arXiv:2304.03185v1 [stat.ML] 6 Apr 2023

基于高斯核的成对排序†

关航磊、史磊

复旦大学数学科学学院，上海市现代应用数学重点实验室，中国上海，200433

摘要

基于高斯核的正则化成对排序是一种前沿学习算法。尽管其应用范围广泛，但目前仍缺乏严谨的理论论证来支持此类排序估计量的性能。本文旨在通过为正则化成对排序建立新的预言机不等式来弥补这一不足。借助这些预言机不等式，在输入域满足一般盒计数维数假设，并结合噪声条件或标准光滑性条件的情况下，我们推导出高斯排序估计量的快速学习速率。我们的理论分析改进了现有估计结果，并表明输入空间较低的内在维数有助于使学习速率摆脱维数灾难。

关键词与短语：成对排序，预言机不等式，成对高斯核，盒计数维数，学习速率

1 引言

根据对象的潜在效用、相关性或质量对一组对象进行排序，是统计推断中的一个重要课题。近年来，这一课题在经济学、信息检索、广告和协同过滤等多种领域受到了广泛研究，参见例如 [45, 20, 8]。排序与分类密切相关，后者同样可以表述为一个监督学习问题。然而，二者本质上不同，因为排序旨在得到对象的正确顺序，而不是正确预测对象所属的类别。本文考虑成对排序问题，即比较两个对象。一般而言，我们将与单个对象相关的数据表示为 (x, y)，其中 x 是属于紧集 $\mathcal{X} \subset \mathbb{R}^d$ 的 d维特征向量，而 $y \in \mathcal{Y}$ 是该对象的标签。这里的 $\mathcal{Y} \subset \mathbb{R}$ 由离散值或连续值组成。对于任意两个不同的对象，分别表示为 $z = (x, y)$和 $z' = (x', y')$，我们根据它们的标签赋予一个顺序关系： z 排序高于z' 等价于 $y > y'$。在成对排序任务中，我们基于对象的特征 x 和 x'来预测对象之间的顺序。为此，一个二元函数 $f\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$，称为一个

† 本文所述研究得到国家自然科学基金（项目编号：12171039 和 12061160462）以及上海市科技计划项目［（项目编号：21JC1400600 和 20JC1412700]）资助。电子邮箱：ghlei21@m.fudan.edu.cn（G. Lei），leishi@fudan.edu.cn（L. Shi）。

排序规则，用于进行预测。也就是说，如果 $f(x,x')\geq 0$ 那么我们预测 $y>y'$，这意味着 f 将 z 排序在 z' 之上（这里当 $f(x,x')=0$ 时，我们优先选择 $y>y'$ 以打破平局）。令 P 为定义在 $\mathcal{Z}$ 上的概率分布：$=\mathcal{X}\times\mathcal{Y}$。我们假设这些对象是随机选取的，也就是说，它们由分布 P 的独立同分布（i.i.d.）随机样本描述。排序规则 f 的预测性能由排序风险测度，定义为

$$\begin{aligned}\mathcal{R}(f):=&P\left(\left\{z=(x,y)\in\mathcal{Z},z'=(x',y')\in\mathcal{Z}\mid y>y',f(x,x')<0\right\}\right)\\&+P\left(\left\{z=(x,y)\in\mathcal{Z},z'=(x',y')\in\mathcal{Z}\mid y<y',f(x,x')\geq 0\right\}\right),\end{aligned}\tag{1.1}$$

或等价地，超额排序风险

$$\mathcal{E}(f):=\mathcal{R}(f)-\inf\left\{\mathcal{R}(g)\mid g:\mathcal{X}\times\mathcal{X}\to\mathbb{R}\text{ is Borel measurable}\right\}.\tag{1.2}$$

在本文中，我们通过采用损失 $\phi:\mathcal{Y}\times\mathcal{Y}\times\mathbb{R}\to[0,\infty)$ 和高斯核的正则化成对学习方案来构造排序规则。如果 ϕ 是通过用于分类的基于间隔的损失定义的，即 $\phi(y,y',t)=\psi(\operatorname{sgn}(y-y')t)$，我们将 ϕ 称为基于间隔的损失，其中 $\psi:\mathbb{R}\to[0,\infty)$ 被视为基于间隔的分类损失（参见 [51, 26, 5]）。这里对于 $t\in\mathbb{R}$，如果 $t>0$，则 $\operatorname{sgn}(t)=1$；如果 $t<0$，则 $\operatorname{sgn}(t)=-1$；如果 $t=0$，则 $\operatorname{sgn}(t)=0$。ψ 的典型选择（也是本文的主要考察对象）是合页损失 $\psi_{\text{hinge}}(t):=\max\{1-t,0\}$ 和平方损失 $\psi_{\text{square}}(t):=(1-t)^2$。令 $\mathcal{X}^2:=\mathcal{X}\times\mathcal{X}$。方差为 $\sigma>0$ 的成对高斯核是在 $\mathcal{X}^2\times\mathcal{X}^2$ 上给出的一个函数

$$\begin{aligned}K^\sigma((x,x'),(u,u')):=&\frac{1}{2}\exp\left(-\frac{\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right)\\&-\frac{1}{2}\exp\left(-\frac{\|(x',x)-(u,u')\|_2^2}{\sigma^2}\right),\end{aligned}\tag{1.3}$$

其中，$\|\cdot\|_2$ 表示通常的欧几里得范数。可以验证，K^σ 在 $\mathcal{X}^2\times\mathcal{X}^2$ 上是连续且半正定的（因此是对称的）。根据 [4]，K^σ 唯一确定一个再生核希尔伯特空间（RKHS）$\mathcal{H}_{K^\sigma}$。具体而言，设 $K^\sigma_{(u,u')}:\mathcal{X}^2\to\mathbb{R}$ 是由 $K^\sigma_{(x,x')}(\cdot,\cdot):=K^\sigma((x,x'),(\cdot,\cdot))$ 定义的函数，对任意 $(x,x')\in\mathcal{X}^2$ 均成立。由 K^σ 诱导的再生核希尔伯特空间 $\mathcal{H}_{K^\sigma}$ 是线性张成空间 $\left\{K^\sigma_{(x,x')}\mid(x,x')\in\mathcal{X}^2\right\}$ 的完备化，其内积记为 $\langle\cdot,\cdot\rangle_{K^\sigma}$，并满足再生性质，即对任意 $(x,x')\in\mathcal{X}^2$ 及所有 $f\in\mathcal{H}_{K^\sigma}$，均有 $\langle K^\sigma_{(x,x')},f\rangle_{K^\sigma}=f(x,x')$。此外，$\mathcal{H}_{K^\sigma}$ 由 $\mathcal{X}^2$ 上的斜对称连续函数构成，即 $f(x,x')=-f(x',x)\ \forall f\in\mathcal{H}_{K^\sigma}$。$K^\sigma$ 的构造遵循 [32], 的思路；后者在数据对上引入所谓的非传递核，以刻画成对斜对称关系。令 $\|\cdot\|_{K^\sigma}$ 表示由内积诱导的 $\mathcal{H}_{K^\sigma}$ 的范数。给定来自 P 的独立同分布样本 $\mathbf{z}:=\{(X_i,Y_i)\}_{i=1}^n$，成对学习算法通过求解如下形式的优化问题，在 $\mathcal{H}_{K^\sigma}$ 中寻找排序规则

$$f_{\mathbf{z}}:=f^\phi_{\mathbf{z},\sigma,\lambda}\in\operatorname*{arg\,min}_{f\in\mathcal{H}_{K^\sigma}}\left\{\mathcal{R}^\phi_{\mathbf{z}}(f)+\lambda\|f\|^2_{K^\sigma}\right\},\tag{1.4}$$

其中，$\lambda>0$ 是调优参数，$\mathcal{R}^\phi_{\mathbf{z}}(f)$ 表示给定的经验 ϕ-排序风险

$$\begin{aligned}\mathcal{R}^\phi_{\mathbf{z}}(f):=&\mathbb{E}_{\mathbf{z}}\phi(Y,Y',f(X,X'))\\=&\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}\phi\left(Y_i,Y_j,f(X_i,X_j)\right).\end{aligned}\tag{1.5}$$

使用基于间隔的损失时， $\mathcal{R}_{\mathbf{z}}^{\phi}(f)$ 采用以下形式

$$\begin{aligned}\mathcal{R}_{\mathbf{z}}^{\phi}(f) &= \mathbb{E}_{\mathbf{z}}\psi(\operatorname{sgn}(Y-Y')f(X,X')) \\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\psi\left(\operatorname{sgn}(Y_i-Y_j)f(X_i,X_j)\right).\end{aligned} \tag{1.6}$$

选择 $\mathcal{H}_{K^\sigma}$ 作为候选排序规则族，自然会对 $f_{\mathbf{z}}$ 施加反对称性限制，从而得出能够保证数据中互反关系的排序预测。注意， $\mathcal{R}_{\mathbf{z}}^{\phi}$（$f$）是 ϕ 排序风险的经验类比

$$\begin{aligned}\mathcal{R}^{\phi}(f) :&= \mathbb{E}\phi\left(Y,Y',f(X,X')\right) \\ &= \int_{\mathcal{Z}}\int_{\mathcal{Z}}\phi\left(y,y',f(x,x')\right)dP(x,y)dP(x',y').\end{aligned} \tag{1.7}$$

因此，超额 ϕ-排序风险定义为

$$\mathcal{E}^{\phi}(f) := \mathcal{R}^{\phi}(f) - \inf\left\{\mathcal{R}^{\phi}(g) \mid g:\mathcal{X}^2\to\mathbb{R} \text{ is Borel measurable}\right\}. \tag{1.8}$$

至一个 cer tain 在一定程度上，它也反映了排序规则 f 的泛化性能。

算法 (1.4)，或者更准确地说，基于核的正则化经验风险最小化，是近年来受到理论界广泛关注的学习方案之一。它在应用中的前沿经验性能、相对简单的实现，以及最后但同样重要的一点——其灵活性，共同促成了人们对它日益增长的兴趣。算法 (1.4)的灵活性得益于其两个主要组成部分，即再生核希尔伯特空间 $\mathcal{H}_{K^\sigma}$ 和损失函数 ϕ。更具体地说，损失函数可用于建模学习目标，而 $\mathcal{H}_{K^\sigma}$ 通过改变宽度 σ 来适应目标函数的光滑性，其中包含了分布和输入域的信息。值得注意的是，核可以定义在任意输入域上，因此除了标准的 $\mathbb{R}^d$值数据之外，还能够处理各种类型的数据。例如，使用定义在 $\mathcal{X}\times\mathcal{X}$上的二元核，我们可以定义一般的成对核 K，以建模互反关系或反对称关系。具体而言，设 $G:\ \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 为一个半正定核，且$(\mathcal{H}_G,\langle\cdot,\cdot\rangle_G)$ 为其对应的再生核希尔伯特空间。成对核 $K:\ \mathcal{X}^2\times\mathcal{X}^2\to\mathbb{R}$ 可以定义为

$$K((x,x'),(u,u')) = \langle G_x - G_{x'}, G_u - G_{u'}\rangle_G,$$

其中 $G_x := G(x,\cdot),\ x\in\mathcal{X}$. 有关这些特定成对核的更多详情，请参阅 [32, 29]。此外，与算法 (1.4) 相关的优化问题可化为凸二次优化问题；对于广泛使用的损失函数（如合页损失和平分损失），人们已对其有充分的理解，参见例如 [15, 39]。

非参数排序方法理论研究的主要课题之一是学习速率，即超额排序误差或超额 ϕ-排序风险的收敛速率，它们刻画了这些方法在特定分布类上的排序性能。因此，越来越多的研究致力于为各种排序算法推导可靠的学习速率。关于近期的一些进展，可参见 [12, 33, 10, 50, 25] 及其中的参考文献。然而，对于在何种非平凡条件下能够获得快速学习速率，目前所知甚少。本文在成对排序中引入了适当的噪声条件，并运用局部 Rademacher 分析，为算法 (1.4) 建立了紧致的学习速率。我们的方法将局部 Rademacher 平均与 U统计量理论相结合，

其中结合了剥层法和对称化技巧，以及针对U过程的压缩原理。我们的工作还受到 [18] 近期工作的启发，在该工作中，作者在假设输入域具有较低内在维数的情况下，从理论上分析了高斯支持向量机的性能。据我们所知，我们的论文首次考察了使用高斯核进行正则化成对排序的学习行为。这与现有文献形成对比，后者要么仅考察采用未正则化经验风险最小化的排序，例如 [12, 33], ；要么只处理训练过程中核保持不变的情况，例如 [10, 52, 25, 44]。我们在第3 节中讨论了这些文章的研究发现，并将其与我们的结果进行了比较。我们的主要贡献总结如下。

- 我们建立了一个预言机不等式，用于界定采用一般成对核K 的正则化经验风险最小化（2.1）所产生的估计量的超额 ϕ排序风险（参见定理 1）。我们的结论基于一个容量假设，该假设要求再生核希尔伯特空间 $\mathcal{H}_K$ 满足经验熵条件（参见假设2）。这一成熟的不等式处理收敛分析中的随机部分，并为界定超额 ϕ排序风险提供了一个框架。将这一不等式直接应用于基于边际的损失，可以改进此前关于采用一般成对核进行正则化排序的分析，并得到迄今为止最优的学习速率（详细比较参见定理 2和第 3 节）。

- 受到支持向量机分类分析的启发，例如参见 [36],，我们在成对排序框架下引入了两个噪声条件。第一个条件描述标签中的噪声量，这类似于二元分类中的 Tsybakov 噪声条件（见假设4）。该条件使我们能够建立一个优雅的校准不等式，以 ϕ-排序风险的超额部分界定超额排序风险，并为基于边际的损失建立一个精细的方差界，从而进一步改进分析中的随机部分。第二个条件是关于分布的一个几何假设，使我们能够估计高斯核的逼近性质（见假设5和第5.2小节）。这两个条件在推导快速学习速率方面都发挥着关键作用。

- 我们在输入域满足一般盒计数维数假设，并结合噪声条件或标准正则性条件的情况下，获得了高斯正则化排序算法（1.4）的学习速率（见定理 3 和定理 4）。具体而言，我们首先在如下假设下估计成对高斯核空间 $\mathcal{H}_{K^\sigma}$ 的容量：数据生成分布在输入空间上的边缘分布$\mathcal{X} \subset \mathbb{R}^d$ 支撑于一个上盒计数维数为 $\varrho \in (0, d]$ 的集合上。然后，我们在噪声条件下推导合页损失以及在 Besov 光滑性下推导平方损失的逼近误差界。最后，我们应用成熟的预言机不等式和校准不等式，推导出超额排序风险的快速学习速率。我们表明，输入空间较低的内在维数有助于使这些速率避开维数灾难。

本文其余部分的组织如下。在第 2 节中，我们首先介绍基本记号和必要的假设。然后给出本文的主要结果，包括关于超额 ϕ-排序风险的一般预言机不等式，以及其在采用合页损失和平方损失的高斯排序估计量（1.4）中的应用。第 3 节概述相关工作，并将我们的结果与其他关于成对排序的研究成果进行比较。第 4 节介绍

定理 1中确立的一般预言机不等式的详细证明。在第5 节中，我们应用预言机不等式，为采用合页损失和平方损失的高斯排序估计量推导学习速率，从而给出定理3 和定理4的证明。第 4 节和第 5 节还包括一些其本身就值得关注的初步估计。

2 主要结果

本节陈述我们的主要结果。令 $(\mathcal{H}_K, \langle\cdot,\cdot\rangle_K)$ 为由半正定核K: $\mathcal{X}^2 \times \mathcal{X}^2 \to \mathbb{R}$诱导的、定义在 $\mathcal{X}^2 := \mathcal{X} \times \mathcal{X}$ 上的斜对称函数所构成的再生核希尔伯特空间。第一个结果是定理 1 中建立的预言机不等式，它为正则化估计量的超额 ϕ-排序风险提供上界

$$f_{\mathbf{z}} := f^{\phi}_{\mathbf{z},\lambda} \in \arg\min_{f\in\mathcal{H}_K} \left\{ \mathcal{R}^{\phi}_{\mathbf{z}}(f) + \lambda\|f\|^2_K \right\}, \tag{2.1}$$

其中， ϕ: $\mathcal{Y} \times \mathcal{Y} \times \mathbb{R} \to [0,\infty)$ 是一般损失函数，而 $\mathcal{R}^{\phi}_{\mathbf{z}}(f)$ 由 (1.5) 定义。下文中，我们将用 $f_{\mathbf{z}}$ 表示 (2.1) 中的 $f^{\phi}{}_{\mathbf{z}\,\lambda}$ 或 (1.4) 中的 $f^{\phi}{}_{\mathbf{z}\,\sigma,\ \lambda}$，具体采用哪一个视上下文而定。非参数统计领域的文献已对预言机不等式进行了广泛研究（参见 [21] 及其中的参考文献）。预言机不等式给出统计估计量的风险界，并将其与所谓的预言机进行比较；预言机拥有无限量的观测数据，并在特定函数类上使总体风险最小化。回顾一下， ϕ-排序风险 $\mathcal{R}^{\phi}(f)$ 由 (1.7) 给出。在我们的设定中，预言机是使 $\min_{f\in\mathcal{H}_K} \lambda\|f\|^2_K + \mathcal{R}^{\phi}(f)$ 达到最小值的下确界。预言机不等式给出超额 ϕ-排序风险的界，并可用于建立排序算法的一致性和学习速率。我们将在第 3 节回顾针对不同排序算法已有的预言机不等式。

为陈述由 (2.1) 定义的 $f_{\mathbf{z}}$ 的预言机不等式，我们引入一些必要的记号和假设。将贝叶斯 ϕ 排序规则定义为

$$f^*_{\phi} := \arg\min \left\{ \mathcal{R}^{\phi}(f) \mid f : \mathcal{X}^2 \to \mathbb{R} \text{ is Borel measurable} \right\}. \tag{2.2}$$

下文中，除非另有说明，“可测”均指“博雷尔可测”。给定一个可测函数 f: $\mathcal{X}^2 \to \mathbb{R}$，定义 ϕ_f: $\mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ 和 $Q\phi_f$: $\mathcal{Z} \to \mathbb{R}$ 为

$$\phi_f(z,z') := \phi(y,y',f(x,x')), \tag{2.3}$$

$$Q\phi_f(z) := \mathbb{E}[\phi_f(z,Z')] = \int_{\mathcal{Z}} \phi\left(y,y',f(x,x')\right) dP(x',y'). \tag{2.4}$$

接下来，我们引入截断操作，从而能够为截断随机变量推导上确界范数界和方差界的 Bernstein 条件，这对于建立紧致的集中不等式并推导快速学习速率至关重要。我们称损失函数 ϕ: $\mathcal{Y} \times \mathcal{Y} \times \mathbb{R} \to [0,\infty)$ 可在 $M > 0$ 处截断，如果对于所有 $(y,y',t) \in \mathcal{Y} \times \mathcal{Y} \times \mathbb{R}$，均有 $\phi(y,y',\pi(t)) \le \phi(y,y',t)$，其中

$$\pi(t) := \begin{cases} -M & t < -M \\ t & t \in [-M,M] \\ M & t > M, \end{cases}$$

表示 t 在 $\pm M$ 处截断后的值。对于任意 $f\colon \mathcal{X}^2 \to \mathbb{R}$, $\pi(f)$ 表示将 f 截断到 $[-M, M]$，其定义为 $\pi(f)(x, x') := \pi(f(x, x'))\ \forall (x, x') \in \mathcal{X}^2$。截断的思想已用于二分类文献中，参见例如 [16, 36]。我们首先对一般损失 ϕ 给出以下假设。

假设1。损失 $\phi\colon \mathcal{Y} \times \mathcal{Y} \times \mathbb{R} \to [0, \infty)$ 可以在某个 $M > 0$ 处截断，并且对于所有 $(y, y') \in \mathcal{Y} \times \mathcal{Y}$，$\phi(y, y', \cdot)\colon \mathbb{R} \to [0, \infty)$ 在 $[-M, M]$ 上是凸的且局部 L-Lipschitz。由 (2.2) 定义的贝叶斯 ϕ-排序规则 f^*_ϕ 是可测的且反对称。对于任意反对称函数 f，ϕ_f 由 (2.3) 定义，满足

$$\phi_f(z, z') = \phi_f(z', z), \quad \forall (z, z') \in \mathcal{Z} \times \mathcal{Z}.$$

此外，存在常数 $B > 0$、$\tau \in [0, 1]$ 和 $V \geq B^{2-\tau}$，使得

$$\phi(y, y', t) \leq B, \quad \forall (y, y', t) \in \mathcal{Y} \times \mathcal{Y} \times [-M, M] \tag{2.5}$$

对于所有可测且反对称的 $f\colon \mathcal{X}^2 \to \mathbb{R}$，均成立。

$$\mathbb{E}(Q\phi_{\pi(f)} - Q\phi_{f^*_\phi})^2 \leq V\big(\mathbb{E}(Q\phi_{\pi(f)} - Q\phi_{f^*_\phi})\big)^\tau, \tag{2.6}$$

其中 $Q\phi_{\pi(f)}$ 和 $Q\phi_{f^*_\phi}$ 由 (2.4) 定义。

接下来，我们引入关于 $\mathcal{H}_K$ 的一个容量假设，该假设由 $\mathcal{H}_K$中单位范数球的覆盖数，或等价地熵数来描述。

定义 1. 设 $(\mathcal{T}, \mathrm{d})$ 为伪度量空间，且有 $\varepsilon > 0$。如果对于所有 $t \in \mathcal{T}$，存在一个 $s \in \mathcal{S}$ 使得 $\mathrm{d}(s, t) \leq \varepsilon$，则称 $\mathcal{S} \subset \mathcal{T}$ 为 $\mathcal{T}$ 的 ε-网。$(\mathcal{T}, \mathrm{d})$ 的 ε-覆盖数定义为

$$\mathcal{N}(\mathcal{T}, \mathrm{d}, \varepsilon) := \inf\big\{|\mathcal{S}| : \mathcal{S} \subset \mathcal{T} \ and\ \mathcal{S}\ is\ an\ \varepsilon\text{-}net\ of\ \mathcal{T}\big\},$$

其中$\inf \emptyset = \infty$。对于 $i \in \mathbb{N}$，$(\mathcal{T}, \mathrm{d})$ 的第 i 个熵数定义为

$$e_i(\mathcal{T}, \mathrm{d}) := \inf\left\{\varepsilon > 0 : \exists s_1, \ldots, s_{2^{i-1}} \in \mathcal{T}\ such\ that\ \mathcal{T} \subset \bigcup_{j=1}^{2^{i-1}} \mathcal{B}_{\mathrm{d}}(s_j, \varepsilon)\right\},$$

其中，$\mathcal{B}_{\mathrm{d}}(s, \varepsilon) := \big\{t \in \mathcal{T} : \mathrm{d}(t, s) \leq \varepsilon\big\}$ 表示以 $s \in \mathcal{T}$为中心、半径为ε的闭球。特别地，如果（$\mathcal{T}$,d）是赋范空间（$\mathcal{F}, \|\cdot\|_\mathcal{F}$）的子空间，且度量 d由 $\mathrm{d}(s, t) := \|s - t\|_\mathcal{F}, \forall s, t \in \mathcal{T}$给出，我们记 $\mathcal{N}(\mathcal{T}, \|\cdot\|_\mathcal{F}, \varepsilon) := \mathcal{N}(\mathcal{T}, \mathrm{d}, \varepsilon)$。此外，如果$S\colon \mathcal{F} \to \mathcal{F}'$是赋范空间 $\mathcal{F}$和 $\mathcal{F}'$之间的有界线性算子，我们记$e_i(S) := e_i(S\mathcal{B}_\mathcal{F}, \|\cdot\|_{\mathcal{F}'})$以及 $\mathcal{N}(S, \varepsilon) := \mathcal{N}(S\mathcal{B}_\mathcal{F}, \|\cdot\|_{\mathcal{F}'}, \varepsilon)$，其中 $\mathcal{B}_\mathcal{F}$是$\mathcal{F}$的闭单位球。

令 $P_\mathcal{X}$ 表示 P 在 $\mathcal{X}$上的边际分布。给定来自$P_\mathcal{X}$的独立同分布样本$\mathbf{x} = \{X_i\}_{i=1}^n$，定义两个经验测度：$P^n{}_\mathbf{x} := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$和 $P^n{}_{\mathbf{x}^2} := \frac{1}{n(n-1)}\sum_{i \neq j} \delta_{(X_i, X_j)}$，其中$\delta_{(\cdot)}$是计数测度。我们对再生核希尔伯特空间 $\mathcal{H}_K$给出如下容量假设。

假设2。对于所有 $n \in \mathbb{N}$ 和独立同分布样本$\mathbf{x} = \{X_i\}_{i=1}^n$ 的 $P_\mathcal{X}$，存在常数$0 < p_1$、$p_2 < 1/2$、$a_1 > 0$ 和 $a_2 > 0$，使得

$$e_i(\mathrm{id} : \mathcal{H}_K \to \mathcal{L}_2(P^n_\mathbf{x} \otimes P_\mathcal{X})) \leq a_1 i^{-\frac{1}{2p_1}}, \quad \forall i \in \mathbb{N},$$

$$e_i(\mathrm{id}:\mathcal{H}_K\to\mathcal{L}_2(P^n_{\mathbf{x}^2}))\le a_2 i^{-\frac{1}{2p_2}},\quad \forall i\in\mathbb{N}.$$

这里，$\mathcal{L}_2(P^n_{\mathbf{x}}\otimes P_{\mathcal{X}})$ 和 $\mathcal{L}_2(P^n_{\mathbf{x}^2})$ 是关于 $P^n_{\mathbf{x}}\otimes P_{\mathcal{X}}$ 和 $P^n_{\mathbf{x}^2}$ 的平方可积函数的希尔伯特空间，其中范数对于 $f\in\mathcal{H}_K$ 定义为

$$\|f\|_{\mathcal{L}_2(P^n_{\mathbf{x}}\otimes P_{\mathcal{X}})}:=\left(\frac{1}{n}\sum_{i=1}^n\mathbb{E}_{P_{\mathcal{X}}}\left[f(X_i,X)^2\right]\right)^{\frac{1}{2}},$$
$$\|f\|_{\mathcal{L}_2(P^n_{\mathbf{x}^2})}:=\left(\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}f(X_i,X_j)^2\right)^{\frac{1}{2}}.$$

