arXiv:2303.05731v2 [cs.LG] 3 Apr 2023

张量分解实对数典范阈值的上界及其在贝叶斯推断中的应用

Naoki Yoshida 和 Sumio Watanabe 东京工业大学数学与计算科学系 日本东京目黑区大冈山 2-12-1，邮政信箱 W8-42，邮编 152-8552

摘要

张量分解现已用于数据分析、信息压缩和知识恢复。然而，由于张量分解属于奇异学习机，其数学性质尚未得到充分阐明。本文采用代数几何方法，给出张量分解实对数典范阈值（RLCT）的上界，并从理论上推导其贝叶斯泛化误差。此外，我们还通过数值实验探讨其数学性质。

1 引言

张量分解广泛应用于数据科学和机器学习 [1]。例如，它在信号处理中的贡献分析 [2], 、通过将张量数据转换为矩阵数据进行数据压缩 [3], ，以及通过从矩阵反向推导原始张量数据来实现数据恢复 [4]等方面发挥着核心作用。在许多情况下，众所周知，张量分解本身是 NP难的 [5]。因此，张量分解通常通过贝叶斯推断进行近似计算。然而，由于它属于奇异统计模型，其数学性质尚未得到完全阐明。本文推导其在贝叶斯推断中的泛化性能。

张量分解主要有两种类型：塔克分解和CP分解。对于三维张量 $X \in \mathbb{R}^{IJK}$，塔克分解定义为

$$X_{ijk} = \sum_{h_1=1}^{H_1} \sum_{h_2=1}^{H_2} \sum_{h_3=1}^{H_3} \Lambda_{h_1 h_2 h_3} A_{i h_1} B_{j h_2} C_{k h_3},$$

其中 $A \in \mathbb{R}^{NH_1}$、 $B \in \mathbb{R}^{NH_2}$、 $C \in \mathbb{R}^{NH_3}$ 和 $\Lambda \in \mathbb{R}^{H_1 H_2 H_3}$。 A、 B 和 C 称为因子矩阵， Λ 称为核心张量。CP分解是塔克分解的一种特殊情况，其定义如下：

$$X_{ijk} = \sum_{h=1}^{H} \Lambda_h A_{ih} B_{jh} C_{kh},$$

其中 $A \in \mathbb{R}^{IH}$, $B \in \mathbb{R}^{JH}$, $C \in \mathbb{R}^{KH}$, 并且 $\Lambda \in \mathbb{R}^{H}$。

我们研究一种张量分解统计模型

$$p(X|A,B,C) = \frac{1}{(2\pi)^{IJK/2}} \exp\left(-\frac{1}{2}\|X - A \circ B \circ C\|^2\right), \tag{1}$$

其中 $X \in \mathbb{R}^{I\times J\times K}$、 $A \in \mathbb{R}^{I\times H}$、 $B \in \mathbb{R}^{J\times H}$ 和 $C \in \mathbb{R}^{K\times H}$，而 $\|\ \|^2$ 是平方弗罗贝尼乌斯范数，即各分量的平方和。这里， $A \circ B \circ C$ 是一个 $I \times J \times K$ 张量，其 (i,j,k) 分量定义为

$$(A \circ B \circ C)_{ijk} = \sum_{h=1}^{H} A_{ih}B_{jh}C_{kh}.$$

注意，算子 $\circ$ 不满足结合律。模型 (1) 不可辨识且是奇异的，因为参数 (A, B, C) 到概率分布 $p(x|A, B, C)$ 的映射不是一一对应的。

在统计学习理论中，泛化性能是统计模型最重要的性质之一。设 $X^n = (X_1, X_2, ..., X_n)$ 是一组独立随机变量，其概率分布为 $q(x)$；对于给定的 X^n 、一个统计模型和一个先验分布， $p^*(x|X^n)$ 是 x 的贝叶斯预测分布。那么，泛化误差 G_n 定义为从 $q(x)$ 到 $p^*(x|X^n)$ 的 Kullback-Leibler散度,

$$G_n = KL(q(x)||p^*(x|X^n)).$$

如果模型是正则的，且数据生成分布可由统计模型实现，则 G_n 训练数据的期望值为

$$\mathbb{E}[G_n] = \frac{m}{2n} + o(\frac{1}{n}),$$

其中， m 是参数的数量 [6]。如果模型是奇异的，那么 Fisher 信息矩阵是退化的，因此我们无法使用赤池信息准则[7], 、贝叶斯信息准则[8], 或广义信息准则[9]等信息准则。然而，即使对于奇异模型， $\mathbb{E}[G_n]$ 的渐近行为也可通过奇点消解 [10,11]得到:

$$\mathbb{E}[G_n] = S + \frac{\lambda}{n} + o(\frac{1}{n}), \tag{2}$$

其中， $S = \int q(x) \log \frac{q(x)}{p(x|w_0)}$， $w_0 = \operatorname{argmin}_w KL(q(x)||p(x|w))$，它们是由模型的表达能力决定的常数值； λ 是实对数典范阈值（RLCT），其定义见下一节的定义 2.1。一般来说，RLCT 不大于参数数量的一半，这表明奇异模型的泛化误差小于正则模型。

RLCT 是由模型和先验分布决定的双有理不变量。对于若干统计模型和学习机，RLCT 或其上界已得到明确，例如三层神经网络[12],降秩回归和矩阵分解 [13], 正态混合模型[14],多项式混合模型[15], 以及潜在狄利克雷分配（LDA） [16], 隐马尔可夫模型[17], 和非负矩阵分解 [18]。然而，张量分解模型的 RLCT

