widetilde{K}\subseteq L\cap\tilde{U}} (-1)^{|(L\cap\tilde{U})\setminus\widetilde{K}|} \\
&\qquad \cdot \nu_{L,\tilde{U}}\Big(\prod_{j\in(L\setminus\tilde{U})\cup\widetilde{K}} \{v_{m_j}^{(j)}\} \times \prod_{j\in((L\cap\tilde{U})\setminus\widetilde{K})} (-\infty, v_{m_j}^{(j)}) \times \prod_{j\in\widetilde{K}} (v_{m_j}^{(j)}, +\infty) \times \prod_{\substack{j\in((L\cap\tilde{U})\setminus\widetilde{K})\\ \cup((\tilde{U}\setminus L)\cap S)}} \{v_{m_j}^{(j)}\} \times \mathbb{R}^{|\tilde{U}\setminus S|}\Big) \\
=\ &\Delta_{\mathbf{m}}^{S}(f),
\end{aligned}$$

这表明，(31) 对 $\mu_{L,U}$ 也成立。

此外，根据 $\mu_{L,U}$ 的定义，我们有

$$\begin{aligned}
\sum_{0<|L|+|U|\leqslant s} |\mu_{L,U}|(\mathbb{R}^{|L|+|U|}) &= \sum_{0<|L|+|U|\leqslant s} \sum_{\mathbf{p}\in\prod_{j\in L}[n_j]} \sum_{\mathbf{q}\in\prod_{j\in U}[n_j]} \big|\mu_{L,U}\big(\{(v_{p_j}^{(j)}, j\in L)\times(v_{q_j}^{(j)}, j\in U)\}\big)\big| \\
&\leqslant \sum_{0<|L|+|U|\leqslant s} \sum_{\mathbf{p}\in\prod_{j\in L}[n_j]} \sum_{\mathbf{q}\in\prod_{j\in U}[n_j]} \sum_{\substack{\tilde{U}\supseteq U\\ |\tilde{U}|\leqslant s-|L|}} \sum_{T\subseteq(L\cap\tilde{U})\setminus U} \\
&\qquad |\nu_{L,\tilde{U}}|\Big(\prod_{j\in L\setminus(((L\cap\tilde{U})\setminus U)\setminus T)} \{v_{p_j}^{(j)}\} \times \prod_{j\in((L\cap\tilde{U})\setminus U)\setminus T} \big((-\infty, v_{p_j}^{(j)})\setminus \underline{V}_{p_j}^{(j)}\big)
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad \times \prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}\{v_{p_j}^{(j)}\}\times\prod_{j\in T}\big((v_{p_j}^{(j)},+\infty)\setminus\overline{V}_{p_j}^{(j)}\big)\times\prod_{j\in U}\{v_{q_j}^{(j)}\}\times\prod_{j\in(\widetilde{U}\setminus L)\setminus U}(\mathbb{R}\setminus V^{(j)})\Big)\\
&=\sum_{0<|L|+|\widetilde{U}|\leqslant s}\sum_{U\subseteq\widetilde{U}}\sum_{T\subseteq(L\cap\widetilde{U})\setminus U}\sum_{\mathbf{p}\in\prod_{j\in L}[n_j]}\sum_{\mathbf{q}\in\prod_{j\in U}[n_j]}\\
&\qquad |\nu_{L,\widetilde{U}}|\Big(\prod_{j\in L\setminus(((L\cap\widetilde{U})\setminus U)\setminus T)}\{v_{p_j}^{(j)}\}\times\prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}\big((-\infty,v_{p_j}^{(j)})\setminus\underline{V}_{p_j}^{(j)}\big)\\
&\qquad\qquad \times \prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}\{v_{p_j}^{(j)}\}\times\prod_{j\in T}\big((v_{p_j}^{(j)},+\infty)\setminus\overline{V}_{p_j}^{(j)}\big)\times\prod_{j\in U}\{v_{q_j}^{(j)}\}\times\prod_{j\in(\widetilde{U}\setminus L)\setminus U}(\mathbb{R}\setminus V^{(j)})\Big)\\
&\leqslant\sum_{0<|L|+|\widetilde{U}|\leqslant s}\sum_{U\subseteq\widetilde{U}}\sum_{T\subseteq(L\cap\widetilde{U})\setminus U}\sum_{\mathbf{p}\in\prod_{j\in L}[n_j]}\sum_{\mathbf{q}\in\prod_{j\in U}[n_j]}\\
&\qquad |\nu_{L,\widetilde{U}}|\Big(\prod_{j\in L\setminus(((L\cap\widetilde{U})\setminus U)\setminus T)}\{v_{p_j}^{(j)}\}\times\prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}(\mathbb{R}\setminus V^{(j)})\\
&\qquad\qquad \times \prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}\{v_{p_j}^{(j)}\}\times\prod_{j\in T}(\mathbb{R}\setminus V^{(j)})\times\prod_{j\in U}\{v_{q_j}^{(j)}\}\times\prod_{j\in(\widetilde{U}\setminus L)\setminus U}(\mathbb{R}\setminus V^{(j)})\Big)\\
&=\sum_{0<|L|+|\widetilde{U}|\leqslant s}\sum_{U\subseteq\widetilde{U}}\sum_{T\subseteq(L\cap\widetilde{U})\setminus U}\\
&\qquad |\nu_{L,\widetilde{U}}|\Big(\prod_{j\in L\setminus(((L\cap\widetilde{U})\setminus U)\setminus T)}V^{(j)}\times\prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}(\mathbb{R}\setminus V^{(j)})\\
&\qquad\qquad \times \prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}V^{(j)}\times\prod_{j\in T}(\mathbb{R}\setminus V^{(j)})\times\prod_{j\in U}V^{(j)}\times\prod_{j\in(\widetilde{U}\setminus L)\setminus U}(\mathbb{R}\setminus V^{(j)})\Big)\\
&\leqslant\sum_{0<|L|+|\widetilde{U}|\leqslant s}|\nu_{L,\widetilde{U}}|(\mathbb{R}^{|L|+|\widetilde{U}|}).
\end{aligned}
$$

~这里，在第二个等式中，我们交换了 U 和 U 中求和的顺序；在第二个不等式中，我们利用了以下事实:

$$
(-\infty,v_{p_j}^{(j)})\setminus\underline{V}_{p_j}^{(j)}\subseteq\mathbb{R}\setminus V^{(j)}\quad\text{and}\quad(v_{p_j}^{(j)},+\infty)\setminus\overline{V}_{p_j}^{(j)}\subseteq\mathbb{R}\setminus V^{(j)}\ \text{for all } j.
$$

~最后一个不等式是基于以下观察得出的：对于每个 L 和 U，集合

$$
\begin{aligned}
&\prod_{j\in L\setminus(((L\cap\widetilde{U})\setminus U)\setminus T)}V^{(j)}\times\prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}(\mathbb{R}\setminus V^{(j)})\\
&\qquad\times\prod_{j\in((L\cap\widetilde{U})\setminus U)\setminus T}V^{(j)}\times\prod_{j\in T}(\mathbb{R}\setminus V^{(j)})\times\prod_{j\in U}V^{(j)}\times\prod_{j\in(\widetilde{U}\setminus L)\setminus U}(\mathbb{R}\setminus V^{(j)})
\end{aligned}
$$

~ ~当 U 遍历 U 的子集、T 遍历 $(L\cap U)\setminus U$ 的子集时，两两不相交。这表明，对于 $\mu_{L,\,U}$，(32) 的目标函数不大于对于 $\nu_{L,\,U}$ 时的目标函数，并保证 (32) 的下确界由某些支撑在 (33) 上的离散有符号测度 $\mu_{L,\,U}$ 取得。

假设 $\mu_{L,\,U}$ 是支撑在格点（33）上且满足 (31) 的离散有符号测度。我们可以将这些测度参数化为

$$
\mu_{L,U}\big(\{(v_{p_j}^{(j)},j\in L)\times(v_{q_j}^{(j)},j\in U)\}\big)=\beta_{\mathbf{p},\mathbf{q}}^{L,U}\tag{34}
$$

对于 $L,\ U\subseteq[d]$ 给定 $0<|L|+|U|\leqslant s$、$\mathbf{p}=(p_j,\ \ j\in L)\in\prod_{j\in L}[n_j]$ 和 $\mathbf{q}=(q_j,\ \ j\in U)\in\prod_{j\in U}[n_j]$。在这种参数化下，条件 (31) 可以完全用 $\beta^{L,\,U}_{\mathbf{p}\,\mathbf{q}}$，表示为

$$
\sum_{\substack{L,U:L\subseteq S\subseteq L\cup U\\ |L|+|U|\leqslant s}}(-1)^{|(U\setminus L)\cap S|}\sum_{K\subseteq L\cap U}(-1)^{|(L\cap U)\setminus K|}\sum_{\mathbf{r}\in\prod_{j\in(L\cap U)\setminus K}\underline{V}_{m_j}^{(j)}\times\prod_{j\in K}\overline{V}_{m_j}^{(j)}\times\prod_{U\setminus S}V^{(j)}}
$$

$$\beta^{L,U}_{(m_j j\in(L\setminus U)\cup K;\ r_j j\in(L\cap U)\setminus K)\times(r_j j\in K;\ m_j j\in((L\cap U)\setminus K)\cup((U\setminus L)\cap S);\ r_j j\in U\setminus S)} = \Delta^S_{\mathbf{m}}(f). \tag{35}$$

因此，(32) 变为

$$V^{d,s}_{\infty-\mathrm{XGB}}(f) \geqslant \inf\big\{\|\beta\|_1 : \beta^{L,U}_{\mathbf{p},\mathbf{q}} \text{ satisfy (35)}\big\}.$$

由于 (35) 是一个线性方程组，显然存在使右端达到极小值的极小化元 $\beta^{L,\mathbf{p}\ U,\mathbf{q}}$。令 $\hat{\beta}^{L,\mathbf{p}\ U,\mathbf{q}}$表示其中一个这样的极小化元，并令 $\hat{\mu}_{L,U}$ 为通过 (34) 定义的相应离散有符号测度。

任取一个常数 $c_0 \in \mathbb{R}$。根据构造，函数 $f^{d,s}_{c_0,\{\hat{\mu}_{L,U}\}}$如 (28) 所示为分段常数，并满足 (29)。根据引理 5，存在一个常数 $b \in \mathbb{R}$，使得 $f(\mathbf{x}) = b + f^{d,s}_{c_0,\{\hat{\mu}_{L,U}\}}(\mathbf{x})$对所有 $\mathbf{x} \in \mathbb{R}^d$均成立。定义 $c = b + c_0$，于是有 $f^{d,s}_{c,\{\hat{\mu}_{L,U}\}} \equiv f$。由此可得

$$\|\hat{\beta}\|_1 = \sum_{0<|L|+|U|\leqslant s} \|\hat{\mu}_{L,U}\|_{\mathrm{TV}} \geqslant V^{d,s}_{\infty-\mathrm{XGB}}(f) \geqslant \|\hat{\beta}\|_1,$$

从而

$$V^{d,s}_{\infty-\mathrm{XGB}}(f) = \sum_{0<|L|+|U|\leqslant s} \|\hat{\mu}_{L,U}\|_{\mathrm{TV}}.$$

因此，$V^{d,s}_{\infty-\mathrm{XGB}}(f)$ 定义中的最小值由离散有符号测度取得。这证明了$V^{d,s}_{\infty-\mathrm{XGB}}(f) = V^{d,s}_{\mathrm{XGB}}(f)$。

□

引理 5 的证明。由于交替和泛函是线性的，只需证明：若 $f \in \mathcal{F}^{d,s}_{\mathrm{ST}}$ 如（28）所示为分段常数，并满足

$$\Delta^S_{\mathbf{m}}(f) = 0 \tag{36}$$

对于所有 $\varnothing \neq S \subseteq [d]$，其中 $|S| \leqslant s$ 且 $\mathbf{m} \in \prod_{j\in S}[n_j]$，则 f 是常值函数。

固定这样一个 $f \in \mathcal{F}^{d,s}_{\mathrm{ST}}$。如（30）所示，$\Delta^S_{\mathbf{m}}(f)$ 的展开式涉及对满足 $|L|+|U| \leqslant s$的 $L \subseteq S \subseteq L \cup U$求和。当 $|S| > s$时，该求和为空，因此 $\Delta^S_{\mathbf{m}}(f) = 0$自动成立。于是，（36）实际上对所有非空的 $S \subseteq [d]$都成立。

对于每个 $j \in [d]$，定义

$$w^{(j)}_0 = v^{(j)}_1 - 1 \quad \text{and} \quad w^{(j)}_{m_j} = v^{(j)}_{m_j} \text{ for } m_j \in [n_j],$$

并令 $\boldsymbol{\phi}$表示 f在 $(w^{(1)}_{m_1}、\ldots、w^{(d)}_{m_d})$处的函数值评估向量；也就是说，

$$\phi(\mathbf{m}) = f(w^{(1)}_{m_1}, \ldots, w^{(d)}_{m_d}) \quad \text{for } \mathbf{m} = (m_1, \ldots, m_d) \in \prod_{j=1}^{d}\{0, \ldots, n_j\}.$$

由于 f 如（28）所示为分段常数且右连续，因此向量 $\boldsymbol{\phi}$ 完全确定了 f。此外，出于同样的原因，（36）等价于

$$\Delta^S_{\mathbf{m}}\boldsymbol{\phi} := \sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}} (-1)^{\sum_{j\in S}\delta_j} \cdot \phi\big(m_j - \delta_j, j \in S; 0, j \in S^c\big) = 0 \quad \text{for } \mathbf{m} = (m_j, j \in S) \in \prod_{j\in S}[n_j] \tag{37}$$

对于所有非空的 S。因此，只需证明：若 $\boldsymbol{\phi}$ 满足（37），则 $\boldsymbol{\phi}$ 是一个常数向量 r.

我们通过对 d进行归纳法来证明这一断言。$d = 1$ 的情形显而易见。当 $d = 1$时，(37) 化简为

$$\Delta^{\{1\}}_m\boldsymbol{\phi} = \phi(m) - \phi(m-1) = 0 \quad \text{for } m \in [n_1].$$

因此，在这种情况下，显然 $\boldsymbol{\phi}$ 是一个常向量。假设该断言对 $d-1$成立，下面证明它对 d也成立。对于每个 $m_d \in \{0, \ldots\ldots, n_d\}$，令 $\boldsymbol{\phi}^{(m_d)}$ 表示 $\boldsymbol{\phi}$ 的子向量，其最后一个索引为 m_d；也就是说，

$$\phi^{(m_d)}(m_1, \ldots, m_{d-1}) = \phi(m_1, \ldots, m_{d-1}, m_d) \quad \text{for } (m_1, \ldots, m_{d-1}) \in \prod_{j=1}^{d-1} \{0, \ldots, n_j\}.$$

显然， $\boldsymbol{\phi}^{(0)}$ 满足 $d-1$的条件 **(37)**。注意，对于 $m_d \in [n_d]$,

$$\Delta^{S}_{(m_1,\ldots,m_{d-1})} \boldsymbol{\phi}^{(m_d)} - \Delta^{S}_{(m_1,\ldots,m_{d-1})} \boldsymbol{\phi}^{(m_d-1)} = \Delta^{S\cup\{d\}}_{(m_1,\ldots,m_{d-1},m_d)} \boldsymbol{\phi} = 0$$

对于每个非空的 $S \subseteq [d-1]$ 和 $(m_1, \ldots, m_{d-1}) \in \prod_{j=1}^{d-1}[n_j]$，结论都成立。因此，所有 $\boldsymbol{\phi}^{(m_d)}$ 都满足 $d-1$的条件 **(37)**。根据归纳假设可知，每个 $\boldsymbol{\phi}^{(m_d)}$ 都是一个常向量。最后，在 **(37)** 中取 $S = \{d\}$，得到

$$\Delta^{\{d\}}_{m_d} \boldsymbol{\phi} = \phi(0, \ldots, 0, m_d) - \phi(0, \ldots, 0, m_d - 1) = 0 \quad \text{for } m_d \in [n_d].$$

因此，这些 $\boldsymbol{\phi}^{(m_d)}$ 中的常数对于所有 m_d都相同，这意味着 $\boldsymbol{\phi}$ 是一个常向量 . □

A.1.4 (13) 的证明

在证明中，我们使用实分析中的以下标准结果。

定理 6 (Folland 的定理 3.29 [13])。 假设 $f: \mathbb{R} \to \mathbb{R}$ 具有有限全变差且右连续。那么，存在唯一的常数 $c \in \mathbb{R}$ 和唯一的有限符号博雷尔测度 λ，定义在 $\mathbb{R}$ 上，使得

$$f(x) = c + \int \mathbf{1}(x \geqslant t)\, d\lambda(t) \quad \textit{for } x \in \mathbb{R}. \tag{38}$$

反过来，若 $f: \mathbb{R} \to \mathbb{R}$ 具有 (38) 的形式，则 f 具有有限全变差、右连续，且全变差$(f) = \|\lambda\|_{TV}$。

(13) 的证明 **(13)**。 我们先证明

$$\mathcal{F}^{1,1}_{\infty-\mathrm{ST}} = \big\{f : \mathrm{TV}(f) < \infty \text{ and } f \text{ is right-continuous}\big\}$$

并且

$$V^{1,1}_{\infty-\mathrm{XGB}}(f) = \mathrm{TV}(f)$$

对于 $f \in \mathcal{F}^{1,1}_{\infty-\mathrm{ST}}$。

假设 $f = f^{1,1}_{c,\{\nu_{L,U}\}} \in \mathcal{F}^{1,1}_{\infty-\mathrm{ST}}$。由于 $d = s = 1$，满足 $0 < |L| + |U| \leqslant s$ 的 (L, U) 的可行配对只有 $(\{1\}, \varnothing)$ 和 $(\varnothing, \{1\})$。因此， f可以表示为

$$f(x) = c + \int \mathbf{1}(x \geqslant l)\, d\nu_{\{1\},\varnothing}(l) + \int \mathbf{1}(x < u)\, d\nu_{\varnothing,\{1\}}(u)$$

对于 $x \in \mathbb{R}$。定义有符号博雷尔测度 λ，使得$\lambda = \nu_{\{1\},\varnothing} - \nu_{\varnothing,\{1\}}$。然后，

$$f(x) = c + \nu_{\varnothing,\{1\}}(\mathbb{R}) + \int \mathbf{1}(x \geqslant l)\, d\lambda(l)$$

对于 $x \in \mathbb{R}$，以及

$$\|\lambda\|_{\mathrm{TV}} = |\lambda|(\mathbb{R}) \leqslant |\nu_{\{1\},\varnothing}|(\mathbb{R}) + |\nu_{\varnothing,\{1\}}|(\mathbb{R}) = \|\nu_{\{1\},\varnothing}\|_{\mathrm{TV}} + \|\nu_{\varnothing,\{1\}}\|_{\mathrm{TV}}.$$

因此，每个 $f \in \mathcal{F}^{1,\,\infty^{-}\,\mathrm{ST}}$ 都可以用更简单的表示形式表示

$$f_{c,\lambda}(x) := c + \int \mathbf{1}(x \geqslant l)\, d\lambda(l),$$

其复杂度 $V^{1,\,\infty^{-}\,\mathrm{XGB}}(f)$ 可由下式计算：

$$V^{1,1}_{\infty-\mathrm{XGB}}(f) = \inf\{\|\lambda\|_{\mathrm{TV}} : f_{c,\lambda} \equiv f\}.$$

根据定理 6，这类函数 $f_{c,\,\lambda}$ 恰好是所有有限全变差的右连续函数。此外，满足 $f_{c,\,\lambda} \equiv f$ 的数对 $(c,\ \lambda)$ 是唯一的，并且满足 $\|\lambda\|_{\mathrm{TV}} = \mathrm{TV}(f)$。因此，

$$\mathcal{F}^{1,1}_{\infty-\mathrm{ST}} = \big\{f : \mathrm{TV}(f) < \infty \text{ and } f \text{ is right-continuous}\big\},$$

并且对于每个 $f \in \mathcal{F}^{1,\,\infty^{-}\,\mathrm{ST}}$,

$$V^{1,1}_{\infty-\mathrm{XGB}}(f) = \mathrm{TV}(f).$$

接下来我们证明

$$V^{1,2}_{\infty-\mathrm{XGB}}(f) = \frac{1}{2}\cdot\big(\mathrm{TV}(f) + |\Delta(f)|\big)$$

对于所有 $f \in \mathcal{F}^{1,\,\infty^{-}\,\mathrm{ST}} (= \mathcal{F}^{1,\,\infty^{-}\,\mathrm{ST}})$。通过与上述相同的论证，我们可以证明每个 $f \in \mathcal{F}^{1,\,\infty^{-}\,\mathrm{ST}}$ 都存在如下表示：

$$f_{c,\lambda,\mu}(x) := c + \int \mathbf{1}(x \geqslant l)\, d\lambda(l) + \int \mathbf{1}(l \leqslant x < u)\, d\mu(l,u),$$

其中 λ 和 μ 分别是 $\mathbb{R}$ 和 $\mathbb{R}^2$上的有限符号博雷尔测度，其复杂度 $V^{1,\,\infty^{-}\,\mathrm{XGB}}(f)$ 由下式给出：

$$V^{1,2}_{\infty-\mathrm{XGB}}(f) = \inf\big\{\|\lambda\|_{\mathrm{TV}} + \|\mu\|_{\mathrm{TV}} : f_{c,\lambda,\mu} \equiv f\big\}.$$

首先，假设 $f \equiv f_{c,\,\lambda,\,\mu} \in \mathcal{F}^{1,\,\infty^{-}\,\mathrm{ST}}$。对于每个 $x < y$，我们有

$$\begin{aligned}|f(x) - f(y)| &= \big| -\lambda((x,y]) + \mu\big((-\infty,x]\times(x,y]\big) - \mu\big((x,y]\times(y,+\infty)\big)\big| \\ &\leqslant |\lambda|((x,y]) + |\mu|\big(\mathbb{R}\times(x,y]\big) + |\mu|\big((x,y]\times\mathbb{R}\big),\end{aligned}$$

由此可得

$$\mathrm{TV}(f) \leqslant \|\lambda\|_{\mathrm{TV}} + 2\|\mu\|_{\mathrm{TV}}.$$

此外，我们有

$$|\Delta(f)| = \big|\lim_{x\to+\infty} f(x) - \lim_{x\to-\infty} f(x)\big| = |\lambda(\mathbb{R})| \leqslant \|\lambda\|_{\mathrm{TV}}.$$

结合这两个不等式，得到

$$\frac{1}{2}\cdot\big(\mathrm{TV}(f) + |\Delta(f)|\big) \leqslant \|\lambda\|_{\mathrm{TV}} + \|\mu\|_{\mathrm{TV}}.$$

对所有满足 $f_{c,\,\lambda,\,\mu} \equiv f$ 的 λ 和 μ 取下确界，得到

$$\frac{1}{2}\cdot\big(\mathrm{TV}(f) + |\Delta(f)|\big) \leqslant V^{1,2}_{\infty-\mathrm{XGB}}(f),$$

这就证明了所需恒等式的一个方向。

现在我们证明反向不等式。设 $f \in \mathcal{F}^{1,\infty^{-}\,\mathrm{ST}} (= \mathcal{F}^{1,\infty^{-}\,\mathrm{ST}})$。由于 f 具有有限全变差且右连续，定理 6 保证存在一个常数 $c \in \mathbb{R}$ 以及定义在 $\mathbb{R}$上的有限符号博雷尔测度 λ，使得

$$f(x) = c + \int \mathbf{1}(x \geqslant l)\, d\lambda(l) \quad \text{for } x \in \mathbb{R}$$

并且 $\|\lambda\|_{\mathrm{TV}}$ =全变差(f)。设 $\lambda = \lambda^+ - \lambda^-$为 λ的 Jordan 分解，并设 (P, N)为一个 Hahn分解。不失一般性，我们假设

$$\lambda^+(\mathbb{R}) \geqslant \lambda^-(\mathbb{R}).$$

于是，

$$|\Delta(f)| = \big|\lim_{x\to+\infty} f(x) - \lim_{x\to-\infty} f(x)\big| = |\lambda(\mathbb{R})| = \lambda^+(\mathbb{R}) - \lambda^-(\mathbb{R}),$$

因此，

$$\frac{1}{2}\cdot\big(\mathrm{TV}(f) + |\Delta(f)|\big) = \frac{1}{2}\cdot\Big(\lambda^+(\mathbb{R}) + \lambda^-(\mathbb{R}) + \lambda^+(\mathbb{R}) - \lambda^-(\mathbb{R})\Big) = \lambda^+(\mathbb{R}).$$

~在 $\mathbb{R}$ 上定义一个 Borel 测度 λ，其定义为

$$\widetilde{\lambda}(E) = \Big(1 - \frac{\lambda^-(\mathbb{R})}{\lambda^+(\mathbb{R})}\Big)\cdot\lambda^+(E)$$

~对于博雷尔集 $E \subseteq \mathbb{R}$。假设 $\lambda^+(\mathbb{R}) \geqslant \lambda^-(\mathbb{R})$ 确保对于所有 E， $\lambda(E)$ 均为非负数。接下来，在 $\mathbb{R}^2$ 上定义一个有符号博雷尔测度 μ，其定义为

$$\mu(E_1\times E_2) = \frac{1}{\lambda^+(\mathbb{R})}\cdot\Big(\lambda^+(E_1)\cdot\lambda^-(E_2) - \lambda^-(E_1)\cdot\lambda^+(E_2)\Big)$$