鉴于熵数与覆盖数之间的联系，参见[36], 引理6.21，关于熵数的容量假设可以用度量熵，即覆盖数的对数，重新表述为：存在常数 $0<p_1$、$p_2<1/2$、$a_1>0$ 和 $a_2>0$，使得

$$\log\mathcal{N}(\mathrm{id}:\mathcal{H}_K\to\mathcal{L}_2(P^n_{\mathbf{x}}\otimes P_{\mathcal{X}}),\varepsilon)\le\log(4)a_1^{2p_1}\varepsilon^{-2p_1},$$
$$\log\mathcal{N}(\mathrm{id}:\mathcal{H}_K\to\mathcal{L}_2(P^n_{\mathbf{x}^2}),\varepsilon)\le\log(4)a_2^{2p_2}\varepsilon^{-2p_2}.$$

回顾一下，ϕ-排序风险 $\mathcal{R}^\phi(\cdot)$由式(1.7)给出，而贝叶斯 ϕ-排序规则 f^*_ϕ 由式(2.2)定义。注意，如果损失 ϕ 可以在某个 $M>0$处截断，则有

$$\mathcal{R}^\phi(\pi(f))-\mathcal{R}^\phi(f^*_\phi)\le\mathcal{R}^\phi(f)-\mathcal{R}^\phi(f^*_\phi)$$

对所有 $f:\ \mathcal{X}^2\to\mathbb{R}$。下文中，我们始终可以考虑截断后的估计量，因为将决策函数的取值投影到 $[-M,M]$ 不会增加其超额 ϕ-排序风险。现在，我们可以给出由算法(2.1)在一般损失 ϕ下生成的估计量所满足的预言机不等式。

定理 1. 设 $\mathbf{z}=\{(X_i,Y_i)\}_{i=1}^n$ 是概率分布 P在 $\mathcal{X}\times\mathcal{Y}$上的独立同分布样本。给定损失函数 $\phi:\mathcal{Y}\times\mathcal{Y}\times\mathbb{R}\to[0,\infty)$和可测半正定核$K:\ \mathcal{X}^2\times\mathcal{X}^2\to\mathbb{R}$，以及由反对称函数组成的相关再生核希尔伯特空间 $\mathcal{H}_K$，估计量 $f_{\mathbf{z}}$ 由带有 $\lambda>0$的式（2.1）定义。假设 ϕ 满足假设1，其中$L>0$、$B>0$、$\tau\in[0,1]$和 $V\ge B^{2-\tau}$，且 $\mathcal{H}_K$满足假设2，其中 $0<p_1$、$p_2<1/2$、$a_1\ge B$ 和 $a_2\ge B$。如果存在某个 $f_0\in\mathcal{H}_K$，使其对于某个常数$B_0\ge0$满足$\|\phi_{f_0}\|_\infty\le B_0$，那么对于所有 $n\ge2$和 $t>0$，以至少 $1-(c_0+5)\exp(-t)$ 的概率，有

$$\begin{aligned}
&\lambda\|f_{\mathbf{z}}\|_K^2+\mathcal{R}^\phi(\pi(f_{\mathbf{z}}))-\mathcal{R}^\phi(f^*_\phi)\\
&\le 8\big(\lambda\|f_0\|_K^2+\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f^*_\phi)\big)+6c_1\left(\frac{a_1^{2p_1}V^{1-p_1}L^{2p_1}}{\lambda^{p_1}n}\right)^{\frac{1}{2-\tau-p_1+p_1\tau}}\\
&\quad+\frac{6c_2a_1^{2p_1}B^{1-p_1}L^{2p_1}}{\lambda^{p_1}n}+\frac{3c_3a_1^{2p_1}L^{2p_1}B^{1-p_1}t}{\lambda^{p_1}n}+\frac{3c_4a_2^{2p_2}L^{2p_2}B^{1-p_2}t}{\lambda^{p_2}n}\\
&\quad+\left(\frac{1872Vt}{n}\right)^{\frac{1}{2-\tau}}+\frac{900Bt}{n}+\frac{456B_0t}{n}+\frac{3c_5t}{n}.
\end{aligned}\tag{2.7}$$

其中，c_i、$i=0$、1、2、3、4、5 是与 λ、n 或 t 无关的常数，并在证明中明确给出。

接下来，我们将预言机不等式应用于基于间隔的损失函数 $\phi(y,y',t) = \psi(\mathrm{sgn}(y-y')t)$，其中 $\psi: \mathbb{R} \to [0,\infty)$，且 ψ 满足以下假设。

假设3. 单变量函数 $\psi: \mathbb{R} \to [0,\infty)$ 是凸函数，在 $t=0$处可微，且 $\psi'(0) < 0$， ψ 的最小零点为 1。

假设3中的 ψ 示例包括合页损失 ψ_{hinge}、平方损失 ψ_{square} 以及 r范数合页损失，其中 $1 \le r < \infty$ 定义为 $\psi_r(t) := (\psi_{\mathrm{hinge}}(t))^r$。根据第5.3小节中的引理9，若 ψ 满足假设3，则基于间隔的损失 $\phi(y,y',t) = \psi(\mathrm{sgn}(y-y')t)$可以在 $M=1$ 处截断，且贝叶斯 ϕ-排序规则 f_ϕ^* 可以取为反对称。现在，我们可以给出采用基于间隔的损失时，由算法 (2.1) 生成的估计量的预言机不等式。

定理 2.令 $\mathbf{z} = \{(X_i,Y_i)\}_{i=1}^n$ 为概率分布 P on$\mathcal{X}\times\mathcal{Y}$的独立同分布样本。给定一个基于间隔的损失 $\phi(y,y',t) = \psi(\mathrm{sgn}(y-y')t)$，其中 $\psi: \mathbb{R}\to[0,\infty)$ 满足假设3，以及一个可测正半定核 $K: \mathcal{X}^2\times\mathcal{X}^2 \to \mathbb{R}$，其关联的再生核希尔伯特空间 $\mathcal{H}_K$ 由反对称函数组成，估计量 $f_{\mathbf{z}}$ 由（2.1）定义，其中 $\lambda > 0$。假设存在常数 $L>0$、 $B>0$、 $\tau\in[0,1]$和 $V \ge B^{2-\tau}$，使得 ψ 是局部 L-Lipschitz 的，并且 $\psi(t) \le B$ 在 $t\in[-1,1]$上成立，且假设1的方差界（2.6）在截断参数 $M=1$下成立。如果存在某个 $f_0 \in \mathcal{H}_K$满足 $\|\phi_{f_0}\|_\infty \le B_0$ （其中某个常数为 $B_0 \ge 0$ ），并且 $\mathcal{H}_K$ 满足假设2，其中$0<p_1$、 $p_2<1/2$、 $a_1 \ge B$和 $a_2 \ge B$，则定理 1 的结论成立。

注1. 在定理 2 的假设下，我们进一步假设 $p_1=p_2=p$ ，并且存在常数 $c>0$ 和 $\gamma\in$（$0,1]$ ），使得对所有 $\lambda>0$，逼近误差 $\mathcal{A}$（λ）满足以下界：

$$\mathcal{A}(\lambda) := \inf_{f\in\mathcal{H}_K} \lambda\|f\|_K^2 + \mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*) \le c\lambda^\gamma.$$

于是，取

$$\lambda = n^{-b},\ b = \frac{1}{(2-\tau-p+p\tau)\gamma+p},$$

我们有

$$\mathcal{R}^\phi(\pi(f_{\mathbf{z}})) - \mathcal{R}^\phi(f_\phi^*) \lesssim n^{-\frac{\gamma}{(2-\tau-p+p\tau)\gamma+p}}.$$

该学习速率显著改进了现有的、针对基于间隔损失的成对排序结果。详细比较请见第 3 节。

我们将在第 4 节证明定理 1 和定理 2。得到预言机不等式后，我们现在聚焦于采用成对高斯核 K^σ 和两种典型的基于间隔的损失函数的成对排序方案 (1.4)：

$$\phi_{\mathrm{hinge}}(y,y',t) := \psi_{\mathrm{hinge}}(\mathrm{sgn}(y-y')t) = \max\{1-\mathrm{sgn}(y-y')t, 0\}$$

and

$$\phi_{\mathrm{square}}(y,y',t) := \psi_{\mathrm{square}}(\mathrm{sgn}(y-y')t) = (1-\mathrm{sgn}(y-y')t)^2.$$

回顾一下，超额排序风险 $\mathcal{E}(f)$和超额 ϕ-排序风险 $\mathcal{E}^\phi(f)$分别由 (1.2)和(1.8) 定义。为了使用定理 2 推导超额排序风险的学习速率，仍需解决以下问题：

- 验证假设1对于$\phi_{\text{铰链}}$，$\phi_{\text{平方}}$；
- 构造一个 $f_0 \in \mathcal{H}_{K_\sigma}$ 以界定逼近误差 $\mathcal{A}(\lambda)$；
- 验证再生核希尔伯特空间 $\mathcal{H}_{K_\sigma}$的容量条件（假设2）；
- 建立排序的校准不等式，即通过超额 ϕ-排序风险 $\mathcal{E}^\phi(f)$界定超额排序风险 $\mathcal{E}(f)$。

在第 5 节中，我们将解决这些问题，并在适当的正则性假设下建立相应的估计结果。在本节的其余部分，我们介绍这些假设，以刻画概率分布、目标函数的光滑性以及输入域的内在维数。然后，我们给出配备 ϕ_{hinge} 和 ϕ_{square}的高斯排序估计器（1.4）的快速学习速率结果。

我们首先介绍两个噪声条件来描述概率分布。回顾一下， P 是 $\mathcal{X} \times \mathcal{Y}$上的概率分布。对于不同的（$x$， y）和（x'， y'），定义 P的后验概率，分别记为 η_+、 η_-、 $\eta_=$： $\mathcal{X}^2 \to [0,1]$，则有

$$\begin{aligned}
\eta_+(x,x') &:= P(y > y'|x,x'),\\
\eta_-(x,x') &:= P(y < y'|x,x'),\\
\eta_=(x,x') &:= P(y = y'|x,x').
\end{aligned}$$

然后将 $\mathcal{X}^2$ 划分为

$$\begin{aligned}
\mathcal{X}_+^2 &:= \{(x,x') \in \mathcal{X}^2 : \eta_+(x,x') > \eta_-(x,x')\},\\
\mathcal{X}_-^2 &:= \{(x,x') \in \mathcal{X}^2 : \eta_+(x,x') < \eta_-(x,x')\},\\
\mathcal{X}_=^2 &:= \{(x,x') \in \mathcal{X}^2 : \eta_+(x,x') = \eta_-(x,x')\}.
\end{aligned} \tag{2.8}$$

根据 [12], 命题1，贝叶斯排序规则 f^*_{rank}： $\mathcal{X}^2 \to \mathbb{R}$ 其在所有可测函数中使(1.1)中的排序风险 $\mathcal{R}(f)$最小化，表示为

$$f^*_{\text{rank}}(x,x') = \operatorname{sgn}\big(\eta_+(x,x') - \eta_-(x,x')\big) = \begin{cases} 1 & \text{if } (x,x') \in \mathcal{X}_+^2,\\ 0 & \text{if } (x,x') \in \mathcal{X}_=^2,\\ -1 & \text{if } (x,x') \in \mathcal{X}_-^2. \end{cases}$$

那么，最小排序风险为 $\mathcal{R}(f^*_{\text{rank}}) = \mathbb{E}\min\{\eta_+(X,X'), \eta_-(X,X')\}$。

现在我们引入一个与二分类中的 Tsybakov 噪声条件类似的条件，用来描述标签中的噪声量。为了说明这一条件的动机，我们注意到由

$$\min\big\{\eta_+(x,x'), \eta_-(x,x')\big\} = \frac{\eta_+(x,x') + \eta_-(x,x')}{2} - \frac{\big|\eta_+(x,x') - \eta_-(x,x')\big|}{2}$$

函数$\big|\eta_+(x,x') - \eta_-(x,x')\big|$ 可用于描述比较分布 P标签时的噪声。事实上，在该函数接近 1 的区域，比较 y 和 y'时的噪声很小；而函数值接近 0 仅出现在比较 y 和 y'时噪声水平较高的区域。下面的条件描述了后者区域的大小，其中对于 $t \in (0,1)$，我们采用约定 $t^\infty := 0$ 。回顾一下，$P_{\mathcal{X}}$是 P在 $\mathcal{X}$上的边缘分布。令 $P_{\mathcal{X}}^2 := P_{\mathcal{X}} \otimes P_{\mathcal{X}}$。

假设4. 存在常数 $C_* > 0$ 和 $q \in [0, \infty]$ 使得

$$P_{\mathcal{X}}^2\big(\{(x, x') \in \mathcal{X}^2 : \big|\eta_+(x, x') - \eta_-(x, x')\big| \le t\}\big) \le C_* t^q \tag{2.9}$$

对所有 $t > 0$。

显然，当且仅当 $\big|\eta_+(x, x') - \eta_-(x, x')\big|^{-1}$ 属于洛伦兹空间 $\mathcal{L}_{q,\infty}(P_{\mathcal{X}}^2)$（见 [27]）时，$P_{\mathcal{X}}^2$具有噪声指数 $q > 0$。也容易看出，所有分布都满足 (2.9)，其中 $C_* = 1$ 且 $q = 0$。此外，极端情形 $q = \infty$ 要求$\big|\eta_+(x, x') - \eta_-(x, x')\big|$ 几乎对所有 $(x, x') \in \mathcal{X}^2$都由某个正数下界限定，其中包括无噪声分布的情形，此时 $\big|\eta_+(x, x') - \eta_-(x, x')\big| = 1$ 几乎必然成立。在假设4下，我们将证明 ϕ_{hinge}的方差界 (2.6)。我们还利用这一假设建立一个精细的校准不等式，用 $\mathcal{E}^{\phi}(f)$界定 $\mathcal{E}(f)$。参见命题9和引理11。 [12] 和 [13] 的研究为二分类排序问题引入了另外两个噪声条件。这些噪声条件比我们的条件更强，因此在 (2.6) 中得到更优的方差界指数 τ。我们将在第 5.5 小节详细比较这些噪声条件。

第二个噪声条件是针对分布的一种几何条件，它描述了噪声集合的位置，并使我们能够估计高斯核的逼近误差。根据我们在第5.3小节中的讨论， ϕ_{hinge} 的贝叶斯排序规则可以逐点定义，即给定 $(x, x') \in \mathcal{X}^2$,

$$f^*_{\text{hinge}}(x, x') := \arg\min_{t \in \mathbb{R}} \eta_+(x, x')\psi_{\text{hinge}}(t) + \eta_-(x, x')\psi_{\text{hinge}}(-t).$$

那么， f^*_{hinge} 可以明确表示为

$$f^*_{\text{hinge}}(x, x') = \operatorname{sgn}\big(\eta_+(x, x') - \eta_-(x, x')\big),\ \forall (x, x') \in \mathcal{X}^2, \tag{2.10}$$

这恰好是最优排序规则 $f^*_{\text{排序}}$。注意， f^*_{hinge}通常是阶跃函数。因此，采用经典的光滑性假设来描述 f^*_{hinge}的正则性似乎相当严格。受合页损失二元分类理论分析的启发（参见 [36]），我们引入了第二个噪声条件，称为间隔噪声条件。它将噪声与到决策边界的距离联系起来，并用于界定合页损失和高斯核对应的逼近误差。参见命题 5。

假设5. 存在常数 $C_{**} > 0$ 和 $\beta \ge 0$，使得对所有 $t \ge 0$,

$$\int_{\{(x,x') \in \mathcal{X}^2 : \Delta(x,x') < t\}} |\eta_+(x, x') - \eta_-(x, x')| dP_{\mathcal{X}}(x) dP_{\mathcal{X}}(x') \le C_{**} t^{\beta}. \tag{2.11}$$

此处， $\Delta(x, x')$ 表示到决策边界的距离，定义为

$$\Delta(x, x') := \begin{cases} \operatorname{dist}((x, x'), \mathcal{X}_-^2 \cup \mathcal{X}_=^2) & (x, x') \in \mathcal{X}_+^2 \\ \operatorname{dist}((x, x'), \mathcal{X}_+^2 \cup \mathcal{X}_=^2) & (x, x') \in \mathcal{X}_-^2 \\ 0 & (x, x') \in \mathcal{X}_=^2, \end{cases} \tag{2.12}$$

其中 $\operatorname{dist}(x, A) := \inf_{y \in A} \|x - y\|_2$.

请注意，在条件 (2.11) 中，我们既不施加任何形式的光滑性假设，也不要求 $P_{\mathcal{X}}^2$ 关于勒贝格测度绝对连续。如果 $P_{\mathcal{X}}^2$ 在决策边界 $\mathcal{X}_{=}^2$ 附近的集中度较低，或者该区域特别嘈杂，那么对于较大的指数 β，假设5成立。例如，在 $\mathcal{X}_{+}^2$ 和 $\mathcal{X}_{-}^2$ 之间具有正距离这一极端情况下，我们可以为 β 选取任意大的值。

接下来，我们介绍盒计数维数条件，该条件充分利用了输入数据的低维内在结构。假设2要求再生核希尔伯特空间 $\mathcal{H}_K$ 关于 $\mathcal{L}_2$-半范数 $\|\cdot\|_{\mathcal{L}_2(P^n{}_{\mathbf{X}}\otimes P_{\mathcal{X}})}$ 和 $\|\cdot\|_{\mathcal{L}_2(P^n{}_{\mathbf{X}^2})}$ 的熵数上界，这两者都受上确界范数 $\|\cdot\|_\infty$ 支配。关于上确界范数的高斯再生核希尔伯特空间 $\mathcal{H}_{K^\sigma}$ 的覆盖数已经在文献中得到了深入研究，参见[23, 37],，其中一些研究将其与输入空间的统一覆盖数 $\mathcal{N}(\mathcal{X},\|\cdot\|_\infty,\sigma)$ 联系起来。由于 $\mathcal{N}(\mathcal{X},\|\cdot\|_\infty,\sigma)$ 以输入空间的维数 $\mathcal{X}\subset\mathbb{R}^d$ 呈指数依赖。当 d 较大时，学习速率会遭受维数灾难。对于高维数据而言，这种现象通常不可避免。然而，当数据的内在维数远小于其环境空间的维数时，我们可以推导出快得多的收敛速率。我们通过盒计数维数描述输入域的低维内在结构；盒计数维数是分形几何中用于描述低维内在分形结构的概念，参见例如 [17]。对于度量空间 $(\mathcal{T},\mathrm{d})$ 及其子集 $\mathcal{S}\subset\mathcal{T}$，如果极限

$$\lim_{\varepsilon\to 0}\frac{\log\mathcal{N}(\mathcal{S},\mathrm{d},\varepsilon)}{\log\frac{1}{\varepsilon}}$$

若存在，则称其为 $\mathcal{S}$ 的盒计数维数。若不存在，则定义上盒计数维数

$$\limsup_{\varepsilon\to 0}\frac{\log\mathcal{N}(\mathcal{S},\mathrm{d},\varepsilon)}{\log\frac{1}{\varepsilon}},$$

以及下盒计数维数

$$\liminf_{\varepsilon\to 0}\frac{\log\mathcal{N}(\mathcal{S},\mathrm{d},\varepsilon)}{\log\frac{1}{\varepsilon}}.$$

假设6. 存在一个常数 $\varrho>0$ 使得

$$\limsup_{\varepsilon\to 0}\frac{\log\mathcal{N}(\mathcal{X},\|\cdot\|_\infty,\varepsilon)}{\log\frac{1}{\varepsilon}}\le\varrho, \tag{2.13}$$

或者，等价地，存在常数 $C_{\mathcal{X}}\ge 1$ 和 $\varrho>0$，使得

$$\mathcal{N}(\mathcal{X},\|\cdot\|_\infty,\varepsilon)\le C_{\mathcal{X}}\varepsilon^{-\varrho}. \tag{2.14}$$

所有满足 (2.13) 的 ϱ 的下确界与 $\mathcal{X}$ 的上盒计数维数一致。该假设相当一般，因为它刻画了低维光滑子流形 $\mathcal{M}\subset\mathbb{R}^d$ 的内在维数。当有界的 $\mathcal{X}\subset\mathbb{R}^d$ 具有非空内部时，该假设恰好在 $\varrho=d$ 的情况下成立。在假设6下，我们将证明，所关注算法的学习速率主要取决于输入域的内在维数 ϱ。

备注 2. 在第5.2小节中，我们定义了一个卷积算子 $\mathcal{K}^\sigma * \cdot : \mathcal{L}_2(\mathbb{R}^{2d}) \to \mathcal{H}_{K^\sigma}$，并证明对贝叶斯 ϕ-排序规则 f_ϕ^* 应用该卷积算子可以得到所需的逼近器。因此，需要考虑将 $f_\phi^*: \mathcal{X}^2 \to \mathbb{R}$ 适当地延拓到 $\mathcal{L}_2(\mathbb{R}^{2d})$。在许多情况下，使用零延拓就足够了。然而，如果 $\mathcal{X}$的内在维数 $\varrho < d$较低，那么 $\mathcal{X}$ 可能具有零勒贝格测度，而对 f_ϕ^* 作平凡的零延拓只会在 $\mathcal{L}_2(\mathbb{R}^{2d})$中得到零函数。对于平方损失的情形，我们可以借助惠特尼延拓定理（参见 *[19]*中的定理 2.3.6）解决这一问题：我们直接将 f^*_{square} （定义于式 (2.16)）视为 $\mathcal{L}_2(\mathbb{R}^{2d})$上的函数，并对 f^*_{square} 施加 Besov 光滑性假设。正如 *[18]*中所解释的：若 $\mathcal{X}^2$ 是一个紧致的 C^k-流形，根据惠特尼延拓定理，任意 $f \in C^k(\mathcal{X}^2)$都可以延拓为函数 $\widetilde{f} \in C^k(\mathcal{X}^2_{+\delta})$，其中 $\mathcal{X}^2_{+\delta}$是 $\mathcal{X}^2$ 在 $\mathbb{R}^{2d}$中的 δ-邻域。然而，对于合页损失，情况更具挑战性，因为我们需要将 f^*_{hinge} 延拓到 $\mathcal{L}_2(\mathbb{R}^{2d})$，同时不违反用于界定逼近误差所必需的间隔噪声条件 (5)。据我们所知，目前没有简单的方法可以解决这一问题。在本文中，我们将在命题 5 的证明中给出这种延拓的构造。

在噪声条件，即假设4和假设5，以及假设6下，我们推导出配备 ϕ_{hinge}的高斯排序估计器（1.4）的快速学习速率。令$\mathcal{B}_{\mathbb{R}^d}$ 表示 $\mathbb{R}^d$中的单位球。

定理 3.令 $\mathbf{z} = \{(X_i, Y_i)\}_{i=1}^n$ 为概率分布 P on$\mathcal{X} \times \mathcal{Y}$的独立同分布样本。估计量 $f_{\mathbf{z}}$ 由(1.4)定义，其中 $\lambda > 0, \sigma \in (0,1)$，且 $\phi_{\text{hinge}}(y, y', t) = \max\{1 - \text{sgn}(y - y')t, 0\}$。此外，令 P 对于常数 C_* 和 q 满足假设4，并对于常数 C_{**} 和 β满足假设5。假设存在某个 $r > 0$ ，使得 $\mathcal{X} \subset r\mathcal{B}_{\mathbb{R}^d}$ ，并且对于常数 $C_{\mathcal{X}}$ 和 $\varrho \in (0, d]$满足假设6。那么对于所有 $n \geq 2$、$t > 0$以及 $p \in (0, 1/4]$，以至少 $1 - (c_0 + 5)\exp(-t)$ 的概率，有

$$
\begin{aligned}
&\lambda \|f_{\mathbf{z}}\|^2_{K^\sigma} + \mathcal{R}^{\phi_{\text{hinge}}}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}) \\
&\leq \frac{2^{3d+4} r^{2d}}{\Gamma(d)} \lambda \sigma^{-2d} + \frac{2^{\beta/2+4} C_{**} \Gamma(d + \frac{\beta}{2})}{\Gamma(d)} \sigma^\beta + 36 c_6 \left(\frac{C^*_{\mathcal{X}} C_*^{\frac{1-p}{q+1}}}{\lambda^p p^{2d+1} \sigma^{2\varrho} n} \right)^{\frac{q+1}{q-p+2}} \\
&\quad + \frac{12 C^*_{\mathcal{X}} c_6 (t+1)}{\lambda^p p^{2d+1} \sigma^{2\varrho} n} + \left(\frac{11232 C_*^{\frac{1}{q+1}} t}{n} \right)^{\frac{q+1}{q+2}} + \frac{(2712 + 3c_6)t}{n},
\end{aligned}
\tag{2.15}
$$

其中

$$
C^*_{\mathcal{X}} := 12 C^2_{\mathcal{X}} \binom{4e + 2d}{2d} \frac{(2d+1)^{2d+1}}{2^{2d+1} e^{4d+1}}
$$

且 c_0、c_6 是独立于 λ、σ、n、t 或 p 的常数，其在证明中有明确给出。特别地，选择

$$
\begin{aligned}
&\sigma = n^{-a}, \quad a = \frac{q+1}{\beta(q+2) + 2\varrho(q+1)}; \\
&\lambda = n^{-b}, \quad b \geq \frac{(2d+\beta)(q+1)}{\beta(q+2) + 2\varrho(q+1)}, \quad p = \frac{\log 2}{4 \log n}.
\end{aligned}
$$

那么，对于所有的 $n \geq 2$ 和 $t \geq 1$，我们有

$$
\mathcal{E}(\pi(f_{\mathbf{z}})) \leq \mathcal{E}^{\phi_{\text{hinge}}}(\pi(f_{\mathbf{z}})) \lesssim t n^{-\frac{\beta(q+1)}{\beta(q+2) + 2\varrho(q+1)}} \log^{2d+1} n
$$

其概率至少为 $1-(c_0+5)\exp(-t)$。

对于平方损失的情形， ϕ_{square} 的贝叶斯排序规则按点定义为

$$f^*_{\text{square}}(x,x') := \arg\min_{t\in\mathbb{R}} \eta_+(x,x')\psi_{\text{square}}(t) + \eta_-(x,x')\psi_{\text{square}}(-t)$$

