Proof details

Chanwoo Lee

1 Proofs of Main Theorems

Proof of Lemma **??**. Recall that we denote $\mathcal{E}_k$ as the m-way partition

$$\mathcal{E}_k = \{\bigtimes_{a=1}^{m} z^{-1}(j_a)\colon (j_1,\ldots,j_m)\in[k]^m\},$$

where $z\colon [d]\to[k]$ is the canonical clustering function such that $z(i)=\lceil ki/d\rceil$, for all $i\in[d]$. For a given partition $\bigtimes_{a=1}^{m} z^{-1}(j_a)\in\mathcal{E}_k$, fix any index $(i_1^0,\ldots,i_m^0)\in\bigtimes_{a=1}^{m} z^{-1}(j_a)$. Then, we have

$$\|(i_1,\ldots,i_m)-(i_1^0,\ldots,i_m^0)\|_\infty \leq \frac{d}{k}, \tag{1}$$

for all $(i_1,\ldots,i_m)\in\bigtimes_{a=1}^{m} z^{-1}(j_a)$. We define the blockwise ℓ-degree polynomial tensor $\mathcal{B}$ based on the partition $\mathcal{E}_k$ as

$$\mathcal{B}(i_1,\ldots,i_m)=\mathcal{P}^{j_1,\ldots,j_m}_{\min(\lfloor\alpha\rfloor,\ell)}\left(\frac{i_1-i_1^0}{d},\ldots\frac{i_m-i_m^0}{d}\right), \text{for all } (i_1,\ldots,i_m)\in\bigtimes_{a=1}^{m} z^{-1}(j_a),$$

where $\mathcal{P}^{j_1,\ldots,j_m}_{\min(\lfloor\alpha\rfloor,\ell)}$ is a ℓ-degree polynomial function satisfying

$$\left|f\left(\frac{i_1}{d},\ldots,\frac{i_m}{d}\right)-\mathcal{P}^{j_1,\ldots,j_m}_{\min(\lfloor\alpha\rfloor,\ell)}\left(\frac{i_1-i_1^0}{d},\ldots\frac{i_m-i_m^0}{d}\right)\right|\leq C\left\|\left(\frac{i_1-i_1^0}{d},\ldots,\frac{i_m-i_m^0}{d}\right)\right\|_\infty^{\min(\alpha,\ell+1)}, \tag{2}$$

for all $(i_1,\ldots,i_m)\in\bigtimes_{a=1}^{m} z^{-1}(j_a)$. Notice that we can always find such polynomial function by α-Hölder smoothness of the generating function f. Based on the construction of blockwise ℓ-degree polynomial tensor $\mathcal{B}$, we have

$$\begin{aligned}
\frac{1}{d^m}\|\Theta-\mathcal{B}\|_F^2 &= \frac{1}{d^m}\sum_{(i_1,\ldots,i_m)\in[d]^m}|\Theta(i_1,\ldots,i_m)-\mathcal{B}(i_1,\ldots,i_m)|^2\\
&=\frac{1}{d^m}\sum_{(j_1,\ldots,j_m)\in[k]^m}\ \sum_{(i_1,\ldots,i_m)\in\times_{a=1}^m z^{-1}(j_a)}\left|f\left(\frac{i_1}{d},\ldots,\frac{i_m}{d}\right)-\mathcal{P}^{j_1,\ldots,j_m}_{\min(\lfloor\alpha\rfloor,\ell)}\left(\frac{i_1-i_1^0}{d},\ldots\frac{i_m-i_m^0}{d}\right)\right|^2\\
&\lesssim\frac{1}{d^m}\sum_{(j_1,\ldots,j_m)\in[k]^m}\ \sum_{(i_1,\ldots,i_m)\in\times_{a=1}^m z^{-1}(j_a)}\left\|\left(\frac{i_1-i_1^0}{d},\ldots,\frac{i_m-i_m^0}{d}\right)\right\|_\infty^{2\min(\alpha,\ell+1)}\\
&\leq\frac{1}{k^{2\min(\alpha,\ell+1)}},
\end{aligned}$$

where the first inequality uses (2) and the second inequality is from (1).

□

Proof of Theorem **??**. The proof is similar to theorem 2.1 on note 030721. By Theorem **??**, there exists a blockwise polynomial tensor $\mathcal{B}\in\mathscr{B}(k,\ell)$ such that

$$\|\mathcal{B}-\Theta\|_F^2\lesssim\frac{d^m m^2}{k^{2\min(\alpha,\ell)}}. \tag{3}$$

By the triangle inequality,

$$\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \Theta \circ \pi\|_F^2 \leq 2\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \mathcal{B} \circ \pi\|_F^2 + 2\underbrace{\|\mathcal{B} \circ \pi - \Theta \circ \pi\|_F^2}_{\text{Lemma ??}}. \tag{4}$$