对于博雷尔集 E_1， $E_2 \subseteq \mathbb{R}$。根据构造，

$$\mu(E_1\times E_2) = \begin{cases} \lambda^+(E_1)\cdot\lambda^-(E_2)/\lambda^+(\mathbb{R}) \geqslant 0 & \text{if } E_1\subseteq P, E_2\subseteq N,\\ -\lambda^-(E_1)\cdot\lambda^+(E_2)/\lambda^+(\mathbb{R}) \leqslant 0 & \text{if } E_1\subseteq N, E_2\subseteq P,\\ 0 & \text{otherwise,}\end{cases}$$

因此，

$$\|\mu\|_{\mathrm{TV}} = \mu(P\times N) - \mu(N\times P) = \frac{2\lambda^+(P)\cdot\lambda^-(N)}{\lambda^+(\mathbb{R})} = 2\lambda^-(\mathbb{R}).$$

定义有符号博雷尔测度 $\widetilde{\mu}$ on $\mathbb{R}^2$ 如下:

$$d\widetilde{\mu}(l,u) = 1(l\leqslant u)\cdot d\mu(l,u).$$

由于 μ i 是反对称的，即对于所有博雷尔集 E_1， $E_2 \subseteq \mathbb{R}$，我们有 e

$$\|\widetilde{\mu}\|_{\mathrm{TV}} = \frac{1}{2}\cdot\|\mu\|_{\mathrm{TV}} = \lambda^-(\mathbb{R}).$$

因此，

$$\|\widetilde{\lambda}\|_{\mathrm{TV}} + \|\widetilde{\mu}\|_{\mathrm{TV}} = \widetilde{\lambda}(\mathbb{R}) + \|\widetilde{\mu}\|_{\mathrm{TV}} = \lambda^+(\mathbb{R}) - \lambda^-(\mathbb{R}) + \lambda^-(\mathbb{R}) = \lambda^+(\mathbb{R}).$$

现在，注意到

$$\int \mathbf{1}(x\geqslant l)\, d\widetilde{\lambda}(l) = \int \mathbf{1}(x\geqslant l)\, d\lambda^+(l) - \frac{\lambda^-(\mathbb{R})}{\lambda^+(\mathbb{R})}\cdot\lambda^+((-\infty, x])$$

以及

$$\begin{aligned}\int \mathbf{1}(l\leqslant x<u)\, d\widetilde{\mu}(l,u) &= \int \mathbf{1}(l\leqslant x<u)\, d\mu(l,u) = \mu\big((-\infty,x]\times(x,+\infty)\big)\\ &= \frac{1}{\lambda^+(\mathbb{R})}\cdot\Big(\lambda^+((-\infty,x])\cdot\lambda^-((x,+\infty)) - \lambda^-((-\infty,x])\cdot\lambda^+((x,+\infty))\Big)\\ &= \frac{\lambda^-(\mathbb{R})}{\lambda^+(\mathbb{R})}\cdot\lambda^+((-\infty,x]) - \lambda^-((-\infty,x]) = \frac{\lambda^-(\mathbb{R})}{\lambda^+(\mathbb{R})}\cdot\lambda^+((-\infty,x]) - \int \mathbf{1}(x\geqslant l)\, d\lambda^-(l)\end{aligned}$$

对于每个 $x \in \mathbb{R}$。结合这两个方程可得

$$f_{c,\tilde{\lambda},\tilde{\mu}}(x) = c + \int \mathbf{1}(x \geqslant l)\, d\tilde{\lambda}(l) + \int \mathbf{1}(l \leqslant x < u)\, d\tilde{\mu}(l,u) = c + \int \mathbf{1}(x \geqslant l)\, d\lambda(l) = f(x)$$

对每个 $x \in \mathbb{R}$ 都成立。因此，

$$V^{1,2}_{\infty-\mathrm{XGB}}(f) \leqslant \|\tilde{\lambda}\|_{\mathrm{TV}} + \|\tilde{\mu}\|_{\mathrm{TV}} = \lambda^{+}(\mathbb{R}) = \frac{1}{2} \cdot \big(\mathrm{TV}(f) + |\Delta(f)|\big),$$

这就证明了反向不等式。 □

A.1.5 命题 5 的证明

在命题 5 的证明中，我们使用以下定理，该定理将紧致定义域上具有有限 Hardy–Krause 变差的函数与同一定义域上有限符号博雷尔测度的累积分布函数联系起来。这一结果也将在命题3和命题7的证明中发挥核心作用。

为陈述这一结果，我们首先回顾紧致定义域上 Hardy–Krause 变差的定义。令 $f\colon \prod_{j=1}^{m}[u_j,\ v_j] \to \mathbb{R}$ 以及 一个 $= (a_1,\ \ldots\ldots,\ a_m) \in \prod_{j=1}^{m}\{u_j,\ v_j\}$。对于每个 $S \subseteq [m]$，定义

$$f^{S}_{(a_j, j\in S^c)}(x_j, j \in S) = f(x_j, j \in S; a_j, j \in S^c) \quad \text{for } (x_j, j \in S) \in \mathbb{R}^{|S|}.$$

由于定义域是紧致的，因此无需像（9）中那样取极限。于是，在 $\prod_{j=1}^{m}[u_j,\ v_j]$ 上以 $\mathbf{a}$ 为锚点的 f Hardy–Krause 变差定义为

$$\mathrm{HK}_{\mathbf{a}}\Big(f; \prod_{j=1}^{m}[u_j, v_j]\Big) = \sum_{0<|S|\leqslant m} \mathrm{Vit}\Big(f^{S}_{(a_j, j\in S^c)}; \prod_{j\in S}[u_j, v_j]\Big).$$

定理 7 (Aistleitner和Dick的定理 3 [1])。 **</style}设：** $\prod_{j=1}^{m}[u_j, v_j] \to \mathbb{R}$ 右连续，且相对于锚点 $\mathbf{a} = (u_1, \ldots, u_m)$ 具有有限的 *Hardy–Krause* 变差。那么，存在唯一的有限符号博雷尔测度 ν，定义在 $\prod_{j=1}^{m}[u_j, v_j]$ 上，使得

$$f(x_1, \ldots, x_m) = \nu\Big(\prod_{j=1}^{m}[u_j, x_j]\Big) \quad \textit{for } (x_1, \ldots, x_m) \in \prod_{j=1}^{m}[u_j, v_j]. \tag{39}$$

反过来， 若 f 具有(39)的形式，则 f 相对于锚点$\mathbf{a}$ 具有有限的 *Hardy–Krause* 变差，并且

$$\|\nu\|_{TV} = HK_{\mathbf{a}}\Big(f; \prod_{j=1}^{m}[u_j, v_j]\Big) + |f(\mathbf{a})|.$$

命题 *5* 的证明。 为证明 (15)，只需证明

$$\mathrm{HK}_{\mathbf{a}}(f)/\min(2^s - 1, 2^d) \leqslant V^{d,s}_{\infty-\mathrm{XGB}}(f) \leqslant \mathrm{HK}_{\mathbf{a}}(f)$$

对每个锚点$\mathbf{a} \in \{-\infty, +\infty\}^d$均成立。这里，我们仅对$\mathbf{a} = (-\infty, \ldots, -\infty)$这种情况证明该不等式。相同的论证也适用于其他锚点选择。

回忆一下， $\mathcal{F}^{d,\ {}^{s}}_{\infty-\mathrm{ST}}$是所有形如（7）的函数 $f^{d,\ c\ {}^{s},\ \{\nu_{L,\ U}\}}$的集合。对于每个 $f^{d,\ c\ {}^{s},\ \{\nu_{L,\ U}\}} \in \mathcal{F}^{d,\ {}^{s}}_{\infty-\mathrm{ST}}$，通过修改每个基函数 $b^{L,\ U\ \mathbf{l},\ \mathbf{u}}$

$$\begin{aligned} b^{L,U}_{\mathbf{l},\mathbf{u}}(x_1, \ldots, x_d) &= \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j) \\ &= \prod_{j\in L\setminus U} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U\setminus L} \big(1 - \mathbf{1}(x_j \geqslant u_j)\big) \cdot \prod_{j\in L\cap U} \big(\mathbf{1}(x_j \geqslant l_j) - \mathbf{1}(x_j \geqslant u_j)\big) \cdot \prod_{j\in L\cap U} \mathbf{1}(l_j \leqslant u_j), \end{aligned} \tag{40}$$

我们可以将 $f^{d,\ s}_{c,\{\nu_{L,\ U}\}}$ 表示为

$$f^{d,s}_{c,\{\nu_{L,U}\}}(x_1,\dots,x_d)=f^{d,s}_{b,\{\mu_S\}}(x_1,\dots,x_d):=b+\sum_{0<|S|\leqslant s}\int_{\mathbb{R}^{|S|}}\prod_{j\in S}\mathbf{1}(x_j\geqslant t_j)\,d\mu_S(t_j,j\in S)\tag{41}$$

对于某个 $b\in\mathbb{R}$ 以及定义在 $\mathbb{R}^{|S|}$ 上的有限符号博雷尔测度 μ_S，其中求和遍历所有满足 $|S|\leqslant s$ 的非空 $S\subseteq[d]$。具体而言，每个 μ_S 都与原测度 $\nu_{L,\ U}$ 相关，其关系为

$$\begin{aligned}\mu_S\Big(\prod_{j\in S}E_j\Big)=&\sum_{\substack{L,U:L\subseteq S\subseteq L\cup U\\|L|+|U|\leqslant s}}(-1)^{|(U\setminus L)\cap S|}\sum_{K\subseteq L\cap U}(-1)^{|(L\cap U)\setminus K|}\\&\cdot\bar\nu_{L,U}\Big(\prod_{j\in(L\setminus U)\cup K}E_j\times\mathbb{R}^{|(L\cap U)\setminus K|}\times\mathbb{R}^{|(U\setminus S)\cup K|}\times\prod_{\substack{j\in((L\cap U)\setminus K)\\\cup((U\setminus L)\cap S)}}E_j\Big)\end{aligned}\tag{42}$$

对于任意 $j\in S$，以及博雷尔集 $E_j\subseteq\mathbb{R}$，其中 $\bar\nu_{L,\ U}$ 是定义在 $\mathbb{R}^{|L|+|U|}$ 上的有符号博雷尔测度，其定义为

$$d\bar\nu_{L,U}(\mathbf{l},\mathbf{u})=\prod_{j\in L\cap U}\mathbf{1}(l_j\leqslant u_j)\cdot d\nu_{L,U}(\mathbf{l},\mathbf{u}).$$

μ_S 与 $\nu_{L,\ U}$ 之间的这一关系意味着

$$\begin{aligned}\sum_{0<|S|\leqslant s}|\mu_S|(\mathbb{R}^{|S|})&\leqslant\sum_{0<|S|\leqslant s}\sum_{\substack{L,U:L\subseteq S\subseteq L\cup U\\|L|+|U|\leqslant s}}\sum_{K\subseteq L\cap U}|\bar\nu_{L,U}|(\mathbb{R}^{|L|+|U|})\\&=\sum_{0<|L|+|U|\leqslant s}\sum_{S:L\subseteq S\subseteq L\cup U,S\neq\varnothing}\sum_{K\subseteq L\cap U}|\bar\nu_{L,U}|(\mathbb{R}^{|L|+|U|})\\&=\sum_{0<|L|+|U|\leqslant s}\big(\mathbf{1}(L\neq\varnothing)\cdot2^{|U\setminus L|}+\mathbf{1}(L=\varnothing)\cdot(2^{|U\setminus L|}-1)\big)\cdot2^{|L\cap U|}\cdot|\bar\nu_{L,U}|(\mathbb{R}^{|L|+|U|})\\&\leqslant\min(2^s-1,2^d)\cdot\sum_{0<|L|+|U|\leqslant s}|\nu_{L,U}|(\mathbb{R}^{|L|+|U|}).\end{aligned}\tag{43}$$

定义

$$V_{\mathbf{a}}(f)=\inf\Big\{\sum_{0<|S|\leqslant s}\|\mu_S\|_{\mathrm{TV}}:f^{d,s}_{b,\{\mu_S\}}\equiv f\Big\}\quad\text{for }f\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}.\tag{44}$$

由（43）可得

$$V^{d,s}_{\infty-\mathrm{XGB}}(f)\leqslant V_{\mathbf{a}}(f)\leqslant\min(2^s-1,2^d)\cdot V^{d,s}_{\infty-\mathrm{XGB}}(f)\quad\text{for every }f\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}.$$

因此，只需证明

$$V_{\mathbf{a}}(f)=\mathrm{HK}_{\mathbf{a}}(f)\quad\text{for every }f\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}}.$$

固定 $f\equiv f^{d,\ s}_{b,\ \{\mu_S\}}\in\mathcal{F}^{d,\ s}_{\infty-\mathrm{ST}}$。对于每个非空的 $S\subseteq[d]$，我们有

$$f^S_{(a_j,j\in S^c)}(x_j,j\in S)=b+\sum_{\substack{R:\varnothing\neq R\subseteq S\\|R|\leqslant s}}\int_{\mathbb{R}^{|R|}}\prod_{j\in R}\mathbf{1}(x_j\geqslant t_j)\,d\mu_R(t_j,j\in R).$$

因此，对于每个非空的 $T\subseteq[d]$,

$$\sum_{S\subseteq T}(-1)^{|T|-|S|}\cdot f^S_{(a_j,j\in S^c)}(x_j,j\in S)=\sum_{\substack{R:\varnothing\neq R\subseteq T\\|R|\leqslant s}}\Big(\sum_{S:R\subseteq S\subseteq T}(-1)^{|T|-|S|}\Big)\int_{\mathbb{R}^{|R|}}\prod_{j\in R}\mathbf{1}(x_j\geqslant t_j)\,d\mu_R(t_j,j\in R).$$

除非 $R=T$，否则内层求和为零；若 $R=T$，则其值为 $\mathbf{1}$。因此，如果 $|T|\leqslant s$，

$$\sum_{S\subseteq T}(-1)^{|T|-|S|}\cdot f^{S}_{(a_j,j\in S^c)}(x_j,j\in S)=\int_{\mathbb{R}^{|T|}}\prod_{j\in T}\mathbf{1}(x_j\geqslant t_j)\,d\mu_T(t_j,j\in T),$$

而当 $|T|>s$ 时，该表达式为零。

现在，固定一个非空的 $T\subseteq[d]$，使得 $|T|\leqslant s$。由于

$$\mathrm{Vit}(f^{T}_{(a_j,j\in T^c)})=\mathrm{Vit}\Big((x_j,j\in T)\mapsto\sum_{S\subseteq T}(-1)^{|T|-|S|}\cdot f^{S}_{(a_j,j\in S^c)}(x_j,j\in S)\Big),$$

我们有

$$\begin{aligned}\mathrm{Vit}(f^{T}_{(a_j,j\in T^c)})&=\mathrm{Vit}\Big((x_j,j\in T)\mapsto\int_{\mathbb{R}^{|T|}}\prod_{j\in T}\mathbf{1}(x_j\geqslant t_j)\,d\mu_T(t_j,j\in T)\Big)\\&=\sup_{u_j<v_j,j\in T}\mathrm{Vit}\Big((x_j,j\in T)\mapsto\int_{\mathbb{R}^{|T|}}\prod_{j\in T}\mathbf{1}(x_j\geqslant t_j)\,d\mu_T(t_j,j\in T);\prod_{j\in T}[u_j,v_j]\Big)\\&=\sup_{u_j<v_j,j\in T}\mathrm{Vit}\bigg((x_j,j\in T)\mapsto\mu_T\Big(\prod_{j\in T}(u_j,x_j]\Big);\prod_{j\in T}[u_j,v_j]\bigg).\end{aligned}$$

此外，由定理 7，

$$\begin{aligned}&\mathrm{Vit}\bigg((x_j,j\in T)\mapsto\mu_T\Big(\prod_{j\in T}(u_j,x_j]\Big);\prod_{j\in T}[u_j,v_j]\bigg)\\&=\mathrm{HK}_{(u_j,j\in T)}\bigg((x_j,j\in T)\mapsto\mu_T\Big(\prod_{j\in T}(u_j,x_j]\Big);\prod_{j\in T}[u_j,v_j]\bigg)=|\mu_T|\Big(\prod_{j\in T}(u_j,v_j]\Big).\end{aligned}$$

这里，第一个等式成立是因为该映射在每个包含锚点 $(u_j,\ j\in T)$ 的截面上都为零；也就是说，当某个 $j\in T$ 使得 $x_j=u_j$ 时，它就变为零。因此，

$$\mathrm{Vit}(f^{T}_{(a_j,j\in T^c)})=|\mu_T|(\mathbb{R}^{|T|})=\|\mu_T\|_{\mathrm{TV}}.$$

由于

$$\mathrm{Vit}(f^{T}_{(a_j,j\in T^c)})=0$$

对所有 $|T|>s$，可得

$$\mathrm{HK}_{\mathbf{a}}(f)=\sum_{0<|T|\leqslant d}\mathrm{Vit}(f^{T}_{(a_j,j\in T^c)})=\sum_{0<|T|\leqslant s}\mathrm{Vit}(f^{T}_{(a_j,j\in T^c)})=\sum_{0<|T|\leqslant s}\|\mu_T\|_{\mathrm{TV}}.$$

因此，实际上无需在（44）中取下确界，并且

$$\mathrm{HK}_{\mathbf{a}}(f)=V_{\mathbf{a}}(f),$$

这就完成了（15）的证明。

现在我们来考察（15）中不等式的紧性。固定一个 $\mathbf{a}=(-\infty,\ldots,\ -\infty)$。首先注意，对于函数 $f(x_1,\ldots,\ x_d)=\mathbf{1}(x_1\geqslant 0)$，我们有

$$V^{d,s}_{\infty-\mathrm{XGB}}(f)=1=\mathrm{HK}_{\mathbf{a}}(f).$$

这表明（15）中的右侧不等式是紧的。

为了证明（15）中的左侧不等式也同样是紧的，我们根据 $s \leqslant d$or $s > d$ 分两种情况讨论。在 $s \leqslant d$ 的情况下，考虑函数 $f(x_1, \ldots, x_d) = \mathbf{1}(x_1, \ldots, x_s < 0)$。显然，$V^{d,\ \infty-}{}^{s}\mathrm{XGB}(f) = 1$。此外，由于

$$f(x_1, \ldots, x_d) = \prod_{j=1}^{s} \big(1 - \mathbf{1}(x_j \geqslant 0)\big) = 1 + \sum_{l=1}^{s} (-1)^l \sum_{1 \leqslant j_1 < \cdots < j_l \leqslant s} \mathbf{1}(x_{j_1} \geqslant 0, \ldots, x_{j_l} \geqslant 0),$$

由此可得

$$\mathrm{HK}_{\mathbf{a}}(f) = 2^s - 1.$$

这表明，当 $s \leqslant d$ 时，(15) 中的左侧不等式是紧的。在 $s > d$ 的情况下，考虑函数$f(x_1$、.....、$x_d) = \mathbf{1}(-1 \leqslant x_1$、.....、$x_{s-d} < 0$、$x_{s-d+1}$、.....、$x_d < 0)$。同样显然，$V^{d}$、$\infty-{}^{s}\mathrm{XGB}(f) = 1$。此外，我们可以写成

$$f(x_1, \ldots, x_d) = \prod_{j=1}^{s-d} \big(\mathbf{1}(x_j \geqslant -1) - \mathbf{1}(x_j \geqslant 0)\big) \cdot \prod_{j=s-d+1}^{d} \big(1 - \mathbf{1}(x_j \geqslant 0)\big),$$

由此我们得到

$$\mathrm{HK}_{\mathbf{a}}(f) = 2^d.$$

这表明，当 $s > d$ 时，(15) 中的左侧不等式是紧的。 □

A.1.6 命题 2 的证明

命题 *2* 的证明。首先，(a) 可由每个基函数 b^{L}、U l、$\mathbf{u}$ 的右连续性、控制收敛定理，以及每个符号测度 ν_{L}、U 均为有限测度这一事实推出。

对于 (b) 和 (c)，回想一下 $f^{d}{}_{,b}{}^{s}{}_{,\{\mu_S\}}$ 的另一种表示形式 $f^{d}{}_{,c}{}^{s}{}_{,\{\nu_{L,U}\}}$，该表示形式见式 (41)，并在命题 5 的证明中引入。由于这一表示形式中的求和遍历所有满足 $0 < |S| \leqslant s$ 的 $S \subseteq [d]$，因此当 s 增大时，函数类 $\mathcal{F}^{d}{}_{,\infty-}{}^{s}\mathrm{ST}$ 也随之扩大。这就得到了 (b)。由于 $|S|$ 始终有界于 d，函数类$\mathcal{F}^{d,\infty-}{}^{s}\mathrm{ST}$ 在 $s \geqslant d$ 之后保持不变，这证明了 (c)。

最后，(d) 由定义立即得出。 □

A.1.7 命题 3 的证明

命题 *3* 的证明。 **步骤 1：** $f \in \mathcal{F}^{d}{}_{,\infty-}{}^{d}\mathbf{ST} \Rightarrow f$ **是右连续的，并且对所有** $\mathbf{a} \in\{-\infty, +\infty\}^d$**，** $\mathbf{HK}_{\mathbf{a}}(f) < \infty$ 。

这直接由命题 2 和命题 5 得出。

步骤 2： $f \in \mathcal{F}^{d,\ \infty-}{}^{s}\mathbf{ST} \Rightarrow (11)$ **对所有满足** $|S| > s$ **的** $S \subseteq [d]$ **都成立**。

我们只考虑 $\mathbf{a} = (-\infty, \cdots, -\infty)$ 这一情形；对于其他锚点的选择，论证完全类似。假设 $f \in \mathcal{F}^{d,\ \infty-}{}^{s}\mathrm{ST}$。回顾命题 5 的证明，$f$ 可表示为式 (41) 的另一种形式 $f \equiv f^{d,\ b}{}^{s}{}_{,\ \{\mu_S\}}$，其中 $b \in \mathbb{R}$ 是某个对象，且 μ_S 是定义在 $\mathbb{R}^{|S|}$ 上的有限符号博雷尔测度。

使用 $|T| > s$ 修复 $T \subseteq [d]$。借助这种表示法，我们可以将 $f^{T}_{(a_j,\ j \in T^c)}$ 表示为

$$f^{T}_{(a_j, j \in T^c)}(x_j, j \in T) = b + \sum_{\substack{S: \varnothing \neq S \subseteq T \\ |S| \leqslant s}} \int_{\mathbb{R}^{|S|}} \prod_{j \in S} \mathbf{1}(x_j \geqslant t_j)\, d\mu_S(t_j, j \in S).$$

由此可得

$$\sum_{\boldsymbol{\delta}\in\{0,1\}^{|T|}} (-1)^{\sum_{j\in T}\delta_j}\cdot f^T_{(a_j,j\in T^c)}\big((1-\delta_j)w_j+\delta_j v_j, j\in T\big)$$
$$= \sum_{\substack{S:\varnothing\neq S\subseteq T\\ |S|\leqslant s}} \sum_{\boldsymbol{\delta}_S\in\{0,1\}^{|S|}} \Bigg(\sum_{\boldsymbol{\delta}_{T\setminus S}\in\{0,1\}^{|T\setminus S|}} (-1)^{\sum_{j\in T\setminus S}\delta_j}\Bigg)\cdot(-1)^{\sum_{j\in S}\delta_j}\int_{\mathbb{R}^{|S|}}\prod_{j\in S}\mathbf{1}(x_j\geqslant t_j)\,d\mu_S(t_j,j\in S),$$

其中 $\boldsymbol{\delta}_S=(\delta_j,\ j\in S)$ 且 $\boldsymbol{\delta}_{T\setminus S}=(\delta_j,\ j\in T\setminus S)$。在最后一个表达式中，由于 $|S|\leqslant s<|T|$，最内层的求和总是为零。这证明了 **(11)** 对于 T 成立。

步骤 3： f **是右连续的，并且** $\mathbf{HK_a}(f)<\infty$ **对某个** $\mathbf{a}\in\{-\infty,\ +\infty\}^d \Rightarrow f\in\mathcal{F}^d_{,\ \infty-\mathbf{ST}}$。

假设 $f\colon\ \mathbb{R}^d\to\mathbb{R}$ 是右连续的，并且 $\mathrm{HK}_{\mathbf{a}}(f)<\infty$ 对某个 $\mathbf{a}\in\{-\infty,\ +\infty\}^d$ 成立。固定一个非空的 $S\subseteq[d]$，并对每个整数 $N\geqslant 1$，定义 $g^S_N\colon\ \mathbb{R}^{|S|}\to\mathbb{R}$ 为

$$g^S_N(x_j,j\in S)=\sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}}(-1)^{\sum_{j\in S}\delta_j}\cdot f^S_{(a_j,j\in S^c)}\big((1-\delta_j)x_j+\delta_j(-N),j\in S\big).$$

显然，g^S_N 继承了 f 在坐标 $j\in S$ 上的右连续性。此外，

$$\mathrm{HK}_{(-N,j\in S)}\big(g^S_N;[-N,N]^{|S|}\big)=\mathrm{Vit}\big(g^S_N;[-N,N]^{|S|}\big)=\mathrm{Vit}\big(f^S_{(a_j,j\in S^c)};[-N,N]^{|S|}\big)<\infty.$$

这里，第一个等式是由以下事实得出的：当某个 $j\in S$ 满足 $x_j=-N$ 时，g^S_N 为零。因此，根据定理 **7**，存在唯一的有限符号博雷尔测度 ν^S_N，定义在 $[-N,\ N]^{|S|}$ 上，使得

$$g^S_N(x_j,j\in S)=\nu^S_N\Big(\prod_{j\in S}(-N,x_j]\Big)\quad\text{for }(x_j,j\in S)\in[-N,N]^{|S|}.$$

这里，由于对于某个$j\in S$，g^S_N 会变为零，因此可以将端点 $-N$ 从区间中排除。此外，定理 **7** 还给出

$$|\nu^S_N|([-N,N]^{|S|})=\mathrm{Vit}\big(f^S_{(a_j,j\in S^c)};[-N,N]^{|S|}\big)\leqslant\mathrm{Vit}\big(f^S_{(a_j,j\in S^c)}\big)<\infty.$$

现在，固定整数 $N_2>N_1\geqslant 1$，并注意到

$$g^S_{N_1}(x_j,j\in S)=\sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}}(-1)^{\sum_{j\in S}\delta_j}\cdot g^S_{N_2}\big((1-\delta_j)x_j+\delta_j(-N_1),j\in S\big)=\nu^S_{N_2}\Big(\prod_{j\in S}(-N_1,x_j]\Big)$$

对于每个 $(x_j,j\in S)\in[-N_1,N_1]^{|S|}$。由 $\nu^S_{N_1}$的唯一性可知，这意味着 $\nu^S_{N_2}$ 在$[-N_1,N_1]^{|S|}$ 上的限制与 $\nu^S_{N_1}$ 一致。因此，$\{\nu^S_N\}_{N\geqslant 1}$ 构成一个有限符号博雷尔测度序列，其中当 $N_2>N_1$时，$\nu^S_{N_2}$ 是 $\nu^S_{N_1}$ 的扩展。

利用这个有符号测度序列，我们定义一个有限符号博雷尔测度 ν_S，它定义在 $\mathbb{R}^{|S|}$ 上，并对所有 $N\geqslant 1$ 扩展 ν^S_N。具体而言，我们将 ν_S 定义为

$$\nu_S(E)=\lim_{N\to\infty}\nu^S_N(E\cap[-N,N]^{|S|})\quad\text{for Borel sets }E\subseteq\mathbb{R}^{|S|}.$$

我们首先验证，对于每个博雷尔集 E，$\nu_S(E)$ 都是良定义的。对于整数 $M>N\geqslant 1$，

$$\begin{aligned}\big|\nu^S_M(E\cap[-M,M]^{|S|})-\nu^S_N(E\cap[-N,N]^{|S|})\big|&=\big|\nu^S_M\big(E\cap([-M,M]^{|S|}\setminus[-N,N]^{|S|})\big)\big|\\&\leqslant|\nu^S_M|\big([-M,M]^{|S|}\setminus[-N,N]^{|S|}\big).\end{aligned}$$

由于

$$\sup_{N\geqslant 1}|\nu^S_N|([-N,N]^{|S|})\leqslant\mathrm{Vit}\big(f^S_{(a_j,j\in S^c)}\big)<\infty,$$

对于每个 $\epsilon>0$，都存在一个整数 $N_0\geqslant 1$，使得

$$|\nu_M^S|\big([-M,M]^{|S|}\backslash[-N,N]^{|S|}\big)<\epsilon\quad\text{for all } M>N\geqslant N_0.$$

因此，$\{\nu_N^S(E\cap[-N,\ N]^{|S|})\}_{N\geqslant 1}$构成一个柯西序列，因此，$\nu_S(E)$是定义良好的　　.

