Model-free multiple testing for matrix-valued predictors with false discovery control

Lei Yan
Department of Statistics, Florida State University
and
Xin Zhang*
Department of Statistics, Florida State University
and
Zhigen Zhao
Department of Statistics, Operations, and Data Science, Temple University

April 29, 2026

Abstract

Identifying influential variables in high-dimensional matrix-valued data while controlling the false discovery rate (FDR) is a critical challenge in modern data science. We propose a novel, model-free procedure specifically designed for simultaneous row and column selection in matrix predictor regression. Our approach utilizes folding selection subspaces (FSS) to formulate structured hypotheses and employs data splitting to construct mirror statistics from FSS estimators. This design bypasses restrictive model specification and the need for p-value computation. Key theoretical contributions include establishing the asymptotic distribution of FSS estimators and proving the mirror statistic is asymptotically symmetric with respect to zero under the null hypothesis. Using the symmetry, we develop a multiple hypothesis testing procedure with data-driven thresholds that provably controls the FDR for row and column at the desired level asymptotically. The framework is further extended to control element-wise FDR under specific structural assumptions. Extensive simulations and a real data analysis demonstrate the superior performance of the proposed method over existing approaches across various settings.

Keywords: Mirror statistics, variable selection, dimension reduction

*Corresponding author: Xin Zhang (henry@stat.fsu.edu), whose research for this paper was supported in part by grant DMS-2053697 from the U.S. National Science Foundation. The authors are grateful to the Co-Editor, Associate Editor, and anonymous referees, whose suggestions led to improvement of this work.

1 Introduction

Matrix-valued data are prevalent in modern science, arising in fields such as electroencephalography (EEG) (Li et al. 2010), colorimetric sensor array (Zhong & Suslick 2015), and nuclear magnetic resonance (Zhao & Leng 2014). These data sets are large and structured, with each dimension representing different types of information. For instance, in EEG data, rows correspond to electrodes and columns to time points. A critical challenge lies in analyzing such data while keeping its inherent structure. Simply vectorizing matrices and applying traditional vector-based methods often discards crucial structural information. The high dimensionality typical of these data sets further complicates analysis.

Many variable selection techniques have been proposed for matrix predictors. For example, the penalization method (Zhao & Leng 2014) utilizes structured Lasso (Tibshirani 1996) to select rows/columns but relies on the linear model assumption. Li & Dong (2021) introduced a trace statistic to test the importance of rows/columns and established the selection consistency theory under the constraint that the matrix dimensions remain fixed as the sample size goes to infinity. However, these methods lack formal procedures for quantifying selection uncertainty. Specifically, providing guarantees for controlling the false discovery rate (FDR) among selected variables is a critical gap. Although test statistics from Li & Dong (2021) could be used with standard FDR procedures such as Benjamini-Hochberg (BH) (Benjamini & Hochberg 1995) or Benjamini-Yekutieli (BY) (Benjamini & Yekutieli 2001), formal theoretical guarantees for controlling FDR are lacking, especially in the challenging high-dimensional and model-free settings considered here. Liu et al. (2023) developed hypothesis testing for row and column significance under an additive model for a matrix response and a scalar predictor. They further utilized multiple testing procedures, such as BY, to control FDR. In contrast to these works, we develop an efficient procedure to select active rows and columns of a matrix predictor within a model-free framework

while rigorously controlling the FDR. The new concept and method are generalizable to various tensor data analysis problems such as regression (Zhou et al. 2013, Zhang & Li 2017), sufficient dimension reduction (Li et al. 2010, Wang et al. 2022), clustering (Mai et al. 2022), and discriminant analysis (Wang et al. 2024).

To avoid restrictive assumptions often challenged in real applications, we consider the following general model:

$$Y = f(\mathbf{X}, \epsilon), \tag{1}$$

where $Y \in \mathbb{R}$ is a univariate response, $\mathbf{X} \in \mathbb{R}^{p\times q}$ is the matrix predictor, $f(\cdot) : \mathbb{R}^{p\times q} \times \mathbb{R} \mapsto \mathbb{R}$ is a unknown function, and ϵ is a noise term independent of $\mathbf{X}$ with zero mean. Under high-dimensional settings, we assume sparsity that Y depends on $\mathbf{X}$ only through a subset of its rows and columns. Let $\mathbf{X}_{j\cdot}$ denote the j-th row and $\mathbf{X}_{\cdot k}$ the k-th column of $\mathbf{X}$ for $j \in [p] \coloneqq \{1, \ldots, p\}$ and $k \in [q] \coloneqq \{1, \ldots, q\}$. Let $F(Y|\mathbf{X})$ denote the conditional distribution of Y given $\mathbf{X}$. The active row and column sets are defined as

$$\mathcal{I}_L = \{j \in [p] : F(Y|\mathbf{X}) \text{ functionally depends on } \mathbf{X}_{j\cdot}\},$$
$$\mathcal{I}_R = \{k \in [q] : F(Y|\mathbf{X}) \text{ functionally depends on } \mathbf{X}_{\cdot k}\}.$$

In essence, row j is considered active if $F(Y|\mathbf{X})$ changes when $\mathbf{X}_{j\cdot}$ is varied, and similarly for columns. Then $\mathcal{I}_L^c$ and $\mathcal{I}_R^c$, the complement of $\mathcal{I}_L$ and $\mathcal{I}_R$, are the index sets of all irrelevant rows and columns. Let $\mathbf{X}_{\mathcal{I}_L\mathcal{I}_R}$ denote the submatrix formed by rows in $\mathcal{I}_L$ and columns in $\mathcal{I}_R$. Thus, the active submatrix $\mathbf{X}_{\mathcal{I}_L\mathcal{I}_R}$ contains all predictive information within $\mathbf{X}$ concerning Y. Formally, we have Y is independent of $\mathbf{X}$ given $\mathbf{X}_{\mathcal{I}_L\mathcal{I}_R}$, denoted as $Y \perp\!\!\!\perp \mathbf{X}|\mathbf{X}_{\mathcal{I}_L\mathcal{I}_R}$, where $\perp\!\!\!\perp$ means independence. Our primary goal is to perform multiple hypothesis testing to identify the active rows and columns. Specifically, for all $j \in [p]$ and $k \in [q]$, we test the null hypotheses:

$$\mathcal{H}_j^L : j \in \mathcal{I}_L^c, \quad \text{or} \quad \mathcal{H}_k^R : k \in \mathcal{I}_R^c. \tag{2}$$

Such hypotheses arise naturally in scientific studies, e.g., identifying influential electrodes or relevant time points in EEG data analysis. When testing multiple hypotheses simultaneously, controlling FDR is crucial for reliably detecting true alternatives (Benjamini & Hochberg 1995). The FDR is the expectation of false positive proportion (FDP). For rows, letting $\widehat{\mathcal{I}}_L$ be the rejection set of row hypotheses in (2), the FDR is:

$$\mathrm{FDP}^L = \frac{\#\{j : j \notin \mathcal{I}_L, j \in \widehat{\mathcal{I}}_L\}}{\#\{j \in \widehat{\mathcal{I}}_L\} \vee 1}, \quad \mathrm{FDR}^L = \mathrm{E}[\mathrm{FDP}^L],$$

where $a \vee b$ denotes $\max(a, b)$. An analogous definition applies to FDP^R and FDR^R for columns.

