Method Comparison: Matrix response regression

Jiaxin Hu

May 22, 2021

1 Model comparison (STD)

1.1 Our model

Let $\mathcal{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ denote the order-K observation tensor, and $\boldsymbol{X}_k \in \mathbb{R}^{d_k \times p_k}, k \in [K]$ denote the feature matrices. Assume the entries in $\mathcal{Y}$ are from exponential family. Our model is stated as

$$\mathbb{E}[\mathcal{Y}|\boldsymbol{X}_1, ..., \boldsymbol{X}_K] = f(\mathcal{B} \times \{\boldsymbol{X}_1, ..., \boldsymbol{X}_K\}),$$

where the coefficient tensor $\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ satisfies the low-rank structure

$$\mathcal{B} = \mathcal{C} \times \{\boldsymbol{M}_1, ..., \boldsymbol{M}_K\},$$

with core tensor $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$, $\boldsymbol{M}_k \in \mathbb{R}^{p_k \times r_k}$, and $\boldsymbol{M}_k^T \boldsymbol{M}_k = \boldsymbol{I}_{r_k}$.

The model under the special case with $K = 3, d_1 = d_2 = n, d_3 = N, \boldsymbol{X}_1 = \boldsymbol{X}_2 = \boldsymbol{I}_n$ is

$$\mathbb{E}[\mathcal{Y}|\boldsymbol{X}_3] = f(\mathcal{B} \times_3 \boldsymbol{X}_3),$$

where $\mathcal{B}$ has low-rank structure.

1.2 Tensor envelope regression (Env)

For simplicity, we consider the order-3 case. The envelope method works on the following model.

$$\boldsymbol{Y}_i = \mathcal{B} \times_3 X_i + \epsilon_i, \quad i = 1, ..., d_3,$$

where $\boldsymbol{Y}_i \in \mathbb{R}^{d_1 \times d_2}$ is the observed matrix, $\mathcal{B} \in \mathbb{R}^{d_1 \times d_2 \times p}$ is the coefficient tensor, $X_i \in \mathbb{R}^p$ is the feature vector for the i-th individual, and $\epsilon_i \in \mathbb{R}^{d_1 \times d_2}$ is the mean-zero noise matrix with independent entries and Kronecker covariance structure, i.e., $\text{cov}(\text{vec}(\epsilon_i)) = \Sigma = \Sigma_2 \otimes \Sigma_1$. Stack d_3 individuals together. We rewrite the model in the tensor form.

$$\mathcal{Y} = \mathcal{B} \times_3 \boldsymbol{X} + \epsilon,$$

where $\mathcal{Y} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, $\boldsymbol{X} \in \mathbb{R}^{d_3 \times p}$ and $\epsilon \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is the noise tensor.

According to the generalized sparsity principle, only part of $\mathcal{Y}$ is to the features, which means

$$\boldsymbol{Y}_i \times_1 (\boldsymbol{I}_{d_1} - \boldsymbol{P}_1) \perp \boldsymbol{Y}_i \times_1 \boldsymbol{P}_1 | X_i, \quad \boldsymbol{Y}_i \times_2 (\boldsymbol{I}_{d_2} - \boldsymbol{P}_2) \perp \boldsymbol{Y}_i \times_2 \boldsymbol{P}_2 | X_i, \quad i = 1, ..., d_3,$$

where $\boldsymbol{P}_k$ are projection matrices to the subspaces $\mathcal{S}_k \in \mathbb{R}^{d_k}$ with dimension $\dim(\mathcal{S}_k) = u_k$, for $k = 1, 2$. Under this principle, the coefficient tensor and covariance matrices satisfy

$$\mathcal{B} = \mathcal{C} \times_1 \Gamma_1 \times_2 \Gamma_2 \times_3 \boldsymbol{I}_p, \quad \Sigma_k = \Gamma_k \Omega_k \Gamma_k^T + \Gamma_{k0} \Omega_{k0} \Gamma_{k0}^T,$$