仍未得到明确。本文从理论上推导张量分解模型 RLCT 的上界，并讨论其紧致性。

本文由四部分组成。第二节在主定理中给出了张量分解模型的实对数典范阈值上界。第三节证明主定理。第四节通过与数值实验结果进行比较，考察该上界的紧致性。

2 设置和主要结果

我们假设数据 X 是独立同分布的，其概率密度函数为 $q(x)$。设$X^n = \{X_i\}_{i=1}^n$、$\phi(x)$和 $p(x|w)$分别为训练数据、参数 w的先验分布以及统计模型或学习机。后验分布 $p(w|X^n)$定义为

$$p(w|X^n) = \frac{1}{Z_n}\phi(w)\prod_{i=1}^{n} p(X_i|w),$$

其中， Z_n 是归一化常数。贝叶斯预测分布 p^* 定义为

$$p^*(x|X^n) = \int p(x|w)p(w|X^n)dw.$$

接下来按如下方式定义实对数典范阈值。

定义 2.1（统计模型的实对数典范阈值）。设 $K(w)$为从 $q(x)$到 $p(x|w)$的Kullback-Leibler散度，即 $K(w) = \int q(x)\log(q(x)/p(x|w))dx$。那么，定义于 $\Re(z) > 0$的ζ函数

$$\zeta(z) = \int K(w)^z \phi(w)dw$$

可解析延拓为整个复平面上的唯一亚纯函数，其极点均为实数且为负。设 $-\lambda < 0$ 为ζ函数的最大极点，则 $\lambda > 0$ 是该模型的实对数典范阈值，记作 $\lambda(K, \phi)$。

张量分解模型表示为式 (1)。我们假设存在参数 $w_0 = (A_0, B_0, C_0)$，满足 $p(x|w_0) = q(x)$，其中 $A_0 \in \mathbb{R}^{IH_0}$、 $B_0 \in \mathbb{R}^{JH_0}$、 $C_0 \in \mathbb{R}^{KH_0}$和 $H_0 \leq H$。此外，假设 $\phi(A, B, C)$ 有一个 C^∞ 函数作为上界，且该函数的支撑集包含 w_0。在这些假设下，我们得到以下定理。

定理 2.2（主定理）。张量分解模型的实对数典范阈值满足以下不等式:

$$\lambda \leq \frac{H_0(I+J+K)-2}{2} + \min\{m_1, m_2, m_3\},$$

其中

$$\begin{aligned} m_1 &= F(IJ, K, H - H_0), \\ m_2 &= F(JK, I, H - H_0), \\ m_3 &= F(KI, J, H - H_0), \end{aligned}$$

and

$$F(N,M,H) = \begin{cases} 2(NM+MH+HN)-(N^2+M^2+H^2) & (|N-M| \le H \le N+M \; and \; H+M-N \; is \; even) \\ 2 & \\ \hline \\ \hline \\ (NM+MH+HN)- & +1 \; (N^2+M^2+H^2) \end{cases}$$

$(|N-M| \le H \le N+M \; and \; H+M-N \; is \; odd)$ $\frac{MH}{2}$(
$H \le N-M)$ $\frac{NH}{2}$$(H \le M-N)$ $\frac{NM}{2}$$(M+N \le H)$.

3 主定理的证明

3.1 预备知识

根据实对数典范阈值的基本性质，以下引理成立 [19]。

引理 3.1. 设 $K_1(w)$、 $K_2(w)$ 为非负解析函数， $\phi_1(w)$、$\phi_2(w)$ 为具有紧支集的非负 C^∞-函数。

(a)（不等式）若 $K_1(w) \le K_2(w)$ 且 $\phi_1(w) \ge \phi_2(w)$，则 $\lambda(K_1,\phi_1) \le \lambda(K_2,\phi_2)$。特别地，若 $\phi_1(w) = \phi_2(w)$ 且 $c_1K_1(w) \le K_2(w) \le c_2K_1(w)$，其中 c_1和 c_2 为正常数，则 $\lambda(K_1,\phi_1) = \lambda(K_2,\phi_2)$。

(b)（乘积）若 $w = (w_1,w_2)$、 $K(w) = K_1(w_1)K_2(w_2)$、 $\phi(w) = \phi_1(w_1)\phi_2(w_2)$，则 $\lambda(K,\phi) = \min\{\lambda(K_1,\phi_1)\ \lambda(K_2,\phi_2)\}$。

(c)（和）若 $w = (w_1,w_2)$， $K(w) = K_1(w_1) + K_2(w_2)$， $\phi(w) = \phi_1(w_1)\phi_2(w_2)$，则 $\lambda(K,\phi) = \lambda(K_1,\phi_1) + \lambda(K_2,\phi_2)$。

接下来，我们得到张量分解模型的 $K(w)$。

引理 3.2. 关于张量分解模型， $K(w)$ 表示如下:

$$K(w) = \|A \circ B \circ C - A_0 \circ B_0 \circ C_0\|^2.$$

证明。设 $T = A \circ B \circ C$, $T_0 = A_0 \circ B_0 \circ C_0$。则,

$$\begin{aligned} K(w) &= \int q(x) \log \frac{p(x|A_0,B_0,C_0)}{p(x|A,B,C)} dx \\ &= \int q(x) \left\{ \frac{1}{2} \left(\|x-T\|^2 - \|x-T_0\|^2 \right) \right\} dx \\ &= \frac{1}{2} \left(\|T\|^2 - \|T_0\|^2 \right) + \int \sum_{i,j,k} \{(T_0)_{ijk} - T_{ijk}\} x_{ijk} q(x) dx \\ &= \frac{1}{2} \left(\|T\|^2 - \|T_0\|^2 \right) + \sum_{i,j,k} \{(T_0)_{ijk} - T_{ijk}\} (T_0)_{ijk} \\ &= \frac{1}{2} \|T - T_0\|^2. \end{aligned}$$