This problem is challenging due to the unknown function f and matrix structure. Even in the simpler vector case ($q = 1$), controlling FDR under unknown f is non-trivial. Many procedures (Barber & Candès 2015, Javanmard & Javadi 2019, Ma et al. 2021, Dai et al. 2023*b*,*a*, Xing et al. 2023) have been proposed to control FDR when f is known (e.g., linear models), where test statistics are constructed using estimated coefficients. The model-free case requires different strategies. The Model-X Knockoff (Candès et al. 2018) achieves FDR control by creating knockoff features. But it requires the joint distribution of predictors to be known, which is often unavailable. Another line of work is based on sufficient dimension reduction (SDR) (Li 1991, Cook 2004), including (Zhao & Xing 2022, Yang et al. 2024). Typically, the hypothesis testing problem is reformulated via the SDR framework. These FDR control procedures rely on the central subspace in SDR and asymptotic properties of related statistics. Extending these vector-based methods to handle matrix predictors, without specifying f and the distribution of $\mathbf{X}$, presents significant hurdles.

Dimension Reduction Frameworks. SDR methods often seek a minimal subspace for a vector predictor such that Y is independent of the predictor given its projection onto that subspace. The dimension folding subspace (DFS) framework (Li et al. 2010) extends this to matrix predictors $\mathbf{X}$. DFS finds low-dimensional subspaces spanned by columns of

matrices $\boldsymbol{\alpha} \in \mathbb{R}^{p\times d_p}$ and $\boldsymbol{\beta} \in \mathbb{R}^{q\times d_q}$, such that $Y \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\alpha}^T\mathbf{X}\mathbf{B}$. This achieves dimension reduction via linear combinations of rows and columns. However, our focus is variable selection: identifying active rows and columns, denoted by sets $\mathcal{I}_L$ and $\mathcal{I}_R$. This motivates using folding selection subspaces (FSS), which generalizes variable selection subspace (Yin & Hilafu 2015). The left (row) FSS, denoted $\mathcal{S}^V_{Y|\circ\mathbf{X}}$, is spanned by standard basis vectors corresponding to active rows, while the right (column) FSS, $\mathcal{S}^V_{Y|\mathbf{X}\circ}$, is defined similarly. Unlike DFS which combines rows and columns, FSS directly captures the sparse structure in our problem (1). Thus, FSS provides a natural framework for reformulating and testing hypotheses (2).

Contributions. In this paper, we address the challenge of FDR control for matrix-valued $\mathbf{X}$ without knowing the function f. While existing works in dimension folding mainly focus on estimation, we introduce, to the best of our knowledge, the first rigorous framework for statistical inference based on this concept in the challenging setting where p and q diverge with sample size. Our major contributions are summarized as follows.

- We propose a novel model-free procedure for large-scale simultaneous hypothesis testing on rows and columns of a matrix predictor, achieving FDR control in high dimensions. This contrasts with prior works limited to parametric models or vector predictors.
- We construct test statistics for rows and columns based on folding section subspaces and prove they are asymptotically symmetric about zero under the null hypothesis. Then, we propose a single-step procedure that asymptotically controls the row and column FDR at any given level (refer to Section 3.1 for details).
- We further derive an element-wise FDR control procedure based on our row and column statistics, and provide theoretical and empirical evidence for its validity under certain conditions (the procedure and properties are presented in Section 3.2).

Notations. For two real numbers a and b, we use $a \wedge b$ to represent $\min(a, b)$. For two positive real sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ means there exists a universal constant C such that $a_n \leq Cb_n$ for all n, and $a_n \ll b_n$ or $b_n \gg a_n$ means $a_n/b_n \to 0$ as $n \to \infty$. $a_n \asymp b_n$ represents $a_n \lesssim b_n$ and $a_n \gtrsim b_n$, which means the two sequences are in the same order. In this article, we assume $p \asymp q$. For a $p \times q$ matrix $\mathbf{A} = (A_{ij})$, span($\mathbf{A}$) represents the subspace of $\mathbb{R}^p$ spanned by columns of $\mathbf{A}$, $\mathbf{P}_{\text{span}(\mathbf{A})}$ stands for the projection onto span($\mathbf{A}$), and vec($\mathbf{A}$) is the vectorization of $\mathbf{A}$. Let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ be the smallest and largest eigenvalues of $\mathbf{A}$. The max norm and L_1 norm of $\mathbf{A}$ are defined as $\|\mathbf{A}\|_{\max} = \max_{i,j}|A_{ij}|$ and $\|\mathbf{A}\|_1 = \max_j \sum_i |A_{ij}|$. For a set $\mathcal{I}$, $|\mathcal{I}|$ is the cardinality. The vector $\mathbf{e}_j$ is a unit vector with j-th element 1 and 0 otherwise.

2 Problem setup

We begin by formalizing the folding selection subspaces. Let $p_1 = |\mathcal{I}_L|$, and $q_1 = |\mathcal{I}_R|$. Construct matrices $\mathbf{A} \in \mathbb{R}^{p \times p_1}$ and $\mathbf{B} \in \mathbb{R}^{q \times q_1}$ whose columns are the standard basis vectors $\{\mathbf{e}_j : j \in \mathcal{I}_L\}$ and $\{\mathbf{e}_k : k \in \mathcal{I}_R\}$, respectively. The conditional independence condition $Y \perp\!\!\!\perp \mathbf{X} | \mathbf{X}_{\mathcal{I}_L \mathcal{I}_R}$ from the introduction is equivalent to

$$Y \perp\!\!\!\perp \mathbf{X} | \mathbf{A}^T \mathbf{X} \mathbf{B}. \tag{3}$$

Since the ordering of the standard basis vectors within $\mathbf{A}$ and $\mathbf{B}$ is arbitrary, the specific matrices $\mathbf{A}$ and $\mathbf{B}$ are not unique. But the column spaces, span($\mathbf{A}$) and span($\mathbf{B}$), are uniquely defined by index sets $\mathcal{I}_L$ and $\mathcal{I}_R$. This motivates the following definition of folding selection subspaces:

Definition 2.1. If there exist matrices $\mathbf{A} \in \mathbb{R}^{p \times p_1}$ and $\mathbf{B} \in \mathbb{R}^{q \times q_1}$, where columns of $\mathbf{A}$ and $\mathbf{B}$ consist of standard basis vectors, such that $Y \perp\!\!\!\perp \mathbf{X} | \mathbf{A}^T \mathbf{X} \mathbf{B}$ holds, the column spaces span($\mathbf{A}$) and span($\mathbf{B}$) are called left and right folding selection space. The intersection of

all such left (or right) folding selection space is denoted $\mathcal{S}^V_{Y|\circ\mathbf{X}}$ (or $\mathcal{S}^V_{Y|\mathbf{X}\circ}$). The subspace $\mathcal{S}^V_{Y|\circ\mathbf{X}\circ} = \mathcal{S}^V_{Y|\circ\mathbf{X}} \otimes \mathcal{S}^V_{Y|\mathbf{X}\circ}$ is called the central folding selection space.

