Functional Tensor Regression

Tongyu Li, Fang Yao and Anru R. Zhang

Peking University and Duke University

Supplementary Material

This supplementary material contains the proofs of the main theorems and technical lemmas, and additional numerical results for observations on a non-uniform grid.

S1 Proofs

Proof of Theorem 1

Notice that

$$\begin{aligned}
\|\boldsymbol{\Theta}^{k+1}-\boldsymbol{\Theta}\|_{\mathrm{F}} &= \|\mathcal{H}_{\boldsymbol{r}}(\check{\boldsymbol{\Theta}}^{k+1})-\boldsymbol{\Theta}\|_{\mathrm{F}} \quad \text{by (2.16)} \\
&\le \|\mathcal{H}_{\boldsymbol{r}}(\check{\boldsymbol{\Theta}}^{k+1})-\check{\boldsymbol{\Theta}}^{k+1}\|_{\mathrm{F}}+\|\check{\boldsymbol{\Theta}}^{k+1}-\boldsymbol{\Theta}\|_{\mathrm{F}} \\
&\le \{(D+1)^{1/2}+1\}\|\check{\boldsymbol{\Theta}}^{k+1}-\boldsymbol{\Theta}\|_{\mathrm{F}} \quad \text{by (2.15)} \\
&\le \{(D+1)^{1/2}+1\}R_l^{-1/2}p_0^{1/2}n^{-1/2}\|\mathscr{X}(\check{\boldsymbol{\Theta}}^{k+1}-\boldsymbol{\Theta})\| \quad \text{by (3.19)} \quad .
\end{aligned} \tag{S1}$$

To separate out the noise term, introduce the minimizer with $\boldsymbol{y}-\boldsymbol{\varepsilon}$ in place of $\boldsymbol{y}$, i.e.,

$$\tilde{\boldsymbol{\Theta}}^{k+1} = \mathscr{R}_{\boldsymbol{\Theta}^k}(\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{Z}^*\mathscr{Z}\mathscr{R}_{\boldsymbol{\Theta}^k} + n\rho\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}^k})^{-1}\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{Z}^*(\boldsymbol{y}-\boldsymbol{\varepsilon}).$$

Since $\boldsymbol{y}-\boldsymbol{\varepsilon} = \big(\int_{\mathbb{T}}\langle\boldsymbol{\mathcal{X}}_i(t),\boldsymbol{\mathcal{B}}(t)\rangle\,\mathrm{d}t\big)_{i=1,\cdots,n} = \mathscr{Z}\boldsymbol{\Theta}+\boldsymbol{\delta}$, it follows that

$$\begin{aligned}
&(2n)^{-1}\|\mathscr{Z}(\boldsymbol{\Theta}-\tilde{\boldsymbol{\Theta}}^{k+1})+\boldsymbol{\delta}\|^2 + \rho\langle\mathscr{A}\tilde{\boldsymbol{\Theta}}^{k+1},\tilde{\boldsymbol{\Theta}}^{k+1}\rangle\\
&= (2n)^{-1}\|\boldsymbol{y}-\boldsymbol{\varepsilon}-\mathscr{Z}\tilde{\boldsymbol{\Theta}}^{k+1}\|^2 + \rho\langle\mathscr{A}\tilde{\boldsymbol{\Theta}}^{k+1},\tilde{\boldsymbol{\Theta}}^{k+1}\rangle\\
&\le (2n)^{-1}\|\boldsymbol{y}-\boldsymbol{\varepsilon}-\mathscr{Z}\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\|^2 + \rho\langle\mathscr{A}\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta},\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\rangle \quad \text{recalling (2.11)}\\
&= (2n)^{-1}\|\mathscr{Z}(\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta})+\boldsymbol{\delta}\|^2 + \rho\langle\mathscr{A}\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta},\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\rangle.
\end{aligned}$$

Using the inequalities that

$$\|\mathscr{Z}(\boldsymbol{\Theta}-\tilde{\boldsymbol{\Theta}}^{k+1})+\boldsymbol{\delta}\|^2 \ge 2^{-1}\|\mathscr{Z}(\boldsymbol{\Theta}-\tilde{\boldsymbol{\Theta}}^{k+1})\|^2 - \|\boldsymbol{\delta}\|^2,$$

$$\|\mathscr{Z}(\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta})+\boldsymbol{\delta}\|^2 \le 2\|\mathscr{Z}(\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta})\|^2 + 2\|\boldsymbol{\delta}\|^2,$$