使用引理 3.1(a)即可完成证明。 □

当 $c_1 f(w) \leq g(w) \leq c_2 f(w)$ 成立时，我们定义等价关系 $f(w) \sim g(w)$，其中 c_1 和 c_2 是正常数。根据以下不等式，我们可以将 λ 的上界分为两部分：

$$\begin{aligned}
&\|A \circ B \circ C - A_0 \circ B_0 \circ C_0\|^2 = \sum_{i,j,k} \left(\sum_{h=1}^{H} A_{ih} B_{jh} C_{kh} - \sum_{h=1}^{H_0} (A_0)_{ih} (B_0)_{jh} (C_0)_{kh} \right)^2 \\
\leq & 2 \sum_{i,j,k} \left(\sum_{h=1}^{H_0} (A_{ih} B_{jh} C_{kh} - (A_0)_{ih} (B_0)_{jh} (C_0)_{kh}) \right)^2 \\
& + 2 \sum_{i,j,k} \left(\sum_{h=H_0+1}^{H} A_{ih} B_{jh} C_{kh} \right)^2 .
\end{aligned}$$

不等式右侧的第一项和第二项中的变量彼此分离，因此可以应用引理 3.1(c)。设 λ_1, λ_2 分别为第一项和第二项的实对数典范阈值，则 $\lambda_1 + \lambda_2$ 是 λ的上界。我们将在第 3.2 小节中为求得 λ_1 的上界做准备，并在第 3.3 小节中为求得 λ_2 的上界做准备，最后在第 3.4 小节中完成主定理的证明。

3.2 为求得 λ_1的上界做准备

首先，我们来为求得 λ_1的上界做准备。根据以下两个引理，可以将 A_0, B_0, C_0 化为对角矩阵，从而简化问题。我们按原样记 $(A)_{ih}(1 \leq h \leq H_0)$、$(B)_{jh}(1 \leq h \leq H_0)$、$(C)_{kh}(1 \leq h \leq H_0)$ by A, B, C 。

引理 3.3. 设对 A_0、B_0 和 C_0 进行任意初等行变换后得到 $A_0' = {P_1}^{-1} A_0$、$B_0' = {P_2}^{-1} B_0$ 和 $C_0' = {P_3}^{-1} C_0$。λ_1 等于 $(\|A \circ B \circ C - A_0' \circ B_0' \circ C_0'\|^2, \phi(A, B, C))$ 的实对数典范阈值。

证明. 通过 $A' = {P_1}^{-1} A$ 作变量代换。于是，

$$\begin{aligned}
& \sum_{h=1}^{H_0} (A_{ih} B_{jh} C_{kh} - (A_0)_{ih} (B_0)_{jh} (C_0)_{kh}) \\
= & \sum_{h=1}^{H_0} \left(\left(\sum_l (P_1)_{il} A'_{lh} \right) B_{jh} C_{kh} - \left(\sum_l (P_1)_{il} (A'_0)_{lh} \right) (B_0)_{jh} (C_0)_{kh} \right) \\
= & \sum_l (P_1)_{il} \left(\sum_{h=1}^{H_0} (A'_{lh} B_{jh} C_{kh} - (A'_0)_{lh} (B_0)_{jh} (C_0)_{kh}) \right) .
\end{aligned}$$

令 $T_{ljk} = \sum_{h=1}^{H_0} (A'_{lh} B_{jh} C_{kh} - (A'_0)_{lh} (B_0)_{jh} (C_0)_{kh})$。由于 $(\sum_l (P_1)_{il} T_{ljk})^2$ 关于 $\{T_{ljk}\}_{l=1}^{I}$ 是二次的，因此存在一个正数 c，使得

$$\sum_{i,j,k} \left(\sum_l (P_1)_{il} T_{ljk} \right)^2 \leq c \|T\|^2. \tag{3}$$

在式 (3) 中，取 $\sum_m (P_1)_{lm} T_{mjk}$ 为 T_{ljk}，取 ${P_1}^{-1}$ 为 P_1。那么，左侧为

$$\begin{aligned}
&\sum_{i,j,k} \left(\sum_l ({P_1}^{-1})_{il} \sum_m (P_1)_{lm} T_{mjk} \right)^2 \\
=&\sum_{i,j,k} \left(\sum_m T_{mjk} \sum_l ({P_1}^{-1})_{il} (P_1)_{lm} \right)^2 \\
=&\sum_{i,j,k} \left(\sum_m T_{mjk} I_{im} \right)^2 \\
=&\sum_{i,j,k} T_{ijk}^2 \\
=&\|T\|^2,
\end{aligned}$$

其中 I 是一个单位矩阵。右侧为

$$\begin{aligned}
&c \sum_{l,j,k} \left(\sum_m (P_1)_{lm} T_{mjk} \right)^2 \\
=&c \sum_{i,j,k} \left(\sum_l (P_1)_{il} T_{ljk} \right)^2 .
\end{aligned}$$

因此，我们有 $\sum_{i,\ j,\ k} (\sum_l (P_1)_{il} T_{ljk})^2 \sim \|T\|^2$。由于 P_1 是正则的，通过 $A' = {P_1}^{-1} A$ 进行变量变换所得的雅可比行列式不会变为 0 或 ∞。因此，这不会影响 ζ函数的极点。根据引理 3.1(a)，λ_1 等于 $(\|A' \circ B \circ C - A'_0 \circ B_0 \circ C_0\|^2, \phi(A', B, C))$ 的实对数典范阈值。对 B 和 C 重复相同的操作，即完成证明。

