Statistica Sinica Preprint No: SS-2026-0222	
Title	Efficient Inference for Low-rank Regression via Covariance-Weighted Projection
Manuscript ID	SS-2026-0222
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0222
Complete List of Authors	Baofang Ke, Zihao Song, Weihua Zhao and Lei Wang
Corresponding Authors	Lei Wang
E-mails	lwangstat@nankai.edu.cn
Notice: Accepted author version.	

Efficient Inference for Low-Rank Regression via Covariance-weighted Projection

Baofang Ke[1,2], Zihao Song[3], Weihua Zhao[3] and Lei Wang*[2]

[1]*Fujian Normal University,* [2]*Nankai University and* [3]*Nantong University*

Abstract: Structured linear regression provides a flexible framework for modeling vector-, matrix-, or tensor-valued covariates, while valid and efficient inference for linear functionals of the structured parameter remains challenging, especially under correlated designs, since the optimal inferential direction depends not only on the ambient linear model but also on the local geometry of the parameter space. To address these issues, we propose a general debiasing framework for inference by covariance-weighted projection over local low-rank geometry. Based on this construction, oracle and feasible one-step estimators are developed with asymptotic normality and valid Wald inference. In particular, the proposed method achieves a smaller asymptotic variance than competing approaches without covariance adjustment, the resulting confidence intervals attain a local geometric minimax lower bound, enabling more efficient statistical inference. We further specialize the general framework to low-rank Tucker regression under mode-wise Kronecker covariance structures, where explicit theory and feasible inference procedures are developed. Simulation studies and an application to resting-state EEG data illustrate favorable finite-sample performance.

Corresponding author: Lei Wang, School of Statistics and Data Science & KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, 300071, China. E-mail: lwangstat@nankai.edu.cn.

Key words and phrases: Local low-rank geometry; Debiasing; Covariance-weighted projection; Tucker tensor regression; Efficient inference

1. Introduction

Linear regression is one of the most fundamental tools in statistics. In many contemporary applications, however, the parameter of interest is endowed with additional structure, and the covariates may be vector-, matrix-, or tensor-valued (Koren et al., 2009; Zhou et al., 2013; Mørup et al., 2015). This gives rise to a broad class of structured linear regression problems, in which statistical efficiency and valid inference depend not only on the ambient linear model, but also on the geometry of the parameter space. In this paper, we consider

$$y_i = \langle Z_i, \Theta^* \rangle + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{1.1}$$

which includes classical vector regression, matrix regression, and tensor regression as special cases. Important applications arise when Θ^* satisfies a low-dimensional structural constraint, such as low rank or Tucker-type structure. A large literature has studied estimation of Θ^* under low-rank structure, including nuclear-norm regularization, matrix factorization (Candès and Recht, 2009; Negahban and Wainwright, 2011; Kolda and Bader, 2009; Zhou et al., 2013), Tucker-type tensor methods (Li and Zhang, 2017; Lock, 2018; Hung and Jou, 2019; Kong et al., 2020) and so on. These developments of estimation have greatly expanded the scope of structured regression

with matrix- and tensor-valued covariates.

In many applications, the object of scientific interest is not the full coefficient array itself, but a low-dimensional summary of it. That is, the linear functional

$$\Delta_A \;=\; \langle A, \Theta^\star \rangle,$$

where A is a prespecified loading tensor representing the contrast or summary of interest. Examples include region-level summaries in neuroimaging, structured contrasts across tensor modes, and other interpretable functionals tailored to the scientific question. Although inference for such linear functionals is of substantial practical interest, it remains relatively underdeveloped in the existing literature. To this end, this paper takes a step toward filling this gap.

1.1 Related work

A substantial literature on low-rank matrix and tensor regression has focused on estimation and prediction under structural constraints such as matrix rank, Tucker rank, and CP rank (Koltchinskii et al., 2011; Zhou et al., 2013; Lock, 2018; Li et al., 2018; Zhang et al., 2020). While these works provide important advances in methodology, computation, and error analysis, they do not directly address inference for low-dimensional functionals, nor do they establish the distributional theory needed for uncertainty quantification. As a result, existing estimation procedures do not by themselves yield valid confidence intervals or asymptotically normal estimators for

such targets.

Debiasing has been extensively studied in high-dimensional vector linear models, where a regularized estimator is corrected by a one-step bias-reduction term to recover asymptotic normality for low-dimensional targets (Zhang and Zhang, 2014; van de Geer et al., 2014; Javanmard and Montanari, 2014; Ning and Liu, 2017; Dezeure et al., 2015). In low-rank tensor regression, however, the inferential problem is more delicate because the relevant directions are constrained by the local geometry induced by the low-rank structure. As we show in Proposition 2, directly applying a Euclidean debiasing rule may yield inefficient correction directions. Related geometric approaches to structured inference have also been developed for general high-dimensional linear inverse problems (Cai et al., 2016), but they do not directly identify the covariance-weighted inferential direction arising in our setting. Moreover, existing inference procedures for matrix and tensor regression are developed largely under isotropic or identity-type designs (Xia and Yuan, 2021; Xia et al., 2022; Chen et al., 2019; Xia, 2019; Choi et al., 2024). However, correlated designs prevail in practical settings, and their underlying covariance structure governs both proper bias correction and the statistical efficiency bound. This practical necessity motivates the present work , which introduces a covariance-weighted debiasing framework for structured linear regression, with K-th Tucker tensor regression as primary application scenario.

1.2 Contributions

We develop a debiasing framework for inference on the linear functional Δ_A in low-rank regression by covariance-weighted projection over local low-rank geometry. Our main contributions are as follows.

(1) A general covariance-weighted framework for debiasing linear functionals is proposed, which treats the low-rank parameter space as a smooth manifold and identifies the appropriate debiasing direction through the local low-rank geometry induced by the design covariance, encompassing low-rank Tucker regression. In particular, it also includes the vector linear model as a special Euclidean case, and the covariance-weighted adjustment reduces to the classical Euclidean debiasing direction. We develop a one-step debiasing estimator for the target linear functional based on the sample splitting and establish asymptotic normality for both oracle and feasible procedures, together with asymptotically valid Wald inference. We further show that the resulting confidence intervals attain the local geometric benchmark induced by the oracle debiasing direction.

(2) Our theoretical analysis yields some interesting findings. Compared with Euclidean debiasing and unweighted-projection debiasing procedures, the proposed estimators achieve the smallest asymptotic variance. Moreover, the re-

sulting confidence intervals are locally minimax rate-optimal, which further rigorously elaborates that our method outperforms the Euclidean debiasing and unweighted-projection debiasing methods.

(3) We specialize the general framework to K-th order Tucker tensor regression with mode-wise Kronecker dependence, and low-rank matrix regression is recovered as the special case $K = 2$. We derive explicit projection formulas, construct a feasible inference procedure, and verify model-specific sufficient conditions for asymptotic normality and valid Wald inference.

1.3 Organization and notation

The remainder of the paper is organized as follows. Section 2 develops the covariance-weighted geometric principle, constructs the oracle and feasible one-step estimators, and derives explicit formulas for K-th order Tucker tensor regression under mode-wise Kronecker dependence. General asymptotic theory is established in Section 3, where the required conditions are verified for the Tucker model and the validity and local optimality of the resulting confidence intervals are proved. Section 4 reports finite-sample results for estimation and inference, and Section 5 presents an EEG data analysis. Concluding remarks are given in Section 6. Technical proofs and auxiliary arguments are deferred to Supplementary Material. In this paper, bold lowercase letters denote vectors, bold uppercase letters denote matrices, calligraphic

letters denote tensors, and plain letters denote scalars unless otherwise specified. For matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p\times q}$, let $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{tr}(\mathbf{A}^\top \mathbf{B})$ and $\|\mathbf{A}\|_F^2 = \langle \mathbf{A}, \mathbf{A}\rangle$. For tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{p_1\times\cdots\times p_K}$, let $\langle \mathcal{A}, \mathcal{B}\rangle = \sum_{i_1=1}^{p_1}\cdots\sum_{i_K=1}^{p_K} \mathcal{A}_{i_1,\ldots,i_K}\mathcal{B}_{i_1,\ldots,i_K}$ and $\|\mathcal{A}\|_F^2 = \langle \mathcal{A}, \mathcal{A}\rangle$. Let $\mathbf{A}_j = \mathrm{Mat}_j(\mathcal{A}) \in \mathbb{R}^{p_j\times\prod_{k\neq j} p_k}$ denote the mode-j matricization of $\mathbf{A}$ and $\mathrm{vec}(\cdot)$ denote vectorization. Let $[m] = \{1, \ldots, m\}$ for a positive integer m. We write $c, C > 0$ for generic constants whose values may change from line to line. We also use $\bar{p} = \max_{j\in[K]} p_j$, $\underline{p} = \min_{j\in[K]} p_j$, $\bar{r} = \max_{j\in[K]} r_j$ and $\underline{r} = \min_{j\in[K]} r_j$.

2. Covariance-weighted projection and inference

This section develops the covariance-weighted debiasing construction underlying the proposed inference procedure. The framework is first formulated for a linear model with a structured parameter lying on a smooth manifold, and then specialized to low-Tucker rank K-th order tensor regression under mode-wise Kronecker dependence, with low-rank matrix regression recovered as the case $K = 2$.

2.1 General setup and covariance-weighted geometric principle

We now turn to the construction of the debiasing direction for the target functional $\Delta_A = \langle A, \Theta^*\rangle$. Denote Θ^* as the structured parameter of interest, assumed to lie on a smooth low-rank manifold $\mathcal{M}$. We consider the general linear model introduced in

(1.1),

$$y_i = \langle Z_i, \Theta^* \rangle + \varepsilon_i, \qquad i = 1, \ldots, N,$$

where Z_i may be vector-, matrix-, or tensor-valued depending on the model under consideration, ε_i are independent mean-zero errors with variance σ^2. Denote by $\mathbf{\Sigma} = \mathbb{E}[\mathrm{vec}(Z_i)\mathrm{vec}(Z_i)^\top]$, the covariance operator of the vectorized covariates. The projection principle developed in this section does not require any specific structure on the covariance operator $\mathbf{\Sigma}$, other than positive definiteness. Let $T_{\Theta^*}\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at Θ^*, which characterizes the local geometry of the low-rank manifold around Θ^*. For a given object A, we define the covariance-weighted projection operator $P_{\mathbf{\Sigma},\Theta^*}(A)$ and adjoint operator $P^*_{\mathbf{\Sigma},\Theta^*}(A)$ as

$$P_{\mathbf{\Sigma},\Theta^*}(A) = \arg\min_{B \in T_{\Theta^*}\mathcal{M}} \|A - B\|^2_{\mathbf{\Sigma}}, \tag{2.2}$$

and $\mathrm{vec}\big(P^*_{\mathbf{\Sigma},\Theta^*}(A)\big) = \mathbf{\Sigma}\,\mathrm{vec}\left(P_{\mathbf{\Sigma},\Theta^*}\left(\mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\mathrm{vec}(A))\right)\right)$, where $\|A\|^2_{\mathbf{\Sigma}} = \mathrm{vec}(A)^\top \mathbf{\Sigma}\,\mathrm{vec}(A)$.

