Functional Tensor Regression

Tongyu Li, Fang Yao, Anru R. Zhang

arXiv:2506.09358, 2025

Background

Tensor regression models relationships between tensor-valued covariates and responses.

Three typical forms:

(1) **Scalar-on-Tensor:** $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$, $y \in \mathbb{R}$
$\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$, $\quad y = \langle \mathcal{X}, \mathcal{B} \rangle + \varepsilon$

Background

Tensor regression models relationships between tensor-valued covariates and responses.

Three typical forms:

(1) **Scalar-on-Tensor:** $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$, $y \in \mathbb{R}$
$\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_D}, \quad y = \langle \mathcal{X}, \mathcal{B} \rangle + \varepsilon$

(2) **Tensor-on-Scalar:** $x \in \mathbb{R}^q$, $\mathcal{Y} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$
$\mathcal{Y} = \sum_{j=1}^{q} x_j \mathcal{B}_j + \mathcal{E}, \quad \mathcal{B}_j \in \mathbb{R}^{p_1 \times \cdots \times p_D}$

Background

Tensor regression models relationships between tensor-valued covariates and responses.

Three typical forms:

(1) **Scalar-on-Tensor:** $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$, $y \in \mathbb{R}$
$\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_D}, \quad y = \langle \mathcal{X}, \mathcal{B} \rangle + \varepsilon$

(2) **Tensor-on-Scalar:** $x \in \mathbb{R}^q$, $\mathcal{Y} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$
$\mathcal{Y} = \sum_{j=1}^{q} x_j \mathcal{B}_j + \mathcal{E}, \quad \mathcal{B}_j \in \mathbb{R}^{p_1 \times \cdots \times p_D}$

(3) **Tensor-on-Tensor:** $\mathcal{X} \in \mathbb{R}^{q_1 \times \cdots \times q_K}$, $\mathcal{Y} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$
$\mathcal{B} \in \mathbb{R}^{q_1 \times \cdots \times q_K \times p_1 \times \cdots \times p_D}, \quad \mathcal{Y} = \langle \mathcal{X}, \mathcal{B} \rangle + \mathcal{E}$

This work focuses on **Scalar-on-Tensor regression** with functional covariates.

Why Functional Tensor Regression (FTR)?

- Traditional tensor regression treats all modes as discrete.
- In many applications, one mode is **continuous** (e.g., time, spectrum).
- **Need:** A method that respects **smooth variation** along the continuous mode.

Model Formulation

Functional Tensor Regression (FTR)

$$y = \int_T \langle \mathcal{X}(t), \mathcal{B}(t) \rangle dt + \varepsilon$$

- $\mathcal{X}(t), \mathcal{B}(t) \in \mathbb{R}^{p_1 \times \cdots \times p_D}$: time-indexed tensor covariate and coefficient
- $t \in T$ denotes the functional mode (e.g., time, spectrum)
- $\mathcal{B}(t)$ is modeled as a smooth function:

$$\mathcal{B}(t) = \Theta \times_0 \psi(t)$$

where $\Theta \in \mathbb{R}^{p_0 \times p_1 \times \cdots \times p_D}$ is a coefficient tensor and $\psi(t) \in \mathbb{R}^{p_0}$ is a spline basis vector.

From Continuous Model to Discrete Optimization

Observation: Only n time points $t_1, \dots, t_n$ are observed.

Discrete model:

$$y_i \approx \langle \mathcal{X}_i,\ \Theta \times_0 \psi(t_i) \rangle + \varepsilon_i, \quad i = 1, \dots, n$$

- $\mathcal{X}_i \in \mathbb{R}^{p_1 \times \cdots \times p_D}$: tensor covariate at t_i
- $\psi(t_i) \in \mathbb{R}^{p_0}$: spline basis at t_i
- $\Theta \in \mathbb{R}^{p_0 \times p_1 \times \cdots \times p_D}$: coefficient tensor

Penalized LS with low-rank constraint:

$$\min_{\Theta \in \mathcal{M}_{\mathbf{r}}} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \langle \mathcal{X}_i, \Theta \times_0 \psi(t_i) \rangle \right)^2 + \rho \, \mathcal{P}(\Theta)$$

where $\mathcal{P}(\Theta)$ enforces smoothness along mode-0, and $\mathcal{M}_{\mathbf{r}}$ is the Tucker rank-$(r_0, \dots, r_D)$ manifold.

Riemannian Optimization for FTR

Problem structure:

- Θ has fixed Tucker rank-$(r_0, \ldots, r_D)$ constraint.
- Such tensors form a **smooth manifold** $\mathcal{M}_r$.

Why Riemannian optimization?

