Supplementary Material:
Convergence rate of Bayesian tensor estimator and its minimax optimality

A. Proofs of Lemma 1 and Theorem 1

Let $M_{r,d}$ be a positive number that will be determined later on. Let $\mathcal{F}_{r,d}$ be

$$\mathcal{F}_{r,d} = \{A = [[U^{(1)}, \dots, U^{(K)}]] \mid \max_{i,k} \|U^{(k)}_{:,i}\|_2 \leq \sqrt{M_{r,d}},\ U^{(k)} \in \mathbb{R}^{d \times M_k}\},$$

and $\mathcal{F}_r$ be $\cup_{d=1}^{M} \mathcal{F}_{r,d}$.

$$\|A^*\|_{\max,2} := \min_{\{U^{(k)}\}} \{\max_{i,k} \|U^{(k)}_{:,i}\| \mid A^* = [[U^{(1)}, \dots, U^{(K)}]],\ U^{(k)} \in \mathbb{R}^{d \times M_k}\}.$$

σ_{p,d^*} is denoted by $\sigma_{\mathrm{p}}/\sqrt{d^*}$.

A.1. Proof of Lemma 1

For d^* dimensional vectors $u^{(k)}$ $(k = 1, \dots, K)$, let

$$\langle u^{(1)}, \dots, u^{(K)} \rangle := \sum_{i=1}^{d'} \prod_{k=1}^{K} u^{(k)}_i.$$

Then for $u^{(k)},\ u^{*(k)}$ $(k = 1, \dots, K)$, we have that

$$\begin{aligned} &|\langle u^{(1)}, \dots, u^{(K)} \rangle - \langle u^{*(1)}, \dots, u^{*(K)} \rangle| \\ =&|\sum_{k=1}^{K} \langle u^{*(1)}, \dots, u^{*(k-1)}, u^{(k)} - u^{*(k)}, u^{(k+1)}, \dots, u^{(K)} \rangle| \\ \leq& \sum_{k=1}^{K} \|u^{*(1)}\| \cdots \|u^{*(k-1)}\| \|u^{(k)} - u^{*(k)}\| \|u^{(k+1)}\| \cdots \|u^{(K)}\|. \end{aligned} \tag{7}$$

Therefore, if $\|u^{(k)}\| \leq \|A^*\|_{\max,2} + \sigma_{\mathrm{p}}$, $\|u^{(k)} - u^{*(k)}\| \leq \frac{\epsilon r}{K(\|A^*\|_{\max,2}+\sigma_{\mathrm{p}})^{K-1}}$ and $\|u^{*(k)}\| \leq \|A^*\|_{\max,2}$, then we have

$$|\langle u^{(1)}, \dots, u^{(K)} \rangle - \langle u^{*(1)}, \dots, u^{*(K)} \rangle| \leq \epsilon r.$$

Now, we consider two situations: (i) $\frac{\epsilon r}{K(\|A^*\|_{\max,2}+\sigma_{\mathrm{p}})^{K-1}} \leq \sqrt{d^*}\sigma_{\mathrm{p},d^*}$ and (ii) $\frac{\epsilon r}{K(\|A^*\|_{\max,2}+\sigma_{\mathrm{p}})^{K-1}} \geq \sqrt{d^*}\sigma_{\mathrm{p},d^*}$. (i) If $\frac{\epsilon r}{K(\|A^*\|_{\max,2}+\sigma_{\mathrm{p}})^{K-1}} \leq \sqrt{d^*}\sigma_{\mathrm{p},d^*} (= \sigma_{\mathrm{p}})$ and $\|u^{*(k)}\| \leq \|A^*\|_{\max,2}$, then $\|u^{(k)}\| \leq \|A^*\|_{\max,2} + \sigma_{\mathrm{p}}$. Hence, by Lemma 2, we have that

$$\begin{aligned} &-\log(\Pi(A : \|A - A^*\|_n < \epsilon r | d^*)) \leq -\log(\Pi(A : \|A - A^*\|_\infty < \epsilon r | d^*)) \\ \leq& -\sum_{i_1,\dots,i_K}^{M_1,\dots,M_K} \log\left(\Pi(U^{(k)}_{:,i_k} : \|U^{(k)}_{:,i_k} - U^{*(k)}_{:,i_k}\| \leq \tfrac{\epsilon r}{K(\|A^*\|_{\max,2}+\sigma_{\mathrm{p}})^{K-1}})\right) \\ \leq& \sum_{k=1}^{K} \sum_{i_k=1}^{M_k} \left\{ -d^* \log\left[\frac{\epsilon r}{6\sigma_{\mathrm{p},d^*}\sqrt{d^*}K(\|A^*\|_{\max,2} + \sigma_{\mathrm{p}})^{K-1}}\right] + \frac{\|U^{*(k)}_{:,i_k}\|^2}{2\sigma^2_{\mathrm{p},d^*}} \right\} \\ \leq& -d^*(\sum_{k=1}^{K} M_k) \log\left[\frac{\epsilon r}{6\sigma_{\mathrm{p},d^*}\sqrt{d^*}K(\|A^*\|_{\max,2} + \sigma_{\mathrm{p}})^{K-1}}\right] + \frac{\sum_{k=1}^{K} \|U^{*(k)}\|_F^2}{2\sigma^2_{\mathrm{p},d^*}}. \end{aligned}$$

(ii) On the other hand, if $\frac{\epsilon r}{K(\|A^*\|_{\max,2}+\sigma_{\mathrm{p}})^{K-1}} > \sigma_{\mathrm{p},d^*}\sqrt{d^*}$, we have that

$$-\log(\Pi(A : \|A - A^*\|_n < \epsilon r | d^*)) \leq -d^*(\sum_{k=1}^{K} M_k) \log\left(\frac{\sigma_{\mathrm{p},d^*}\sqrt{d^*}}{6\sigma_{\mathrm{p},d^*}\sqrt{d^*}}\right) + \frac{\sum_{k=1}^{K} \|U^{*(k)}\|_F^2}{2\sigma^2_{\mathrm{p},d^*}}.$$

Combining these inequality, we have that

$$
\begin{aligned}
&-\log(\Pi(A:\|A-A^*\|_n<\epsilon r)) \le -\log(\Pi(A:\|A-A^*\|_n<\epsilon r|d^*)) - \log(\pi(d^*)) \\
\le& -d^*\left(\sum_{k=1}^K M_k\right)\log\left[\frac{1}{6}\left(\frac{\epsilon r}{\sigma_{\rm p}K(\|A^*\|_{\max,2}+\sigma_{\rm p})^{K-1}}\wedge 1\right)\right] + \frac{\sum_{k=1}^K \|U^{*(k)}\|_F^2}{2\sigma_{{\rm p},d^*}^2} - \log(\pi(d^*)) \\
\le& d^*\left(\sum_{k=1}^K M_k\right)\log\left[\frac{6}{\xi}\left(\frac{\sigma_{\rm p}^K K(\frac{\|A^*\|_{\max,2}}{\sigma_{\rm p}}+1)^{K-1}}{\epsilon r}\vee 1\right)\right] + \frac{\sum_{k=1}^K \|U^{*(k)}\|_F^2}{2\sigma_{{\rm p},d^*}^2}. \qquad (8)
\end{aligned}
$$

The last line is by the definition of the unnormalized prior $\pi(d) = \xi^{d(\sum_k M_k)}$. Later on we let $\epsilon = 1/\sqrt{n}$. This gives Lemma 1.

A.2. Proof of Theorem 1

In this section, we fix $X_{1:n}$ and suppose it is not a random variable.

To prove the theorem, we utilize the technique developed by (van der Vaart & van Zanten, 2011). Their technique is originally developed to show the posterior convergence of Gaussian process regression and is based on theories by (Ghosal et al., 2000) for the posterior convergence of non-parametric Bayes models. Although our situation is of parametric model, their technique is useful because ours is high dimensional singular model in which a standard asymptotic statistics for parametric models does not work.

For a set of tensors $\mathcal{F}_r$, an event $\mathcal{A}_r$ and a test ϕ_r (all of which are dependent on a positive real number $r > 0$), it holds that, for $\epsilon > 0$,

$$
\begin{aligned}
&{\rm E}\left[\int \|A-A^*\|_n^2 \Pi({\rm d}A|Y_{1:n})\right] \\
=&{\rm E}\left[32\epsilon^2\int_{r>0} r\Pi(\|A-A^*\|_n \ge 4\epsilon r|Y_{1:n}){\rm d}r\right] \\
\le& 32\epsilon^2\int_{r>0} r^K\big\{{\rm E}[\phi_r] + P(\mathcal{A}_r^c) \\
&\quad + {\rm E}[(1-\phi_r)\mathbf{1}_{\mathcal{A}_r}\Pi(A\in\mathcal{F}_r^c|Y_{1:n})] \\
&\quad + {\rm E}[(1-\phi_r)\mathbf{1}_{\mathcal{A}_r}\Pi(A\in\mathcal{F}_r:\|A-A^*\|_n^2\ge 4\epsilon r^{2K}|Y_{1:n})]\big\}{\rm d}r^K \\
=:& 32\epsilon^2\int_{r>0} r^K(A_r+B_r+C_r+D_r){\rm d}r^K. \qquad (9)
\end{aligned}
$$

We give an upper bound of A_r, B_r, C_r and D_r in the following.