接下来，我们证明 ν_S 是可数可加的。设 $E=\cup_{k\geqslant 1}E_k$，其中 E_k为互不相交的博雷尔集。对于整数 $N_2>N_1\geqslant 1$，由于 $\nu_{N_2}^S$ 扩张了 $\nu_{N_1}^S$，因此将 $|\nu_{N_2}^S|$（$\nu_{N_2}^S$的全变差）限制在 $[-N_1,\ N_1]^{|S|}$上也与 $|\nu_{N_1}^S|$一致。因此，对于每个 $k\geqslant 1$,

$$\big\{|\nu_N^S|(E_k\cap[-N,N]^{|S|})\big\}_{N\geqslant 1}$$

是一个非负数递增序列。根据单调收敛定理，我们有　　ve

$$\begin{aligned}\sum_{k\geqslant 1}\lim_{N\to\infty}|\nu_N^S|(E_k\cap[-N,N]^{|S|})&=\lim_{N\to\infty}\sum_{k\geqslant 1}|\nu_N^S|(E_k\cap[-N,N]^{|S|})\\&=\lim_{N\to\infty}|\nu_N^S|(E\cap[-N,N]^{|S|})\leqslant\sup_{N\geqslant 1}|\nu_N^S|([-N,N]^{|S|})\leqslant\mathrm{Vit}\big(f^S_{(a_j,j\in S^c)}\big)<\infty.\end{aligned}$$

此外，由于

$$\big|\nu_N^S\big(E_k\cap[-N,N]^{|S|}\big)\big|\leqslant|\nu_N^S|\big(E_k\cap[-N,N]^{|S|}\big)\leqslant\lim_{N\to\infty}|\nu_N^S|\big(E_k\cap[-N,N]^{|S|}\big)$$

对于每个 k 和 N，由控制收敛定理可得

$$\begin{aligned}\nu_S(E)&=\lim_{N\to\infty}\nu_N^S(E\cap[-N,N]^{|S|})=\lim_{N\to\infty}\sum_{k\geqslant 1}\nu_N^S(E_k\cap[-N,N]^{|S|})\\&=\sum_{k\geqslant 1}\lim_{N\to\infty}\nu_N^S(E_k\cap[-N,N]^{|S|})=\sum_{k\geqslant 1}\nu_S(E_k).\end{aligned}$$

这就证明了 ν_S 是可数可加的。

对于每个 $N\geqslant 1$，由 ν_S的定义可知，对于任意博雷尔集 $E\subseteq[-N,\ N]^{|S|}$，都有$\nu_S(E)=\nu_N^S(E)$。此外，

$$|\nu_S|(\mathbb{R}^{|S|})=\lim_{N\to\infty}|\nu_S|([-N,N]^{|S|})=\lim_{N\to\infty}|\nu_N^S|([-N,N]^{|S|})\leqslant\mathrm{Vit}\big(f^S_{(a_j,j\in S^c)}\big)<\infty.$$

He因此，ν_S 是 $\mathbb{R}^{|S|}$ 上的有限符号博雷尔测度，并且对所有 $N\geqslant 1$都扩张了 ν_N^S，如所需。

定义 $N_{\mathbf{a}}=\{j\in[d]\colon\ a_j=-\infty\}$。对于每个整数 $N\geqslant 1$，有

$$\begin{aligned}&\sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}}(-1)^{\sum_{j\in S}\delta_j}\cdot f^S_{(a_j,j\in S^c)}\big((1-\delta_j)x_j+\delta_j(-N),j\in S\cap N_{\mathbf{a}};(1-\delta_j)x_j+\delta_jN,j\in S\backslash N_{\mathbf{a}}\big)\\&=\sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}}(-1)^{\sum_{j\in S}\delta_j}\cdot g^S_N\big((1-\delta_j)x_j+\delta_j(-N),j\in S\cap N_{\mathbf{a}};(1-\delta_j)x_j+\delta_jN,j\in S\backslash N_{\mathbf{a}}\big)\\&=(-1)^{|S\backslash N_{\mathbf{a}}|}\cdot\nu_S\Big(\prod_{j\in S\cap N_{\mathbf{a}}}(-N,x_j]\times\prod_{j\in S\backslash N_{\mathbf{a}}}(x_j,N]\Big).\end{aligned}$$

令 $N\to\infty$ 趋于极限，得到

$$\sum_{R\subseteq S}(-1)^{|S|-|R|}\cdot f^R_{(a_j,j\in R^c)}(x_j,j\in R)=\widetilde{\nu}_S\Big(\prod_{j\in S\cap N_{\mathbf{a}}}(-\infty,x_j]\times\prod_{j\in S\backslash N_{\mathbf{a}}}(x_j,+\infty)\Big)$$

其中，$\widetilde{\nu}_S$ 是由 $\widetilde{\nu}_S=(-1)^{|S\backslash N_{\mathbf{a}}|}\cdot\nu_S$定义在 $\mathbb{R}^{|S|}$ 上的有符号博雷尔测度。此外，由于

$$f(x_1,\ldots,x_d)=\lim_{\mathbf{z}\to\mathbf{a}}f(\mathbf{z})+\sum_{S:\varnothing\neq S\subseteq[d]}\sum_{R\subseteq S}(-1)^{|S|-|R|}\cdot f^R_{(a_j,j\in R^c)}(x_j,j\in R)\quad\text{for }(x_1,\ldots,x_d)\in\mathbb{R}^d,$$

由此可得

$$f(x_1,\dots,x_d)=\lim_{\mathbf{z}\to\mathbf{a}}f(\mathbf{z})+\sum_{S:\varnothing\neq S\subseteq[d]}\widetilde{\nu}_S\Big(\prod_{j\in S\cap N_{\mathbf{a}}}(-\infty,x_j]\times\prod_{j\in S\setminus N_{\mathbf{a}}}(x_j,+\infty)\Big)$$
$$=\lim_{\mathbf{z}\to\mathbf{a}}f(\mathbf{z})+\sum_{S:\varnothing\neq S\subseteq[d]}\int_{\mathbb{R}^{|S|}}\Big(\prod_{j\in S\cap N_{\mathbf{a}}}\mathbf{1}(x_j\geqslant t_j)\Big)\cdot\Big(\prod_{j\in S\setminus N_{\mathbf{a}}}\mathbf{1}(x_j<t_j)\Big)\,d\widetilde{\nu}_S(t_j,j\in S)$$

对于所有 $(x_1,\dots,x_d)\in\mathbb{R}^d$。这证明了 $f\in\mathcal{F}^{d,\overset{d}{\infty-\mathrm{ST}}}$。

步骤 4： $f\in\mathcal{F}^{d,\overset{d}{\infty-\mathbf{ST}}}$ **并且** (11) **对所有满足** $|S|>s\Rightarrow f\in\mathcal{F}^{d,\overset{s}{\infty-\mathbf{ST}}}$ **的** $S\subseteq[d]$ **都成立**。

现在，我们假设 f 还满足条件 **(11)**，即对所有满足 $|S|>s$的 $S\subseteq[d]$ 都成立。由于当 $s\geqslant d$ 时附加条件显然成立，我们假设 $s<d$。这里我们仅展示 $\mathbf{a}=(-\infty,\cdots,-\infty)$ 这一情形的论证；其他锚点的证明完全类似。

由于 $f\in\mathcal{F}^{d,\overset{d}{\infty-\mathrm{ST}}}$，$f$ 可表示为某个 $b\in\mathbb{R}$ 与定义在 $\mathbb{R}^{|S|}$ 上的有限符号博雷尔测度 μ_S 的替代表达式 **(41)**。对于每个满足 $|S|>s$ 的 $S\subseteq[d]$，我们有

$$\sum_{\boldsymbol{\delta}\in\{0,1\}^{|S|}}(-1)^{\sum_{j\in S}\delta_j}\cdot f^S_{(a_j,j\in S^c)}\big((1-\delta_j)v_j+\delta_j u_j,j\in S\big)=\mu_S\Big(\prod_{j\in S}(u_j,v_j]\Big)$$

对于所有 $u_j<v_j$，$j\in S$。因此，条件 **(11)** 表明，对于所有此类 S 以及所有 $u_j<v_j$，$j\in S$，

$$\mu_S\Big(\prod_{j\in S}(u_j,v_j]\Big)=0.$$

根据 Dynkin 的 π-λ 定理，可知 $\mu_S=0$。因此，可以从 **(41)** 中删去所有关于 μ_S 且满足 $|S|>s$ 的积分，从而得出 $f\in\mathcal{F}^{d,\overset{s}{\infty-\mathrm{ST}}}$。 □

A.1.8 命题 6 的证明

命题 *6* 的证明。固定 $j_0\in[d]$ 和 $t_{j_0}\in\mathbb{R}$。按照命题陈述中给出的方式定义 $g:\mathbb{R}^d\to\mathbb{R}$，其中 $j=j_0$。由对称性，只需证明 $V^{d,\overset{s}{\infty-\mathrm{XGB}}}(g)\leqslant V^{d,\overset{s}{\infty-\mathrm{XGB}}}(f)$。

假设 $f\equiv f^{d,\overset{s}{c},\{\nu_{L,U}\}}$。对于每个满足 $0<|L|+|U|\leqslant s$ 的 L，$U\subseteq[d]$，我们在 $\mathbb{R}^{|L|+|U|}$ 上定义一个有符号博雷尔测度 $\mu_{L,U}$，如下所示。若 $j_0\notin L\cup U$，则令 $\mu_{L,U}=\nu_{L,U}$。若 $j_0\in L\backslash U$，则将 $\mu_{L,U}$ 定义为 $\nu_{L\backslash\{j_0\},U\cup\{j_0\}}$ 在该映射下的推前

$$\big((l_j,j\in L\backslash\{j_0\}),(u_j,j\in U\cup\{j_0\})\big)\mapsto\big((l_j,j\in L\backslash\{j_0\};t_{j_0}-u_{j_0}),(u_j,j\in U)\big),$$

若 $j_0\in U\backslash L$，则定义 $\mu_{L,U}$ 为 $\nu_{L\cup\{j_0\},U\backslash\{j_0\}}$ 在该映射下的推前

$$\big((l_j,j\in L\cup\{j_0\}),(u_j,j\in U\backslash\{j_0\})\big)\mapsto\big((l_j,j\in L),(u_j,j\in U\backslash\{j_0\};t_{j_0}-l_{j_0})\big),$$

且若 $j_0\in L\cap U$，则定义 $\mu_{L,U}$ 为 $\nu_{L,U}$ 在该映射下的推前

$$\big((l_j,j\in L),(u_j,j\in U)\big)\mapsto\big((l_j,j\in L\backslash\{j_0\};t_{j_0}-u_{j_0}),(u_j,j\in U\backslash\{j_0\};t_{j_0}-l_{j_0})\big).$$

根据这些定义，可以直接验证 $f^{d,\overset{s}{c},\{\mu_{L,U}\}}\equiv g$。此外，符号测度 $\mu_{L,U}$ 与符号测度 $\nu_{L,U}$之间存在一一对应关系，且对应的符号测度具有相同的全变差。因此，

$$V^{d,s}_{\infty-\mathrm{XGB}}(g)\leqslant\sum_{0<|L|+|U|\leqslant s}\|\mu_{L,U}\|_{\mathrm{TV}}$$

$$
\begin{aligned}
&= \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\notin L\cup U}} \|\mu_{L,U}\|_{\mathrm{TV}} + \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\in L\setminus U}} \|\mu_{L,U}\|_{\mathrm{TV}} + \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\in U\setminus L}} \|\mu_{L,U}\|_{\mathrm{TV}} + \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\in L\cap U}} \|\mu_{L,U}\|_{\mathrm{TV}} \\
&= \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\notin L\cup U}} \|\nu_{L,U}\|_{\mathrm{TV}} + \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\in L\setminus U}} \|\nu_{L\setminus\{j_0\},U\cup\{j_0\}}\|_{\mathrm{TV}} \\
&\quad + \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\in U\setminus L}} \|\nu_{L\cup\{j_0\},U\setminus\{j_0\}}\|_{\mathrm{TV}} + \sum_{\substack{0<|L|+|U|\leqslant s\\ j_0\in L\cap U}} \|\nu_{L,U}\|_{\mathrm{TV}} = \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}}.
\end{aligned}
$$

Ta对所有表示取下确界 $f^{d,\ c}{}^{s,\ \{\nu_{L,\ U}\}}$，得出 f，可得 $V^{d,\ \infty-}{}^{s}_{\mathrm{XGB}}(g) \leqslant V^{d,\ \infty-}{}^{s}_{\mathrm{XGB}}(f\quad)$. □

A.2 第4节中定理和引理的证明

A.2.1 定理 3 的证明

由于定理的后半部分可直接由引理 1 得出，因此我们这里只证明关于存在性的前半部分。定理 2 的证明完全类似。

定理 *3* 的证明。 当有符号测度 $\nu_{L,\ U}$ 满足引理 1 的条件 (a) 时，相应的函数 $f^{d,\ c}{}^{s,\ \{\nu_{L,\ U}\}}$ 可以写成

$$
f^{d,s}_{c,\{\nu_{L,U}\}}(x_1,\ldots,x_d) = c + \sum_{(L,U,\mathbf{p},\mathbf{q})\in J} \beta^{L,U}_{\mathbf{p},\mathbf{q}} \cdot \prod_{j\in L} \mathbf{1}\big(x_j \geqslant (v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2\big) \cdot \prod_{j\in U} \mathbf{1}\big(x_j < (v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2\big)
$$

其中

$$
J = \left\{(L,U,\mathbf{p},\mathbf{q}) : L,U\subseteq[d], 0<|L|+|U|\leqslant s, \mathbf{p}\in\prod_{j\in L}[n_j-1], \text{ and } \mathbf{q}\in\prod_{j\in U}[n_j-1]\right\} \tag{45}
$$

and

$$
\beta^{L,U}_{\mathbf{p},\mathbf{q}} = \nu_{L,U}\Big(\big\{\big((v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2, j\in L; (v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2, j\in U\big)\big\}\Big)
$$

对每个 $(L,\ U,\ \mathbf{p},\ \mathbf{q})\in J$，其中 $\mathbf{p} = (p_j,\ \ j\in L)$ 且 $\mathbf{q} = (q_j,\ \ j\in U)$。

令 $(\hat{c}\ (\hat{}_{\beta^{L,\ \mathbf{p}\ \mathbf{q}}_{\cdot},\ U},\ \ (L,\ \ U,\ \ \mathbf{p},\ \ \mathbf{q})\in J))$ 为有限维优化问题的一个解

$$
\begin{aligned}
&\operatorname{argmin} \sum_{i=1}^{n} \left(y_i - c - \sum_{(L,U,\mathbf{p},\mathbf{q})\in J} \beta^{L,U}_{\mathbf{p},\mathbf{q}} \cdot \prod_{j\in L} \mathbf{1}\big(x^{(i)}_j \geqslant (v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2\big) \cdot \prod_{j\in U} \mathbf{1}\big(x^{(i)}_j < (v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2\big)\right)^2 \\
&\quad \text{s.t.} \sum_{(L,U,\mathbf{p},\mathbf{q})\in J} |\beta^{L,U}_{\mathbf{p},\mathbf{q}}| \leqslant V.
\end{aligned}
$$

此类解的存在性是显然的。定义 $\hat{}f^{d,\ n}{}^{s}_{V,}:\ \mathbb{R}^d\to\mathbb{R}$ 如下

$$
\hat{f}^{d,s}_{n,V}(x_1,\ldots,x_d) = \hat{c} + \sum_{(L,U,\mathbf{p},\mathbf{q})\in J} \hat{\beta}^{L,U}_{\mathbf{p},\mathbf{q}} \cdot \prod_{j\in L} \mathbf{1}\big(x_j \geqslant (v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2\big) \cdot \prod_{j\in U} \mathbf{1}\big(x_j < (v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2\big).
$$

根据构造，$\hat{}f^{d,\ n}{}^{s}_{,\ V}\in\mathcal{F}^{d}_{,\ s\ \mathrm{ST}\bullet}$，它是问题 (16) 的一个解。此外，引理 1 表明它也是问题 (3) 的一个解。

□

A.2.2 引理 1 的证明

我们在证明中使用以下引理。该引理紧接在引理 1 的证明之后给出。

引理 6. 对于每个 $f^{d,s}_{c,\{\nu_{L,U}\}} \in \mathcal{F}^{d,s}_{\infty-ST}$，存在 $f^{d,s}_{b,\{\mu_{L,U}\}} \in \mathcal{F}^{d,s}_{ST}$，其离散有符号测度 $\mu_{L,U}$ 支撑在格点 (17) 上，使得

(a) $f^{d,s}_{b,\{\mu_{L,U}\}}(\mathbf{x}^{(i)}) = f^{d,s}_{c,\{\nu_{L,U}\}}(\mathbf{x}^{(i)})$ *for* $i = 1, \ldots, n$

(b)

$$\sum_{0<|L|+|U|\leqslant s} \|\mu_{L,U}\|_{TV} \leqslant \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{TV}.$$

引理 *1* 的证明。对于 $z_1, \ldots, z_n \in \mathbb{R}$，定义

$$V^{d,s}_{\infty-\mathrm{XGB}}(z_1, \ldots, z_n) = \inf\left\{ \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} : f^{d,s}_{c,\{\nu_{L,U}\}}(\mathbf{x}^{(i)}) = z_i \text{ for } i = 1, \ldots, n \right\}.$$

为简 便起见，我们省略对设计点 $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ 的依赖。根据定义 ，

$$V^{d,s}_{\infty-\mathrm{XGB}}(f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(n)})) \leqslant V^{d,s}_{\infty-\mathrm{XGB}}(f) \quad \text{for every } f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}.$$

根据引理 6，对于每个 $z_1, \ldots, z_n$，我们有

$$V^{d,s}_{\infty-\mathrm{XGB}}(z_1, \ldots, z_n) = \inf\left\{ \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} : f^{d,s}_{c,\{\nu_{L,U}\}}(\mathbf{x}^{(i)}) = z_i \text{ for } i = 1, \ldots, n, \right.$$
$$\left. \text{with } \nu_{L,U} \text{ supported on the lattices (17)} \right\}.$$

回顾式（45）中的指标集 J，并用$\mathbf{x}$ 表示其元素为

$$X_{i,(L,U,\mathbf{p},\mathbf{q})} = \prod_{j\in L} \mathbf{1}\big(x^{(i)}_j \geqslant (v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2\big) \cdot \prod_{j\in U} \mathbf{1}\big(x^{(i)}_j < (v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2\big)$$

其中 $i = 1, \ldots, n$，且 $(L, U, \mathbf{p}, \mathbf{q}) \in J$。我们将支撑在格点（17）上的离散有符号测度 $\nu_{L,U}$ 参数化为

$$\nu_{L,U}\big(\{((v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2, j \in L) \times ((v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2, j \in U)\}\big) = \beta^{L,U}_{\mathbf{p},\mathbf{q}} \tag{46}$$

其中 $L, U \subseteq [d]$，且 $0 < |L| + |U| \leqslant s$，$\mathbf{p} = (p_j, j \in L) \in \prod_{j\in L}[n_j - 1]$，以及$\mathbf{q} = (q_j, j \in U) \in \prod_{j\in U}[n_j - 1]$。利用此参数化，我们可以将 $V^{d,s}_{\infty-\mathrm{XGB}}(z_1, \ldots, z_n)$ 表示为

$$V^{d,s}_{\infty-\mathrm{XGB}}(z_1, \ldots, z_n) = \inf\big\{\|\boldsymbol{\beta}\|_1 : \mathbf{x}\boldsymbol{\beta} = \mathbf{z} - c\mathbf{1} \text{ for some } c \in \mathbb{R}\big\}, \tag{47}$$

其中$\mathbf{z} = (z_1, \ldots, z_n)$，且$\mathbf{1}$ 是全 $\mathbf{1}$ 向量。

接下来，我们证明若集合

$$\mathcal{D}_{\mathbf{z}} := \big\{\boldsymbol{\beta} \in \mathbb{R}^{|J|} : \mathbf{x}\boldsymbol{\beta} = \mathbf{z} - c\mathbf{1} \text{ for some } c \in \mathbb{R}\big\} \tag{48}$$

该集合非空；显然，当$\mathbf{z} = (f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(n)}))$ 对于某个 $f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$时，情况确实如此。此时，存在一个 $\hat{\boldsymbol{\beta}}$，使其在 (47) 中达到最小值。固定一个 $\mathbf{z} \in \mathbb{R}^n$，使得 $\mathcal{D}_{\mathbf{z}}$ 非空，并假设 $\mathbf{x}\boldsymbol{\beta}_0 = \mathbf{z} - c_0\mathbf{1}$ 对于某个 $\boldsymbol{\beta}_0$ 和 c_0。显然，若进一步限制 $\boldsymbol{\beta}$ 满足 $\|\boldsymbol{\beta}\|_1 \leqslant \|\boldsymbol{\beta}_0\|_1$，则 (47) 右侧的下确界保持不变：

$$V^{d,s}_{\infty-\mathrm{XGB}}(z_1, \ldots, z_n) = \inf\big\{\|\boldsymbol{\beta}\|_1 : \mathbf{x}\boldsymbol{\beta} = \mathbf{z} - c\mathbf{1} \text{ for some } c \in \mathbb{R} \text{ and } \|\boldsymbol{\beta}\|_1 \leqslant \|\boldsymbol{\beta}_0\|_1\big\}.$$

也很容易验证该集合

$$\mathcal{D}_{\mathbf{z}} \cap \{\boldsymbol{\beta} \in \mathbb{R}^{|J|} : \|\boldsymbol{\beta}\|_1 \leqslant \|\hat{\boldsymbol{\beta}}_0\|_1\}$$

非空、闭且有界。由于映射 $\boldsymbol{\beta} \mapsto \|\boldsymbol{\beta}\|_1$ 连续，因此存在一个 $\hat{\boldsymbol{\beta}}$，使其在 (47) 中取得最小值。