□

引理 3.4。设对 A'_0、B'_0 和 C'_0 进行任意初等列变换后分别得到 $A''_0 = A'_0 {Q_1}^{-1}$、$B''_0 = B'_0 {Q_2}^{-1}$ 和 $C''_0 = C'_0 {Q_3}^{-1}$。λ_1 小于 $\left(\sum_{i,j,k} \sum_{n,m,l} (A_{in} B_{jm} C_{kl} - (A''_0)_{in} (B''_0)_{jm} (C''_0)_{kl})^2, \phi(A, B, C) \right)$ 的实对数典范阈值。

证明。通过 $A' = A {Q_1}^{-1}$, $B' = B {Q_2}^{-1}$, $C' = C {Q_3}^{-1}$ 进行变量替换。则

$$\begin{aligned}
&\left(\sum_{h=1}^{H_0} (A_{ih} B_{jh} C_{kh} - (A'_0)_{ih} (B'_0)_{jh} (C'_0)_{kh}) \right)^2 \\
=&\left(\sum_{h=1}^{H_0} \left(\sum_n A'_{in} (Q_1)_{nh} \sum_m B'_{jm} (Q_2)_{mh} \sum_l C'_{kl} (Q_3)_{lh} \right.\right. \\
&\left.\left. - \sum_n (A''_0)_{in} (Q_1)_{nh} \sum_m (B''_0)_{jm} (Q_2)_{mh} \sum_l (C''_0)_{kl} (Q_3)_{lh} \right) \right)^2 \\
=&\left(\sum_{n,m,l,h} (Q_1)_{nh} (Q_2)_{mh} (Q_3)_{lh} \left(A'_{in} B'_{jm} C'_{kl} - (A''_0)_{in} (B''_0)_{jm} (C''_0)_{kl} \right) \right)^2 .
\end{aligned}$$

令 $D_{nml} = \sum_h (Q_1)_{nh}(Q_2)_{mh}(Q_3)_{lh}$, $T_{(ijk)\,nml} = A'_{in}B'_{jm}C'_{kl} - (A''_0)_{in}(B''_0)_{jm}(C''_0)_{kl}$。则，存在一个满足下式的正常数 c

$$\left(\sum_{n,m,l} D_{nml}T^{(ijk)}_{nml}\right)^2 \le c\sum_{n,m,l}\left(T^{(ijk)}_{nml}\right)^2.$$

由于 Q_1、Q_2 和 Q_3 都是正则的，利用引理 3.1(a) 即可完成证明。 □

根据引理 3.3 和 3.4，λ_1 的一个上界是 $\left(\sum_{i,j,k}\sum_{n,m,l}(A_{in}B_{jm}C_{kl} - (A''_0)_{in}(B''_0)_{jm}(C''_0)_{kl})^2, \phi(A, B, C)\right)$ 的实对数典范阈值，其中 A''_0、B''_0 和 C''_0 分别是 A_0、B_0、C_0 的对角化。我们将这个实对数典范阈值记为 λ'_1。令 $K'(A,B,C) = \sum_{i,j,k}\sum_{n,m,l}(A_{in}B_{jm}C_{kl} - (A''_0)_{in}(B''_0)_{jm}(C''_0)_{kl})^2$。令 r_A、r_B、r_C 分别为 A_0、B_0、C_0 的秩。那么 A''_0、B''_0 和 C''_0 的非零元素分别只到 r_{Ath}、r_{Bth} 和 r_{Cth} 个分量为止。因此，我们令

$$\begin{aligned}
G_a &= \{(i,n) \mid 1 \le i = n \le r_A\}\\
G_b &= \{(j,m) \mid 1 \le j = m \le r_B\}\\
G_c &= \{(k,l) \mid 1 \le k = l \le r_C\}.
\end{aligned}$$

那么，

$$\begin{aligned}
&K'(A,B,C)\\
&= \sum_{(i,n)\notin G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{(i,n)\in G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{(i,n)\notin G_a}\sum_{(j,m)\in G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{(i,n)\notin G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\in G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{(i,n)\in G_a}\sum_{(j,m)\in G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{(i,n)\notin G_a}\sum_{(j,m)\in G_b}\sum_{(k,l)\in G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{(i,n)\in G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\in G_c}(A_{in}B_{jm}C_{kl})^2\\
&+ \sum_{i=1}^{r_A}\sum_{j=1}^{r_B}\sum_{k=1}^{r_C}(A_{ii}B_{jj}C_{kk} - (A''_0)_{ii}(B''_0)_{jj}(C''_0)_{kk})^2, \qquad (4)
\end{aligned}$$

由此得到以下引理:

引理 3.5. 设

$$
\begin{aligned}
&K_1'(A_1, B_1, C_1)\\
=&\sum_{(i,n)\notin G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\\
+&\sum_{(j,m)\notin G_b}\sum_{(k,l)\notin G_c}(B_{jm}C_{kl})^2+\sum_{(i,n)\notin G_a}\sum_{(k,l)\notin G_c}(A_{in}C_{kl})^2+\sum_{(i,n)\notin G_a}\sum_{(j,m)\notin G_b}(A_{in}B_{jm})^2\\
+&\sum_{(k,l)\notin G_c}(C_{kl})^2+\sum_{(i,n)\notin G_a}(A_{in})^2+\sum_{(j,m)\notin G_b}(B_{jm})^2,\\
&K_2'(A_2, B_2, C_2)\\
=&\sum_{i=1}^{r_A}\sum_{j=1}^{r_B}\sum_{k=1}^{r_C}(A_{ii}B_{jj}C_{kk}-(A_0'')_{ii}(B_0'')_{jj}(C_0'')_{kk})^2,
\end{aligned}
$$