Step 1. The probability distribution of $Y_{1:n}$ with a true tensor A (that means $Y_i = \langle X_i, A\rangle + \epsilon_i$) is denoted by $P_{n,A}$. The expectation of a function f with respect to $P_{n,A}$ is denoted by $P_{n,A}(f)$.

For arbitrary $r' > 0$, define $C_{j,r',d} = \{A \in \mathcal{F}_{r,d} \mid jr' \le \sqrt{n}\|A-A^*\|_n \le (j+1)r'\}$. We construct a maximum cardinality set $\Theta_{j,r',d} \subset C_{j,r',d}$ such that each $A, A' \in \Theta_{j,r',d}$ satisfies $\sqrt{n}\|A-A'\|_n \ge jr'/2$. The cardinality of $\Theta_{j,r',d}$ is equal to $D(jr'/2, C_{j,r',d}, \sqrt{n}\|\cdot\|_n)$ [3]. Then, one can construct a test $\phi_{j,d}$ such that

$$
\begin{aligned}
&P_{n,A^*}\phi_{j,d} \le D(jr'/2, C_{j,r,d}, \sqrt{n}\|\cdot\|_n)e^{-\frac{1}{2}(\frac{jr'}{2}+q)^2} \le D(jr'/2, C_{j,r,d}, \sqrt{n}\|\cdot\|_n)e^{-\frac{1}{8}j^2r'^2-\frac{1}{2}q^2}, \\
&\sup_{A\in C_{j,r',d}} P_{n,A}(1-\phi_{j,d}) \le e^{-\frac{1}{2}\max(\frac{jr'}{2}-q,0)^2} \le e^{-\frac{j^2r'^2}{16}+\frac{q^2}{2}},
\end{aligned}
$$

for any $q > 0$ (see (van der Vaart & van Zanten, 2011) for the details). For each d, we construct a test ϕ_d as $\phi_d = \max_j \phi_{j,d}$. Then we have

$$
P_{n,A^*}\phi_d \le \sum_{j\ge1} D(jr'/2, C_{j,r',d}, \sqrt{n}\|\cdot\|_n)e^{-\frac{1}{8}j^2r'^2-\frac{1}{2}q^2} \le \sum_{j\ge1} D(r'/2, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n)e^{-\frac{1}{8}j^2r'^2-\frac{1}{2}q^2}
$$

[3] For a normed space $\mathcal{F}$ attached with a norm $\|\cdot\|$, the ϵ-packing number is denoted by $D(\epsilon, \mathcal{F}, \|\cdot\|)$.

$$\leq 9D(r'/2, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n)e^{-\frac{1}{8}r'^2 - \frac{1}{2}q^2}.$$

Here, by setting $\frac{1}{2}q^2 = \log(D(r'/2, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))$, we have

$$P_{n,A^*}\phi_d \leq 9e^{-\frac{1}{8}r'^2},$$
$$\sup_{A \in \mathcal{F}_{r,d}} P_{n,A}(1-\phi_d) \leq e^{-\frac{1}{16}r'^2 + \log(D(r'/2, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))}.$$

Finally, we construct a test ϕ as the maximum of ϕ_d, that is, $\phi = \max_{d \geq 1} \phi_d$. Then, we have

$$P_{n,A^*}\phi \leq 9e^{-\frac{1}{8}r'^2 + \log(d_{\max})}$$
$$\sup_{A \in \mathcal{F}_{r,d}} P_{n,A}(1-\phi) \leq e^{-\frac{1}{16}r'^2 + \log(D(r'/2, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))},$$

for all $d \geq 1$.

Substituting $4\sqrt{n}\epsilon r^K$ into r', we obtain

$$P_{n,A^*}\phi \leq 9e^{-2n\epsilon^2 r^{2K} + \log(d_{\max})} \tag{10}$$
$$\sup_{A \in \mathcal{F}_{r,d}} P_{n,A}(1-\phi) \leq e^{-n\epsilon^2 r^{2K} + \log(D(r'/2, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))}. \tag{11}$$

We define

$$A_r = P_{n,A^*}\phi.$$

From now on, we denote by ϕ_r the test constructed above to indicate that the test is associated to a specific r.

Step 2.

By Lemma 14 of (van der Vaart & van Zanten, 2011) and its proof, one can show that, for any $r > 0$,

$$P\left[\int \frac{p_{n,A}}{p_{n,A^*}} \mathrm{d}\Pi(A) \geq e^{-\frac{n\epsilon^2 r^2}{2} - \sqrt{n}\epsilon r x}\Pi(A : \|A - A^*\|_n < \epsilon r)\right] \leq e^{-\frac{x^2}{2}}.$$

Therefore, there exists an even $\mathcal{A}_{1,r}$ such that

$$P_{A^*}(\mathcal{A}_{1,r}^c) \leq e^{-n\epsilon^2 r^2/8},$$

and, on the event $\mathcal{A}_{1,r}$, it holds that

$$\int \frac{p_{n,A}}{p_{n,A^*}} \mathrm{d}\Pi(A) \geq e^{-n\epsilon^2 r^2}\Pi(A : \|A - A^*\|_n < \epsilon r).$$

Moreover, it can be checked in a similar way that there exists an event $\mathcal{A}_{2,r}$ such that, for a some fixed constant $c_\xi < \frac{1}{2}$ (which will be determined later),

$$P_{A^*}(\mathcal{A}_{2,r}^c) \leq \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{c_\xi}\sqrt{n}\epsilon r^K}{2} + \frac{\log(\Pi(A : \|A - A^*\|_n < \sqrt{c_\xi}\epsilon r^K))}{\sqrt{c_\xi}\sqrt{n}\epsilon r^K} \vee 0\right)^2\right\}, \tag{12}$$

and, on the event $\mathcal{A}_{2,r}$, it holds that

$$\int \frac{p_{n,A}}{p_{n,A^*}} \mathrm{d}\Pi(A) \geq e^{-c_\xi n\epsilon^2 r^{2K}}.$$

Here, we note that, by a simple calculation, the RHS of Eq. (12) is bounded by

$$e^{-\frac{1}{16}c_\xi n\epsilon^2 r^{2K}},$$

if $c_\xi n\epsilon^2 r^{2K} \geq -8\log(\Pi(A : \|A - A^*\|_n < \sqrt{c_\xi}\epsilon r^K))$.

Now, define $\mathcal{A}_r = \mathcal{A}_{1,r} \cap \mathcal{A}_{2,r}$, then we have

$$B_r = P_{A^*}(\mathcal{A}_r^c) \leq e^{-n\epsilon^2 r^2/8} + \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{c_\xi}\sqrt{n}\epsilon r^K}{2} + \frac{\log(\Pi(A : \|A - A^*\|_n < \sqrt{c_\xi}\epsilon r^K))}{\sqrt{c_\xi}\sqrt{n}\epsilon r^K} \vee 0\right)^2\right\}.$$

Step 3.

Since

$$\begin{aligned}
&\{\{U^{(k)}\}_{k=1}^K \in \mathbb{R}^{d\times M_1} \times \cdots \times \mathbb{R}^{d\times M_K} \mid \max_{k,i} \|U^{(k)}_{:,i}\|_2 \geq \sqrt{M_{r,d}}\} \\
\subseteq&\{\{U^{(k)}\}_{k=1}^K \in \mathbb{R}^{d\times M_1} \times \cdots \times \mathbb{R}^{d\times M_K} \mid \sum_{k=1}^K \|U^{(k)}\|_2^2 \geq M_{r,d}\},
\end{aligned}$$

Proposition 1 yields the bound of the prior probability of $\mathcal{F}_r^c$ as

$$\Pi(\mathcal{F}_r^c) = \sum_{d=1}^M \Pi(\mathcal{F}_{r,d}^c)\pi(d) \leq \sum_{d=1}^M \exp\left[\frac{d\sum_k M_k}{2} + \frac{d\sum_k M_k}{2}\log\left(\frac{M_{r,d}}{d\sum_k M_k \sigma_{\mathrm{p},d}^2}\right) - \frac{M_{r,d}}{2\sigma_{\mathrm{p},d}^2}\right]\pi(d).$$

Therefore, its posterior probability in the event $\mathcal{A}_r$ is bounded as

$$\begin{aligned}
C_r &= P_{n,A^*}[\Pi(\mathcal{F}_r^c|Y_{1:n})\mathbf{1}_{\mathcal{A}_r}(1-\phi_r)] \\
&\leq \sum_{d=1}^M \exp\left[n\epsilon^2 r^2 + \Xi(\epsilon r) + \frac{d\sum_k M_k}{2} + \frac{d\sum_k M_k}{2}\log\left(\frac{M_{r,d}}{d\sum_k M_k\sigma_{\mathrm{p},d}^2}\right) - \frac{M_{r,d}}{2\sigma_{\mathrm{p},d}^2}\right]\pi(d).
\end{aligned}$$

Step 4.