As discussed in the introduction, the FSS framework relates to dimension folding subspace (Li et al. 2010). The subspace $\text{span}(\boldsymbol{\alpha}) \otimes \text{span}(\boldsymbol{\beta})$ is called the central DFS, where $\text{span}(\boldsymbol{\alpha})$ and $\text{span}(\boldsymbol{\beta})$ are the left and right DFS. By definition, we have $\text{span}(\boldsymbol{\alpha}) \subset \mathcal{S}^V_{Y|\circ\mathbf{X}}$ and $\text{span}(\boldsymbol{\beta}) \subset \mathcal{S}^V_{Y|\mathbf{X}\circ}$. If $\text{span}(\boldsymbol{\alpha})$ and $\text{span}(\boldsymbol{\beta})$ exist, $\mathcal{S}^V_{Y|\circ\mathbf{X}}$ and $\mathcal{S}^V_{Y|\mathbf{X}\circ}$ exist and unique. A key difference relates to invariance properties. While DFS is invariant under non-singular linear transformations, a sparse solution in one scale could result in a non-sparse solution in another. Therefore, the sparsity of the left and right folding selection subspace should be interpreted relative to the coordinate system.

Using FSS, the hypothesis defined in (2) is equivalent to

$$\mathcal{H}^L_j : \mathbf{P}_{\text{span}(\mathbf{e}_j)}\mathcal{S}^V_{Y|\circ\mathbf{X}} = \mathbf{0}, \quad \text{or} \quad \mathcal{H}^R_k : \mathbf{P}_{\text{span}(\mathbf{e}_k)}\mathcal{S}^V_{Y|\mathbf{X}\circ} = \mathbf{0}. \tag{4}$$

In other words, declaring $\mathbf{X}_{j\cdot}$ or $\mathbf{X}_{\cdot k}$ to be inactive is equivalent to stating the j-th or k-th row of a basis matrix for left or right FSS consists of all zeros. Assume $\mathbf{X}$ follows a matrix normal distribution, $\mathbf{X} \sim \mathcal{MN}_{p,q}(\mathbf{0}, \mathbf{U}, \mathbf{V})$, with precision matrices $\boldsymbol{\Omega} = \mathbf{U}^{-1}$ and $\boldsymbol{\Gamma} = \mathbf{V}^{-1}$. The left and right FSS can be estimated using precision matrices and expectation of $Y\mathbf{X}$, according to the following result.

Lemma 2.2. *If* $Y \perp\!\!\!\perp \mathbf{X}|\mathbf{A}^T\mathbf{X}\mathbf{B}$ *and* $\mathbf{X} \sim \mathcal{MN}_{p,q}(\mathbf{0}, \mathbf{U}, \mathbf{V})$*, then* $\text{span}(\boldsymbol{\Omega}\,\text{E}[Y\mathbf{X}]) \subseteq \text{span}(\mathbf{A})$ *and* $\text{span}(\boldsymbol{\Gamma}\,\text{E}[Y\mathbf{X}^T]) \subseteq \text{span}(\mathbf{B})$.

The expectation $\text{E}[Y\mathbf{X}]$ captures information about the relationship between Y and $\mathbf{X}$, conceptually linking it to the principal Hessian matrix in classical SDR (Li 1992). This type of expectation enables applications like interaction detection (Tang et al. 2020) and liquid association (Li et al. 2023). Lemma 2.2 suggests that left and right FSS can be partially identifiable. When the equalities $\text{span}(\boldsymbol{\Omega}\,\text{E}[Y\mathbf{X}]) = \mathcal{S}^V_{Y|\circ\mathbf{X}}$ and $\text{span}(\boldsymbol{\Omega}\,\text{E}[Y\mathbf{X}^T]) = \mathcal{S}^V_{Y|\mathbf{X}\circ}$

hold, the coverage condition or the exhaustiveness is said to be satisfied. The exhaustiveness does not hold universally for all SDR and dimension folding methods. Even when this property holds, methods like directional regression (Li & Wang 2007) can fail in practice due to unmet assumptions. In practice, verifying the exhaustiveness requires additional diagnostic procedures such as conditional independence tests (Chen et al. 2015, Li et al. 2025). Since our focus in this work is on FDR control, we assume the coverage condition holds, following the literature (e.g., Cook & Ni 2006).

Remark 2.3. Our proposed statistics in Section 3.1 rely on the expectation $\mathrm{E}[Y\mathbf{X}]$. As in classical SDR, if the conditional mean function $\mathrm{E}[Y \mid \mathbf{X}]$ is symmetric and $\mathbf{X}$ is symmetrically distributed about zero, $\mathrm{E}[Y\mathbf{X}] = \mathbf{0}$ and all methods based on it could fail. In this case, the coverage condition or the exhaustiveness for the folding selection subspaces is violated. By assuming the coverage condition holds, we exclude this degenerate scenario. We leave the development of methods that remain effective under symmetric link functions for future work.

The normality assumption on $\mathbf{X}$ is used to apply Stein's Lemma in the proof, same as the popular principal Hessian directions method (Li 1992) in SDR. Landsman & Nešlehová (2008) extended Stein's Lemma to the broader class of elliptical distributions. This result suggests that Lemma 2.2 remains meaningful and retains practical utility when predictors deviate from normality. Simulation results in Section 4 demonstrate the robustness of our method under such deviations.

In Section A of the Supplement, we derive a central limit theorem for the sample estimators of the folding selection subspaces via Gaussian approximation theory (Chernozhukov et al. 2013, 2017). While this theorem is not used directly in the theoretical development of the mirror statistics in Section 3.1, the technical framework established in its proof provides the essential tools for verifying the mirror statistics property in Theorem 3.1.

3 Multiple testing using mirror statistics

3.1 Row and column testing

We aim to test hypotheses $\mathcal{H}_j^L$ and $\mathcal{H}_k^R$ defined in (4) simultaneously, controlling FDR at a target level. Our approach utilizes a data splitting strategy and mirror statistics (Dai et al. 2023*a*, Zhao & Xing 2022). First, we randomly split the data set $\mathcal{D} = \{Y_i, \mathbf{X}_i\}_{i=1}^n$ into two disjoint subsets of equal size, $\mathcal{D}_1$ and $\mathcal{D}_2$. For each subset $m \in \{1, 2\}$, we estimate the precision matrices $\widehat{\mathbf{\Omega}}^{(m)}$ and $\widehat{\mathbf{\Gamma}}^{(m)}$ using only data from $\mathcal{D}_m$, and compute the sample expectations $\widehat{\mathrm{E}}^{(m)}[Y\mathbf{X}] = |\mathcal{D}_m|^{-1}\sum_{i\in\mathcal{D}_m} Y_i\mathbf{X}_i$. We then construct estimators $\widehat{\mathbf{A}}^{(m)} = \widehat{\mathbf{\Omega}}^{(m)}\widehat{\mathrm{E}}^{(m)}[Y\mathbf{X}]$ and $\widehat{\mathbf{B}}^{(m)} = \widehat{\mathbf{\Gamma}}^{(m)}\widehat{\mathrm{E}}^{(m)}[Y\mathbf{X}]^T$ for left and right FSS. Using these, we define the row and column mirror statistics:

$$u_j = \sum_{l=1}^{q} \widehat{A}_{jl}^{(1)}\widehat{A}_{jl}^{(2)}, \quad v_k = \sum_{l=1}^{p} \widehat{B}_{kl}^{(1)}\widehat{B}_{kl}^{(2)}, \quad j = 1, \ldots, p, k = 1, \ldots, q. \tag{5}$$

By definition, row j is inactive ($\mathcal{H}_j^L$ is true) if and only if $\mathbf{P}_{\mathrm{span}(\mathbf{e}_j)}\mathcal{S}^V_{Y|\circ\mathbf{X}} = \mathbf{0}$, where $\mathbf{P}_{\mathrm{span}(\mathbf{e}_j)}$ denotes the projection onto j-th standard basis vector. This condition implies that every vector in $\mathcal{S}^V_{Y|\circ\mathbf{X}}$ has a zero in the j-th coordinate. Since $\mathbf{A}$ spans this subspace, the j-row $\mathbf{A}_{j\cdot}$ must be zero. The estimators should thus oscillate independently around zero. Conversely, if row j is inactive ($\mathcal{H}_j^L$ is false), both vectors $\widehat{\mathbf{A}}_{j\cdot}^{(1)}$ and $\widehat{\mathbf{A}}_{j\cdot}^{(2)}$ center around a common non-zero vector $\mathbf{A}_{j\cdot}$, causing u_j to tend to take positive values. Similar intuition applies to v_k for columns. The following theorem provides a rigorous argument for the above reasoning by showing the sampling distributions of u_j and v_k are asymptotically symmetric about zero under the null hypothesis.