(3.17) and (3.21), we obtain that

$$\|\mathscr{Z}(\boldsymbol{\Theta}-\tilde{\boldsymbol{\Theta}}^{k+1})\|^2 \le 4nR_u p_0^{-1}\|\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\|_{\mathrm{F}}^2 + 6\|\boldsymbol{\delta}\|^2 + 8n\rho C_m. \tag{S2}$$

Besides, by writing

$$\mathscr{Z}(\check{\boldsymbol{\Theta}}^{k+1}-\tilde{\boldsymbol{\Theta}}^{k+1}) = \mathscr{Z}\mathscr{R}_{\boldsymbol{\Theta}^k}\big(\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{Z}^*\mathscr{Z}\mathscr{R}_{\boldsymbol{\Theta}^k} + n\rho\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}^k}\big)^{-1}\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{Z}^*\boldsymbol{\varepsilon},$$

it can be seen that

$$\begin{aligned}\|\mathscr{X}(\check{\boldsymbol{\Theta}}^{k+1}-\tilde{\boldsymbol{\Theta}}^{k+1})\|^2 &\le \boldsymbol{\varepsilon}^\top\mathscr{X}\mathscr{R}_{\boldsymbol{\Theta}^k}\left(\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{X}^*\mathscr{X}\mathscr{R}_{\boldsymbol{\Theta}^k}+n\rho\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}^k}\right)^{-1}\mathscr{R}^*_{\boldsymbol{\Theta}^k}\mathscr{X}^*\boldsymbol{\varepsilon}\\ &\le p_0(c_mn\rho)^{-1}\|\mathscr{P}_{\boldsymbol{\Theta}^k}\mathscr{X}^*\boldsymbol{\varepsilon}\|_{\mathrm{F}}^2 \quad \text{by (2.8)}\\ &\le c_m^{-1}p_0n^{-1}\rho^{-1}\left\|(\mathscr{X}^*\boldsymbol{\varepsilon})_{\max(2\boldsymbol{r})}\right\|_{\mathrm{F}}^2\end{aligned} \tag{S3}$$

where the last step invokes the fact that $\mathrm{T}_{\boldsymbol{\Theta}^k}\mathbb{M}_{\boldsymbol{r}}\subset\bigcup_{\boldsymbol{s}\le 2\boldsymbol{r}}\mathbb{M}_{\boldsymbol{s}}$ (Luo and Zhang, 2023, Lemma 2). The combination of (S2) and (S3) leads to

$$\begin{aligned}&\|\mathscr{X}(\check{\boldsymbol{\Theta}}^{k+1}-\boldsymbol{\Theta})\|^2\le 2\|\mathscr{X}(\boldsymbol{\Theta}-\tilde{\boldsymbol{\Theta}}^{k+1})\|^2+2\|\mathscr{X}(\check{\boldsymbol{\Theta}}^{k+1}-\tilde{\boldsymbol{\Theta}}^{k+1})\|^2\\ &\le 8nR_up_0^{-1}\|\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\|_{\mathrm{F}}^2+12\|\boldsymbol{\delta}\|^2+16n\rho C_m+2c_m^{-1}p_0n^{-1}\rho^{-1}\left\|(\mathscr{X}^*\boldsymbol{\varepsilon})_{\max(2\boldsymbol{r})}\right\|_{\mathrm{F}}^2,\end{aligned}$$

so

$$\begin{aligned}\|\mathscr{X}(\check{\boldsymbol{\Theta}}^{k+1}-\boldsymbol{\Theta})\| &\le 8^{1/2}n^{1/2}R_u^{1/2}p_0^{-1/2}\|\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\|_{\mathrm{F}}\\ &\quad+\left\{12\|\boldsymbol{\delta}\|^2+16n\rho C_m+2c_m^{-1}p_0n^{-1}\rho^{-1}\left\|(\mathscr{X}^*\boldsymbol{\varepsilon})_{\max(2\boldsymbol{r})}\right\|_{\mathrm{F}}^2\right\}^{1/2}.\end{aligned}$$

Plugging this into (S1), we have

$$\|\boldsymbol{\Theta}^{k+1}-\boldsymbol{\Theta}\|_{\mathrm{F}}\le\{(D+1)^{1/2}+1\}(8^{1/2}R_u^{1/2}R_l^{-1/2}\|\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\|_{\mathrm{F}}+R_l^{-1/2}\eta).$$