利用上述结果，我们现在证明该引理。固定 $f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$，并令 $\hat{\boldsymbol{\beta}}$ 为针对 $\mathbf{z} = (f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(n)}))$ 的 (47) 的极小化解。令 $\hat{c}$ 为相应的 (48) 中的常数，并令 $\nu_{L,U}$ 表示通过 (46) 与 $\hat{\boldsymbol{\beta}}$ 相关联的有符号博雷尔测度。根据构造，$f^{d,s}_{\hat{c},\{\nu_{L,U}\}} \in \mathcal{F}^{d,s}_{\mathrm{ST}}$ 满足引理的前两个条件。此外，由于

$$\begin{aligned} V^{d,s}_{\infty-\mathrm{XGB}}(f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(n)})) &= \|\hat{\boldsymbol{\beta}}\|_1 = \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} \\ &\geqslant V^{d,s}_{\infty-\mathrm{XGB}}(f^{d,s}_{\hat{c},\{\nu_{L,U}\}}) \geqslant V^{d,s}_{\infty-\mathrm{XGB}}(f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(n)})), \end{aligned}$$

我们有

$$V^{d,s}_{\infty-\mathrm{XGB}}(f^{d,s}_{\hat{c},\{\nu_{L,U}\}}) = \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} = V^{d,s}_{\infty-\mathrm{XGB}}(f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(n)})) \leqslant V^{d,s}_{\infty-\mathrm{XGB}}(f).$$

因此，$f^{d,s}_{\hat{c},\{\nu_{L,U}\}} \in \mathcal{F}^{d,s}_{\mathrm{ST}}$ 是一个满足所有所需性质的函数。 □

引理 6 的证明。对于每个 $j \in [d]$，定义

$$I^{(j)}_{m_j} = (v^{(j)}_{m_j}, v^{(j)}_{m_j+1}] \qquad \text{for } m_j \in [n_j - 1],$$

and

$$\underline{I}^{(j)}_{m_j} = (v^{(j)}_1, v^{(j)}_{m_j}] \text{ and } \overline{I}^{(j)}_{m_j} = (v^{(j)}_{m_j}, v^{(j)}_{n_j}] \qquad \text{for } m_j \in [n_j].$$

此外，令

$$\underline{O}^{(j)} = (-\infty, v^{(j)}_1] \text{ and } \overline{O}^{(j)} = (v^{(j)}_{n_j}, +\infty).$$

使用这些记号，我们定义离散有符号测度 $\mu_{L,U}$ 和常数 b 如下。对于 $L, U \subseteq [d]$且满足 $0 < |L| + |U| \leqslant s$的情形，令 $\mu_{L,U}$ 为支撑在格点 (17) 上的离散有符号测度，其定义为

$$\begin{aligned} &\mu_{L,U}\big(\{((v^{(j)}_{p_j} + v^{(j)}_{p_j+1})/2, j \in L) \times ((v^{(j)}_{q_j} + v^{(j)}_{q_j+1})/2, j \in U)\}\big) \\ &\quad = \sum_{\substack{\tilde{L},\tilde{U}:\tilde{L}\supseteq L,\tilde{U}\supseteq U \\ |\tilde{L}|+|\tilde{U}|\leqslant s}} \nu_{\tilde{L},\tilde{U}}\Big(\prod_{j\in L} I^{(j)}_{p_j} \times \prod_{j\in\tilde{L}\setminus L} \underline{O}^{(j)} \times \prod_{j\in U} I^{(j)}_{q_j} \times \prod_{j\in\tilde{U}\setminus U} \overline{O}^{(j)}\Big) \end{aligned}$$

对于 $(p_j, j \in L) \in \prod_{j\in L}[n_j - 1]$ 和 $(q_j, j \in U) \in \prod_{j\in U}[n_j - 1]$。此外，令

$$b = c + \sum_{0<|L|+|U|\leqslant s} \nu_{L,U}\Big(\prod_{j\in L} \underline{O}^{(j)} \times \prod_{j\in U} \overline{O}^{(j)}\Big).$$

根据构造，对于每个 $(m_1, \ldots, m_d) \in \prod_{j=1}^d [n_j]$，我们有

$$\begin{aligned} &\int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}\big(v^{(j)}_{m_j} \geqslant l_j\big) \cdot \prod_{j\in U} \mathbf{1}\big(v^{(j)}_{m_j} < u_j\big)\, d\mu_{L,U}(\mathbf{l}, \mathbf{u}) \\ &\quad = \sum_{\mathbf{r}\in \prod_{j\in L}\{1,\ldots,m_j-1\}\times\prod_{j\in U}\{m_j,\ldots,n_j-1\}} \mu_{L,U}\big(\{((v^{(j)}_{r_j} + v^{(j)}_{r_j+1})/2, j \in L) \times ((v^{(j)}_{r_j} + v^{(j)}_{r_j+1})/2, j \in U)\}\big) \\ &\quad = \sum_{\substack{\tilde{L},\tilde{U}:\tilde{L}\supseteq L,\tilde{U}\supseteq U \\ |\tilde{L}|+|\tilde{U}|\leqslant s}} \nu_{\tilde{L},\tilde{U}}\Big(\prod_{j\in L} \underline{I}^{(j)}_{m_j} \times \prod_{j\in\tilde{L}\setminus L} \underline{O}^{(j)} \times \prod_{j\in U} \overline{I}^{(j)}_{m_j} \times \prod_{j\in\tilde{U}\setminus U} \overline{O}^{(j)}\Big). \end{aligned}$$

由此可知，对于每个 $(m_1,\dots,m_d)\in\prod_{j=1}^d[n_j]$，

$$
\begin{aligned}
f^{d,s}_{b,\{\mu_{L,U}\}}(v^{(1)}_{m_1},\dots,v^{(d)}_{m_d}) &= b+\sum_{0<|L|+|U|\leqslant s}\sum_{\substack{\tilde L,\tilde U:\tilde L\supseteq L,\tilde U\supseteq U\\ |\tilde L|+|\tilde U|\leqslant s}}\nu_{\tilde L,\tilde U}\Big(\prod_{j\in L}\underline{I}^{(j)}_{m_j}\times\prod_{j\in\tilde L\setminus L}\underline{O}^{(j)}\times\prod_{j\in U}\overline{I}^{(j)}_{m_j}\times\prod_{j\in\tilde U\setminus U}\overline{O}^{(j)}\Big)\\
&= c+\sum_{0<|\tilde L|+|\tilde U|\leqslant s}\sum_{L\subseteq\tilde L}\sum_{U\subseteq\tilde U}\nu_{\tilde L,\tilde U}\Big(\prod_{j\in L}\underline{I}^{(j)}_{m_j}\times\prod_{j\in\tilde L\setminus L}\underline{O}^{(j)}\times\prod_{j\in U}\overline{I}^{(j)}_{m_j}\times\prod_{j\in\tilde U\setminus U}\overline{O}^{(j)}\Big)\\
&= c+\sum_{0<|\tilde L|+|\tilde U|\leqslant s}\nu_{\tilde L,\tilde U}\Big(\prod_{j\in\tilde L}\big(\underline{I}^{(j)}_{m_j}\cup\underline{O}^{(j)}\big)\times\prod_{j\in\tilde U}\big(\overline{I}^{(j)}_{m_j}\cup\overline{O}^{(j)}\big)\Big)=f^{d,s}_{c,\{\nu_{L,U}\}}(v^{(1)}_{m_1},\dots,v^{(d)}_{m_d}),
\end{aligned}
$$

这意味着，在所有设计点 $\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}$ 处，$f^{d,s}_{,b,\{\mu_{L,U}\}}$ 与 $f^{d,s}_{,c,\{\nu_{L,U}\}}$ 一致。此外，

$$
\begin{aligned}
\sum_{0<|L|+|U|\leqslant s}|\mu_{L,U}|(\mathbb{R}^{|L|+|U|}) &\leqslant \sum_{0<|L|+|U|\leqslant s}\sum_{\mathbf{p}\in\prod_{j\in L}[n_j-1]}\sum_{\mathbf{q}\in\prod_{j\in U}[n_j-1]}\sum_{\substack{\tilde L,\tilde U:\tilde L\supseteq L,\tilde U\supseteq U\\ |\tilde L|+|\tilde U|\leqslant s}}\\
&\qquad |\nu_{\tilde L,\tilde U}|\Big(\prod_{j\in L}I^{(j)}_{p_j}\times\prod_{j\in\tilde L\setminus L}\underline{O}^{(j)}\times\prod_{j\in U}I^{(j)}_{q_j}\times\prod_{j\in\tilde U\setminus U}\overline{O}^{(j)}\Big)\\
&= \sum_{0<|L|+|U|\leqslant s}\sum_{\substack{\tilde L,\tilde U:\tilde L\supseteq L,\tilde U\supseteq U\\ |\tilde L|+|\tilde U|\leqslant s}}|\nu_{\tilde L,\tilde U}|\Big(\prod_{j\in L}\underline{I}^{(j)}_{n_j}\times\prod_{j\in\tilde L\setminus L}\underline{O}^{(j)}\times\prod_{j\in U}\overline{I}^{(j)}_{1}\times\prod_{j\in\tilde U\setminus U}\overline{O}^{(j)}\Big)\\
&\leqslant \sum_{0<|\tilde L|+|\tilde U|\leqslant s}\sum_{L\subseteq\tilde L}\sum_{U\subseteq\tilde U}|\nu_{\tilde L,\tilde U}|\Big(\prod_{j\in L}\underline{I}^{(j)}_{n_j}\times\prod_{j\in\tilde L\setminus L}\underline{O}^{(j)}\times\prod_{j\in U}\overline{I}^{(j)}_{1}\times\prod_{j\in\tilde U\setminus U}\overline{O}^{(j)}\Big)\\
&= \sum_{0<|\tilde L|+|\tilde U|\leqslant s}|\nu_{\tilde L,\tilde U}|\Big(\prod_{j\in\tilde L}\big(\underline{I}^{(j)}_{n_j}\cup\underline{O}^{(j)}\big)\times\prod_{j\in\tilde U}\big(\overline{I}^{(j)}_{1}\cup\overline{O}^{(j)}\big)\Big)\leqslant\sum_{0<|\tilde L|+|\tilde U|\leqslant s}|\nu_{\tilde L,\tilde U}|(\mathbb{R}^{|\tilde L|+|\tilde U|}).
\end{aligned}
$$

这证明了 $f^{d,s}_{,b,\{\mu_{L,U}\}}$ 是所需的函数，并满足引理的条件。 □

A.3 第5节中定理、引理和推论的证明

A.3.1 定理4的证明

我们将使用经验过程理论中的以下三个结果来证明该定理。定理 8 给出了乘子经验过程期望上确界的矩不等式。引理 7 根据括号熵积分界定带有拉德马赫乘子的经验过程的期望上确界。定理 9 则将控制一般乘子经验过程期望上确界的问题，简化为带有拉德马赫乘子的情形。定理 8 和定理 9 是一般性结果，而引理 7 则更贴合我们的研究设定。我们在附录A.5.1中给出引理 7 的证明。

定理 8 (Gin´等人的命题 3.1 [16])。 假设 $\mathcal{F}$ 是从 $\mathcal{X}$到 $\mathbb{R}$的可数函数族。假设 $\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}$ 在 $\mathcal{X}$上服从分布 P且独立同分布，并且 $\xi_1,\dots,\xi_n$ 是相互独立的均值为零的随机变量，且与 $\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}$独立。那么，存在一个常数 $C>0$使得

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n\xi_i f(\mathbf{x}^{(i)})\Big|^p\Big]\leqslant C^p\Bigg[\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n\xi_i f(\mathbf{x}^{(i)})\Big|\Big]^p+p^{p/2}n^{p/2}\Big(\sup_{f\in\mathcal{F}}\|f\|_{P,2}\Big)^p\cdot\max_i\|\xi_i\|_2^p\\
+p^p\mathbb{E}\Big[\max_i\Big(|\xi_i|^p\cdot\sup_{f\in\mathcal{F}}|f(\mathbf{x}^{(i)})|^p\Big)\Big]\Bigg]
\end{aligned}
$$

对每个 $p\geqslant 1$。这里，$\|\cdot\|_{P,\,2}$ 定义为

$$
\|f\|_{P,2}=\big(\mathbb{E}_{X\sim P}[f^2(X)]\big)^{1/2}.
$$

引理 7。 假设 $\mathbf{x}^{(1)}$、…、$\mathbf{x}^{(k)}$ 是定义在 $\mathbb{R}^d$ 上、密度为 p_0 的独立同分布随机变量，且令 ϵ_1、…、ϵ_k 为相互独立的拉德马赫随机变量，并且与 $\mathbf{x}^{(1)}$、…、$\mathbf{x}^{(k)}$ 独立。设 $\mathcal{F}$ 是从 $\mathbb{R}^d$ 到 $\mathbb{R}$ 的可数函数族，并假设存在 t、$D>0$，使得 $\|f\|_{p_0,\,2}\leqslant t$ 且对所有 $f\in\mathcal{F}$ 都有 $\|f\|_\infty\leqslant D$。那么，

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big]\leqslant CJ_{[\,]}(t,\mathcal{F},\|\cdot\|_{p_0,2})\cdot\Big(1+D\cdot\frac{J_{[\,]}(t,\mathcal{F},\|\cdot\|_{p_0,2})}{t^2\sqrt{k}}\Big),$$

其中 C 是一个普适常数，且 $J_{[\,]}(t,\mathcal{F},\|\cdot\|_{p_0,\,2})$ 是定义如下的括号熵积分

$$J_{[\,]}(t,\mathcal{F},\|\cdot\|_{p_0,2})=\int_0^t\sqrt{1+\log N_{[\,]}(\epsilon,\mathcal{F},\|\cdot\|_{p_0,2})}\,d\epsilon,$$

其中 $N_{[\,]}(\epsilon,\mathcal{F},\|\cdot\|_{p_0,\,2})$ 表示相对于 $\|\cdot\|_{p_0,\,2}$ 、$\mathcal{F}$ 的 ϵ-括号数。

定理 9 (Han 和 Wellner 的推论 1 [19])。 设 $\mathcal{F}_1,\ldots,\mathcal{F}_n$ 是从 $\mathcal{X}$ 到 $\mathbb{R}$ 的可数函数集合，使得对每个 $1\leqslant k\leqslant n$ 都有 $\mathcal{F}_k\supseteq\mathcal{F}_n$。假设 $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}$ 是定义在 $\mathcal{X}$ 上的置换不变随机变量，并设 $\xi_1,\ldots,\xi_n$ 是独立同分布的均值为零的随机变量，且与 $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}$ 相互独立。假设存在 $p\geqslant 1$ 和 $C>0$，使得

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}_k}\Big|\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big]\leqslant Ck^{1/p}$$

对每个 $1\leqslant k\leqslant n$，其中 $\epsilon_1,\ldots,\epsilon_n$ 是相互独立的拉德马赫随机变量，且与 $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}$ 相互独立。那么，对每个 $q\geqslant 1$,

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}_n}\Big|\sum_{i=1}^{n}\xi_i f(\mathbf{x}^{(i)})\Big|\Big]\leqslant 4C\|\xi_1\|_{\min(p,q),1}\cdot k^{1/\min(p,q)},$$

其中，对于每个 $r\geqslant 1$,

$$\|\xi_1\|_{r,1}:=\int_0^\infty\mathbb{P}(|\xi_1|>t)^{1/r}\,dt.$$

注 3. 上述结果假设函数类是可数的，以确保期望中的上确界具有可测性。对于不可数函数类 $\mathcal{F}$ 以及由 $\mathcal{F}$ 索引的随机过程 $(\Phi(f):\ f\in\mathcal{F})$，上确界 $\sup_{f\in\mathcal{F}}\Phi(f)$ 可能不可测。

在定理 *4* 的证明中，为避免此类可测性问题，我们将 Φ 在 $\mathcal{F}$ 上的期望上确界定义为

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\Phi(f)\Big]:=\sup\Big\{\mathbb{E}\Big[\sup_{f\in\mathcal{G}}\Phi(f)\Big]:\mathcal{G}\subseteq\mathcal{F}\ \textit{is countable}\Big\},$$

遵循 *Talagrand [29]*。类似地，对于任意 $c\in\mathbb{R}$，我们定义

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\Phi(f)\geqslant c\Big):=\sup\Big\{\mathbb{P}\Big(\sup_{f\in\mathcal{G}}\Phi(f)>c\Big):\mathcal{G}\subseteq\mathcal{F}\ \textit{is countable}\Big\}.$$

采用这些定义后，我们可以避免可测性问题，而且上述定理和引理也适用于不可数函数类。

定理 *4* 的证明。令 $\mathcal{F}_{\mathbf{M}}(V)$ 表示满足以下条件的形式 **(7)** 的所有函数 $f^{d,\,c,\,\{\nu_{L,\,U}\}}\in\mathcal{F}^{d,\,\infty\text{-ST}}_{s,\,s}$ 的集合：

(a) $\nu_{L,U}$ are supported on $\prod_{j\in L}(-M_j/2,M_j/2]\times\prod_{j\in U}(-M_j/2,M_j/2]$

(b)

$$\sum_{L,U:0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} \leqslant V.$$

从 $\hat{f}^{d,\ n}_{\ V}{}^{s}$ 的定义可知,

$$\hat{f}^{d,s}_{n,V} \in \mathcal{F}_{\mathbf{M}}(V) \subseteq \{f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}} : V^{d,s}_{\infty-\mathrm{XGB}}(f) \leqslant V\}.$$

此外，在附录 A.5.2 中证明的以下引理保证了 $f_{0,\mathbf{M}} \in \mathcal{F}_{\mathbf{M}}(V)$ 使得$f_{0,\mathbf{M}}(\cdot) = f_0(\cdot)$ 在 $\prod_{j=1}^{d}[-M_j/2, M_j/2]$ 上成立。

引理 8. 对于每个 $f \in \mathcal{F}^{d,\ s}_{,\infty-ST}$满足 $V^{d,\ s}_{,\infty-XGB}(f) < V$的情形，都存在 $f_{\mathbf{M}} \in \mathcal{F}_{\mathbf{M}}(V)$，使得 $f_{\mathbf{M}}(\cdot) = f(\cdot)$ 在 $\prod_{j=1}^{d}[-M_j/2, M_j/2]$ 上成立。

对于每个 $t > 0$，定义

$$B(V,t) = \{f \in \mathcal{F}_{\mathbf{M}}(V) : \|f\|_{p_0,2} \leqslant t\}.$$

为简洁起见，我们省略 $B(V,t)$ 对 $\mathbf{M} = (M_1, \ldots, M_d)$ 的依赖关系。以下引理在附录 A.5.3 中证明，为 $B(V,t)$ 给出了括号熵积分界，这将在整个证明过程中发挥关键作用。

引理 9。 存在一个常数 $C_{B,\ s} > 0$，其取决于 B 和 s，使得对于所有 $t > 0$,

$$J_{[\ ]}(t, B(V,t), \|\cdot\|_{p_0,2}) \leqslant C_{B,s} d^{\bar{s}} (1+\log d)^{\bar{s}-1} \left(t \log\left(2 + \frac{V}{t}\right) + V^{1/2} t^{1/2} \Big[\log\left(2 + \frac{V}{t}\right) \Big]^{\bar{s}-1} \right).$$

现在，假设我们有 $t_n > 4\|f_0 - f^*\|_{p_0,\ 2}$，使得对于每个 $r \geqslant 1$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{f \in B(V, rt_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i f(\mathbf{x}^{(i)}) \Big| \Big] \leqslant r\sqrt{n} t_n^2/(V+1), \\
&\mathbb{E}\Big[\sup_{f \in B(V, rt_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i f(\mathbf{x}^{(i)}) \Big| \Big] \leqslant r\sqrt{n} t_n^2/(V+1), \text{ and} \\
&\mathbb{E}\Big[\sup_{f \in B(V, rt_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i f(\mathbf{x}^{(i)}) \cdot (f_0 - f^*)(\mathbf{x}^{(i)}) \Big| \Big] \leqslant r\sqrt{n} t_n^2/(V+1),
\end{aligned} \tag{49}$$

其中 ϵ_i 是独立于 $\mathbf{x}^{(i)}$ 的拉德马赫随机变量，期望取遍 $\mathbf{x}^{(i)}$、ϵ_i 和 ξ_i。接下来，我们将首先说明如何利用这些关于期望上确界的界来获得 $\hat{f}^{d,\ n}_{\ V}{}^{s}$，的风险界。满足上述不等式的 t_n 的值将在下一步中指定，随后将推导出更精确的风险界。下标 n 强调 t_n 取决于 n，而其对其他参数的依赖关系为记号简洁起见予以省略。

我们首先旨在对 $\mathbb{P}(\|\hat{f}^{d,\ n}_{\ ,\ V}{}^{s} - f_0\|_{p_0,\ 2} > t)$ 在 $t \geqslant t_n$ 下进行界定。固定 $r \geqslant 1$，并对每个整数 $j \geqslant 2$ 定义

$$\mathcal{F}_j = \big\{ f \in \mathcal{F}_{\mathbf{M}}(V) : 2^{j-2} r t_n < \|f - f_{0,\mathbf{M}}\|_{p_0,2} < 2^j r t_n \big\}.$$

根据构造,

$$\mathbb{P}\big(\|\hat{f}^{d,s}_{n,V} - f_0\|_{p_0,2} > r t_n\big) = \mathbb{P}\big(\|\hat{f}^{d,s}_{n,V} - f_{0,\mathbf{M}}\|_{p_0,2} > r t_n\big) \leqslant \sum_{j=2}^{\infty} \mathbb{P}\big(\hat{f}^{d,s}_{n,V} \in \mathcal{F}_j\big).$$

接下来，令 $(M_n(f) :\ f \in \mathcal{F}_{\mathbf{M}}(V))$ 表示由以下方式定义的随机过程

$$M_n(f) = \frac{2}{n} \sum_{i=1}^{n} \xi_i (f - f^*)(\mathbf{x}^{(i)}) - \frac{1}{n} \sum_{i=1}^{n} (f - f^*)^2(\mathbf{x}^{(i)})$$

并按以下方式定义 $(M(f)\colon f\in\mathcal{F}_{\mathbf{M}}(V))$

$$M(f)=-\|f-f^*\|_{p_0,2}^2.$$

由于

$$M_n(f)=\frac{2}{n}\sum_{i=1}^n\xi_i(f-f^*)(\mathbf{x}^{(i)})-\frac{1}{n}\sum_{i=1}^n(f-f^*)^2(\mathbf{x}^{(i)})=-\frac{1}{n}\sum_{i=1}^n\big(y_i-f(\mathbf{x}^{(i)})\big)^2+\frac{1}{n}\sum_{i=1}^n\xi_i^2,$$

我们有

$$M_n(\hat{f}_{n,V}^{d,s})-M_n(f_{0,\mathbf{M}})\geqslant 0,$$

由于 $\hat{f}_{n,V}^{d,s}$ 在 $\mathcal{F}_{\mathbf{M}}(V)$ 上使最小二乘损失最小。此外,

$$\begin{aligned}M_n(f)-M_n(f_{0,\mathbf{M}})=&\frac{2}{n}\sum_{i=1}^n\xi_i(f-f_{0,\mathbf{M}})(\mathbf{x}^{(i)})-\frac{1}{n}\sum_{i=1}^n(f-f_{0,\mathbf{M}})^2(\mathbf{x}^{(i)})\\&-\frac{2}{n}\sum_{i=1}^n(f-f_{0,\mathbf{M}})(\mathbf{x}^{(i)})\cdot(f_{0,\mathbf{M}}-f^*)(\mathbf{x}^{(i)})\end{aligned}$$

and

$$M(f)-M(f_{0,\mathbf{M}})=-\|f-f_{0,\mathbf{M}}\|_{p_0,2}^2-2\mathbb{E}_{\mathbf{x}\sim p_0}\big[(f-f_{0,\mathbf{M}})(\mathbf{x})\cdot(f_{0,\mathbf{M}}-f^*)(\mathbf{x})\big].$$

假设 $t_n>4\|f_0-f^*\|_{p_0,\,2}=4\|f_{0,\,\mathbf{M}}-f^*\|_{p_0,\,2}$意味着对于每个 $f\in\mathcal{F}_{\quad j}$,

$$\begin{aligned}-(M(f)-M(f_{0,\mathbf{M}}))&\geqslant\|f-f_{0,\mathbf{M}}\|_{p_0,2}\cdot\big(\|f-f_{0,\mathbf{M}}\|_{p_0,2}-2\|f_{0,\mathbf{M}}-f^*\|_{p_0,2}\big)\\&\geqslant 2^{j-2}rt_n\cdot\Big(2^{j-2}rt_n-\frac{t_n}{2}\Big)\geqslant 2^{2j-5}r^2t_n^2.\end{aligned}$$