Proposition 1 records the basic algebraic properties of the covariance-weighted projection $P_{\mathbf{\Sigma},\Theta^*}$ and its adjoint $P^*_{\mathbf{\Sigma},\Theta^*}$. These identities underlie the debiasing construction below, in particular the oracle and feasible error decompositions and the analysis of the plug-in projection term. Figure 1 gives a geometric illustration of the covariance-weighted projection, together with its unweighted counterpart.

Proposition 1. For any matrices or tensors A and B of compatible dimensions, the following hold: (a) for any $C \in T_{\Theta^*}\mathcal{M}$, $P_{\mathbf{\Sigma},\Theta^*}(C) = C$; (b) $P_{\mathbf{\Sigma},\Theta^*}(A + B) =$

$P_{\mathbf{\Sigma},\Theta^*}(A) + P_{\mathbf{\Sigma},\Theta^*}(B)$; and (c) $\langle P_{\mathbf{\Sigma},\Theta^*}(A), B\rangle = \langle A, P^*_{\mathbf{\Sigma},\Theta^*}(B)\rangle$.

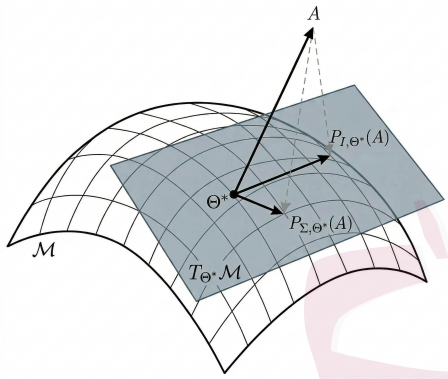

Figure 1: Geometric illustration of the covariance-weighted and unweighted projections onto the tangent space $T_{\Theta^*}\mathcal{M}$ at the true parameter Θ^*. For a loading A, the covariance-weighted projection $P_{\mathbf{\Sigma},\Theta^*}(A)$ is the element of $T_{\Theta^*}\mathcal{M}$ that minimizes the $\mathbf{\Sigma}$-weighted distance to A, whereas $P_{I,\Theta^*}(A)$ denotes the corresponding unweighted projection.

2.2 One-step debiasing: oracle and feasible constructions

We now use the geometric objects introduced in Section 2.1 to construct an estimator for the target functional $\Delta_A = \langle A, \Theta^*\rangle$. For a given loading A, define the projection-adjusted loading $\Phi(A)$ as

$$\Phi(A) = P_{\mathbf{\Sigma},\Theta^*}(\mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\mathrm{vec}(A))),$$

which serves as the debiasing direction for target function Δ_A. Subsequently, we define the oracle one-step debiased estimator

$$\widehat{\Delta}_A^{\mathrm{or}} = \langle A, P_{\mathbf{\Sigma},\Theta^*}(\widehat{\Theta})\rangle + \frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle\big(y_i - \langle Z_i, \widehat{\Theta}\rangle\big). \tag{2.3}$$

where $\widehat{\Theta}$ is a consistent initial estimate. The error decomposition has the following form

$$\widehat{\Delta}_A^{\mathrm{or}} - \Delta_A = \underbrace{\frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle\varepsilon_i}_{\text{Stochastic error}} + \underbrace{\Big\langle P^*_{\mathbf{\Sigma},\Theta^*}(A) - \frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle Z_i, \widehat{\Theta} - \Theta^*\Big\rangle}_{\text{Remainder term}}. \tag{2.4}$$

With such constructed $\Phi(A)$, the remainder term satisfies $\mathbb{E}[\langle \Phi(A), Z_i\rangle Z_i] = P^*_{\mathbf{\Sigma},\Theta^*}(A)$, which indicates that the remainder term can be asymptotically negligible compared with the stochastic term under some mild conditions.

The oracle estimator (2.3) relies on the population projection-adjusted loading $\Phi(A)$ and the true tangent space $T_{\Theta^*}\mathcal{M}$, which are usually unknown in practice. Operators $P_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}_{\mathcal{M}}}$, $P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}_{\mathcal{M}}}$, and $\widehat{\Phi}$ are defined by replacing $\mathbf{\Sigma}$ and Θ^* with the estimated nuisance quantities $\widehat{\mathbf{\Sigma}}$ and $\widehat{\Theta}_{\mathcal{M}}$, where $\widehat{\Theta}_{\mathcal{M}} \in \mathcal{M}$ is a manifold-adapted version of the consistent initial estimator $\widehat{\Theta}$. Given the estimated nuisance quantities from another independent dataset, we define the feasible debiased estimator by

$$\widehat{\Delta}_A^{\mathrm{fe}} = \langle A, P_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}_{\mathcal{M}}}(\widehat{\Theta})\rangle + \frac{1}{N}\sum_{i=1}^{N}\langle \widehat{\Phi}(A), Z_i\rangle\big(y_i - \langle Z_i, \widehat{\Theta}\rangle\big). \tag{2.5}$$

The error decomposition has the following form

$$\widehat{\Delta}_A^{\mathrm{fe}} - \Delta_A = \underbrace{\frac{1}{N}\sum_{i=1}^{N}\langle\widehat{\Phi}(A), Z_i\rangle\varepsilon_i}_{\text{Stochastic error}} + \underbrace{\Big\langle P^*_{\widehat{\boldsymbol{\Sigma}},\widehat{\Theta}_{\mathcal{M}}}(A) - \frac{1}{N}\sum_{i=1}^{N}\langle\widehat{\Phi}(A), Z_i\rangle Z_i, \widehat{\Theta} - \Theta^*\Big\rangle}_{\text{Remainder term}} + \underbrace{\langle P^*_{\widehat{\boldsymbol{\Sigma}},\widehat{\Theta}_{\mathcal{M}}}(A) - A, \Theta^*\rangle}_{\text{Feasibility term}}. \tag{2.6}$$

Compared with the oracle decomposition in (2.4), the feasible estimator involves an additional feasibility term arising from the use of an estimated projection operator. This term would vanish if the population tangent-space projection were available, but controlling its sample analogue is nontrivial in general. In the model-specific analysis below, we show that this feasibility term is asymptotically negligible.

Remark 1. *[Euclidean benchmark and geometric intuition] The oracle estimator in* (2.3) *admits the plug-in representation* $\widehat{\Delta}_A^{\mathrm{or}} = \langle A, \widetilde{\Theta}\rangle$ *with*

$$\widetilde{\Theta} = P_{\boldsymbol{\Sigma},\Theta^*}(\widehat{\Theta}) + \frac{1}{N}\sum_{i=1}^{N}\Phi(Z_i)\big(y_i - \langle Z_i, \widehat{\Theta}\rangle\big). \tag{2.7}$$

Consider the classic vector linear model with $y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}^* + \varepsilon_i, i = 1, \ldots, N$, *where* $\boldsymbol{x}_i \in \mathbb{R}^p$, $\boldsymbol{\beta}^* \in \mathbb{R}^p$, $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i^\top]$ *and the parameter space is the full Euclidean space. The tangent space is* $T_{\beta^*}\mathcal{M} = \mathbb{R}^p$, *and hence the covariance-weighted projection operator* (2.2) *is the identity. Therefore,* $\Phi(\boldsymbol{x}_i) = \boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i$ *and the one-step debiased estimaor* (2.7) *reduces to the classic debiased estimator*

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i\big(y_i - \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}\big).$$

The low-rank setting differs because the local parameter space is restricted to the tangent space of the low-rank manifold. As a result, the ambient correction direction

$\boldsymbol{\Sigma}^{-1}vec(A)$ *generally contains components outside the locally admissible space. The proposed estimator removes these components by projecting the correction direction onto the tangent space under the covariance-induced metric, thereby yielding a more efficient inferential direction.*

2.3 Tucker tensor regression under mode-wise Kronecker covariance

We specialize the general covariance-weighted debiasing framework to Tucker tensor regression by taking Z_i as $\mathcal{X}_i \in \mathbb{R}^{p_1\times\cdots\times p_K}$, Θ^* as $\mathcal{T}^* \in \mathbb{R}^{p_1\times\cdots\times p_K}$, and denoting $\mathcal{M}_{\boldsymbol{r}}$ as the manifold of tensors with Tucker rank $\boldsymbol{r} = (r_1, \cdots, r_K)$, respectively. Assume

$$\mathcal{T}^* = \mathcal{G}^* \times_1 \mathbf{U}_1^* \times_2 \cdots \times_K \mathbf{U}_K^*,$$

where $\mathcal{G}^* \in \mathbb{R}^{r_1\times\cdots\times r_K}$ is the core tensor and $\mathbf{U}_j^* \in \mathbb{R}^{p_j\times r_j}$ has orthonormal columns for $j \in [K]$. The tangent space $T_{\mathcal{M}}\mathcal{M}_{\mathbf{r}}$ at $\mathcal{T} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \mathbf{U}_K \in \mathcal{M}_{\mathbf{r}}$ can be parameterized as

$$T_{\mathcal{M}}\mathcal{M}_{\mathbf{r}} = \Big\{\dot{\mathcal{G}} \times_{k=1}^{K} \mathbf{U}_k\mathbf{U}_k + \sum_{k=1}^{K} \mathcal{G} \times_k \dot{\mathbf{U}}_k \times_{j\neq k} \mathbf{U}_j : \dot{\mathbf{U}}_k^\top \mathbf{U}_k = \mathbf{0}\Big\}. \tag{2.8}$$

Our inferential target is the linear functional $\Delta_{\mathcal{A}} = \langle \mathcal{A}, \mathcal{T}^* \rangle$ for a given loading tensor $\mathcal{A} \in \mathbb{R}^{p_1\times\cdots\times p_K}$. To obtain explicit formulas for the covariance-weighted projection operator, we assume that the vectorized design covariance has the mode-wise Kronecker form $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{[K]} \otimes \cdots \otimes \boldsymbol{\Sigma}^{[1]}$, where each $\boldsymbol{\Sigma}^{[k]} \in \mathbb{R}^{p_k\times p_k}, k \in [K]$ is positive definite.