- Avoids explicit nuclear norm $\Rightarrow$ no costly SVD truncation.
- Exploits manifold geometry for faster convergence.

Riemannian Optimization for Fixed-Tucker-Rank FTR

Functional Riemannian Gauss–Newton (FRGN)

Input: responses y_i, tensors $\mathcal{X}_i$, basis $\psi(t_i)$, penalty $\mathcal{P}(\Theta)$, Tucker rank $\mathbf{r}$, max iters K
Output: $\widehat{\Theta}$

Initialize: $\Theta^0 \leftarrow \mathcal{H}_{\mathbf{r}}(Z^* y)$ (*T-HOSVD warm start*)
for $k = 0, 1, \ldots, K-1$ **do**

1) **Euclidean gradient:** $g_k \leftarrow \nabla_\Theta f(\Theta^k)$
2) **Project to tangent space:** $\xi_k \leftarrow P_{T_{\Theta^k}\mathcal{M}_{\mathbf{r}}}(-g_k)$
3) **Gauss–Newton step:** solve subproblem on $T_{\Theta^k}\mathcal{M}_{\mathbf{r}}$ via Riemannian CG / trust-region
4) **Retraction:** $\Theta^{k+1} \leftarrow \mathcal{H}_{\mathbf{r}}(\Theta^k + \xi_k)$

end for

Note: QR 与 W_d 用于切空间构造；$\mathcal{H}_{\mathbf{r}}$ 是 T-HOSVD 作为 retraction.

Theoretical Results of FRGN

- **Error Bound:** Under assumptions (1) and (2), the estimator $\widehat{\Theta}$ satisfies

$$\|\widehat{\Theta} - \Theta^*\|_F \leq C \cdot \phi(n, p_d, r_d),$$

  where C is a constant and $\phi(\cdot)$ depends on sample size n, tensor dimensions p_d, and ranks r_d.

- **Quadratic Convergence Rate:** The Functional Riemannian Gauss–Newton method achieves

$$\|\Theta^{k+1} - \Theta^*\|_F \leq C_q \|\Theta^k - \Theta^*\|_F^2$$

  near the optimum, **outperforming first-order methods** (linear convergence).

fMRI 实验

Dataset

- **Source:** ADHD-200 开源数据集
- **Samples:** $n = 50$ 人

Tensor Covariates

- Resting-state fMRI time series $X_i(t) \in \mathbb{R}^{50\times8\times8\times4}$
  - 50 time frames (采集间隔 2 s)
  - Spatial down-sampling to $8 \times 8 \times 4$ blocks
- 5 个额外协变量: 年龄、性别、头动参数（mean FD）等，m_i 表示

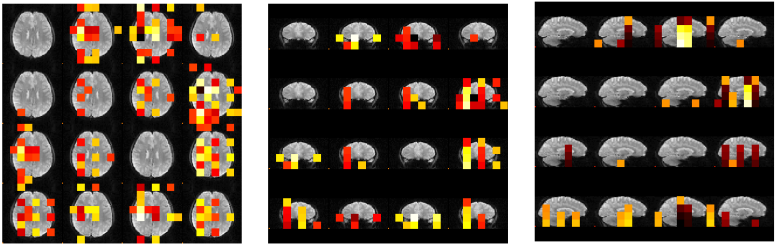
Model Configuration

- Functional Tensor Regression (FTR):

$$y_i = m_i^\top \gamma + \int \langle X_i(t), \mathcal{B}(t) \rangle \, dt + \varepsilon_i$$

- *Time smoothness:* 自然三次样条基函数个数 $K = 5$
- *Low Tucker rank:* $(r_1, r_2, r_3, r_4) = (2, 2, 2, 2)$

Experiments



Coronal Plane Axial Plane Sagittal Plane

Figure 3: Functional tensor regression applied to the ADHD data. Plotted are slices from three spatial dimensions where only coefficients with a magnitude larger than their 80% quantile are displayed. A brighter color means a larger value.

Experiments

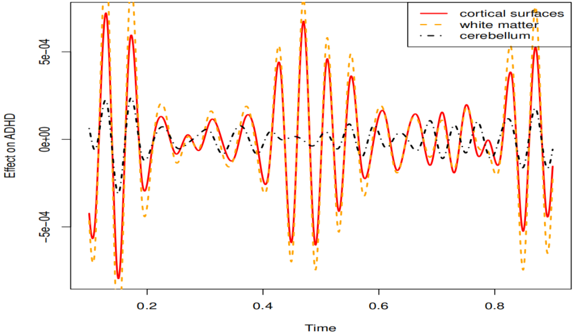


Figure 4: Estimated effects of different regions of the brain on ADHD along time.

Code: github.com/kellty/ftreg.