Due to Luo and Zhang (2023, Lemma 9),

$$\|\boldsymbol{\Theta}-\mathscr{P}_{\boldsymbol{\Theta}^k}\boldsymbol{\Theta}\|_{\mathrm{F}}\le(D+1)\lambda_{\min}^{-1}\|\boldsymbol{\Theta}^k-\boldsymbol{\Theta}\|_{\mathrm{F}}^2,$$

and thus the proof is complete. □

Proof of Corollary 1

The argument is similar to the proof of Luo and Zhang (2023, Corollary 1), so we omit it here. □

Proof of Theorem 2

By writing the Riemann sum as the integral of a piece-wise constant function and recalling (2.5), it can be easily seen that the approximation errors satisfy that $|\delta_i| = \mathcal{O}_{\mathrm{pr}}(p_0^{-\kappa} C_m^{1/2})$, and thus

$$\|\boldsymbol{\delta}\|^2/n = \mathcal{O}_{\mathrm{pr}}(p_0^{-2\kappa} C_m).$$

Following the proof of Luo and Zhang (2023, Lemma 12),

$$\begin{aligned}\big\|(\mathscr{X}^*\boldsymbol{\varepsilon})_{\max(2\boldsymbol{r})}\big\|_{\mathrm{F}} &= \max_{\boldsymbol{U}_d\in\mathbb{O}_{p_d,2r_d},\, d=0,1,\cdots,D}\ \max_{\mathcal{T}\in\mathbb{R}^{p_0\times p_1\times\cdots\times p_D}:\|\mathcal{T}\|_{\mathrm{F}}\le 1} \langle \mathcal{T}\times_{d=0}^{D}(U_dU_d^\top), \mathscr{X}^*\boldsymbol{\varepsilon}\rangle \\ &= \max_{\boldsymbol{U}_d\in\mathbb{O}_{p_d,2r_d},\, d=0,1,\cdots,D}\ \max_{\mathcal{T}\in\mathbb{R}^{p_0\times p_1\times\cdots\times p_D}:\|\mathcal{T}\|_{\mathrm{F}}\le 1} \boldsymbol{\varepsilon}^\top\mathscr{X}\{\mathcal{T}\times_{d=0}^{D}(U_dU_d^\top)\}.\end{aligned}$$

Since Lemma 1 provides the bounds for the operator norm of $\mathscr{X}$, the proof of Han et al. (2022, Theorem 4.2) implies that

$$\big\|(\mathscr{X}^*\boldsymbol{\varepsilon})_{\max(2\boldsymbol{r})}\big\|_{\mathrm{F}}^2 = \mathcal{O}_{\mathrm{pr}}\left\{np_0^{-1}\left(\sum_{d=0}^{D} p_d r_d + \prod_{d=0}^{D} r_d\right)\right\}.$$

Plugging these into (3.23) leads to the desired result, and (3.24) follows directly. The proof is then complete. □

Proof of Theorem 3

After rescaling the regression model using (3.17) and (3.19), it follows from Luo and Zhang (2023, Lemma 12) and the proof of Luo and Zhang (2023, Theorem 2) that when $n = \prod_{d=0}^{D} p_d$ we can choose $\boldsymbol{\mathcal{E}}_0 = \boldsymbol{\mathcal{E}}_{\max(2\boldsymbol{r})}$ for some $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{p_0 \times p_1 \times \cdots \times p_D}$ with i.i.d. $\mathcal{N}(0,1)$ entries. In what follows, c_1 and c_2 denote universal positive constants. By Sudakov's minoration inequality (Vershynin, 2018, Theorem 7.4.1),

$$
\begin{aligned}
E(\|\boldsymbol{\mathcal{E}}_{\max(2\boldsymbol{r})}\|_{\mathrm{F}}) &= E\big(\max_{\boldsymbol{U}_d \in \mathbb{O}_{p_d, 2r_d},\, d=0,1,\cdots,D} \|\boldsymbol{\mathcal{E}} \times_{d=0}^{D} \boldsymbol{U}_d^\top\|_{\mathrm{F}}\big) \\
&\geq c_1 \epsilon \{\log N(\prod \mathbb{O}_{p_d, 2r_d}, \epsilon)\}^{1/2}, \quad \epsilon > 0
\end{aligned}
$$