Let $\mathbf{P}_{\mathbf{U}_j^*}^{\mathbf{\Sigma}^{[j]}} = \mathbf{U}_j^*\{(\mathbf{U}_j^*)^\top \mathbf{\Sigma}^{[j]} \mathbf{U}_j^*\}^{-1}(\mathbf{U}_j^*)^\top$ for $j \in [K]$. The abstract operators $P_{\mathbf{\Sigma},\Theta^*}$ and $P^*_{\mathbf{\Sigma},\Theta^*}$ introduced in Section 2.2 specialize to $P_{\mathbf{\Sigma},\mathcal{T}^*}$ and $P^*_{\mathbf{\Sigma},\mathcal{T}^*}$. For any $\mathcal{A} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$, they admit the explicit forms

$$P_{\mathbf{\Sigma},\mathcal{T}^*}(\mathcal{A}) = \mathcal{A} \times_{k=1}^K \mathbf{P}_{\mathbf{U}_k^*}^{\mathbf{\Sigma}^{[k]}} \mathbf{\Sigma}^{[k]} \; + \; \sum_{j=1}^K \mathcal{G}^* \times_{k \neq j} \mathbf{U}_k^* \times_j \mathbf{W}_{j1}, \tag{2.9}$$

$$P^*_{\mathbf{\Sigma},\mathcal{T}^*}(\mathcal{A}) = \mathcal{A} \times_{k=1}^K \mathbf{\Sigma}^{[k]} \mathbf{P}_{\mathbf{U}_k^*}^{\mathbf{\Sigma}^{[k]}} \; + \; \sum_{j=1}^K \mathcal{G}^* \times_{k \neq j} \mathbf{\Sigma}^{[k]} \mathbf{U}_k^* \times_j \mathbf{W}_{j2}, \tag{2.10}$$

where $\mathbf{W}_{j1}$ and $\mathbf{W}_{j2}$ are defined as

$$\mathbf{W}_{j1} = \left(\mathbf{I}_{p_j} - \mathbf{P}_{\mathbf{U}_j^*}^{\mathbf{\Sigma}^{[j]}} \mathbf{\Sigma}^{[j]}\right) \mathbf{A}_j \left(\bigotimes_{k \neq j} \mathbf{\Sigma}^{[k]} \mathbf{U}_k^*\right) (\mathbf{G}_j^*)^\top \left(\mathbf{G}_j^* \Big[\bigotimes_{k \neq j} (\mathbf{U}_k^*)^\top \mathbf{\Sigma}^{[k]} \mathbf{U}_k^*\Big] (\mathbf{G}_j^*)^\top\right)^{-1},$$

$$\mathbf{W}_{j2} = \left(\mathbf{I}_{p_j} - \mathbf{\Sigma}^{[j]} \mathbf{P}_{\mathbf{U}_j^*}^{\mathbf{\Sigma}^{[j]}}\right) \mathbf{A}_j \left(\bigotimes_{k \neq j} \mathbf{U}_k^*\right) (\mathbf{G}_j^*)^\top \left(\mathbf{G}_j^* \Big[\bigotimes_{k \neq j} (\mathbf{U}_k^*)^\top \mathbf{\Sigma}^{[k]} \mathbf{U}_k^*\Big] (\mathbf{G}_j^*)^\top\right)^{-1},$$

and $\mathbf{G}_j^*$ denotes the mode-j matricization of $\mathcal{G}^*$. Define the oracle projection-adjusted loading $\Phi(\mathcal{A}) = P_{\mathbf{\Sigma},\mathcal{T}^*}(\mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\mathrm{vec}(\mathcal{A})))$. Therefore, the oracle one-step estimator for the linear functional $\Delta_{\mathcal{A}} = \langle \mathcal{A}, \mathcal{T}^* \rangle$ takes the same form as (2.3) by plugging-in $P_{\mathbf{\Sigma},\mathcal{T}^*}$ and $\Phi(\mathcal{A})$.

To derive the feasible one-step debiased estimator without requiring an external independent dataset, we use sample splitting to decouple nuisance estimation from the one-step correction. Assume $N = 2n$ and partition the sample $\mathcal{D} = \{(y_i, \mathcal{X}_i)\}_{i=1}^{2n}$ into two disjoint subsamples $\mathcal{D}_1 = \{(y_i, \mathcal{X}_i)\}_{i=1}^{n}$ and $\mathcal{D}_2 = \{(y_i, \mathcal{X}_i)\}_{i=n+1}^{2n}$. We first construct an estimator based on $\mathcal{D}_1$, with all nuisance quantities estimated from the

auxiliary sample $\mathcal{D}_2$. Define

$$\widetilde{\mathbf{\Sigma}}^{[j]} = \frac{1}{|\mathcal{D}_2|}\sum_{i\in\mathcal{D}_2} \mathrm{Mat}_j(\mathcal{X}_i)^{\otimes 2},\ j\in[K], \quad b_n = \frac{1}{|\mathcal{D}_2|}\sum_{i\in\mathcal{D}_2}\|\mathcal{X}_i\|_F^2, \tag{2.11}$$

and normalize

$$\widehat{\mathbf{\Sigma}}^{[j]} = b_n^{-(K-1)/K}\widetilde{\mathbf{\Sigma}}^{[j]}, \quad j\in[K]. \tag{2.12}$$

Thus, $\mathbf{\Sigma}$ is estimated by

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}^{[K]}\otimes\cdots\otimes\widehat{\mathbf{\Sigma}}^{[1]}. \tag{2.13}$$

Let $\widehat{\mathcal{T}}$ be an initial estimator computed from $\mathcal{D}_2$, and let $\widehat{\mathcal{T}}_{\boldsymbol{r}} = \widehat{\mathcal{G}}\times_{k=1}^K \widehat{\boldsymbol{U}}_k$ be its Tucker-rank $\boldsymbol{r}=(r_1,\ldots,r_K)$ approximation obtained via the HOSVD algorithm (De Lathauwer et al., 2000). We then use the plug-in operator $P_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}_{\boldsymbol{r}}}$ to estimate $P_{\mathbf{\Sigma},\mathcal{T}^*}$, and define $\widehat{\Phi}(\mathcal{A}) = P_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}_{\boldsymbol{r}}}(\mathrm{vec}^{-1}(\widehat{\mathbf{\Sigma}}^{-1}\mathrm{vec}(\mathcal{A})))$. The resulting feasible projection-based estimator based on $\mathcal{D}_1$ is

$$\widehat{\Delta}_{\mathcal{A}}^{(1)} = \big\langle \mathcal{A},\, P_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}_{\boldsymbol{r}}}(\widehat{\mathcal{T}})\big\rangle + \frac{1}{|\mathcal{D}_1|}\sum_{i\in\mathcal{D}_1}\langle\widehat{\Phi}(\mathcal{A}),\,\mathcal{X}_i\rangle(y_i-\langle\mathcal{X}_i,\widehat{\mathcal{T}}\rangle). \tag{2.14}$$

Exchanging the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$ yields another feasible estimator $\widehat{\Delta}_{\mathcal{A}}^{(2)}$, and the final estimator is

$$\widehat{\Delta}_{\mathcal{A}} = \frac{1}{2}\big(\widehat{\Delta}_{\mathcal{A}}^{(1)} + \widehat{\Delta}_{\mathcal{A}}^{(2)}\big). \tag{2.15}$$

The algorithm is summarized in Algorithm 1.

For the inference of $\Delta_{\mathcal{A}}$, Corollary 1 shows that

$$\sqrt{N}V^{-1/2}(\widehat{\Delta}_{\mathcal{A}} - \Delta_{\mathcal{A}}) \xrightarrow{d} N(0,1),$$

Algorithm 1 Feasible Covariance-Weighted One-Step Debiased Estimate

Input: Independent datasets $\mathcal{D}_1$ and $\mathcal{D}_2$; initial estimators $\widehat{\mathcal{T}}^{(1)}$ and $\widehat{\mathcal{T}}^{(2)}$ computed from $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively.

Output: Debiased estimator $\widehat{\Delta}_{\mathcal{A}}$

1: Perform HOSVD on $\widehat{\mathcal{T}}^{(2)}$ such that $\widehat{\mathcal{T}}_{\boldsymbol{r}} = \widehat{\mathcal{G}} \times_{k=1}^{K} \widehat{\boldsymbol{U}}_k$, where $\widehat{\mathcal{G}} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$

2: Compute covariance estimators $\widehat{\boldsymbol{\Sigma}}^{[j]}, j \in [K]$ and $\widehat{\boldsymbol{\Sigma}}$ based on $\mathcal{D}_2$ according to (2.11)–(2.13).

3: Compute projection-adjusted loading $\widehat{\Phi}(\mathcal{A}) = P_{\widehat{\boldsymbol{\Sigma}}, \widehat{\mathcal{T}}_{\boldsymbol{r}}}\big(\mathrm{vec}^{-1}(\widehat{\boldsymbol{\Sigma}}^{-1}\mathrm{vec}(\mathcal{A}))\big)$.

4: Compute $\widehat{\Delta}_{\mathcal{A}}^{(1)}$ using (2.14)

5: Repeat Steps 1–4, switching the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$ and obtain $\widehat{\Delta}_{\mathcal{A}}^{(2)}$.

6: **return** Final estimator: $\widehat{\Delta}_{\mathcal{A}} = (\widehat{\Delta}_{\mathcal{A}}^{(1)} + \widehat{\Delta}_{\mathcal{A}}^{(2)})/2$.

where $V = \sigma^2 \|\Phi(\mathcal{A})\|_{\boldsymbol{\Sigma}}^2$. We estimate V by $\widehat{V} = \widehat{\sigma}^2 \|\widehat{\Phi}(\mathcal{A})\|_{\widehat{\boldsymbol{\Sigma}}}^2$ with $\widehat{\sigma}^2 = \frac{1}{N}\sum_{i\in\mathcal{D}_1}\big(y_i - \langle \mathcal{X}_i, \widehat{\mathcal{T}}^{(2)}\rangle\big)^2 + \frac{1}{N}\sum_{i\in\mathcal{D}_2}\big(y_i - \langle \mathcal{X}_i, \widehat{\mathcal{T}}^{(1)}\rangle\big)^2$. This yields the $(1-\alpha)$-level Wald confidence interval

$$\mathrm{CI}_{\alpha}(\mathcal{A}, \mathcal{D}) = \left[\widehat{\Delta}_{\mathcal{A}} - z_{\alpha/2}\frac{\widehat{V}^{1/2}}{\sqrt{N}},\ \widehat{\Delta}_{\mathcal{A}} + z_{\alpha/2}\frac{\widehat{V}^{1/2}}{\sqrt{N}}\right].$$

Remark 2. *[Matrix regression as the special case $K = 2$] The low-rank matrix regression model is recovered from the Tucker formulation by taking $K = 2$. In this case, $\mathcal{T}^*$ reduces to a rank-r coefficient matrix, and the mode-wise Kronecker covariance $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{[2]} \otimes \boldsymbol{\Sigma}^{[1]}$ becomes the usual Kronecker covariance for matrix-valued*

covariates. Accordingly, the covariance-weighted projection operators, the projection-adjusted loading, and the feasible one-step debiased estimator are obtained by specializing the general formulas to $K = 2$. We record the resulting matrix expressions separately in Supplementary Material for reference.