and

$$\phi(A, B, C) = \phi_1(A_1, B_1, C_1)\phi_2(A_2, B_2, C_2),$$

其中， A_1、 B_1、 C_1 是 A、B、C 的分量，它们出现在 K_1' 的表述中； A_2、 B_2、 C_2 则属于 K_2'。于是， $\lambda_1' = \lambda_{1-1}' + \lambda_{1-2}'$，其中 λ_{1-1}' 是 $(K_1'(A_1, B_1, C_1)$ $\phi_1(A_1, B_1, C_1))$ 的实对数典范阈值， λ_{1-2}' 是 $(K_2'(A_2, B_2, C_2)$ $\phi_2(A_2, B_2, C_2))$ 的实对数典范阈值。

证明。 我们只需考虑 $K(w) = 0$ 的邻域，因为如果$K(w) > 0$，则不会影响ζ函数的极点。当方程 (4) 的最后一项接近 0 时， B_{jj} 且 C_{kk} 不为 0。因此我们有 $A_{ii} = \frac{(A_0'')_{ii}(B_0'')_{jj}(C_0'')_{kk}+\epsilon}{B_{jj}C_{kk}}$，其中 ϵ 是一个接近 0 的值 · 由于.....的支撑集 $\phi(A, B, C)$是紧的，因此存在一个正数 M，使得 $|A_{ii}|, |B_{jj}|, |C_{kk}| \leq M$。此外，$(A_0'')_{ii}$，$(B_0'')_{jj}$，$(C_0'')_{kk}$不为 0，因此 A_{ii}也不为 0。因此存在一个正数 m，使得 $m \leq |A_{ii}|$。$|B_{jj}|$ 和 $|C_{kk}|$ 也有正的下界和上界。于是，

$$
\begin{aligned}
&\sum_{(i,n)\in G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\sim\sum_{(j,m)\notin G_b}\sum_{(k,l)\notin G_c}(B_{jm}C_{kl})^2\\
&\sum_{(i,n)\notin G_a}\sum_{(j,m)\in G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\sim\sum_{(i,n)\notin G_a}\sum_{(k,l)\notin G_c}(A_{in}C_{kl})^2\\
&\sum_{(i,n)\notin G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\in G_c}(A_{in}B_{jm}C_{kl})^2\sim\sum_{(i,n)\notin G_a}\sum_{(j,m)\notin G_b}(A_{in}B_{jm})^2\\
&\sum_{(i,n)\in G_a}\sum_{(j,m)\in G_b}\sum_{(k,l)\notin G_c}(A_{in}B_{jm}C_{kl})^2\sim\sum_{(k,l)\notin G_c}(C_{kl})^2\\
&\sum_{(i,n)\notin G_a}\sum_{(j,m)\in G_b}\sum_{(k,l)\in G_c}(A_{in}B_{jm}C_{kl})^2\sim\sum_{(i,n)\notin G_a}(A_{in})^2\\
&\sum_{(i,n)\in G_a}\sum_{(j,m)\notin G_b}\sum_{(k,l)\in G_c}(A_{in}B_{jm}C_{kl})^2\sim\sum_{(j,m)\notin G_b}(B_{jm})^2.
\end{aligned}
$$

由引理 3.1(a)， $(K'(A, B, C)$ $\phi(A, B, C))$ 的实对数典范阈值与 $(K_1'(A_1, B_1, C_1)$ $+ K_2'(A_2, B_2, C_2)$ $\phi(A, B, C))$ 的相同。 $K_1'(A_1, B_1, C_1)$ 与 $K_2'(A_2, B_2, C_2)$ 的变量相互独立，因此由引理 3.1(c)即可完成证明。 □

下一个引理引自 [20]。

引理 3.6。设 x_i、y_j 是绝对值有限的变量（i、$j \in \mathbb{N}$）。设 a_i、b_j 为常数。设 $f_{ij} = x_i y_j - a_i b_j$。则，$\forall M$、$N \geq 2 \in \mathbb{N}$,

$$\sum_{i=1}^{M}\sum_{j=1}^{N} f_{ij}^2 \sim \sum_{i=2}^{M} f_{i1}^2 + \sum_{j=2}^{N} f_{1j}^2 + f_{11}^2.$$

利用上述引理，可以很容易地得到以下推论：

推论 3.1。设 x_i、y_j、z_k 是绝对值有限的变量（i、j、$k \in \mathbb{N}$）。设 a_i、b_j、c_k 为常数。设 $f_{ijk} = x_i y_j z_k - a_i b_j c_k$。则，$\forall I$、$J$、$K \geq 2 \in \mathbb{N}$,

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} f_{ijk}^2 \sim \sum_{i=2}^{I} f_{i11}^2 + \sum_{j=2}^{J} f_{1j1}^2 + \sum_{k=2}^{K} f_{11k}^2 + f_{111}^2.$$

证明。令 $y'_{(j-1)K+k} = y_j z_k$、$b'_{(j-1)K+k} = b_j c_k$ 和 $g_{il} = x_i y'_l - a_i b'_l$。则 $f_{ijk} = g_{il}$。在引理 3.6 中将 g_{il} 作为 f_{ij}，

$$\begin{aligned}\sum_{i=1}^{I}\sum_{l=1}^{JK} g_{il}^2 &\sim \sum_{i=2}^{I} g_{i1}^2 + \sum_{l=2}^{JK} g_{1l}^2 + g_{11}^2 \\ &= \sum_{i=2}^{I} f_{i11}^2 + \sum_{2\leq(j-1)K+k\leq JK} f_{1jk}^2 + f_{111}^2 \\ &= \sum_{i=2}^{I} f_{i11}^2 + \sum_{j=1}^{J}\sum_{k=1}^{K} f_{1jk}^2.\end{aligned}$$