where $N(\prod \mathbb{O}_{p_d, 2r_d}, \epsilon)$ is the ϵ-covering number of $\prod_{d=0}^{D} \mathbb{O}_{p_d, 2r_d}$. It can be seen that

$$
N(\prod \mathbb{O}_{p_d, 2r_d}, \epsilon) \geq N(\prod \mathbb{O}_{p_d, r_d}, \epsilon) \geq (c_2/\epsilon)^{\sum_{d=0}^{D}(p_d - r_d) r_d},
$$

and thus $E(\|\boldsymbol{\mathcal{E}}_{\max(2\boldsymbol{r})}\|_{\mathrm{F}}) \geq 2c_0 \sum_{d=0}^{D}(p_d - r_d) r_d$. Besides, with some fixed $\boldsymbol{U}_d \in \mathbb{O}_{p_d, 2r_d}$,

$$
E(\|\boldsymbol{\mathcal{E}}_{\max(2\boldsymbol{r})}\|_{\mathrm{F}}^2) \geq E(\|\boldsymbol{\mathcal{E}} \times_{d=0}^{D} \boldsymbol{U}_d^\top\|_{\mathrm{F}}^2) = \prod_{d=0}^{D} (2r_d).
$$

Combining the two bounds completes the proof. $\square$

Proof of Lemma 1

By the law of large numbers,

$$n^{-1}\|\mathscr{Z}\boldsymbol{\Upsilon}\|^2 = n^{-1}(\langle \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\Upsilon}\rangle^2 + \cdots + \langle \boldsymbol{\mathcal{Z}}_n, \boldsymbol{\Upsilon}\rangle^2) \to E(\langle \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\Upsilon}\rangle^2)$$

in probability as $n \to \infty$. Hence we only need to bound $E(\langle \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\Upsilon}\rangle^2)$. For convenience, let $\boldsymbol{\Gamma}_\ell$ and $\boldsymbol{\mathcal{E}}$ be the $p_0 \times p_1 \times \cdots \times p_D$ tensors such that $[\boldsymbol{\Gamma}_\ell]_{k,j_1,\ldots,j_D} = [\boldsymbol{\Xi}_{k+(\ell-1)p_0}]_{j_1,\ldots,j_D}$ and $[\boldsymbol{\mathcal{E}}]_{j,j_1,\ldots,j_D} = [\boldsymbol{\mathcal{E}}_{1j}]_{j_1,\ldots,j_D}$, and let $\boldsymbol{\Delta}^{1/2}$ be the diagonal matrix constructed from $(\Delta t_j)^{1/2}$, $j = 1, \cdots, p_0$. We have

$$\begin{aligned}
E(\langle \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\Upsilon}\rangle^2) &= E\left\{\left\langle \sum_{\ell=1}^{\infty} \boldsymbol{\Gamma}_\ell \times_0 (\boldsymbol{\Delta}^{1/2}\boldsymbol{\Phi}_\ell) + \boldsymbol{\mathcal{E}} \times_0 \boldsymbol{\Delta}, \boldsymbol{\Upsilon}\right\rangle^2\right\} \quad \text{by (2.1) and (2.2)} \\
&= \sum_{\ell=1}^{\infty} E\{\langle \boldsymbol{\Gamma}_\ell \times_0 (\boldsymbol{\Delta}^{1/2}\boldsymbol{\Phi}_\ell), \boldsymbol{\Upsilon}\rangle^2\} + E(\langle \boldsymbol{\mathcal{E}} \times_0 \boldsymbol{\Delta}, \boldsymbol{\Upsilon}\rangle^2) \quad \text{by uncorrelatedness} \\
&= \sum_{\ell=1}^{\infty} E\{\langle \boldsymbol{\Gamma}_\ell, \boldsymbol{\Upsilon} \times_0 (\boldsymbol{\Phi}_\ell^\top \boldsymbol{\Delta}^{1/2})\rangle^2\} + E(\langle \boldsymbol{\mathcal{E}}, \boldsymbol{\Upsilon} \times_0 \boldsymbol{\Delta}\rangle^2) \quad \text{by algebra} \\
&\le \sum_{\ell=1}^{\infty} \frac{A}{\{1+(\ell-1)p_0\}^a} \left\|\boldsymbol{\Upsilon} \times_0 \boldsymbol{\Phi}_\ell^\top \times_0 \boldsymbol{\Delta}^{1/2}\right\|_{\mathrm{F}}^2 + \sigma_X^2 \|\boldsymbol{\Upsilon} \times_0 \boldsymbol{\Delta}\|_{\mathrm{F}}^2 \quad \text{by Assumptions 1 and 3} \\
&\le \sum_{\ell=1}^{\infty} \frac{A}{\{1+(\ell-1)p_0\}^a} \frac{C_0}{p_0} \|\boldsymbol{\Upsilon} \times_0 \boldsymbol{\Phi}_\ell^\top\|_{\mathrm{F}}^2 + \sigma_X^2 \frac{C_0^2}{p_0^2} \|\boldsymbol{\Upsilon}\|_{\mathrm{F}}^2 \quad \text{by (2.5)} \\
&\le \sum_{\ell=1}^{\infty} \frac{A}{\{1+(\ell-1)p_0\}^a} \frac{C_0}{p_0} C_\varphi \|\boldsymbol{\Upsilon}\|_{\mathrm{F}}^2 + \sigma_X^2 \frac{C_0^2}{p_0^2} \|\boldsymbol{\Upsilon}\|_{\mathrm{F}}^2 \quad \text{by Assumption 2} \\
&\le \left\{\left(1 + \frac{a}{a-1}\frac{1}{p_0^a}\right) A C_0 C_\varphi + \frac{C_0^2 \sigma_X^2}{p_0}\right\} \frac{\|\boldsymbol{\Upsilon}\|_{\mathrm{F}}^2}{p_0},
\end{aligned}$$