Here, D_r is evaluated. Remind that D_r is defined as

$$D_r = P_{n,A^*}[\Pi(A \in \mathcal{F}_r : \|A - A^*\|_n > 4\epsilon r^K|Y_{1:n})(1-\phi_r)\mathbf{1}_{\mathcal{A}_r}].$$

Since $\mathcal{A}_r = \mathcal{A}_{1,r} \cap \mathcal{A}_{2,r} \subseteq \mathcal{A}_{2,r}$, we have

$$\begin{aligned}
D_r &\leq P_{n,A^*}\left[\sum_d \int_{A\in\mathcal{F}_{r,d}:\|A-A^*\|_n > 4\epsilon r^K} p_{n,A}/p_{n,A^*}\mathrm{d}\Pi(A|d)\exp(c_\xi n\epsilon^2 r^{2K})\pi(d)(1-\phi_r)\mathbf{1}_{\mathcal{A}_r}\right] \\
&= \sum_d \int_{A\in\mathcal{F}_{r,d}:\|A-A^*\|_n>4\epsilon r^K} P_{n,A}[(1-\phi_r)\mathbf{1}_{\mathcal{A}_r}]\exp(c_\xi n\epsilon^2 r^{2K})\mathrm{d}\Pi(A|d)\pi(d).
\end{aligned}$$

Therefore, using $P_{n,A}[(1-\phi_r)\mathbf{1}_{\mathcal{A}_r}] \leq 1$, the summand in the RHS is bounded by

$$\pi(d)\exp(c_\xi n\epsilon^2 r^{2K}). \tag{13}$$

Simultaneously, Eq. (11) gives another upper bound of the summand as

$$\begin{aligned}
&\int_{A\in\mathcal{F}_{r,d}:\|A-A^*\|_n>4\epsilon r^K} e^{c_\xi n\epsilon^2 r^{2K} - n\epsilon^2 r^{2K} + \log(D(2\sqrt{n}\epsilon r^K, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))}\mathrm{d}\Pi(A|d)\pi(d) \\
\leq&\pi(d)\exp\left\{-\frac{1}{2}n\epsilon^2 r^{2K} + \log(D(2\sqrt{n}\epsilon r^K, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))\right\},
\end{aligned} \tag{14}$$

for $r \geq 1$, where we used $c_\xi < \frac{1}{2}$.

We evaluate the packing number $\log(D(2\sqrt{n}\epsilon r^K, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n))$. It is known that the packing number of unit ball in d-dimensional Euclidean space is bounded by

$$D(\epsilon, B_d(1), \|\cdot\|) \leq \left(\frac{4+\epsilon}{\epsilon}\right)^d.$$

Here $B_d(R)$ denotes the ball with the radius R in d-dimensional Euclidean space.

Similar to Eq. (7), the L_2-norm between two tensors $A = [[U^{(1)}, \dots, U^{(K)}]]$ and $A' = [[U'^{(1)}, \dots, U'^{(K)}]]$ can be bounded by

$$\begin{aligned}
\|A - A'\|_2^2 = &\sum_{i_1, i_2, \dots, i_K} (\langle U^{(1)}_{:,i_1}, \dots, U^{(K)}_{:,i_K}\rangle - \langle U'^{(1)}_{:,i_1}, \dots, U'^{(K)}_{:,i_K}\rangle)^2 \\
= &\sum_{i_1, i_2, \dots, i_K} \left(\sum_{k=1}^K \langle U^{(1)}_{:,i_1}, \dots, U^{(k-1)}_{:,i_{k-1}}, U^{(k)}_{:,i_k} - U'^{(k)}_{:,i_k}, \dots, U'^{(K)}_{:,i_K}\rangle\right)^2 \\
\leq &K \sum_{i_1, i_2, \dots, i_K} \sum_{k=1}^K \langle U^{(1)}_{:,i_1}, \dots, U^{(k-1)}_{:,i_{k-1}}, U^{(k)}_{:,i_k} - U'^{(k)}_{:,i_k}, U'^{(k+1)}_{:,i_{k+1}}, \dots, U'^{(K)}_{:,i_K}\rangle^2 \\
\leq &K \sum_{k=1}^K \sum_{i_1, i_2, \dots, i_K} \|U^{(1)}_{:,i_1}\|_2^2 \times \dots \times \|U^{(k-1)}_{:,i_{k-1}}\|_2^2 \times \|U^{(k)}_{:,i_k} - U'^{(k)}_{:,i_k}\|_2^2 \times \|U'^{(k+1)}_{:,i_{k+1}}\|_2^2 \times \dots \times \|U'^{(K)}_{:,i_K}\|_2^2 \\
\leq &K \sum_{k=1}^K \|U^{(1)}\|_2^2 \times \dots \times \|U^{(k-1)}\|_2^2 \|U^{(k)} - U'^{(k)}\|_2^2 \times \|U'^{(k+1)}\|_2^2 \times \dots \times \|U'^{(K)}\|_2^2.
\end{aligned}$$

If $A, A' \in \mathcal{F}_{r,d}$, then the RHS is further bounded by

$$\|A - A'\|_2^2 \leq K M_{r,d}^{K-1} \sum_{k=1}^K \|U^{(k)} - U'^{(k)}\|_2^2. \tag{15}$$

Thus, if $2\epsilon r^K \leq K M_{r,d}^{K/2}$, using the relation (15) and $\sqrt{n}\|A - A'\|_n \leq \sqrt{n}\|A - A'\|_\infty \leq \sqrt{n}\|A - A'\|_2$, we have that

$$\begin{aligned}
&\log(D(2\sqrt{n}\epsilon r^K, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n)) \leq \log(D(2\epsilon r^K, \mathcal{F}_{r,d}, \|\cdot\|_2)) \\
&\leq \log(D(2\epsilon r^K/(K M_{r,d}^{(K-1)/2}), B_{d\sum_k M_k}(\sqrt{M_{r,d}}), \|\cdot\|_2)) \\
&\leq d\left(\sum_{k=1}^K M_k\right) \log\left(\frac{4 + \frac{2\epsilon r^K}{K M_{r,d}^{K/2}}}{\frac{2\epsilon r^K}{K M_{r,d}^{K/2}}}\right) \leq d(\sum_{k=1}^K M_k) \log\left(\frac{3K M_{r,d}^{K/2}}{\epsilon r^K}\right),
\end{aligned} \tag{16}$$

otherwise, $\log(D(2\sqrt{n}\epsilon r^K, \mathcal{F}_{r,d}, \sqrt{n}\|\cdot\|_n)) = 0$.

Combining Eqs. (13) and (14) with Eq. (16) results in the following upper bound of D_r:

$$D_r \leq \sum_d \pi(d) \min\left\{\exp(c_\xi n\epsilon^2 r^{2K}), \exp\left[-\frac{1}{2}n\epsilon^2 r^{2K} + d\left(\sum_{k=1}^K M_k\right) \log\left(\frac{3K M_{r,d}^{K/2}}{\epsilon r^K}\right)\right]\right\} \tag{17}$$

for $r \geq 1$.

Step 5.

Here, we establish the assertion by combining the bounds of A_r, B_r, C_r and D_r obtained above. Set $\epsilon = \frac{1}{\sqrt{n}}$ and $M_{r,d} = \frac{4}{\min(d^*, d)}\Xi(1/\sqrt{n})\sigma_{\mathrm{p}}^2 r^2$. Then, we have that, for all $d \geq d^*$,

$$d\left(\sum_{k=1}^K M_k\right) \log\left(\frac{3K M_{r,d}^{K/2}}{\epsilon r^K}\right) \leq d\left(\sum_{k=1}^K M_k\right) \log\left(3K\sqrt{n}\left(\frac{4\sigma_{\mathrm{p}}^2 \Xi(\frac{1}{\sqrt{n}})}{d^*}\right)^{\frac{K}{2}}\right).$$

Recall that $C_{n,K} = 3K\sqrt{n}\left(\frac{4\sigma_{\mathrm{p}}^2\Xi(\frac{1}{\sqrt{n}})}{d^*}\right)^{\frac{K}{2}}$, and $c_\xi = \min\{|\log(\xi)|/\log(C_{n,K}),1\}/4$. Let $\tilde{r}_d$ be such that $-\frac{1}{2}\tilde{r}_d^{2K} + d\left(\sum_{k=1}^K M_k\right)\log(C_{n,K}) = 0$, that is, $\tilde{r}_d = 2^{1/2K}[d(\sum_k M_k)\log(C_{n,K})]^{1/2K}$.