其可明确表示为

$$f^*_{\text{square}}(x,x') := \frac{\eta_+(x,x')-\eta_-(x,x')}{\eta_+(x,x')+\eta_-(x,x')},\ \forall (x,x')\in\mathcal{X}^2. \tag{2.16}$$

函数 f^*_{square} 具有继承自 $\eta_+(x,x')$ 和 $\eta_-(x,x')$ 的光滑性，因此我们可以更直接地推导出逼近误差界。这里我们引入 Besov 光滑性的记号，这一记号也见于 [22, 40, 41, 18]。给定函数 $f:\ \mathbb{R}^{2d}\to\mathbb{R}, h\in\mathbb{R}^{2d}$ 和 $s\in\mathbb{N}$，定义差分算子的 s 次迭代应用。

$$\Delta_h^s f(x) := \sum_{j=0}^{s}(-1)^{s-j}\binom{s}{j} f(x+jh).$$

给定 $\mathcal{X}^2\subset\mathbb{R}^{2d}$ 上的测度 μ，定义 s 阶光滑模

$$\omega_{s,\mathcal{L}_2(\mu)}(f,t) := \sup_{\|h\|_2\le t}\|\Delta_h^s f\|_{\mathcal{L}_2(\mu)}.$$

最后，给定一个 $\alpha>0$ 和集合 $s=\lfloor\alpha\rfloor+1$，定义半范数

$$|f|_{\mathcal{B}^\alpha_{2,\infty}(\mu)} := \sup_{t>0} t^{-\alpha}\omega_{s,\mathcal{L}_2(\mu)}(f,t).$$

在假设4、假设6以及 f^*_{square} 的 Besov 光滑性假设下，我们为配备 ϕ_{square} 的高斯排序估计器 (1.4) 推导出快速学习速率。

定理 4.设 $\mathbf{z}=\{(X_i,Y_i)\}_{i=1}^n$ 是概率分布 P on$\mathcal{X}\times\mathcal{Y}$的独立同分布样本。估计量 $f_{\mathbf{z}}$ 由(1.4)定义，其中 $\lambda>0,\sigma\in$(0,1)，且 $\phi_{\text{square}}(y,y',t)$ =$(1-\text{sgn}(y-y')t)^2$。假设假设6对常数 $C_{\mathcal{X}}$ 和 $\varrho\in(0,d]$成立。进一步假设f^*平方 $\in\mathcal{L}_2(\mathbb{R}^{2d})$，并且$|f^*$平方$|_{\mathcal{B}^\alpha_2,\infty}(P^2_{\mathcal{X}})<\infty$对某个 $\alpha>0$成立。那么对于所有 $n\ge 2, t>0$和 $p\in(0,1/4]$，以至少 $1-(c_0+5)\exp(-t)$ 的概率，有

$$\begin{aligned}
&\lambda\|f_{\mathbf{z}}\|^2_{K^\sigma}+\mathcal{R}^{\phi_{\text{square}}}(\pi(f_{\mathbf{z}}))-\mathcal{R}^{\phi_{\text{square}}}(f^*_{\text{square}})\\
\le\ &\frac{2^{2s+3}\|f^*_{\text{square}}\|^2_{\mathcal{L}_2(\mathbb{R}^{2d})}}{\pi^d}\lambda\sigma^{-2d}+2^{3-\alpha}\left(\frac{\Gamma\left(d+\frac{\alpha}{2}\right)}{\Gamma(d)}\right)^2|f^*_{\text{square}}|^2_{\mathcal{B}^\alpha_{2,\infty}(P^2_{\mathcal{X}})}\sigma^{2\alpha}\\
&+\frac{C^*_{\mathcal{X}}(96c_6+72c_6t)}{\lambda^p p^{2d+1}\sigma^{2\varrho}n}+\frac{(33552+1824\cdot 2^{2s}+3c_6)t}{n},
\end{aligned}\tag{2.17}$$

其中

$$C^*_{\mathcal{X}} := 12C^2_{\mathcal{X}}\binom{4e+2d}{2d}\frac{(2d+1)^{2d+1}}{2^{2d+1}e^{4d+1}}$$

且 c_0,c_6 是与 λ,σ,n,t 或 p 无关的常数，它们在证明中有明确给出。特别地，选择

$$\sigma=n^{-a},\quad a=\frac{1}{2\alpha+2\varrho};$$

$$\lambda = n^{-b}, \quad b \geq \frac{\alpha + d}{\alpha + \varrho}, \quad p = \frac{\log 2}{4 \log n}.$$

那么对于所有 $n \geq 2, t \geq 1$，我们有

$$\mathcal{E}(\pi(f_{\mathbf{z}})) \lesssim \sqrt{\mathcal{E}^{\phi_{\text{square}}}(\pi(f_{\mathbf{z}}))} \lesssim \sqrt{t} n^{-\frac{\alpha}{2(\alpha+\varrho)}} \log^{d+\frac{1}{2}} n$$

以至少 $1 - (c_0 + 5) \exp(-t)$ 的概率。此外，如果 P还满足带有 $q > 0$ 的假设 4，那么以至少 $1 - (c_0 + 5) \exp(-t)$ 的概率，有

$$\mathcal{E}(\pi(f_{\mathbf{z}})) \lesssim t^{\frac{q+1}{q+2}} n^{-\frac{(q+1)\alpha}{(q+2)(\alpha+\varrho)}} \log^{\frac{(q+1)(2d+1)}{q+2}} n.$$

最后，我们想指出，根据 [18], 中的命题 3.2，满足 $|f|_{\mathcal{B}^{\alpha}_{2,\infty}(P^2_{\mathcal{X}})} < \infty$ 的函数类 f确实非常丰富，即使 $\mathcal{X}$ 的上盒计数维数严格小于 d。例如，若 f 是定义在 $\mathcal{X}^2$ 上的 H¨older α-连续函数，那么对于每个满足 supp$\mu \subset \mathcal{X}^2$ 的测度 μ，都有 $|f|_{\mathcal{B}^{\alpha}_{2,\infty}(\mu)} < \infty$。在建立一些预备结果后，定理 3 和定理 4 的证明将在第 5 节给出。

3 关于相关工作的讨论

在本节中，我们将收敛分析与文献中的一些已有结果进行比较。本文涉及多个快速发展的研究领域，因此我们只列出特别相关的文献。更多参考文献可见于这些文献中。

本文研究旨在学习双变量排序规则 $f: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 的成对排序模型。与基于评分的排序模型相比，我们的模型更为一般。在基于评分的模型中，我们学习评分函数 $s: \mathcal{X} \to \mathbb{R}$，并将排序规则构造为 $f(x, x') = s(x) - s(x')$。一些流行的排序算法，包括 [20], 中的 RankSVM、[7] 中的 RankNet 和 [30, 14, 31] 中的 RankRLS，都与基于评分的排序模型密切相关。针对基于评分的排序模型的误差分析，[1] 通过算法稳定性为标签 Y 为二元值的二部排序问题证明了泛化界。[11] 从算子逼近的角度，为 RankRLS 推导出了与容量无关的学习速率。随后，[52] 利用覆盖数和 U统计量的 Hoeffding 分解，对 RankRLS 开展了与容量相关的泛化分析。

在一些学习任务中，例如学习两个对象之间的二元关系，我们可以为每个对象分配一个顶点，并用一条边表示两个顶点之间的二元关系，这自然形成了一个图结构。记 $\mathcal{X}$ 为顶点集，$\mathcal{X}^2$ 为边集。半正定核 $G: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 诱导出一个再生核希尔伯特空间 $\mathcal{H}_G$，在其中我们可以学习定义在顶点集上的函数 $f: \mathcal{X} \to \mathbb{R}$。然而，在这类图学习任务中，如果我们想要确定二元关系，成对学习更加适用，因为在这种情况下，我们需要学习定义在边集上的函数 $f: \mathcal{X}^2 \to \mathbb{R}$。事实上，定义在顶点之间的核函数 G 可以通过克罗内克积诱导出所谓的克罗内克积成对核 $K: \mathcal{X}^2 \times \mathcal{X}^2 \to \mathbb{R}$，其定义为 $K((x, x'), (u', u')) := G(x, u)G(x', u')$。克罗内克核岭回归（KKRR）是一种基于克罗内克积成对核 K的成对学习算法，其中诱导的正则化经验最小化问题在再生核希尔伯特空间 $\mathcal{H}_K$中求解。在第 5.1 小节中，我们将看到，再生核希尔伯特空间 $\mathcal{H}_K$ 由定义在 $\mathcal{X}^2$ 上的函数组成，并且我们可以将任意 $f \in \mathcal{H}_K$ 分解为其对称部分和

反对称部分，对应于 $\mathcal{H}_K$ 的两个子空间的直和，并且也导致了成对核 K 的分解。成对高斯核 K^σ 确实是 $\mathcal{X}^2$ 上定义为 $\widetilde{K}^\sigma((x,x'),(u,u')) := \exp\left(-\|(x,x') - (u,u')\|_2^2/\sigma^2\right)$ 的传统高斯核 $\widetilde{K}^\sigma$ 的反对称部分，因此相应的再生核希尔伯特空间 $\mathcal{H}_{K^\sigma}$是 $\mathcal{H}_{\widetilde{K}^\sigma}$ 的直和中的一个子空间。尽管 $\mathcal{H}_{K^\sigma}$是一个较小的假设空间，但在仅学习反对称关系时，仍然可以保持核的通用性。也就是说，如果克罗内克积成对核 K 能够在 $\mathcal{X}^2$ 上以任意高的精度逼近任意连续函数，那么成对核的反对称部分也能够以任意高的精度逼近任意连续的反对称函数，参见 [42] 的定理 III.4。因此，当我们使用包括KKRR 在内的成对学习算法来学习数据之间的反对称关系时，只需将估计量限制在假设空间$\mathcal{H}_{K^\sigma}$ 中即可。使用核的对称部分来学习对称关系时，也有类似的观察结果。更多详情可参阅 [6, 42, 29, 43] 。

由于成对排序模型更为一般，已有一些理论框架用于分析其收敛行为。排序与分类或回归在收敛分析方面的主要区别在于，在排序问题中，收敛分析的随机部分涉及二阶 U-过程，而不是独立同分布随机变量的求和。 [12] 和 [33] 将 Hoeffding 分解应用于未正则化排序算法，并获得了比 $\frac{1}{\sqrt{n}}$更快的泛化界。Hoeffding 分解将排序算法的样本误差分解为基于独立同分布随机变量之和的经验项和一个退化 U 过程。在我们的工作中，我们充分利用 Talagrand 不等式、局部 Rademacher 分析以及假设空间的容量信息，分别对经验项和退化部分推导出紧的界（见第 4 节）。对于正则化成对排序， [9] 研究了 ℓ_1-范数正则化的 SVM 排序，并在类似于假设4的噪声条件下建立了学习速率。 [34] 证明了高维情形下带有 Lasso 惩罚项的参数化成对排序的预言机不等式。与我们的工作不同， [10, 25, 44] 研究了采用固定核的正则化成对排序，并假设逼近误差具有 γ-衰减速率，从而推导出学习速率。我们进一步强调我们的结果与 [12, 33, 10, 25, 44]工作的差异。[12]提出了这样一个问题：对于排序中的超额风险，是否能够获得具有快速速率的泛化界。他们对具有非凸 $0-1$ 损失的未正则化排序估计量给出了肯定答案，即 $\phi(y,y',t) = \psi_{0-1}(\mathrm{sgn}(y-y')t)$，其中$\psi_{0-1}(t) := \mathbb{I}_{[0,\infty)}(t)$，并将凸风险最小化的情形留待未来研究。 [33] 的工作推进了这一研究方向，建立了采用凸间隔损失时超额排序风险的泛化界，其速率优于 $\frac{1}{\sqrt{n}}$ 。这两篇论文都在一般假设空间中构造排序估计量，并要求最优排序规则属于该函数空间。因此，它们无需分析逼近误差。此外，[12], 还提出了一种噪声条件；不同噪声条件之间的比较将推迟到第 5.5 小节。[10, 44] 的工作还发展了依赖容量的经验过程技术，用于分析正则化排序估计量，其中容量条件基于上确界范数 $\|\cdot\|_\infty$，这比本文采用的经验 $\mathcal{L}_2$-半范数限制更强。此外， [10]仅关注合页损失，并推导出 $n^{-\frac{\gamma}{2-\tau+p}()^{\gamma+p}}$ 的速率，该速率慢于注1中推导的 $n^{-\frac{\gamma}{2-\tau-p+p\tau}()^{\gamma+p}}$ 速率。 [25] 研究了与我们类似的一般损失函数，并在假设空间不满足任何容量条件、偏差和方差也不满足 Bernstein 条件的情况下，推导出了 $n^{-\frac{\gamma}{\gamma+1}}$ 的速率。我们注意到，即使在方差界 (2.6) 得到平凡满足的情形 $\tau = 0$

下，我们给出的速率 $n^{-\frac{\gamma}{2-p}}()^{\gamma+p}$ 仍然快于 $n^{-\frac{\gamma}{\gamma+1}}$，前提是 $\gamma < 1$。 [44] 在回归情形下考虑了最小二乘损失 $\phi_{\text{ls}}(y, y', t) := (y - y' - t)^2$。对于 $\phi_{\text{ls}}(y, y', t)$损失，我们可以将针对平方损失所发展的分析应用于 $n^{-\frac{\gamma}{\gamma+p}}$的情形，从而得到其速率；该速率远快于 [44]通过 $n^{-\frac{2\gamma}{p+1}}()^{\gamma+4}$ 所获得的速率。我们还注意到，速率 $n^{-\frac{\gamma}{\gamma+p}}$ 实际上达到了 [38]在 $\mathbb{R}^{2d}$上逼近函数时推导出的众所周知的极小极大下界速率。从这一角度看，定理 1 和定理 2 中建立的预言机不等式比现有结果更一般，但能够对超额风险给出更紧的界。此外， [35] 表明，在描述高斯再生核希尔伯特空间的逼近能力时，逼近误差的 γ-衰减假设确实是一个非常严格的假设，因此，采用变化高斯核的排序设定无法通过 [10, 25, 44],的工作中所发展的方法处理，因为该工作仅关注采用固定成对核的正则化排序。

数据的内在维数可用于改善学习速率对维数的依赖。描述内在维数结构时，最常见的概念或许是假设 $\mathcal{X} \subset \mathbb{R}^d$ 是一个低维子流形，参见例如 [47, 48,49]。本文采用的更一般概念基于盒计数维数，它大大推广了流形假设。据我们所知，本文首次研究了在数据内在维数与其环境空间维数之差较大的假设下，正则化高斯排序估计量的学习行为。

总之，本文在相当一般的假设下，研究了具有一般凸损失的正则化成对排序问题。本文建立的预言机不等式可构建一个简洁的收敛分析框架，从而显著改善现有研究的学习速率。此外，该不等式还使我们能够推导出高斯排序估计量的快速学习速率，并通过对数据采用低内在维数假设来避免维数灾难。

4 一般预言机不等式的证明

本节将详细证明定理 1 中给出的一般预言机不等式及其在定理 2 中的变体。按照标准的误差分解过程，我们采用 Hoeffding 分解，以及基于 U 过程和局部 Rademacher 分析的集中估计，来界定误差项中的随机部分。

4.1 误差分解与 Hoeffding 分解

给定一个 $f_0 \in \mathcal{H}_K$，根据 $f_{\mathbf{z}}$ 的定义，我们有 $\mathcal{R}^{\phi}_{\mathbf{z}}(f_{\mathbf{z}}) + \lambda\|f_{\mathbf{z}}\|^2_K \le \mathcal{R}^{\phi}_{\mathbf{z}}(f_0) + \lambda\|f_0\|^2_K$. 然后

$$\begin{aligned}
&\lambda\|f_{\mathbf{z}}\|^2_K + \mathcal{R}^{\phi}(f_{\mathbf{z}}) - \mathcal{R}^{\phi}(f^*_{\phi}) \\
\le\ &\big(\lambda\|f_0\|^2_K + \mathcal{R}^{\phi}(f_0) - \mathcal{R}^{\phi}(f^*_{\phi})\big) + \big(\mathcal{R}^{\phi}_{\mathbf{z}}(f_0) - \mathcal{R}^{\phi}_{\mathbf{z}}(f^*_{\phi}) - \mathcal{R}^{\phi}(f_0) + \mathcal{R}^{\phi}(f^*_{\phi})\big) \\
&+ \big(\mathcal{R}^{\phi}(f_{\mathbf{z}}) - \mathcal{R}^{\phi}(f^*_{\phi}) - \mathcal{R}^{\phi}_{\mathbf{z}}(f_{\mathbf{z}}) + \mathcal{R}^{\phi}_{\mathbf{z}}(f^*_{\phi})\big).
\end{aligned}$$

对于可在某个 $M > 0$ 处截断的损失函数 ϕ，由于 $\phi(y, y', \pi(t)) \le \phi(y, y', t)$，我们有 $\mathcal{R}^{\phi}_{\mathbf{z}}(\pi(f_{\mathbf{z}})) \le \mathcal{R}^{\phi}_{\mathbf{z}}(f_{\mathbf{z}})$，这意味着 $\mathcal{R}^{\phi}_{\mathbf{z}}(\pi(f_{\mathbf{z}})) + \lambda\|f_{\mathbf{z}}\|^2_K \le \mathcal{R}^{\phi}_{\mathbf{z}}(f_0) + \lambda\|f_0\|^2_K$ 对于

任意 $f_0 \in \mathcal{H}_K$。因此，我们可以通过以下误差分解对 $\lambda\|f_{\mathbf{z}}\|_K^2 + \mathcal{R}^\phi(\pi(f_{\mathbf{z}})) - \mathcal{R}^\phi(f_\phi^*)$ 给出界

$$\lambda\|f_{\mathbf{z}}\|_K^2 + \mathcal{R}^\phi(\pi(f_{\mathbf{z}})) - \mathcal{R}^\phi(f_\phi^*) \le \mathcal{A}(\lambda, f_0) + \mathcal{S}_{\mathbf{z}}(f_0) + \mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}})) \tag{4.1}$$

其中

$$\begin{aligned}
\mathcal{A}(\lambda, f_0) &:= \lambda\|f_0\|_K^2 + \mathcal{R}^\phi(f_0) - \mathcal{R}^\phi(f_\phi^*),\\
\mathcal{S}_{\mathbf{z}}(f_0) &:= \mathcal{R}_{\mathbf{z}}^\phi(f_0) - \mathcal{R}_{\mathbf{z}}^\phi(f_\phi^*) - \mathcal{R}^\phi(f_0) + \mathcal{R}^\phi(f_\phi^*),\\
\mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}})) &:= \mathcal{R}^\phi(\pi(f_{\mathbf{z}})) - \mathcal{R}^\phi(f_\phi^*) - \mathcal{R}_{\mathbf{z}}^\phi(\pi(f_{\mathbf{z}})) + \mathcal{R}_{\mathbf{z}}^\phi(f_\phi^*).
\end{aligned}$$

第一项 $\mathcal{A}(\lambda, f_0)$反映了使用 f_0 来逼近 f_ϕ^* 所产生的逼近误差。另两项 $\mathcal{S}_{\mathbf{z}}(f_0)$ 和 $\mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}}))$ 构成了误差分解中的随机部分。

在假设1下，我们可以通过 U-统计量的Bernstein不等式来界定 $\mathcal{S}_{\mathbf{z}}(f_0)$，参见下文所述的[2],。

引理1. 设$\mathbf{z} = \{Z_i\}_{i=1}^n$ 为一个概率分布 P在可测空间 $\mathcal{Z}$上的独立同分布样本， $h: \mathcal{Z}\times\mathcal{Z} \to \mathbb{R}$ 为一个对称可测函数，满足 $\mathbb{E}[h(Z, Z')] = 0$ 且$\|h\|_\infty = b$。具有核 h 的 U-统计量定义为

$$U_{\mathbf{z}}(h) = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\ne i} h(Z_i, Z_j).$$

设 $f: \mathcal{Z}\to\mathbb{R}$定义为 $f(z) := \mathbb{E}[h(z, Z')]$和 $\zeta^2 := \mathbb{E}[f(Z)^2]$。那么对于所有 $t > 0$，我们有

$$P\left(\sqrt{n}|U_{\mathbf{z}}(h)| \ge t\right) \le 4\exp\left(-\frac{t^2}{8\zeta^2 + \left(\frac{64}{\sqrt{n-1}} + \frac{1}{3\sqrt{n}}\right)bt}\right),$$

这意味着

$$P\left(U_{\mathbf{z}}(h) \ge \sqrt{\frac{8\zeta^2 t}{n}} + \frac{150bt}{n}\right) \le 2\exp(-t). \tag{4.2}$$

回顾一下， $\mathcal{S}_{\mathbf{z}}(f_0) = \mathcal{R}_{\mathbf{z}}^\phi(f_0) - \mathcal{R}_{\mathbf{z}}^\phi(f_\phi^*) - \mathcal{R}^\phi(f_0) + \mathcal{R}^\phi(f_\phi^*)$。

命题1. 给定满足$\|\phi_{f_0}\|_\infty \le B_0$ 的 $f_0 \in \mathcal{H}_K$，其中某个常数为 $B_0 \ge 0$。若假设1成立，则以至少 $1 - 4\exp(-t)$ 的概率，

$$\mathcal{S}_{\mathbf{z}}(f_0) \le \mathcal{R}^\phi(f_0) - \mathcal{R}^\phi(f_\phi^*) + \left(\frac{8Vt}{n}\right)^{\frac{1}{2-\tau}} + \frac{300Bt}{n} + \frac{152B_0t}{n}. \tag{4.3}$$

证明. 定义 v_f: 将 $\mathcal{Z}\times\mathcal{Z}\to\mathbb{R}$ 定义为 $v_f := \phi_f - \phi_{f_\phi^*}$，则

$$\mathcal{S}_{\mathbf{z}}(f_0) = \mathcal{R}_{\mathbf{z}}^\phi(f_0) - \mathcal{R}_{\mathbf{z}}^\phi(f_\phi^*) - \mathcal{R}^\phi(f_0) + \mathcal{R}^\phi(f_\phi^*) = \mathbb{E}_{\mathbf{z}}v_{f_0} - \mathbb{E}v_{f_0}.$$

我们进一步将 $\mathbb{E}_{\mathbf{z}}v_{f_0} - \mathbb{E}v_{f_0}$分解为两项，引入截断 $\pi(f_0)$，其定义为

$$\mathbb{E}_{\mathbf{z}}v_{f_0} - \mathbb{E}v_{f_0} = [\mathbb{E}_{\mathbf{z}}(v_{f_0} - v_{\pi(f_0)}) - \mathbb{E}(v_{f_0} - v_{\pi(f_0)})] + (\mathbb{E}_{\mathbf{z}}v_{\pi(f_0)} - \mathbb{E}v_{\pi(f_0)}).$$

上面的每一项都是一个 U-统计量与其期望之差，因此我们可以应用引理1中的Bernstein不等式对其进行界定。为此，我们需要为相关的U统计量建立上确界范数界和方差界。

我们首先对 $\mathbb{E}_{\mathbf{z}}(v_{f_0} - v_{\pi(f_0)}) - \mathbb{E}(v_{f_0} - v_{\pi(f_0)})$ 进行界定。由于

$$\phi_{f_0}(z, z') - \phi_{\pi(f_0)}(z, z') = \phi(y, y', f_0(x, x')) - \phi(y, y', \pi(f_0)(x, x')) \geq 0$$

对所有 $(z, z') \in \mathcal{Z} \times \mathcal{Z}$，我们有

$$v_{f_0} - v_{\pi(f_0)} = \phi_{f_0} - \phi_{\pi(f_0)} : \mathcal{Z} \times \mathcal{Z} \to [0, B_0]$$

and

$$\|v_{f_0} - v_{\pi(f_0)} - \mathbb{E}(v_{f_0} - v_{\pi(f_0)})\|_\infty \leq B_0.$$

定义 $g: \mathcal{Z} \to \mathbb{R}$ 为

$$g(z) := \mathbb{E}[v_{f_0}(z, Z') - v_{\pi(f_0)}(z, Z') - \mathbb{E}(v_{f_0} - v_{\pi(f_0)})].$$

Then

$$\mathbb{E}[g(Z)^2] \leq \mathbb{E}\left[\left(\mathbb{E}[v_{f_0}(Z, Z') - v_{\pi(f_0)}(Z, Z')]\right)^2\right] \leq B_0 \mathbb{E}(v_{f_0} - v_{\pi(f_0)}).$$

现在，我们将 Bernstein不等式 (4.2) 应用于零均值对称函数 $v_{f_0} - v_{\pi(f_0)} - \mathbb{E}(v_{f_0} - v_{\pi(f_0)})$，可得

$$P\left(\mathbb{E}_{\mathbf{z}}(v_{f_0} - v_{\pi(f_0)}) - \mathbb{E}(v_{f_0} - v_{\pi(f_0)}) \geq \sqrt{\frac{8B_0\mathbb{E}(v_{f_0} - v_{\pi(f_0)})t}{n}} + \frac{150B_0t}{n}\right) \leq 2\exp(-t).$$

根据基本不等式 $\sqrt{ab} \leq (a+b)/2$，我们有

$$\sqrt{\frac{8B_0\mathbb{E}(v_{f_0} - v_{\pi(f_0)})t}{n}} \leq \mathbb{E}(v_{f_0} - v_{\pi(f_0)}) + \frac{2B_0t}{n}.$$

因此，以至少 $1 - 2\exp(-t)$ 的概率，有

$$\mathbb{E}_{\mathbf{z}}(v_{f_0} - v_{\pi(f_0)}) - \mathbb{E}(v_{f_0} - v_{\pi(f_0)}) \leq \mathbb{E}(v_{f_0} - v_{\pi(f_0)}) + \frac{152B_0t}{n}. \tag{4.4}$$