where the last step uses the fact that $\sum_{\ell=2}^{\infty}\{1/p_0 + (\ell-1)\}^{-a} < 1 + \int_1^\infty u^{-a}\,\mathrm{d}u = a/(a-1)$. This completes the proof of (3.17). In a similar manner, writing $\boldsymbol{\Gamma}_0$ for the $r_0 \times p_1 \times \cdots \times p_D$ tensor such that $[\boldsymbol{\Gamma}_0]_{k,j_1,\ldots,j_D} =$

$[\boldsymbol{\Xi}_k]_{j_1,\dots,j_D}$, it can be seen that

$$\begin{aligned}
E(\langle \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\Upsilon}\rangle^2) &\geq E(\langle \boldsymbol{\Gamma}_0, \boldsymbol{\Upsilon} \times_0 (\boldsymbol{\Phi}_0^\top \boldsymbol{\Delta}^{1/2})\rangle^2) \\
&\geq \frac{1}{Ar_0^a} \left\| \boldsymbol{\Upsilon} \times_0 \boldsymbol{\Phi}_0^\top \times_0 \boldsymbol{\Delta}^{1/2} \right\|_{\mathrm{F}}^2 \quad \text{by Assumption 1} \\
&\geq \frac{1}{Ar_0^a} \frac{1}{C_0 p_0} \left\| \boldsymbol{\Upsilon} \times_0 \boldsymbol{\Phi}_0^\top \right\|_{\mathrm{F}}^2 \quad \text{by (2.5)} \\
&\geq \frac{c^2}{AC_0 r_0^a} \frac{\|\boldsymbol{\Upsilon}\|_{\mathrm{F}}^2}{p_0} \quad \text{by (3.18)} \quad .
\end{aligned}$$

This completes the proof of (3.19). $\square$

Proof of Lemma 2

Recall that $\boldsymbol{U}_0\boldsymbol{U}_0^\top + \boldsymbol{U}_{0\perp}\boldsymbol{U}_{0\perp}^\top = \boldsymbol{I}_{r_0}$. Using the triangle inequality,