where $\mathcal{C} \in \mathbb{R}^{u_1 \times u_2 \times p}$ is the core tensor, $\Gamma_k \in \mathbb{R}^{d_k \times u_k}$ are the bases of $\mathcal{S}_k$, and $\Omega_k \in \mathbb{R}^{u_k \times u_k}, \Omega_{k0} \in \mathbb{R}^{(d_k-u_k)\times(d_k-u_k)}$. Notice that $\mathcal{B}$ has a special tucker decomposition structure with rank $\boldsymbol{r} = c(u_1, u_2, p)$ and factor matrices $\Gamma_1, \Gamma_2, \boldsymbol{I}_p$. The tensor envelope is define as the intersection of all the reducing subspace $\mathcal{S}_1 \otimes \mathcal{S}_2$ which contains the span space $span(\text{unfold}_3(\mathcal{B}))$.

2 Simulation Set up

2.1 Basic points

1. STD and Env should have similar performance when $u_1 = r_1, u_2 = r_2$ and $r_3 = p$.

2. STD should have better performance when more feature matrices on different modes are available.

3. The effect from sample size should be checked.

2.2 Simulation results

Experiment 1: PMSE(Cor) vs number of informative modes

In this experiment, we consider the case with $d_1 = d_2 = d_3 = 20$, $p = 20 * 0.4 = 8$. We consider the signal $\alpha = 3$(low) and $\alpha = 6$(high), and rank $\boldsymbol{r} = (r_1, r_2, r_3) = (3, 3, p)$(low) and $\boldsymbol{r} = (4, 5, p)$(high). The data is generated from the gaussian model. We fit the STD with true parameters and fit the envelope method with $(u_1, u_2) = (r_1, r_2)$. For each setting, we duplicate the estimation for 10 times. See Figure 1 and 2 for simulation results.

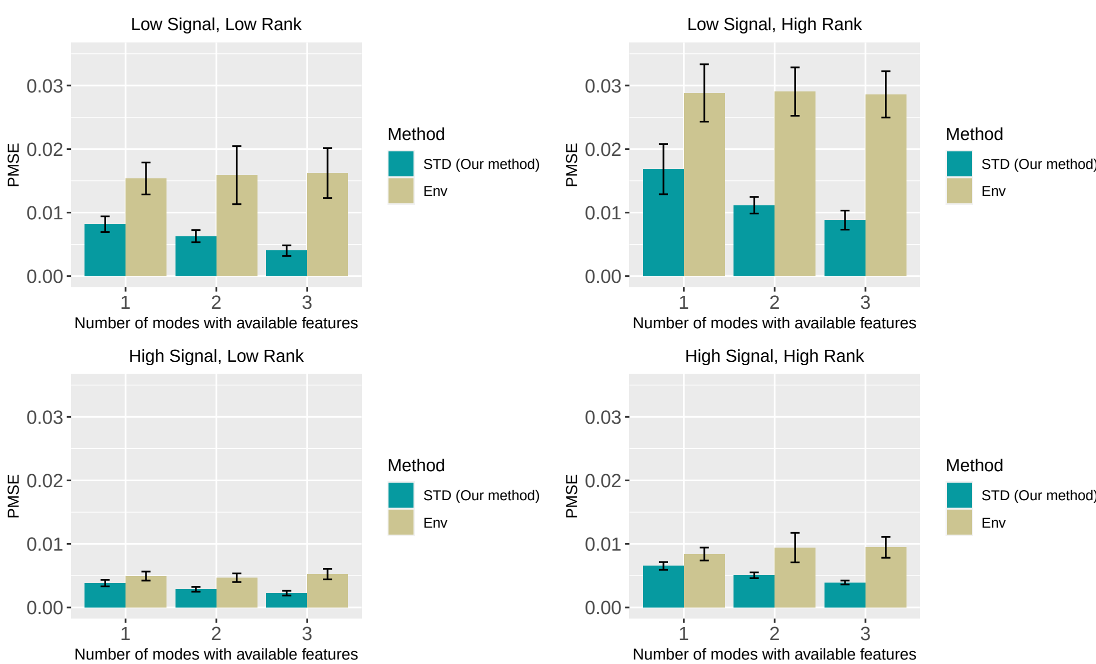

Figure 1: Comparison between STD method and Envelope method. The y-axis is the PMSE and x-axis is the number of informative modes. Consider rank $\boldsymbol{r} = (3, 3, p)$(low) and $\boldsymbol{r} = (4, 5, p)$(high), and signal $\alpha = 3$(low) and $\alpha = 6$(high).