为了界定第二项 $\mathbb{E}_{\mathbf{z}}v_{\pi(f_0)} - \mathbb{E}v_{\pi(f_0)}$，定义 $u: \mathcal{Z} \to \mathbb{R}$ 为

$$u(z) := \mathbb{E}[v_{\pi(f_0)}(z, Z') - \mathbb{E}v_{\pi(f_0)}].$$

根据假设1中的(2.5)和(2.6)，我们有

$$v_{\pi(f_0)} = \phi_{\pi(f_0)} - \phi_{f_\phi^*} \in [-B, B],$$

$$\|v_{\pi(f_0)} - \mathbb{E}v_{\pi(f_0)}\|_\infty \leq 2B,$$

and

$$\mathbb{E}[u(Z)^2] \leq \mathbb{E}\left[\left(\mathbb{E}[v_{\pi(f_0)}(Z, Z')]\right)^2\right] = \mathbb{E}\left(Q\phi_{\pi(f_0)}(Z) - Q\phi_{f_\phi^*}(Z)\right)^2 \leq V(\mathbb{E}v_{\pi(f_0)})^\tau.$$

我们将 Bernstein不等式应用于零均值对称函数 $v_{\pi(f_0)}-\mathbb{E}v_{\pi(f_0)}$，从而证明了

$$P\left(\mathbb{E}_{\mathbf{z}}(v_{\pi(f_0)})-\mathbb{E}v_{\pi(f_0)}\geq\sqrt{\frac{8V(\mathbb{E}v_{\pi(f_0)})^{\tau}t}{n}}+\frac{300Bt}{n}\right)\leq 2\exp(-t).$$

如果 $\tau>0$，我们通过令 $ab\leq a^q/q+b^p/p$ 使用 Young 不等式

$$a=\sqrt{\frac{2^{3-\tau}\tau^{\tau}Vt}{n}},\quad b=\left(\frac{2\mathbb{E}v_{\pi(f_0)}}{\tau}\right)^{\tau/2},\quad q=\frac{2}{2-\tau},\quad p=\frac{2}{\tau},$$

从而得到

$$\sqrt{\frac{8V(\mathbb{E}v_{\pi(f_0)})^{\tau}t}{n}}=ab\leq\frac{2-\tau}{2}\cdot\left(\frac{2^{3-\tau}\tau^{\tau}Vt}{n}\right)^{\frac{1}{2-\tau}}+\mathbb{E}v_{\pi(f_0)}\leq\left(\frac{8Vt}{n}\right)^{\frac{1}{2-\tau}}+\mathbb{E}v_{\pi(f_0)}.$$

由于$\mathbb{E}v_{\pi(f_0)}\geq 0$，该不等式对 $\tau=0$ 也成立。因此，以至少$1-2\exp(-t)$的概率，有

$$\mathbb{E}_{\mathbf{z}}(v_{\pi(f_0)})-\mathbb{E}v_{\pi(f_0)}\leq\mathbb{E}v_{\pi(f_0)}+\left(\frac{8Vt}{n}\right)^{\frac{1}{2-\tau}}+\frac{300Bt}{n}.\tag{4.5}$$

因此，结合（4.4）和（4.5），我们得出所需的界（4.3）。证明至此完成。

□

对 $\mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}}))$进行界定更为复杂。回顾,

$$\mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}}))=\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}}))-\mathcal{R}^{\phi}(f_{\phi}^{*})-\mathcal{R}_{\mathbf{z}}^{\phi}(\pi(f_{\mathbf{z}}))+\mathcal{R}_{\mathbf{z}}^{\phi}(f_{\phi}^{*}).$$

我们将 U-统计量的Hoeffding分解（参见 [24]）应用于 $\mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}}))$，其形式为

$$\begin{aligned}&\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}}))-\mathcal{R}^{\phi}(f_{\phi}^{*})-\mathcal{R}_{\mathbf{z}}^{\phi}(\pi(f_{\mathbf{z}}))+\mathcal{R}_{\mathbf{z}}^{\phi}(f_{\phi}^{*})\\&=2Q_{\mathbf{z}}[\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}}))-\mathcal{R}^{\phi}(f_{\phi}^{*})-Q\phi_{\pi(f_{\mathbf{z}})}+Q\phi_{f_{\phi}^{*}}]-U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})}-h_{f_{\phi}^{*}})\\&=:2\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}}))-\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}})).\end{aligned}\tag{4.6}$$

这里的泛函 $Q_{\mathbf{z}}$、$U_{\mathbf{z}}$ 和函数 $h_f:\ \mathcal{Z}\times\mathcal{Z}\to\mathbb{R}$ 定义如下

$$\begin{aligned}Q_{\mathbf{z}}(g)&:=\frac{1}{n}\sum_{i=1}^{n}g(Z_i),\forall g:\mathcal{Z}\to\mathbb{R},\\U_{\mathbf{z}}(g')&:=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}g'(Z_i,Z_j),\forall g':\mathcal{Z}\times\mathcal{Z}\to\mathbb{R},\\h_f(z,z')&:=\phi_f(z,z')-Q\phi_f(z)-Q\phi_f(z')+\mathcal{R}^{\phi}(f).\end{aligned}$$

Hoeffding 分解将 $\mathcal{S}_{\mathbf{z}}(\pi(f_{\mathbf{z}}))$ 拆分为一组独立同分布随机变量 $\mathcal{S}_{\mathbf{z},\ 1}(\pi(f_{\mathbf{z}}))$ 之和，称为经验项；以及一个退化的 U-统计量 $\mathcal{S}_{\mathbf{z},\ 2}(\pi(f_{\mathbf{z}}))$，称为退化项，其条件期望为零

$$\mathbb{E}[h_{\pi(f_{\mathbf{z}})}(Z,Z')-h_{f_{\phi}^{*}}(Z,Z')|Z=z]=0.$$

我们分别在以下两个小节中推导经验项和退化项的界。

4.2 对经验项进行界定

在本小节中，我们估计经验项。

$$\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}})) = Q_{\mathbf{z}}[\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi}(f_{\phi}^{*}) - Q\phi_{\pi(f_{\mathbf{z}})} + Q\phi_{f_{\phi}^{*}}]$$

它还可以进一步由以 $\mathcal{H}_K$ 为指标的经验过程的上确界范数来界定，即

$$\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}})) \leq \sup_{f\in\mathcal{H}_K} Q_{\mathbf{z}}[\mathcal{R}^{\phi}(\pi(f)) - \mathcal{R}^{\phi}(f_{\phi}^{*}) - Q\phi_{\pi(f)} + Q\phi_{f_{\phi}^{*}}].$$

在本小节余下部分，我们将结合 Talagrand 浓缩不等式与局部 Rademacher 平均，为经验过程的上确界范数推导一个紧界。我们的估计思路可见 [36]，第 7 章，其中作者针对支持向量机的分类建立了精细的预言机不等式。

定义 $s_f, g_{f,\ r}\colon\ \mathcal{Z}\to\mathbb{R}$ 为 $s_f\colon = Q\phi_f - Q\phi_{f_{\phi}^{*}}$，其中 $Q\phi_f$ 由 (2.4) 给出，且

$$g_{f,r} := \frac{\mathbb{E}s_{\pi(f)} - s_{\pi(f)}}{\lambda\|f\|_K^2 + \mathbb{E}s_{\pi(f)} + r}$$

对所有 $f\in\mathcal{H}_K$ 和 $r>0.$，在假设1下，我们应用 Talagrand 不等式得到以下引理。

引理 2. 若假设1成立，则对所有 $t>0$，以至少 $1-\exp(-t)$ 的概率，

$$\sup_{f\in\mathcal{H}_K}|Q_{\mathbf{z}}(g_{f,r})| \leq \frac{5}{4}\mathbb{E}\sup_{f\in\mathcal{H}_K}|Q_{\mathbf{z}}(g_{f,r})| + \sqrt{\frac{2Vt}{nr^{2-\tau}}} + \frac{28Bt}{3nr}.$$

证明。与命题1的证明类似，根据假设1中的 (2.5)，显然 $\|g_{f,\ r}\|_\infty \leq 2Br^{-1}$。对于方差界，若 $\tau>0$，利用 Young 不等式，我们可以证明 $(qa)^{2/q}(pb)^{2/p}\leq(a+b)^2$，然后令 $a=r, b=\mathbb{E}s_{\pi(f)}$，$q=2/(2-\tau)$ $p=2/\tau$，可得

$$\mathbb{E}g_{f,r}^2 \leq \frac{\mathbb{E}s_{\pi(f)}^2}{(\mathbb{E}s_{\pi(f)}+r)^2} \leq \frac{(2-\tau)^{2-\tau}\tau^{\tau}V}{4r^{2-\tau}} \leq Vr^{\tau-2}$$

这对 $\tau=0$ 也成立。因此，我们可以将 Talagrand 不等式（参见[36], 中的定理 7.5）应用于 $(g_{f,\ r})_{f\in\mathcal{H}_K}$，从而得到所需的界。证明完毕。 □

为了推导 $\sup_{f\in\mathcal{H}_K}|Q_{\mathbf{z}}(g_{f,\ r})|$ 的上界，我们采用剥层法、对称化和 Dudley 链式法等标准工具来估计 $\mathbb{E}\sup_{f\in\mathcal{H}_K}|Q_{\mathbf{z}}(g_{f,\ r})|$。定义

$$\begin{aligned}
r^* &:= \inf\{\lambda\|f\|_K^2 + \mathbb{E}s_{\pi(f)} : f\in\mathcal{H}_K\},\\
\mathcal{F}_r &:= \{f\in\mathcal{H}_K : \lambda\|f\|_K^2 + \mathbb{E}s_{\pi(f)} \leq r\},\ \forall r>r^*,\\
\mathcal{S}_r &:= \{s_{\pi(f)} : f\in\mathcal{F}_r\},\ \forall r>r^*,
\end{aligned}$$

以及 $\mathcal{S}_r$ 的经验 Rademacher 复杂度

$$\mathrm{Rad}(\mathcal{S}_r,\mathbf{z}) := \mathbb{E}_{\varepsilon}\sup_{f\in\mathcal{F}_r}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i s_{\pi(f)}(Z_i)\right|,$$

其中 $\{\varepsilon_i\}_{i=1}^n$ 是拉德马赫序列，即 $\{\varepsilon_i\}_{i=1}^n$ 是从 $\{-1,1\}$中均匀选取的独立同分布随机变量。以下两个引理给出了剥层法和对称化论证，以处理 $\mathbb{E}\sup_{f\in\mathcal{F}_r}|Q_{\mathbf{z}}(\mathbb{E}s_{\pi(f)}-s_{\pi(f)})|$；根据 [36] 中的定理 7.7 和命题 7.10 即可证明。

引理 3. 设 $\varphi:(r^*,\infty)\to[0,\infty)$，使得 $\varphi(4r)\le 2\varphi(r)$ 并且

$$\mathbb{E}\sup_{f\in\mathcal{F}_r}|Q_{\mathbf{z}}(\mathbb{E}s_{\pi(f)}-s_{\pi(f)})|\le\varphi(r)$$

对所有 $r>r^*$。然后对于所有 $r>r^*$，我们有

$$\mathbb{E}\sup_{f\in\mathcal{H}_K}|Q_{\mathbf{z}}(g_{f,r})|\le\frac{4\varphi(r)}{r}.$$

引理 4。

$$\mathbb{E}\sup_{f\in\mathcal{F}_r}|Q_{\mathbf{z}}(\mathbb{E}s_{\pi(f)}-s_{\pi(f)})|\le 2\mathbb{E}\,\mathrm{Rad}(\mathcal{S}_r,\mathbf{z}).$$

然后，我们应用 [36] 的定理 7.16，推导出以下引理。

引理 5。如果假设1和假设2成立，则存在常数 $k_1>0$ 和$k_2>0$ 使得

$$\begin{aligned}\mathbb{E}\,\mathrm{Rad}(\mathcal{S}_r,\mathbf{z})\le\max\Bigg\{&2^{p_1}k_1a_1^{p_1}\left(\frac{r}{\lambda}\right)^{\frac{p_1}{2}}(Vr^\tau)^{\frac{1-p_1}{2}}L^{p_1}n^{-\frac12},\\&4^{\frac{p_1}{1+p_1}}k_2a_1^{\frac{2p_1}{1+p_1}}\left(\frac{r}{\lambda}\right)^{\frac{p_1}{1+p_1}}B^{\frac{1-p_1}{1+p_1}}L^{\frac{2p_1}{1+p_1}}n^{-\frac{1}{1+p_1}}\Bigg\}.\end{aligned}$$

证明。我们首先建立 $s_{\pi(f)}\in\mathcal{S}_r$. 的上确界范数界和方差界。根据假设 1，我们有 $\|s_{\pi(f)}\|_\infty\le B$，并且

$$\mathbb{E}(s_{\pi(f)}^2)=\mathbb{E}(Q\phi_{\pi(f)}-Q\phi_{f_\phi^*})^2\le V\big(\mathbb{E}(Q\phi_{\pi(f)}-Q\phi_{f_\phi^*})\big)^\tau=V(\mathbb{E}s_{\pi(f)})^\tau\le Vr^\tau.$$

对于 $\mathbf{z}=\{(X_i,Y_i)\}_{i=1}^n$，在 $\mathcal{S}_r$ 上，经验范数$\|\cdot\|_{\mathcal{L}_2(P_{\mathbf{z}}^n)}$定义为

$$\|s_{\pi(f)}\|_{\mathcal{L}_2(P_{\mathbf{z}}^n)}:=Q_{\mathbf{z}}^{\frac12}(s_{\pi(f)}^2)=\left(\frac1n\sum_{i=1}^n s_{\pi(f)}(Z_i)^2\right)^{\frac12}.$$

根据 [36], 的定理 7.16，剩下需要给 $\mathbb{E}e_i(\mathrm{id}:\mathcal{S}_r\to\mathcal{L}_2(P_{\mathbf{z}}^n))$ 找一个界。对于所有 $(y,y')\in\mathcal{Y}\times\mathcal{Y}$，函数 $\phi(y,y',\cdot)$ 在 $[-M,M]$ 上局部满足 L-Lipschitz 条件。因此，对于所有 $s_{\pi(f)},s_{\pi(f')}\in\mathcal{S}_r$，以及 $\mathbf{z}=\{(X_i,Y_i)\}_{i=1}^n$，我们有

$$\begin{aligned}\|s_{\pi(f)}-s_{\pi(f')}\|_{\mathcal{L}_2(P_{\mathbf{z}}^n)}^2&=\frac1n\sum_{i=1}^n\big(s_{\pi(f)}(Z_i)-s_{\pi(f')}(Z_i)\big)^2\\&=\frac1n\sum_{i=1}^n\left(\mathbb{E}\Big[\phi(Y_i,Y',\pi(f)(X_i,X'))-\phi(Y_i,Y',\pi(f')(X_i,X'))\Big]\right)^2\\&\le\frac{L^2}{n}\sum_{i=1}^n\left(\mathbb{E}\Big|\pi(f)(X_i,X')-\pi(f')(X_i,X')\Big|\right)^2\end{aligned}$$

$$\leq \frac{L^2}{n}\sum_{i=1}^{n}\mathbb{E}\left|\pi(f)(X_i,X')-\pi(f')(X_i,X')\right|^2$$
$$\leq L^2\|f-f'\|^2_{\mathcal{L}_2(P_{\mathbf{x}}^n\otimes P_{\mathcal{X}})}.$$

因此，$(\mathcal{F}_r,\|\cdot\|_{\mathcal{L}_2(P_{\mathbf{x}}^n\otimes P_{\mathcal{X}})})$ 的一个 ε-网会诱导出 $(\mathcal{S}_r,\|\cdot\|_{\mathcal{L}_2(P^n{}_{\mathbf{z}})})$ 的一个 $L\varepsilon$-网。此外，注意到 $\mathcal{F}_r\subset(r/\lambda)^{1/2}\mathcal{B}_{\mathcal{H}_K}$。结合假设2，我们得到

$$\begin{aligned}\mathbb{E}e_i(\mathrm{id}:\mathcal{S}_r\to\mathcal{L}_2(P_{\mathbf{z}}^n)) &\leq L\mathbb{E}e_i(\mathrm{id}:\mathcal{F}_r\to\mathcal{L}_2(P_{\mathbf{x}}^n\otimes P_{\mathcal{X}}))\\ &\leq 2L(r/\lambda)^{1/2}\mathbb{E}e_i(\mathrm{id}:\mathcal{H}_K\to\mathcal{L}_2(P_{\mathbf{x}}^n\otimes P_{\mathcal{X}}))\\ &\leq 2L(r/\lambda)^{1/2}a_1 i^{-\frac{1}{2p_1}}.\end{aligned}$$

现在我们应用 [36] 中的定理 7.16，并推导出所需的界。证明至此完成。

□

借助前述引理，我们现在可以为经验项 $\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}}))$ 建立一个上界。

命题2。如果假设1和假设2成立，则存在常数

$$c_1:=\frac{\left(2^{p_1}\cdot 270k_1\right)^{\frac{2}{2-\tau-p_1+p_1\tau}}}{3},\quad c_2:=\frac{2^{1+3p_1}\left(135k_2\right)^{1+p_1}}{3}$$

使得对于所有 $t>0$、$n\geq 72t$ 和 $f_0\in\mathcal{H}_K$，以至少 $1-\exp(-t)$ 的概率，

$$\begin{aligned}\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}}))\leq{}&\frac{1}{3}\left(\lambda\|f_{\mathbf{z}}\|_K^2+\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}}))-\mathcal{R}^{\phi}(f_{\phi}^*)\right)+\frac{1}{3}\left(\lambda\|f_0\|_K^2+\mathcal{R}^{\phi}(f_0)-\mathcal{R}^{\phi}(f_{\phi}^*)\right)\\&+c_1\left(\frac{a_1^{2p_1}V^{1-p_1}L^{2p_1}}{\lambda^{p_1}n}\right)^{\frac{1}{2-\tau-p_1+p_1\tau}}+\frac{c_2a_1^{2p_1}B^{1-p_1}L^{2p_1}}{\lambda^{p_1}n}+\left(\frac{24Vt}{n}\right)^{\frac{1}{2-\tau}}.\end{aligned}$$

证明。根据引理 5，我们有

$$\begin{aligned}\mathbb{E}\,\mathrm{Rad}(\mathcal{S}_r,\mathbf{z})\leq\max\Bigg\{&2^{p_1}k_1a_1^{p_1}\left(\frac{r}{\lambda}\right)^{\frac{p_1}{2}}(Vr^{\tau})^{\frac{1-p_1}{2}}L^{p_1}n^{-\frac{1}{2}},\\&4^{\frac{p_1}{1+p_1}}k_2a_1^{\frac{2p_1}{1+p_1}}\left(\frac{r}{\lambda}\right)^{\frac{p_1}{1+p_1}}B^{\frac{1-p_1}{1+p_1}}L^{\frac{2p_1}{1+p_1}}n^{-\frac{1}{1+p_1}}\Bigg\}.\end{aligned}$$

定义

$$\begin{aligned}\varphi(r):=\max\Bigg\{&2^{p_1+1}k_1a_1^{p_1}\left(\frac{r}{\lambda}\right)^{\frac{p_1}{2}}(Vr^{\tau})^{\frac{1-p_1}{2}}L^{p_1}n^{-\frac{1}{2}},\\&2^{\frac{1+3p_1}{1+p_1}}k_2a_1^{\frac{2p_1}{1+p_1}}\left(\frac{r}{\lambda}\right)^{\frac{p_1}{1+p_1}}B^{\frac{1-p_1}{1+p_1}}L^{\frac{2p_1}{1+p_1}}n^{-\frac{1}{1+p_1}}\Bigg\}.\end{aligned}$$

可以验证，$\varphi(4r)\leq 2\varphi(r)$。由引理 2、引理 3 和引理 4 可知，对所有 $t>0$，我们有

$$P\left(\sup_{f\in\mathcal{H}_K}|Q_{\mathbf{z}}(g_{f,r})|\leq\frac{5\varphi(r)}{r}+\sqrt{\frac{2Vt}{r^{2-\tau}n}}+\frac{28Bt}{3nr}\right)\geq 1-\exp(-t).$$

对于 $f_{\mathbf{z}} \in \mathcal{H}_K$，根据 $g_{f_{\mathbf{z}},\, r}$ 的定义，以至少 $1-\exp(-t)$ 的概率，

$$\begin{aligned}
&Q_{\mathbf{z}}[\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi}(f_{\phi}^{*}) - Q\phi_{\pi(f_{\mathbf{z}})} + Q\phi_{f_{\phi}^{*}}] \\
&\leq (\lambda\|f_{\mathbf{z}}\|_K^2 + \mathcal{R}^{\phi}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi}(f_{\phi}^{*}) + r)\left(\frac{5\varphi(r)}{r} + \sqrt{\frac{2Vt}{r^{2-\tau}n}} + \frac{28Bt}{3nr}\right).
\end{aligned}$$

如果存在一个

$$r \geq \max\left\{135\varphi(r), \left(\frac{72Vt}{n}\right)^{\frac{1}{2-\tau}}, r^{*}\right\},$$

然后，利用简单的代数运算和假设 $n \geq 72t$，我们有

$$\frac{5\varphi(r)}{r} \leq \frac{1}{27}, \quad \sqrt{\frac{2Vt}{r^{2-\tau}n}} \leq \frac{1}{6}, \quad \frac{28Bt}{3nr} \leq \frac{28V^{\frac{1}{2-\tau}}t}{3nr} \leq \frac{7}{54}.$$

因此，以至少 $1-\exp(-t)$ 的概率，

$$\begin{aligned}
&Q_{\mathbf{z}}[\mathcal{R}^{\phi}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi}(f_{\phi}^{*}) - Q\phi_{\pi(f_{\mathbf{z}})} + Q\phi_{f_{\phi}^{*}}] \\
&\leq \frac{1}{3}(\lambda\|f_{\mathbf{z}}\|_K^2 + \mathcal{R}^{\phi}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi}(f_{\phi}^{*}) + r).
\end{aligned} \tag{4.7}$$

还需要找到这样的 r。根据 $\varphi(r)$ 的定义，如果

$$\begin{aligned}
r \geq &\left(\frac{4^{p_1}\left(270k_1\right)^2 a_1^{2p_1} V^{1-p_1} L^{2p_1}}{\lambda^{p_1} n}\right)^{\frac{1}{2-\tau-p_1+p_1\tau}} \\
&+ \frac{2^{1+3p_1}\left(135k_2\right)^{1+p_1} a_1^{2p_1} B^{1-p_1} L^{2p_1}}{\lambda^{p_1} n},
\end{aligned}$$

我们有 $r \geq 135\varphi(r)$。此外，注意到 $r^{*} \leq \lambda\|f_0\|_K^2 + \mathcal{R}^{\phi}(f_0) - \mathcal{R}^{\phi}(f_{\phi}^{*})$，只需选择

$$\begin{aligned}
r = &\lambda\|f_0\|_K^2 + \mathcal{R}^{\phi}(f_0) - \mathcal{R}^{\phi}(f_{\phi}^{*}) + \left(\frac{4^{p_1}\left(270k_1\right)^2 a_1^{2p_1} V^{1-p_1} L^{2p_1}}{\lambda^{p_1} n}\right)^{\frac{1}{2-\tau-p_1+p_1\tau}} \\
&+ \frac{2^{1+3p_1}\left(135k_2\right)^{1+p_1} a_1^{2p_1} B^{1-p_1} L^{2p_1}}{\lambda^{p_1} n} + \left(\frac{72Vt}{n}\right)^{\frac{1}{2-\tau}}.
\end{aligned}$$

将此 r 代入（4.7）并选择

$$c_1 = \frac{\left(2^{p_1}\cdot 270k_1\right)^{\frac{2}{2-\tau-p_1+p_1\tau}}}{3}, \quad c_2 = \frac{2^{1+3p_1}\left(135k_2\right)^{1+p_1}}{3}.$$

命题2的证明至此完成。 □

这里我们指出，命题2中的条件 $n \geq 72t$ 并非必要，我们将把 $n < 72t$ 的情形留到下一小节中一般预言机不等式的最终证明。

4.3 对退化项的界定

在本小节中，我们使用对称化和 Dudley 链式法的论证来界定退化项

$$\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}})) = U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})} - h_{f_\phi^*}).$$

如果假设1成立，由于 $\mathcal{R}_{\mathbf{z}}^{\phi}(f_{\mathbf{z}}) + \lambda\|f_{\mathbf{z}}\|_K^2 \le \mathcal{R}_{\mathbf{z}}^{\phi}(0) \le B$，我们有 $\|f_{\mathbf{z}}\|_K \le \sqrt{B/\lambda}$。回顾拉德马赫序列 $\{\varepsilon_i\}_{i=1}^n$，定义

$$\mathcal{F} := \{f \in \mathcal{H}_K : \|f\|_K \le \sqrt{B/\lambda}\},$$
$$S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*}) := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \ne i} \varepsilon_i \varepsilon_j \big(h_{\pi(f)}(Z_i, Z_j) - h_{f_\phi^*}(Z_i, Z_j)\big).$$

那么 $f_{\mathbf{z}} \in \mathcal{F}$。

引理6. 存在常数 $C_1 > 0$、$C_2 > 0$ 使得，对于所有 $\xi > 0$

$$\mathbb{E} \exp\left(\xi \sqrt{\sup_{f\in\mathcal{F}} |(n-1)U_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|}\right)$$
$$\le C_1 \mathbb{E}\mathbb{E}_\varepsilon \exp\left(C_2 \xi \sqrt{\sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|}\right)$$

证明。定义 $\varphi_1(t) := \exp(\xi\sqrt{t})$，并且

$$\varphi_2(t) := \begin{cases} e & 0 \le t \le \xi^{-2} \\ \exp(\xi\sqrt{t}) & t \ge \xi^{-2}. \end{cases}$$

然后 φ_2 是凸的、非递减的，并且以下不等式成立：

$$\varphi_2(t)e^{-1} \le \varphi_1(t) \le \varphi_2(t). \tag{4.8}$$

我们可以在具有 φ_2 的 U-过程理论中应用对称化论证，参见 [24]，的备注 3.5.4，以证明存在常数 $C_1' > 0$ 和 $C_2' > 0$，使得

$$\mathbb{E}\varphi_2(\sup_{f\in\mathcal{F}} |(n-1)U_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|) \le C_1' \mathbb{E}\mathbb{E}_\varepsilon \varphi_2\big(C_2' \sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|\big).$$

结合不等式 (4.8)，即可完成证明。 □

引理7。存在常数 $C_3 > 0$、$C_4 > 0$，使得对于所有 $\xi > 0$,

$$C_1 \mathbb{E}\mathbb{E}_\varepsilon \exp\left(C_2 \xi \sqrt{\sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|}\right)$$
$$\le C_3 \mathbb{E} \exp\left(C_4 \xi^2 \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|\right).$$

证明。定义一个二阶拉ademacher混沌 $R_1 := C_2^2\xi^2(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})$，并且

$$\|R_1\| := \sup_{f\in\mathcal{F}} C_2^2\xi^2(n-1)|S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|.$$

根据 [3], 的公式 (3.5)，存在常数 $C_3' > 0$、$C_4' > 0$，使得

$$\mathbb{E}_\varepsilon \exp(\|R_1\|^{1/2}) \le C_3' \exp(C_4'(\mathbb{E}_\varepsilon\|R_1\|^2)^{1/2}). \tag{4.9}$$

此外，根据 H¨older 不等式和 [3] 的公式 (3.4)，我们有

$$\begin{aligned}(\mathbb{E}_\varepsilon\|R_1\|^2)^{1/2} &= (\mathbb{E}_\varepsilon\|R_1\|^{1/2}\|R_1\|^{3/2})^{1/2}\\ &\le (\mathbb{E}_\varepsilon\|R_1\|)^{1/4}(\mathbb{E}_\varepsilon\|R_1\|^3)^{1/4}\\ &\le 2^{3/4}(\mathbb{E}_\varepsilon\|R_1\|)^{1/4}(\mathbb{E}_\varepsilon\|R_1\|^2)^{3/8},\end{aligned}$$

这等价于

$$(\mathbb{E}_\varepsilon\|R_1\|^2)^{1/2} \le 8\mathbb{E}_\varepsilon\|R_1\|.$$

结合（4.9），我们完成了证明。 □

Next we bound $\mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|$ by the chaining argument.