因此, [2]

$$\begin{aligned}\mathbb{P}\big(\hat{f}_{n,V}^{d,s}\in\mathcal{F}_j\big)&\leqslant\mathbb{P}\Big(\sup_{f\in\mathcal{F}_j}\big(M_n(f)-M_n(f_{0,\mathbf{M}})\big)\geqslant 0\Big)\\&\leqslant\mathbb{P}\Big(\sup_{f\in\mathcal{F}_j}\big((M_n(f)-M_n(f_{0,\mathbf{M}}))-(M(f)-M(f_{0,\mathbf{M}}))\big)\geqslant 2^{2j-5}r^2t_n^2\Big)\\&\leqslant\mathbb{P}\Big(\sup_{f\in\mathcal{F}_j}\Big|\frac{1}{n}\sum_{i=1}^n\xi_i(f-f_{0,\mathbf{M}})(\mathbf{x}^{(i)})\Big|\geqslant 2^{2j-7}r^2t_n^2\Big)\\&\quad+\mathbb{P}\Big(\sup_{f\in\mathcal{F}_j}\Big|\frac{1}{n}\sum_{i=1}^n(f-f_{0,\mathbf{M}})^2(\mathbf{x}^{(i)})-\|f-f_{0,\mathbf{M}}\|_{p_0,2}^2\Big|\geqslant 2^{2j-7}r^2t_n^2\Big)\\&\quad+\mathbb{P}\Big(\sup_{f\in\mathcal{F}_j}\Big|\frac{1}{n}\sum_{i=1}^n(f-f_{0,\mathbf{M}})(\mathbf{x}^{(i)})\cdot(f_{0,\mathbf{M}}-f^*)(\mathbf{x}^{(i)})\\&\qquad\qquad-\mathbb{E}_{\mathbf{x}\sim p_0}\big[(f-f_{0,\mathbf{M}})(\mathbf{x})\cdot(f_{0,\mathbf{M}}-f^*)(\mathbf{x})\big]\Big|\geqslant 2^{2j-8}r^2t_n^2\Big)\\&\leqslant\mathbb{P}\Big(\sup_{f\in B(2V,2^jrt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_if(\mathbf{x}^{(i)})\Big|\geqslant 2^{2j-7}r^2\sqrt{n}t_n^2\Big)\\&\quad+\mathbb{P}\Big(\sup_{f\in B(2V,2^jrt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-7}r^2\sqrt{n}t_n^2\Big)\qquad(50)\\&\quad+\mathbb{P}\Big(\sup_{f\in B(2V,2^jrt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(f(\mathbf{x}^{(i)})\cdot(f_{0,\mathbf{M}}-f^*)(\mathbf{x}^{(i)})\\&\qquad\qquad-\mathbb{E}_{\mathbf{x}\sim p_0}\big[f(\mathbf{x})\cdot(f_{0,\mathbf{M}}-f^*)(\mathbf{x})\big]\Big)\Big|\geqslant 2^{2j-8}r^2\sqrt{n}t_n^2\Big),\end{aligned}$$

[2]由于我们在注 3 中引入了这些定义以避免可测性问题，因此需要更加谨慎地论证第一个不等式。证明之后的注释中给出了更为详细的论证。

其中，第二个不等式利用了对于所有 $f \in \mathcal{F}_j$，$-(M(f) - M(f_{0,\ \mathrm{M}})) \geqslant 2^{2j-5} r^2 t_n^2$；最后一个不等式则是因为对于所有 $f \in \mathcal{F}_j$，$f - f_{0,\ \mathrm{M}} \in B(2V,\ 2^j r t_n)$。

接下来，我们对（50）右侧的每一项进行界定。作为准备工作，我们证明存在一个常数 $C > 0$，使得对于每个 $f \in B(V,\ t)$都有 $\|f\|_\infty \leqslant C(V+t)$。设 $f \in B(V,\ t)$ 具有如下形式

$$f(x_1, \ldots, x_d) = c + \sum_{0 < |L| + |U| \leqslant s} \int_{\mathbb{R}^{|L|+|U|}} \prod_{j \in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j \in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l}, \mathbf{u})$$

其中

$$\sum_{0 < |L| + |U| \leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} \leqslant V.$$

由于这些有符号测度的全变差之和被 V 所界定，因此上述 f 表示式中的第二项的绝对值一致地被 V 所界定。因此，根据柯西不等式，

$$t^2 \geqslant \|f\|_{p_0,2}^2 = \int_{\prod_{j=1}^d [-M_j/2, M_j/2]} f^2(\mathbf{x}) \cdot p_0(\mathbf{x})\, d\mathbf{x} \geqslant \int_{\prod_{j=1}^d [-M_j/2, M_j/2]} \Big(\frac{c^2}{2} - V^2 \Big) \cdot p_0(\mathbf{x})\, d\mathbf{x} = \frac{c^2}{2} - V^2.$$

由此可得

$$\|f\|_\infty \leqslant |c| + V \leqslant C(V+t)$$

对于某个普适常数 $C > 0$。

现在我们对（50）右侧的第一项进行界定。根据马尔可夫不等式，

$$\mathbb{P}\Big(\sup_{f \in B(2V, 2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(\mathbf{x}^{(i)}) \Big| \geqslant 2^{2j-7} r^2 \sqrt{n} t_n^2 \Big) \leqslant \frac{1}{2^{6j-21} r^6 n^{3/2} t_n^6} \cdot \mathbb{E}\Big[\sup_{f \in B(2V, 2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(\mathbf{x}^{(i)}) \Big|^3 \Big].$$

为了对右侧的期望进行界定，我们应用定理 8，得到

$$\begin{aligned} \mathbb{E}\Big[\sup_{f \in B(2V, 2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(\mathbf{x}^{(i)}) \Big|^3 \Big] &\leqslant C \cdot \mathbb{E}\Big[\sup_{f \in B(2V, 2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(\mathbf{x}^{(i)}) \Big| \Big]^3 + C \cdot 2^{3j} r^3 t_n^3 \|\xi_1\|_2^3 \\ &\quad + C n^{-3/2} \cdot \mathbb{E}\Big[\max_i \Big(|\xi_i|^3 \cdot \sup_{f \in B(2V, 2^j r t_n)} |f(\mathbf{x}^{(i)})|^3 \Big) \Big]. \end{aligned} \tag{51}$$

利用 (49) 中的不等式

$$\max_i |\xi_i|^3 \leqslant \sum_{i=1}^n |\xi_i|^3,$$

以及初步结果

$$\|f\|_\infty \leqslant C(V + 2^j r t_n) \ \text{ for all } f \in B(2V, 2^j r t_n), \tag{52}$$

由 (51) 可推得

$$\mathbb{E}\Big[\sup_{f \in B(2V, 2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(\mathbf{x}^{(i)}) \Big|^3 \Big] \leqslant C \cdot 2^{3j} r^3 n^{3/2} t_n^6 + C \cdot 2^{3j} r^3 t_n^3 \|\xi_1\|_2^3 + C n^{-1/2} \|\xi_1\|_3^3 (V^3 + 2^{3j} r^3 t_n^3).$$

将其代回马尔可夫不等式给出的界中，我们得到

$$\mathbb{P}\Big(\sup_{f \in B(2V, 2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(\mathbf{x}^{(i)}) \Big| \geqslant 2^{2j-7} r^2 \sqrt{n} t_n^2 \Big) \leqslant \frac{C}{2^{3j} r^3} + \frac{C \|\xi_1\|_2^3}{2^{3j} r^3 n^{3/2} t_n^3} + \frac{C \|\xi_1\|_3^3 V^3}{2^{6j} r^6 n^2 t_n^6} + \frac{C \|\xi_1\|_3^3}{2^{3j} r^3 n^2 t_n^3}.$$

接下来，我们对式 (50) 右侧的第二项进行界定。根据是否满足 $2^j rt_n \leqslant V$，我们分两种情况讨论。首先，假设 $2^j rt_n \leqslant V$。根据马尔可夫不等式，

$$\begin{aligned}&\mathbb{P}\Big(\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-7}r^2\sqrt{n}t_n^2\Big)\\&\qquad\leqslant\frac{1}{2^{6j-21}r^6n^{3/2}t_n^6}\cdot\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|^3\Big].\end{aligned}\tag{53}$$

此外，根据对称化的标准论证（例如，参见 van der Vaart 和 Wellner [36, 引理 2.3.1]以及 van de Geer [33,定理 16.1]），

$$\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|^3\Big]\leqslant 8\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i f^2(\mathbf{x}^{(i)})\Big|^3\Big],\tag{54}$$

其中 ϵ_i 是与$\mathbf{x}^{(i)}$独立的拉德马赫随机变量。应用定理 8，我们可以将右侧的期望值界定为

$$\begin{aligned}\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i f^2(\mathbf{x}^{(i)})\Big|^3\Big]&\leqslant C\cdot\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i f^2(\mathbf{x}^{(i)})\Big|\Big]^3\\&\quad+C\Big(\sup_{f\in B(2V,2^j rt_n)}\|f^2\|_{p_0,2}\Big)^3+Cn^{-3/2}\cdot\mathbb{E}\Big[\max_i\sup_{f\in B(2V,2^j rt_n)}|f(\mathbf{x}^{(i)})|^6\Big].\end{aligned}$$

压缩原理（例如，参见 van der Vaart 和 Wellner [36, 命题 A.3.2] 以及 Ledoux 和 Talagrand [22, 定理 4.12]），结合式 (49) 和式 (52)，可得

$$\begin{aligned}\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i f^2(\mathbf{x}^{(i)})\Big|\Big]&\leqslant C(V+2^j rt_n)\cdot\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big]\\&\leqslant C(V+2^j rt_n)\cdot 2^j r\sqrt{n}t_n^2/V.\end{aligned}$$

此外，我们有

$$\sup_{f\in B(2V,2^j rt_n)}\|f^2\|_{p_0,2}\leqslant\sup_{f\in B(2V,2^j rt_n)}\|f\|_\infty\cdot\sup_{f\in B(2V,2^j rt_n)}\|f\|_{p_0,2}\leqslant C(V+2^j rt_n)\cdot 2^j rt_n.$$

因此，

$$\begin{aligned}&\mathbb{E}\Big[\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i f^2(\mathbf{x}^{(i)})\Big|^3\Big]\\&\quad\leqslant C(V+2^j rt_n)^3\cdot 2^{3j}r^3n^{3/2}t_n^6/V^3+C(V+2^j rt_n)^3\cdot 2^{3j}r^3t_n^3+Cn^{-3/2}(V+2^j rt_n)^6.\end{aligned}$$

将此结果与式 (53) 和式 (54) 结合，可得

$$\begin{aligned}&\mathbb{P}\Big(\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-7}r^2\sqrt{n}t_n^2\Big)\\&\quad\leqslant\frac{C(V+2^j rt_n)^3}{2^{3j}r^3V^3}+\frac{C(V+2^j rt_n)^3}{2^{3j}r^3n^{3/2}t_n^3}+\frac{C(V+2^j rt_n)^6}{2^{6j}r^6n^3t_n^6}\leqslant\frac{C}{2^{3j}r^3}+\frac{CV^3}{2^{3j}r^3n^{3/2}t_n^3}+\frac{CV^6}{2^{6j}r^6n^3t_n^6}.\end{aligned}$$

接下来，假设 $2^j rt_n > V$。对于每个 $f\in B(2V,\ 2^j rt_n)$，如初步步骤中所示，我们可以将 f 分解为 $f=c+g$，其中 c 是一个常数，满足 $|c|\leqslant C(V+2^j rt_n)$，而 g 是一个绝对值一致有界于 $2V$的函数。利用这一分解，我们可以写出

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)=2c\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g(\mathbf{x}^{(i)})-\mathbb{E}_{X\sim p_0}[g(X)]\big)+\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g^2(\mathbf{x}^{(i)})-\|g\|_{p_0,2}^2\big).$$

由此可得

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-7}r^2\sqrt{n}t_n^2\Big)\\
&\quad\leqslant\mathbb{P}\Big(\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g(\mathbf{x}^{(i)})-\mathbb{E}_{X\sim p_0}[g(X)]\big)\Big|\geqslant\frac{2^{2j-8}r^2\sqrt{n}t_n^2}{C(V+2^j rt_n)}\Big)\\
&\qquad+\mathbb{P}\Big(\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g^2(\mathbf{x}^{(i)})-\|g\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-8}r^2\sqrt{n}t_n^2\Big).
\end{aligned}\tag{55}$$

根据马尔可夫不等式，(55) 的第一项有如下界：

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g(\mathbf{x}^{(i)})-\mathbb{E}_{X\sim p_0}[g(X)]\big)\Big|\geqslant\frac{2^{2j-8}r^2\sqrt{n}t_n^2}{C(V+2^j rt_n)}\Big)\\
&\quad\leqslant\frac{C(V+2^j rt_n)^3}{2^{6j-24}r^6n^{3/2}t_n^6}\cdot\mathbb{E}\Big[\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g(\mathbf{x}^{(i)})-\mathbb{E}_{X\sim p_0}[g(X)]\big)\Big|^3\Big].
\end{aligned}\tag{56}$$

此外，根据标准的对称化论证，

$$\mathbb{E}\Big[\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g(\mathbf{x}^{(i)})-\mathbb{E}_{X\sim p_0}[g(X)]\big)\Big|^3\Big]\leqslant 8\mathbb{E}\Big[\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i g(\mathbf{x}^{(i)})\Big|^3\Big].\tag{57}$$

利用定理 8，右侧的期望有如下界：

$$\mathbb{E}\Big[\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i g(\mathbf{x}^{(i)})\Big|^3\Big]\leqslant C\cdot\mathbb{E}\Big[\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i g(\mathbf{x}^{(i)})\Big|\Big]^3+CV^3.\tag{58}$$

应用引理 7 和引理 9，可得

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i g(\mathbf{x}^{(i)})\Big|\Big]=\mathbb{E}\Big[\sup_{\substack{g\in B(2V,2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i g(\mathbf{x}^{(i)})\Big|\Big]\\
&\quad\leqslant CJ_{[\,]}(2V,B(2V,2V),\|\cdot\|_{p_0,2})\cdot\Big(1+2V\cdot\frac{J_{[\,]}(2V,B(2V,2V),\|\cdot\|_{p_0,2})}{4V^2\sqrt{n}}\Big)\leqslant C_{B,s}a_{d,s}^2V,
\end{aligned}$$

其中 $a_{d,\ s}:=d^{\bar{s}}(1+\log d)^{\bar{s}-1}$。结合 (56)、(57) 和 (58)，得到

$$\mathbb{P}\Big(\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g(\mathbf{x}^{(i)})-\mathbb{E}_{X\sim p_0}[g(X)]\big)\Big|\geqslant\frac{2^{2j-8}r^2\sqrt{n}t_n^2}{C(V+2^j rt_n)}\Big)\leqslant\frac{C_{B,s}a_{d,s}^6V^3(V+2^j rt_n)^3}{2^{6j-24}r^6n^{3/2}t_n^6}\leqslant\frac{C_{B,s}a_{d,s}^6V^3}{2^{3j}r^3n^{3/2}t_n^3},$$

其中，最后一个不等式由假设 $2^j rt_n>V$ 得出。通过类似的论证，(55) 中的第二项有如下界：

$$\mathbb{P}\Big(\sup_{\substack{g\in F_{\mathbf{M}}(2V)\\ \|g\|_\infty\leqslant 2V}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(g^2(\mathbf{x}^{(i)})-\|g\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-8}r^2\sqrt{n}t_n^2\Big)\leqslant\frac{C_{B,s}a_{d,s}^6V^6}{2^{6j}r^6n^{3/2}t_n^6}\leqslant\frac{C_{B,s}a_{d,s}^6V^3}{2^{3j}r^3n^{3/2}t_n^3}.$$

将这些界代回（55）式可得

$$\mathbb{P}\Big(\sup_{f\in B(2V,2^j rt_n)}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(f^2(\mathbf{x}^{(i)})-\|f\|_{p_0,2}^2\big)\Big|\geqslant 2^{2j-7}r^2\sqrt{n}t_n^2\Big)\leqslant\frac{C_{B,s}a_{d,s}^6V^3}{2^{3j}r^3n^{3/2}t_n^3}.$$

因此，无论 $2^j r t_n \leqslant V$ 是否成立，我们都有

$$\mathbb{P}\Big(\sup_{f\in B(2V,2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(f^2(\mathbf{x}^{(i)}) - \|f\|_{p_0,2}^2 \big) \Big| \geqslant 2^{2j-7} r^2 \sqrt{n} t_n^2 \Big) \leqslant \frac{C}{2^{3j} r^3} + \frac{C_{B,s} a_{d,s}^6 V^3}{2^{3j} r^3 n^{3/2} t_n^3} + \frac{C V^6}{2^{6j} r^6 n^3 t_n^6}.$$

(50) 右侧的第三项可以用与 $2^j r t_n \leqslant V$ 情形下第二项类似的方法进行界定。依次应用马尔可夫不等式、对称化论证和定理 8，可得

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in B(2V,2^j r t_n)} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(f(\mathbf{x}^{(i)}) \cdot (f_{0,\mathbf{M}} - f^*)(\mathbf{x}^{(i)}) - \mathbb{E}_{\mathbf{x}\sim p_0}\big[f(\mathbf{x}) \cdot (f_{0,\mathbf{M}} - f^*)(\mathbf{x}) \big] \Big) \Big| \geqslant 2^{2j-8} r^2 \sqrt{n} t_n^2 \Big) \\
&\quad \leqslant \frac{C}{2^{3j} r^3} + \frac{C \|f_0 - f^*\|_{\infty,\mathbf{M}}^3}{2^{3j} r^3 n^{3/2} t_n^3} + \frac{C \|f_0 - f^*\|_{\infty,\mathbf{M}}^3 V^3}{2^{6j} r^6 n^3 t_n^6},
\end{aligned}$$

其中 $\|\cdot\|_{\infty,\,\mathbf{M}}$ 表示在 $\prod_{j=1}^{d} [-M_j/2, M_j/2]$ 上的上确界范数

$$\|g\|_{\infty,\mathbf{M}} := \sup_{\mathbf{x}\in \prod_{j=1}^{d} [-M_j/2, M_j/2]} |g(\mathbf{x})|.$$

因此，我们有

$$\begin{aligned}
&\mathbb{P}\big(\|\hat{f}_{n,V}^{d,s} - f_0\|_{p_0,2} > r t_n \big) = \sum_{j=2}^{\infty} \mathbb{P}\big(\hat{f}_{n,V}^{d,s} \in \mathcal{F}_j \big) \\
&\quad \leqslant \sum_{j=2}^{\infty} \Big[\frac{C}{2^{3j} r^3} + \frac{C(\|\xi_1\|_2^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3)}{2^{3j} r^3 n^{3/2} t_n^3} + \frac{C \|\xi_1\|_3^3 V^3}{2^{6j} r^6 n^2 t_n^6} + \frac{C \|\xi_1\|_3^3}{2^{3j} r^3 n^2 t_n^3} \\
&\qquad\qquad + \frac{C_{B,s} a_{d,s}^6 V^3}{2^{3j} r^3 n^{3/2} t_n^3} + \frac{C V^3 (V^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3)}{2^{6j} r^6 n^3 t_n^6} \Big] \\
&\quad \leqslant \frac{C}{r^3} + \frac{C(\|\xi_1\|_2^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3) + C_{B,s} a_{d,s}^6 V^3}{r^3 n^{3/2} t_n^3} + \frac{C \|\xi_1\|_3^3 V^3}{r^6 n^2 t_n^6} + \frac{C \|\xi_1\|_3^3}{r^3 n^2 t_n^3} + \frac{C V^3 (V^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3)}{r^6 n^3 t_n^6}.
\end{aligned}$$

代入 $r = t/t_n$, ，得到

$$\begin{aligned}
\mathbb{P}\big(\|\hat{f}_{n,V}^{d,s} - f_0\|_{p_0,2} > t \big) &\leqslant \frac{C t_n^3}{t^3} + \frac{C(\|\xi_1\|_2^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3) + C_{B,s} a_{d,s}^6 V^3}{n^{3/2} t^3} \\
&\quad + \frac{C \|\xi_1\|_3^3 V^3}{n^2 t^6} + \frac{C \|\xi_1\|_3^3}{n^2 t^3} + \frac{C V^3 (V^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3)}{n^3 t^6},
\end{aligned}$$

这对所有 $t \geqslant t_n$ 都成立。因此，对每个 $t \geqslant 2t_n$，我们有

$$\begin{aligned}
&\mathbb{P}\big(\|\hat{f}_{n,V}^{d,s} - f^*\|_{p_0,2} > t \big) \leqslant \mathbb{P}\big(\|\hat{f}_{n,V}^{d,s} - f_0\|_{p_0,2} > t - \|f_0 - f^*\|_{p_0,2} \big) \leqslant \mathbb{P}\Big(\|\hat{f}_{n,V}^{d,s} - f_0\|_{p_0,2} > \frac{t}{2} \Big) \\
&\quad \leqslant \frac{C t_n^3}{t^3} + \frac{C(\|\xi_1\|_2^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3) + C_{B,s} a_{d,s}^6 V^3}{n^{3/2} t^3} + \frac{C \|\xi_1\|_3^3 V^3}{n^2 t^6} + \frac{C \|\xi_1\|_3^3}{n^2 t^3} + \frac{C V^3 (V^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3)}{n^3 t^6},
\end{aligned}$$

其中第二个不等式由假设 $t_n > 4\|f_0 - f^*\|_{p_0,\ 2}$ 得出。由于

$$\int_a^{\infty} 2y \cdot \mathbb{P}(Y \geqslant y) dy = \mathbb{E}[(Y^2 - a^2)_+],$$

其中，$(\cdot)_+$ 表示正部，因此可得

$$\begin{aligned}
\mathbb{E}\big[\|\hat{f}_{n,V}^{d,s} - f^*\|_{p_0,2}^2 \big] &\leqslant 4t_n^2 + \int_{2t_n}^{\infty} 2t \cdot \mathbb{P}\big(\|\hat{f}_{n,V}^{d,s} - f^*\|_{p_0,2} > t \big)\, dt \\
&\leqslant C t_n^2 + \frac{C(\|\xi_1\|_2^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3) + C_{B,s} a_{d,s}^6 V^3}{n^{3/2} t_n} + \frac{C \|\xi_1\|_3^3 V^3}{n^2 t_n^4} \\
&\quad + \frac{C \|\xi_1\|_3^3}{n^2 t_n} + \frac{C V^3 (V^3 + \|f_0 - f^*\|_{\infty,\mathbf{M}}^3)}{n^3 t_n^4}.
\end{aligned} \tag{59}$$

我们刚刚看到，一旦我们利用某个 $t_n > 4\|f_0 - f^*\|_{p_0,2}$ 建立期望上确界的界（49），就可以像上面的公式那样，用 t_n 来界定 $\hat{f}^{d,n}_{V,s}$ 的风险。因此，我们接下来的目标是找出一个满足（49）的合适 t_n。为此，我们先界定

$$\mathbb{E}\Big[\sup_{f\in B(V,t)}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big]$$

对每个 $t>0$ 和 $k=1,\ldots\ldots,n$，然后应用定理 9，将此界转化为关于 ξ_i 的期望上确界。

固定 $k\in\{1,\ldots\ldots,n\}$。由于对每个 $f\in B(V,t)$ 都有 $\|f\|_\infty\leqslant C(V+t)$，由引理 7 可得

$$\mathbb{E}\Big[\sup_{f\in B(V,t)}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big]\leqslant CJ_{[\,]}(t,B(V,t),\|\cdot\|_{p_0,2})\cdot\Big(1+C(V+t)\cdot\frac{J_{[\,]}(t,B(V,t),\|\cdot\|_{p_0,2})}{t^2\sqrt{k}}\Big).$$

应用引理 9 中的熵积分界，我们得到

$$\begin{aligned}
\mathbb{E}\Big[\sup_{f\in B(V,t)}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big] &\leqslant C_{B,s}a_{d,s}\Big(t\log\Big(2+\frac{V}{t}\Big)+V^{1/2}t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}\Big)\\
&\quad\cdot\Big[1+C_{B,s}a_{d,s}(V+t)k^{-1/2}t^{-2}\Big(t\log\Big(2+\frac{V}{t}\Big)+V^{1/2}t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}\Big)\Big]\\
&\leqslant C_{B,s}a_{d,s}t\log\Big(2+\frac{V}{t}\Big)+C_{B,s}a_{d,s}V^{1/2}t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}+C_{B,s}a_{d,s}^2k^{-1/2}V^{3/2}t^{-1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}}\\
&\quad+C_{B,s}a_{d,s}^2k^{-1/2}V^2t^{-1}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{2(\bar{s}-1)}+C_{B,s}a_{d,s}^2k^{-1/2}t\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{2}\\
&\quad+C_{B,s}a_{d,s}^2k^{-1/2}V^{1/2}t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}}+C_{B,s}a_{d,s}^2k^{-1/2}V\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\max(2,2(\bar{s}-1))}.
\end{aligned}\tag{60}$$

回顾一下，$a_{d,s}=d^s(1+\log d)^{\bar{s}-1}$。令 $\Psi:\ \mathbb{R}\to\mathbb{R}$ 表示由式 (60) 右侧给出的函数。直接计算可知，如果

$$\begin{aligned}
t\geqslant\max\Big(&C_{B,s}a_{d,s}(V+1)k^{-1/2}\log(2+k),C_{B,s}a_{d,s}^{2/3}(V+1)k^{-1/3}[\log(2+k)]^{2(\bar{s}-1)/3},\\
&C_{B,s}a_{d,s}^{4/5}(V+1)k^{-2/5}[\log(2+k)]^{2\bar{s}/5},C_{B,s}a_{d,s}^{2}(V+1)k^{-1}[\log(2+k)]^{2},\\
&C_{B,s}a_{d,s}^{4/3}(V+1)k^{-2/3}[\log(2+k)]^{2\bar{s}/3},C_{B,s}a_{d,s}(V+1)k^{-1/2}[\log(2+k)]^{\max(1,\bar{s}-1)}\Big),
\end{aligned}$$

then

$$\Psi(t)\leqslant\sqrt{k}t^2/(V+1).$$

为简化这个最大值，注意到对于适当的常数 $C,\ C_s>0$，有

$$\begin{aligned}
&\log(2+x)\leqslant x^{1/6}\ \text{ for all } x\geqslant C,\\
&[\log(2+x)]^{2\bar{s}/5}\leqslant x^{1/15}\ \text{ for all } x\geqslant C_s,\\
&[\log(2+x)]^{2}\leqslant x^{2/3}\ \text{ for all } x\geqslant C,\\
&[\log(2+x)]^{2\bar{s}/3}\leqslant x^{1/3}\ \text{ for all } x\geqslant C_s,\ \text{and}\\
&[\log(2+x)]^{\max(1,\bar{s}-1)}\leqslant x^{1/6}\ \text{ for all } x\geqslant C_s.
\end{aligned}$$