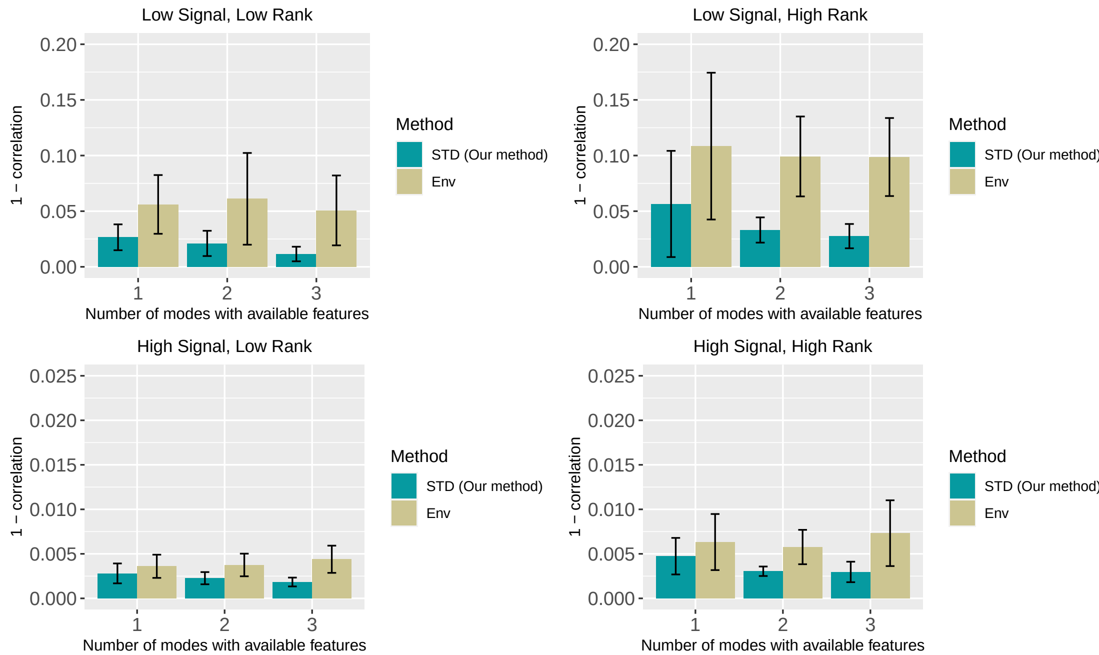

Figure 2: Comparison between STD method and Envelope method. The y-axis is the $1 - \mathrm{cor}(\hat{Y}, Y)$ and x-axis is the number of informative modes. Consider rank $\boldsymbol{r} = (3, 3, p)$(low) and $\boldsymbol{r} = (4, 5, p)$(high), and signal $\alpha = 3$(low) and $\alpha = 6$(high).

Experiment 2: PMSE(Cor) vs sample size

In this experiment, we consider the case with $d_1 = d_2 = 20$ and d_3 varies in $(20, 40, 60, 80, 100)$. Let $p = 0.4d_3$. We consider the signal $\alpha = 3$(low) and $\alpha = 6$(high), and rank $\boldsymbol{r} = (r_1, r_2, r_3) = (3, 3, 3)$(low) and $\boldsymbol{r} = (4, 5, 6)$(high). The data is generated from the gaussian model.We fit the STD with true parameters and fit the envelope method with $(u_1, u_2) = (r_1, r_2)$. For each setting, we duplicate the estimation for 10 times. See Figure 3 and 4 for simulation results.

Conclusion:

Figure 1 shows that our STD model exhibits a substantial reduction in error as the number of informative modes increases. However, the error for Env remains unchanged since Env only incorporates one informative feature matrix on the third mode. The feature matrices on the first two modes will not benefit the estimation of the coefficient tensor.