Theorem 3.1. *Assume Assumptions B.1 and B.2 in Section B of Supplement hold and* $s_1\log(p) \vee s_2\log(q) = o(\sqrt{n})$. *If the null hypothesis is true, then for any* $t > 0$, *it holds that*

$$\sup_t |\Pr(u_j < -t) - \Pr(u_j > t)| \leq C\frac{s_1\log(p)}{\sqrt{nq}}, \sup_t |\Pr(v_k < -t) - \Pr(v_k > t)| \leq C\frac{s_2\log(q)}{\sqrt{np}}.$$

The above asymptotic symmetry property enables the development of an FDR control procedure using the mirror statistics u_j and v_k. We consider a single-step rule: reject $\mathcal{H}_j^L$ if $u_j > \tau_a$ and reject $\mathcal{H}_k^R$ if $v_k > \tau_b$ for some threshold τ_a and τ_b. Define the decision rules $\delta_j^L = \mathbb{1}(u_j > t)$ and $\delta_k^R = \mathbb{1}(v_k > t)$ for any value t. Then the false discovery proportions at threshold t for hypothesis (4) are

$$\mathrm{FDP}^L(t) = \frac{\sum_{j \in \mathcal{I}_L^c} \delta_j^L}{\sum_j \delta_j^L \vee 1}, \quad \mathrm{FDP}^R(t) = \frac{\sum_{k \in \mathcal{I}_R^c} \delta_k^R}{\sum_k \delta_k^R \vee 1}.$$

Theorem 3.1 suggests that under the null hypothesis, $\Pr(u_j > t) \approx \Pr(u_j < -t)$ and $\Pr(v_k > t) \approx \Pr(v_k < -t)$, motivating the estimation of the number of false positives using the count of statistics below $-t$. This leads to the estimated false discovery proportion:

$$\widehat{\mathrm{FDP}}^L(t) = \frac{\sum_j \mathbb{1}(u_j < -t)}{\sum_j \mathbb{1}(u_j > t) \vee 1}, \quad \widehat{\mathrm{FDP}}^R(t) = \frac{\sum_k \mathbb{1}(v_k < -t)}{\sum_k \mathbb{1}(v_k > t) \vee 1}.$$

Note that $\widehat{\mathrm{FDP}}^L(t)$ and $\widehat{\mathrm{FDP}}^R(t)$ tend to slightly overestimate the true false discovery proportions. However, this overestimation is acceptable or even desirable for two reasons: (1) it avoids an excessive number of false positives, and (2) $\Pr(u_j < -t)$ (or $\Pr(v_k < -t)$) is negligible when $\mathcal{H}_j^L$ (or $\mathcal{H}_k^R$) is false. For any designated level a for rows and b for columns, we choose data-driven thresholds τ_a and τ_b by controlling the estimated FDP:

$$\tau_a = \min\{t : \widehat{\mathrm{FDP}}^L(t) \le a\}, \quad \tau_b = \min\{t : \widehat{\mathrm{FDP}}^R(t) \le b\}. \tag{6}$$

The full procedure is summarized in Algorithm 1.

Algorithm 1: Model-Free Multiple Testing for Matrix-Valued Predictors

split sample into two parts $\mathcal{D}_1$ and $\mathcal{D}_2$; and compute $\widehat{\mathbf{A}}_1$ and $\widehat{\mathbf{B}}_1$ using $\mathcal{D}_1$, $\widehat{\mathbf{A}}_2$ and $\widehat{\mathbf{B}}_2$ using $\mathcal{D}_2$;

for $j = 1, \ldots, p, k = 1, \ldots, q$ **do**

  compute $u_j = \sum_{l=1}^{q} \widehat{A}_{jl}^{(1)} \widehat{A}_{jl}^{(2)}, v_k = \sum_{l=1}^{p} \widehat{B}_{kl}^{(1)} \widehat{B}_{kl}^{(2)}$;

end

Find the thresholds τ_a and τ_b such that

$$\tau_a = \min\{t : \widehat{\mathrm{FDP}}^{L}(t) \le a\}, \quad \tau_b = \min\{t : \widehat{\mathrm{FDP}}^{R}(t) \le b\};$$

for $j = 1, \ldots, p$ **do**

  reject $\mathcal{H}_j^L$ if $u_j > \tau_a$;

end

for $k = 1, \ldots, q$ **do**

  reject $\mathcal{H}_k^R$ if $v_k > \tau_b$;

end

We now establish the asymptotic FDR control properties of the proposed testing procedure. Define the true FDP based on mirror statistics as $\mathrm{FDP}_{\dagger}^{L}(t) = \sum_{j \in \mathcal{I}_L^c} \mathbb{1}(u_j < -t) / [\sum_j \mathbb{1}(u_j > t) \vee 1]$, and similarly for $\mathrm{FDP}_{\dagger}^{R}(t)$.

Theorem 3.2. *Let* $\tau_a^* = \inf\{t : \mathrm{FDP}_{\dagger}^{L}(t) \le a\}$ *and* $\tau_b^* = \inf\{t : \mathrm{FDP}_{\dagger}^{R}(t) \le b\}$. *For any target levels* $a, b \in (0, 1)$, *assume the pointwise limits* $\mathrm{FDP}_{\infty}^{L}(t)$ *and* $\mathrm{FDP}_{\infty}^{R}(t)$ *of* $\mathrm{FDP}^{L}(t)$ *and* $\mathrm{FDP}^{R}(t)$ *exist for all* $t > 0$. *Further suppose there exist constants* $t_a, t_b > 0$ *such that* $\mathrm{FDP}_{\infty}^{L}(t_a) \le a$ *and* $\mathrm{FDP}_{\infty}^{R}(t_b) \le b$. *Under Assumptions B.1-B.4 in Section B in Supplement and assuming* $s_1 \log(p) \vee s_2 \log(q) = o(\sqrt{n})$, *we have*

$$\limsup_{n,p \to \infty} \mathrm{FDR}^{L}(\tau_a^*) \le a, \quad \limsup_{n,q \to \infty} \mathrm{FDR}^{R}(\tau_b^*) \le b.$$

Theorem 3.2 establishes the FDR control when using the oracle cutoffs τ_a^* and τ_b^* instead of

their estimators τ_a and τ_b defined in (6). By construction, the empirical cutoffs overestimate τ_a^* and τ_b^*. With additional assumptions, these results can be extended to thresholds τ_a and τ_b. To this end, define $\mathcal{S}^L_{1,\text{strong}}$ and $\mathcal{S}^R_{1,\text{strong}}$ as the largest subset of $\mathcal{I}^L$ and $\mathcal{I}^R$ such that $\mathbb{1}\left(\min_{j\in\mathcal{S}^L_{1,\text{strong}}} u_j > t_1\right) \xrightarrow{P} 1, \mathbb{1}\left(\min_{k\in\mathcal{S}^R_{1,\text{strong}}} v_k > t_2\right) \xrightarrow{P} 1$, for some $t_1, t_2 > 0$. Let $p_{1,\text{strong}} = |\mathcal{S}^L_{1,\text{strong}}|$ and $q_{1,\text{strong}} = |\mathcal{S}^R_{1,\text{strong}}|$. A similar notion of a strong signal set appears in (Dai et al. 2023*b*) for generalized linear models, where variables in the strong signal set are those with sufficiently large coefficients. We need an additional Assumption B.5 in Section B of Supplement on the strong signal set for the following theorem.