By the upper bound (9), we have that

$$\mathrm{E}\left[\int \|A - A^*\|_n^2 \mathrm{d}\Pi(A|Y_{1:n})\right] \leq 32\epsilon^2 + 32\epsilon^2 \int_{r>1} r^K(A_r + B_r + C_r + D_r)\mathrm{d}r^K.$$

We are going to bound each term in the integral.

$$\int_{r>1} r^K A_r \mathrm{d}r^K \leq \int_{r>1} r^K \min\{9\exp(-2r^{2K} + \log(d_{\max})), 1\}\mathrm{d}r^K \leq \frac{\log(d_{\max})}{2} + \frac{9}{4}.$$

By Lemma 3, the integral related to B_r is bounded by

$$\begin{aligned}
\int_{r>1} r^K B_r \mathrm{d}r^K \leq 8\Xi(\sqrt{c_\xi}\epsilon)/c_\xi + \int_{r^K > \sqrt{8\Xi(\sqrt{c_\xi}\epsilon)/c_\xi}} r^K \left[\exp(-r^2/8) + \exp(-c_\xi r^{2K}/16)\right]\mathrm{d}r^K \\
\leq 8\frac{\Xi(\sqrt{c_\xi}\epsilon)}{c_\xi} + \int_{r^K>1} r^K \exp(-r^2/8)\mathrm{d}r^K \\
+ \frac{1}{32}\exp\left(-\frac{\Xi(\sqrt{c_\xi}\epsilon)}{2c_\xi}\right).
\end{aligned}$$

Here, the second term of the RHS is bounded by

$$\int_{r^K>1} r^K \exp(-r^2/8)\mathrm{d}r^K \leq \frac{1}{2}\sum_{j=1}^K 8^j \frac{K!}{(K-j)!}\exp\left(-\frac{1}{8}\right) \leq 8^K(K+1)!.$$

Similarly, by Lemma 3, the integral related to C_r is bounded by

$$\begin{aligned}
\int_{r>1} r^K C_r \mathrm{d}r^K &\leq \int_{r>1} r^K \sum_{d=1}^M \exp\left[-\frac{d}{\min(d,d^*)}\Xi(\epsilon)r^2\right]\pi(d)\mathrm{d}r^K \\
&\leq \frac{1}{2}\sum_{j=0}^{K-1}\frac{K\cdots(K-j)}{\Xi(1/\sqrt{n})^{j+1}} \leq K.
\end{aligned}$$

Next, we bound the integral corresponding to D_r. By the definition of c_ξ and $\tilde{r}_d$, for all $r \leq \tilde{r}_d$, it holds that

$$\log[\pi(d)\exp(c_\xi r^{2K})] \leq d'(\sum_k M_k)\log(\xi) + c_\xi \tilde{r}_d^{2K} \leq d(\sum_k M_k)\log(\xi)/2,$$

(remind that we are using unnormalized prior $\pi(d)$). On the other hand, for all $r \geq \tilde{r}_d$, it holds that

$$\exp\left\{-n\epsilon^2 r^{2K}/2 + d\left(\sum_{k=1}^K M_k\right)\log\left(\frac{3KM_{r,d}^{K/2}}{\epsilon r^K}\right)\right\} \leq \exp\left[-(r^{2K} - \tilde{r}_d^{2K})/2\right].$$

Therefore, Eq. (17) becomes

$$\begin{aligned}
&\int_{r>1} r^K D_r \mathrm{d}r^K = \int_{r=1}^{\tilde{r}_{d^*}} r^K D_r \mathrm{d}r^K + \int_{r>\tilde{r}_{d^*}} r^K D_r \mathrm{d}r^K \\
\leq& \frac{1}{2}\tilde{r}_{d^*}^{2K} + \sum_{d=1}^{d^*-1}\pi(d)\int_{\tilde{r}_{d^*}\leq r} r^K \exp(-(r^{2K} - \tilde{r}_d^{2K}))\mathrm{d}r^K
\end{aligned}$$

$$+\sum_{d=d^*}^{M}\left\{\int_{\tilde{r}_{d^*}\le r\le \tilde{r}_d} r^K \mathrm{d}r^K \exp\left[\frac{1}{2}d(\sum\nolimits_k M_k)\log(\xi)\right] + \pi(d)\int_{\tilde{r}_d\le r} r^K \exp\left[-(r^{2K}-\tilde{r}_d^{2K})\right]\mathrm{d}r^K\right\} \tag{18}$$

Here the second term of the RHS can be evaluated as

$$\sum_{d=1}^{d^*-1}\pi(d)\int_{\tilde{r}_{d^*}\le r} r^K \exp(-(r^{2K}-\tilde{r}_d^{2K}))\mathrm{d}r^K \le \sum_{d=1}^{d^*-1}\pi(d).$$

The third term is evaluated as

$$\begin{aligned}
&\sum_{d=d^*}^{M}\int_{\tilde{r}_{d^*}\le r\le\tilde{r}_d} r^K\mathrm{d}r^K\exp\left[\frac{1}{2}d\left(\sum\nolimits_k M_k\right)\log(\xi)\right]\\
\le&\sum_{d=d^*}^{M}\frac{1}{2}\tilde{r}_d^{2K}\exp\left[\frac{1}{2}d\left(\sum\nolimits_k M_k\right)\log(\xi)\right]\\
=&\sum_{d=d^*}^{M} d(\sum\nolimits_k M_k)\log(C_{n,K})\exp\left[\frac{1}{2}d\left(\sum_k M_k\right)\log(\xi)\right]\\
\le&\int_{d\ge d^*}(d+1)(\sum\nolimits_k M_k)\log(C_{n,K})\exp\left[\frac{1}{2}d\left(\sum_k M_k\right)\log(\xi)\right]\mathrm{d}d\\
\le&\frac{2\sum_k M_k(d^*+2)}{\sum_k M_k|\log(\xi)|}\log(C_{n,K})\exp\left[\frac{d^*(\sum_k M_k)}{2}\log(\xi)\right]\\
\le&\frac{2(d^*+2)}{|\log(\xi)|}\log(C_{n,K})
\end{aligned}$$

The fourth term is bounded in a similar way to the second term as

$$\sum_{d=d^*}^{M}\pi(d)\int_{\tilde{r}_d\le r} r^K\exp(-(r^{2K}-\tilde{r}_d^{2K}))\mathrm{d}r^K\le\sum_{d=d^*}^{M}\pi(d).$$

Thus Eq. (18) is upper bounded by

$$\begin{aligned}
&\int_{r\ge1} r^K D_r \mathrm{d}r^K \le \frac{1}{2}\tilde{r}_{d^*}^{2K} + 2 + \frac{2(d^*+2)}{|\log(\xi)|}\log(C_{n,K})\\
=&d^*\left(\sum\nolimits_k M_k\right)\log(C_{n,K}) + 2 + \frac{2(d^*+2)}{|\log(\xi)|}\log(C_{n,K}).
\end{aligned}$$

Combining all inequalities yields that there exists a universal constant C such that

$$\begin{aligned}
&\mathrm{E}\left[\int \|A-A^*\|_n^2\mathrm{d}\Pi(A|Y_{1:n})\right]\\
\le&\frac{C}{n}\left(\frac{\log(d_{\max})}{2}+K+8\frac{\Xi(\sqrt{\frac{c_\xi}{n}})}{c_\xi}+8^K(K+1)!+d^*\left(\sum\nolimits_k M_k\right)\log(C_{n,K})\right.\\
&\left.+\frac{2(d^*+2)}{|\log(\xi)|}\log(C_{n,K})+1\right).
\end{aligned}\tag{19}$$

B. Rejection Sampling (Proof of Theorem 2)

In this section, we prove Theorem 2. By assumption, we have $R\ge 2\|A^*\|_\infty$ and $\frac{R}{2}\ge 1/\sqrt{n}$. Therefore, if $\epsilon r\ge 1/\sqrt{n}$, we have that

$$-\log(\Pi(A:\|A-A^*\|_n\le\epsilon r,\|A\|_\infty\le R))\le\Xi(\frac{R}{2}\wedge\epsilon r)\le\Xi(1/\sqrt{n}). \tag{20}$$

Now, for any non-negative measurable function $f : \mathbb{R}^{M_1\times\cdots\times M_K}$, we have that

$$\int f(A)\Pi(\mathrm{d}A|\|A\|_\infty \le R, D_n) = \frac{\int f(A)\mathbf{1}[\|A\|_\infty \le R]\Pi(\mathrm{d}A|D_n)}{\Pi(\|A\|_\infty \le R|Y_{1:n})} = \frac{\int f(A)\mathbf{1}[\|A\|_\infty \le R]\Pi(\mathrm{d}A|D_n)}{\Pi(\|A\|_\infty \le R|Y_{1:n})} = \frac{\int f(A)\mathbf{1}[\|A\|_\infty \le R]\frac{p_{n,A}}{p_{n,A^*}}\Pi(\mathrm{d}A)}{\int \mathbf{1}[\|A\|_\infty \le R]\frac{p_{n,A}}{p_{n,A^*}}\Pi(\mathrm{d}A)}.$$