$f_{1jk} = x_1 y_j z_k$，因此令 $x'_k = x_1 z_k$、$a'_k = a_1 c_k$、$h_{kj} = x'_k y_j - a'_k b_j$，可得 $f_{1jk} = h_{kj}$。将 h_{kj} 作为引理 3.6 中的 f_{ij}，

$$\sum_{k=1}^{K}\sum_{j=1}^{J} h_{kj}^2 \sim \sum_{k=2}^{K} h_{k1}^2 + \sum_{j=2}^{J} h_{1j}^2 + h_{11}^2$$

成立。因此，$\sum_{j=1}^{J}\sum_{k=1}^{K} f_{1jk}^2 \sim \sum_{j=2}^{J} f_{1j1}^2 + \sum_{k=2}^{K} f_{11k}^2 + f_{111}^2$，证明完毕。

□

3.3 为求 λ_2 的上界做准备

其次，我们通过以下两个引理，为求得 λ_2 的上界做准备。

引理 3.7。如果 ζ 函数中的积分区间变小，实对数典范阈值就会变大。

下一个引理由 [13]的结果得出。

引理 3.8. 设 $B \in \mathbb{R}^{NH}$和 $A \in \mathbb{R}^{HM}$。那么，$(\|BA\|^2, \phi(A,B))$ 的实对数典范阈值为

$$F(N,M,H) =$$
$$\begin{cases} \frac{2(NM+MH+HN)-(N^2+M^2+H^2)}{8}(|N-M| \le H \le N+M \, and \, H+M-N \, is \, even) \\ \frac{2(NM+MH+HN)-(N^2+M^2+H^2)+1}{8}(|N-M| \le H \le N+M \, and \, H+M-N \, is \, odd) \\ \frac{MH}{2}(H \le N-M) \\ \frac{NH}{2}(H \le M-N) \\ \frac{NM}{2}(M+N \le H), \end{cases}$$

其中 $\|\cdot\|$ 是弗罗贝尼乌斯范数。

3.4 主定理的证明

证明。我们分三部分证明主定理。

（步骤 1）计算 λ'_{1-1}

λ'_{1-1} 是通过以 [10, 11, 21]为中心对 K'_1 进行爆破得到的： $\{A_1 = B_1 = C_1 = 0\}$。由变量的对称性，我们只需考虑以下情形:

$$\begin{cases} a_{11}^{(1)} = \nu \\ a_{ij}^{(1)} = \nu a_{ij}^{(1)}((i,j) \neq (1,1)) \\ B_1 = \nu B_1, C_1 = \nu C_1, \end{cases}$$

其中 $a_{11}^{(1)}$ 是 A_1的一个分量。因此我们得到

$$K'_1(A_1,B_1,C_1) = \nu^2(1 + (positive \, variable)),$$

这称为正规交叉。该变换的雅可比行列式为

$$\nu^{(I+J+K)H-(r_A+r_B+r_C)-1},$$

因此，$(K'_1(A_1,B_1,C_1)\ \phi_1(A_1,B_1,C_1))$ 的实对数典范阈值为,

$$\lambda'_{1-1} = \frac{(I+J+K)H-(r_A+r_B+r_C)}{2}.$$

（步骤 2）计算 λ'_{1-2}

λ'_{1-2} 由推论 3.1 得到。考虑分别在推论 3.1 中使用 A_{ii}, B_{jj}, C_{kk},$(A''_0)_{ii}$, $(B''_0)_{jj}$, $(C''_0)_{kk}$ 作为 x_i, y_j, z_k, a_i, b_j, c_k。则

$$\begin{aligned} K'_2(A_2,B_2,C_2) \sim & \sum_{i=2}^{r_A}(A_{ii}B_{11}C_{11} - (A''_0)_{ii}(B''_0)_{11}(C''_0)_{11})^2 \\ & + \sum_{j=2}^{r_B}(A_{11}B_{jj}C_{11} - (A''_0)_{11}(B''_0)_{jj}(C''_0)_{11})^2 \\ & + \sum_{k=2}^{r_C}(A_{11}B_{11}C_{kk} - (A''_0)_{11}(B''_0)_{11}(C''_0)_{kk})^2 \\ & + (A_{11}B_{11}C_{11} - (A''_0)_{11}(B''_0)_{11}(C''_0)_{11})^2 \end{aligned}$$

成立。这里，考虑以下变换:

$$\begin{cases} A'_{ijk} = A_{ii}B_{jj}C_{kk} - (A''_0)_{ii}(B''_0)_{jj}(C''_0)_{kk} \\ B'_j = B_{jj} \\ C'_k = C_{kk}. \end{cases}$$

Then

$$K'_2(A_2, B_2, C_2) \sim \sum_{i=2}^{r_A}(A'_{i11})^2 + \sum_{j=2}^{r_B}(A'_{1j1})^2 + \sum_{k=2}^{r_C}(A'_{11k})^2 + (A'_{111})^2 \qquad (5)$$

成立。该变换的雅可比行列式为

$$\left|\frac{\partial(A'_{ijk}, B'_j, C'_k)}{\partial(A_{ii}, B_{jj}, C_{kk})}\right| = \begin{vmatrix} B_{jj}C_{kk} & A_{ii}C_{kk} & A_{ii}B_{jj} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = B_{jj}C_{kk} \neq 0, \infty,$$