Theorem 3.3. *For any target levels $a, b \in (0,1)$, under Assumption B.1-B.5 in Section B of Supplement and assuming $s_1 \log(p) \vee s_2 \log(q) = o(\sqrt{n})$, we have*

$$\limsup_{n,p\to\infty} \text{FDR}^L(\tau_a) \le a, \quad \limsup_{n,q\to\infty} \text{FDR}^R(\tau_b) \le b.$$

Remark 3.4. The theoretical analysis in this section, including the techniques used to establish asymptotic symmetry of mirror statistics (Theorem 3.1) and to prove asymptotic FDR control in high dimensions (Theorems 3.2), builds upon the framework developed in (Zhao & Xing 2022) for vector predictors. (Zhao & Xing 2022) applied sliced inverse regression (SIR) (Li 1991) to construct angular balanced statistics (ABS), and proposed a model-free testing procedure using ABS (MTA). Our work generalizes this approach to matrix-valued predictors, extending their method to a broader class of problems.

3.2 Element-wise testing

Beyond selecting active rows and columns, statistics u_j and v_k can be combined to infer the importance of individual elements X_{jk}. Specifically, consider the hypothesis

$$\mathcal{H}_{jk} : \mathbf{P}_{\text{span}(\mathbf{e}_j)}\mathcal{S}^V_{Y|\circ\mathbf{X}} = \mathbf{0} \text{ or } \mathbf{P}_{\text{span}(\mathbf{e}_k)}\mathcal{S}^V_{Y|\mathbf{X}\circ} = \mathbf{0}, \tag{7}$$

where $\mathcal{H}_{jk}$ is rejected if $u_j > \tau_a$ and $v_k > \tau_b$. Rejecting $\mathcal{H}_{jk}$ suggests both row j and column k are active. However, this does not necessarily imply the specific element $\mathbf{X}_{jk}$ is active. A hypothesis directly addressing the importance of element $\mathbf{X}_{jk}$ requires considering the vectorized predictor $\text{vec}(\mathbf{X})$ and its central variable selection subspace (CVSS), denoted $\mathcal{S}^V_{Y|\text{vec}(\mathbf{X})}$ (Yin & Hilafu 2015). The CVSS is the minimal subspace satisfying $Y \perp\!\!\!\perp \text{vec}(\mathbf{X})|\mathbf{P}_{\mathcal{S}^V_{Y|\text{vec}(\mathbf{X})}}\text{vec}(\mathbf{X})$. Then, the element-wise hypothesis is:

$$\widetilde{\mathcal{H}}_{jk} : \mathbf{P}_{\text{span}(\widetilde{\mathbf{e}}_{j+(k-1)p})}\mathcal{S}^V_{Y|\text{vec}(\mathbf{X})} = \mathbf{0}, \tag{8}$$

where $\widetilde{\mathbf{e}}_{j+(k-1)p}$ is a standard basis vector in $\mathbb{R}^{pq}$. Comparing the two hypotheses, $\mathcal{H}_{jk}$ being true implies $\widetilde{\mathcal{H}}_{jk}$ is true (if row j or column k is inactive, element (j,k) must be inactive). However, the alternative $\mathcal{H}^a_{jk}$ (row j active AND column k active) does not necessarily imply the alternative $\widetilde{\mathcal{H}}^a_{jk}$ (element (j,k) active). This discrepancy arises because the CVSS $\mathcal{S}^V_{Y|\text{vec}(\mathbf{X})}$ does not, in general, possess a Kronecker product structure. Corollary 3.5 provides the necessary and sufficient condition for the equivalence of $\mathcal{H}_{jk}$ and $\widetilde{\mathcal{H}}_{jk}$.

Corollary 3.5. *The hypotheses $\mathcal{H}_{jk}$ and $\widetilde{\mathcal{H}}_{jk}$ are equivalent if and only if the central variable selection space $\mathcal{S}^V_{Y|\text{vec}(\mathbf{X})} = \mathcal{S}^V_{Y|\circ\mathbf{X}} \otimes \mathcal{S}^V_{Y|\mathbf{X}\circ}$.*

In other words, the equivalence holds if and only if the active elements form a submatrix after row and column permutations. That is, they need not be aligned as a submatrix in their original positions. For example, if the active elements are at positions (1, 1), (1, 3), (3, 1), and (3, 3) in a 4×4 matrix, this condition holds since swapping column 3 with column 2 and row 3 with row 2 would yield a 2×2 submatrix.

Under the equivalence condition in Corollary 3.5, we can analyze the FDR of the element-wise selection rule $\boldsymbol{\delta} = (\delta_{jk}), \delta_{jk} = \mathbb{1}(u_j > \tau_a, v_k > \tau_b)$. Let $\theta_{jk} = \theta^L_j\theta^R_k$ be the indicator of whether the hypothesis $\mathcal{H}_{jk}$ is false, and $\mathcal{S}_0 = \{(j,k) : \theta_{jk} = 0\}$ be the set true null

hypotheses. The FDP of this procedure is

$$\mathrm{FDP}(t_1, t_2) = \frac{\sum_{(j,k):\theta_{jk}\in\mathcal{S}_0} \delta_{jk}}{\sum_{(j,k)} \delta_{jk} \vee 1}.$$

The following result provides FDR control by leveraging the separate row/column controls.

Corollary 3.6. *If* $\mathcal{S}^V_{Y|\mathrm{vec}(\mathbf{X})} = \mathcal{S}^V_{Y|\circ\mathbf{X}} \otimes \mathcal{S}^V_{Y|\mathbf{X}\circ}$, *we have*

$$\limsup_{n,p,q\to\infty} \mathrm{FDR}(\tau_a^*, \tau_b^*) \le a + b - ab, \;\; \textit{or} \;\; \limsup_{n,p,q\to\infty} \mathrm{FDR}(\tau_a, \tau_b) \le a + b - ab,$$

assuming the same assumptions of Theorem 3.2 or Theorem 3.3, respectively.

The bound $a + b - ab = 1 - (1 - a)(1 - b)$ combines the FDR control levels from row and column selections. To achieve an overall element-wise FDR control at a target level $\gamma \in (0, 1)$, one can set the same levels $a = b = 1 - \sqrt{1 - \gamma}$ for the row and column tests in Algorithm 1, assuming no prior preference between rows and columns. If domain knowledge or constraints suggest stricter FDR control along rows than columns, then a and b can be adjusted accordingly, as long as $a + b - ab = \gamma$.