引理8. 如果假设1和假设2成立，则存在一个常数 $C_5 > 0$ 使得

$$\begin{aligned}&\mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|\\ &\le (\log 4)C_5 B\left[\left(\frac{3a_1^2L^2}{2\lambda B}\right)^{p_1}\frac{4^{1-2p_1}}{1-2p_1} + \left(\frac{3a_2^2L^2}{4\lambda B}\right)^{p_2}\frac{4^{1-2p_2}}{1-2p_2}\right] + \frac{(1+\log 4\sqrt{3})C_5B}{4}.\end{aligned}$$

证明。定义

$$\mathcal{G} := \{h_{\pi(f)} - h_{f_\phi^*} : f \in \mathcal{F}\} \cup \{0 : \mathcal{Z}\times\mathcal{Z} \to \mathbb{R}\}.$$

由假设1中的 (2.5)，$\|h\|_\infty \le 4B$ 对所有 $h \in \mathcal{G}$ 均成立。现在给定 $\mathbf{z} = \{(X_i, Y_i)\}_{i=1}^n$，定义随机过程 $J_{\mathbf{z}}$ 以及伪度量 ρ，其定义在 $\mathcal{G}$ 上，如下：

$$\begin{aligned}J_{\mathbf{z}}(h) &:= \frac{1}{4Bn}\sum_{i=1}^n\sum_{j\ne i}\varepsilon_i\varepsilon_j h(Z_i, Z_j), \quad h \in \mathcal{G},\\ \rho(h,h') &:= \frac{1}{4B}\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\ne i}\big(h(Z_i,Z_j) - h'(Z_i,Z_j)\big)^2}, \quad h, h' \in \mathcal{G}.\end{aligned}$$

我们将对 $\{J_{\mathbf{z}}(h)$ 应用适用于 U-过程的链式引理：$h \in \mathcal{G}\}$，参见 [28] 的引理 5。为此，我们需要找到一个凸的、严格递增的函数 $\varphi : [0,\infty) \to [0,\infty)$，使得 $0 \le \varphi(0) \le 1$，并且对所有 $\rho(h,h') > 0$，

$$\mathbb{E}_\varepsilon\varphi\left(\frac{|J_{\mathbf{z}}(h) - J_{\mathbf{z}}(h')|}{\rho(h,h')}\right) \le 1.$$

定义二阶拉ademacher混沌 $R_2 := J_{\mathbf{z}}(h) - J_{\mathbf{z}}(h')$。根据 [24], 的推论 3.2.6，存在一个与 h、h'、n 无关的常数 $\kappa > 0$，使得

$$\mathbb{E}_\varepsilon \exp\left(\frac{|R_2|}{\kappa(\mathbb{E}_\varepsilon |R_2|^2)^{1/2}}\right) \le e. \tag{4.10}$$

请注意,

$$\begin{aligned}
\mathbb{E}_\varepsilon |R_2|^2 &= \frac{1}{16B^2n^2}\mathbb{E}_\varepsilon\bigg(\sum_{i\neq j}\varepsilon_i\varepsilon_j\big(h(Z_i,Z_j)-h'(Z_i,Z_j)\big)\bigg)^2 \\
&= \frac{1}{16B^2n^2}\bigg(\sum_{i=k\neq j=l}\big(h(Z_i,Z_j)-h'(Z_i,Z_j)\big)\big(h(Z_k,Z_l)-h'(Z_k,Z_l)\big) \\
&\quad + \sum_{i=l\neq j=k}\big(h(Z_i,Z_j)-h'(Z_i,Z_j)\big)\big(h(Z_k,Z_l)-h'(Z_k,Z_l)\big)\bigg) \\
&= \frac{1}{8B^2n^2}\sum_{i\neq j}\big(h(Z_i,Z_j)-h'(Z_i,Z_j)\big)^2 \\
&\le 2\rho(h,h')^2.
\end{aligned} \tag{4.11}$$

因此，如果我们选择 $\varphi(t) = \exp(\frac{t}{\sqrt{2}\kappa} - 1)$，根据 (4.10) 和 (4.11)，我们有

$$\begin{aligned}
\mathbb{E}_\varepsilon\varphi\left(\frac{|J_{\mathbf{z}}(h)-J_{\mathbf{z}}(h')|}{\rho(h,h')}\right) &\le \mathbb{E}_\varepsilon\varphi\left(\frac{\sqrt{2}|R_2|}{(\mathbb{E}_\varepsilon|R_2|^2)^{1/2}}\right) \\
&= \mathbb{E}_\varepsilon\exp\left(\frac{|R_2|}{\kappa(\mathbb{E}_\varepsilon|R_2|^2)^{1/2}} - 1\right) \le 1.
\end{aligned}$$

固定 $h' = 0 \in \mathcal{G}$，显然 $\sup_{h\in\mathcal{G}} \rho(h, 0) \le 1$，并且

$$4B\mathbb{E}_\varepsilon \sup_{h\in\mathcal{G}} |J_{\mathbf{z}}(h) - J_{\mathbf{z}}(0)| = \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})|.$$

因此，根据 [28], 的引理 5，我们得到

$$\begin{aligned}
\mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} |(n-1)S_{\mathbf{z}}(h_{\pi(f)} - h_{f_\phi^*})| &\le 32B\int_0^{\frac{1}{4}} \varphi^{-1}(\mathcal{N}(\mathcal{G},\rho,\delta))\,d\delta \\
&\le C_5 B\int_0^{\frac{1}{4}} \log\mathcal{N}(\mathcal{G},\rho,\delta)\,d\delta,
\end{aligned} \tag{4.12}$$

其中 $C_5 = 64\sqrt{2}\kappa$。接下来需要界定度量熵 $\log\,\mathcal{N}(\mathcal{G},\rho,\delta)$。我们记为

$$\begin{aligned}
&16B^2\rho^2(h,h') \\
&= \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}\big(h(Z_i,Z_j)-h'(Z_i,Z_j)\big)^2 \\
&\le \frac{4}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}\Big[\big(\phi_{\pi(f)}(Z_i,Z_j)-\phi_{\pi(f')}(Z_i,Z_j)\big)^2 \\
&\quad + \big(Q\phi_{\pi(f)}(Z_i) - Q\phi_{\pi(f')}(Z_i)\big)^2 + \big(Q\phi_{\pi(f)}(Z_j) - Q\phi_{\pi(f')}(Z_j)\big)^2
\end{aligned}$$

$$
\begin{aligned}
&\quad + \big(\mathcal{R}^{\phi}(\pi(f)) - \mathcal{R}^{\phi}(\pi(f'))\big)^2\Big] \\
&\le \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \Big[L^2 \big(\pi(f)(X_i, X_j) - \pi(f')(X_i, X_j)\big)^2 \\
&\quad + L^2 \mathbb{E}\big(\pi(f)(X_i, X') - \pi(f')(X_i, X')\big)^2 + L^2 \mathbb{E}\big(\pi(f)(X_j, X') - \pi(f')(X_j, X')\big)^2 \\
&\quad + \big(\mathcal{R}^{\phi}(\pi(f)) - \mathcal{R}^{\phi}(\pi(f'))\big)^2\Big] \\
&\le 4L^2 \|f - f'\|^2_{\mathcal{L}_2(P^n_{\mathbf{x}^2})} + 8L^2 \|f - f'\|^2_{\mathcal{L}_2(P^n_{\mathbf{x}} \otimes P_{\mathcal{X}})} + 4\big(\mathcal{R}^{\phi}(\pi(f)) - \mathcal{R}^{\phi}(\pi(f'))\big)^2,
\end{aligned}
$$

其中，第二个不等式源于以下假设： ϕ 关于第三个参数是 L-利普希茨的。这意味着

$$
\begin{aligned}
\mathcal{N}(\mathcal{G}, \rho, \delta) &\le \mathcal{N}\left(\mathcal{F}, \|\cdot\|_{\mathcal{L}_2(P^n_{\mathbf{x}^2})}, \frac{2B\delta}{\sqrt{3}L}\right) \cdot \mathcal{N}\left(\mathcal{F}, \|\cdot\|_{\mathcal{L}_2(P^n_{\mathbf{x}} \otimes P_{\mathcal{X}})}, \frac{\sqrt{2}B\delta}{\sqrt{3}L}\right) \cdot \mathcal{N}\left([0, B], |\cdot|, \frac{2B\delta}{\sqrt{3}}\right) \\
&= \mathcal{N}\left(\mathrm{id} : \mathcal{H}_K \to \mathcal{L}_2(P^n_{\mathbf{x}^2}), 2\sqrt{\frac{\lambda B}{3}}\frac{\delta}{L}\right) \cdot \mathcal{N}\left(\mathrm{id} : \mathcal{H}_K \to \mathcal{L}_2(P^n_{\mathbf{x}} \otimes P_{\mathcal{X}}), \sqrt{\frac{2\lambda B}{3}}\frac{\delta}{L}\right) \\
&\quad \cdot \mathcal{N}\left([0, B], |\cdot|, \frac{2B\delta}{\sqrt{3}}\right).
\end{aligned}
$$

根据假设2，我们有

$$
\log \mathcal{N}(\mathcal{G}, \rho, \delta) \le \log 4 \cdot \left(\frac{3a_1^2 L^2}{2\lambda B \delta^2}\right)^{p_1} + \log 4 \cdot \left(\frac{3a_2^2 L^2}{4\lambda B \delta^2}\right)^{p_2} + \log \frac{\sqrt{3}}{\delta}.
$$

由于 $p_1, p_2 \in (0, 1/2)$，对右侧积分，并由 (4.12) 得到所需的界。证明至此完成。

□

借助前面的引理，我们现在可以建立退化项 $\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}}))$ 的上界。

命题3。如果假设1和假设2成立，则存在常数

$$
c_0 := C_3, \quad c_3 := \frac{(64 \log 2) C_4 C_5 3^{p_1}}{32^{p_1}(1 - 2p_1)}, \quad c_4 := \frac{(64 \log 2) C_4 C_5 3^{p_2}}{64^{p_2}(1 - 2p_2)}, \quad c_5 := (2 + \log 48) C_4 C_5,
$$

使得对于所有 $t > 0$ 和 $n \ge 2$，以至少 $1 - c_0 \exp(-t)$ 的概率，我们有

$$
|\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}}))| \le \frac{c_3 a_1^{2p_1} L^{2p_1} B^{1-p_1} t}{\lambda^{p_1} n} + \frac{c_4 a_2^{2p_2} L^{2p_2} B^{1-p_2} t}{\lambda^{p_2} n} + \frac{c_5 t}{n}.
$$

证明。回忆 $\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}})) = U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})} - h_{f^*_\phi})$。记

$$
\omega := (\log 4) C_5 B \left[\left(\frac{3a_1^2 L^2}{2\lambda B}\right)^{p_1} \frac{4^{1-2p_1}}{1 - 2p_1} + \left(\frac{3a_2^2 L^2}{4\lambda B}\right)^{p_2} \frac{4^{1-2p_2}}{1 - 2p_2}\right] + \frac{(1 + \log 4\sqrt{3}) C_5 B}{4}.
$$

结合引理6、引理7和引理8，我们得到

$$
\mathbb{E} \exp\left(\xi \sqrt{|(n-1) U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})} - h_{f^*_\phi})|}\right)
$$

$$\leq C_1\mathbb{E}\mathbb{E}_\varepsilon \exp\left(C_2\xi\sqrt{\sup_{f\in\mathcal{F}}|(n-1)S_{\mathbf{z}}(h_{\pi(f)}-h_{f_\phi^*})|}\right)$$
$$\leq C_3\mathbb{E}\exp\left(C_4\xi^2\mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}}|(n-1)S_{\mathbf{z}}(h_{\pi(f)}-h_{f_\phi^*})|\right) \leq C_3\exp\left(C_4\xi^2\omega\right).$$

根据马尔可夫不等式，我们得出，对于所有 $t'>0$ 和 $\xi>0$，有

$$P\left(|(n-1)U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})}-h_{f_\phi^*})|\geq t'\right)\leq C_3\exp(C_4\xi^2\omega-\xi\sqrt{t'}).$$

令 $t'=4C_4\omega t$ 和 $\xi=\frac{\sqrt{t'}}{2C_4\omega}$，我们有

$$\begin{aligned}&P\left(|U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})}-h_{f_\phi^*})|\geq\frac{8C_4\omega t}{n}\right)\\&\leq P\left(|(n-1)U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})}-h_{f_\phi^*})|\geq 4C_4\omega t\right)\\&\leq C_3\exp(-t).\end{aligned}$$

这就完成了命题3的证明。 □

现在我们可以证明定理1，它给出了超额 ϕ-排序风险的一般预言机不等式，以及定理 2 中针对基于间隔的损失的变体。

定理 1 的证明。根据误差分解 (4.1) 和 Hoeffding 分解 (4.6)，我们有

$$\begin{aligned}&\lambda\|f_{\mathbf{z}}\|_K^2+\mathcal{R}^\phi(\pi(f_{\mathbf{z}}))-\mathcal{R}^\phi(f_\phi^*)\\&\leq\left(\lambda\|f_0\|_K^2+\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f_\phi^*)\right)+\left(\mathcal{R}_{\mathbf{z}}^\phi(f_0)-\mathcal{R}_{\mathbf{z}}^\phi(f_\phi^*)-\mathcal{R}^\phi(f_0)+\mathcal{R}^\phi(f_\phi^*)\right)\\&\quad+2Q_{\mathbf{z}}[\mathcal{R}^\phi(\pi(f_{\mathbf{z}}))-\mathcal{R}^\phi(f_\phi^*)-Q\phi_{\pi(f_{\mathbf{z}})}+Q\phi_{f_\phi^*}]-U_{\mathbf{z}}(h_{\pi(f_{\mathbf{z}})}-h_{f_\phi^*})\\&=\mathcal{A}(\lambda,f_0)+\mathcal{S}_{\mathbf{z}}(f_0)+2\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}}))-\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}})).\end{aligned}$$

若 $n<72t$，根据假设1的(2.5)和 $\|\phi_{f_0}\|_\infty\leq B_0$，我们有

$$\begin{aligned}&\lambda\|f_{\mathbf{z}}\|_K^2+\mathcal{R}^\phi(\pi(f_{\mathbf{z}}))-\mathcal{R}^\phi(f_\phi^*)\\&\leq\left(\lambda\|f_0\|_K^2+\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f_\phi^*)\right)+(B_0+B)+(2B+2B)+4B\\&=\left(\lambda\|f_0\|_K^2+\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f_\phi^*)\right)+9B+B_0\\&\leq 8\left(\lambda\|f_0\|_K^2+\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f_\phi^*)\right)+\frac{900Bt}{n}+\frac{456B_0t}{n}.\end{aligned}$$

因此结论立即得出。

[{"id": 0, "output": "如果 $n\geq 72t$，根据命题1，以至多 4 exp(−t) 的概率，有以下结论成立"}]

$$\mathcal{S}_{\mathbf{z}}(f_0)\geq\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f_\phi^*)+\left(\frac{8Vt}{n}\right)^{\frac{1}{2-\tau}}+\frac{300Bt}{n}+\frac{152B_0t}{n}.$$

根据命题2，以至多 exp(−t) 的概率，有

$$\mathcal{S}_{\mathbf{z},1}(\pi(f_{\mathbf{z}}))\geq\frac{1}{3}\left(\lambda\|f_{\mathbf{z}}\|_K^2+\mathcal{R}^\phi(\pi(f_{\mathbf{z}}))-\mathcal{R}^\phi(f_\phi^*)\right)+\frac{1}{3}\left(\lambda\|f_0\|_K^2+\mathcal{R}^\phi(f_0)-\mathcal{R}^\phi(f_\phi^*)\right)$$

$$+ c_1 \left(\frac{a_1^{2p_1} V^{1-p_1} L^{2p_1}}{\lambda^{p_1} n} \right)^{\frac{1}{2-\tau-p_1+p_1\tau}} + \frac{c_2 a_1^{2p_1} B^{1-p_1} L^{2p_1}}{\lambda^{p_1} n} + \left(\frac{24Vt}{n} \right)^{\frac{1}{2-\tau}}.$$

根据命题3，以至多 $c_0\exp(-t)$ 的概率，有

$$|\mathcal{S}_{\mathbf{z},2}(\pi(f_{\mathbf{z}}))| \geq \frac{c_3 a_1^{2p_1} L^{2p_1} B^{1-p_1} t}{\lambda^{p_1} n} + \frac{c_4 a_2^{2p_2} L^{2p_2} B^{1-p_2} t}{\lambda^{p_2} n} + \frac{c_5 t}{n}$$

最后，综合上述所有估计，以至少 $1 - (c_0 + 5)\exp(-t)$ 的概率，我们有

$$\begin{aligned}
&\lambda \|f_{\mathbf{z}}\|_K^2 + \mathcal{R}^\phi(\pi(f_{\mathbf{z}})) - \mathcal{R}^\phi(f_\phi^*) \\
&\leq \left(\lambda \|f_0\|_K^2 + \mathcal{R}^\phi(f_0) - \mathcal{R}^\phi(f_\phi^*)\right) + \mathcal{R}^\phi(f_0) - \mathcal{R}^\phi(f_\phi^*) + \left(\frac{8Vt}{n}\right)^{\frac{1}{2-\tau}} + \frac{300Bt}{n} + \frac{152B_0 t}{n} \\
&\quad + \frac{2}{3}\left(\lambda \|f_{\mathbf{z}}\|_K^2 + \mathcal{R}^\phi(\pi(f_{\mathbf{z}})) - \mathcal{R}^\phi(f_\phi^*)\right) + \frac{2}{3}\left(\lambda \|f_0\|_K^2 + \mathcal{R}^\phi(f_0) - \mathcal{R}^\phi(f_\phi^*)\right) \\
&\quad + 2c_1 \left(\frac{a_1^{2p_1} V^{1-p_1} L^{2p_1}}{\lambda^{p_1} n} \right)^{\frac{1}{2-\tau-p_1+p_1\tau}} + \frac{2c_2 a_1^{2p_1} B^{1-p_1} L^{2p_1}}{\lambda^{p_1} n} + \left(\frac{96Vt}{n} \right)^{\frac{1}{2-\tau}} \\
&\quad + \frac{c_3 a_1^{2p_1} L^{2p_1} B^{1-p_1} t}{\lambda^{p_1} n} + \frac{c_4 a_2^{2p_2} L^{2p_2} B^{1-p_2} t}{\lambda^{p_2} n} + \frac{c_5 t}{n}.
\end{aligned}$$

证明随后完成。 □

定理 2 的证明。利用关于 ψ 的条件以及第5.3小节中进行的分析，我们看到假设1得到满足，因此应用定理 1 得出结论。

□

5 推导高斯排序估计器的学习率

本节中，我们将应用预言机不等式，推导采用合页损失和平方损失的高斯排序估计器 (1.4) 的快速学习速率。首先，在数据生成分布在输入空间 $\mathcal{X} \subset \mathbb{R}^d$ 上的边缘分布支撑于一个上盒计数维数为 $\varrho \in (0, d]$ 的集合这一假设下，我们估计成对高斯核空间 $\mathcal{H}_{K^\sigma}$ 的容量。然后，我们推导噪声条件下合页损失的逼近误差界，以及 Besov 光滑性下平方损失的逼近误差界。最后，我们应用成熟的预言机不等式和校准不等式，推导超额排序风险的快速学习速率。本节末尾，我们还比较了不同的噪声条件。

5.1 成对高斯核空间的熵数估计

在本小节中，我们写作 $\mathcal{H}_K(\mathcal{X}^2)$，以强调由成对再生核 K 诱导的再生核希尔伯特空间由定义在 $\mathcal{X}^2 = \mathcal{X} \times \mathcal{X}$ 上的函数组成，或者等价地， K 是

定义于或限制在 $\mathcal{X}^2 \times \mathcal{X}^2$ 上，其中 $\mathcal{X}$ 将在上下文中指定。方差为 $\sigma > 0$ 的 $\mathbb{R}^{2d} \times \mathbb{R}^{2d}$ 上的传统高斯核定义为

$$\widetilde{K}^{\sigma}((x,x'),(u,u')) := \exp\left(-\frac{\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) \tag{5.1}$$

并且我们将由 $\mathcal{X}^2$ 限制所诱导的再生核希尔伯特空间记为 $\mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)$，参见 [4]。注意，我们要求排序估计量是反对称的。直接选择 $\mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)$ 作为假设空间并不合适。然而，我们可以将 $\widetilde{f} \in \mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)$分解为对称部分$\widetilde{}(f-f)$ 和反对称部分 f:

$$\widetilde{f}(x,x') = \frac{\widetilde{f}(x,x')+\widetilde{f}(x',x)}{2} + \frac{\widetilde{f}(x,x')-\widetilde{f}(x',x)}{2} =: (\widetilde{f}-f)+f \tag{5.2}$$

从而得到 $\mathcal{L}_2(P_{\mathcal{X}}^2)$ 中 $\mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)$ 的正交分解，并且还得到传统高斯核的一个分解，该分解由

$$\begin{aligned}\widetilde{K}^{\sigma}((x,x')(u,u')) = &\frac{\widetilde{K}^{\sigma}((x,x'),(u,u'))+\widetilde{K}^{\sigma}((x',x),(u,u'))}{2}\\ &+\frac{\widetilde{K}^{\sigma}((x,x'),(u,u'))-\widetilde{K}^{\sigma}((x',x),(u,u'))}{2}.\end{aligned}$$

$\widetilde{}\,K^{\sigma}((x,x')(u,u'))$ 的反对称部分给出了本文所考虑的形式为 (1.3) 的成对高斯核 $K^{\sigma}((x,x'),(u,u'))$。根据再生核的性质，参见 [4], $\mathcal{H}_{K^{\sigma}}(\mathcal{X}^2)$是 $\mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)$ 的子空间，并且对于所有 $f \in \mathcal{H}_{K^{\sigma}}(\mathcal{X}^2)$ 都有 $\|f\|_{K^{\sigma}} = \|f\|_{\widetilde{K}^{\sigma}}$。因此，我们可以利用针对 $\mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)$ 所建立的现有估计，来进行 $\mathcal{H}_{K^{\sigma}}(\mathcal{X}^2)$ 的熵数估计。具体而言， [23] 表明

$$\log \mathcal{N}(\mathrm{id}: \mathcal{H}_{\widetilde{K}^{\sigma}}([0,1]^{2d}) \to C([0,1]^{2d}), \varepsilon) \asymp \frac{\left(\log \frac{1}{\varepsilon}\right)^{2d+1}}{\left(\log\log \frac{1}{\varepsilon}\right)^{2d}}, \quad \varepsilon \to 0.$$

[37] 中的一项改进分析阐明了常数对 σ 和底层空间 $\mathcal{X} \subset \mathbb{R}^d$的关键依赖关系:

$$\begin{aligned}&\log \mathcal{N}(\mathrm{id}: \mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2) \to C(\mathcal{X}^2), \varepsilon)\\ &\le \mathcal{N}(\mathcal{X}^2, \|\cdot\|_{\infty}, \sigma) \cdot \binom{4e+2d}{2d} e^{-2d} \frac{\left(\log \frac{1}{\varepsilon}\right)^{2d+1}}{\left(\log\log \frac{1}{\varepsilon}\right)^{2d}}.\end{aligned} \tag{5.3}$$

基于上述结果，在 $\mathcal{X}$ 的内在维数假设下，我们得到以下命题。

命题 4. 若假设6成立，则 $\mathcal{H}_{K^{\sigma}}(\mathcal{X}^2)$对所有 $p_1 = p_2 = p \in (0, 1/2)$ 以及 $a_1 = a_2 = a := (C_{\mathcal{X}}^{*} p^{-2d-1} \sigma^{-2\varrho})^{\frac{1}{2p}}$. 满足假设2

证明。首先我们证明

$$\mathcal{N}(\mathrm{id}: \mathcal{H}_{K^{\sigma}}(\mathcal{X}^2) \to C(\mathcal{X}^2), \varepsilon) \le \mathcal{N}(\mathrm{id}: \mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2) \to C(\mathcal{X}^2), \varepsilon).$$

$\widetilde{}$ $\widetilde{}$回顾一下， $\mathcal{B}_{\mathcal{F}}$表示赋范空间 $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ 的单位球。给定 $(\mathcal{B}_{\mathcal{H}_{\widetilde{K}^{\sigma}}(\mathcal{X}^2)}\ \|\cdot\|_{\infty})$ 的一个 ε-网 $\{f_1, \ldots, f_m\}$，根据分解 (5.2)，它诱导出一组反对称函数