We define the event $\mathcal{A}_r$ as in the proof of Theorem 1. Then, we have the same upper bound of B_r as in the previous section. By Eq. (20), on the event $\mathcal{A}_r$,

$$\int \mathbf{1}[\|A\|_\infty \le R]\frac{p_{n,A}}{p_{n,A^*}}\Pi(\mathrm{d}A) \ge \min\{\exp(r^2 + \Xi(1/\sqrt{n})), \exp(c_\xi r^{2K})\}^{-1}.$$

Other quantities such as A_r, C_r, D_r are also bounded in the same manner because

$$\int f(A)\mathbf{1}[\|A\|_\infty \le M]\frac{p_{n,A}}{p_{n,A^*}}\Pi(\mathrm{d}A) \le \int f(A)\frac{p_{n,A}}{p_{n,A^*}}\Pi(\mathrm{d}A).$$

Thus, the conditional posterior mean of the squared error $\mathrm{E}\left[\int \|A - A^*\|_n^2 \mathrm{d}\Pi(A|\|A\|_\infty \le M, Y_{1:n})\right]$ can be bounded by the same quantity as the RHS of Eq. (19).

Now, we are going to bound the out-of-sample predictive error:

$$\mathrm{E}_{D_n}\left[\int \|A - A^*\|_{L_2(P(X))}^2 \mathrm{d}\Pi(A|\|A\|_\infty \le R, D_n)\right]. \tag{21}$$

By the assumption that $\|X\|_1 \le 1$ a.s., we have that

$$|\langle X, A - A^*\rangle| \le 2R.$$

Now, we bound the expected error (21) by $I + II + III$ where, for $\eta = \epsilon\max\{4R, 1\} = \frac{1}{\sqrt{n}}\max\{4R, 1\}$, I, II and III are defined by

$$I = \mathrm{E}_{D_n}\left[\int_0^\infty \eta^2 r \mathbf{1}_{\mathcal{A}_r^c}\mathrm{d}r\right],$$

$$II = \mathrm{E}_{D_n}\left[\int_0^\infty \eta^2 r \mathbf{1}_{\mathcal{A}_r}\Pi(A : \sqrt{2}\|A - A^*\|_n > \eta r \mid \|A\|_\infty \le R, D_n)\mathrm{d}r\right],$$

$$III = \mathrm{E}_{D_n}\left[\int_0^\infty \eta^2 r \mathbf{1}_{\mathcal{A}_r}\Pi(A : \|A - A^*\|_{L_2(P(X))} > \eta r \ge \sqrt{2}\|A - A^*\|_n \mid \|A\|_\infty \le R, D_n)\mathrm{d}r\right].$$

Here, I and II are bounded by

$$I + II \le C\eta^2\Bigg(\log(d_{\max}) + K + \frac{\Xi(\sqrt{\frac{c_\xi}{n}})}{c_\xi} + 8^K(K+1)! + d^*\left(\textstyle\sum_k M_k\right)\log(C_{n,K}) + \frac{d^*+2}{|\log(\xi)|}\log(C_{n,K}) + C\Bigg), \tag{22}$$

as in Eq. (19).

Next, we bound the term III. To bound this, we need to evaluate the difference between the empirical norm $\|A - A^*\|_n$ and the expected norm $\|A - A^*\|_{L_2(P(X))}$, which can be done by Bernstein's inequality:

$$P\left(\|A - A^*\|_{L_2(P(X))} \ge \sqrt{2}\|A - A^*\|_n\right) \le \exp\left(-\frac{n\|A - A^*\|_{L_2(P(X))}^4}{2(v + \|A - A^*\|_\infty^2/3)}\right),$$

where $v = \mathrm{E}_X[(\langle X, A - A^* \rangle^2 - \mathrm{E}_{\tilde{X}}[\langle \tilde{X}, A - A^* \rangle^2])^2]$. Now $v \le \mathrm{E}_X[\langle X, A - A^* \rangle^4] \le \|A - A^*\|_\infty^2 \mathrm{E}_X[\langle X, A - A^* \rangle^2] = \|A - A^*\|_\infty^2 \|A - A^*\|_{L_2(P(X))}^2$. This yields that

$$P\left(\|A - A^*\|_{L_2(P(X))} \ge \sqrt{2}\|A - A^*\|_n\right) \le \exp\left[-\frac{n}{2}\left(\frac{\|A - A^*\|_{L_2(P(X))}}{\|A - A^*\|_\infty}\right)^2\right].$$

If $\|A - A^*\|_\infty \le 2R$, then the RHS is further bounded by $\exp\left(-\frac{n\|A - A^*\|_{L_2(P(X))}^2}{8R^2}\right)$.

To evaluate III, we evaluate the expectation of the posterior inside the integral:

$$\begin{aligned}
&\mathrm{E}_{D_n}\left[\mathbf{1}_{\mathcal{A}_r}\Pi(A : \|A - A^*\|_{L_2(P(X))} > \eta r \ge \sqrt{2}\|A - A^*\|_n \mid \|A\|_\infty \le R, D_n)\right] \\
\le &\mathrm{E}_{X_{1:n}}\left[\exp(r^2 + \Xi(\epsilon))\Pi(A : \|A - A^*\|_{L_2(P(X))} > \eta r \ge \sqrt{2}\|A - A^*\|_n, \|A\|_\infty \le R \mid X_{1:n})\right] \\
\le &\exp(r^2 + \Xi(1/\sqrt{n}))\int_A \mathbf{1}[\|A - A^*\| > \eta r, \|A\|_\infty \le R] P(\|A - A^*\|_{L_2(P(X))} > \eta r \ge \sqrt{2}\|A - A^*\|_n)\Pi(\mathrm{d}A) \\
\le &\exp(r^2 + \Xi(1/\sqrt{n}))\int_A \mathbf{1}[\|A - A^*\| > \eta r]\exp\left(-\frac{n\|A - A^*\|_{L_2(P(X))}^2}{8R^2}\right)\Pi(\mathrm{d}A) \\
\le &\exp\left(r^2 + \Xi(1/\sqrt{n}) - \frac{n\eta^2 r^2}{8R^2}\right).
\end{aligned}$$

Therefore, we have that

$$\begin{aligned}
III \le &\int_0^\infty \eta^2 r \min\left\{1, \exp\left(r^2 + \Xi(1/\sqrt{n}) - \frac{n\eta^2 r^2}{8R^2}\right)\right\}\mathrm{d}r \\
\le &\frac{1}{2}\eta^2\Xi(1/\sqrt{n}) + \int_{r > \sqrt{\Xi(1/\sqrt{n})}}^\infty \eta^2 r \exp\left(-r^2\right)\mathrm{d}r \\
\le &C\eta^2\Xi(1/\sqrt{n}),
\end{aligned} \tag{23}$$

where $\eta = \frac{1}{\sqrt{n}}\max\{4R, 1\}$ is used in the second inequality.

Combining the inequalities (22) and (23), we obtain that

$$\begin{aligned}
&\mathrm{E}_{D_n}\left[\int \|A - A^*\|_{L_2(P(X))}^2 \mathrm{d}\Pi(A | \|A\|_\infty \le R, D_n)\right] \\
\le &\frac{C\max\{R^2, 1\}}{n}\left(\frac{\log(d_{\max})}{2} + 8^K(K+1)! + \frac{\Xi(\sqrt{\frac{c_\xi}{n}})}{c_\xi} + \Xi(1/\sqrt{n})\right. \\
&\left. + d^*\left(\sum\nolimits_k M_k\right)\log(C_{n,K}) + \frac{d^* + 2}{|\log(\xi)|}\log(C_{n,K}) + C\right),
\end{aligned}$$

where C is a universal constant. Noticing $\frac{\Xi(\sqrt{\frac{c_\xi}{n}})}{c_\xi} \ge \Xi(1/\sqrt{n})$, we obtain the assertion.