When the equivalence condition does not hold, the procedure does not necessarily control the FDR for the true element-wise hypothesis $\widetilde{\mathcal{H}}_{jk}$. Let $\widetilde{\mathcal{S}}_0$ be the set where $\widetilde{\mathcal{H}}_{jk}$ is true. The FDP is

$$\widetilde{\mathrm{FDP}}(t_1, t_2) = \frac{\sum_{(j,k):\theta_{jk}\in\mathcal{S}_0} \delta_{jk} + \sum_{(j,k):\theta_{jk}\in\widetilde{\mathcal{S}}_0\setminus\mathcal{S}_0} \delta_{jk}}{\sum_{(j,k)} \delta_{jk} \vee 1} \coloneqq \mathrm{FDP}(t_1, t_2) + \zeta(t_1, t_2),$$

where $\zeta(t_1, t_2)$ captures the portion of false discoveries that cannot be estimated using only the row and column test statistics. Since $\zeta(t_1, t_2)$ is unknown and non-negative, using the estimator $\widehat{\mathrm{FDP}}(t_1, t_2)$ as a proxy for the true element-wise $\widetilde{\mathrm{FDP}}(t_1, t_2)$ results in an underestimation. However, recall that $\widehat{\mathrm{FDP}}(t_1, t_2)$ tends to be a conservative estimator, often overestimating $\mathrm{FDP}(t_1, t_2)$. This overestimation can partially mitigate the underestimation introduced by ignoring $\zeta(t_1, t_2)$.

Beyond this theoretical issue, a practical challenge is that our method may select extra inactive variables, which is a common issue in structured selection methods. To address this, we recommend a two-stage strategy: first, use our method to select a much smaller submatrix; then apply vector-based variable selection to further eliminate inactive features. This approach leverages the efficiency of our method (as shown in Figure 2 in Section 4), making large-scale inference (e.g., 10,000 hypotheses) computationally feasible.

4 Simulations

4.1 Simulation settings and competing methods

We simulate data from three settings: one linear model and two nonlinear models. The predictors are generated from a matrix normal distribution $\mathcal{MN}_{p,q}(\mathbf{0}, \mathbf{U}, \mathbf{V})$, with $p = q$ and $p_1 = q_1$. For all models, we examine two covariance structures: (i) autoregressive (AR) covariance, where $U_{ij} = V_{ij} = \rho^{|i-j|}$; and (ii) compound symmetry (CS) covariance, where $U_{ij} = V_{ij} = \rho^{\mathbb{1}(i \neq j)}$. The noise term ϵ is generated from the standard normal distribution. We study two sparsity patterns for the active variables: (i) a single submatrix, and (ii) four smaller submatrices located at the corners. Figure 1 illustrates these patterns.

Model 1. $Y = \langle \mathbf{B}, \mathbf{X} \rangle + \epsilon$, where $\langle \mathbf{B}, \mathbf{X} \rangle = \mathrm{tr}(\mathbf{X}^T \mathbf{X})$, $\mathbf{B}_{1:p_1, 1:p_1} = 1, B_{jk} = 0$ otherwise.

Model 2. $Y = 3 \sum_{j=1}^{p_1} \sin(X_{jj}) + \sum_{j \neq k}^{p_1} X_{jk} + \epsilon$, combining a sine component on the diagonal and a linear component on the off-diagonals.

Model 3. $Y = \sum_{(j,k) \in \mathcal{G}_1} \exp(0.5 X_{jk}) + \sum_{(j,k) \in \mathcal{G}_2} \exp(0.5 X_{jk}) + \mathrm{sign}(\sum_{(j,k) \in \mathcal{G}_3} X_{jk}) + \mathrm{sign}(\sum_{(j,k) \in \mathcal{G}_4} X_{jk}) + \epsilon$. Here, $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$ and $\mathcal{G}_4$ are index sets of $\lfloor p_1/2 \rfloor \times \lfloor q_1/2 \rfloor$ active submatrices at four corners of $\mathbf{X}$, where the explicit form is in Section F of the Supplement.

We also consider an inverse regression model with a binary response in Section F of the

Supplement.

Figure 1: Active-variable patterns: four corner submatrices (Model 3, left) and a single top-left submatrix (Models 1–2, right).

We evaluate the performance of our procedure against existing methods for both row/column selection and element selection tasks. For row/column selection, we compare with MVSA+BH, which applies the BH procedure to p-values obtained from the asymptotic model-free test in (Li & Dong 2021). For element selection, which involves vectorizing the matrix predictor, we consider a broader set of competitors: dimension reduction coordinate (DRC) test (Cook 2004), Hilbert-Schmidt independence criterion (HSIC) test (Gretton et al. 2007), both followed by BH for FDR control; MTA; and the model-X Knockoff with statistics from SIR (Knockoff-SIR) and Lasso (Knockoff-Lasso). DRC and HSIC are followed by BH for FDR control. Our method and MVSA+BH are adapted for element-wise comparison by selecting an element (j, k) if both j-th row and k-th column are active, as described in Section 3.2. When required, covariance matrices for our and competing methods are estimated using the iterative maximum likelihood algorithm of Dutilleul (1999).

For all simulations, the target FDR level is 0.2 if not stated explicitly, and results are averaged over 100 replications. Empirical FDR, power, and computation time are reported. In the next subsection, we present: (i) row selection results for $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$ under different FDR levels in Table 1; (ii) element selection results for $(n, p, p_1, q, q_1) =$

$(1000, 20, 6, 20, 6)$ in Table 2, noting that many vector-based competitors require $n > pq$; (iii) computation time comparison in Figure 2, showing our method is at least 110 faster than MVSA+BH; and (iv) scalability at $p = q = 100$ and robustness to non-normal predictors in Table 3.

4.2 Results

Table 1 summarizes row-wise FDR and power for $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$ under FDR levels (0.20, 0.15, 0.10, 0.05). The results under the CS covariance structure are summarized in Table S.3. Since $p = q$ and $p_1 = q_1$, the column-wise results are similar (see Table S.1 in the Supplement). Overall, our method controls FDR well at levels 0.20 and 0.15 with high power across AR and CS structures for all three models. At the more stringent levels 0.10 and 0.05, we observe mild FDR inflation in several settings. However, the deviations are modest and the procedure still has large power. For comparison, MVSA+BH achieves good FDR control across these levels. But, it is substantially more conservative when ρ is large, leading to a significant loss of power. Especially for Model 3, it fails to select any rows.

Table 2 presents element-wise selection results for $(n, p, p_1, q, q_1) = (1000, 20, 6, 20, 6)$. Our method controls FDR near the target with high power across AR and CS. The FDR is approximately 14-20% and the power is around 95% for most settings. DRC+BH attains high power but inflated FDR, which is often greater than 50% and up to 90%. HSIC+BH also exhibits substantial FDR inflation under CS, which is often greater than 80%. MTA and Knockoff-SIR generally fail to control the FDR. Knockoff-Lasso controls FDR at 0.2, but its power collapses for Model 3 at higher ρ. MVSA+BH fails to select any variables under Model 3. Overall, our method demonstrates the best performance, achieving FDR control and competitive power.

Table 1: Row-wise FDR and power (tuples are (FDR, Power) in percent) for Models 1-–3 under AR when $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$.