利用这些不等式，我们可以对最大值中的各项作如下界定：

$$\begin{aligned}
&k^{-1/2}\log(2+k)\leqslant Ck^{-1/2}+k^{-1/2}k^{1/6}\leqslant Ck^{-1/3},\\
&k^{-2/5}[\log(2+k)]^{2\bar{s}/5}\leqslant C_sk^{-2/5}+k^{-2/5}k^{1/15}\leqslant C_sk^{-1/3},\\
&k^{-1}[\log(2+k)]^{2}\leqslant Ck^{-1}+k^{-1}k^{2/3}\leqslant Ck^{-1/3},\\
&k^{-2/3}[\log(2+k)]^{2\bar{s}/3}\leqslant C_sk^{-2/3}+k^{-2/3}k^{1/3}\leqslant C_sk^{-1/3},\ \text{and}\\
&k^{-1/2}[\log(2+k)]^{\max(1,\bar{s}-1)}\leqslant C_sk^{-1/2}+k^{-1/2}k^{1/6}\leqslant C_sk^{-1/3}.
\end{aligned}$$

因此，如果我们设定

$$\widetilde{t}_k = C_{B,s} a_{d,s}^2 (V+1) k^{-1/3} [\log(2+n)]^{2(\bar{s}-1)/3},$$

于是我们有

$$\Psi(\widetilde{t}_k) \leqslant \sqrt{k}\widetilde{t}_k^2/(V+1).$$

由于映射 $t \mapsto \Psi(t)/t$ 是递减的，因此

$$\mathbb{E}\Big[\sup_{f\in B(V,r\widetilde{t}_k)} \Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big] \leqslant \Psi(r\widetilde{t}_k) \leqslant r\Psi(\widetilde{t}_k) \leqslant r\sqrt{k}\widetilde{t}_k^2/(V+1)$$

对于每个 $r \geqslant 1$，均成立。

我们刚刚证明，对于每个 $k = 1$、.....、 n,

$$\mathbb{E}\Big[\sup_{f\in B(V,r\widetilde{t}_k)} \Big|\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big] \leqslant rk\widetilde{t}_k^2/(V+1)$$

对于每个 $r \geqslant 1$，均成立。根据定理 9，此界可推广到 ξ_i 的期望上确界，从而得到

$$\mathbb{E}\Big[\sup_{f\in B(V,r\widetilde{t}_n)} \Big|\sum_{i=1}^{n}\xi_i f(\mathbf{x}^{(i)})\Big|\Big] \leqslant 4\|\xi_1\|_{3,1}\cdot rn\widetilde{t}_n^2/(V+1)$$

对于每个 $r \geqslant 1$，均成立。因此，将 $\widetilde{t}_n$ 重新定义为乘以因子 $(1+4\|\xi_1\|_{3,\ 1})$，即可得到式 **(49)** 中的前两个不等式（取 $t_n = \widetilde{t}_n$），适用于 $r \geqslant 1$。

现在我们对其进行界定

$$\mathbb{E}\Big[\sup_{f\in B(V,t)} \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i f(\mathbf{x}^{(i)})\cdot(f_0-f^*)(\mathbf{x}^{(i)})\Big|\Big]$$

对每个 $t > 0$。只需稍作修改，按照引理 7 的证明，我们便可证明

$$\begin{aligned}&\mathbb{E}\Big[\sup_{f\in B(V,t)} \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i f(\mathbf{x}^{(i)})\cdot(f_0-f^*)(\mathbf{x}^{(i)})\Big|\Big]\\ &\quad\leqslant C\|f_0-f^*\|_{\infty,\mathbf{M}}\cdot J_{[\ ]}(t,B(V,t),\|\cdot\|_{p_0,2})\cdot\Big(1+C(V+t)\cdot\frac{J_{[\ ]}(t,B(V,t),\|\cdot\|_{p_0,2})}{t^2\sqrt{n}}\Big).\end{aligned}$$

因此，重复上述计算可得，若我们将 t_n 定义为

$$\bar{t}_n = (1+\|f_0-f^*\|_{\infty,\mathbf{M}})\cdot C_{B,s}a_{d,s}^2(V+1)n^{-1/3}[\log(2+n)]^{2(\bar{s}-1)/3},$$

then

$$\mathbb{E}\Big[\sup_{f\in B(V,\bar{t}_n)} \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i f(\mathbf{x}^{(i)})\cdot(f_0-f^*)(\mathbf{x}^{(i)})\Big|\Big] \leqslant \sqrt{n}t_n^2/(V+1),$$

由此，对于所有 $r \geqslant 1$，即可推出 **(49)** 中的最后一个不等式（其中 $t_n = \bar{t}_n$）。

利用 t_n 和 t_n，我们将 t_n 定义为

$$t_n = 4\|f_0-f^*\|_{p_0,2} + \max(\widetilde{t}_n, \bar{t}_n).$$

那么，对于每个 $r \geqslant 1$,

$$\mathbb{E}\Big[\sup_{f\in B(V,rt_n)} \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big] \leqslant (rt_n/\widetilde{t}_n)\cdot\sqrt{n}\widetilde{t}_n^2/(V+1) \leqslant r\sqrt{n}t_n^2/(V+1)$$

其余两个不等式（49）也可用同样的论证得出。因此， t_n 对所有 $r \geqslant 1$ 都满足（49）中的所有不等式，我们利用这一点 t_n 推导 $\hat{f}^{d,\ s}_{n,\ V,}$ 的风险界。

最后一步，我们将 t_n 代入 (59)，得到

$$\mathbb{E}\big[\|\hat{f}^{d,s}_{n,V} - f^*\|^2_{p_0,2}\big] \leqslant C\|f_0 - f^*\|^2_{p_0,2} + a^4_{d,s}(V+1)^2\Big[C_{B,s}(1 + \|\xi_1\|^2_{3,1} + \|f_0 - f^*\|^2_{\infty,\mathbf{M}}) \cdot n^{-2/3}[\log(2+n)]^{4(\bar{s}-1)/3} + O(n^{-2/3})\Big],$$

当 $s \geqslant 2$时。另一方面，若 $s = 1$，则得到

$$\begin{aligned}\mathbb{E}\big[\|\hat{f}^{d,s}_{n,V} - f^*\|^2_{p_0,2}\big] &\leqslant C\|f_0 - f^*\|^2_{p_0,2} + a^4_{d,1}(V+1)^2\Big[C\Big(\frac{\max(\widetilde{t}_n, \bar{t}_n)}{a^2_{d,1}(V+1)}\Big)^2 \\ &\qquad + \frac{C\|\xi_1\|^3_3 V^3}{a^4_{d,1}(V+1)^2 n^2 (\max(\widetilde{t}_n, \bar{t}_n))^4} + o(n^{-2/3})\Big] \\ &= C\|f_0 - f^*\|^2_{p_0,2} + a^4_{d,1}(V+1)^2 \cdot O(n^{-2/3}),\end{aligned}$$

其中， $O(\cdot)$ 所依赖的常数因子取决于 B， s、 ξ_i 的矩，以及 $\|f_0 - f^*\|_{\infty,\ \mathbf{M}}$。 □

注4. 对于 (50) 中的第一个不等式，我们实际上需要证明

$$\mathbb{P}\big(\hat{f}^{d,s}_{n,V} \in \mathcal{F}_j\big) \leqslant \sup\Big\{\mathbb{P}\Big(\sup_{f\in\mathcal{G}}\big(M_n(f) - M_n(f_{0,\mathbf{M}})\big) \geqslant 0\Big) : \mathcal{G} \subseteq \mathcal{F}_j \text{ is countable}\Big\},$$

因为 $\mathcal{F}_j$ 可能不可数。这里，我们对此给出一个更严谨的论证。

对于每个整数 $N \geqslant 1$，令 $\mathcal{G}_N$表示由所有满足以下两个条件的 $\mathcal{F}_j$ 构成的子集族： $f^{d,\ c\ \ ^s,\ \{\nu_{L,\ U}\}}$（形式为 (7)）：

(a) $\nu_{L,U}$ *are supported on* $\prod_{j\in L}((1/N)\mathbb{Z} \cap (-M_j/2, M_j/2]) \times \prod_{j\in U}((1/N)\mathbb{Z} \cap (-M_j/2, M_j/2])$*, where* $(1/N)\mathbb{Z} := \{m/N : m \in \mathbb{Z}\}$

(b) $c \in \mathbb{Q}$并且 $\nu_{L,\ U}(\{(p_j, j \in L) \times (q_j, j \in U)\}) \in \mathbb{Q}$对每个 $(p_j, j \in L) \times (q_j, j \in U) \in \mathbb{R}^{|L|+|U\ \ |}$.

显然，每个 $\mathcal{G}_N$ 都是可数的，因此 $\mathcal{G} := \cup_{N\geqslant 1}\mathcal{G}_N$ 也可数。由于 $\hat{f}^{d,\ n\ \overset{s}{V},}$ 由有限支撑的离散有符号测度构造而成，因此很容易证明，存在一个序列 $\{g_N\}_{N\geqslant 1}$，满足$g_N \in \mathcal{G}_N \subseteq \mathcal{G}$，使得 $g_N(\mathbf{x}) \to \hat{f}^{d,\ n\ \ ^s,}_{V}(\mathbf{x})$当 $N \to \infty$ 时，对每个$\mathbf{x} \in \mathbb{R}^d$都成立。因此，如果 $\hat{f}^{d,\ n\ \ ^s,\ V} \in \mathcal{F}_j$，那么

$$\sup_{f\in\mathcal{G}}\big(M_n(f) - M_n(f_{0,\mathbf{M}})\big) \geqslant \lim_{N\to\infty}\big(M_n(g_N) - M_n(f_{0,\mathbf{M}})\big) = M_n(\hat{f}^{d,s}_{n,V}) - M_n(f_{0,\mathbf{M}}) \geqslant 0.$$

因此，

$$\begin{aligned}\mathbb{P}\big(\hat{f}^{d,s}_{n,V} \in \mathcal{F}_j\big) &\leqslant \mathbb{P}\Big(\sup_{f\in\mathcal{G}}\big(M_n(f) - M_n(f_{0,\mathbf{M}})\big) \geqslant 0\Big) \\ &\leqslant \sup\Big\{\mathbb{P}\Big(\sup_{f\in\mathcal{H}}\big(M_n(f) - M_n(f_{0,\mathbf{M}})\big) \geqslant 0\Big) : \mathcal{H} \subseteq \mathcal{F}_j \text{ is countable}\Big\}.\end{aligned}$$

A.3.2 引理2的证明

引理 *2* 的证明。对于每个 $f^{d,\ c\ \ ^s,\ \{\nu_{L,\ U}\}} \in B(V,\ t)$，通过按照 (40) 修改每个基函数 $b^{L,\ U\mathbf{l},\ \mathbf{u}}$，我们可以将 $f^{d,\ c\ \ ^s,\ \{\nu_{L,\ U}\}}$表示为

$$f^{d,s}_{c,\{\nu_{L,U}\}}(x_1,\ldots,x_d) = f^{d,s}_{b,\{\mu_S\}}(x_1,\ldots,x_d) := b + \sum_{0<|S|\leqslant s}\int_{\mathbb{R}^{|S|}}\prod_{j\in S}\mathbf{1}(x_j \geqslant l_j)\, d\mu_S(l_j, j \in S) \tag{61}$$

对于某个 $b \in \mathbb{R}$ 以及定义在 $\mathbb{R}^{|S|}$ 上、与原测度 $\nu_{L,\,U}$ 通过 (42) 相关联的有限符号博雷尔测度 μ_S。这里，求和遍历所有满足 $|S| \leqslant s$ 的非空子集 $S \subseteq [d]$。由于每个 $\nu_{L,\,U}$ 都支撑于 $\prod_{j\in L}(-M_j/2, M_j/2] \times \prod_{j\in U}(-M_j/2, M_j/2],$，关系 (42) 表明每个 μ_S 都支撑于 $\prod_{j\in S}(-M_j/2, M_j/2]$。此外，根据 (43),

$$\sum_{0<|S|\leqslant s} \|\mu_S\|_{\mathrm{TV}} \leqslant \min(2^s - 1, 2^d) \cdot \sum_{0<|L|+|U|\leqslant s} \|\nu_{L,U}\|_{\mathrm{TV}} \leqslant (2^s - 1)V \leqslant C_s V.$$

因此，若我们将 $\widetilde{B}(V,\ t)$ 定义为所有满足以下条件的函数 $f^{d,\ b}{}^{s}{}_{,\ \{\mu_S\}} \in \mathcal{F}^{d}{}_{,\ \infty\text{-ST}}^{s}$ 的集合：$\|f^{d,\ b}{}^{s}{}_{,\ \{\mu_S\}}\|_{p_0,\ 2} \leqslant t$ 使得 μ_S 支撑于 $\prod_{j\in S}(-M_j/2, M_j/2]$ 并满足

$$\sum_{0<|S|\leqslant s} \|\mu_S\|_{\mathrm{TV}} \leqslant V,$$

~那么我们有 $B(V,\ t) \subseteq B(C_s V,\ t)$。

~现在我们将 $B(V,\ t)$ 拆分成若干部分，计算每一部分的括区熵，然后将它们组合起来~，从而得到 $B(V,\ t)$的括区熵界，进而得到 $B(V,\ t)$的括区熵界。对于每个 $f^{d,\ b}{}^{s}{}_{,\ \{\mu_S\}} \in \widetilde{B}(V,\ t)$，通过重复定理 4 证明中的论证（使用柯西不等式），我们可以证明，对于某个常数 $C > 0$，有 $|b| \leqslant C(V + t)$。令 $K = \lfloor C(V + t)/\epsilon \rfloor,$，并且对于每个 $k = -(K + 1),\ \ldots,\ K$，令 $\mathcal{G}_k$ 表示形如（61）的所有满足 $k\epsilon \leqslant b \leqslant (k + 1)\epsilon$ 的函数 $f^{d,\ b}{}^{s}{}_{,\ \{\mu_S\}}$ 的集合，其中有符号博雷尔测度 μ_S 支撑于 $\prod_{j\in S}(-M_j/2, M_j/2]$ 且满足 $\sum_{0<|S|\leqslant s} \|\mu_S\|_{\mathrm{TV}} \leqslant V$。显然,

$$\widetilde{B}(V,t) \subseteq \bigcup_{k=-(K+1),\ldots,K} \mathcal{G}_k,$$

因此,

$$\begin{aligned}\log N_{[\ ]}(\epsilon, \widetilde{B}(V,t), \|\cdot\|_{p_0,2}) &\leqslant \log\Big(\sum_{k=-(K+1),\ldots,K} N_{[\ ]}(\epsilon, \mathcal{G}_k, \|\cdot\|_{p_0,2})\Big) \qquad (62)\\ &\leqslant \log\Big(2 + \frac{C(V+t)}{\epsilon}\Big) + \sup_k \log N_{[\ ]}(\epsilon, \mathcal{G}_k, \|\cdot\|_{p_0,2}) \leqslant \log\Big(2 + \frac{C(V+t)}{\epsilon}\Big) + \log N_{[\ ]}(\epsilon, \mathcal{G}_0, \|\cdot\|_{p_0,2}).\end{aligned}$$

现在，令 $\mathcal{G}_\varnothing$ 表示取值于 $[0$ 的 $\mathbb{R}^d$ 上所有常值函数的集合，$\epsilon]$。此外，对于每个满足 $|S| \leqslant s$ 的非空 $S \subseteq [d]$，定义 $\mathcal{G}_S$ 为 $\mathbb{R}^d$ 上所有形如以下形式的函数的集合

$$(x_1, \ldots, x_d) \mapsto \int_{\prod_{j\in S}[-M_j/2, M_j/2]} \prod_{j\in S} \mathbf{1}(x_j \geqslant l_j)\, d\mu_S(l_j, j \in S),$$

其中 μ_S 是 $\prod_{j\in S}[-M_j/2, M_j/2]$ 上满足 $\|\mu_S\|_{\mathrm{TV}} \leqslant V$ 的有限符号博雷尔测度。根据定义,

$$\mathcal{G}_0 \subseteq \mathcal{G}_\varnothing + \bigdoublebarwedge_{0<|S|\leqslant s} \mathcal{G}_S,$$

其中 $A + B = \{a + b :\ a \in A,\ b \in B\}$。由此可得

$$\log N_{[\ ]}(\epsilon, \mathcal{G}_0, \|\cdot\|_{p_0,2}) \leqslant C_s(1 + \log d) + \sum_{0<|S|\leqslant s} \log N_{[\ ]}\big(\epsilon/(C_s d^{\bar{s}}), \mathcal{G}_S, \|\cdot\|_{p_0,2}\big), \qquad (63)$$

其中 $\bar{s} = \min(s,\ d)$。

~接下来，对于每个非空的 $S \subseteq [d]$ 且满足 $|S| \leqslant s$ 的情况，令 $\mathcal{G}_S$ 表示原始定义域 $\mathbb{R}^d$ 在截断区间 $\prod_{j\in S}[-M_j/2, M_j/2]$ 上、通过限制到由 S 编号的坐标而得到的所有函数的集合，其形式为

$$(x_j, j \in S) \mapsto \int_{\prod_{j\in S}[-M_j/2, M_j/2]} \prod_{j\in S} \mathbf{1}(x_j \geqslant l_j)\, d\mu_S(l_j, j \in S),$$

其中， μ_S 的定义如 $\mathcal{G}_S$中所述。由于 p_0 由 $B/\prod_{j=1}^d M_j$, 一致有界，因此我们有

$$N_{[\ ]}(\epsilon, \mathcal{G}_S, \|\cdot\|_{p_0,2}) \leqslant N_{[\ ]}\Big(\Big(\frac{\prod_{j\in S} M_j}{B}\Big)^{1/2}\epsilon, \widetilde{\mathcal{G}}_S, \|\cdot\|_2\Big). \tag{64}$$

~此外，对于每个满足 $|S| \leqslant s$的非空 $S \subseteq [d]$ ，令 $\mathcal{G}_S$ 表示缩放后的定义域 $[0,\ 1]^{|S|}$ 上所有具有以下形式的函数的集合

$$(x_j, j \in S) \mapsto \int_{[0,1]^{|S|}} \prod_{j\in S} \mathbf{1}(x_j \geqslant l_j)\, d\mu_S(l_j, j \in S),$$

其中， μ_S 是定义在 $[0,\ 1]^{|S|}$ 上的有限符号博雷尔测度，且满足 $\|\mu_S\|_{\mathrm{TV}} \leqslant V$。通过一个简单的缩放论证，可以很容易地验证

$$N_{[\ ]}(\epsilon, \widetilde{\mathcal{G}}_S, \|\cdot\|_2) \leqslant N_{[\ ]}\Big(\Big(\frac{1}{\prod_{j\in S} M_j}\Big)^{1/2}\epsilon, \bar{\mathcal{G}}_S, \|\cdot\|_2\Big) = N_{[\ ]}\Big(\Big(\frac{1}{\prod_{j\in S} M_j}\Big)^{1/2}\epsilon, \bar{\mathcal{G}}_{[|S|]}, \|\cdot\|_2\Big). \tag{65}$$

接下来，对于每个整数 $m \geqslant 1$ 和 $R > 0$，令 $\mathcal{H}_m(R)$ 表示定义在 $[0,\ 1]^m$ 上所有具有以下形式的函数的集合

$$(x_1, \ldots, x_m) \mapsto \int_{[0,1]^m} \prod_{j=1}^m \mathbf{1}(x_j \geqslant l_j)\, d\mu(l_1, \ldots, l_m)$$

其中， μ 是定义在 $[0,\ 1]^m$ 上的有限博雷尔测度（非符号测度），且满足 $\|\mu\|_{\mathrm{TV}} \leqslant R$。Gao 在[15,定理 1.1]中证明了

$$\log N_{[\ ]}(\epsilon, \mathcal{H}_m(R), \|\cdot\|_2) \leqslant C_m\Big(2 + \frac{R}{\epsilon}\Big)\Big[\log\Big(2 + \frac{R}{\epsilon}\Big)\Big]^{2(m-1)}.$$

根据有符号测度的 Jordan分解,

$$\bar{\mathcal{G}}_{[|S|]} \subseteq \mathcal{H}_{|S|}(V) - \mathcal{H}_{|S|}(V),$$

其中 $A - B = \{a - b\colon\ a \in A,\ b \in B\}$。由此可得

$$\log N_{[\ ]}(\epsilon, \bar{\mathcal{G}}_{[|S|]}, \|\cdot\|_2) \leqslant 2\log N_{[\ ]}(\frac{\epsilon}{2}, \mathcal{H}_{|S|}(V), \|\cdot\|_2) \leqslant C_{|S|}\Big(2 + \frac{2V}{\epsilon}\Big)\Big[\log\Big(2 + \frac{2V}{\epsilon}\Big)\Big]^{2(|S|-1)}.$$

将此界依次代回（65）、（64）、（63）和（62），可得

$$\log N_{[\ ]}(\epsilon, \widetilde{B}(V,t), \|\cdot\|_{p_0,2}) \leqslant \log\Big(2 + \frac{C(V+t)}{\epsilon}\Big) + C_{B,s}d^{2\bar{s}}(1+\log d)^{2(\bar{s}-1)}\Big(2 + \frac{V}{\epsilon}\Big)\Big[\log\Big(2 + \frac{V}{\epsilon}\Big)\Big]^{2(\bar{s}-1)}.$$

~最后，由于 $B(V,\ t) \subseteq B(C_sV,\ t)$, 我们得到

$\log\ N_{[\ ]}(\epsilon,\ B(V,\ t),\ \|\cdot\|_{p_0,\ 2}) \leqslant \log\Big(2 + \dfrac{C_s(V+t)}{\epsilon}\Big) + C_{B,\ s}d^{2\bar{s}}(1+\log\ d)^{2(\bar{s}-1)}\Big(2 + \dfrac{V}{\epsilon}\Big)\Big[\log\Big(2 + \dfrac{V}{\epsilon}\Big)\Big]^{2(\bar{s}-1)}$，由此完成证明。

□

A.3.3 Corollary 1 的证明

推论 *1* 的证明。 从定理 4 的证明可知，$\hat{}f^{d,\ s}_{n\ V}$的风险界关于 连续。因此，推论中所需的界可直接由 $\hat{}$，$^{s}_{n\ V}$的风险界以及以下不等式推出：

$$\mathfrak{M}^{d,s}_{n,V} \leqslant \liminf_{\epsilon\to 0+} \sup_{\substack{f^*\in\mathcal{F}^{d,s}_{\infty-\mathrm{ST}} \\ V^{d,s}_{\infty-\mathrm{XGB}}(f^*)\leqslant V}} \mathbb{E}\|\hat{f}^{d,s}_{n,V+\epsilon} - f^*\|^2_{p_0,2}.$$

□

A.3.4 定理 5 的证明

我们对定理 5 的证明建立在 Fang 等人 [12, 定理 46], 的证明思路之上，而后者本身又受到 Blei 等人 [4, 第 4] 节中相关思想的启发。与 Fang 等人 [12, 定理 46], 一样，我们以如下形式使用 Assouad 引理。

引理 10 （van der Vaart 的引理 24.3 [35] 和 Ki 等人的引理 11.20 [21]）。 设 q 为正整数，并且有 $f_{\boldsymbol{\eta}} \in \mathcal{F}^{d,\ ^{s}_{\infty-ST}}$其中 $V^{d,\ ^{s}_{\infty-XGB}}(f_{\boldsymbol{\eta}}) \leqslant V$对每个 $\boldsymbol{\eta} \in \{-1$和 $1\}^q$均成立。于是，对于极小极大风险 $\mathfrak{M}^{d,\ ^{s}_{n\ V}}$，我们有如下下界：

$$\mathfrak{M}^{d,s}_{n,V} \geqslant \frac{q}{8}\cdot \min_{\boldsymbol{\eta}\neq\boldsymbol{\eta}'} \frac{\|f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'}\|^2_{p_0,2}}{H(\boldsymbol{\eta},\boldsymbol{\eta}')} \cdot \min_{H(\boldsymbol{\eta},\boldsymbol{\eta}')=1} \left(1 - \sqrt{\frac{1}{2}\mathbb{E}\big[K(\mathbb{P}_{f_{\boldsymbol{\eta}}}, \mathbb{P}_{f_{\boldsymbol{\eta}'}})\big]}\right).$$

这里， $H(\cdot$和 $\cdot)$表示汉明距离 $H(\boldsymbol{\eta}$、 $\boldsymbol{\eta}'): = \sum_{j=1}^{q} 1\{\eta_j \neq \eta'_j\}$和 $\mathbb{P}_f$表示在 $(\mathbf{x}^{(1)}$、…、$\mathbf{x}^{(n)})$给定 $f^* = f$ 时， $(y_1$、…、 $y_n)$的概率分布，而 $K(\cdot$和 $\cdot)$表示两个概率分布之间的 *Kullback* 散度。

定理 *5* 的证明。取一个整数 l 作为

$$l = \left\lceil \frac{1}{3\log 2}\Big\{ \log\Big(\frac{C_{B,\bar{s}} n V^2}{\sigma^2}\Big) - (\bar{s}-1)\log\log\Big(\frac{C_{B,\bar{s}} n V^2}{\sigma^2}\Big)\Big\}\right\rceil$$