3. Theoretical analysis

3.1 General asymptotic theory

We list the assumptions needed for the asymptotic normality results in this section.

Assumption 1 (Noise)**.** The errors $\{\varepsilon_i\}$ are independent of $\{Z_i\}$, mean zero, and sub-Gaussian with variance σ^2.

Assumption 2 (Covariance and norm equivalence)**.** Assume $\{Z_i\}_{i=1}^N$ are i.i.d. and $\mathbb{E}\,\mathrm{vec}(Z_i) = 0$. Let $\mathbf{\Sigma} = \mathrm{Cov}(\mathrm{vec}(Z_i))$ be positive definite, and assume its eigenvalues are bounded away from 0 and ∞: there exist constants $0 < \lambda_{\min} \le \lambda_{\max} < \infty$ such that $\lambda_{\min} \le \lambda_{\min}(\mathbf{\Sigma}) \le \lambda_{\max}(\mathbf{\Sigma}) \le \lambda_{\max}$.

Assumption 3 (Sub-Gaussian design)**.** Assume $\mathrm{vec}(Z_i)$ is sub-Gaussian with respect to $\|\cdot\|_{\mathbf{\Sigma}}$: there exists $\kappa > 0$ such that for all $u \in \mathbb{R}^p$, $\mathbb{E}\exp\big(\langle u, \mathrm{vec}(Z_i)\rangle\big) \le \exp(\kappa^2\|u\|_{\mathbf{\Sigma}}^2/2)$.

The oracle and feasible decompositions in (2.4) and (2.6) share the same leading stochastic term, while differing in the remainder induced by nuisance estimation and

plug-in projection. We first state the oracle benchmark.

Theorem 1 (Asymptotic normality of $\widehat{\Delta}_A^{\mathrm{or}}$)**.** Assume Assumptions (1)–(3) hold, $\widehat{\Theta}$ is independent of $\{(y_i, Z_i)\}_{i=1}^N$, and $\|\widehat{\Theta} - \Theta^*\|_{\mathbf{\Sigma}} = o_p(1)$, then we have $\mathbb{E}(\widehat{\Delta}_A^{\mathrm{or}}) = \Delta_A$. Define $V := \mathrm{Var}(\langle \Phi(A), Z_1 \rangle \varepsilon_1) = \sigma^2 \|\Phi(A)\|_{\mathbf{\Sigma}}^2 > 0$. As $N \to \infty$, we have

$$\sqrt{N}\, V^{-1/2} \big(\widehat{\Delta}_A^{\mathrm{or}} - \Delta_A \big) \xrightarrow{d} N(0, 1).$$

To obtain a computable estimator, we replace unknown quantities by consistent estimators. To justify concentration steps, we employ sample splitting or cross-fitting so that the nuisance estimators are independent of the sample $\{(y_i, Z_i)\}_{i=1}^N$ used in the debiasing step.

Theorem 2 (Asymptotic normality of $\widehat{\Delta}_A^{\mathrm{fe}}$)**.** Assume Assumptions (1)–(3) hold. Suppose $(\widehat{\Phi}(A), \widehat{\mathbf{\Sigma}}, \widehat{\Theta})$ are independent of $\{(y_i, Z_i)\}_{i=1}^N$, $\|\widehat{\Theta} - \Theta^*\|_{\mathbf{\Sigma}} = o_p(1)$, $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = O_p(n^{-1/2})$, and $\|\widehat{\Phi}(A) - \Phi(A)\|_{\mathbf{\Sigma}} = o_p\big(\|\Phi(A)\|_{\mathbf{\Sigma}}\big)$, we have $\mathbb{E}(\widehat{\Delta}_A^{\mathrm{fe}}) = \Delta_A$. In addition, assume

$$\big\langle P^*_{\widehat{\mathbf{\Sigma}}, \widehat{\Theta}}(A) - A,\, \Theta^* \big\rangle = o_p\big(N^{-1/2} \sigma \|\Phi(A)\|_{\mathbf{\Sigma}}\big), \tag{3.16}$$

Then with $V := \mathrm{Var}(\langle \Phi(A), Z_1 \rangle \varepsilon_1) = \sigma^2 \|\Phi(A)\|_{\mathbf{\Sigma}}^2$, we have

$$\sqrt{N}\, V^{-1/2} \big(\widehat{\Delta}_A^{\mathrm{fe}} - \Delta_A \big) \xrightarrow{d} N(0, 1), \qquad N \to \infty.$$

Remark 3. *When $\widehat{\Theta}$ in (3.16) is replaced by Θ^*, the term $\langle P^*_{\widehat{\mathbf{\Sigma}}, \Theta^*}(A) - A, \Theta^* \rangle$ vanishes exactly. Indeed, since $\Theta^* \in T_{\Theta^*}\mathcal{M}$, Proposition 1(a) yields $P_{\widehat{\mathbf{\Sigma}}, \Theta^*}(\Theta^*) = \Theta^*$. Hence,*

by Proposition 1(c),

$$\langle P^*_{\widehat{\mathbf{\Sigma}},\Theta^*}(A), \Theta^* \rangle = \langle A, P_{\widehat{\mathbf{\Sigma}},\Theta^*}(\Theta^*) \rangle = \langle A, \Theta^* \rangle,$$

and therefore $\langle P^*_{\widehat{\mathbf{\Sigma}},\Theta^*}(A) - A, \Theta^* \rangle = 0$. *Thus, the extra term in* (3.16) *is purely a plug-in effect, arising from the sample analogue of the projection operator and, in particular, from estimating the local tangent-space geometry.*

Remark 4. *The first-order validity of the proposed feasible estimator is preserved under sample splitting, covariance estimation and HOSVD/SVD-based initialization, provided that the corresponding consistency and rate conditions in Theorem 2 and Corollary 1 are satisfied. Initialization error and signal strength are coupled through the factor* $\lambda^{-2}\|\widehat{\mathcal{T}} - \mathcal{T}^\star\|_F^2 \eta_{\boldsymbol{p},\boldsymbol{r}} \bar{r}/\sqrt{p}$*; hence, a smaller* λ *requires a more accurate initial estimator. In contrast, rank under-specification may introduce a persistent approximation bias that cannot generally be removed by improving the initialization. See also the rank-sensitivity analysis in Supplementary Material.*

Comparison with two benchmark debiasing estimators To clarify the efficiency gain from incorporating both low-rank geometry and covariance weighting, we compare the proposed estimator with two natural benchmark procedures. The first is the Euclidean debiased estimator obtained by ignoring the low-rank geometry,

$$\widehat{\Delta}_A^{\mathrm{euc}} = \langle A, \widehat{\Theta} \rangle + \frac{1}{N} \sum_{i=1}^{N} \big\langle \mathrm{vec}(A), \mathbf{\Sigma}^{-1} \mathrm{vec}(Z_i) \big\rangle \big(y_i - \langle Z_i, \widehat{\Theta} \rangle \big), \tag{3.17}$$

and the second is the unweighted-projection debiased estimator

$$\widetilde{\Delta}_A^{\mathrm{or}} = \langle A, P_{\boldsymbol{I},\Theta^*}(\widehat{\Theta})\rangle + \frac{1}{N}\sum_{i=1}^{N}\langle \bar{\Phi}(A), Z_i\rangle\big(y_i - \langle Z_i, \widehat{\Theta}\rangle\big), \tag{3.18}$$

where $\mathrm{vec}(\bar{\Phi}(A)) = \boldsymbol{\Sigma}^{-1}\mathrm{vec}(P_{\boldsymbol{I},\Theta^*}(A))$.

Proposition 2. Under the assumptions of Theorem 1, the Euclidean estimator in (3.17), the unweighted-projection estimator in (3.18), and the proposed oracle estimator $\widehat{\Delta}_A^{\mathrm{or}}$ have asymptotic variances

$$\mathrm{Var}(\widehat{\Delta}_A^{\mathrm{euc}}) = \frac{\sigma^2}{N}\|\boldsymbol{\Sigma}^{-1}\mathrm{vec}(A)\|_{\boldsymbol{\Sigma}}^2,\ \mathrm{Var}(\widetilde{\Delta}_A^{\mathrm{or}}) = \frac{\sigma^2}{N}\|\bar{\Phi}(A)\|_{\boldsymbol{\Sigma}}^2,\ \text{and } \mathrm{Var}(\widehat{\Delta}_A^{\mathrm{or}}) = \frac{\sigma^2}{N}\|\Phi(A)\|_{\boldsymbol{\Sigma}}^2,$$

respectively. Moreover,

$$\mathrm{Var}(\widehat{\Delta}_A^{\mathrm{or}}) \le \mathrm{Var}(\widetilde{\Delta}_A^{\mathrm{or}}) \le \mathrm{Var}(\widehat{\Delta}_A^{\mathrm{euc}}).$$

The first inequality shows the gain from incorporating the covariance structure into the tangent-space projection, relative to an unweighted geometric correction. The second shows the gain from restricting the correction direction to the locally admissible tangent space, relative to the ambient Euclidean benchmark. We next show that this advantage is not only relative: the resulting confidence intervals also attain the local geometric benchmark induced by the oracle debiasing direction.

Local optimality of confidence intervals We next show that this efficiency gain is not merely relative to the Euclidean benchmark: the resulting confidence intervals

attain the local geometric benchmark induced by the oracle debiasing direction $\Phi(A)$, and are therefore locally rate-optimal. For the local minimax formulation, extend the earlier notation and write $\Delta_A(\Theta) = \langle A, \Theta \rangle$, for $\Theta \in \mathcal{U} \subset \mathcal{M}$, where $\mathcal{U}$ is a local parameter space containing a relative neighborhood of Θ^* in $\mathcal{M}$. For $0 < \alpha < 1/2$, define the class of $(1-\alpha)$ confidence intervals

$$\mathcal{I}_\alpha(\mathcal{U}, A) = \Big\{ \mathrm{CI} = [\ell(\mathcal{D}), u(\mathcal{D})] : \inf_{\Theta \in \mathcal{U}} \mathbb{P}_\Theta\big(\Delta_A(\Theta) \in \mathrm{CI}\big) \geq 1 - \alpha \Big\},$$

where $\mathcal{D} = \{(y_i, Z_i)\}_{i=1}^N$ and $L(\mathrm{CI}) := u(\mathcal{D}) - \ell(\mathcal{D})$ denotes the interval length.