C. Rejection Sampling II (Proof of Theorem 3)

In this section, the proof of Theorem 3 is given. Let

$$\begin{aligned}
&\mathcal{U}_{R,d} = \{(U^{(1)}, \dots, U^{(K)}) \in \mathbb{R}^{d \times M_1} \times \cdots \times \mathbb{R}^{d \times M_K} \mid \|U_{:,j}^{(k)}\|_2 \le R \ (1 \le k \le d_{\max}, 1 \le j \le M_k).\}, \\
&\mathcal{U}_R = \bigcup_d \mathcal{U}_{R,d}.
\end{aligned}$$

Define

$$\mathcal{F}_{r,d} = \left\{A \in \mathbb{R}^{M_1 \times \cdots \times M_K} \mid \|A\|_{\max,2} \le R\right\}.$$

For $\boldsymbol{U} = (U^{(1)}, \dots, U^{(K)})$, we denote by $A_{\boldsymbol{U}} := [[U^{(1)}, \dots, U^{(K)}]]$. Since the rejection sampling considered here is defined on a factorization $A = [[U^{(1)}, \dots, U^{(K)}]]$, the quantities I, II, III are redefined as

$$I = \mathrm{E}_{D_n}\left[\int_0^\infty \eta^2 r \mathbf{1}_{\mathcal{A}_r^c} \mathrm{d}r\right],$$
$$II = \mathrm{E}_{D_n}\left[\int_0^\infty \eta^2 r \mathbf{1}_{\mathcal{A}_r} \Pi(\boldsymbol{U} : \sqrt{2}\|A_{\boldsymbol{U}} - A^*\|_n > \eta r \mid \boldsymbol{U} \in \mathcal{U}_R, D_n) \mathrm{d}r\right],$$
$$III = \mathrm{E}_{D_n}\left[\int_0^\infty \eta^2 r \mathbf{1}_{\mathcal{A}_r} \Pi(\boldsymbol{U} : \|A_{\boldsymbol{U}} - A^*\|_{L_2(P(X))} > \eta r \geq \sqrt{2}\|A_{\boldsymbol{U}} - A^*\|_n \mid \boldsymbol{U} \in \mathcal{U}_R, D_n) \mathrm{d}r\right].$$

Since $R \geq \|A^*\|_{\max,2} + \sigma_{\mathrm{p}}$ is satisfied, one can apply the same upper-bound evaluation of $\Pi(\boldsymbol{U} : \|A_{\boldsymbol{U}} - A^*\|_n < \epsilon, A_{\boldsymbol{U}} \in \mathcal{U}_R | d^*)$ as Eq. (8).

Then, in the same event $\mathcal{A}_r$ as before, we have that

$$\int \mathbf{1}[\|A\|_\infty \leq R] \frac{p_{n,A}}{p_{n,A^*}} \Pi(\mathrm{d}A) \geq \min\{\exp(r^2 + \Xi(1/\sqrt{n})), \exp(c_\xi r^{2K})\}^{-1},$$

by Eq. (20).

A_r, D_r can be bounded in a similar fashion, except we use R instead of $M_{r,d}$ and $\mathcal{U}_R$ instead of $\mathcal{F}_r$. Now, in this case, C_r could be zero. That is,

$$C_r = \mathrm{E}_{Y_{1:n}}[\Pi(\mathcal{U}_R^c | \{U^{(k)}\}_k \in \mathcal{U}_R, Y_{1:n}) \mathbf{1}_{\mathcal{A}_r}] = 0.$$

Thus, by resetting $M_{r,d} = R$ and, accordingly, re-defining

$$C_{n,K} = 3K\sqrt{n} R^{\frac{K}{2}},$$

then we have the same bound of $I + II$ as Eq. (22) except the definition of $C_{n,K}$.

Because $\|A_{\boldsymbol{U}}\|_\infty \leq R^K$ for $\boldsymbol{U} \in \mathcal{U}_R$, III is bounded by $\frac{C}{n} R^{2K} \Xi(\epsilon)$.

Therefore, we again obtain that

$$\mathrm{E}_{D_n}\left[\int \|A_U - A^*\|_{L_2(P(X))}^2 \mathrm{d}\Pi(A_U \mid U \in \mathcal{U}_R, D_n)\right]$$
$$\leq \frac{C \max\{R^{2K}, 1\}}{n} \Bigg(\log(d_{\max}) + 8^K (K+1)! + \frac{\Xi(\sqrt{\frac{c_\xi}{n}})}{c_\xi} + \Xi(\frac{1}{\sqrt{n}})$$
$$+ d^* \left(\sum\nolimits_k M_k\right) \log(K\sqrt{n} R^{\frac{K}{2}}) + \frac{d^* + 2}{|\log(\xi)|} \log(K\sqrt{n} R^{\frac{K}{2}}) + C \Bigg).$$

This yields the assertion.

D. Auxiliary lemmas

Proposition 1. *Tail bound of χ^2 distribution Let χ_k be the chi-square random variable with freedom k. Then the tail probability is bounded as*

$$P(\chi_k \geq k + 2\sqrt{xk} + 2x) \leq \exp(-x)$$
$$P(\chi_k \geq x) \leq \exp\left[\frac{k}{2} + \frac{k}{2} \log\left(\frac{x}{k}\right) - \frac{x}{2}\right],$$

for $x \geq k$.

The proofs of the first assertion and the second one are given in Lemma 1 of (Laurent & Massart, 2000) and Lemma 2.2 of (Dasgupta & Gupta, 2002) respectively.

Lemma 2. *The small ball probability of d-dimensional Gaussian random variable $g \sim N(0, \sigma \mathrm{I}_d)$ is lower bounded as*

$$P\left(\|g\| \leq \epsilon\right) \geq \left(\frac{\epsilon}{3\sigma d^{\frac{1}{2}}}\right)^d,$$

for all $\epsilon \leq \sigma\sqrt{d}$.

Proof. The lower bound is obtained in an almost same manner as the proof of Proposition 2.3 of (Li & Linde, 1999). If $\delta \leq 1$, then we have

$$P(|g_j| \leq \delta\sigma) = \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^{\delta\sigma} \frac{2\delta}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathrm{d}x \geq \frac{2\delta}{\sqrt{2\pi}} \exp(-\frac{1}{2}) \geq \frac{\delta}{3}.$$

Therefore, for ϵ such that $\epsilon \leq \sigma\sqrt{d}$, it holds that

$$\begin{aligned} P\left(\|g\| \leq \epsilon\right) &\geq P\left(\sqrt{d}\max_j |g_j| \leq \epsilon\right) \geq \prod_{j=1}^d P\left(\sqrt{d}|g_j| \leq \epsilon\right) \\ &\geq \left(\frac{\epsilon}{3\sigma d^{\frac{1}{2}}}\right)^d. \end{aligned}$$

□

Lemma 3. *For all $a > 0, c > 0, K \geq 1$, it holds that*

$$\int_{x \geq a} x \exp\big(-cx^{\frac{2}{K}}\big)\mathrm{d}x = \frac{1}{2}\sum_{i=1}^K \frac{K\cdots(K-i+1)}{c^i} a^{(K-i)2/K} \exp(-ca^{2/K}).$$

Proof.

$$\begin{aligned} &\int_{x \geq a} x \exp(-cx^{\frac{2}{K}})\mathrm{d}x \\ =&\frac{K}{2}\int_{x \geq a^{2/K}} x^{K-1}\exp(-cx)\mathrm{d}x \\ =&\frac{K}{2}\left[-\frac{1}{c}x^{K-1}\exp(-cx)\right]_{a^{2/K}}^{\infty} + \frac{K(K-1)}{2c}\int_{a^{2/K}}^{\infty} x^{K-2}\exp(-cx)\mathrm{d}x \\ =&\frac{K}{2c}a^{(K-1)2/K}\exp(-ca^{2/K}) + \frac{K(K-1)}{2c}\int_{a^{2/K}}^{\infty} x^{K-2}\exp(-cx)\mathrm{d}x. \end{aligned}$$

Then, applying recursive argument we obtain the assertion. □

Lemma 4. *For all $K \geq 2$, we have*

$$\sum_{i=1}^d u_i^{(1)} u_i^{(2)} \cdots u_i^{(K)} \leq \prod_{k=1}^K \|u^{(k)}\|_K.$$

In particular,

$$\sum_{i=1}^d u_i^{(1)} u_i^{(2)} \cdots u_i^{(K)} \leq \prod_{k=1}^K \|u^{(k)}\|_2.$$

Proof. By Hölder's inequality, it holds that

$$\sum_{i=1}^d u_i^{(1)} u_i^{(2)} \cdots u_i^{(K)} \leq \|u^{(1)}\|_K \|(u_i^{(2)} \cdots u_i^{(K)})_{i=1}^d\|_{K/(K-1)}.$$

Applying Hölder's inequality again, we observe that

$$\|(u_i^{(2)}\cdots u_i^{(K)})_{i=1}^d\|_{K/(K-1)} = \left(\sum_{i=1}^d |u_i^{(2)}\cdots u_i^{(K)}|^{K/(K-1)}\right)^{(K-1)/K}$$
$$\leq \left[\left(\sum_{i=1}^d |u^{(2)}|^K\right)^{1/K-1}\left(\sum_{i=1}^d |u_i^{(3)}\cdots u_i^{(K)}|^{K/(K-2)}\right)^{(K-2)/(K-1)}\right]^{(K-1)/K}$$
$$\leq \left[\left(\sum_{i=1}^d |u_i^{(2)}|^K\right)^{1/(K-1)}\left(\sum_{i=1}^d |u_i^{(3)}\cdots u_i^{(K)}|^{K/(K-2)}\right)^{(K-2)/(K-1)}\right]^{(K-1)/K}$$
$$= \|u^{(2)}\|_K \left\|(u_i^{(3)}\cdots u_i^{(K)})_{i=1}^d\right\|_{K/(K-2)}.$$

Then by applying the same argument recursively we obtain the assertion.