$$\begin{aligned}
\|\boldsymbol{\Upsilon} \times_0 \boldsymbol{\Phi}^\top\|_{\mathrm{F}} &= \|\boldsymbol{\Upsilon} \times_0 (\boldsymbol{\Phi}^\top \boldsymbol{U}_0\boldsymbol{U}_0^\top) + \boldsymbol{\Upsilon} \times_0 (\boldsymbol{\Phi}^\top \boldsymbol{U}_{0\perp}\boldsymbol{U}_{0\perp}^\top)\|_{\mathrm{F}} \\
&\geq \|\boldsymbol{\Upsilon} \times_0 (\boldsymbol{\Phi}^\top \boldsymbol{U}_0\boldsymbol{U}_0^\top)\|_{\mathrm{F}} - \|\boldsymbol{\Upsilon} \times_0 (\boldsymbol{\Phi}^\top \boldsymbol{U}_{0\perp}\boldsymbol{U}_{0\perp}^\top)\|_{\mathrm{F}} \\
&\geq \sigma_{r_0}(\boldsymbol{\Phi}^\top \boldsymbol{U}_0)\|\boldsymbol{\Upsilon} \times_0 \boldsymbol{U}_0^\top\|_{\mathrm{F}} - \sigma_1(\boldsymbol{\Phi}^\top \boldsymbol{U}_{0\perp})\|\boldsymbol{\Upsilon} \times_0 \boldsymbol{U}_{0\perp}^\top\|_{\mathrm{F}} \\
&= \sigma_{r_0}(\boldsymbol{\Phi}^\top \boldsymbol{U}_0)\|\boldsymbol{\Upsilon} \times_0 (\boldsymbol{U}_0\boldsymbol{U}_0^\top)\|_{\mathrm{F}} - \sigma_1(\boldsymbol{\Phi}^\top \boldsymbol{U}_{0\perp})\|\boldsymbol{\Upsilon} \times_0 (\boldsymbol{U}_{0\perp}\boldsymbol{U}_{0\perp}^\top)\|_{\mathrm{F}} \\
&= \sigma_{r_0}(\boldsymbol{\Phi}^\top \boldsymbol{U}_0)\alpha\|\boldsymbol{\Upsilon}\|_{\mathrm{F}} - \sigma_1(\boldsymbol{\Phi}^\top \boldsymbol{U}_{0\perp})(1-\alpha^2)^{1/2}\|\boldsymbol{\Upsilon}\|_{\mathrm{F}}.
\end{aligned}$$

This completes the proof. $\square$

Proof of Proposition 1

Recalling (2.12), we have

$$\begin{aligned}\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}},(\boldsymbol{D}_d)_{d=0}^D\big) &= \mathscr{A}\Big\{\boldsymbol{\mathcal{C}}\times_{d=0}^D \boldsymbol{U}_d + \sum_{d=0}^D \mathscr{T}_d(\boldsymbol{U}_{d\perp}\boldsymbol{D}_d\boldsymbol{W}_d^\top)\Big\}\\ &= \boldsymbol{\mathcal{C}}\times_0(\boldsymbol{A}\boldsymbol{U}_0)\times_{d=1}^D \boldsymbol{U}_d + \sum_{d=0}^D \mathscr{T}_d(\boldsymbol{U}_{d\perp}\boldsymbol{D}_d\boldsymbol{W}_d^\top)\times_0 \boldsymbol{A}.\end{aligned}$$

The core part of $\mathscr{R}_{\boldsymbol{\Theta}}^*\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}},(\boldsymbol{D}_d)_{d=0}^D\big)$ is

$$\begin{aligned}&\Big\{\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}},(\boldsymbol{D}_d)_{d=0}^D\big)\Big\}\times_{d=0}^D \boldsymbol{U}_d^\top\\ &= \boldsymbol{\mathcal{C}}\times_0(\boldsymbol{U}_0^\top\boldsymbol{A}\boldsymbol{U}_0)\times_{d=1}^D(\boldsymbol{U}_d^\top\boldsymbol{U}_d) + \mathscr{T}_0(\boldsymbol{U}_{0\perp}\boldsymbol{D}_0\boldsymbol{W}_0^\top)\times_0(\boldsymbol{U}_0^\top\boldsymbol{A})\times_{d=1}^D \boldsymbol{U}_d^\top\\ &\qquad + \sum_{d=1}^D \mathscr{T}_d(\boldsymbol{U}_d^\top\boldsymbol{U}_{d\perp}\boldsymbol{D}_d\boldsymbol{W}_d^\top)\times_0(\boldsymbol{U}_0^\top\boldsymbol{A})\times_{e\neq 0,d}\boldsymbol{U}_e^\top\\ &= \boldsymbol{\mathcal{C}}\times_0(\boldsymbol{U}_0^\top\boldsymbol{A}\boldsymbol{U}_0) + \mathscr{T}_0(\boldsymbol{D}_0\boldsymbol{V}_0^\top)\times_0(\boldsymbol{U}_0^\top\boldsymbol{A}\boldsymbol{U}_{0\perp})\times_{d=1}^D(\boldsymbol{U}_d^\top\boldsymbol{U}_d)\\ &= \boldsymbol{\mathcal{C}}\times_0(\boldsymbol{U}_0^\top\boldsymbol{A}\boldsymbol{U}_0) + \mathscr{T}_0(\boldsymbol{D}_0\boldsymbol{V}_0^\top)\times_0(\boldsymbol{U}_0^\top\boldsymbol{A}\boldsymbol{U}_{0\perp}).\end{aligned}$$