Theorem 3 (A unified local geometric lower bound for confidence-interval length)**.** Assume $\mathcal{M}$ is a C^1 embedded submanifold. Under model (1.1), assume Assumptions 1–2 hold and, in addition, $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$. Fix $\Theta^* \in \mathcal{U} \subset \mathcal{M}$. Then for any fixed loading object A and any $0 < \alpha < 1/2$, there exists a constant $c > 0$ depending only on α such that

$$\inf_{\mathrm{CI} \in \mathcal{I}_\alpha(\mathcal{U}, A)} \sup_{\Theta \in \mathcal{U}} \mathbb{E}_\Theta L(\mathrm{CI}) \geq c \, \frac{\sigma}{\sqrt{N}} \, \|\Phi(A)\|_{\mathbf{\Sigma}}.$$

The lower bound in Theorem 3 is therefore of the same geometric order as the Wald interval length implied by Theorems 1 and 2, where $V = \sigma^2 \|\Phi(A)\|_{\mathbf{\Sigma}}^2$. Consequently, the projection-based confidence intervals are locally minimax rate-optimal.

Remark 5. *The local minimax lower bound in Theorem 3 and the variance ordering in Proposition 2 are established for a general C^1 embedded manifold $\mathcal{M}$ and any positive-definite design covariance $\mathbf{\Sigma}$. Their forms are therefore not tied to the*

Tucker structure or Kronecker covariance. The corresponding asymptotic variances and lower-bound value depend on $(\mathcal{M}, \mathbf{\Sigma}, A)$ *through the covariance-weighted projection* $\Phi(A)$*. The Tucker–Kronecker setting is used only to obtain explicit projection formulas and verifiable sufficient conditions.*

3.2 Verification for Tucker tensor regression

We verify the assumptions of Theorem 2 under the tensor linear regression model with Kronecker covariance. The only nonstandard step is the control of the feasibility term induced by the plug-in tangent-space projection.

Assumption 4 (Structured sub-Gaussian design)**.** Assume $\mathcal{X}_i = \mathcal{S}_i \times_{k=1}^{K} \mathbf{L}_k$, where $\mathbf{L}_k \in \mathbb{R}^{p_k \times p_k}$ for $k \in [K]$ are deterministic full-rank matrices. The entries of $\mathcal{S}_i$ are i.i.d. with mean 0, variance 1, and are sub-Gaussian with parameter $\kappa > 0$: for all $t \in \mathbb{R}$, $\mathbb{E}\exp\big(t\,(\mathcal{S}_i)_{j_1,\ldots,j_K}\big) \le \exp(\kappa^2 t^2/2)$, with $j_k \in [p_k]$ for all $k \in [K]$.

Assumption 5 (Signal strength)**.** Assume the minimum nonzero singular value across all unfoldings satisfies $\lambda := \lambda_{\min}(\mathcal{T}^*) = \min_{1\le j\le K} \sigma_{r_j}(\mathbf{T}_j^*) \ge c$, for some constant $c > 0$.

Assumption 6 (Initial singular subspace estimation)**.** For each $j \in [K]$, let $\mathbf{P}_{\boldsymbol{U}_j} := \boldsymbol{U}_j(\boldsymbol{U}_j^\top \boldsymbol{U}_j)^{-1}\boldsymbol{U}_j^\top$ and $\mathbf{P}_{\widehat{\boldsymbol{U}}_j} := \widehat{\boldsymbol{U}}_j(\widehat{\boldsymbol{U}}_j^\top \widehat{\boldsymbol{U}}_j)^{-1}\widehat{\boldsymbol{U}}_j^\top$ denote the projection matrices onto the column spaces of $\boldsymbol{U}_j$ and $\widehat{\boldsymbol{U}}_j$, respectively. Assume that $\|\mathbf{P}_{\widehat{\boldsymbol{U}}_j} - \mathbf{P}_{\boldsymbol{U}_j}\|_2 = o_p(1)$, for $j \in [K]$.

Assumption 7 (Dimension-driven alignment)**.** Let $\bar{\mathcal{A}} = \mathcal{A} \times_k (\mathbf{\Sigma}^{[k]})^{1/2} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ and write $\bar{\mathbf{A}}_j = \mathrm{Mat}_j(\bar{\mathcal{A}})$. For each $j \in [K]$, let $\boldsymbol{U}_j, \widetilde{\boldsymbol{U}}_j \in \mathbb{R}^{p_j \times r_j}$ have orthonormal columns, and let $\boldsymbol{V}_{-j}, \widetilde{\boldsymbol{V}}_{-j} \in \mathbb{R}^{(\prod_{k \neq j} p_k) \times r_j}$ have orthonormal columns. Assume there exists $\eta_{\boldsymbol{p},\boldsymbol{r}} \geq 1$ such that for all j,

$$\|\boldsymbol{U}_j^\top \bar{\mathbf{A}}_j \boldsymbol{V}_{-j}\|_F \;\leq\; \eta_{\boldsymbol{p},\boldsymbol{r}} \cdot \max\left\{ \sqrt{\frac{r_j}{p_j}}\, \|\widetilde{\boldsymbol{U}}_j^\top \bar{\mathbf{A}}_j\|_F,\; \sqrt{\frac{r_j}{\prod_{k \neq j} p_k}}\, \|\bar{\mathbf{A}}_j \widetilde{\boldsymbol{V}}_{-j}\|_F \right\}.$$

Under Assumption 4, $\mathrm{vec}(\mathcal{X}_i)$ is sub-Gaussian with covariance $\mathbf{\Sigma} = \mathbf{\Sigma}^{[K]} \otimes \cdots \otimes \mathbf{\Sigma}^{[1]}$, so Assumption 3 holds. Assumptions 5 and 6 are standard in low-rank inference. The signal strength λ ensures local identifiability of the tangent space, while Assumption 6 follows from $\|\widehat{\mathbf{T}}_k - \mathbf{T}_k^*\|_2 = o_p(\lambda)$ for all $k \in [K]$ via Davis–Kahan/Wedin perturbation bounds. Assumption 7 is a dimension-adjusted compatibility condition between the covariance-transformed loading $\bar{\mathbf{A}}_j$ and the perturbation subspaces arising from singular-subspace estimation. The factor $\eta_{\boldsymbol{p},\boldsymbol{r}}$ measures the deviation from the dimensional contraction expected when the loading is projected onto low-dimensional subspaces. It may remain of constant order when $\bar{\mathbf{A}}_j$ is sufficiently diffuse relative to these perturbation directions. Sparsity or localization of $\mathcal{A}$ alone does not determine whether the condition holds. A localized loading can still satisfy the condition when the relevant one-sided projections are nondegenerate, whereas strong simultaneous alignment with the perturbation subspaces may cause $\eta_{\boldsymbol{p},\boldsymbol{r}}$ to grow. Such growth is permitted as long as the rate condition in Corollary 1 remains valid.

Corollary 1 (Asymptotic normality of $\widehat{\Delta}^{\mathrm{fe}}_{\mathcal{A}}$)**.** Assume that Assumptions 1–3, and 4–7 hold. In addition, assume

$$\lambda^{-2}\|\widehat{\mathcal{T}}-\mathcal{T}^*\|_F^2\cdot\eta_{\boldsymbol{p},\boldsymbol{r}}\frac{\bar{r}}{\sqrt{\underline{p}}}=o_p(N^{-1/2}\sigma), \tag{3.19}$$

Then the feasible estimator $\widehat{\Delta}_{\mathcal{A}}$ defined in (2.15) satisfies $\sqrt{N}\,V^{-1/2}\big(\widehat{\Delta}_{\mathcal{A}}-\Delta_{\mathcal{A}}\big)\xrightarrow{d}N(0,1)$.

Remark 6. *Suppose the initial error satisfying* $\|\widehat{\mathcal{T}}-\mathcal{T}^*\|_F^2=O_p(N^{-1}\sigma^2\mathrm{df}(\boldsymbol{p},\boldsymbol{r}))$, *with* $\mathrm{df}(\boldsymbol{p},\boldsymbol{r})=\prod_{j=1}^K r_j+\sum_{j=1}^K r_j(p_j-r_j)$. *Then the feasibility condition* (3.19) *is simplified into*

$$\frac{\sigma}{\lambda^2}\cdot\eta_{\boldsymbol{p},\boldsymbol{r}}\,\frac{\bar{r}\,\mathrm{df}(\boldsymbol{p},\boldsymbol{r})}{\sqrt{\underline{p}}\,\sqrt{N}}=o(1).$$

In the balanced regime $p_j\asymp d$ *and* $r_j\asymp r$ *for all* $j\in[K]$, *this is implied whenever* $N\gg d^{-1}\eta^2_{\boldsymbol{p},\boldsymbol{r}}\,r^2\,(Kdr+r^K)^2$, *in which case the plug-in error is asymptotically negligible and the feasible estimator inherits the same first-order limit as the oracle one.*

Corollary 1 establishes asymptotic normality with scaling by the population variance V. To obtain a feasible pivot, we next show that $\widehat{V}$ is consistent for V, which justifies studentization and the resulting Wald procedure.

Corollary 2 (Valid Wald inference)**.** Under the assumptions of Corollary 1,

$$\sqrt{N}\,\widehat{V}^{-1/2}\,(\widehat{\Delta}_{\mathcal{A}}-\Delta_{\mathcal{A}})\xrightarrow{d}N(0,1).$$

Hence the usual Wald test for $H_0 : \Delta_{\mathcal{A}} = \Delta_{\mathcal{A},0}$ has asymptotic size α, and an asymptotically valid $(1-\alpha)$ confidence interval for $\Delta_{\mathcal{A}}$ is $\widehat{\Delta}_{\mathcal{A}} \pm z_{\alpha/2}\sqrt{\widehat{V}/N}$.

4. Simulation studies

This section studies the finite-sample behavior of the proposed covariance-weighted one-step debiased estimator in matrix trace regression $(K = 2)$ and low Tucker-rank tensor regression with $K = 3$, respectively. We report bias, root mean squared error (RMSE), empirical coverage, and average interval length.