The second assertion is obvious because $\|u\|_2 \leq \|u\|_K$ for all $K \geq 2$. □

E. Minimax optimality (Proof of Theorem 4)

The proof of Proof of Theorem 4 is given as follows.

Proof. Let $\mathcal{T} \in \mathbb{R}^{M_1\times\cdots\times M_K}$ be a set of tensors. The δ packing number $\mathcal{M}(\delta, \mathcal{T}, L_2(P(X)))$ of the set $\mathcal{T}$ with respect to $L_2(P(X))$ norm is the largest number of tensors $\{A_1,\ldots,A_{\mathcal{M}}\} \subseteq \mathcal{T}$ such that $\|A_i - A_j\|_{L_2(P(X))} \geq \delta$ for all $i \neq j$. On the other hand, the δ covering number $\mathcal{N}(\delta, \mathcal{T}, L_2(P(X)))$ of the set $\mathcal{T}$ with respect to $L_2(P(X))$ norm is the smallest number of tensors $\{A_1,\ldots,A_{\mathcal{N}}\} \subseteq \mathcal{T}$ such that $\forall A \in \mathcal{T}$ there exists A_i and $\|A_i - A_j\|_{L_2(P(X))} \leq \delta$ is satisfied. It is easily checked that

$$\mathcal{N}(\delta/2, \mathcal{T}, L_2(P(X))) \leq \mathcal{M}(\delta, \mathcal{T}, L_2(P(X))) \leq \mathcal{N}(\delta, \mathcal{T}, L_2(P(X))). \tag{24}$$

For a given $\delta_n > 0$ and $\varepsilon_n > 0$, let Q be the δ_n packing number $\mathcal{M}(\delta_n, \mathcal{T}_R, L_2(P(X)))$ of $\mathcal{T}_R$ and N be the ε_n covering number $\mathcal{N}(\varepsilon_n, \mathcal{T}_R, L_2(P(X)))$ of $\mathcal{T}_R$. As shown in (Yang & Barron, 1999), we utilize the following inequality:

$$\inf_{\hat{A}} \sup_{A^*\in\mathcal{T}_R} \mathrm{E}[\|\hat{A} - A^*\|^2_{L_2(P(X))}] \geq \inf_{\hat{A}} \sup_{A^*\in\mathcal{T}_R} \frac{\delta_n^2}{2} P[\|\hat{A} - A^*\|^2_{L_2(P(X))} \geq \delta_n^2/2]$$
$$\geq \frac{\delta_n^2}{2}\left(1 - \frac{\log(N) + \frac{n}{2\sigma^2}\varepsilon_n^2 + \log(2)}{\log(Q)}\right).$$

Thus by taking δ_n and ε_n to satisfy

$$\frac{n}{2\sigma^2}\varepsilon_n^2 \leq \log(N), \tag{25a}$$
$$8\log(N) \leq \log(Q), \tag{25b}$$
$$4\log(2) \leq \log(Q), \tag{25c}$$

the minimax rate is lower bounded by $\frac{\delta_n^2}{4}$.

Note that by the assumption on the distribution, we have that $\|A\|^2_{L_2(P(X))} = \frac{1}{\prod_{k=1}^K M_k}\sum_{i_1,i_2,\ldots,i_K} A^2_{i_1,\ldots,i_K} = \frac{1}{\prod_{k=1}^K M_k}\|A\|_2^2$, $\langle A, A'\rangle_{L_2(P(X))} = \frac{1}{\prod_{k=1}^K M_k}\sum_{i_1,i_2,\ldots,i_K} A_{i_1,\ldots,i_K}A'_{i_1,\ldots,i_K}$ and $\|X\|_1 \leq 1$.

We construct a δ covering set of $\mathcal{T}_R$ in the following. Let $\mathcal{E}_R$ be a δ covering of the ball with radius R. It is known that we can take $\mathcal{E}_R$ as

$$|\mathcal{E}_R| \geq \left(\frac{R}{\delta}\right)^{d^*}.$$

Let

$$\mathcal{E}_R^M := \{(\mu_1,\ldots,\mu_M) \in \mathcal{E}_R \times\cdots\times \mathcal{E}_R\} \subset \mathbb{R}^{d^*\times M}.$$

Here there is a $\sqrt{\delta^2 MR^2/4}$ packing set $\tilde{\mathcal{U}}_M$ of $\mathcal{E}_R^M$ by Lemma **??** such that $|\tilde{\mathcal{U}}_M| \geq \left(1/\delta^{d^*}\right)^{M/4}$. Based on $\tilde{\mathcal{U}}_M$, $\mathcal{U}_M$ is defined by extending the elements in $\tilde{\mathcal{U}}_M$ to $\mathbb{R}^{d^* \times M}$. Let

$$\hat{\mathcal{U}}_M := \{U \in \mathbb{R}^{d^* \times M} \mid 4\sqrt{d}\tilde{U} \in \tilde{\mathcal{U}}_M, U_{i,j} = \tilde{U}_{i,j} + c_i \text{ where } c_i \text{ is chosen so that } \|U_{i,:}\|^2 = MR^2/d^*\}.$$

By construction, we have that $|\hat{\mathcal{U}}_M| \geq \left(1/\delta^{d^*}\right)^{M/4}$. By introducing $(c_i)_{i=1}^{d^*}$, $U,\ U' \in \hat{\mathcal{U}}$ corresponding to $\tilde{U},\ \tilde{U}' \in \tilde{\mathcal{U}}_M$ s.t. $\tilde{U} \neq \tilde{U}'$ (that is, $\|\tilde{U} - \tilde{U}'\|_2^2 \geq R^2\delta^2 M/4$) could be so close to each other that $\|U - U'\|_2^2 < R^2\delta^2 M/64$. To deal with this problem, we identify elements $\tilde{U}$ and $\tilde{U}'$ for which there exists $(c_i)_{i=1}^{d^*} \in \mathbb{R}^{d^*}$ such that $\sum_{i=1}^{d^*}\sum_{j=1}^{M}(\tilde{U}_{i,j} - \tilde{U}'_{i,j} - c_i)^2 < MR^2\delta^2/4$. We pick up a representative from the equivalence class induced from the identification defined above, and denote by $\mathcal{U}_M$ the set consisting of the representatives. Then $|\mathcal{U}_M| \geq \frac{\left(1/\delta^{d^*}\right)^{M/4}}{\left(1/(\delta/8)^{d^*}\right)} = 8^{-d^*}\left(1/\delta^{d^*}\right)^{M/4-1}$. Then, by construction, we have $\forall U, U' \in \mathcal{U}_M$ s.t. $U \neq U'$ satisfy $\|U - U'\|_2^2 \geq R^2\delta^2 M/64$, $\|U_{j,:}\|_2 = R\sqrt{M/d}$, and $\|U_{j,:} - U'_{j,:}\|_2 \leq R\sqrt{M/d}$ $(\forall j)$. In particular, there exists j such that $\|U_{j,:} - U'_{j,:}\|_2^2 \geq R^2\delta^2 M/(64d^*)$.

For $\mathcal{U}_M$, we define its "shifted" set $\bar{\mathcal{U}}_M = \left\{\sqrt{d^*} \times \begin{pmatrix} u_1^\top & 0\cdots0 & \cdots & 0\cdots0 \\ 0\cdots0 & u_2^\top & \cdots & 0\cdots0 \\ \vdots & \vdots & \ddots & \vdots \\ 0\cdots0 & 0\cdots0 & \cdots & u_{d^*}^\top \end{pmatrix} \mid \begin{pmatrix} u_1^\top \\ u_2^\top \\ \vdots \\ u_{d^*}^\top \end{pmatrix} \in \mathcal{U}_M\right\}$. Let

$$\hat{\mathcal{T}} = \left\{[[U^{(1)}, \dots, U^{(K)}]] \mid U^{(k)} \in \mathcal{U}_{M_k}\ (k = 1, \dots, K-1),\ U^{(K)} \in \bar{\mathcal{U}}_{M_K/d^*}\right\}.$$

From now on, we show that $\hat{\mathcal{T}}$ is an ϵ packing of $\mathcal{T}_R$ with respect to $L_2(P(X))$ for some $\epsilon > 0$. By the construction, it is easily seen that $\hat{\mathcal{T}} \subset \mathcal{T}_R$. By the definition of $\bar{\mathcal{U}}_{M_K/d^*}$, each element A in $\hat{\mathcal{T}}$ can be orthogonally decomposed into

$$A = \sum_{d=1}^{d^*} u_d^{(1)} \otimes u_d^{(2)} \otimes \cdots \otimes u_d^{(K)},$$

where each component $u_d^{(1)} \otimes u_d^{(2)} \otimes \cdots \otimes u_d^{(K)}$ $(d = 1, \dots, d^*)$ is mutually orthogonal. Moreover, for two tensors A and A' in $\hat{\mathcal{T}}$, we may take such decomposition so that

$$\|A - A\|_2^2 = \sum_{d=1}^{d^*} \|(u_d^{(1)} \otimes \cdots \otimes u_d^{(K)}) - (v_d^{(1)} \otimes \cdots \otimes v_d^{(K)})\|_2^2$$

where $A = \sum_{d=1}^{d^*} u_d^{(1)} \otimes \cdots \otimes u_d^{(K)}$ and $A' = \sum_{d=1}^{d^*} v_d^{(1)} \otimes \cdots \otimes v_d^{(K)}$. Therefore, we may focus on rank one tensors.