The 0-mode part of $\mathscr{R}_{\boldsymbol{\Theta}}^*\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}},(\boldsymbol{D}_d)_{d=0}^D\big)$ is

$$\begin{aligned}&\boldsymbol{U}_{0\perp}^\top\mathscr{M}_0\big\{\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}},(\boldsymbol{D}_d)_{d=0}^D\big)\big\}\boldsymbol{W}_0\\ &= \boldsymbol{U}_{0\perp}^\top\Big[\boldsymbol{A}\boldsymbol{U}_0\mathscr{M}_0(\boldsymbol{\mathcal{C}})(\otimes_{d=D}^1\boldsymbol{U}_d^\top) + \boldsymbol{A}\boldsymbol{U}_{0\perp}\boldsymbol{D}_0\boldsymbol{W}_0^\top + \sum_{d=1}^D \boldsymbol{A}\mathscr{M}_0\{\mathscr{T}_d(\boldsymbol{U}_{d\perp}\boldsymbol{D}_d\boldsymbol{W}_d^\top)\}\Big]\boldsymbol{W}_0\\ &= \boldsymbol{U}_{0\perp}^\top\boldsymbol{A}\boldsymbol{U}_0\mathscr{M}_0(\boldsymbol{\mathcal{C}})\boldsymbol{V}_0 + \boldsymbol{U}_{0\perp}^\top\boldsymbol{A}\boldsymbol{U}_{0\perp}\boldsymbol{D}_0 + \boldsymbol{U}_{0\perp}^\top\sum_{d=1}^D \boldsymbol{A}\mathscr{M}_0\big\{\mathscr{T}_d(\boldsymbol{U}_d^\top\boldsymbol{U}_{d\perp}\boldsymbol{D}_d\boldsymbol{W}_d^\top)\times_{e\neq 0,d}\boldsymbol{U}_e^\top\big\}\boldsymbol{V}_0\\ &= \boldsymbol{U}_{0\perp}^\top\boldsymbol{A}\boldsymbol{U}_0\mathscr{M}_0(\boldsymbol{\mathcal{C}})\boldsymbol{V}_0 + \boldsymbol{U}_{0\perp}^\top\boldsymbol{A}\boldsymbol{U}_{0\perp}\boldsymbol{D}_0.\end{aligned}$$

For $d \neq 0$, the d-mode part of $\mathscr{R}^*_{\boldsymbol{\Theta}} \mathscr{A} \mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}}, (\boldsymbol{D}_d)_{d=0}^D\big)$ is