4.1 Matrix linear regression

For matrix trace regression, we generate covariates as $\boldsymbol{X}_i = \mathbf{L}_1 \boldsymbol{S}_i \mathbf{L}_2^\top$, $i = 1, \ldots, N$, where $N = 2n$ and the entries of $\boldsymbol{S}_i \in \mathbb{R}^{p\times q}$ are i.i.d. standard normal. We set $p = 30$, $q = 25$, $r = 2$, and $\sigma = 1$, and generate the response from

$$y_i = \langle \boldsymbol{X}_i, \boldsymbol{\Theta}^* \rangle + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

The true coefficient matrix is $\boldsymbol{\Theta}^* = \boldsymbol{U}^* \boldsymbol{V}^{*\top}$, where $\boldsymbol{U}^* \in \mathbb{R}^{p\times r}$ and $\boldsymbol{V}^* \in \mathbb{R}^{q\times r}$ are obtained by orthonormalizing independent Gaussian matrices. We consider two covariance designs. (a) Under identity covariance, $\mathbf{L}_1 = \boldsymbol{I}_p$ and $\mathbf{L}_2 = \boldsymbol{I}_q$ such that $\boldsymbol{\Sigma} = \boldsymbol{I}_q \otimes \boldsymbol{I}_p$. (b) Under general Kronecker covariance, we take $\mathbf{L}_1 = \boldsymbol{U}_p \boldsymbol{\Lambda}_p^{1/2}$, and

$\mathbf{L}_2 = \boldsymbol{U}_q \boldsymbol{\Lambda}_q^{1/2}$, where $\boldsymbol{U}_p$ and $\boldsymbol{U}_q$ are random orthogonal matrices and

$$\boldsymbol{\Lambda}_k = \mathrm{diag}\Big(\underbrace{2, \ldots, 2}_{\lfloor k/4 \rfloor}, \underbrace{1, \ldots, 1}_{k - 2\lfloor k/4 \rfloor}, \underbrace{0.5, \ldots, 0.5}_{\lfloor k/4 \rfloor} \Big), \qquad k \in \{p, q\}. \tag{4.20}$$

Then $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{[2]} \otimes \boldsymbol{\Sigma}^{[1]}$ with $\boldsymbol{\Sigma}^{[1]} = \mathbf{L}_1 \mathbf{L}_1^\top$ and $\boldsymbol{\Sigma}^{[2]} = \mathbf{L}_2 \mathbf{L}_2^\top$. For the target functional $\Delta_{\mathbf{A}} = \langle \mathbf{A}, \boldsymbol{\Theta}^* \rangle$, we consider three representative loading matrices corresponding to sparse, structured low-rank, and dense directions. Specifically, (i) $\mathbf{A}_1$ has a single nonzero entry equal to one at $(1, 1)$; (ii) $\mathbf{A}_2 = \boldsymbol{a}_p \boldsymbol{a}_q^\top / \|\boldsymbol{a}_p \boldsymbol{a}_q^\top\|_F$, where $\boldsymbol{a}_p \in \mathbb{R}^p$ has its first $2\lfloor p^{1/4} \rfloor$ entries equal to one and the remaining entries zero, and $\boldsymbol{a}_q$ is defined analogously; and (iii) $\mathbf{A}_3$ has i.i.d. $\mathrm{Unif}(0, 1)$ entries and is normalized to Frobenius norm one. These loadings represent sparse, structured low-rank, and dense directions, respectively.

We compare the proposed feasible covariance-weighted one-step debiased estimator (**CW-DB**), the unweighted one-step debiased estimator constructed by (3.18) (**UW-DB**), the two-step debiasing method (**TS-DB**) of Ke et al. (2025), the split and non-split power-iteration refinements (**PI-S** and **PI-NS**) adapted from Xu et al. (2023), and the vectorized debiasing baseline (**Vec-DB**) based on (Zhang and Zhang, 2014). In all simulations, the initial estimator is constructed by applying a rank-r truncation to an empirical moment of the form $|\mathcal{J}|^{-1} \sum_{i \in \mathcal{J}} \boldsymbol{X}_i y_i$, where $\mathcal{J}$ denotes the index set for either the full dataset $\mathcal{D}$ or a split sample $\mathcal{D}_1, \mathcal{D}_2$, depending on the method. To standardize initialization quality across regimes, we calibrate the

resulting estimator to have Frobenius error approximately $0.3\sqrt{rp/n}$. The sampling regime is indexed by $\sqrt{rp/n} \in \{0.4, 0.2, 0.1\}$, a quantity proportional to the canonical Frobenius error scale in low-rank estimation and decreasing with sample size. This scaling is also consistent with the feasibility condition in Remark 6. Each configuration is replicated 500 times.

Table 1 reports the bias and RMSE. Across all settings, the biases are small for all debiasing procedures, consistent with their asymptotic unbiasedness. Under identity covariance, all structured methods except Vec-DB have similar RMSEs, which agrees with the theory that they share the same asymptotic variance lower bound in this setting. Under general Kronecker covariance, CW-DB and TS-DB are typically the most accurate procedures, with CW-DB performing better in the large majority of settings. Interestingly, UW-DB performs comparably to the two IT-based procedures, suggesting that exploiting the low-rank structure already yields a substantial gain. Taken together, the simulations further support CW-DB as the primary method, especially under dependent designs.

The inferential results are consistent with the asymptotic theory. Figure S1 in Supplementary Material shows that the studentized statistic $\sqrt{N}, \widehat{V}^{-1/2}(\widehat{\Delta}_{\mathbf{A}} - \Delta_{\mathbf{A}})$ is already close to Gaussian in moderate samples, with the approximation improving as $\sqrt{rp/n}$ decreases. Table 2 shows empirical coverage close to the nominal 95% level, while the average interval length decreases monotonically as $\sqrt{rp/n}$ decreases.

Table 1: Bias ($\times 10^3$) and RMSE ($\times 10^2$) for matrix trace regression.

A	$\sqrt{rp/n}$	Statistic	Method					
			CW-DB	UW-DB	TS-DB	PI-NS	PI-S	Vec-DB
Identity covariance								
A_1	0.4	Bias	0.118	0.110	0.441	0.383	0.047	1.119
		RMSE	1.590	1.595	1.680	**1.524**	1.599	3.751
	0.2	Bias	0.081	0.079	0.081	0.113	0.060	-0.516
		RMSE	0.729	0.729	0.735	**0.726**	0.732	1.881
	0.1	Bias	0.140	0.143	0.110	0.134	0.134	0.248
		RMSE	**0.395**	0.396	0.397	0.397	0.397	0.909
A_2	0.4	Bias	-0.373	-0.440	-0.599	-2.123	-0.942	-0.457
		RMSE	1.483	1.489	1.505	**1.435**	1.498	3.490
	0.2	Bias	-0.271	-0.267	-0.382	-0.689	-0.303	-0.362
		RMSE	0.670	0.669	**0.661**	0.665	0.670	1.845
	0.1	Bias	0.142	0.141	0.170	0.062	0.142	-0.431
		RMSE	**0.355**	0.356	0.358	0.356	0.356	0.970
A_3	0.4	Bias	-0.773	-0.798	-1.379	-2.062	-1.195	-0.044
		RMSE	1.339	1.344	1.407	**1.324**	1.372	3.528
	0.2	Bias	-0.061	-0.059	-0.162	-0.419	-0.093	-0.092
		RMSE	0.666	0.667	0.664	**0.661**	0.670	1.872
	0.1	Bias	0.068	0.067	0.050	-0.001	0.071	-0.215
		RMSE	0.337	0.337	0.337	**0.336**	0.338	0.930
General Kronecker covariance								
A_1	0.4	Bias	0.208	0.027	-0.360	2.054	0.267	0.127
		RMSE	**1.643**	1.756	1.741	1.710	1.767	4.460
	0.2	Bias	0.248	0.104	0.270	0.602	0.190	-1.202
		RMSE	**0.771**	0.826	0.790	0.829	0.831	2.180
	0.1	Bias	0.020	0.012	0.017	0.126	-0.003	-0.586
		RMSE	0.412	0.443	**0.412**	0.444	0.443	1.111
A_2	0.4	Bias	0.430	0.517	0.595	0.560	0.293	-0.115
		RMSE	**1.488**	1.628	1.504	1.626	1.669	4.500
	0.2	Bias	0.194	0.168	0.255	0.216	0.169	0.791
		RMSE	**0.758**	0.821	0.770	0.820	0.831	2.254
	0.1	Bias	0.326	0.358	0.344	0.351	0.343	0.003
		RMSE	**0.354**	0.383	0.355	0.386	0.385	1.104
A_3	0.4	Bias	0.764	1.294	0.622	0.642	1.048	0.976
		RMSE	**1.389**	1.505	1.462	1.502	1.557	3.616
	0.2	Bias	0.354	0.498	0.343	0.310	0.501	0.212
		RMSE	**0.718**	0.773	0.743	0.787	0.784	1.989
	0.1	Bias	-0.080	0.067	-0.071	0.010	0.076	0.013
		RMSE	**0.348**	0.386	0.352	0.390	0.388	0.977

Table 2: Empirical coverage probabilities and average lengths of nominal 95% Wald confidence intervals in matrix regression.

A	$\sqrt{rp/n}$	Identity covariance		General Kronecker covariance	
		CP	Avg. length	CP	Avg. length
1	0.4	0.948	0.0596	0.956	0.0653
	0.2	0.956	0.0295	0.948	0.0320
	0.1	0.944	0.0147	0.950	0.0159
2	0.4	0.950	0.0572	0.948	0.0593
	0.2	0.960	0.0283	0.932	0.0293
	0.1	0.952	0.0141	0.966	0.0146
3	0.4	0.950	0.0530	0.946	0.0558
	0.2	0.944	0.0262	0.950	0.0276
	0.1	0.944	0.0131	0.958	0.0138

Intervals under general Kronecker covariance are slightly longer than those under identity covariance, which is consistent with the additional uncertainty induced by covariance estimation. Overall, these results indicate that CW-DB provides accurate Wald inference in the matrix model.

4.2 Tucker tensor linear regression

For three-way Tucker tensor regression, we generate covariates by $\mathcal{X}_i = \mathcal{S}_i \times_1 \mathbf{L}_1 \times_2 \mathbf{L}_2 \times_3 \mathbf{L}_3$, $i = 1, \ldots, N$, where $N = 2n$ and the entries of $\mathcal{S}_i \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ are i.i.d. standard normal. The true tensor coefficient is $\mathcal{T}^* = \mathcal{G}^* \times_1 \boldsymbol{U}_1^* \times_2 \boldsymbol{U}_2^* \times_3 \boldsymbol{U}_3^*$ with $\boldsymbol{U}_1^*, \boldsymbol{U}_2^*, \boldsymbol{U}_3^* \in \mathbb{R}^{p_j \times r_j}$ are orthonormalized Gaussian matrices and the core tensor $\mathcal{G}^* \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is diagonal, with the only nonzero entries given by $\mathcal{G}^*_{111} = \mathcal{G}^*_{222} = 1$.

The response is generated from

$$y_i = \langle \mathcal{X}_i, \mathcal{T}^* \rangle + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

We set $p_1 = p_2 = p_3 = p = 30$, $r_1 = r_2 = r_3 = r = 2$, and $\sigma = 1$.