If the norms of two vectors x and y are same, we have that

$$\begin{aligned} &\|x\|_2 = \|y\|_2 \\ \Rightarrow &\|x - y + y\|_2^2 = \|x - y\|_2^2 + 2\langle x - y, y\rangle + \|y\|_2^2 = \|y\|_2^2 \\ \Rightarrow &\langle x - y, y\rangle = -\frac{1}{2}\|x - y\|_2^2. \end{aligned} \tag{26}$$

Therefore, for $u_1, u_1' \in \mathbb{R}_{M_1}$ and $U, U' \in \mathbb{R}^{M_2 \times \cdots \times M_K}$ such that $\|u_1\|_2^2 = \|u_1'\|_2^2 = R^2 M_1/d$ and $\|U\|_2^2 = \|U'\|_2^2 = \prod_{k=2}^K (R^2 M_k/d^*) (= (R^2/d^*)^{K-1}\check{M})$ and $\|u_1 - u_1'\|_2 \geq M_1\delta^2 R^2/(d^* 64)$ and $\|u_1 - u_1'\|_2 \leq R\sqrt{M_1}$. Then we have that

$$\begin{aligned} &\|u_1 \otimes U - u_1' \otimes U'\|_2^2 \\ =&\|(u_1 - u_1') \otimes U'\|^2 + \|u_1' \otimes (U - U')\|^2 + \|(u_1 - u_1') \otimes (U - U')\|^2 \\ &+ 2\langle (u_1 - u_1') \otimes U', u_1' \otimes (U - U')\rangle + 2\langle (u_1 - u_1') \otimes U', (u_1 - u_1') \otimes (U - U')\rangle \\ &+ 2\langle u_1' \otimes (U - U'), (u_1 - u_1') \otimes (U - U')\rangle \\ =&\|u_1 - u_1'\|^2\|U'\|^2 + \|u_1'\|^2\|U - U'\|^2 + \|u_1 - u_1'\|^2\|U - U'\|^2 \end{aligned}$$

$$
\begin{aligned}
&+2\langle u_1-u_1',u_1'\rangle\langle U',U-U'\rangle+2\|u_1-u_1'\|^2\langle U',(U-U')\rangle+2\langle u_1-u_1',u_1'\rangle\|U-U'\|^2\\
=&\|u_1-u_1'\|^2(R^2/d^*)^{K-1}\check{M}+R^2M_1/d^*\|U-U'\|^2+\|u_1-u_1'\|^2\|U-U'\|^2\\
&+\frac{1}{2}\|u_1-u_1'\|^2\|U-U'\|^2-\|u_1-u_1'\|^2\|U-U'\|^2-\|u_1-u_1'\|^2\|U-U'\|^2\\
&(\because \text{Eq. (26)})\\
=&\|u_1-u_1'\|^2(R^2/d^*)^{K-1}\check{M}+(R^2M_1/d^*)\|U-U'\|^2-\frac{1}{2}\|u_1-u_1'\|^2\|U-U'\|^2\\
=&\|u_1-u_1'\|^2(R^2/d^*)^{K-1}\check{M}+\|U-U'\|^2(R^2M_1/d^*-\frac{1}{2}\|u_1-u_1'\|^2).
\end{aligned}
$$

Since $\|u_1-u_1'\|^2\le R^2M_1/d^*$, the right hand side is lower bounded by

$$
(R^2/d^*)^K\frac{M_1\check{M}\delta^2}{64}=\frac{(R^2/d^*)^K\prod_{k=1}^K M_k\delta^2}{64}.
$$

This implies that $\hat{\mathcal{T}}$ is a $\sqrt{\frac{\prod_{k=1}^K M_k\delta^2}{64}}(R/\sqrt{d^*})^K$ packing set with respect to $\|\cdot\|_2$ such that

$$
|\hat{\mathcal{T}}|\ge 8^{-Kd^*}\prod_{k=1}^K\left(\frac{1}{\delta^{d^*}}\right)^{M_k/4-1}.
$$

Simultaneously $\hat{\mathcal{T}}$ is a $(R/\sqrt{d^*})^K\frac{\delta}{8}$ packing set with respect to $\|\cdot\|_{L_2(P(X))}$.

Therefore, we may take $\delta_n^2=C\min\left\{\sigma^2\frac{d^*(\sum_{k=1}^K M_k)}{n},(R^2/d^*)^K\right\}$ and $\epsilon_n^2=c\delta_n^2$ with appropriate constants C and c so that the conditions (25) are satisfied. Hence, we obtain the assertion. □

Lemma 5 (Varshamov-Gilbert bound). *Suppose that we are given a finite set $\mathcal{E}\in\mathbb{R}^d$ such that $|\mathcal{E}|\ge a$ and $\forall\mu,\mu'\in\mathcal{E}$ s.t. $\mu\ne\mu'$ satisfies $\|\mu-\mu'\|\ge\delta$. Let $m\ge 4$, then there exists a finite set $\{U^{(0)},\dots,U^{(M)}\}\subset\mathcal{E}^m$ such that*

$$
\|U^{(j)}-U^{(j')}\|^2\ge m\delta^2/4,\ \ (0\le j<k\le M)
$$

and

$$
M\ge a^{m/4}.
$$

Proof. Pick up $U^{(0)}\in\mathcal{E}^m$ arbitrary. Let $D=\lfloor m/4\rfloor$. $\Omega_1=\{U\in\mathcal{E}^m\mid\|U-U^{(0)}\|_2^2>\delta^2D\}$ and pick up one element $U^{(1)}$ from Ω_1 and construct $\Omega_2=\{U\in\Omega_1\mid\|U-U^{(1)}\|_2^2>\delta^2D\}$, and recursively we define $\Omega_j=\{U\in\Omega_{j-1}\mid\|U-U^{(j-1)}\|_2^2>\delta^2D\}$. M is the smallest integer such that $\Omega_{M+1}=\emptyset$.

Let $A_j=\Omega_j\setminus\Omega_{j+1}$ ($j=0,\dots,M$). Since $A_j\subseteq\{U\in\mathcal{E}^m\mid\|U-U^{(j-1)}\|_2^2\le\delta^2D\}$ and $\|U-U'\|^2=\sum_{j=1}^m\|U_{:,j}-U'_{:,j}\|^2$, we have that

$$
|A_j|\le\sum_{i=0}^D(a-1)^i\binom{m}{i}.
$$

Since A_j $(j=0,\dots,M)$ are disjoint and their union is $\mathcal{E}^m$, we have

$$
\sum_{j=0}^M|A_j|=|\mathcal{E}^m|=a^m.
$$

Hence

$$
(M+1)\sum_{i=0}^D(a-1)^i\binom{m}{i}\ge a^m.
$$

Equivalently, we have that

$$
M+1\ge\frac{a^m}{\sum_{i=0}^D(a-1)^i\binom{m}{i}}.
$$

To evaluate the (inverse of) RHS $\frac{\sum_{i=0}^{D}(a-1)^i\binom{m}{i}}{a^m}$, we consider an i.i.d. sequence $(X_i)_{i=1}^m$ where $P(X_i=1)=(a-1)/a$ and $P(X_i=0)=1/a$. Then, by the Chernoff-Hoeffding bound, we have that

$$\begin{aligned}\frac{\sum_{i=0}^{D}(a-1)^i\binom{m}{i}}{a^m} &= P(\sum_{i=1}^{m} X_i \leq D) \\ &\leq \left(\frac{(a-1)/a}{D/m}\right)^D \left(\frac{1/a}{1-D/m}\right)^{m-D} \\ &\leq 4^{-m/4}(2/a)^{m/2} = (1/a)^{m/2}.\end{aligned}$$

Therefore, $M+1 \geq a^{m/2} \geq a^{m/4}+1$.

□