Figure 3 implies that both STD and Env have decreasing prediction error with respect to the effective sample size when only the feature matrix on the third mode is available. Our STD outperforms under all the combinations of signal and rank. Note that Env assumes full-rankness on third mode. The outperformance of our STD indicates the advantage of assuming the low-rankness on all modes.

Experiment 3: PMSE vs dim with binary and poisson obeservations

In this example, we consider the case with $d_1 = d_2 = 20$ and d_3 varies in $(20, 40, 60, 80, 100)$. Let $p = 0.4d_3$. We consider the signal $\alpha = 3$(low) and $\alpha = 6$(high), and rank $\boldsymbol{r} = (r_1, r_2, r_3) = (3, 3, 3)$(low) and $\boldsymbol{r} = (4, 5, 6)$(high). The data is generated from the binary and poisson model.We fit the STD with true parameters and fit the envelope method with $(u_1, u_2) = (r_1, r_2)$. For each setting, we

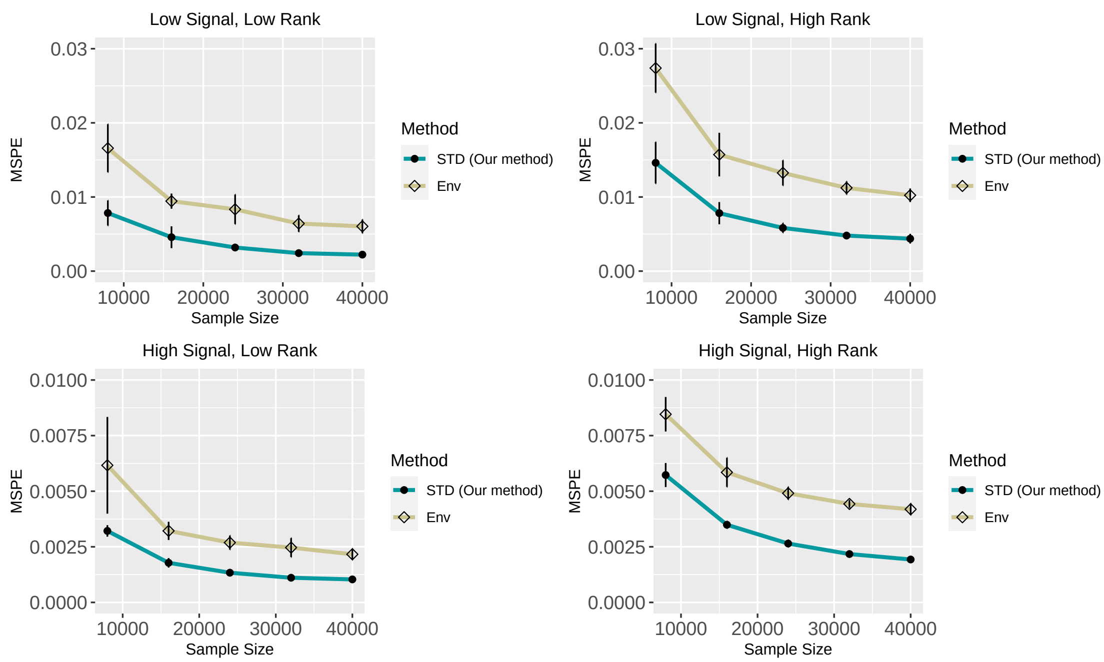

Figure 3: Comparison between STD method and Envelope method. The y-axis is the PMSE and x-axis is the sample size. Consider rank $\boldsymbol{r} = (3,3,3)$(low) and $\boldsymbol{r} = (4,5,6)$(high), and signal $\alpha = 3$(low) and $\alpha = 6$(high).