As in the matrix case, we study two covariance settings. For the identity design, we set $\mathbf{L}_1 = \mathbf{L}_2 = \mathbf{L}_3 = \boldsymbol{I}_p$. For the general Kronecker design, we let $\mathbf{L}_j = \boldsymbol{Q}_j \boldsymbol{\Lambda}^{1/2}$ for $j = 1, 2, 3$, where each $\boldsymbol{Q}_j$ is a random orthogonal matrix and

$$\boldsymbol{\Lambda} = \operatorname{diag}\Big(\underbrace{2, \ldots, 2}_{\lfloor p/4 \rfloor}, \underbrace{1, \ldots, 1}_{p - 2\lfloor p/4 \rfloor}, \underbrace{0.5, \ldots, 0.5}_{\lfloor p/4 \rfloor} \Big).$$

Each mode covariance matrix $\boldsymbol{\Sigma}^{[j]} = \mathbf{L}_j \mathbf{L}_j^\top$ is unknown and is estimated by (2.11)–(2.12), together with the induced Kronecker covariance (2.13). For the target functional $\Delta_{\mathcal{A}} = \langle \mathcal{A}, \mathcal{T}^* \rangle$, we consider three loading tensors: (i) $\mathcal{A}_1$ has a single nonzero entry equal to one at $(1, 1, 1)$; (ii) $\mathcal{A}_2 = \boldsymbol{a}^{\otimes 3} / \|\boldsymbol{a}^{\otimes 3}\|_F$, where $\boldsymbol{a} \in \mathbb{R}^p$ has its first $2\lfloor p^{1/4} \rfloor$ entries equal to one and the remaining entries zero; and (iii) $\mathcal{A}_3$ has i.i.d. standard normal entries and is normalized to Frobenius norm one. Similarly, these again represent sparse, structured, and dense directions. We compare CW-DB with UW-DB and two iterative projection refinements, IT-S and IT-NS, together with the vectorized baseline Vec-DB. For all methods, initialization is based on a Tucker low-rank approximation of the empirical moment $|\mathcal{J}|^{-1} \sum_{i \in \mathcal{J}} \mathcal{X}_i y_i$, where $\mathcal{J}$ denotes either the full dataset or a split sample, depending on the method. The resulting estimator is calibrated to have Frobenius error approximately $0.3\sqrt{rp/n}$. We consider

Table 3: Bias ($\times 10^3$) and RMSE ($\times 10^3$) for tensor regression with $K = 3$.

A	$\sqrt{rp/n}$	Statistic	Method				
			CW-DB	UW-DB	IT-S	IT-NS	Vec-DB
			Identity covariance				
A_1	0.4	Bias	0.181	0.183	0.180	0.044	-3.467
		RMSE	**2.719**	2.721	2.855	3.108	38.903
	0.2	Bias	-0.034	-0.033	-0.062	-0.039	-0.653
		RMSE	**1.381**	1.382	1.401	1.405	17.763
	0.1	Bias	-0.003	-0.003	-0.002	0.003	-0.040
		RMSE	**0.701**	0.701	0.704	0.705	9.318
A_2	0.4	Bias	-0.024	-0.025	-0.242	0.039	-0.712
		RMSE	3.382	**3.381**	3.485	3.754	41.573
	0.2	Bias	0.065	0.064	0.056	0.084	-0.197
		RMSE	1.835	1.835	1.853	**1.826**	17.871
	0.1	Bias	0.023	0.023	0.024	0.028	-0.144
		RMSE	0.874	0.874	0.881	**0.873**	8.947
A_3	0.4	Bias	-0.011	-0.011	0.214	-0.145	2.343
		RMSE	**2.706**	2.706	2.764	3.002	41.202
	0.2	Bias	0.040	0.040	0.058	0.037	0.084
		RMSE	**1.280**	1.281	1.296	1.316	18.243
	0.1	Bias	0.061	0.062	0.065	0.059	0.246
		RMSE	0.659	0.659	0.662	**0.656**	9.002
			General Kronecker covariance				
A_1	0.4	Bias	0.061	-0.018	-0.226	1.012	-0.393
		RMSE	**1.763**	2.355	2.476	3.156	46.163
	0.2	Bias	-0.033	0.004	-0.015	0.184	-0.262
		RMSE	**0.869**	1.118	1.166	1.188	21.249
	0.1	Bias	-0.020	-0.040	-0.044	0.007	-0.646
		RMSE	**0.416**	0.548	0.552	0.545	10.126
A_2	0.4	Bias	-0.131	-0.223	-0.566	-1.034	-0.409
		RMSE	**1.806**	2.299	2.505	2.747	51.431
	0.2	Bias	-0.037	-0.030	-0.090	-0.188	0.005
		RMSE	**0.930**	1.184	1.206	1.195	21.460
	0.1	Bias	-0.018	-0.058	-0.070	-0.089	0.008
		RMSE	**0.461**	0.580	0.580	0.576	9.734
A_3	0.4	Bias	-0.016	-0.091	-0.303	-0.346	-0.742
		RMSE	**3.443**	4.343	4.425	5.129	51.178
	0.2	Bias	-0.076	-0.077	-0.039	-0.156	0.849
		RMSE	**1.653**	2.105	2.144	2.126	20.506
	0.1	Bias	0.008	-0.002	0.018	-0.022	-0.286
		RMSE	**0.837**	1.017	1.024	1.013	10.354

Table 4: Empirical coverage probabilities and average lengths of nominal 95% Wald confidence intervals in Tucker tensor regression.

A	$\sqrt{rp/n}$	Identity covariance		General Kronecker covariance	
		CP	Avg. length	CP	Avg. length
1	0.4	0.960	0.0110	0.940	0.0068
	0.2	0.958	0.0055	0.948	0.0034
	0.1	0.954	0.0027	0.958	0.0017
2	0.4	0.960	0.0137	0.962	0.0074
	0.2	0.948	0.0068	0.944	0.0036
	0.1	0.950	0.0034	0.942	0.0018
3	0.4	0.938	0.0104	0.946	0.0133
	0.2	0.966	0.0052	0.954	0.0065
	0.1	0.946	0.0026	0.942	0.0033

$\sqrt{rp/n} \in \{0.4, 0.2, 0.1\}$ and each configuration is repeated 500 times.

Table 3 reports the bias and RMSE for tensor regression with $K = 3$. The main conclusions are similar to those for matrix trace regression. All structured procedures substantially outperform Vec-DB, and under identity covariance their RMSEs are close. Under general Kronecker covariance, CW-DB stands out more clearly: it attains the smallest RMSE in essentially all settings, whereas UW-DB and the two IT-based procedures are less accurate but remain comparable to one another. These results further support CW-DB as the preferred procedure for tensor regression, particularly under dependent designs.

The results for Tucker tensor regression with $K = 3$ are broadly similar to those for matrix regression. Table 4 and Figure S2 in Supplementary Material show

coverage close to the nominal 95% level across settings. The intervals remain well calibrated under both identity and general Kronecker covariance, further supporting the validity of the proposed Wald inference for CW-DB in the Tucker model.

5. EEG data analysis

We illustrate the proposed inference procedures using resting-state EEG data from the Healthy Brain Network (`https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/`). The response variable is age, and the covariates are region-by-frequency summaries derived from eyes-closed (EC) and eyes-open (EO) recordings. The raw data come from HBN releases R1–R4. After data preprocessing and quality control, the analysis includes 397 subjects aged from 5.04 to 21.00 years. Band power is computed over eight canonical frequency bands and averaged within ten prespecified ROI groups, yielding a 10×8 matrix covariate for each condition. Stacking EC and EO further gives a $10 \times 8 \times 2$ tensor covariate. The ROI definitions are reported in Table 5. All inferential targets are prespecified linear functionals and all loading matrices/tensors are normalized to have Frobenius norm one. For the matrix model, we consider ROI-slice, band-slice, and global loadings. For the tensor model, we additionally consider shared and contrast loadings along the EC/EO mode, corresponding respectively to $(1, 1)/\sqrt{2}$ and $(1, -1)/\sqrt{2}$. Tuning parameters are selected by cross-validation.

Table 5: Coordinate-based aggregation from the 129-channel EGI montage to the 10 ROI groups. ROI-level band power is obtained by averaging the channel-level band power within each group.

ROI group	n	Channels
anterior-midline	12	Cz, E5, E6, E7, E11, E12, E14, E15, E16, E17, E21, E106
frontal-left	13	E13, E18, E19, E20, E22, E23, E24, E25, E26, E27, E28, E32, E33
frontal-right	13	E1, E2, E3, E4, E8, E9, E10, E112, E117, E118, E122, E123, E124
central-left	13	E29, E30, E31, E34, E35, E36, E37, E39, E40, E41, E42, E45, E46
central-right	13	E80, E87, E93, E102, E103, E104, E105, E108, E109, E110, E111, E115, E116
temporal-left	10	E38, E43, E44, E48, E49, E56, E63, E68, E127, E128
temporal-right	10	E94, E99, E107, E113, E114, E119, E120, E121, E125, E126
posterior-left	18	E47, E50, E51, E52, E53, E54, E57, E58, E59, E60, E61, E64, E65, E66, E67, E69, E70, E73
posterior-midline	9	E55, E62, E71, E72, E74, E75, E76, E81, E82
posterior-right	18	E77, E78, E79, E83, E84, E85, E86, E88, E89, E90, E91, E92, E95, E96, E97, E98, E100, E101

5.1 Matrix log-contrast model.

We first fit a low-rank matrix regression model using the log-contrast covariate

$$\log_{10}(\mathrm{EC}) - \log_{10}(\mathrm{EO}).$$

Five-fold cross-validation selects rank one. Figure 2 reports point estimates and nominal 95% confidence intervals for the prespecified ROI-slice, band-slice, and global loadings. The results indicate a localized rather than diffuse age association. The anterior-midline ROI loading is significantly positive, whereas the central-right ROI loading is significantly negative. At the frequency level, only the alpha2 loading is significantly positive after averaging over ROIs. In contrast, the global loading is not

significantly different from zero. Selected intervals are reported in Table 6.

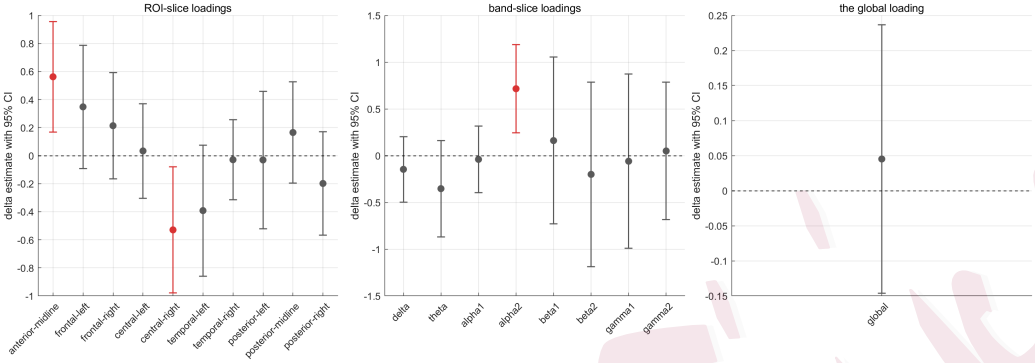

Figure 2: Projection-based inference for prespecified loadings in the rank-one matrix model with covariate $\log_{10}(\mathrm{EC}) - \log_{10}(\mathrm{EO})$. Shown are point estimates and nominal 95% confidence intervals for ROI-slice, band-slice, and global loadings.