因此，它不会影响 ζ函数的极点。由于式（5）右侧各项中的变量彼此独立，且先验分布 $\phi(A, B, C)$ 也不会影响极点，因此我们可以使用引理 3.1(c)。 $K(w) = w^2$ 的实对数典范阈值为 $1/2$，因此得到

$$\lambda'_{1-2} = \frac{r_A + r_B + r_C - 2}{2}.$$

（步骤 3）计算 λ_2 的上界 根据引理 3.7, $\left(\int \left(\sum_{i',j,k}\left(\sum_{h=H_0+1}^{H} A_{ih}B_{jh}C_{kh}\right)^2\right) dw, \phi(A, B, C)\right)$ 的实对数典范阈值小于 $\left(\int_{B's\ components \neq 0}\left(\sum_{i',j,k}\left(\sum_{h=H_0+1}^{H} A_{ih}B_{jh}C_{kh}\right)^2\right) dw, \phi(A, B, C)\right)$ 的实对数典范阈值。在所有 B 的分量均不为 0 的条件下，考虑以下变换:

$$\begin{cases} A'_{i(J-1)+j,h} = A_{ih}B_{jh} \\ B'_{jh} = B_{jh} \\ C'_{kh} = C_{kh}. \end{cases}$$

该变换的雅可比行列式为

$$\left|\frac{\partial(A'_{i(J-1)+j,h}, B'_{jh}, C'_{kh})}{\partial(A_{ih}, B_{jh}, C_{kh})}\right| = \begin{vmatrix} B_{jh} & A_{ih} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = B_{jh} \neq 0,$$

因此，这对 ζ函数的极点没有影响。我们来计算 $\left(\sum_{i=1}^{IJ}\sum_{k=1}^{K}\left(\sum_{h=H_0+1}^{H} A'_{ih}C'_{kh}\right)^2 = \|A'C'^T\|, \phi(A, B, C)\right)$ 的实对数典范阈值，其中 T 表示转置。根据引理 3.8，它是 $F(IJ, K, H - H_0)$。值得注意的是，即使 $H - H_0 = 0$，结论仍然成立，因为 $F(N, M, 0) = 0$。根据对称性，$F(JK, I, H - H_0)$ 和 $F(KI, J, H - H_0)$ 也都是 λ_2 的上界。因此， λ_2 的上界为

$$\min\{F(IJ, K, H - H_0), F(JK, I, H - H_0), F(KI, J, H - H_0)\}.$$

设 $m_1 = F(IJ, K, H - H_0)$、 $m_2 = F(JK, I, H - H_0)$ 和 $m_3 = F(KI, J, H - H_0)$。根据（步骤 1）至（步骤 3）， λ 的上界为

$$\begin{aligned}\lambda \leq \lambda_1 + \lambda_2 \leq & \lambda'_{1-1} + \lambda'_{1-2} + \min\{m_1, m_2, m_3\} \\ = & \frac{H_0(I + J + K) - 2}{2} + \min\{m_1, m_2, m_3\}.\end{aligned}$$

□

4 讨论

4.1 数值实验与界的紧致性

我们来考察张量分解的实对数典范阈值的精确值，并将其与主定理中的界进行比较。我们通过实验计算 $\mathbb{E}[G_n]$ 的近似值。具体而言，

$$\begin{aligned}&\mathbb{E}[G_n] \\ =&\mathbb{E}_{X\sim p(X|A_0,B_0,C_0)}\left[\mathbb{E}_{x\sim p(x|A_0,B_0,C_0)}\left[\log\frac{p(x|A_0,B_0,C_0)}{\mathbb{E}_{(A,B,C)\sim p(A,B,C|X^n)}\left[p(x|A,B,C)\right]}\right]\right]\end{aligned}$$

成立，并且第一和第二个期望可以分别通过从 $\mathcal{N}((A_0 \circ B_0 \circ C_0)_{ijk}, 1)$ 中抽样 X 和 x 的分量来近似计算，其中 $\mathcal{N}$ 表示正态分布；第三个期望则通过从后验分布中抽样 (A, B, C) 来近似计算。后验分布通过马尔可夫链蒙特卡罗方法生成。我们为第一个期望抽取10个数据，为第二个抽取10000个，为第三个抽取1000个。利用以下推论可以得到 λ 的近似值:

推论4.1. 若存在一个满足以下条件的参数 w_0

$$q(x) = p(x|w_0),$$

则式(2)中的 $S = 0$ ，即,

$$\mathbb{E}[G_n] = \frac{\lambda}{n} + o(\frac{1}{n}).$$

利用推论4.1，我们设定 $n = 100$ ，并通过$n\mathbb{E}[G_n]$近似计算 λ 。我们在条件 $I = J = K =$ 2、3、4和 $H =$2$H_0 =$ 2、4、6、8、10下进行实验。对于每种情况，我们随机改变真实参数（A_0、 B_0、 C_0）5次，并计算误差。

结果如表1所示。将上界记为 λ_B。当$I = J = K = 2$ 或 $I = J = K = 3$时，当 H 较大时，我们的上界 λ_B 非常宽松：精确值 λ 小得出乎意料。这个结果可以用ζ函数的极点数量来解释。随着 H 增大，满足 $A\circ B\circ C = A_0\circ B_0\circ C_0$ 的组合（A、 B、 C）的数量也会增加，这意味着极点数量增加。这些因素导致实对数典范阈值下降。此外，当$I = J = K$ 增大时，这种现象很难发生。这仅仅是因为当张量 $A\circ B\circ C$的分量很多时， $A \circ B \circ C = A_0 \circ B_0 \circ C_0$ 难以满足。