Therefore, it suffices to bound $\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \mathcal{B} \circ \pi\|_F^2$. By the global optimality of least-square estimator, we have

$$\begin{aligned}
\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \mathcal{B} \circ \pi\|_F &\leq \left\langle \frac{\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \mathcal{B} \circ \pi}{\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \mathcal{B} \circ \pi\|_F},\ \mathcal{E} + (\Theta \circ \pi - \mathcal{B} \circ \pi) \right\rangle \\
&\leq \sup_{\pi,\pi' \colon [d]\to[d]} \sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle + \underbrace{\|\mathcal{B} \circ \pi - \Theta \circ \pi\|_F}_{\text{Lemma ??}}.
\end{aligned}$$

Now, for fixed π, π', the space embedding $\mathscr{B}(k,\ell) \subset \mathbb{R}^{(\ell+m)^\ell k^m}$ implies the space embedding $\{(\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi) \colon \mathcal{B}, \mathcal{B}' \in \mathscr{B}(k,\ell)\} \subset \mathbb{R}^{2(\ell+m)^\ell k^m}$. To be specific, let $\boldsymbol{P}$ and $\boldsymbol{P}'$ be permutation matrices corresponding to permutations π and π' respectively. We express vectorized blockwise degree-ℓ polynomial tensors, $\mathrm{vec}(\mathcal{B})$ and $\mathrm{vec}(\mathcal{B}')$ by $\boldsymbol{X\beta}$ and $\boldsymbol{X\beta}'$ respectively, where $\boldsymbol{X} \in \mathbb{R}^{d^m \times k^m(k+m)^\ell}$ is a design matrices consisting of covariates $(1/d, \ldots, d/d)^m$ and $\boldsymbol{\beta}$ and $\boldsymbol{\beta}' \in \mathbb{R}^{k^m(k+m)^\ell}$ are corresponding coefficient vectors. Notice that the number of coefficients for ℓ-polynomial m-multivariate function is $\binom{\ell+m}{\ell}$. We choose to use $(k+m)^\ell$ coefficients for each block for the notational simplicity. Therefore, we rewrite the inner product,

$$\begin{aligned}
\left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle &= \left\langle \frac{(\boldsymbol{P}')^{\otimes m}\mathrm{vec}(\mathcal{B}') - (\boldsymbol{P})^{\otimes m}\mathrm{vec}(\mathcal{B})}{\|(\boldsymbol{P}')^{\otimes m}\mathrm{vec}(\mathcal{B}') - (\boldsymbol{P})^{\otimes m}\mathrm{vec}(\mathcal{B})\|_F}, \mathcal{E} \right\rangle \\
&= \left\langle \frac{(\boldsymbol{P}')^{\otimes m}\boldsymbol{X\beta}' - (\boldsymbol{P})^{\otimes m}\boldsymbol{X\beta}}{\|(\boldsymbol{P}')^{\otimes m}\boldsymbol{X\beta}' - (\boldsymbol{P})^{\otimes m}\boldsymbol{X\beta}\|_F}, \mathcal{E} \right\rangle \\
&= \left\langle \frac{\boldsymbol{Ac}}{\|\boldsymbol{Ac}\|_F}, \mathcal{E} \right\rangle,
\end{aligned}$$

where we define $\boldsymbol{A} := \begin{pmatrix} \boldsymbol{P}' & -\boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{X} & 0 \\ 0 & \boldsymbol{X} \end{pmatrix} \in \mathbb{R}^{d^m \times 2k^m(k+m)^\ell}$ and $\boldsymbol{c} := \begin{pmatrix} \boldsymbol{\beta}' \\ \boldsymbol{\beta} \end{pmatrix} \in \mathbb{R}^{2k^m(k+m)^\ell}$. By Lemma 2.1, we have

$$\sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle \leq \sup_{\boldsymbol{c} \in \mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\ e \right\rangle, \tag{5}$$

where $e \in \mathbb{R}^{2k^m(k+m)^\ell}$ is a vector consisting of i.i.d. sub-Gaussian entries with variance proxy σ^2. Therefore, the union bound of Gaussian maxima over countable set $\{\pi, \pi' \colon [d] \to [d]\}$, we obtain

$$\begin{aligned}
\mathbb{P}\left(\sup_{\pi,\pi' \colon [d]\to[d]} \sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle \geq t \right) &\leq \sum_{\pi,\pi' \in [d]^d} \mathbb{P}\left(\sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle \geq t \right) \\
&\leq d^d \mathbb{P}\left(\sup_{\boldsymbol{c} \in \mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\ e \right\rangle \geq t \right) \\
&\leq \exp\left(-\frac{t^2}{8\sigma^2} + k^m(\ell+m)^\ell \log 6 + d\log d \right),
\end{aligned}$$

where the second inequality is from (5) and the last inequality is from Theorem 1.19 [1]. Setting $t = C\sigma\sqrt{k^m(\ell+m)^\ell + d\log d}$ for sufficiently large $C > 0$ gives,

$$\sup_{\pi,\pi' \colon [d]\to[d]} \sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle \lesssim \sigma\sqrt{k^m(\ell+m)^\ell + d\log d}, \tag{6}$$

with high probability.