duplicate the estimation for 10 times. See Figure 5. The PMSE is calculated as following

$$PMES_{\text{binary, STD}} = \text{mean}\left(\frac{1}{1+\exp(-\hat{\mathcal{B}} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})} - \frac{1}{1+\exp(-\mathcal{B}_{true} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})}\right)^2$$

$$PMES_{\text{binary, Env}} = \text{mean}\left(\hat{\mathcal{Y}} - \frac{1}{1+\exp(-\mathcal{B}_{true} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})}\right)^2$$

$$PMES_{\text{poisson, STD}} = \text{mean}\left(\exp(\hat{\mathcal{B}} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) - \exp(\mathcal{B}_{true} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})\right)^2$$

$$PMES_{\text{poisson, Env}} = \text{mean}\left(\hat{\mathcal{Y}} - \exp(\mathcal{B}_{true} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})\right)^2.$$

Further, we consider the response transformation for Envelope method. For binary data, we consider the transformation $\mathcal{Y}_{ijk,trans} = \mathcal{Y}_{ijk}$ if $\mathcal{Y}_{ijk} = 1$, and $\mathcal{Y}_{ijk,trans} = -1$ if $\mathcal{Y}_{ijk} = 0$; for poisson data, we consider the transformation $\mathcal{Y}_{ijk,trans} = \log(\mathcal{Y}_{ijk} + 1)$. Since the expectation of $\mathcal{Y}$ has changed due to the transformation, the corresponding PMSE calculations are changed as

$$PMES_{\text{binary, Env}} = \text{mean}\left(\hat{\mathcal{Y}} + 1 - 2\frac{1}{1+\exp(-\mathcal{B}_{true} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})}\right)^2$$

$$PMES_{\text{poisson, Env}} = \text{mean}\left(\hat{\mathcal{Y}} - \mathcal{B}_{true} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}\right)^2,$$

where $\hat{\mathcal{Y}}$ is the fitted value by Envelope regression. See Figure 6.

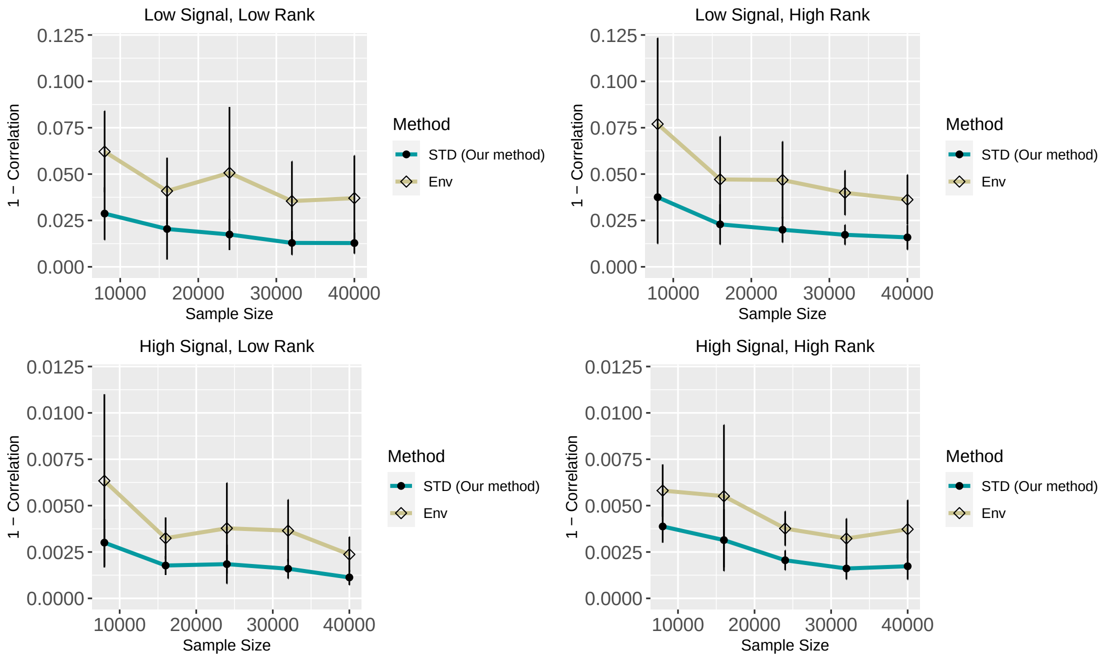

Figure 4: Comparison between STD method and Envelope method. The y-axis is the $1 - \mathrm{cor}(\hat{Y}, Y)$ and x-axis is the sample size. Consider rank $\boldsymbol{r} = (3, 3, 3)$(low) and $\boldsymbol{r} = (4, 5, 6)$(high), and signal $\alpha = 3$(low) and $\alpha = 6$(high).