表1：数值计算值与理论上界的比较（$H = 2H_0$）。

		$H_0 = 1$	$H_0 = 2$	$H_0 = 3$	$H_0 = 4$	$H_0 = 5$
I=J=K=2	λ_B	3.00	7.00	11.00	14.50	18.00
	λ	2.68 ± 0.57	4.31 ± 0.58	3.82 ± 0.84	3.76 ± 0.39	4.06 ± 0.59
I=J=K=3	λ_B	5.00	11.00	17.00	23.00	29.00
	λ	5.00 ± 1.54	9.73 ± 1.00	13.49 ± 2.20	12.06 ± 0.75	12.68 ± 2.14
I=J=K=4	λ_B	7.00	15.00	23.00	31.00	39.00
	λ	6.50 ± 1.11	13.66 ± 0.81	20.13 ± 1.98	29.43 ± 2.25	32.66 ± 2.08

4.2 未来研究

这项研究未来的一个方向是寻找更紧的上界。从数值实验中我们发现，上界还不够紧，尤其是在 H 较大且ζ函数中有许多极点时。我们对 λ_1 的上界为 $H_0 (I+J+K)_2^{-2}$，而显然的上界为$H_0(I+J+K)2$。

另一个研究方向是推广这些假设。我们在第2节中为主定理作出了许多假设，因此考虑更一般的条件也很重要。

5 结论

本文推导了张量分解模型的实对数典范阈值上界，该上界体现了贝叶斯泛化误差，并通过与数值实验进行比较，考察了其紧致性。

致谢

本研究部分受日本学术振兴会科学研究费助成事业科学研究（C）资助，项目编号为 21K12025。

参考文献

[1] T.G.Kolda, B.W.Bader, "张量分解及其应用", SIAM评论, vol.51, No.3, pp.455-500, 2009, doi.org/10.1137/07070111X.[2] L.D.Lathauwer, J.Vandewalle, "高阶信号处理中的降维与多线性代数中的秩-(R1,R2,...,RN)约简", 线性代数及其应用, vol.391, pp.31-55, 2004, doi.org/10.1016/j.laa.2004.01.016.[3] A.Karami, M.Yazdi, G.Mercier, "使用离散小波变换和塔克分解压缩高光谱图像", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol.5, issue.2, pp.444-450, 2012.

[4] X.Chen, Z.He, J.Wang, ”通过结合奇异值分解的张量分解发现时空交通速度模式并恢复不完整数据”, Transportion Research Part C: Emerging Technologies, vol.86, pp.59-77, 2018, doi.org/10.1016/j.trc.2017.10.023.[5] C.J.Hillar, L.-H.Lim, ”大多数张量问题都是NP难的”, Journal of the ACM, vol.60, issue.6, no.45, pp.1-39, 2013, doi.org/10.1145/2512329.[6] S.Watanabe, ”非正则学习机的代数分析”, Advances in Neural I nformation Proceeding Systems, no.12, pp.356-362, 2000, 丹佛, 美国.[7] H.Akaike, ”统计模型识别的新视角”, IEEE Trans.Autom.Control, vol.19, no.6, pp.716-723, 1974.[8] Schwarz, Gideon E., ”估计模型的维数”, Annals of Statistics, 6 (2): pp.461-464, 1978.[9] K.Takeuchi, ”信息统计量的分布与最优模型的判据”, Mathematical Science, no.153, pp.12-18, 1976.[10] S.Watanabe, ”奇异统计估计的代数分析”, Lecture Notes on Computer Science, vol.1720, pp.39-50, 1999.[11] S.Watanabe, ”分层学习机的代数几何方法”, 神经网络, Vol.14, issue.8, pp.1049-1060, 2001.[12] M.Aoyagi, S.Watanabe, ”奇点消解与分层神经网络贝叶斯估计的泛化误差”, IEICE Trans, pp.2112-2124, 2006.[13] M.Aoyagi, S.Watanabe, ”贝叶斯估计中降秩回归的随机复杂度”, 神经网络, vol.18, pp.924-933, 2005.[14]K.Yamazaki, S.Watanabe, ”混合模型中的奇点与随机复杂度上界”, 神经网络, vol.16, pp.1029-1038, 2003, doi.org/10.1016/S0893-6080(03)00005-4.[15] T.Watanabe, S.Watanabe, ”多项式混合模型中贝叶斯泛化误差的渐近行为”, IEICE Technical Report, vol.119, no.360, pp.1-8, 2020.[16] N.Hayashi, ”潜在狄利克雷分配中贝叶斯泛化误差的精确渐近形式”, 神经网络, vol.137, pp.127-137, 2021, doi.org/10.1016/j.neunet.2021.01.024.[17]K.Yamazaki, ”隐马尔可夫模型的代数几何与随机复杂度”, Neurocomputing, vol.69, pp.62-84, 2005.[18] N.Hayashi, S.Watanabe, ”非负矩阵分解实对数典范阈值的更紧上界及其在贝叶斯推断中的应用”, IEEE SSCI, pp.1-8, 2017, 檀香山, 夏威夷, 美国, doi:10.1109/SSCI.2017.8280811.

[19] S.Watanabe.《代数几何与统计学习理论》，剑桥大学出版社，2009。
[20] N.Hayashi, S.Watanabe，《非负矩阵分解中贝叶斯泛化误差的上界》，Neurocomputing，第266C卷，第29期，第21-28页，2017，doi: 10.1016/j.neucom.2017.04.068。[21] S.Watanabe，《不可识别学习机的代数分析》，Neural Computation，第13卷，第4期，第899-933页，2001。