$\sim \quad \sim \quad \sim\{f_1,\ldots,f_m\} \subset \mathcal{B}_{\mathcal{H}_{K^\sigma}(\mathcal{X}^2)}$。令 $f \in \mathcal{B}_{\mathcal{H}_{K^\sigma}(\mathcal{X}^2)} \subset \mathcal{B}_{\mathcal{H}_{\widetilde{K}^\sigma}(\mathcal{X}^2)}$。存在一个 $f_i \in \{f_1,\ldots,f_m\}^{\sim}$，使得 $\|f_i - f\|_\infty \le \varepsilon$。注意， f 是反对称的。因此，根据分解 (5.2)，我们有 $\|f_i - f\|_\infty \le \|f_i - f\|_\infty \le \varepsilon$。于是， $\{f_1,\ldots,f_m\}$ 是 $(B_{\mathcal{H}_{K^\sigma}(\mathcal{X}^2)} \ \|\cdot\|_\infty)$ 的一个 ε-网。由 (2.14) 和 (5.3)，对所有 $p \in (0,1/2)$，我们有如下估计。

$$
\begin{aligned}
&\log \mathcal{N}(\mathrm{id} : \mathcal{H}_{K^\sigma}(\mathcal{X}^2) \to C(\mathcal{X}^2), \varepsilon) \\
\le\ &\log \mathcal{N}(\mathrm{id} : \mathcal{H}_{\widetilde{K}^\sigma}(\mathcal{X}^2) \to C(\mathcal{X}^2), \varepsilon) \\
\le\ &\mathcal{N}(\mathcal{X}^2, \|\cdot\|_\infty, \sigma) \cdot \binom{4e+2d}{2d} e^{-2d} \frac{\left(\log \frac{4}{\varepsilon}\right)^{2d+1}}{\left(\log\log \frac{4}{\varepsilon}\right)^{2d}} \\
\le\ &C_{\mathcal{X}}^2 \sigma^{-2\varrho} \binom{4e+2d}{2d} e^{-2d} \frac{\left(\log \frac{4}{\varepsilon}\right)^{2d+1}}{\left(\log\log \frac{4}{\varepsilon}\right)^{2d}} \\
\le\ &C_{\mathcal{X}}^2 \sigma^{-2\varrho} \binom{4e+2d}{2d} e^{-2d} \left(\log \frac{4}{\varepsilon}\right)^{2d+1} \\
\le\ &C_{\mathcal{X}}^2 \sigma^{-2\varrho} \binom{4e+2d}{2d} e^{-2d} \left(\frac{2d+1}{2p}\right)^{2d+1} 4^{2p} e^{-2d-1} \varepsilon^{-2p} \\
\le\ &4C_{\mathcal{X}}^2 \binom{4e+2d}{2d} \frac{(2d+1)^{2d+1}}{2^{2d+1} e^{4d+1}} p^{-2d-1} \sigma^{-2\varrho} \varepsilon^{-2p}.
\end{aligned}
$$

接下来，我们利用上述估计为熵数建立一个界，参见 [36], 习题6.8，其表达式为

$$
e_i(\mathrm{id} : \mathcal{H}_{K^\sigma}(\mathcal{X}^2) \to C(\mathcal{X}^2)) \le (C_{\mathcal{X}}^* p^{-2d-1} \sigma^{-2\varrho})^{\frac{1}{2p}} i^{-\frac{1}{2p}}
$$

其中

$$
C_{\mathcal{X}}^* := 12 C_{\mathcal{X}}^2 \binom{4e+2d}{2d} \frac{(2d+1)^{2d+1}}{2^{2d+1} e^{4d+1}}.
$$

注意， $\mathcal{L}_2$-半范数$\|\cdot\|_{\mathcal{L}_2(P^n_{\mathbf{x}} \otimes P_{\mathcal{X}})}$ 和$\|\cdot\|_{\mathcal{L}_2(P^n{}_{\mathbf{x}^2})}$ 都由上确界范数$\|\cdot\|_\infty$控制。因此，假设2对所有 $p_1 = p_2 = p \in (0,1/2)$和 $a_1 = a_2 = a := (C_{\mathcal{X}}^* p^{-2d-1} \sigma^{-2\varrho})^{\frac{1}{2p}}$. 都成立。证明至此完成。 □

5.2 逼近误差的界

在本小节中，我们旨在构造一个合适的 $f_0 \in \mathcal{H}_{K^\sigma}$ ，并为 $\phi = \phi_{\mathrm{hinge}}$ 和 $\phi = \phi_{\mathrm{square}}$给出逼近误差 $\lambda \|f_0\|_{K^\sigma}^2 + \mathcal{R}^\phi(f_0) - \mathcal{R}^\phi(f_\phi^*)$的界。受[36, 46], 的工作启发，定义

$$
\widetilde{\mathcal{K}}^\sigma(x,x') := \left(\frac{2}{\sigma\sqrt{\pi}}\right)^d \exp\left(-\frac{2\|(x,x')\|_2^2}{\sigma^2}\right)
$$

以及一个算子 $\widetilde{\mathcal{K}}^\sigma * \cdot$ 在 $\mathcal{L}_2(\mathbb{R}^{2d})$ 上，作为

$$
\widetilde{\mathcal{K}}^\sigma * f(x,x') := \int_{\mathbb{R}^{2d}} \widetilde{\mathcal{K}}^\sigma(u,u') f((x,x') + (u,u')) d(u,u'), \forall f \in \mathcal{L}_2(\mathbb{R}^{2d}).
$$

可以验证

$$
\widetilde{\mathcal{K}}^\sigma * f(x,x')
$$

$$= \left(\frac{1}{\sigma\sqrt{\pi}}\right)^d \int_{\mathbb{R}^{2d}} \widetilde{K}^\sigma((x,x'),(u,u')) f\left(\left(1-\frac{1}{\sqrt{2}}\right)(x,x') + \frac{1}{\sqrt{2}}(u,u')\right) d(u,u')$$

其中，K^σ 是由（5.1）给出的传统高斯核。因此，$\mathcal{K}^\sigma * \cdot$ 本质上是一个卷积算子，它是从 $\mathcal{L}_2(\mathbb{R}^{2d})$ 到 $\mathcal{H}_{\widetilde{K}^\sigma}$ 的度量满射，即 $\widetilde{\mathcal{K}}^\sigma * f \in \mathcal{H}_{\widetilde{K}^\sigma}$ 和 $\|\widetilde{\mathcal{K}}^\sigma * f\|_{\widetilde{K}^\sigma} \leq \|f\|_{\mathcal{L}_2(\mathbb{R}^{2d})}$。这促使我们考虑定义在 $\mathcal{L}_2(\mathbb{R}^{2d})$ 上的以下算子：

$$\mathcal{K}^\sigma * f(x,x') := \frac{1}{2}\int_{\mathbb{R}^{2d}} \widetilde{\mathcal{K}}^\sigma(u,u')\Big(f((x,x')+(u,u')) - f((x',x)+(u,u'))\Big) d(u,u').$$

事实上，对于所有 $f \in \mathcal{L}_2(\mathbb{R}^{2d})$，$\mathcal{K}^\sigma * f(x,x')$都是反对称的，即 $\mathcal{K}^\sigma * f(x,x') = -\mathcal{K}^\sigma * f(x',x)$，且如果 f是反对称的，则 $\mathcal{K}^\sigma * f(x,x') = \mathcal{K}^\sigma * f(x,x')$。此外，如果 f是对称的，则有$\mathcal{K}^\sigma * f(x,x') = 0$。对于所有 $\widetilde{f} \in \mathcal{L}_2(\mathbb{R}^{2d})$，$\widetilde{f} = \widetilde{f} - f + f$，根据分解（5.2），我们将其写为（），其中 f 是反对称的，而 $f - f$ 是对称的。于是我们有

$$\|\mathcal{K}^\sigma * \widetilde{f}\|_{K^\sigma} = \|\mathcal{K}^\sigma * f\|_{K^\sigma} = \|\widetilde{\mathcal{K}}^\sigma * f\|_{\widetilde{K}^\sigma} \leq \|f\|_{\mathcal{L}_2(\mathbb{R}^{2d})} \leq \|\widetilde{f}\|_{\mathcal{L}_2(\mathbb{R}^{2d})}.$$

因此，$\mathcal{K}^\sigma * \cdot$ 是从 $\mathcal{L}_2(\mathbb{R}^{2d})$到 $\mathcal{H}_{K^\sigma}$ 的度量满射。注意，我们需要对 $\phi = \phi_{\text{hinge}}$ 和 $\phi = \phi_{\text{square}}$的贝叶斯排序规则 f_ϕ^* 进行近似，该规则是反对称的。接下来，我们将对反对称函数应用卷积算子 $\mathcal{K}^\sigma * \cdot$，以构造 $\mathcal{H}_{K^\sigma}$ 中 f_ϕ^* 的高效近似，然后进行近似分析。

我们首先处理合页损失。在间隔噪声条件假设5下，我们证明以下命题，以界定逼近误差。

命题 5。将 $\mathcal{B}_{\mathbb{R}^{2d}}$ 记为 $\mathbb{R}^{2d}$ 中的单位闭球。假设 $\mathcal{X}^2 \subset r\mathcal{B}_{\mathbb{R}^{2d}}$ 对于 $r > 0$，且假设5 在 $C_{**} > 0$ 和 $\beta > 0$ 下成立。那么对于所有 $\lambda > 0$、$\sigma > 0$，存在一个 $f_0 \in \mathcal{H}_{K^\sigma}$ 使得 $\|f_0\|_\infty \leq 1$ 且

$$\lambda\|f_0\|_{K^\sigma}^2 + \mathcal{R}^{\phi_{\text{hinge}}}(f_0) - \mathcal{R}^{\phi_{\text{hinge}}}(f_{\text{hinge}}^*) \leq \frac{2^{2d+1}r^{2d}}{\Gamma(d)}\lambda\sigma^{-2d} + \frac{2^{\beta/2+1}C_{**}\Gamma(d+\frac{\beta}{2})}{\Gamma(d)}\sigma^\beta.$$

证明。请回忆，合页损失的贝叶斯排序规则 f_{hinge}^*：$\mathcal{X}^2 \to \mathbb{R}$ 具有以下形式

$$f_{\text{hinge}}^*(x,x') = \text{sgn}(\eta_+(x,x') - \eta_-(x,x')).$$

将 $\text{vol}(\cdot)$ 记为欧几里得空间上的勒贝格测度。注意，如果 $\text{vol}(\mathcal{X}) = 0$，例如，当 $\mathcal{X}$ 是一个上盒计数维数为 $\varrho < d$的子流形时，对 f_ϕ^* 进行平凡的零延拓，在 $\mathcal{L}_2(\mathbb{R}^{2d})$ 中只会得到零函数。因此，在我们使用 $\mathcal{K}^\sigma * f_{\text{hinge}}^*$构造逼近器之前，需要以非平凡的方式将 f^*_{hinge}延拓到一个具有正测度的定义域，同时仍然保持间隔噪声条件。

为此，我们将后验概率 $\eta_+, \eta_-, \eta_=$：$\mathcal{X}^2 \to [0,1]$ 扩展到$2r\mathcal{B}_{\mathbb{R}^{2d}}$。对于所有 $(x,x') \in \mathcal{X}^2$，稍微滥用记号，记 $\mathcal{B}_{\mathbb{R}^{2d}}((x,x'), \Delta(x,x')/2)$ 为以 (x,x') 为中心、半径为 $\Delta(x,x')/2$的开球，其中 $\Delta(x,x')$ 定义为 (2.12)。由于$\mathcal{X}^2 \subset r\mathcal{B}_{\mathbb{R}^{2d}}$，我们有 $\Delta(x,x') \leq 2r$ 以及 $\mathcal{B}_{\mathbb{R}^{2d}}((x,x'), \Delta(x,x')/2) \subset 2r\mathcal{B}_{\mathbb{R}^{2d}}$。对于所有 $(u,u') \in 2r\mathcal{B}_{\mathbb{R}^{2d}} \backslash \mathcal{X}^2$，定义扩展为

$$\eta_+(u,u') = 1, \eta_-(u,u') = \eta_=(u,u') = 0,$$

$$\forall (u,u') \in \bigcup_{(x,x')\in\mathcal{X}_+^2} \mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\Delta(x,x')/2),$$
$$\eta_-(u,u') = 1, \eta_+(u,u') = \eta_=(u,u') = 0,$$
$$\forall (u,u') \in \bigcup_{(x,x')\in\mathcal{X}_-^2} \mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\Delta(x,x')/2),$$
$$\eta_=(u,u') = 1, \eta_+(u,u') = \eta_-(u,u') = 0,$$
$$\forall (u,u') \in 2r\mathcal{B}_{\mathbb{R}^{2d}} \setminus \bigcup_{(x,x')\in\mathcal{X}_+^2\cup\mathcal{X}_-^2} \mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\Delta(x,x')/2).$$

由于 $\bigcup_{(x,x')\in\mathcal{X}_+^2} \mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\Delta(x,x')/2)$ 和 $\bigcup_{(x,x')\in\mathcal{X}_-^2} \mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\Delta(x,x')/2)$ 根据 $\Delta(x,x')$ 的定义和三角不等式具有空交集，该扩展定义良好。使用扩展后的 $\eta_+,\eta_-,\eta_=$： $2r\mathcal{B}_{\mathbb{R}^{2d}} \to [0,1]$, 定义

$$(2r\mathcal{B}_{\mathbb{R}^{2d}})_+ := \{(x,x') \in 2r\mathcal{B}_{\mathbb{R}^{2d}} : \eta_+(x,x') > \eta_-(x,x')\},$$
$$(2r\mathcal{B}_{\mathbb{R}^{2d}})_- := \{(x,x') \in 2r\mathcal{B}_{\mathbb{R}^{2d}} : \eta_+(x,x') < \eta_-(x,x')\},$$
$$(2r\mathcal{B}_{\mathbb{R}^{2d}})_= := \{(x,x') \in 2r\mathcal{B}_{\mathbb{R}^{2d}} : \eta_+(x,x') = \eta_-(x,x')\},$$

以及一个新的距离函数 $\widetilde{\Delta}$： $\mathcal{X}^2 \to \mathbb{R}$ 作为

$$\widetilde{\Delta}(x,x') := \begin{cases} \mathrm{dist}((x,x'),(2r\mathcal{B}_{\mathbb{R}^{2d}})_- \cup (2r\mathcal{B}_{\mathbb{R}^{2d}})_=) & (x,x') \in \mathcal{X}_+^2 \\ \mathrm{dist}((x,x'),(2r\mathcal{B}_{\mathbb{R}^{2d}})_+ \cup (2r\mathcal{B}_{\mathbb{R}^{2d}})_=) & (x,x') \in \mathcal{X}_-^2 \\ 0 & (x,x') \in \mathcal{X}_=^2. \end{cases}$$

这里，$\mathcal{X}_+^2$、$\mathcal{X}_-^2$ 和 $\mathcal{X}_=^2$ 在（2.8）中定义，其中使用了原始的 η_+、η_- 和 $\eta_=$。

~我们声称，使用重新定义的 Δ后，间隔噪声条件以幂指数 β和常数 $2^\beta C_{**}$成立。为此，可以注意到，对于所有（x， x'）$\in$ $\mathcal{X}_+^2$，如果~Δ（x， x'）$<$ Δ（x， x'）/2，那么存在一个（u， u'）$\in$ $(2r\mathcal{B}_{\mathbb{R}^{2d}})_-$ 或（u， u'）$\in$ $(2r\mathcal{B}_{\mathbb{R}^{2d}})_=$，使得 dist（（$x$， x'），（u， u'））$<$ Δ（x， x'）/2。如果（u， u'）$\in$ $(2r\mathcal{B}_{\mathbb{R}^{2d}})_-$，那么必然存在某个（$v$， v'）$\in$ $\mathcal{X}_-^2$，使得（u， u'）$\in$ $\mathcal{B}_{\mathbb{R}^{2d}}$（（$v$， v'） Δ（v， v'）/2）。由三角不等式可得

$$\begin{aligned} \mathrm{dist}((x,x'),(v,v')) &\le \mathrm{dist}((x,x'),(u,u')) + \mathrm{dist}((u,u'),(v,v')) \\ &< \Delta(x,x')/2 + \Delta(v,v')/2, \end{aligned}$$

这与 Δ 的定义矛盾。如果 $(u,u') \in (2r\mathcal{B}_{\mathbb{R}^{2d}})_=$，那么

$$(u,u') \in 2r\mathcal{B}_{\mathbb{R}^{2d}} \setminus \bigcup_{(v,v')\in\mathcal{X}_+^2\cup\mathcal{X}_-^2} \mathcal{B}_{\mathbb{R}^{2d}}((v,v'),\Delta(v,v')/2),$$

这也与以下事实相矛盾：$\mathrm{dist}((x,x'),(u,u')) < \Delta(x,x')/2$。按照同样的论证，可以讨论 $(x,x') \in \mathcal{X}_-^2$ 的情形。因此，我们得出，对于所有 $(x,x') \in \mathcal{X}^2$， $\widetilde{\Delta}(x,x') \ge \Delta(x,x')/2$。于是， $\widetilde{\Delta}$ 满足间隔噪声条件，其形式为

$$\begin{aligned} &\int_{\{(x,x')\in\mathcal{X}^2:\widetilde{\Delta}(x,x')<t\}} |\eta_+(x,x') - \eta_-(x,x')| dP_{\mathcal{X}}^2(x,x') \\ \le &\int_{\{(x,x')\in\mathcal{X}^2:\Delta(x,x')<2t\}} |\eta_+(x,x') - \eta_-(x,x')| dP_{\mathcal{X}}^2(x,x') \\ \le &\, 2^\beta C_{**} t^\beta. \end{aligned} \tag{5.4}$$

现在我们扩展原始的 f^*_{hinge} as

$$f^*_{\text{hinge}}(x,x') = \mathbb{I}_{2r\mathcal{B}_{\mathbb{R}^{2d}}}(x,x')\text{sgn}(\eta_+(x,x') - \eta_-(x,x')).$$

可以验证，f^*_{hinge}保持反对称，并且f^*铰链$\in \mathcal{L}_2(\mathbb{R}^{2d})$与$\|f^*$铰链$\|_{\mathcal{L}_2(\mathbb{R}^{2d})} \leq \sqrt{\text{vol}(2r\mathcal{B}_{\mathbb{R}^{2d}})} = (2r)^d\sqrt{\text{vol}(\mathcal{B}_{\mathbb{R}^{2d}})}$。我们选择$\widetilde{f_0 \in \mathcal{H}_{K^\sigma}}$作为

$$f_0(x,x') = (\pi\sigma^2)^{-d/2}\widetilde{\mathcal{K}^\sigma} * f^*_{\text{hinge}}(x,x').$$

由于$\widetilde{\mathcal{K}^\sigma} * \cdot$是度量满射，

$$\|f_0\|_{K^\sigma} \leq (\pi\sigma^2)^{-d/2}\|f^*_{\text{hinge}}\|_{\mathcal{L}_2(\mathbb{R}^{2d})} \leq (\pi\sigma^2)^{-d/2}(2r)^d\sqrt{\text{vol}(\mathcal{B}_{\mathbb{R}^{2d}})}. \tag{5.5}$$

此外，根据杨氏卷积不等式，有

$$\|f_0\|_\infty \leq \frac{2^d}{\sigma^{2d}\pi^d}\|f^*_{\text{hinge}}\|_\infty \int_{\mathbb{R}^{2d}} \exp\left(-\frac{2\|(x,x')\|_2^2}{\sigma^2}\right) d(x,x') = 1. \tag{5.6}$$

还需界定超额风险 $\mathcal{R}^{\phi_{\text{hinge}}}(f_0) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}})$。固定一个$(x,x') \in \mathcal{X}^2_+ \subset \widetilde{(2r\mathcal{B}_{\mathbb{R}^{2d}})_+}$。对于所有 $(u,u') \in \mathcal{B}_{\mathbb{R}^{2d}}((x,x')\ \Delta(x,x'))$，有$(u,u') \in (2r\mathcal{B}_{\mathbb{R}^{2d}})_+$。因此

$$\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x')) \subset (2r\mathcal{B}_{\mathbb{R}^{2d}})_+, \quad (2r\mathcal{B}_{\mathbb{R}^{2d}})_- \subset \mathbb{R}^{2d}\backslash\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x'))$$

and

$$\begin{aligned}
& f_0(x,x') \\
=& \left(\frac{2}{\pi\sigma^2}\right)^d \int_{\mathbb{R}^{2d}} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) f^*_{\text{hinge}}(u,u')d(u,u') \\
=& \left(\frac{2}{\pi\sigma^2}\right)^d \left(\int_{(2r\mathcal{B}_{\mathbb{R}^{2d}})_+} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u')\right. \\
& \left. - \int_{(2r\mathcal{B}_{\mathbb{R}^{2d}})_-} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u')\right) \\
\geq& \left(\frac{2}{\pi\sigma^2}\right)^d \left(\int_{\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x'))} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u')\right. \\
& \left. - \int_{\mathbb{R}^{2d}\backslash\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x'))} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u')\right) \\
=& 2\left(\frac{2}{\pi\sigma^2}\right)^d \int_{\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x'))} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u') - 1.
\end{aligned}$$

由(5.6) 和 $f^*_{\text{hinge}}(x,x') = 1$，我们推导出

$$\begin{aligned}
& |f_0(x,x') - f^*_{\text{hinge}}(x,x')| \\
=& 1 - f_0(x,x') \\
\leq& 2 - 2\left(\frac{2}{\pi\sigma^2}\right)^d \int_{\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x'))} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u')
\end{aligned}$$

$$
\begin{aligned}
&= 2\left(\frac{2}{\pi\sigma^2}\right)^d \int_{\mathbb{R}^{2d}\setminus\mathcal{B}_{\mathbb{R}^{2d}}((x,x'),\widetilde{\Delta}(x,x'))} \exp\left(-\frac{2\|(x,x')-(u,u')\|_2^2}{\sigma^2}\right) d(u,u') \\
&= \frac{2^{d+2}}{\sigma^{2d}\Gamma(d)} \int_{\widetilde{\Delta}(x,x')}^{\infty} \exp(-2\sigma^{-2}\rho^2)\rho^{2d-1}d\rho \\
&= \frac{2}{\Gamma(d)} \int_{2\widetilde{\Delta}(x,x')^2\sigma^{-2}}^{\infty} e^{-\rho}\rho^{d-1}d\rho.
\end{aligned}
$$

这个不等式也通过相同的论证对所有 $(x,x') \in \mathcal{X}_-^2$ 成立。请注意,

$$
\begin{aligned}
\mathcal{R}^{\phi_{\text{hinge}}}(f_0) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}) &= \mathbb{E}[\text{sgn}(Y-Y')\cdot(f^*_{\text{hinge}}(X,X') - f_0(X,X'))] \\
&= \mathbb{E}_{P^2_{\mathcal{X}}}[(\eta_+(X,X') - \eta_-(X,X'))\cdot(f^*_{\text{hinge}}(X,X') - f_0(X,X'))],
\end{aligned}
$$

我们应用我们的断言（5.4），得到

$$
\begin{aligned}
&\mathcal{R}^{\phi_{\text{hinge}}}(f_0) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}) \\
&= \int_{\mathcal{X}_-^2\cup\mathcal{X}_+^2} |f_0(x,x') - f^*_{\text{hinge}}(x,x')| \cdot |\eta_+(x,x') - \eta_-(x,x')| dP^2_{\mathcal{X}}(x,x') \\
&\le \frac{2}{\Gamma(d)} \int_{\mathcal{X}_-^2\cup\mathcal{X}_+^2} \int_{2\widetilde{\Delta}(x,x')^2\sigma^{-2}}^{\infty} e^{-\rho}\rho^{d-1}|\eta_+(x,x') - \eta_-(x,x')| d\rho dP^2_{\mathcal{X}}(x,x') \\
&= \frac{2}{\Gamma(d)} \int_0^{\infty} e^{-\rho}\rho^{d-1} \int_{\mathcal{X}_+^2\cup\mathcal{X}_-^2} \mathbb{I}_{[0,\sigma(\rho/2)^{1/2})}(\widetilde{\Delta}(x,x'))|\eta_+(x,x') - \eta_-(x,x')| dP^2_{\mathcal{X}}(x,x') d\rho \\
&\le \frac{2^{1+\beta/2}C_{**}\sigma^{\beta}}{\Gamma(d)} \int_0^{\infty} e^{-\rho}\rho^{d-1+\beta/2} d\rho \\
&= \frac{2^{1+\beta/2}C_{**}\Gamma(d+\frac{\beta}{2})}{\Gamma(d)}\sigma^{\beta}.
\end{aligned}
$$

结合（5.5）和（5.6），我们得出 $f_0 \in \mathcal{H}_{K^\sigma}$、$\|f_0\|_\infty \le 1$ 和

$$
\lambda\|f_0\|^2_{K^\sigma} + \mathcal{R}^{\phi_{\text{hinge}}}(f_0) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}) \le \frac{2^{2d+1}r^{2d}}{\Gamma(d)}\lambda\sigma^{-2d} + \frac{2^{1+\beta/2}C_{**}\Gamma(d+\frac{\beta}{2})}{\Gamma(d)}\sigma^{\beta}.
$$

证明至此完成。 □

接下来，我们将在贝叶斯排序规则满足标准 Besov 光滑性条件下，建立平方损失的逼近误差估计

$$
f^*_{\text{square}}(x,x') = \frac{\eta_+(x,x') - \eta_-(x,x')}{\eta_+(x,x') + \eta_-(x,x')}.
$$

由于备注 2 中的讨论，我们可以将 f^*_{square} 视为 $\mathcal{L}_2(\mathbb{R}^{2d})$ 中的一个函数，或者等价地，f^*_{square} 是 $\mathcal{L}_2(\mathbb{R}^{2d})$ 中某个函数在 $\mathcal{X}^2$ 上的限制。

命题 6. 假设 $f^*{}_{\text{平方}} \in \mathcal{L}_2(\mathbb{R}^{2d})$ 和 $|f^*{}_{\text{平方}}|_{\mathcal{B}^{\alpha}_{2,\infty}(P^2_{\mathcal{X}})} < \infty$ 对于某个 $\alpha > 0$ 成立。那么对于所有 $\lambda > 0$、$\sigma > 0$，存在一个 $f_0 \in \mathcal{H}_{K^\sigma}$ 使得 $\|f_0\|_\infty \le 2^s$ 并且

$$
\begin{aligned}
&\lambda\|f_0\|^2_{K^\sigma} + \mathcal{R}^{\phi_{\text{square}}}(f_0) - \mathcal{R}^{\phi_{\text{square}}}(f^*_{\text{square}}) \\
&\le \frac{2^{2s}\|f^*_{\text{square}}\|^2_{\mathcal{L}_2(\mathbb{R}^{2d})}}{\pi^d}\lambda\sigma^{-2d} + \frac{|f^*_{\text{square}}|^2_{\mathcal{B}^{\alpha}_{2,\infty}(P^2_{\mathcal{X}})}}{2^{\alpha}}\left(\frac{\Gamma\left(d+\frac{\alpha}{2}\right)}{\Gamma(d)}\right)^2\sigma^{2\alpha}.
\end{aligned}
$$