$$
\begin{aligned}
&\boldsymbol{U}_{d\perp}^\top \mathscr{M}_d\big\{\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}}, (\boldsymbol{D}_d)_{d=0}^D\big)\big\}\boldsymbol{W}_d \\
&= \mathscr{M}_d\big[\,\big\{\mathscr{A}\mathscr{R}_{\boldsymbol{\Theta}}\big(\boldsymbol{\mathcal{C}}, (\boldsymbol{D}_d)_{d=0}^D\big)\big\} \times_d \boldsymbol{U}_{d\perp}^\top \times_{e\neq d} \boldsymbol{U}_e^\top\big]\boldsymbol{V}_d \\
&= \mathscr{M}_d\Big\{\boldsymbol{\mathcal{C}} \times_0 (\boldsymbol{U}_0^\top \boldsymbol{A}\boldsymbol{U}_0) \times_d (\boldsymbol{U}_{d\perp}^\top \boldsymbol{U}_d) \times_{e\neq 0,d} (\boldsymbol{U}_e^\top \boldsymbol{U}_e) \\
&\qquad + \sum_{c=0}^D \mathscr{T}_c(\boldsymbol{U}_{c\perp}\boldsymbol{D}_c\boldsymbol{W}_c^\top) \times_0 (\boldsymbol{U}_0^\top \boldsymbol{A}) \times_d \boldsymbol{U}_{d\perp}^\top \times_{e\neq 0,d} \boldsymbol{U}_e^\top\Big\}\boldsymbol{V}_d \\
&= \sum_{c=0}^D \mathscr{M}_d\Big\{\big(\mathscr{T}_c(\boldsymbol{U}_{c\perp}\boldsymbol{D}_c\boldsymbol{V}_c^\top) \times_{e\neq c} \boldsymbol{U}_e\big) \times_0 (\boldsymbol{U}_0^\top \boldsymbol{A}) \times_d \boldsymbol{U}_{d\perp}^\top \times_{e\neq 0,d} \boldsymbol{U}_e^\top\Big\}\boldsymbol{V}_d \\
&= \mathscr{M}_d\big\{\mathscr{T}_d(\boldsymbol{D}_d\boldsymbol{V}_d^\top) \times_0 (\boldsymbol{U}_0^\top \boldsymbol{A}\boldsymbol{U}_0) \times_d (\boldsymbol{U}_{d\perp}^\top \boldsymbol{U}_{d\perp}) \times_{e\neq 0,d} (\boldsymbol{U}_e^\top \boldsymbol{U}_e)\big\}\boldsymbol{V}_d \\
&\quad + \sum_{c\neq d} \mathscr{M}_d\Big\{\big(\mathscr{T}_c(\boldsymbol{U}_{c\perp}\boldsymbol{D}_c\boldsymbol{V}_c^\top) \times_{e\neq c,d} \boldsymbol{U}_e\big) \times_0 (\boldsymbol{U}_0^\top \boldsymbol{A}) \times_d (\boldsymbol{U}_{d\perp}^\top \boldsymbol{U}_d) \times_{e\neq 0,d} \boldsymbol{U}_e^\top\Big\}\boldsymbol{V}_d \\
&= \mathscr{M}_d\big\{\mathscr{T}_d(\boldsymbol{D}_d\boldsymbol{V}_d^\top) \times_0 (\boldsymbol{U}_0^\top \boldsymbol{A}\boldsymbol{U}_0)\big\}\boldsymbol{V}_d.
\end{aligned}
$$

The proof is then complete. □

S2 Additional Numerical Results

Here we present an example of non-uniform grid points. The setup follows the first simulation in Section 4, except we now choose an unbalanced set of t_j's. Specifically, $5p_0/12$ measurements are equally spaced within $(0, 1/3)$, another $5p_0/12$ measurements are equally spaced within $(2/3, 1)$, and $p_0/6$ measurements are equally spaced within $(1/3, 2/3)$. Despite this imbalance, the estimation performs similarly to the uniform case, as shown below, analogous to the results in Figure 1.

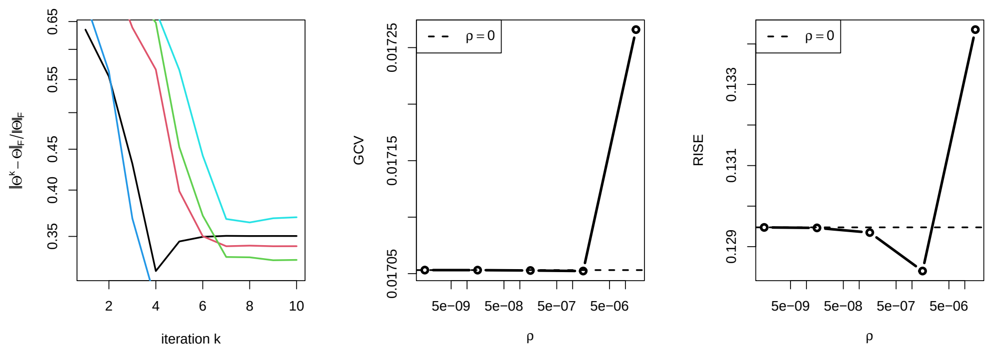

Figure S1: Left: Convergence performance of the functional Riemannian Gauss–Newton algorithm. Middle and Right: GCV and RISE versus the tuning parameter ρ. Displayed are averages based on 100 Monte Carlo replications of $(\boldsymbol{\mathcal{X}}_i, y_i)_{i=1,\ldots,500}$ with measurements on a non-uniform grid.

References

Han, R., R. Willett, and A. R. Zhang (2022). An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics 50*(1), 1–29.

Luo, Y. and A. R. Zhang (2023). Low-rank tensor estimation via riemannian gauss-newton: Statistical optimality and second-order convergence. *Journal of Machine Learning Research 24*(381), 1–48.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.