其中 $C_{B,\ \bar{s}} = B2^{-4\bar{s}+1}(6\log\ 2)^{\bar{s}-1}\cdot(\bar{s}-1)!$ 且 $\lceil x\rceil$ 表示大于或等于 x的最小整数。这样选取 l 可确保

$$2^{-l} \leqslant \Big(\frac{\sigma^2}{C_{B,\bar{s}} n V^2}\Big)^{1/3}\left[\log\Big(\frac{C_{B,\bar{s}} n V^2}{\sigma^2}\Big)\right]^{(\bar{s}-1)/3} < 2^{-l+1}. \tag{66}$$

定义

$$P_l = \left\{(p_1,\ldots,p_{\bar{s}}) \in \mathbb{Z}^{\bar{s}}_{\geqslant 0} : \sum_{j=1}^{\bar{s}} p_j = l\right\},$$

并且，对于每个 $\mathbf{p} = (p_1,\ldots,p_{\bar{s}}) \in P_l$，令

$$I_{\mathbf{p}} = \big\{(i_1,\ldots,i_{\bar{s}}) : i_j \in [2^{p_j}] \text{ for each } j \in [\bar{s}]\big\}.$$

回顾一下，对于每个整数 $m \geqslant 1$，都有 $[m] = \{1,\ldots,m\}$。显然，对于每个 $\mathbf{p} \in P_l$，都有 $|I_{\mathbf{p}}| = 2^l$，并且

$$|P_l| = \binom{\bar{s}+l-1}{\bar{s}-1} \geqslant \frac{l^{\bar{s}-1}}{(\bar{s}-1)!}.$$

接下来，定义

$$Q = \big\{(\mathbf{p},\mathbf{i}) : \mathbf{p} \in P_l \text{ and } \mathbf{i} \in I_{\mathbf{p}}\big\}$$

并令 $q = |Q| = |P_l| \cdot 2^l$。在这个证明中，函数将由向量 $\boldsymbol{\eta} \in \{-1,\ 1\}^q$ 作为索引，其分量由集合 Q 索引。

对于整数 $m \geqslant 1$ 和 $k \in [2^m]$，记 $\psi_{m,k}$ 为定义在 $(0,1)$ 上的实值函数，其定义如下：

$$\psi_{m,k}(x) = \begin{cases} 1 & \text{if } x \in \big((k-1)2^{-m}, (k-3/4)2^{-m}\big) \cup \big((k-1/4)2^{-m}, k2^{-m}\big), \\ -1 & \text{if } x \in \big((k-3/4)2^{-m}, (k-1/4)2^{-m}\big), \\ 0 & \text{otherwise.} \end{cases}$$

利用这些函数 $\psi_{m,k}$，我们构造 $f_{\boldsymbol{\eta}} \in \mathcal{F}^{d,\ {}^{s}}_{\infty-\mathrm{ST}}$，其中 $\boldsymbol{\eta} \in \{-1,1\}^q$，如下所示，并将用其结合引理 10 证明极小极大风险 $\mathfrak{M}^{d,\ {}^{s}}_{n,V}$的下界。对于每个 $\boldsymbol{\eta} \in \{-1,1\}^q$，令 $\nu_{\boldsymbol{\eta}}$为定义在 $\prod_{j=1}^{\bar{s}}(-M_j/2, M_j/2)$ 上的有符号博雷尔测度，其定义如下：

$$d\nu_{\boldsymbol{\eta}}(\mathbf{t}) = \frac{1}{M_1 \cdots M_{\bar{s}}} \cdot \frac{V}{\sqrt{|P_l|}} \sum_{\mathbf{p} \in P_l} \sum_{\mathbf{i} \in I_{\mathbf{p}}} \eta_{\mathbf{p},\mathbf{i}} \Big(\prod_{j=1}^{\bar{s}} \psi_{p_j,i_j}\Big(\frac{t_j}{M_j} + \frac{1}{2}\Big)\Big)\, d\mathbf{t},$$

并定义 $f_{\boldsymbol{\eta}} : \mathbb{R}^d \to \mathbb{R}$ 如下：

$$f_{\boldsymbol{\eta}}(x_1, \ldots, x_d) = \int_{\prod_{j=1}^{\bar{s}}(-M_j/2, M_j/2)} \prod_{j=1}^{\bar{s}} \mathbf{1}(x_j \geqslant t_j)\, d\nu_{\boldsymbol{\eta}}(\mathbf{t}).$$

显然，对于每个 $\boldsymbol{\eta} \in \{-1,1\}^q$，都有 $f_{\boldsymbol{\eta}} \in \mathcal{F}^{d,\ {}^{s}}_{\infty-\mathrm{ST}}$。以下引理概述了证明中所需函数 $f_{\boldsymbol{\eta}}$ 的关键性质，其证明见附录 A.5.4。

引理 11. 对于每个 $\boldsymbol{\eta} \in \{-1,1\}^q$，$f_{\boldsymbol{\eta}}$ 的复杂度有上界 V，即：

$$V^{d,s}_{\infty-XGB}(f_{\boldsymbol{\eta}}) \leqslant V. \tag{67}$$

我们还有

$$\max_{H(\boldsymbol{\eta},\boldsymbol{\eta}')=1} \|f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'}\|^2_{p_0,2} \leqslant \frac{BV^2}{|P_l|} \cdot 2^{-3l-4\bar{s}+2} \tag{68}$$

and

$$\min_{\boldsymbol{\eta} \neq \boldsymbol{\eta}'} \frac{\|f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'}\|^2_{p_0,2}}{H(\boldsymbol{\eta},\boldsymbol{\eta}')} \geqslant \frac{bV^2}{|P_l|} \cdot 2^{-3l-6\bar{s}+2}. \tag{69}$$

不难验证，对于 $\boldsymbol{\eta}, \boldsymbol{\eta}' \in \{-1,1\}^q$，$\mathbb{P}_{f_{\boldsymbol{\eta}}}$ 和 $\mathbb{P}_{f_{\boldsymbol{\eta}'}}$ 之间的 Kullback 散度可以计算为

$$K(\mathbb{P}_{f_{\boldsymbol{\eta}}}, \mathbb{P}_{f_{\boldsymbol{\eta}'}}) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} \big(f_{\boldsymbol{\eta}}(\mathbf{x}^{(i)}) - f_{\boldsymbol{\eta}'}(\mathbf{x}^{(i)})\big)^2.$$

因此，由 (68) 得到

$$\max_{H(\boldsymbol{\eta},\boldsymbol{\eta}')=1} \mathbb{E}\big[K(\mathbb{P}_{f_{\boldsymbol{\eta}}}, \mathbb{P}_{f_{\boldsymbol{\eta}'}})\big] = \frac{n}{2\sigma^2} \cdot \max_{H(\boldsymbol{\eta},\boldsymbol{\eta}')=1} \|f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'}\|^2_{p_0,2} \leqslant \frac{BnV^2}{\sigma^2|P_l|} \cdot 2^{-3l-4\bar{s}+1}. \tag{70}$$

Ap应用引理 10，并结合 (69) 和 (70)，我们可以将极小极大风险 $\mathfrak{M}^{d,\ {}^{s}}_{n,V}$ 界定为

$$\begin{aligned} \mathfrak{M}^{d,s}_{n,V} &\geqslant \frac{q}{8} \cdot \frac{bV^2}{|P_l|} \cdot 2^{-3l-6\bar{s}+2} \left(1 - \sqrt{\frac{BnV^2}{\sigma^2|P_l|} \cdot 2^{-3l-4\bar{s}}}\right) \\ &\geqslant bV^2 2^{-2l-6\bar{s}-1} \left(1 - \sqrt{\frac{1}{2(6\log 2)^{\bar{s}-1}} \cdot \frac{C_{B,\bar{s}} nV^2}{\sigma^2} \cdot \frac{2^{-3l}}{l^{\bar{s}-1}}}\right). \end{aligned}$$

回顾一下， $q=|P_l|\cdot 2^l$, $|P_l|\geqslant l^{\bar{s}-1}/(\bar{s}-1)!$ ，以及 $C_{B,\ \bar{s}}=B2^{-4\bar{s}+1}(6\log\ 2)^{\bar{s}-1}\cdot(\bar{s}-1)!$ 我们在证明开头对 l 的选择意味着

$$\begin{aligned}\frac{1}{2(6\log 2)^{\bar{s}-1}}\cdot\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\cdot\frac{2^{-3l}}{l^{\bar{s}-1}}&\leqslant\frac{1}{2}\cdot\left[\frac{\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)}{2\Big(\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)-(\bar{s}-1)\log\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)\Big)}\right]^{\bar{s}-1}\\&\leqslant\frac{1}{2}\cdot\left[2\left(1-(\bar{s}-1)\frac{\log\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)}{\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)}\right)\right]^{-(\bar{s}-1)}\\&\leqslant\frac{1}{2}\cdot\left[2\left(1-(\bar{s}-1)\Big\{\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)\Big\}^{-\frac{1}{2}}\right)\right]^{-(\bar{s}-1)}.\end{aligned}$$

这里，第一个不等式由（66）可得，

$$l\geqslant\frac{1}{3\log 2}\left\{\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)-(\bar{s}-1)\log\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)\right\},$$

最后一个不等式则来自不等式 $\log\log\ x/\log\ x\leqslant(\log\ x)^{-1/2},$ ，该不等式对所有 $x>1$ 均成立。如果我们假设

$$n\geqslant\frac{e^{4\bar{s}^2}}{C_{B,\bar{s}}}\cdot\frac{\sigma^2}{V^2},$$

then

$$\frac{1}{2(6\log 2)^{\bar{s}-1}}\cdot\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\cdot\frac{2^{-3l}}{l^{\bar{s}-1}}\leqslant\frac{1}{2}\cdot\left(1+\frac{1}{\bar{s}}\right)^{-(\bar{s}-1)}\leqslant\frac{1}{2},$$

从而，我们有

$$\begin{aligned}\mathfrak{M}_{n,V}^{d,s}&\geqslant\left(1-\sqrt{\frac{1}{2}}\right)\cdot bV^2 2^{-2l-6\bar{s}-1}\geqslant\left(1-\sqrt{\frac{1}{2}}\right)\cdot bV^2 2^{-6\bar{s}-3}\cdot\left(\frac{\sigma^2}{C_{B,\bar{s}}nV^2}\right)^{2/3}\left[\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)\right]^{2(\bar{s}-1)/3}\\&\geqslant C_{b,B,\bar{s}}\left(\frac{\sigma^2V}{n}\right)^{2/3}\left[\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)\right]^{2(\bar{s}-1)/3},\end{aligned}$$

其中 $C_{b、\ B、\ \bar{s}}=(1-\ 1/\sqrt{2})\cdot b2^{-6\bar{s}-3}C_{B,\bar{s}}^{-2/3}$ 。这里，第二个不等式再次使用了（66）。最后，由于

$$\log\left(\frac{C_{B,\bar{s}}nV^2}{\sigma^2}\right)\geqslant\frac{1}{2}\log\left(\frac{nV^2}{\sigma^2}\right)$$

在进一步假设的条件下， $n\geqslant(1/C_{B,\bar{s}}^2)\cdot(\sigma^2/V^2)$,

$$n\geqslant\max\left\{\frac{e^{4\bar{s}^2}}{C_{B,\bar{s}}}\cdot\frac{\sigma^2}{V^2},\frac{1}{C_{B,\bar{s}}^2}\cdot\frac{\sigma^2}{V^2}\right\},$$

我们可以推导出下界

$$\mathfrak{M}_{n,V}^{d,s}\geqslant C'_{b,B,\bar{s}}\left(\frac{\sigma^2V}{n}\right)^{2/3}\left[\log\left(\frac{nV^2}{\sigma^2}\right)\right]^{2(\bar{s}-1)/3}$$

其中 $C'_{b,\ B,\ \bar{s}}=C_{b,\ B,\ \bar{s}}\cdot 2^{-2(\bar{s}-1)/3}$。 □

A.4 第6节中命题和引理的证明

A.4.1 命题7的证明

命题 7 的证明。假设 $f_{\mathbf{a}}\in\mathcal{F}^{d,\overset{s}{\infty-\mathrm{ST}}}$ 对于 $\mathbf{a}\in\{-\infty,+\infty\}^d$ 以及 $\sum_{\mathbf{a}\in\{-\infty,+\infty\}^d}f_{\mathbf{a}}\equiv f$ 成立。重复命题 5 证明中的论证可知，对于每个 $g\in\mathcal{F}^{d,\overset{s}{\infty-\mathrm{ST}}}$,

$$\widetilde{V}_{\infty-\mathrm{XGB}}^{d,s}(g)\leqslant V_{\mathbf{a}}(g)=\mathrm{HK}_{\mathbf{a}}(g),$$

其中 $V_{\mathbf{a}}(\cdot)$ 的定义与（44）中相同。将此不等式应用于每个 $f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$，得到

$$\sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} \mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) \geqslant \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} \widetilde{V}^{d,s}_{\infty-\mathrm{XGB}}(f_{\mathbf{a}}) \geqslant \widetilde{V}^{d,s}_{\infty-\mathrm{XGB}}(f),$$

这就证明了所需等式的一个方向：

$$\widetilde{V}^{d,s}_{\infty-\mathrm{XGB}}(f) \leqslant \inf\Big\{ \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} \mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) : \sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} f_{\mathbf{a}} \equiv f,\ f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\ \forall \mathbf{a} \Big\}.$$

现在我们来证明反向不等式。假设 $f \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$ 表示为

$$f(x_1,\dots,x_d) = c + \sum_{\substack{L,U:L\cap U=\varnothing \\ 0<|L|+|U|\leqslant s}} \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l},\mathbf{u})$$

对于某些 $c \in \mathbb{R}$ 以及 signed Borel 测度 $\nu_{L,U}$，定义在 $\mathbb{R}^{|L|+|U|}$ 上。对于每个 $\mathbf{a} \in \{-\infty, +\infty\}^d$，定义 $f_{\mathbf{a}} \colon \mathbb{R}^d \to \mathbb{R}$，使得

$$f_{\mathbf{a}}(x_1,\dots,x_d) = c \cdot \prod_{j=1}^d \mathbf{1}(a_j = -\infty) + \sum_{\substack{L,U:L\cap U=\varnothing \\ 0<|L|+|U|\leqslant s}} \prod_{j\in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j\in U} \mathbf{1}(a_j = +\infty)$$
$$\cdot \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l},\mathbf{u}).$$

显然，对于所有$\mathbf{a} \in \{-\infty, +\infty\}^d$，都有 $f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}$。此外，由于对 $\nu_{L,U}$ 上的每个积分，都恰有一个 $\mathbf{a}$ 的取值能使所有相乘的指示函数都等于一，因此我们有

$$\sum_{\mathbf{a}\in\{-\infty,+\infty\}^d} f_{\mathbf{a}} \equiv f.$$

固定$\mathbf{a} \in \{-\infty, +\infty\}^d$。对于每个非空的 $S \subseteq [d]$，我们有

$$f^S_{(a_j, j\in S^c)}(x_j, j\in S) = c \cdot \prod_{j=1}^d \mathbf{1}(a_j = -\infty) + \sum_{\substack{L,U\subseteq S:L\cap U=\varnothing \\ 0<|L|+|U|\leqslant s}} \prod_{j\in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j\in U} \mathbf{1}(a_j = +\infty)$$
$$\cdot \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l},\mathbf{u}).$$

因此，对于每个非空的 $T \subseteq [d]$，

$$\begin{aligned}
&\sum_{S\subseteq T} (-1)^{|T|-|S|} \cdot f^S_{(a_j, j\in S^c)}(x_j, j\in S) \\
&= \sum_{S\subseteq T} (-1)^{|T|-|S|} \sum_{\substack{L,U\subseteq S:L\cap U=\varnothing \\ 0<|L|+|U|\leqslant s}} \prod_{j\in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j\in U} \mathbf{1}(a_j = +\infty) \\
&\qquad \cdot \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l},\mathbf{u}) \\
&= \sum_{\substack{L,U:L\cap U=\varnothing \\ 0<|L|+|U|\leqslant s}} \Big(\sum_{S:L\cup U\subseteq S\subseteq T} (-1)^{|T|-|S|} \Big) \prod_{j\in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j\in U} \mathbf{1}(a_j = +\infty) \\
&\qquad \cdot \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l},\mathbf{u}) \\
&= \sum_{\substack{L,U:L\cap U=\varnothing, L\cup U=T \\ 0<|L|+|U|\leqslant s}} \prod_{j\in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j\in U} \mathbf{1}(a_j = +\infty) \cdot \int_{\mathbb{R}^{|L|+|U|}} \prod_{j\in L} \mathbf{1}(x_j \geqslant l_j) \cdot \prod_{j\in U} \mathbf{1}(x_j < u_j)\, d\nu_{L,U}(\mathbf{l},\mathbf{u})
\end{aligned}$$

若 $|T| \leqslant s$，否则它为零。对于每个满足 $|T| \leqslant s$ 的非空 $T \subseteq [d]$，由于至多存在一对 $(L,\ U)$ 满足 $L \cap U = \varnothing$ 和 $L \cup U = T$，使得

$$\prod_{j \in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j \in U} \mathbf{1}(a_j = +\infty) = 1,$$

重复命题 5 证明中的计算，可得

$$\begin{aligned}\mathrm{Vit}(f^T_{(a_j, j \in T^c)}) &= \mathrm{Vit}\Big((x_j, j \in T) \mapsto \sum_{S \subseteq T} (-1)^{|T|-|S|} \cdot f^S_{(a_j, j \in S^c)}(x_j, j \in S)\Big) \\ &= \sum_{\substack{L,U: L \cap U = \varnothing, L \cup U = T \\ 0 < |L| + |U| \leqslant s}} \prod_{j \in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j \in U} \mathbf{1}(a_j = +\infty) \cdot \|\nu_{L,U}\|_{\mathrm{TV}}.\end{aligned}$$

因此，

$$\begin{aligned}\mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) &= \sum_{0 < |T| \leqslant d} \mathrm{Vit}(f^T_{(a_j, j \in T^c)}) = \sum_{0 < |T| \leqslant s} \mathrm{Vit}(f^T_{(a_j, j \in T^c)}) \\ &= \sum_{0 < |T| \leqslant s} \sum_{\substack{L,U: L \cap U = \varnothing, L \cup U = T \\ 0 < |L| + |U| \leqslant s}} \prod_{j \in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j \in U} \mathbf{1}(a_j = +\infty) \cdot \|\nu_{L,U}\|_{\mathrm{TV}} \\ &= \sum_{\substack{L,U: L \cap U = \varnothing \\ 0 < |L| + |U| \leqslant s}} \prod_{j \in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j \in U} \mathbf{1}(a_j = +\infty) \cdot \|\nu_{L,U}\|_{\mathrm{TV}}.\end{aligned}$$

将上述恒等式对所有 $\mathbf{a} \in \{-\infty, +\infty\}^d$ 求和，得到

$$\begin{aligned}&\inf\Big\{ \sum_{\mathbf{a} \in \{-\infty,+\infty\}^d} \mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) : \sum_{\mathbf{a} \in \{-\infty,+\infty\}^d} f_{\mathbf{a}} \equiv f,\ f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\ \forall \mathbf{a} \Big\} \leqslant \sum_{\mathbf{a} \in \{-\infty,+\infty\}^d} \mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) \\ &= \sum_{\substack{L,U: L \cap U = \varnothing \\ 0 < |L| + |U| \leqslant s}} \sum_{\mathbf{a} \in \{-\infty,+\infty\}^d} \Big(\prod_{j \in U^c} \mathbf{1}(a_j = -\infty) \cdot \prod_{j \in U} \mathbf{1}(a_j = +\infty) \Big) \cdot \|\nu_{L,U}\|_{\mathrm{TV}} = \sum_{\substack{L,U: L \cap U = \varnothing \\ 0 < |L| + |U| \leqslant s}} \|\nu_{L,U}\|_{\mathrm{TV}}.\end{aligned}$$

对 f 的所有可能表示 $f^{d,\ s}_{,c\ ,\{\nu_{L,U}\}}$ 取下确界，得到

$$\inf\Big\{ \sum_{\mathbf{a} \in \{-\infty,+\infty\}^d} \mathrm{HK}_{\mathbf{a}}(f_{\mathbf{a}}) : \sum_{\mathbf{a} \in \{-\infty,+\infty\}^d} f_{\mathbf{a}} \equiv f,\ f_{\mathbf{a}} \in \mathcal{F}^{d,s}_{\infty-\mathrm{ST}}\ \forall \mathbf{a} \Big\} \leqslant \widetilde{V}^{d,s}_{\infty-\mathrm{XGB}}(f),$$

由此完成证明。 □

A.4.2 引理3的证明

引理 *3* 的证明。设 $f \in \mathcal{F}^{d,\ s}_{\mathrm{ST}}$ 以及 $f = \sum_{k=1}^K f_k$，其中每个 f_k 都是具有右连续分裂且深度至多为 s的回归树。令$\mathbf{w}_k$ 表示 f_k的叶权重向量。

固定一个整数 $L \geqslant 1$。对于每个 k，定义回归树 $g_{k,\ L} := (1/L) f_k$, ，其通过将 f_k 的每个叶节点权重缩放 $1/L$ 倍，同时保持相同的树结构而得到。利用这些 $g_{k,\ L}$，我们可以将 f 表示为 KL 棵树之和：

$$f = \sum_{k=1}^K \sum_{l=1}^L g_{k,L}.$$

对于这一表示，叶节点权重的 p^{th} 次幂之和等于

$$\sum_{k=1}^K \sum_{l=1}^L \|(1/L) \cdot \mathbf{w}_k\|_p^p = \frac{1}{L^{p-1}} \sum_{k=1}^K \|\mathbf{w}_k\|_p^p.$$

由于 $p > 1$，当 $L \to \infty$ 时，该量趋于 0。这证明了 $V^{d,\ s}_{,\ \mathrm{XGB}}(f;\ p) = 0.$ □

A.5 附录 A.3 中引理的证明

A.5.1 引理 7 的证明

引理 7 是以下更一般结果的推论，该结果涉及伯恩斯坦范数。对于一个在 $\mathcal{X}$ 上服从分布 P 的随机变量X，以及一个函数 $f: \mathcal{X} \to \mathbb{R}$， f 的伯恩斯坦范数定义为

$$\|f\|_{P,B} = \Big(2\mathbb{E}_P\big[\exp(|f(X)|) - 1 - |f(X)|\big]\Big)^{1/2}.$$

尽管$\|\cdot\|_{P,\ B}$不满足齐次性或三角不等式，因此实际上并不是一个范数，但通常仍称其为范数，并且仍可用来衡量函数的“大小”。

引理 12 (van der Vaart 和 Wellner 的引理 3.4.3 [36])。 假设$\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ 独立同分布，服从 P上的分布，其中 $\mathcal{F}$ 是从 $\mathcal{X}$ 到 $\mathbb{R}$ 的可数函数族，且对所有 $f \in \mathcal{F}$，$\|f\|_{P,\ B} \leqslant \delta$ 。那么，存在一个常数 $C > 0$ ，使得

$$\mathbb{E}_P\Big[\sup_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k} f(\mathbf{x}^{(i)})\Big|\Big] \leqslant C J_{[\ ]}(\delta, \mathcal{F}, \|\cdot\|_{P,B}) \cdot \Big(1 + \frac{J_{[\ ]}(\delta, \mathcal{F}, \|\cdot\|_{P,B})}{\delta^2\sqrt{k}}\Big).$$

引理 7 的证明。 令 P为$\mathbf{x}^{(i)}, \epsilon_i)$ 在 $\mathbb{R}^d \times \{-1, 1\}$ 上的分布，并令 $\mathcal{G}$表示 $\mathbb{R}^d \times \{-1, 1\}$上所有函数 Φ_f的集合，每个 $f \in \mathcal{F}$对应一个函数，定义如下：

$$\Phi_f(\mathbf{x}, \epsilon) = \frac{\epsilon f(\mathbf{x})}{2D} \quad \text{for } (\mathbf{x}, \epsilon) \in \mathbb{R}^d \times \{-1, 1\}.$$

对于每个 $\Phi_f \in \mathcal{G}$，我们有

$$\begin{aligned}\|\Phi_f\|_{P,B} &= \Big(2\mathbb{E}_{(\mathbf{x},\epsilon)\sim P}\Big[\exp\Big(\Big|\frac{\epsilon f(\mathbf{x})}{2D}\Big|\Big) - 1 - \Big|\frac{\epsilon f(\mathbf{x})}{2D}\Big|\Big]\Big)^{1/2} = \Big(2\sum_{m=2}^{\infty}\frac{1}{m!}\cdot\mathbb{E}_{\mathbf{x}\sim p_0}\Big[\Big|\frac{f(\mathbf{x})}{2D}\Big|^m\Big]\Big)^{1/2} \\ &\leqslant \Big(2\sum_{m=2}^{\infty}\frac{1}{m!}\cdot\mathbb{E}_{\mathbf{x}\sim p_0}\Big[\Big|\frac{f(\mathbf{x})}{2D}\Big|^2\Big]\Big)^{1/2} \leqslant \frac{(e-2)^{1/2}t}{2^{1/2}D} := \frac{at}{2D},\end{aligned}$$