~证明。定义卷积算子 $\mathcal{K}^\sigma$ 的 s 重应用为

$$\widetilde{\mathcal{K}}_s^\sigma * \cdot := \sum_{j=1}^{s} (-1)^{1-j} \binom{s}{j} (j\sigma\sqrt{\pi})^{-d} \widetilde{\mathcal{K}}^{j\sigma} * \cdot$$

并定义逼近器 $f_0 \in \mathcal{H}_{K^\sigma}$ 为

$$\begin{aligned}
&f_0(x,x') := \widetilde{\mathcal{K}}_s^\sigma * f^*_{\text{square}}(x,x') \\
=& \int_{\mathbb{R}^{2d}} \sum_{j=1}^{s} (-1)^{1-j} \binom{s}{j} \left(\frac{2}{\pi j^2\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{(j\sigma)^2}\right) f^*_{\text{square}}((x,x')+(u,u'))d(u,u') \\
=& \int_{\mathbb{R}^{2d}} \sum_{j=1}^{s} (-1)^{1-j} \binom{s}{j} \left(\frac{2}{\pi\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2}\right) f^*_{\text{square}}((x,x')+j(u,u'))d(u,u').
\end{aligned}$$

于是我们有

$$\begin{aligned}
&f_0(x,x') - f^*_{\text{square}}(x,x') \\
=& \widetilde{\mathcal{K}}_s^\sigma * f^*_{\text{square}}(x,x') - \int_{\mathbb{R}^{2d}} (\sigma\sqrt{\pi})^{-d} \mathcal{K}^\sigma(u,u') f^*_{\text{square}}(x,x') d(u,u') \\
=& \int_{\mathbb{R}^{2d}} \sum_{j=0}^{s} (-1)^{1-j} \binom{s}{j} \left(\frac{2}{\pi\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2}\right) f^*_{\text{square}}((x,x')+j(u,u'))d(u,u') \\
=& \int_{\mathbb{R}^{2d}} \sum_{j=0}^{s} (-1)^{1-j} \binom{s}{j} \left(\frac{2}{\pi\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2}\right) f^*_{\text{square}}((x,x')+j(u,u'))d(u,u') \\
=& \int_{\mathbb{R}^{2d}} (-1)^{1-s} \left(\frac{2}{\pi\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2}\right) \Delta^s_{(u,u')} f^*_{\text{square}}(x,x') d(u,u').
\end{aligned}$$

注意，根据 f^*_{square} 的定义，

$$\begin{aligned}
&\mathcal{R}^{\phi_{\text{square}}}(f_0) - \mathcal{R}^{\phi_{\text{square}}}(f^*_{\text{square}}) \\
=& \mathbb{E}[(1-\text{sgn}(Y-Y')f_0(X,X'))^2 - (1-\text{sgn}(Y-Y')f^*_{\text{square}}(X,X'))^2] \\
=& \mathbb{E}_{P^2_{\mathcal{X}}}\left[\left(\eta_+(2-f_0-f^*_{\text{square}}) - \eta_-(2+f_0+f^*_{\text{square}})\right)\cdot\left(f^*_{\text{square}}-f_0\right)\right] \\
=& \mathbb{E}_{P^2_{\mathcal{X}}}\left[(\eta_+ + \eta_-)\cdot(f^*_{\text{square}}-f_0)^2\right] \\
\leq& \|f^*_{\text{square}} - f_0\|^2_{\mathcal{L}_2(P^2_{\mathcal{X}})}.
\end{aligned} \tag{5.7}$$

根据闵可夫斯基积分不等式，我们有

$$\begin{aligned}
&\mathcal{R}^{\phi_{\text{square}}}(f_0) - \mathcal{R}^{\phi_{\text{square}}}(f^*_{\text{square}}) \\
\leq& \|f^*_{\text{square}} - f_0\|^2_{\mathcal{L}_2(P^2_{\mathcal{X}})} \\
=& \int_{\mathcal{X}^2} \left| \int_{\mathbb{R}^{2d}} (-1)^{1-s} \left(\frac{2}{\pi\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2}\right) \Delta^s_{(u,u')} f^*_{\text{square}}(x,x') d(u,u') \right|^2 dP^2_{\mathcal{X}}(x,x') \\
\leq& \left(\int_{\mathbb{R}^{2d}} \left(\int_{\mathcal{X}^2} \left(\left(\frac{2}{\pi\sigma^2}\right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2}\right) \Delta^s_{(u,u')} f^*_{\text{square}}(x,x') \right)^2 dP^2_{\mathcal{X}}(x,x') \right)^{\frac{1}{2}} d(u,u') \right)^2
\end{aligned}$$

$$
\begin{aligned}
&= \left(\int_{\mathbb{R}^{2d}} \left(\frac{2}{\pi\sigma^2} \right)^d \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2} \right) \|\Delta^s_{(u,u')} f^*_{\text{square}}\|_{\mathcal{L}_2(P^2_{\mathcal{X}})} d(u,u') \right)^2 \\
&\leq |f^*_{\text{square}}|^2_{\mathcal{B}^\alpha_{2,\infty}(P^2_{\mathcal{X}})} \left(\frac{2}{\pi\sigma^2} \right)^{2d} \left(\int_{\mathbb{R}^{2d}} \exp\left(-\frac{2\|(u,u')\|_2^2}{\sigma^2} \right) \|(u,u')\|_2^\alpha d(u,u') \right)^2 \\
&= |f^*_{\text{square}}|^2_{\mathcal{B}^\alpha_{2,\infty}(P^2_{\mathcal{X}})} 2^{-\alpha} \left(\frac{\Gamma(d+\frac{\alpha}{2})}{\Gamma(d)} \right)^2 \sigma^{2\alpha}.
\end{aligned}
$$

根据 [36], 中关于 $0 < \sigma_1 < \sigma_2 < \infty$的命题 4.46，空间 $\mathcal{H}_{\widetilde{K}^{\sigma_2}}$ 连续嵌入到 $\mathcal{H}_{\widetilde{K}^{\sigma_1}}$ 中，且算子范数满足：

$$
\left\| \text{id} : \mathcal{H}_{\widetilde{K}^{\sigma_2}} \to \mathcal{H}_{\widetilde{K}^{\sigma_1}} \right\| \leq \left(\frac{\sigma_2}{\sigma_1} \right)^d.
$$

于是我们可以将$\|f_0\|_{K^\sigma}$ 界定为

$$
\begin{aligned}
\|f_0\|_{K^\sigma} &\leq \sum_{j=1}^s \binom{s}{j} (j\sigma\sqrt{\pi})^{-d} \|\widetilde{\mathcal{K}}^{j\sigma} * f^*_{\text{square}}\|_{K^\sigma} \\
&= \sum_{j=1}^s \binom{s}{j} (j\sigma\sqrt{\pi})^{-d} \|\widetilde{\mathcal{K}}^{j\sigma} * f^*_{\text{square}}\|_{\widetilde{K}^\sigma} \\
&\leq \sum_{j=1}^s \binom{s}{j} (\sigma\sqrt{\pi})^{-d} \|\widetilde{\mathcal{K}}^{j\sigma} * f^*_{\text{square}}\|_{\widetilde{K}^{j\sigma}} \\
&\leq 2^s (\sigma\sqrt{\pi})^{-d} \|f^*_{\text{square}}\|_{\mathcal{L}_2(\mathbb{R}^{2d})}.
\end{aligned}
$$

最后，根据杨氏卷积不等式，我们有

$$
\begin{aligned}
\|f_0\|_\infty &\leq \|f^*_{\text{square}}\|_\infty \left\| \sum_{j=1}^s (-1)^{1-j} \binom{s}{j} (j\sigma\sqrt{\pi})^{-d} \widetilde{\mathcal{K}}^{j\sigma} \right\|_{L_1(\mathbb{R}^{2d})} \\
&\leq \sum_{j=1}^s \binom{s}{j} (j\sigma\sqrt{\pi})^{-d} \|\widetilde{\mathcal{K}}^{j\sigma}\|_{L_1(\mathbb{R}^{2d})} \\
&= \sum_{j=1}^s \binom{s}{j} \\
&\leq 2^s.
\end{aligned}
$$

结合所有这些估计，我们完成了证明。 □

5.3 成对排序的校准不等式

在本小节中，我们为采用基于间隔的损失 $\phi(y,y',t) = \psi(\text{sgn}(y-y')t)$ 的成对排序建立校准不等式。我们的讨论不仅限于 $\phi = \phi_{\text{hinge}}$ and$\phi = \phi_{\text{square}}$，还可以推广到更一般的基于间隔的损失研究。到目前为止，

我们的估计足以推导出超额 ϕ-排序风险的学习速率。为了通过超额 ϕ-排序风险来界定超额排序风险 $\mathcal{E}(f)$，还需建立所谓的校准不等式或比较定理。二元分类中的校准不等式已在大量文献中得到广泛研究，参见例如 [51, 5, 16, 36] 及其中的参考文献。受这些研究的启发，本小节旨在成对排序情形下构建类似的校准不等式。具体而言，对于 $\phi = \phi_{\text{hinge}}$，命题7中建立了Zhang不等式在成对排序中的一个类似版本。对于满足假设3且 $\psi''(0) > 0$ 的那些 ψ，其中包括 $\phi = \phi_{\text{square}}$ 这一特殊情形，我们证明超额排序风险可以由超额 ϕ-排序风险的平方根界定（乘以某个常数），而假设4中的噪声条件还可以进一步细化这一上界。

命题 7。对于所有可测的 $f\colon \mathcal{X}^2 \to \mathbb{R}$，我们有

$$\mathcal{R}(f) - \mathcal{R}(f^*_{\text{rank}}) \le \mathcal{R}^{\phi_{\text{hinge}}}(f) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}).$$

证明。考虑截断函数 $\pi(f)\colon \mathcal{X}^2 \to [-1,1]$。由于合页损失可以在 $t=1$ 处截断，且截断函数不会改变排序风险，因此只需对所有可测的 $f\colon \mathcal{X}^2 \to [-1,1]$证明该命题。注意，对于所有可测的 $f\colon \mathcal{X}^2 \to [-1,1]$，有

$$\mathcal{R}^{\phi_{\text{hinge}}}(f) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}) = \mathbb{E}_{P^2_{\mathcal{X}}}[(\eta_+(X,X') - \eta_-(X,X')) \cdot (f^*_{\text{hinge}}(X,X') - f(X,X'))].$$

Then

$$\begin{aligned}
&\mathcal{R}(f) - \mathcal{R}(f^*_{\text{rank}}) \\
&= \mathbb{E}_{P^2_{\mathcal{X}}}\left[\eta_+ \mathbb{I}_{(-\infty,0)}(f) + \eta_- \mathbb{I}_{[0,\infty)}(f) - \min\{\eta_+, \eta_-\}\right] \\
&\le \mathbb{E}_{P^2_{\mathcal{X}}}\left[(\eta_+ - \eta_-) \cdot (f^*_{\text{hinge}} - f)\right] \\
&= \mathcal{R}^{\phi_{\text{hinge}}}(f) - \mathcal{R}^{\phi_{\text{hinge}}}(f^*_{\text{hinge}}),
\end{aligned}$$

其中，该不等式可分别通过考虑 $\eta_+ > \eta_-$、$\eta_+ = \eta_-$ 和 $\eta_+ < \eta_-$这些情形推导得出。至此，证明完成。 □

为了继续分析一般的基于间隔的损失，我们进一步引入一些定义和符号。给定一个满足假设3的 ψ，固定一个 $(x,x') \in \mathcal{X}^2$，我们将函数 $\Psi = \Psi_{(x,x')}\colon \mathbb{R} \to [0,\infty)$ 定义为

$$\Psi_{(x,x')}(t) := \eta_+(x,x')\psi(t) + \eta_-(x,x')\psi(-t) + \eta_=(x,x')\psi(0).$$

由于 ψ 是凸函数，因此左导数 $\psi'_-(t) \ge 0$ 适用于 $t \in (1,\infty)$，右导数$\psi'_+(t) < 0$ 适用于 $t \in (-\infty,1)$。因此我们可以定义

$$\begin{aligned}
f^*_-(x,x') &:= \sup\left\{t \in \mathbb{R} \mid \Psi'_-(t) < 0\right\}, \\
f^*_+(x,x') &:= \inf\left\{t \in \mathbb{R} \mid \Psi'_+(t) > 0\right\}.
\end{aligned}$$

由于 Ψ 也是凸函数，因此我们有 $f^*_-(x,x') \le f^*_+(x,x')$。以下引理可用与 [16] 中定理 10.8 和引理 10.10 相同的论证证明。回顾一下，$\mathcal{X}^2_+$、$\mathcal{X}^2_-$和 $\mathcal{X}^2_=$ 定义于 (2.8) 中。

引理9. 给定 $\phi(y,y',t) = \psi(\mathrm{sgn}(y-y')t)$ 和 $(x,x') \in \mathcal{X}^2$。令 f_ϕ^*表示 Bayesϕ-排序规则。若 ψ 满足假设3，则以下结论成立。

1. $\Psi_{(x,x')}$在 $(-\infty,\ f_-^*(x,\ x')]$,上严格递减，在 $[f_+^*(x,\ x')\ \infty$上严格递增，并在 $[f_-^*(x,\ x')\ f_+^*(x,\ x')]$上保持常数。

2. $f_\phi^*(x,\ x')$是 $\Psi_{(x,x')}$ 的一个极小值点，即 $f_\phi^*(x,\ x') = \arg\min_{t\in\mathbb{R}} \Psi_{(x,x')}(t)$，且可取 $[f_-^*(x,\ x')\ f_+^*(x,\ x')]$中的任意值。

3. *There holds* $\begin{cases} 0 \le f_-^*(x,x') \le f_\phi^*(x,x'), & (x,x') \in \mathcal{X}_+^2, \\ f_\phi^*(x,x') \le f_+^*(x,x') \le 0, & (x,x') \in \mathcal{X}_-^2, \\ f_-^*(x,x') \le 0 \le f_+^*(x,x'), & (x,x') \in \mathcal{X}_=^2. \end{cases}$

4. $f_-^*(x,\ x') \le 1$和 $f_+^*(x,\ x') \ge -1$。

引理10. 若 ψ 满足假设3且 $\psi''(0) > 0$。则存在一个仅依赖于 ψ 的常数 $C_\psi > 0$，使得对所有 $(x,\ x') \in \mathcal{X}^2$

$$(\eta_+(x,x') - \eta_-(x,x'))^2 \le C_\psi\big(\Psi_{(x,x')}(0) - \Psi_{(x,x')}(f_\phi^*(x,x'))\big).$$

根据引理9，我们还可以看到，ϕ 在 $t=1$ 处被截断，并且贝叶斯 ϕ-排序规则可以逐点定义，且可取为反对称的。于是，我们可以建立成对排序的一般校准不等式。

命题8. 若 ψ 满足假设3且 $\psi''(0) > 0$。令 f_ϕ^* 为贝叶斯 ϕ-排序规则。则存在一个仅依赖于 ψ 的常数 $c_\psi > 0$，使得对于所有可测的f： $\mathcal{X}^2 \to \mathbb{R}$，我们有

$$\mathcal{R}(f) - \mathcal{R}(f_{\mathrm{rank}}^*) \le c_\psi \sqrt{\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*)}.$$

证明。给定 f： $\mathcal{X}^2 \to \mathbb{R}$，定义

$$\begin{aligned} \mathcal{X}_f^2 :=& \big\{(x,x') \in \mathcal{X}^2 \mid (x,x') \in \mathcal{X}_+^2 \text{ and } f(x,x') < 0\big\} \\ & \bigcup \big\{(x,x') \in \mathcal{X}^2 \mid (x,x') \in \mathcal{X}_-^2 \text{ and } f(x,x') \ge 0\big\}. \end{aligned}$$

然后，根据柯西-施瓦茨不等式和引理10，我们有

$$\begin{aligned} \mathcal{R}(f) - \mathcal{R}(f_{\mathrm{rank}}^*) &= \int_{\mathcal{X}_f^2} |\eta_+(x,x') - \eta_-(x,x')| dP_{\mathcal{X}}^2(x,x') \\ &\le \left(\int_{\mathcal{X}_f^2} |\eta_+(x,x') - \eta_-(x,x')|^2 dP_{\mathcal{X}}^2(x,x') \right)^{\frac{1}{2}} \\ &\le \left(C_\psi \int_{\mathcal{X}_f^2} \Psi_{(x,x')}(0) - \Psi_{(x,x')}(f_\phi^*(x,x')) dP_{\mathcal{X}}^2(x,x') \right)^{\frac{1}{2}}. \end{aligned}$$

对于 $(x,x') \in \mathcal{X}_f$，若 $(x,x') \in \mathcal{X}_+^2$，则 $f(x,x') < 0$且 $f_{\mathrm{rank}}^*(x,x') = 1$。由引理9， $\Psi_{(x,x')}$在 $(-\infty,0]$ 上严格递减，因此 $\Psi_{(x,x')}(0) \le \Psi_{(x,x')}(f(x,x'))$。沿用相同的 $_1$ 论证，不等式对 $(x,x') \in \mathcal{X}_-^2$也成立。因此，令 $c_\psi := C_\psi^{\frac{1}{2}}$，我们得到

$$\mathcal{R}(f) - \mathcal{R}(f_{\mathrm{rank}}^*) \le \left(C_\psi \int_{\mathcal{X}_f^2} \Psi_{(x,x')}(0) - \Psi_{(x,x')}(f_\phi^*(x,x')) dP_{\mathcal{X}}^2(x,x') \right)^{\frac{1}{2}}$$

$$\leq \left(C_\psi \int_{\mathcal{X}_f^2} \Psi_{(x,x')}(f(x,x')) - \Psi_{(x,x')}(f_\phi^*(x,x')) dP_{\mathcal{X}}^2(x,x') \right)^{\frac{1}{2}}$$
$$\leq c_\psi \sqrt{\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*)}.$$

证明至此完成。 □

此外，在假设4中的噪声条件下，可以建立一个更精细的预言机不等式。

命题9. 假设 ψ 满足假设3， $\psi''(0) > 0$ 且假设4对 $C_* > 0$ 和 $q \in [0, \infty]$成立。令 f_ϕ^* 为 Bayes ϕ-排序规则。则存在一个常数$c_{\psi,q} > 0$ 仅取决于 ψ、q 和 C_*，使得对于所有可测的 $f\colon \mathcal{X}^2 \to \mathbb{R}$,

$$\mathcal{R}(f) - \mathcal{R}(f_{\text{rank}}^*) \leq c_{\psi,q} (\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*))^{\frac{q+1}{q+2}}.$$

证明。对于 $t > 0$，我们记

$$U_t := \{(x,x') \in \mathcal{X}^2 : |\eta_+(x,x') - \eta_-(x,x')| \leq t\},$$
$$V_t := \{(x,x') \in \mathcal{X}^2 : |\eta_+(x,x') - \eta_-(x,x')| > t\}.$$

回顾 $\mathcal{X}_f^2$，它在命题9的证明中定义。根据假设4以及上一个证明中的分析，我们有

$$\begin{aligned}
&\mathcal{R}(f) - \mathcal{R}(f_{\text{rank}}^*) \\
=& \int_{\mathcal{X}_f^2} |\eta_+(x,x') - \eta_-(x,x')| dP_{\mathcal{X}}^2(x,x') \\
=& \int_{\mathcal{X}_f^2 \cap U_t} |\eta_+(x,x') - \eta_-(x,x')| dP_{\mathcal{X}}^2(x,x') + \int_{\mathcal{X}_f^2 \cap V_t} |\eta_+(x,x') - \eta_-(x,x')| dP_{\mathcal{X}}^2(x,x') \\
\leq& t P_{\mathcal{X}}^2(U_t) + \frac{1}{t} \int_{\mathcal{X}_f^2 \cap V_t} |\eta_+(x,x') - \eta_-(x,x')|^2 dP_{\mathcal{X}}^2(x,x') \\
\leq& C_* t^{q+1} + \frac{C_\psi(\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*))}{t}.
\end{aligned}$$

那么，最优选择

$$t := \left(\frac{C_\psi(\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*))}{(q+1)C_*} \right)^{\frac{1}{q+2}}$$

由此可得 久久
精品国产

$$\mathcal{R}(f) - \mathcal{R}(f_{\text{rank}}^*) \leq C_* \left(\frac{C_\psi(\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*))}{(q+1)C_*} \right)^{\frac{q+1}{q+2}}$$
$$+ (q+1)^{\frac{1}{q+2}} C_*^{\frac{1}{q+2}} C_\psi^{\frac{q+1}{q+2}} (\mathcal{R}^\phi(f) - \mathcal{R}^\phi(f_\phi^*))^{\frac{q+1}{q+2}}.$$

通过取 $c_{\psi,q} := (q+1)^{-\frac{q+1}{q+2}} C_*^{\frac{1}{q+2}} C_\psi^{\frac{q+1}{q+2}} + (q+1)^{\frac{1}{q+2}} C_*^{\frac{1}{q+2}} C_\psi^{\frac{q+1}{q+2}}$，我们完成了证明。 □

我们看到，对于 $q=0$，在假设4平凡成立的情况下，命题9 退化为命题 8。当 $q>0$ 时，相较于命题 8 中给出的前一个校准不等式，该校准不等式得到了改进。特别地，当 $q=\infty$ 时，我们得到一个与命题 7 中针对 ϕ_{hinge} 所建立的不等式类似的不等式。

5.4 采用合页损失和平方损失的高斯排序估计量的学习速率

现在我们可以证明带有 $\phi=\phi_{\text{hinge}}$ 和 $\phi=\phi_{\text{square}}$的高斯排序估计量的预言机不等式，进而推导出学习速率。为此，只需验证假设1，尤其是方差界(2.6)。对于合页损失，可以利用假设4中的噪声条件来建立方差界。

引理11. 若假设4对 $C_*>0$ 和 $q\in[0,\infty]$ 成立，则对于所有可测的 $f\colon\ \mathcal{X}^2\to[-1,1]$，有

$$\mathbb{E}(Q\phi_{\text{hinge}}f-Q\phi_{\text{hinge}}f^*_{\text{hinge}})^2\le V\big(\mathbb{E}(Q\phi_{\text{hinge}}f-Q\phi_{\text{hinge}}f^*_{\text{hinge}})\big)^{\tau},$$

where $V=2^{\frac{q+2}{q+1}}C_*^{\frac{1}{q+1}}q^{\frac{1}{q+1}}(1+q^{-1})$ *and* $\tau=\frac{q}{q+1}$.