5.2 Stacked EC/EO tensor model.

We next stack the log-transformed EC and EO matrices and fit a Tucker tensor regression model. Cross-validation selects rank $(2, 5, 2)$. The tensor formulation allows us to separate shared EC/EO effects from EC–EO contrast effects. Combining these condition-mode loadings with global, band-specific, or ROI-group loadings yields interpretable functionals for inference. Figure 3 summarizes the resulting confidence intervals. The global, condition-slice, shared-global, and contrast-global summaries are not significantly different from zero. In contrast, several band-level shared and contrast summaries are significant. In particular, the shared component shows signif-

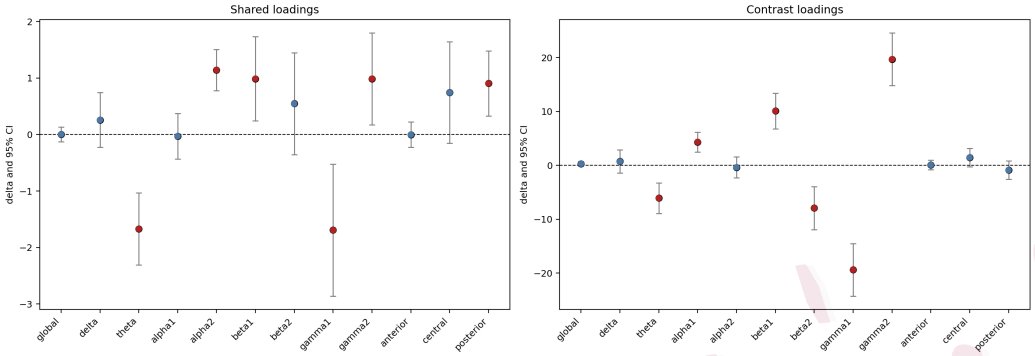

Figure 3: Projection-based inference for prespecified shared and contrast loadings in the Tucker tensor model with rank $(2, 5, 2)$. Point estimates and nominal 95% confidence intervals are reported for global, band-specific, and ROI-group summaries.

icant effects in theta, alpha2, beta1, gamma1 and gamma2, while the contrast component shows significant effects in theta, alpha1, beta1, beta2, gamma1, and gamma2. Among the broader ROI-group summaries, only the posterior shared loading is significantly positive. These findings suggest that the age-related association is mainly concentrated in frequency-specific summaries rather than in global or condition-level averages.

Taken together, the matrix and tensor analyses give a consistent qualitative conclusion: the age-related EEG association is structured rather than diffuse. The global summaries are not significant, whereas selected spatial or frequency-specific summaries are distinguishable from zero. This pattern is broadly consistent with developmental resting-state EEG studies reporting age-related variation in specific

Table 6: Selected projection-based inference results for the HBN EEG analysis. Panel A reports point estimates and nominal 95% confidence intervals for prespecified loadings under the rank-1 matrix model with covariate $\log_{10}(\mathrm{EC}) - \log_{10}(\mathrm{EO})$. Panel B reports the corresponding results for prespecified shared and contrast loadings under the Tucker tensor model with fixed rank $(2, 5, 2)$. All loadings are normalized to have Frobenius norm one before inference. The reported intervals are pointwise and are not adjusted for multiplicity.

Loading	Description	Estimate	95% CI
Panel A: Matrix log-contrast model, rank 1			
ROI slice	anterior-midline	0.562	[0.169, 0.956]
ROI slice	central-right	−0.529	[−0.979, −0.080]
Band slice	alpha2	0.715	[0.243, 1.187]
Global	all elements	0.045	[−0.146, 0.237]
Panel B: Tucker tensor model, fixed rank $(2, 5, 2)$			
Shared global	$(1, 1)/\sqrt{2}$ over condition mode	0.003	[−0.126, 0.133]
Contrast global	$(1, -1)/\sqrt{2}$ over condition mode	0.282	[−0.110, 0.673]
ROI slice	central-left	−0.611	[−1.027, −0.194]
Shared ROI group	posterior	0.907	[0.329, 1.484]
Shared band	theta	−1.671	[−2.307, −1.034]
Shared band	alpha2	1.142	[0.775, 1.510]
Shared band	gamma1	−1.692	[−2.861, −0.524]
Contrast band	alpha1	4.340	[2.503, 6.176]
Contrast band	beta1	10.112	[6.785, 13.438]
Contrast band	gamma1	−19.391	[−24.283, −14.500]

frequency bands, especially theta/alpha-related activity, together with topographic and condition-dependent structure (Cellier et al., 2021; Petro et al., 2022; McSweeney et al., 2023; Wilkinson et al., 2024)

6. Conclusion

We have developed a unified covariance-weighted projection framework for inference on linear functionals in low-rank matrix and tensor regression with dependent designs. The debiasing direction should reflect both the local geometry of the low-rank parameter space and the covariance structure of the covariates. This leads to a projection-adjusted loading and a one-step debiased estimator, formulated abstractly for low-rank manifold models and made explicit for low-rank matrix regression and Tucker tensor regression under Kronecker-type covariance structures.

Supplementary Material

It contains proofs of Propositions, Theorems and additional simulation results.

Acknowledgements

The authors are grateful to the anonymous referees, associate editor, and editor for their helpful comments. Lei Wang's research was supported by the National Natural Science Foundation of China (12271272). The corresponding author is Lei Wang.

References

Cai, T. T., T. Liang, and A. Rakhlin (2016). Geometric inference for general high-dimensional linear inverse problems. *The Annals of Statistics 44*(4), 1536–1563.

Candès, E. J. and B. Recht (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics 9*(6), 717–772.

Cellier, D., J. Riddle, I. T. Petersen, and K. Hwang (2021). The development of theta and alpha neural oscillations from ages 3 to 24 years. *Developmental Cognitive Neuroscience 50*, 100969.

Chen, Y., J. Fan, C. Ma, and Y. Yan (2019). Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences 116*(46), 22931–22937.

Choi, J., H. Kwon, and Y. Liao (2024). Inference for low-rank completion without sample splitting with application to treatment effect estimation. *Journal of Econometrics 240*(1), 105682.

De Lathauwer, L., B. De Moor, and J. Vandewalle (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications 21*(4), 1253–1278.

Dezeure, R., P. Bühlmann, L. Meier, and N. Meinshausen (2015). High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Statistical Sci-*

ence 30(4), 533–558.

Hung, H. and Z.-Y. Jou (2019). A low rank-based estimation-testing procedure for matrix-covariate regression. *Statistica Sinica 29*(2), 1025–1046.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research 15*(82), 2869–2909.

Ke, B., W. Zhao, and L. Wang (2025). Statistical inference for matrix-vector linear regression without debiasing under kronecker covariance structure. *Statistics and Computing 35*(6), 201.

Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM Review 51*(3), 455–500.

Koltchinskii, V., K. Lounici, and A. B. Tsybakov (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics 39*(5), 2302–2329.

Kong, D., C.-H. Zhang, and J. Lv (2020). L2RM: Low-rank linear regression models for high-dimensional matrix responses. *Journal of the American Statistical Association 115*(529), 403–424.

Koren, Y., R. Bell, and C. Volinsky (2009). Matrix factorization techniques for recommender systems. *Computer 42*(8), 30–37.

Li, L. and X. Zhang (2017). Parsimonious tensor response regression. *Journal of the*

American Statistical Association 112(519), 1131–1146.

Li, X., D. Xu, H. Zhou, and L. Li (2018). Tucker tensor regression and neuroimaging analysis. *Statistics in Biosciences 10*(3), 520–545.

Lock, E. F. (2018). Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics 27*(3), 638–647.

McSweeney, M., S. Morales, E. A. Valadez, G. A. Buzzell, L. Yoder, W. P. Fifer, N. Pini, L. C. Shuffrey, A. J. Elliott, J. R. Isler, and N. A. Fox (2023). Age-related trends in aperiodic EEG activity and alpha oscillations during early- to middle-childhood. *NeuroImage 269*, 119925.

Mørup, M., L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred (2015). Tensor decomposition of EEG signals: A brief review. *Journal of Neuroscience Methods 248*, 59–69.

Negahban, S. N. and M. J. Wainwright (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics 39*(2), 1069–1097.

Ning, Y. and H. Liu (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics 45*(1), 158–195.

Petro, N. M., L. R. Ott, S. H. Penhale, M. P. Rempe, C. M. Embury, G. Picci, Y.-P. Wang, J. M. Stephen, V. D. Calhoun, and T. W. Wilson (2022). Eyes-closed

versus eyes-open differences in spontaneous neural dynamics during development. *NeuroImage 258*, 119337.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

Wilkinson, C. L., L. D. Yankowitz, J. Y. Chao, R. Gutiérrez, J. L. Rhoades, S. Shinnar, P. L. Purdon, and C. A. Nelson (2024). Developmental trajectories of EEG aperiodic and periodic components in children 2–44 months of age. *Nature Communications 15*(1), 5788.

Xia, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Transactions on Information Theory 65*(11), 7437–7459.

Xia, D. and M. Yuan (2021). Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society Series B: Statistical Methodology 83*(1), 58–77.

Xia, D., A. R. Zhang, and Y. Zhou (2022). Inference for low-rank tensors—no need to debias. *The Annals of Statistics 50*(2), 1220–1245.

Xu, X., Y. Shen, Y. Chi, and C. Ma (2023). The power of preconditioning in overparameterized low-rank matrix sensing. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 38611–38654. PMLR.

Zhang, A. R., Y. Luo, G. Raskutti, and M. Yuan (2020). ISLET: Fast and optimal

low-rank tensor regression via importance sketching. *SIAM Journal on Mathematics of Data Science 2*(2), 444–479.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(1), 217–242.

Zhou, H., L. Li, and H. Zhu (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association 108*(502), 540–552.

Baofang Ke

School of Mathematics and Statistics, Fujian Normal University, Fujian, China. School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, China. E-mail: kebaofang@outlook.com

Zihao Song

School of Mathematics and Statistics, Nantong University, Jiangsu Nantong, 226019, China. E-mail: zihaosongntu@126.com

Weihua Zhao

School of Mathematics and Statistics, Nantong University, Jiangsu Nantong, 226019, China. E-mail: zhaowhstat@163.com

Lei Wang

School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai Uni-

versity, Tianjin, China. E-mail: lwangstat@nankai.edu.cn