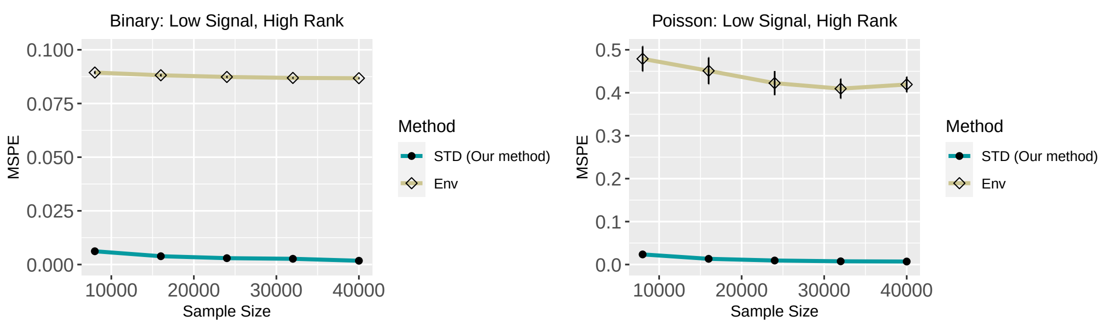

Figure 5: Comparison between STD method and Envelope method. The y-axis is the PMSE calculated as mentioned and x-axis is the sample size. Consider rank $\boldsymbol{r} = (4, 5, 6)$(high), and signal $\alpha = 3$(low). Data in left panel is generated by binary model, and the data in right panel is generated by poisson model.

2.3 Sanity check

Here we generate the data from Envelope model. Consider the case $d_1 = d_2 = d_3 = 20, p = 0.4d_3 = 8$ and envelope dimension $u = (3, 3)$ and $u = (4, 5)$. We fit the envelope model with true parameters, and fit our STD model with $(r_1, r_2) = u$ and choose r_3 in $(3, 6, 9)$ for best performance. See Figure 7.

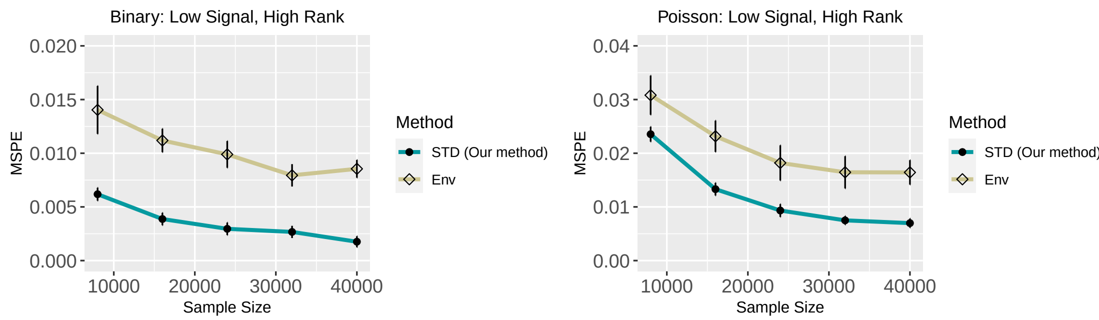

Figure 6: Comparison between STD method and Envelope method. The y-axis is the PMSE calculated as mentioned and x-axis is the sample size. Consider rank $\boldsymbol{r} = (4, 5, 6)$(high), and signal $\alpha = 3$(low). Data in left panel is generated by binary model, and the data in right panel is generated by poisson model. The responses for Env are transformed.