Model	FDR Level	Method	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
Model 1	0.2	Proposed	(21.2, 100)	(17.4, 100)	(17.8, 95.3)
		MVSA+BH	(15.9, 100)	(13.8, 100)	(15.0, 72.2)
	0.15	Proposed	(16.2, 100)	(13.1, 100)	(13.6, 93.8)
		MVSA+BH	(11.9, 100)	(9.8, 99.9)	(11.2, 65.7)
	0.1	Proposed	(14.7, 100)	(11.8, 100)	(10.9, 90.9)
		MVSA+BH	(7.6, 100)	(5.8, 99.9)	(6.0, 59.2)
	0.05	Proposed	(8.9, 100)	(5.6, 100)	(6.8, 89.7)
		MVSA+BH	(3.6, 100)	(2.8, 99.7)	(3.6, 48.7)
Model 2	0.2	Proposed	(21.1, 100)	(16.6, 100)	(17.2, 95.2)
		MVSA+BH	(15.4, 100)	(13.5, 99.9)	(15.8, 72.9)
	0.15	Proposed	(16.2, 100)	(12.1, 100)	(13.2, 93.6)
		MVSA+BH	(12.0, 100)	(9.5, 99.9)	(10.8, 65.9)
	0.1	Proposed	(15.0, 100)	(10.9, 100)	(10.4, 90.7)
		MVSA+BH	(8.1, 100)	(6.0, 99.7)	(6.9, 59.1)
	0.05	Proposed	(7.1, 100)	(5.5, 100)	(7.3, 89.6)
		MVSA+BH	(3.7, 100)	(2.3, 99.5)	(3.2, 49.1)
Model 3	0.2	Proposed	(20.4, 100)	(21.9, 100)	(19.6, 96.5)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
	0.15	Proposed	(13.2, 100)	(15.2, 99.9)	(13.0, 94.3)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
	0.1	Proposed	(12.6, 100)	(13.8, 99.8)	(10.3, 90.8)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
	0.05	Proposed	(6.0, 100)	(7.8, 99.8)	(6.7, 89.9)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)

Table 2: Element-wise FDR and power (tuples are (FDR, Power) in percent) for Models 1–3 with $(n, p, p_1, q, q_1) = (1000, 20, 6, 20, 6)$ under AR and CS covariance.

Model	Method	AR			CS		
		$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
Model 1	Proposed	(18.2, 100)	(18.9, 100)	(20.4, 94.6)	(13.6, 100)	(18.7, 100)	(19.1, 91.3)
	DRC+BH	(53.3, 100)	(52.9, 100)	(52.7, 99.9)	(52.9, 100)	(52.9, 100)	(54.2, 92.8)
	HSIC+BH	(21.8, 73.1)	(20.8, 95.4)	(27.3, 99.8)	(82.8, 99.9)	(91.0, 100)	(91.0, 100)
	MTA	(20.8, 100)	(22.7, 99.2)	(24.6, 43.7)	(22.6, 99.9)	(22.3, 73.7)	(38.0, 9.1)
	Knockoff-SIR	(27.4, 98.0)	(28.1, 98.3)	(30.8, 92.8)	(27.9, 96.8)	(31.3, 93.0)	(30.7, 67.1)
	Knockoff-Lasso	(18.8, 99.2)	(17.0, 97.9)	(11.7, 93.0)	(18.7, 99.5)	(21.8, 99.4)	(18.2, 99.2)
	MVSA+BH	(13.5, 100)	(10.5, 99.8)	(16.2, 81.7)	(11.5, 100)	(15.0, 94.7)	(15.1, 32.6)
Model 2	Proposed	(17.7, 100)	(20.0, 100)	(19.7, 94.6)	(15.5, 100)	(18.9, 100)	(18.5, 90.7)
	DRC+BH	(87.9, 95.4)	(89.4, 82.4)	(90.6, 71.7)	(89.2, 84.8)	(90.4, 73.2)	(90.9, 68.8)
	HSIC+BH	(21.6, 60.2)	(18.5, 90.9)	(19.1, 99.1)	(80.5, 99.3)	(90.9, 100)	(91.0, 100)
	MTA	(21.4, 99.0)	(23.1, 87.0)	(29.5, 26.9)	(22.1, 93.5)	(20.7, 49.8)	(36.0, 8.6)
	Knockoff-SIR	(27.4, 96.5)	(25.4, 96.9)	(28.3, 87.3)	(28.2, 94.6)	(28.9, 88.2)	(32.1, 57.0)
	Knockoff-Lasso	(18.1, 98.9)	(17.3, 97.5)	(10.3, 89.7)	(19.4, 99.3)	(19.8, 98.9)	(16.1, 97.2)
	MVSA+BH	(13.3, 100)	(10.0, 99.8)	(14.5, 80.1)	(13.6, 100)	(15.4, 94.4)	(15.4, 33.7)
Model 3	Proposed	(16.7, 100.0)	(20.0, 99.8)	(20.2, 94.8)	(13.5, 100)	(19.2, 99.2)	(16.4, 87.1)
	DRC+BH	(88.8, 89.4)	(89.3, 83.9)	(90.3, 75.1)	(89.2, 84.8)	(90.2, 74.7)	(90.9, 69.7)
	HSIC+BH	(25.0, 56.1)	(31.7, 58.8)	(57.8, 63.2)	(87.0, 100)	(91.0, 100)	(91.0, 100)
	MTA	(21.4, 51.6)	(27.3, 27.4)	(67.1, 4.6)	(23.0, 41.8)	(25.1, 17.2)	(62.6, 4.2)
	Knockoff-SIR	(29.1, 75.8)	(30.4, 62.0)	(31.5, 42.2)	(29.1, 70.2)	(31.9, 53.4)	(33.6, 31.2)
	Knockoff-Lasso	(17.0, 91.6)	(10.5, 78.1)	(2.7, 41.1)	(13.7, 89.0)	(11.6, 74.6)	(3.8, 46.3)
	MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)

Beyond improved FDR control, our method is computationally efficient. Figure 2 compares runtime while holding $n = 2000$, AR covariance, and $\rho = 0.4$ fixed and varying $pq \in \{400, 900, 1600, 2500, 3600, 4900\}(p = q)$. Our method delivers a persistent order-of-magnitude speedup, about 190-280 times faster than MVSA+BH across sizes. For element-wise testing, Table S.5 in the Supplement shows that our method is between 5 and 1500 times faster than the vectorized competitors, making large-scale testing practically feasible.

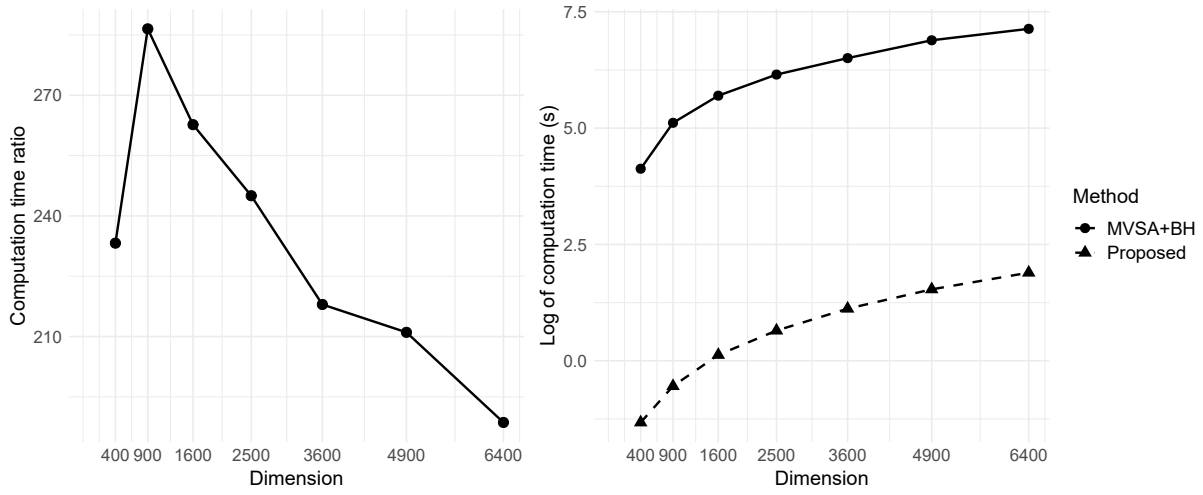

Figure 2: Computation time comparison between our method and MVSA+BH under Model 3 with $n = 2000, p_1 = q_1 = 10$ and AR covariance structure with $\rho = 0.4$. With $p = q = 50$ (the setting of Table 1), Table S.4 in the Supplement shows a 180-380 times speedup compared to MVSA+BH.