证明。回想在命题 9 的证明中定义的 U_t 和 V_t，其定义为

$$U_t=\big\{(x,x')\in\mathcal{X}^2:\big|\eta_+(x,x')-\eta_-(x,x')\big|\le t\big\},$$
$$V_t=\big\{(x,x')\in\mathcal{X}^2:\big|\eta_+(x,x')-\eta_-(x,x')\big|>t\big\}.$$

那么我们有

$$\begin{aligned}
&\mathbb{E}(Q\phi_{\text{hinge}}f-Q\phi_{\text{hinge}}f^*_{\text{hinge}})^2\\
&=\int_{\mathcal{X}\times\mathcal{Y}}\left(\int_{\mathcal{X}\times\mathcal{Y}}\phi_{\text{hinge}}(y,y',f(x,x'))-\phi_{\text{hinge}}(y,y',f^*_{\text{hinge}}(x,x'))dP(x',y')\right)^2dP(x,y)\\
&\le\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}}\left(\phi_{\text{hinge}}(y,y',f(x,x'))-\phi_{\text{hinge}}(y,y',f^*_{\text{hinge}}(x,x'))\right)^2dP(x',y')dP(x,y)\\
&\le\int_{\mathcal{X}^2}|f(x,x')-f^*_{\text{hinge}}(x,x')|^2dP^2_{\mathcal{X}}(x,x')\\
&\le 2\int_{\mathcal{X}^2}|f(x,x')-f^*_{\text{hinge}}(x,x')|dP^2_{\mathcal{X}}(x,x')\\
&=2\int_{U_t\cup V_t}|f(x,x')-f^*_{\text{hinge}}(x,x')|dP^2_{\mathcal{X}}(x,x')\\
&\le 4P^2_{\mathcal{X}}(U_t)+\frac{2}{t}\int_{V_t}|\eta_+(x,x')-\eta_-(x,x')|\cdot|f(x,x')-f^*_{\text{hinge}}(x,x')|dP^2_{\mathcal{X}}(x,x')\\
&\le 4C_*t^q+2t^{-1}\mathbb{E}(Q\phi_{\text{hinge}}f-Q\phi_{\text{hinge}}f^*_{\text{hinge}})\\
&\le 2^{\frac{q+2}{q+1}}C_*^{\frac{1}{q+1}}q^{\frac{1}{q+1}}(1+q^{-1})\big(\mathbb{E}(Q\phi_{\text{hinge}}f-Q\phi_{\text{hinge}}f^*_{\text{hinge}})\big)^{\frac{q}{q+1}}.
\end{aligned}$$

在最后一个不等式中，我们选择最优的

$$t:=\left(\frac{\mathbb{E}(Q\phi_{\text{hinge}}f-Q\phi_{\text{hinge}}f^*_{\text{hinge}})}{2qC_*}\right)^{\frac{1}{q+1}}$$

将总和最小化。证明至此完成。 □

现在我们可以证明定理 3，它给出了$\phi = \phi_{\text{hinge}}$的预言机不等式和学习速率。

定理 3 的证明。结合命题 4、命题 5 和命题 11，直接应用定理 2，其中 $L = 1, B = 2$, $B_0 = 2, \tau = \frac{q}{q+1}, V = 2^{\frac{q+2}{q+1}} C_*^{\frac{1}{q+1}} q^{\frac{1}{q+1}} (1+q^{-1}) \leq 6 C_*^{\frac{1}{q+1}}, p_1 = p_2 = p, a_1 = a_2 = (C_{\mathcal{X}}^* p^{-2d-1} \sigma^{-2\varrho})^{\frac{1}{2p}}$，即可得到

$$
\begin{aligned}
&\lambda \|f_{\mathbf{z}}\|_{K^\sigma}^2 + \mathcal{R}^{\phi_{\text{hinge}}}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi_{\text{hinge}}}(f_{\text{hinge}}^*) \\
\leq\ & \frac{2^{3d+4} r^{2d}}{\Gamma(d)} \lambda \sigma^{-2d} + \frac{2^{\beta/2+4} C_{**} \Gamma(d+\frac{\beta}{2})}{\Gamma(d)} \sigma^\beta + 36 c_1 \left(\frac{C_{\mathcal{X}}^* C_*^{\frac{1-p}{q+1}}}{\lambda^p p^{2d+1} \sigma^{2\varrho} n} \right)^{\frac{q+1}{q-p+2}} \\
&+ \frac{6 C_{\mathcal{X}}^* (2c_2 + c_3 t + c_4 t)}{\lambda^p p^{2d+1} \sigma^{2\varrho} n} + \left(\frac{11232 C_*^{\frac{1}{q+1}} t}{n} \right)^{\frac{q+1}{q+2}} + \frac{(2712 + 3c_5) t}{n}.
\end{aligned}
$$

回顾命题3的证明，并注意到常数 c_3 和 c_4 取决于 p。当$p \in (0, 1/4]$时，可以验证 c_3 和 c_4 均有一致的上界。取 $c_6 = \sup\{c_i : 1 \leq i \leq 5, 0 < p \leq 1/4\}$，即可完成预言机不等式的证明。

为推导学习速率，对所有 $t \geq 1$，记

$$
\mathcal{R}^{\phi_{\text{hinge}}}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi_{\text{hinge}}}(f_{\text{hinge}}^*) \lesssim \lambda \sigma^{-2d} + \sigma^\beta + \left(\frac{t}{\lambda^p p^{2d+1} \sigma^{2\varrho} n} \right)^{\frac{q+1}{q-p+2}} + \left(\frac{t}{n} \right)^{\frac{q+1}{q+2}}.
$$

采用定理陈述中指定的选择，我们有

$$
\lambda < 1, \quad \sigma^{2\varrho} n \geq 1, \quad \frac{q+1}{q+2} \leq \frac{q+1}{q-p+2} \leq 1,
$$

由此可得

$$
\begin{aligned}
\left(\frac{t}{p^{2d+1} \sigma^{2d} \lambda^p n} \right)^{\frac{q+1}{q-p+2}} &\leq t p^{-2d-1} \lambda^{-p} (\sigma^{2\varrho} n)^{-\frac{q+1}{q+2}} \\
&\lesssim t \lambda^{-1/\log n} (\sigma^{2\varrho} n)^{-\frac{q+1}{q+2}} \log^{2d+1} n \\
&= t e^b n^{-\frac{\beta(q+1)}{\beta(q+2)+2\varrho(q+1)}} \log^{2d+1} n,
\end{aligned}
$$

而其他项的收敛速率都比这更快。因此，对所有 $n \geq 2, t \geq 1$，根据命题 7，以至少 $1 - (c_0 + 5) \exp(-t)$ 的概率，我们有

$$
\mathcal{E}(\pi(f_{\mathbf{z}})) \leq \mathcal{E}^{\phi_{\text{hinge}}}(\pi(f_{\mathbf{z}})) \lesssim t n^{-\frac{\beta(q+1)}{\beta(q+2)+2\varrho(q+1)}} \log^{2d+1} n.
$$

至此，证明完成。 □

对于 $\phi = \phi_{\text{square}}$，方差界可以直接证明。

引理 12. 对所有可测的 $f\colon\ \mathcal{X}^2 \to [-1,1]$ 我们有

$$\mathbb{E}(Q\phi_{\mathrm{square}f} - Q\phi_{\mathrm{square}f^*_{\mathrm{square}}})^2 \le 16\mathbb{E}(Q\phi_{\mathrm{square}f} - Q\phi_{\mathrm{square}f^*_{\mathrm{square}}}).$$

证明。回忆我们在 (5.7) 中所做的计算，可得

$$\begin{aligned}
&\mathbb{E}(Q\phi_{\mathrm{square}f} - Q\phi_{\mathrm{square}f^*_{\mathrm{square}}})^2 \\
&= \int_{\mathcal{X}\times\mathcal{Y}} \left(\int_{\mathcal{X}\times\mathcal{Y}} \phi_{\mathrm{square}}(y,y',f(x,x')) - \phi_{\mathrm{square}}(y,y',f^*_{\mathrm{square}}(x,x'))dP(x',y') \right)^2 dP(x,y) \\
&\le \int_{\mathcal{X}\times\mathcal{Y}} \int_{\mathcal{X}\times\mathcal{Y}} \left(\phi_{\mathrm{square}}(y,y',f(x,x')) - \phi_{\mathrm{square}}(y,y',f^*_{\mathrm{square}}(x,x')) \right)^2 dP(x',y')dP(x,y) \\
&= \mathbb{E}_{P^2_{\mathcal{X}}}\left[\eta_+(2-f-f^*_{\mathrm{square}})^2(f^*_{\mathrm{square}}-f)^2 + \eta_-(2+f+f^*_{\mathrm{square}})^2(f^*_{\mathrm{square}}-f)^2\right] \\
&\le 16\mathbb{E}_{P^2_{\mathcal{X}}}\left[(\eta_+ + \eta_-)(f^*_{\mathrm{square}}-f)^2\right] \\
&= 16(\mathcal{R}^{\phi_{\mathrm{square}}}(f) - \mathcal{R}^{\phi_{\mathrm{square}}}(f^*_{\mathrm{square}})) \\
&= 16\mathbb{E}(Q\phi_{\mathrm{square}f} - Q\phi_{\mathrm{square}f^*_{\mathrm{square}}}).
\end{aligned}$$

至此，证明完成。 □

本小节末尾，我们证明定理 4，该定理给出了 $\phi = \phi_{\mathrm{square}}$ 的预言机不等式和学习速率。

定理 4 的证明。结合命题 4、命题 6 和命题 12，我们可以将定理 2 应用于 $L = 4, B = 4,$ $B_0 = 2^{2s+2}, \tau = 1, V = 16, p_1 = p_2 = p, a_1 = a_2 = (C^*_{\mathcal{X}} p^{-2d-1}\sigma^{-2\varrho})^{\frac{1}{2p}}$，以及我们在定理 3 的证明中引入的 c_6，从而得到

$$\begin{aligned}
&\lambda\|f_{\mathbf{z}}\|^2_{K^\sigma} + \mathcal{R}^{\phi_{\mathrm{square}}}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi_{\mathrm{square}}}(f^*_{\mathrm{square}}) \\
&\le \frac{2^{2s+3}\|f^*_{\mathrm{square}}\|^2_{\mathcal{L}_2(\mathbb{R}^{2d})}}{\pi^d}\lambda\sigma^{-2d} + 2^{3-\alpha}\left(\frac{\Gamma\left(d+\frac{\alpha}{2}\right)}{\Gamma(d)}\right)^2 |f^*_{\mathrm{square}}|^2_{\mathcal{B}^\alpha_{2,\infty}(P^2_{\mathcal{X}})}\sigma^{2\alpha} \\
&\quad + \frac{C^*_{\mathcal{X}}(96c_6 + 72c_6 t)}{\lambda^p p^{2d+1}\sigma^{2\varrho}n} + \frac{(33552 + 1824\cdot 2^{2s} + 3c_6)t}{n}.
\end{aligned}$$

为推导学习速率，对所有 $t \ge 1$，写出

$$\mathcal{R}^{\phi_{\mathrm{square}}}(\pi(f_{\mathbf{z}})) - \mathcal{R}^{\phi_{\mathrm{square}}}(f^*_{\mathrm{square}}) \lesssim \lambda\sigma^{-2d} + \sigma^{2\alpha} + \frac{t}{\lambda^p p^{2d+1}\sigma^{2\varrho}n} + \frac{t}{n}.$$

根据定理陈述中指定的选择，我们有

$$\lambda < 1, \quad \sigma^{2\varrho}n \ge 1,$$

which leads to

$$\frac{t}{p^{2d+1}\sigma^{2\varrho}\lambda^p n} \lesssim t\lambda^{-1/\log n}(\sigma^{2\varrho}n)^{-1}\log^{2d+1} n = e^b n^{\frac{\alpha}{\alpha+\varrho}}\log^{2d+1} n,$$

而其他项的收敛速率比这更快。因此，对于所有 $n \ge 2$、$t \ge 1$，根据命题 8，以至少为 $1-(c_0+5)\exp(-t)$ 的概率，我们有

$$\mathcal{E}(\pi(f_{\mathbf{z}})) \lesssim \sqrt{\mathcal{E}^{\phi_{\mathrm{square}}}(\pi(f_{\mathbf{z}}))} \lesssim \sqrt{t}n^{-\frac{\alpha}{2(\alpha+\varrho)}}\log^{d+\frac{1}{2}} n.$$

此外，如果 P 还满足假设4，根据命题9，以至少 $1-(c_0+5)\exp(-t)$ 的概率，我们有

$$\mathcal{E}(\pi(f_{\mathbf{z}})) \lesssim (\mathcal{E}^{\phi_{\mathrm{square}}}(\pi(f_{\mathbf{z}})))^{\frac{q+1}{q+2}} \lesssim t^{\frac{q+1}{q+2}} n^{-\frac{(q+1)\alpha}{(q+2)(\alpha+\varrho)}} \log^{\frac{(q+1)(2d+1)}{q+2}} n.$$

证明至此完成。 □

5.5 噪声条件的比较

本小节中，我们对不同的噪声条件进行比较。令 TN(q) 表示假设4中的噪声条件。 [12] 和 [13] 在二部排序的背景下提出了另外两种噪声条件，即全局低噪声条件 LN(q) 和 NA(q)，它们在成对排序设定下可改写为以下条件。

假设7（全局低噪声条件 LN(q)）。存在常数 $C_{\mathrm{LN}}>0$ 和$q\in[0,\infty]$，使得对所有 $x\in\mathcal{X}$,

$$\mathbb{E}_{X'}[|\eta_+(x,X')-\eta_-(x,X')|^{-q}]\le C_{\mathrm{LN}}.$$

假设8（全局低噪声条件 NA(q)）。存在常数 $C_{\mathrm{NA}}>0$和 $q\in[0,\infty]$，使得对所有 $x\in\mathcal{X}$和 $t>0$,

$$P_{\mathcal{X}}(\{x'\in\mathcal{X}:|\eta_+(x,x')-\eta_-(x,x')|\le t\})\le C_{\mathrm{NA}}t^q.$$

类似于 [13] 中命题3的证明（该命题描述了 LN(q) 与 NA(q) 之间的联系），我们可以证明，对于所有 $q'<q$，LN(q) 蕴含 NA(q)，而 NA(q) 蕴含 LN(q')，因此 NA(q) 可视为一个稍弱的条件。

显然，NA(q)蕴含 TN(q)，下面给出一个说明该蕴含关系是尖锐的例子。考虑 $P_{\mathcal{X}}$ 为 $\mathcal{X}=[0,1]$ 上的均匀分布，并令标签为 $Y=X+\epsilon$，其中 ϵ 是均匀分布于 $[0,1]$上的随机噪声。条件分布 $P_{Y|X}$与 $P_{\mathcal{X}}$共同生成 $\mathcal{X}\times\mathcal{Y}=[0,1]\times[0,2]$上的分布 P。简单计算表明，$|\eta_+(x,x')-\eta_-(x,x')|=2|x-x'|-|x-x'|^2$，因此集合 $\{(x,x')\in\mathcal{X}^2:|\eta_+(x,x')-\eta_-(x,x')|\le t\}$ 是沿对角线的带状域，其宽度和概率测度渐近为 $O(t)$，当 $t\to 0$时。因此， P 满足 NA(1) 和 TN(1)，但对于所有 $q>1$ 均不满足TN(q)。这也是一个 P 不满足LN(1)，而对所有 $q<1$均满足 LN(q) 的例子。

另一方面，一般而言，TN(q)不能推出对于任意 $q'\in(0,q]$ 的 NA(q')，因为全局低噪声条件是在每个 $x\in\mathcal{X}$ 处陈述的，而TN(q)只要求乘积概率测度具有快速衰减速度。我们可以轻易地在 $\mathcal{X}^2$ 上构造一个集合递减序列，其测度快速衰减，但它们在某个点$x\in\mathcal{X}$处的截面测度仍保持为一个正的常数。

注意到 LN(q)和NA(q) 严格强于 TN(q) 后，人们可能会预期，全局低噪声条件可以推导出一个更大的方差界指数 τ，相比于 TN(q) 下引理 11 中所述的 $\tau=\frac{q}{q+1}$。事实上，我们有以下结果。

命题10。以下断言成立。

1. 如果 P满足带有 $q \in [0,1]$ 的 LN(q)，那么对于所有可测的 $f\colon\ \mathcal{X}^2 \to [-1,1]$ 我们有

$$\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})^2 \le V\big(\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})\big)^\tau,$$

其中 $V = 4C_{\text{LN}}$ 和 $\tau = q$。

2. 若 P满足 NA(q)，且 $q \in [0,\infty]$，则对于所有可测的 $f\colon\ \mathcal{X}^2 \to [-1,1]$，我们有

$$\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})^2 \le V\big(\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})\big)^\tau,$$

where $V = 2^{\frac{2q+3}{2q+1}} C_{\text{NA}}^{\frac{2}{2q+1}} q^{\frac{1}{2q+1}} (2+q^{-1})$ *and* $\tau = \frac{2q}{2q+1}$.

3. 根据全局低噪声条件可推导出的最佳方差指数 τ 如下所示:

τ	$q \in [0,1/2]$	$q \in (1/2,1]$	$q \in (1,\infty]$
LN(q)	$2q/(2q+1)$	q	1
NA(q)	$2q/(2q+1)$	$(1/2,q)$	1

证明。如果 LN(q) 在 $q \in [0,1]$ 下成立，根据柯西-施瓦茨不等式和詹森不等式，我们有

$$\begin{aligned}
&\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})^2 \\
&\le \mathbb{E}_X(\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')|)^2 \\
&\le C_{\text{LN}}\mathbb{E}_X(\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')|^2 \cdot |\eta_+(X,X') - \eta_-(X,X')|^q) \\
&\le 4C_{\text{LN}}\mathbb{E}_X(\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')|^q \cdot |\eta_+(X,X') - \eta_-(X,X')|^q) \\
&\le 4C_{\text{LN}}(\mathbb{E}_X\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')| \cdot |\eta_+(X,X') - \eta_-(X,X')|)^q \\
&= 4C_{\text{LN}}\big(\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})\big)^q.
\end{aligned}$$

若 NA(q) 与 $q \in [0,\infty]$ 一同成立，则我们有

$$\begin{aligned}
&\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})^2 \\
&\le \mathbb{E}_X(\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')| \cdot \\
&\quad (\mathbb{I}_{\{|\eta_+(X,X') - \eta_-(X,X')| \le t\}} + \mathbb{I}_{\{|\eta_+(X,X') - \eta_-(X,X')| > t\}}))^2 \\
&\le 2\mathbb{E}_X(\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')| \cdot \mathbb{I}_{\{|\eta_+(X,X') - \eta_-(X,X')| \le t\}})^2 \\
&\quad + 2\mathbb{E}_X(\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')| \cdot \mathbb{I}_{\{|\eta_+(X,X') - \eta_-(X,X')| > t\}})^2 \\
&\le 8C_{\text{NA}}^2 t^{2q} + 2\mathbb{E}_X\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')|^2 \cdot \mathbb{I}_{\{|\eta_+(X,X') - \eta_-(X,X')| > t\}} \\
&\le 8C_{\text{NA}}^2 t^{2q} + \frac{4}{t}\mathbb{E}_X\mathbb{E}_{X'}|f(X,X') - f^*_{\text{hinge}}(X,X')| \cdot |\eta_+(X,X') - \eta_-(X,X')| \\
&\le 8C_{\text{NA}}^2 t^{2q} + \frac{4}{t}\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}}) \\
&\le 2^{\frac{2q+3}{2q+1}} C_{\text{NA}}^{\frac{2}{2q+1}} q^{\frac{1}{2q+1}} (2+q^{-1}) \big(\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})\big)^{\frac{2q}{2q+1}}.
\end{aligned}$$

在最后一个不等式中，我们选择最优的

$$t := \left(\frac{\mathbb{E}(Q\phi_{\text{hinge}f} - Q\phi_{\text{hinge}f^*_{\text{hinge}}})}{4qC_{\text{NA}}^2}\right)^{\frac{1}{2q+1}}$$

以使总和最小。

对于最佳方差指数，我们利用 LN(q) 与 NA(q) 之间的联系。若$q \in [0, 1/2]$，LN(q) 蕴含 NA(q)，且 $2q/(2q+1) \geq q$，因此 $\tau = 2q/(2q+1)$。若 $q \in (1/2, 1]$，NA(q) 蕴含对所有 $q' < q$ 均有 LN(q')，因此 τ 可以任意接近 q，其中常数 V 取决于 τ。若 $q \in (1, \infty]$，LN(q) 或 NA(q) 均蕴含 LN(1)，因此 $\tau = 1$。由此证明完毕。 □

从上述讨论中可以看出，与假设4中的噪声条件相比，全局低噪声条件在相同的噪声指数 q 下总能推出更精确的指数 τ，因而能够推出更快的学习速率。可以在具体实例中仔细考察这些噪声条件。

参考文献

[1] Shivani Agarwal 和 Partha Niyogi。通过算法稳定性获得排序算法的泛化界。《机器学习研究杂志》，10(16):441–474，2009。[2] Miguel A. Arcones。U统计量和U过程的 Bernstein 型不等式。《统计与概率快报》，22(3):239–247，1995。[3] Miguel A. Arcones 和 Evarist Gin´e。由 Vapnik–˘Cervonenkis 函数类索引的 U过程及其在带估计参数的 U统计量渐近性和自助法中的应用。《随机过程及其应用》，52(1):17–38，1994。[4] N. Aronszajn。再生核理论。《美国数学会汇刊》，68(3):337–404，1950。[5] Peter L Bartlett、Michael I Jordan 和 Jon D McAuliffe。凸性、分类与风险界。《美国统计学会杂志》，101(473):138–156，2006。[6] A. Ben-Hur 和 W. S. Noble。用于预测蛋白质—蛋白质相互作用的核方法。《生物信息学》，21(增刊1):i38–i46，2005。[7] Chris Burges、Tal Shaked、Erin Renshaw、Ari Lazier、Matt Deeds、Nicole Hamilton 和 Greg Hullender。使用梯度下降学习排序。载于《第22届国际机器学习会议论文集》，ICML ’ 05，第89–96页，美国纽约州纽约市，2005年。计算机协会。[8] Yunbo Cao、Jun Xu、Tie-Yan Liu、Hang Li、Yalou Huang 和 Hsiao-Wuen Hon。调整排序支持向量机以进行文档检索。载于《第29届年度国际 ACM SIGIR 信息检索研究与发展会议论文集》，SIGIR ’ 06，第186–193页，美国纽约州纽约市，2006年。计算机协会。[9] Di-Rong Chen 和 Shou-You Huang。具有指数强混合序列的用于排序的 L1 范数支持向量机。《国际小波、多分辨率与信息处理杂志》，12(05):1461001，2014。[10] Heng Chen 和 Di-Rong Chen。用于排序的支持向量机的学习速率。《逼近理论杂志》，188:57–68，2014。

[11] Hong Chen。正则化排序算法的收敛速率。《逼近理论杂志》，164(12):1513–1519，2012。[12] St´ephan Cl´emen¸con、G´abor Lugosi 和 NicolasVayatis。U统计量的排序与经验最小化。《统计年刊》，36(2):844–874，2008。[13] St´ephan Cl´emen¸con和Sylvain Robbiano。二部排序与插件规则的极小极大学习速率。载于《机器学习》国际会议，第441–448页，美国，2011年。[14] Corinna Cortes、Mehryar Mohri 和Ashish Rastogi。保持幅度的排序算法。载于第24届《机器学习》国际会议论文集，ICML'07，第169–176页，美国纽约州纽约市，2007年。计算机协会。[15] Corinna Cortes 和 Vladimir Vapnik。支持向量网络。《机器学习》，20(3):273–297，1995。[16] FelipeCucker和DingXuanZhou。学习理论：一种逼近理论视角。剑桥大学出版社，2007。[17] Kenneth Falconer。分形几何：数学基础与应用。John Wiley & Sons，2004。[18] ThomasHamm和IngoSteinwart。针对低内在维数数据的支持向量机自适应学习速率。《统计年刊》，49(6):3153–3180，2021。[19] LarsH¨ormander。线性偏微分算子分析I：分布理论与傅里叶分析。Springer，2015。[20] Thorsten Joachims。利用点击流数据优化搜索引擎。载于第八届ACM SIGKDD知识发现与数据挖掘国际会议论文集，KDD'02，第133–142页，美国纽约州纽约市，2002年。计算机协会。[21] Iain M. Johnstone。预言机不等式与非参数函数估计。Documenta Mathematica，III:267–278，1998。[22] G. Kerkyacharian和D. Picard。Besov空间中的密度估计。《统计与概率快报》，13(1):15–24，1992。[23] ThomasK¨uhn。高斯再生核希尔伯特空间的覆盖数。《复杂性杂志》，27(5):489–499，2011。[24] V´ıctor H. Dela Pe˜na、Victor H.De la Pe˜na、Victor delaPe˜na和EvaristGin´e。解耦：从依赖到独立。施普林格科学与商业媒体，1999。[25] Yunwen Lei、Shao-BoLin 和 Ke Tang。正则化成对学习的泛化界。载于第二十七届国际人工智能联合会议论文集，第2376–2382页，瑞典斯德哥尔摩，2018年。国际人工智能联合会议组织。

[26] Yi Lin. 分类中基于间隔的损失函数注记。《统计与概率快报》，68(1):73–82，2004。[27] G. G. Lorentz.《函数逼近》。Holt, Rinehart and Winston，1966。[28] Deborah Nolan 和 David Pollard. U过程：收敛速率。《统计年刊》，15(2):780–799，1987。[29] Tapio Pahikkala、Antti Airola、Michiel Stock、Bernard De Baets 和 Willem Waegeman. 面向关系数据条件排序的高效正则化最小二乘算法。《机器学习》，93(2-3):321–356，2013。[30] Tapio Pahikkala、Evgeni Tsivtsivadze、Antti Airola、Jorma Boberg 和 Tapio Salakoski. 使用成对正则化最小二乘进行学习排序。载于 SIGIR 2007 信息检索中的学习排序研讨会，第27–33页，2007。[31] Tapio Pahikkala、Evgeni Tsivtsivadze、Antti Airola、Jouni J¨arvinen 和 Jorma Boberg. 一种从偏好图中学习排序的高效算法。《机器学习》，75(1):129–165，2009。[32] Tapio Pahikkala、Willem Waegeman、Evgeni Tsivtsivadze、Tapio Salakoski 和 Bernard De Baets. 使用核方法学习非传递互反关系。《欧洲运筹学杂志》，206(3):676–685，2010。[33] Wojciech Rejchel. 关于排序和泛化界。《机器学习研究杂志》，13(46):1373–1392，2012。[34] Wojciech Rejchel. 排序与U过程的预言机不等式及Lasso 惩罚项。《神经计算》，239:214–222，2017。[35] Steve Smale 和 Ding-Xuan Zhou. 估计学习理论中的逼近误差。《分析与应用》，01(01):17–41，2003。[36] Ingo Steinwart 和 Andreas Christmann.《支持向量机》。Springer Science & Business Media，2008。[37] Ingo Steinwart 和 Simon Fischer. 深入考察高斯核的覆盖数界。《复杂性杂志》，62:101513，2021。[38] Charles J. Stone. 非参数回归的最优全局收敛速率。《统计年刊》，10(4):1040–1053，1982。[39] J.A.K. Suykens 和 J. Vandewalle. 最小二乘支持向量机分类器。《神经处理快报》，9(3):293–300，1999。[40] Taiji Suzuki. 深度ReLU网络在Besov空间和混合光滑Besov空间中学习的自适应性：最优速率与维数灾难。载于国际学习表示会议，2019。[41] Kazuma Tsuji 和 Taiji Suzuki. 变指数Besov空间回归问题中深度学习的估计误差分析。《电子统计学杂志》，15(1):1869–1908，2021。

[42] Willem Waegeman、Tapio Pahikkala、Antti Airola、Tapio Salakoski、Michiel Stock 和 Bernard De Baets. 一种从数据中学习分级关系的基于核的框架。IEEE《模糊系统汇刊》，20(6):1090–1101，2012。[43] Chendi Wang. 连续与离散数据的回归学习。博士论文，香港理工大学，2021。[44] Shuhua Wang 和 Baohuai Sheng. 强凸损失下核正则化成对学习的误差分析。已被《计算数学基础》接收。[45] S. K.M. Wong 和 Y. Y. Yao. 信息检索中的线性结构。载于第11届 ACM 信息检索研究与发展年度国际会议论文集，SIGIR '88，第219–232页，美国纽约州纽约，1988。计算机协会。[46] Dao-Hong Xiang 和 Ding-Xuan Zhou. 使用高斯函数和凸损失进行分类。《机器学习研究杂志》，10(49):1447–1468，2009。[47] Yun Yang 和 David B. Dunson. 贝叶斯流形回归。《统计年刊》，44(2):876–905，2016。[48] Gui-Bo Ye 和 Ding-Xuan Zhou. 黎曼流形上的高斯函数学习与逼近。《计算数学进展》，29(3):291–310，2008。[49] Gui-Bo Ye 和 Ding-Xuan Zhou. 黎曼流形上基于高斯函数的SVM学习与lp逼近。《分析与应用》，07(03):309–339，2009。[50] Yiming Ying 和 Ding-Xuan Zhou. 在线成对学习算法。《神经计算》，28(4):743–777，2016。[51] Tong Zhang. 基于凸风险最小化的分类方法的统计行为与一致性。《统计年刊》，32(1):56–85，2004。[52] Yulong Zhao、Jun Fan 和 Lei Shi. 正则化最小二乘排序算法的学习速率。《分析与应用》，15(06):815–836，2017。