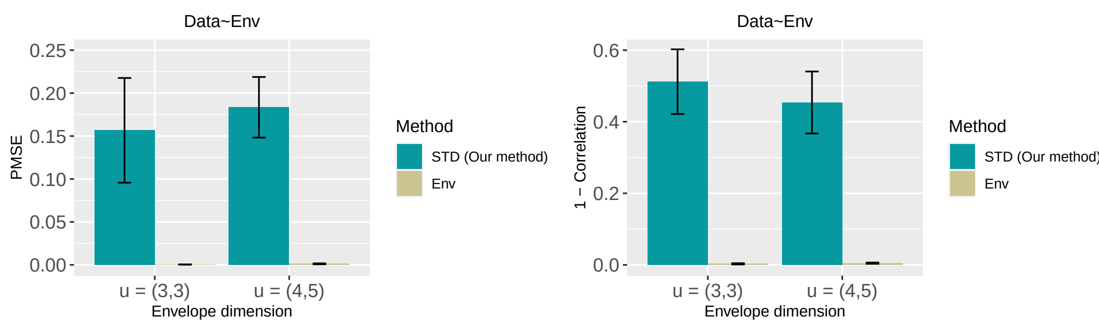

Figure 7: Sanity check for envelope method. Data is generated by the envelope method with different envelope dimensions u. Left panel reports the PMSE performance, and the right panel reports the correlation performance.

Here, we generate the data manually and control the covariance matrices of the noise. We consider the setting $d_1 = d_2 = d_3 = 20$, the feature dimension $p = 5$ and the envelope dimension $(u_1, u_2) = (3, 3)$(low) and $(u_1, u_2) = (4, 5)$(high). The entries of core tensor follows $Unif(-\alpha, \alpha)$, where α is denoted as *signal level*. The covariance tensor is generated as $\mathcal{Z} \times_1 \Sigma_1 \times_2 \Sigma_2$, where $\mathcal{Z} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is tensor with i.i.d. entries from $N(0, 1)$. By the envelope model,

$$\Sigma_k = \Gamma_k \Omega_k \Gamma_k^T + \Gamma_{0k} \Omega_{0k} \Gamma_{0k}^T + \boldsymbol{I}_{d_k}, \quad k = 1, 2$$

, where Γ_k has orthogonal columns and Γ_{0k} has orthogonal columns in $C(\Gamma_k)^{\perp}$, $\Omega_k = AA^T$ with $A \in \mathbb{R}^{u_k \times u_k}$, $\Omega_{0k} = A_0 A_0^T$ with $A_0 \in \mathbb{R}^{(d_k - u_k) \times (d_k - u_k)}$, the entries of A, A_0 i.i.d. follow $Unif(-\gamma, \gamma)$, and γ is denoted as *correlation level*. If γ is near to 0, then Σ_k is near to identity matrix and thus mimics the i.i.d. noise case. In simulation, we consider signal $\alpha = 3$(low) and $\alpha = 6$(high), and γ varies in $(0, 0.15, 0.25)$. See Figures 8 and 9.