Finally, combining the inequalities (3), (4) and (6) yields the desired conclusion

$$\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \Theta \circ \pi\|_F^2 \lesssim \sigma^2 \left(k^m(\ell+m)^\ell + d\log d\right) + \frac{d^m m^2}{k^{2\min(\alpha,\ell)}}.$$

$\square$

2 Technical Lemmas

Lemma 2.1 (Gaussian maxima under full embedding)**.** Let $\boldsymbol{A} \in \mathbb{R}^{d_1 \times d_2}$ be a deterministic matrix with rank $r \le \min(d_1, d_2)$. Let $\boldsymbol{y} \in \mathbb{R}^{d_2}$ be a sub-Gaussian random vector with variance proxy σ^2. Then, there exists a sub-Gaussian random vector $\boldsymbol{x} \in \mathbb{R}^r$ with variance prosy σ^2 such that

$$\max_{\boldsymbol{c}\in\mathbb{R}^{d_2}} \left\langle \frac{\boldsymbol{Ac}}{\|\boldsymbol{Ac}\|_2}, \boldsymbol{y} \right\rangle = \max_{\boldsymbol{c}\in\mathbb{R}^{r}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2}, \boldsymbol{x} \right\rangle.$$

Proof. Let $\boldsymbol{u}_i \in \mathbb{R}^{d_1}, \boldsymbol{v}_j \in \mathbb{R}^{d_2}$ singular vectors and $\lambda_i \in \mathbb{R}$ be singular values of of $\boldsymbol{A}$ such that $\boldsymbol{A} = \sum_{i=1}^r \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T$. Then for any $\boldsymbol{c} \in \mathbb{R}^r$, we have

$$\boldsymbol{Ac} = \sum_{i=1}^r \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T \boldsymbol{c} = \sum_{i=1}^r \lambda_i (\boldsymbol{v}_i^T \boldsymbol{c}) \boldsymbol{u}_i = \sum_{i=1}^r \alpha_i \boldsymbol{u}_i,$$

where $\boldsymbol{\alpha}(\boldsymbol{c}) = (\alpha_1, \ldots, \alpha_r)^T := \left(\lambda_1(\boldsymbol{v}_1^T\boldsymbol{c}), \ldots, \lambda_r(\boldsymbol{v}_r^T\boldsymbol{c})\right)^T \in \mathbb{R}^r$. Notice that $\boldsymbol{\alpha}(\boldsymbol{c})$ covers $\mathbb{R}^r$ in the sense that $\{\boldsymbol{\alpha}(\boldsymbol{c}) \colon \boldsymbol{c} \in \mathbb{R}^r\} = \mathbb{R}^r$. Therefore, we have

$$\begin{aligned}
\max_{\boldsymbol{c}\in\mathbb{R}^r} \left\langle \frac{\boldsymbol{Ac}}{\|\boldsymbol{Ac}\|_2}, \boldsymbol{y} \right\rangle &= \max_{\boldsymbol{c}\in\mathbb{R}^r} \sum_{i=1}^r \frac{\alpha_i}{\|\boldsymbol{\alpha}(\boldsymbol{c})\|_2} \boldsymbol{u}_i^T \boldsymbol{y} \\
&= \max_{\boldsymbol{c}\in\mathbb{R}^r} \left\langle \frac{\boldsymbol{\alpha}(\boldsymbol{c})}{\|\boldsymbol{\alpha}(\boldsymbol{c})\|_2}, \boldsymbol{x} \right\rangle \\
&= \max_{\boldsymbol{c}\in\mathbb{R}^r} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2}, \boldsymbol{x} \right\rangle,
\end{aligned}$$

where we define $\boldsymbol{x} = (\boldsymbol{u}_1^T\boldsymbol{y}, \ldots, \boldsymbol{u}_r^T\boldsymbol{y})^T \in \mathbb{R}^r$. Since $\boldsymbol{u}_i^T\boldsymbol{y}$ is sub-Gaussian with variance proxy σ^2 because of orthonormality of $\boldsymbol{u}_i$, the proof is completed. $\square$

Remark 1. Let $\boldsymbol{x}, \in \mathbb{R}^r, \boldsymbol{y} \in \mathbb{R}^d$ be Gaussian random vectors whose entries are i.i.d. drawn from $N(0, \sigma^2)$. Define two Gaussian maximums

$$F(\boldsymbol{x}) \stackrel{\text{def}}{=} \max_{\boldsymbol{c}\in\mathbb{R}^r} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2}, \boldsymbol{x} \right\rangle, \qquad G(\boldsymbol{x}) \stackrel{\text{def}}{=} \max_{\boldsymbol{c}\in\mathbb{R}^r} \left\langle \frac{\boldsymbol{Ac}}{\|\boldsymbol{Ac}\|_2}, \boldsymbol{y} \right\rangle,$$

Then $F(\boldsymbol{x}) = G(\boldsymbol{y})$ in distribution. This holds because $(\boldsymbol{u}_1^T\boldsymbol{y}, \ldots, \boldsymbol{u}_r^T\boldsymbol{y})$ is again Gaussian random vectors whose entries are i.i.d. drawn from $N(0, \sigma^2)$.

proof lemma1, 2 and theorem 3

References

[1] Jan-Christian Hitter Phillippe Rigollet. High dimensional statistics. *Lecture notes for course 18S997*, 2015.