其中，第一个不等式利用了这样一个事实： $\|f\|_\infty \leqslant D$；第二个不等式则由以下事实得出：$\|f\|_{p_0,\ 2} \leqslant t$。将引理 12 应用于 $\mathcal{G}$ 和 $\delta = at/2D,$ ，我们得到

$$\begin{aligned}\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i f(\mathbf{x}^{(i)})\Big|\Big] &= 2D\cdot\mathbb{E}\Big[\sup_{\Phi_f\in\mathcal{G}}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\Phi_f(\mathbf{x}^{(i)},\epsilon_i)\Big|\Big] \\ &\leqslant 2D\cdot C J_{[\ ]}\Big(\frac{at}{2D}, \mathcal{G}, \|\cdot\|_{P,B}\Big)\cdot\Big(1 + \frac{J_{[\ ]}(\frac{at}{2D}, \mathcal{G}, \|\cdot\|_{P,B})}{(\frac{at}{2D})^2\sqrt{k}}\Big).\end{aligned} \tag{71}$$

接下来，我们将 $\mathcal{G}$ 在伯恩斯坦范数下的括号熵积分与 $\mathcal{F}$在$\|\cdot\|_{p_0,\ 2}$范数下的括号熵积分联系起来。固定 $\Phi_f \in \mathcal{G}$，令 $[f_1,\ f_2]$为一个包含 f的括号。由于 $\|f\|_\infty \leqslant D$，如有必要，可将 f_1替换为$\mathbf{x} \mapsto \max(\min(f_1(\mathbf{x}),\ D),\ -D)$，从而假设 $\|f_1\|_\infty \leqslant D$。类似地，我们假设 $\|f_2\|_\infty \leqslant D$。定义 $\Phi_1,\ \Phi_2: \mathbb{R}^d \times \{-1,\ 1\} \to \mathbb{R}$ 为

$$\Phi_1(\mathbf{x}, \epsilon) = f_1(\mathbf{x})\cdot\mathbf{1}(\epsilon = 1) - f_2(\mathbf{x})\cdot\mathbf{1}(\epsilon = -1) \quad \text{for } (\mathbf{x}, \epsilon) \in \mathbb{R}^d \times \{-1, 1\}$$

and

$$\Phi_2(\mathbf{x}, \epsilon) = f_2(\mathbf{x})\cdot\mathbf{1}(\epsilon = 1) - f_1(\mathbf{x})\cdot\mathbf{1}(\epsilon = -1) \quad \text{for } (\mathbf{x}, \epsilon) \in \mathbb{R}^d \times \{-1, 1\}.$$

显然，括号 $[\Phi_1,\ \Phi_2]$ 包含 Φ_f。此外，由于 $\|f_2-f_1\|_\infty \leqslant 2D$，根据上述相同的论证，我们得到

$$\|\Phi_2-\Phi_1\|_{P,B}=\left(2\mathbb{E}_{\mathbf{x}\sim p_0}\Big[\exp\Big(\Big|\frac{(f_2-f_1)(\mathbf{x})}{2D}\Big|\Big)-1-\Big|\frac{(f_2-f_1)(\mathbf{x})}{2D}\Big|\Big]\right)^{1/2}\leqslant\frac{a}{2D}\cdot\|f_2-f_1\|_{p_0,2}.$$

由此可得

$$N_{[\,]}\Big(\frac{a\epsilon}{2D},\mathcal{G},\|\cdot\|_{P,B}\Big)\leqslant N_{[\,]}(\epsilon,\mathcal{F},\|\cdot\|_{p_0,2})\quad\text{for }\epsilon>0.$$

因此，

$$\begin{aligned}J_{[\,]}\Big(\frac{at}{2D},\mathcal{G},\|\cdot\|_{P,B}\Big)&=\int_0^{at/2D}\sqrt{1+N_{[\,]}(\epsilon,\mathcal{G},\|\cdot\|_{P,B})}\,d\epsilon=\frac{a}{2D}\int_0^t\sqrt{1+N_{[\,]}\Big(\frac{a\epsilon}{2D},\mathcal{G},\|\cdot\|_{P,B}\Big)}\,d\epsilon\\&\leqslant\frac{a}{2D}\cdot J_{[\,]}(t,\mathcal{F},\|\cdot\|_{p_0,2}).\end{aligned}$$

将此界代入式 (71)，证明即告完成。 □

A.5.2 引理 8 的证明

我们使用以下引理，它保证可以将有符号测度的支撑限制在$\prod_{j\in L}(-M_j/2,M_j/2]\times\prod_{j\in U}(-M_j/2,M_j/2]$内，而不改变函数在 $\prod_{j=1}^d[-M_j/2,M_j/2]$ 上的取值。证明略去，因为可用与引理 6 类似的方法证明。

引理 13。对于每个 $f^{d,s}_{c,\{\nu_{L,U}\}}$，都存在 $f^{d,s}_{b,\{\mu_{L,U}\}}$，使得有限符号博雷尔测度 $\mu_{L,U}$ 支撑于 $\prod_{j\in L}(-M_j/2,M_j/2]\times\prod_{j\in U}(-M_j/2,M_j/2]$，并满足

(a) $f^{d,s}_{b,\{\mu_{L,U}\}}(\cdot)=f^{d,s}_{c,\{\nu_{L,U}\}}(\cdot)$ *on* $\prod_{j=1}^d[-M_j/2,M_j/2]$

(b)

$$\sum_{0<|L|+|U|\leqslant s}\|\mu_{L,U}\|_{TV}\leqslant\sum_{0<|L|+|U|\leqslant s}\|\nu_{L,U}\|_{TV}.$$

引理 8 的证明。假设 $f\in\mathcal{F}^{d,s}_{,\infty-\mathrm{ST}}$ 满足 $V^{d,s}_{,\infty-\mathrm{XGB}}(f)<V$。根据复杂度度量$V^{d,s}_{,\infty-\mathrm{XGB}}(\cdot)$的定义，存在 $f^{d,s}_{,c,\{\nu_{L,U}\}}\in\mathcal{F}^{d,s}_{,\infty-\mathrm{ST}}$使得 $f^{d,s}_{,c,\{\nu_{L,U}\}}\equiv f$并且

$$\sum_{0<|L|+|U|\leqslant s}\|\nu_{L,U}\|_{\mathrm{TV}}\leqslant V.$$

引理 13 还保证存在 $f^{d,s}_{,b,\{\mu_{L,U}\}}$，以及有限符号博雷尔测度 $\mu_{L,U}$，其支撑于 $\prod_{j\in L}(-M_j/2,M_j/2]\times\prod_{j\in U}(-M_j/2,M_j/2]$并满足该引理的条件 (a) 和 (b)。根据条件 (a)，$f^{d,s}_{,b,\{\mu_{L,U}\}}(\cdot)=f^{d,s}_{,c,\{\nu_{L,U}\}}(\cdot)=f(\cdot)$在 $\prod_{j=1}^d[-M_j/2,M_j/2]$ 上。此外，根据条件 (b)，

$$\sum_{0<|L|+|U|\leqslant s}\|\mu_{L,U}\|_{\mathrm{TV}}\leqslant\sum_{0<|L|+|U|\leqslant s}\|\nu_{L,U}\|_{\mathrm{TV}}\leqslant V.$$

因此，$f^{d,s}_{,b,\{\mu_{L,U}\}}$是满足引理 8 条件的所需函数。 □

A.5.3 引理9的证明

引理 *9* 的证明。 我们使用引理 2 中为 $B(V,\ t)$ 建立的括号熵界，得到

$$J_{[\ ]}(t, B(V,t), \|\cdot\|_{p_0,2}) = \int_0^t \sqrt{1+\log N_{[\ ]}(\epsilon, B(V,t), \|\cdot\|_{p_0,2})}\, d\epsilon$$
$$\leqslant t + \int_0^t \Big[\log\Big(2+\frac{C_s(V+t)}{\epsilon}\Big)\Big]^{1/2} d\epsilon + C_{B,s} d^{\bar{s}} (1+\log d)^{\bar{s}-1} \int_0^t \Big(2+\frac{V}{\epsilon}\Big)^{1/2}\Big[\log\Big(2+\frac{V}{\epsilon}\Big)\Big]^{\bar{s}-1} d\epsilon.$$

为处理右侧的积分，我们使用以下引理，它可由分部积分直接推出（例如，参见 Ki 等人 [21, 引理 11.7]）。

引理14。对于 $u > t$,

$$\int_0^t \Big[\log\Big(\frac{u}{\epsilon}\Big)\Big]^{1/2} d\epsilon = \frac{t}{2\sqrt{\tau}} \cdot (1+2\tau)$$

and

$$\int_0^t \Big(\frac{u}{\epsilon}\Big)^{1/2}\Big[\log\Big(\frac{u}{\epsilon}\Big)\Big]^{k} d\epsilon \leqslant C_k u^{1/2} t^{1/2} (1+\tau^k),$$

其中 τ =对数(u/t) ，且 C_k 是一个依赖于 k的常数。

根据引理14中的第一个不等式,

$$\int_0^t \Big[\log\Big(2+\frac{C_s(V+t)}{\epsilon}\Big)\Big]^{1/2} d\epsilon \leqslant \int_0^t \Big[\log\Big(\frac{2t+C_s(V+t)}{\epsilon}\Big)\Big]^{1/2} d\epsilon$$
$$\leqslant Ct\Big[1+2\log\Big(\frac{2t+C_s(V+t)}{t}\Big)\Big] \leqslant C_s t \log\Big(2+\frac{V}{t}\Big).$$

此外，根据引理14中的第二个不等式以及不等式 $(x+y)^{1/2} \leqslant x^{1/2}+y^{1/2},$ ，我们有

$$\int_0^t \Big(2+\frac{V}{\epsilon}\Big)^{1/2}\Big[\log\Big(2+\frac{V}{\epsilon}\Big)\Big]^{\bar{s}-1} d\epsilon \leqslant \int_0^t \Big(\frac{2t+V}{\epsilon}\Big)^{1/2}\Big[\log\Big(\frac{2t+V}{\epsilon}\Big)\Big]^{\bar{s}-1} d\epsilon$$
$$\leqslant C_s(2t+V)^{1/2} t^{1/2}\Big(1+\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}\Big) \leqslant C_s t\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1} + C_s V^{1/2} t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}.$$

结合这些界可得

$$J_{[\ ]}(t, B(V,t), \|\cdot\|_{p_0,2}) \leqslant C_s t \log\Big(2+\frac{V}{t}\Big) + C_{B,s} d^{\bar{s}} (1+\log d)^{\bar{s}-1} t\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}$$
$$+ C_{B,s} d^{\bar{s}} (1+\log d)^{\bar{s}-1} V^{1/2} t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}.$$

若 $t \leqslant V$,

$$t\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1} \leqslant V^{1/2} t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1},$$

否则,

$$t\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1} \leqslant C_s t \leqslant C_s t \log\Big(2+\frac{V}{t}\Big).$$

因此，在这两种情况下，我们有

$$J_{[\ ]}(t, B(V,t), \|\cdot\|_{p_0,2}) \leqslant C_{B,s} d^{\bar{s}} (1+\log d)^{\bar{s}-1}\Big(t\log\Big(2+\frac{V}{t}\Big) + V^{1/2} t^{1/2}\Big[\log\Big(2+\frac{V}{t}\Big)\Big]^{\bar{s}-1}\Big).$$

□

A.5.4 引理11的证明

证明 (67)。 根据定义，对于每个 $\boldsymbol{\eta} \in \{-1,\ 1\}^q$，我们有

$$
\begin{aligned}
|\nu_{\boldsymbol{\eta}}|\Big(\prod_{j=1}^{\bar{s}}\Big(-\frac{M_j}{2},\frac{M_j}{2}\Big)\Big) &= \frac{1}{M_1\cdots M_{\bar{s}}}\cdot\frac{V}{\sqrt{|P_l|}}\int_{\prod_{j=1}^{\bar{s}}(-M_j/2,M_j/2)}\Big|\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\prod_{j=1}^{\bar{s}}\psi_{p_j,i_j}\Big(\frac{t_j}{M_j}+\frac{1}{2}\Big)\Big|\,d\mathbf{t}\\
&= \frac{V}{\sqrt{|P_l|}}\int_{(0,1)^{\bar{s}}}\Big|\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\prod_{j=1}^{\bar{s}}\psi_{p_j,i_j}(t_j)\Big|\,d\mathbf{t} \leqslant \frac{V}{\sqrt{|P_l|}}\bigg(\int_{(0,1)^{\bar{s}}}\Big(\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\prod_{j=1}^{\bar{s}}\psi_{p_j,i_j}(t_j)\Big)^2 d\mathbf{t}\bigg)^{1/2}\\
&= \frac{V}{\sqrt{|P_l|}}\bigg(\int_{(0,1)^{\bar{s}}}\sum_{\mathbf{p}\in P_l}\Big(\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\prod_{j=1}^{\bar{s}}\psi_{p_j,i_j}(t_j)\Big)^2 d\mathbf{t}\bigg)^{1/2} = \frac{V}{\sqrt{|P_l|}}\bigg(\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\int_{(0,1)^{\bar{s}}}\Big(\prod_{j=1}^{\bar{s}}\psi_{p_j,i_j}(t_j)\Big)^2 d\mathbf{t}\bigg)^{1/2}\\
&= \frac{V}{\sqrt{|P_l|}}\bigg(\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\prod_{j=1}^{\bar{s}}\int_0^1(\psi_{p_j,i_j}(t_j))^2\,dt_j\bigg)^{1/2} = \frac{V}{\sqrt{|P_l|}}\bigg(\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\prod_{j=1}^{\bar{s}}2^{-p_j}\bigg)^{1/2} = V.
\end{aligned}
$$

这里，不等式来自柯西不等式，第三个等式根据以下事实得出：

$$\int_0^1 \psi_{m,k}(x)\psi_{m',k'}(x)\,dx = 0$$

对于不同的 m 和 m'，第四个等式则是由于 $\psi_{m,\,k}\psi_{m,\,k'} \equiv 0$，前提是 $k \neq k'$。这证明了对于每个 $\boldsymbol{\eta} \in \{-1,\ 1\}^q$，都有 $V^{d,\,\infty}_{-\,\mathrm{XGB}}(f_{\boldsymbol{\eta}}) \leqslant V$。 □

(68)的证明。对于一个整数 $m \geqslant 1$ 和 $k \in [2^m]$，令 $\Psi_{m,\,k}$ 为定义在 $[0,\ 1]$ 上的实值函数，其定义为

$$\Psi_{m,k}(x) = \int_0^x \psi_{m,k}(t)\,dt.$$

很容易验证

$$
\begin{aligned}
&\text{(i) } \Psi_{m,k}(x) = 0 \ \text{ if } x \leqslant (k-1)2^{-m} \text{ or } x \geqslant k2^{-m}\\
&\text{(ii) } \Psi_{m,k}(x+2^{-m-1}) = -\Psi_{m,k}(x) \ \text{ for } x \in [(k-1)2^{-m},(k-1/2)2^{-m}]\\
&\text{(iii) } |\Psi_{m,k}(x)| \leqslant 2^{-m-2} \ \text{ for all } x \in [0,1].
\end{aligned}
\tag{72}
$$

此外，对于每个 $(x_1,\ldots,x_{\bar{s}}) \in [0,1]^{\bar{s}}$，我们有

$$
\begin{aligned}
&f_{\boldsymbol{\eta}}\Big(M_1x_1-\frac{M_1}{2},\ldots,M_{\bar{s}}x_{\bar{s}}-\frac{M_{\bar{s}}}{2}\Big)\\
&\quad= \frac{1}{M_1\cdots M_{\bar{s}}}\cdot\frac{V}{\sqrt{|P_l|}}\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\int_{\prod_{j=1}^{\bar{s}}(-M_j/2,M_j/2)}\prod_{j=1}^{\bar{s}}\Big[\mathbf{1}\Big(M_jx_j-\frac{M_j}{2}\geqslant t_j\Big)\cdot\psi_{p_j,i_j}\Big(\frac{t_j}{M_j}+\frac{1}{2}\Big)\Big]\,d\mathbf{t}\\
&\quad= \frac{V}{\sqrt{|P_l|}}\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\cdot\int_{(0,1)^{\bar{s}}}\prod_{j=1}^{\bar{s}}\big[\mathbf{1}(x_j\geqslant t_j)\cdot\psi_{p_j,i_j}(t_j)\big]\,d\mathbf{t}\\
&\quad= \frac{V}{\sqrt{|P_l|}}\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\prod_{j=1}^{\bar{s}}\int_0^{x_j}\psi_{p_j,i_j}(t_j)\,dt_j = \frac{V}{\sqrt{|P_l|}}\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\eta_{\mathbf{p},\mathbf{i}}\cdot\prod_{j=1}^{\bar{s}}\Psi_{p_j,i_j}(x_j).
\end{aligned}
\tag{73}
$$

我们利用方程 (73) 证明 (68)。假设给定 $\boldsymbol{\eta},\boldsymbol{\eta}' \in \{-1,1\}^q$，其中 $H(\boldsymbol{\eta},\boldsymbol{\eta}') = 1$，且 $(\mathbf{p},\mathbf{i})$ 是 Q 中满足 $\eta_{\mathbf{p},\mathbf{i}} \neq \eta'_{\mathbf{p},\mathbf{i}}$ 的唯一元素。于是有

$$(f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'})\Big(M_1x_1-\frac{M_1}{2},\ldots,M_{\bar{s}}x_{\bar{s}}-\frac{M_{\bar{s}}}{2}\Big) = \frac{V}{\sqrt{|P_l|}}\cdot(\eta_{\mathbf{p},\mathbf{i}}-\eta'_{\mathbf{p},\mathbf{i}})\prod_{j=1}^{\bar{s}}\Psi_{p_j,i_j}(x_j),$$

由此可得

$$\begin{aligned}\|f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'}\|_{p_0,2}^2 &\leqslant \frac{B}{M_1\cdots M_d}\cdot\int_{\prod_{j=1}^d[-M_j/2,M_j/2]}\big((f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'})(x_1,\ldots,x_{\bar{s}})\big)^2\,d\mathbf{x}\\ &= B\cdot\int_{[0,1]^{\bar{s}}}\Big((f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'})\Big(M_1x_1-\frac{M_1}{2},\ldots,M_{\bar{s}}x_{\bar{s}}-\frac{M_{\bar{s}}}{2}\Big)\Big)^2\,d\mathbf{x}\\ &= \frac{4BV^2}{|P_l|}\cdot\prod_{j=1}^{\bar{s}}\int_0^1\big(\Psi_{p_j,i_j}(x_j)\big)^2\,dx_j \leqslant \frac{4BV^2}{|P_l|}\cdot\prod_{j=1}^{\bar{s}}2^{-p_j}\cdot 2^{-2p_j-4}=\frac{BV^2}{|P_l|}\cdot 2^{-3l-4\bar{s}+2}.\end{aligned}$$

回忆一下，对于第一个不等式，$B=M_1\cdots M_d\cdot\sup_{\mathbf{x}}p_0(\mathbf{x})$。 □

证明 (69)。固定 $\boldsymbol{\eta}\neq\boldsymbol{\eta}'\in\{-1,1\}^q$。对于整数 $m\geqslant 1$ 和 $k\in[2^m]$，令 $h_{m,k}$ 为定义在 $[0,1]$ 上的实值函数，定义如下：

$$h_{m,k}(x)=\begin{cases}2^{m/2} & \text{if } (k-1)2^{-m}<x<\big(k-1/2\big)2^{-m},\\ -2^{m/2} & \text{if } \big(k-1/2\big)2^{-m}<x<k2^{-m},\\ 0 & \text{otherwise,}\end{cases}$$

并且，对于每个 $(\mathbf{p},\mathbf{i})\in Q$，令 $H_{\mathbf{p},\mathbf{i}}$ 为定义在 $[0,1]^{\bar{s}}$ 上的实值函数，定义如下：

$$H_{\mathbf{p},\mathbf{i}}(x_1,\ldots,x_{\bar{s}})=\prod_{j=1}^{\bar{s}}h_{p_j,i_j}(x_j).$$

容易验证，$\{H_{\mathbf{p},\mathbf{i}}\colon(\mathbf{p},\mathbf{i})\in Q\}$ 是 $L^2([0,1]^{\bar{s}})$ 中的一个标准正交集。考虑函数$g_{\boldsymbol{\eta},\boldsymbol{\eta}'}\colon[0,1]^{\bar{s}}\to\mathbb{R}$，定义如下：

$$g_{\boldsymbol{\eta},\boldsymbol{\eta}'}(x_1,\ldots,x_{\bar{s}})=\frac{V}{\sqrt{|P_l|}}\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\big(\eta_{\mathbf{p},\mathbf{i}}-\eta'_{\mathbf{p},\mathbf{i}}\big)\prod_{j=1}^{\bar{s}}\Psi_{p_j,i_j}(x_j).$$

由于

$$b=M_1\cdots M_d\cdot\inf_{\mathbf{x}\in\prod_{j=1}^d[-M_j/2,M_j/2]}p_0(\mathbf{x})>0,$$

我们有

$$\begin{aligned}\|f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'}\|_{p_0,2}^2 &\geqslant \frac{b}{M_1\cdots M_d}\cdot\int_{\prod_{j=1}^d[-M_j/2,M_j/2]}\big((f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'})(x_1,\ldots,x_{\bar{s}})\big)^2\,d\mathbf{x}\\ &= b\cdot\int_{[0,1]^{\bar{s}}}\Big((f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'})\Big(M_1x_1-\frac{M_1}{2},\ldots,M_{\bar{s}}x_{\bar{s}}-\frac{M_{\bar{s}}}{2}\Big)\Big)^2\,d\mathbf{x}=b\|g_{\boldsymbol{\eta},\boldsymbol{\eta}'}\|_2^2,\end{aligned}$$

其中 $\|\cdot\|_2$ 表示 L^2 范数。关于最后一个等式，请回顾 (68) 的证明中的 (73)。由贝塞尔不等式可知，

$$\|f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'}\|_{p_0,2}^2\geqslant b\|g_{\boldsymbol{\eta},\boldsymbol{\eta}'}\|_2^2\geqslant b\sum_{\mathbf{p}'\in P_l}\sum_{\mathbf{i}'\in I_{\mathbf{p}'}}\langle g_{\boldsymbol{\eta},\boldsymbol{\eta}'},H_{\mathbf{p}',\mathbf{i}'}\rangle^2,\tag{74}$$

其中 $\langle\cdot,\ \cdot\rangle$ 表示 L^2 内积。

注意，对于每个 $(\mathbf{p}',\mathbf{i}')\in Q$，

$$\langle g_{\boldsymbol{\eta},\boldsymbol{\eta}'},H_{\mathbf{p}',\mathbf{i}'}\rangle=\frac{V}{\sqrt{|P_l|}}\cdot\sum_{\mathbf{p}\in P_l}\sum_{\mathbf{i}\in I_{\mathbf{p}}}\big(\eta_{\mathbf{p},\mathbf{i}}-\eta'_{\mathbf{p},\mathbf{i}}\big)\prod_{j=1}^{\bar{s}}\langle\Psi_{p_j,i_j},h_{p'_j,i'_j}\rangle.\tag{75}$$

通过证明对于 $(\mathbf{p},\ \mathbf{i}),\ (\mathbf{p}',\mathbf{i}') \in Q$，我们完成证明，

$$\prod_{j=1}^{\bar{s}} \langle \Psi_{p_j,i_j}, h_{p'_j,i'_j} \rangle = \begin{cases} 2^{-3l/2-3\bar{s}} & \text{if } (\mathbf{p},\mathbf{i}) = (\mathbf{p}',\mathbf{i}') \\ 0 & \text{otherwise.} \end{cases} \tag{76}$$

一旦证明了 (76)，将其与 (74) 和 (75) 结合，即可推出

$$\|f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'}\|_{p_0,2}^2 \geqslant \frac{bV^2}{|P_l|} \cdot 2^{-3l-6\bar{s}} \sum_{\mathbf{p}\in P_l} \sum_{\mathbf{i}\in I_{\mathbf{p}}} (\eta_{\mathbf{p},\mathbf{i}} - \eta'_{\mathbf{p},\mathbf{i}})^2 = \frac{bV^2}{|P_l|} \cdot 2^{-3l-6\bar{s}+2} \cdot H(\boldsymbol{\eta},\boldsymbol{\eta}'),$$

由此可直接得到 (69)。我们先考虑 $(\mathbf{p},\mathbf{i}) \neq (\mathbf{p}',\mathbf{i}')$ 的情况。若 $\mathbf{p} \neq \mathbf{p}',$，则由于 $\sum_{j=1}^{\bar{s}} p_j = l = \sum_{j=1}^{\bar{s}} p'_j$，存在 $j \in [\bar{s}]$ 使得 $p_j > p'_j$。此时，$h_{p'_j,i'_j}$ 在 $((i_j-1)2^{-p_j}, i_j 2^{-p_j})$ 上为常数，因此由 (72) 可得 $\langle \Psi_{p_j,i_j}, h_{p'_j,i'_j} \rangle = 0$。若 $\mathbf{p} = \mathbf{p}',$，则 i 和 i' 必定不同，因此存在 $j \in [\bar{s}]$ 使得 $i_j \neq i'_j$。此时，对于所有 $x \in [0,1]$，都有 $\Psi_{p_j,i_j}(x) \cdot h_{p'_j,i'_j}(x) = 0$，显然 $\langle \Psi_{p_j,i_j}, h_{p'_j,i'_j} \rangle = 0$。对于 $(\mathbf{p},\mathbf{i}) = (\mathbf{p}',\mathbf{i}')$ 的情况，(76) 可由以下事实推出：

$$\langle \Psi_{p_j,i_j}, h_{p_j,i_j} \rangle = \int_{(i_j-1)2^{-p_j}}^{i_j 2^{-p_j}} \Psi_{p_j,i_j}(x) \cdot h_{p_j,i_j}(x)\, dx = 2^{-3p_j/2-3}$$

对每个 $j \in [\bar{s}]$。 □