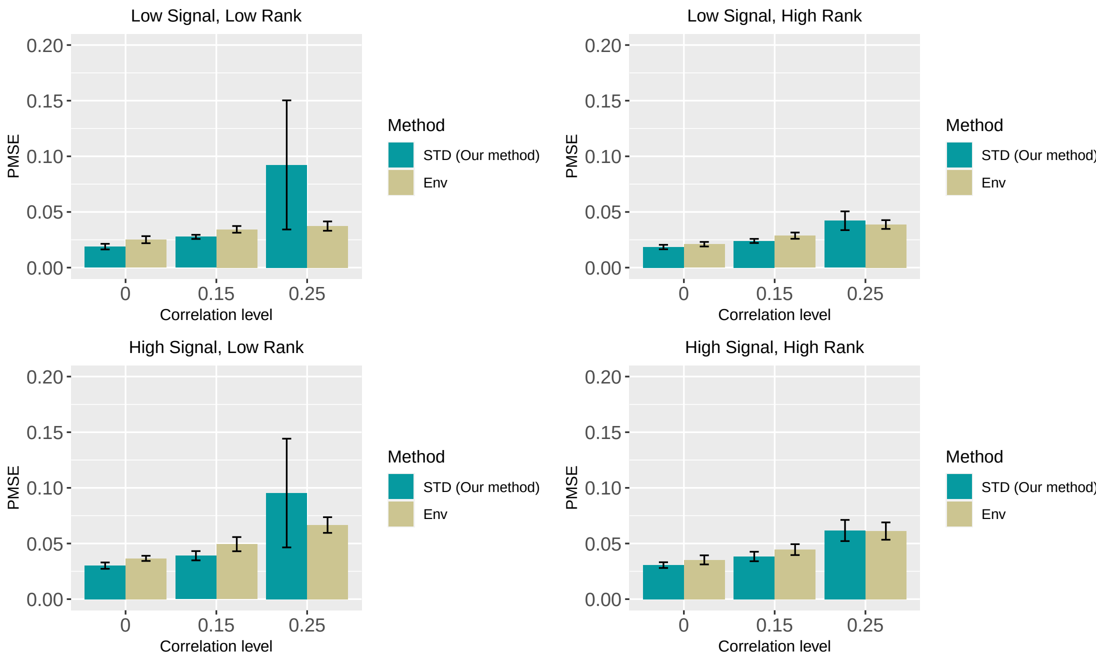

Figure 8: Sanity check for envelope method. Data is generated by envelope method mentioned above. The y-axis is the PMSE and x-axis is the correlation level γ.

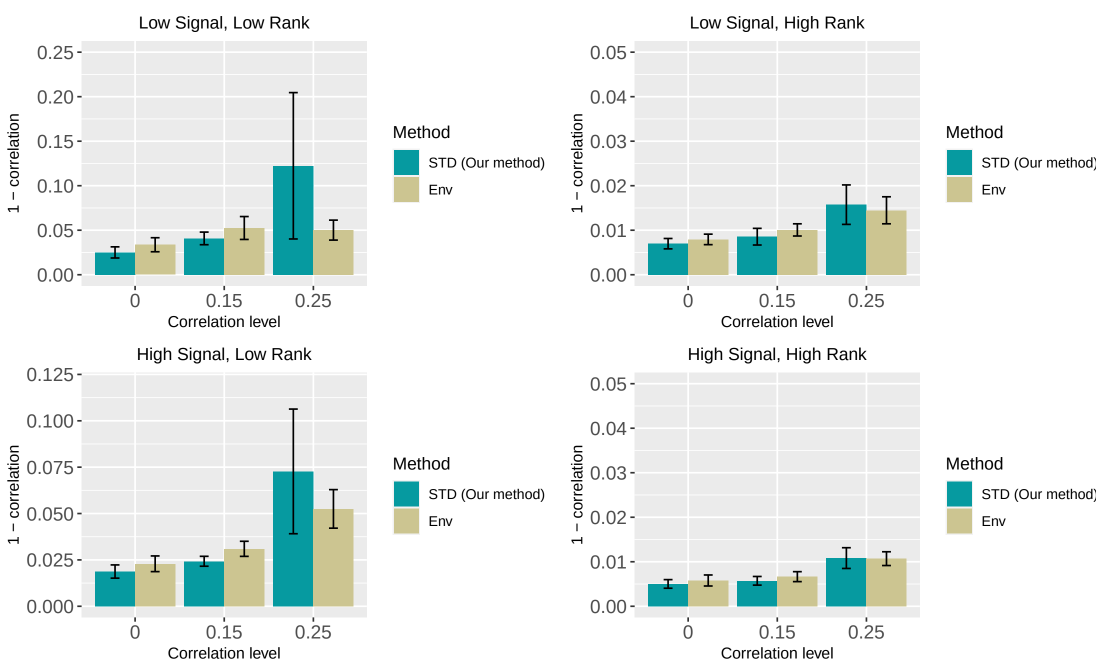

Figure 9: Sanity check for envelope method. Data is generated by envelope method mentioned above. The y-axis is $1 - Correlation$ and x-axis is the correlation level γ.