Finally, we assess both scalability and robustness of our method. With $(p, q) = (100, 100)$ and $(n, p_1, q_1) = (4000, 10, 10)$, we obtain 10,000 element-wise hypotheses. To assess robustness to deviations from the matrix-normal assumption, we generate predictors from several non-Gaussian distributions. Specifically, we set $\mathbf{X} = \mathbf{U}^{1/2}\mathbf{Z}\mathbf{V}^{1/2}$ where the entries of $\mathbf{Z}$ are i.i.d. samples from: (i) uniform on $[-2, 2]$, (ii) Laplace$(0, 1)$, and (iii) logistic$(0, 1)$. Matrices $\mathbf{U}$ and $\mathbf{V}$ follow the AR covariance structure. Table 3 reports results for Model 3.

Comparing $p = 50$ and $p = 100$, we find that the performance at $p = 100$ is similar to the lower-dimensional case. Competing methods are omitted at this scale since they are computationally infeasible or require $n > pq$. Our method maintains FDR close to the nominal 20% with high power across non-normal distributions. Mild FDR inflation appears in some cases (e.g., row- and element-wise test at $\rho = 0.4$), but deviations are modest.

Table 3: Scalability and robustness for Model 3 under AR covariance (tuples are (FDR, Power) in percent).

(n, p)	Distribution	Element-wise test			Row-wise test		
		$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
(2000, 50)	Uniform	(23.6, 100)	(26.8, 99.7)	(22.2, 83.2)	(19.3, 100)	(22.7, 100)	(18.9, 94.9)
(2000, 50)	Laplace	(23.4, 88.8)	(22.4, 93.8)	(21.4, 67.8)	(21.7, 96.3)	(18.8, 98.8)	(17.9, 87.3)
(2000, 50)	Logistic	(20.6, 88.1)	(27.9, 92.8)	(20.1, 62.1)	(19.6, 95.4)	(23.4, 97.7)	(18.2, 84.7)
(2000, 50)	Normal	(22.7, 100)	(25.7, 99.6)	(21.0, 82.5)	(20.4, 100)	(21.9, 100)	(19.6, 96.5)
(4000, 100)	Normal	(23.1, 100)	(26.4, 100)	(25.7, 97.7)	(17.9, 100)	(19.8, 100)	(21.7, 99.7)

5 Real data analysis

We analyze the publicly available EEG seizure data from Temple University Hospital (Shah et al. 2018, Obeid & Picone 2018). The dataset includes 19 electrodes $\times$ 50 time points EEG signals ($\mathbf{X}$) from 219 subjects (169 control, 50 seizure). The subject's group serves as the response Y for our model-free procedure. Since the single data split can yield unstable selections in real applications, we adopt the multiple data splitting approach (Dai et al. 2023*a*) to aggregate results across 100 data splits. For this analysis, the target FDR level is set to 0.2. First, our method is applied to identify active time points. Subject-level EEG signal curves (50 time points) are obtained by averaging across electrodes, followed by group-level averaging for seizure and control groups. We plot the two group average curves

in the left panel of Figure 3, where orange vertical lines represent selected time points. It is clear that two groups have significant differences at selected time points. Second, for electrode selection, C3, P3, and P8 are identified (right panel of Figure 3), which are known to be informative for seizure detection. C3 (left sensorimotor cortex) is linked to focal seizure origins (Iwama et al. 2023, Baumer et al. 2020, Inoue et al. 2021). P3 (left parietal) and P8 (right parietal) are relevant to parietal lobe epilepsy and serve as onset zones for various seizure types (Ristić et al. 2012). This electrode triplet was also identified as key for seizure prediction in (Ra et al. 2021), highlighting our method's capacity to extract clinically relevant electrodes.

To further evaluate selection robustness, we conduct an experiment with added noise variables. Specifically, after normalizing the data to unit standard deviation, we augment the original 19×50 data by adding 10 noise electrodes and 20 noise time points, with entries drawn independently from $N(0, 1)$, resulting in 29×70 matrices. Applying our method with multiple splits selects the same electrodes (C3, P3, P8) as before. The selected time points shift slightly to 18, 21, 39, 43, and 45 (compared to 3, 21, 32, 43 before). Figure 4 illustrates the mirror statistics for electrodes and time points from a representative split. As expected, statistics for non-selected variables, including the added noise dimensions, are approximately symmetric around zero and close to zero, clearly separated from the selected variables' statistics. In Section G of the Supplement, we transform the EEG signals to the frequency domain and reanalyze the data. The selected electrodes (F3, P7, O1, P4) are spatially close to (C3, P3, P8). The selected frequencies fall in the Theta (4-8 Hz) and Alpha (8-12 Hz) bands. That section also provides the full implementation details for the EEG case study.

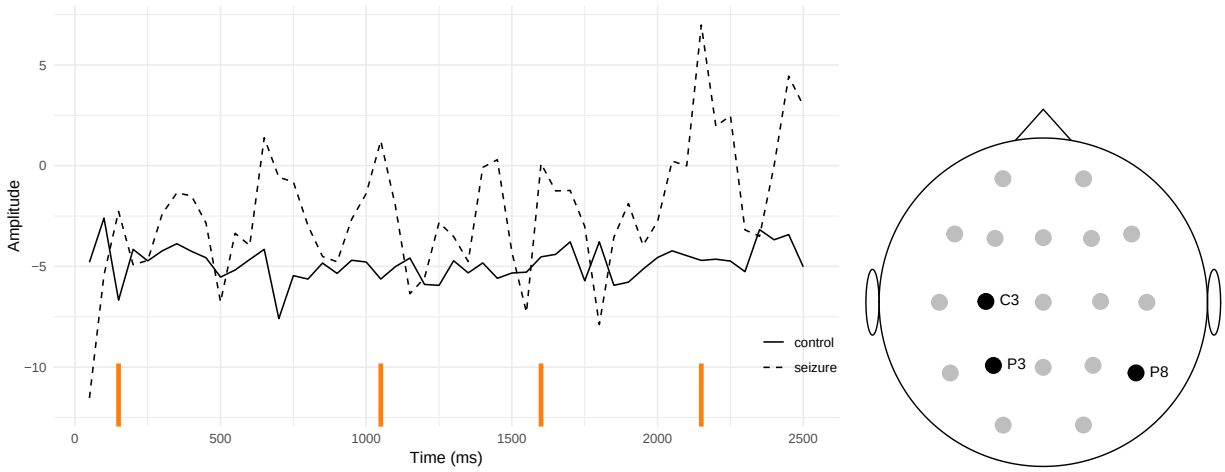

Figure 3: Left panel: group average EEG signal curves, with selected active time points denoted by orange vertical lines. Right panel: selected electrodes (filled dots).

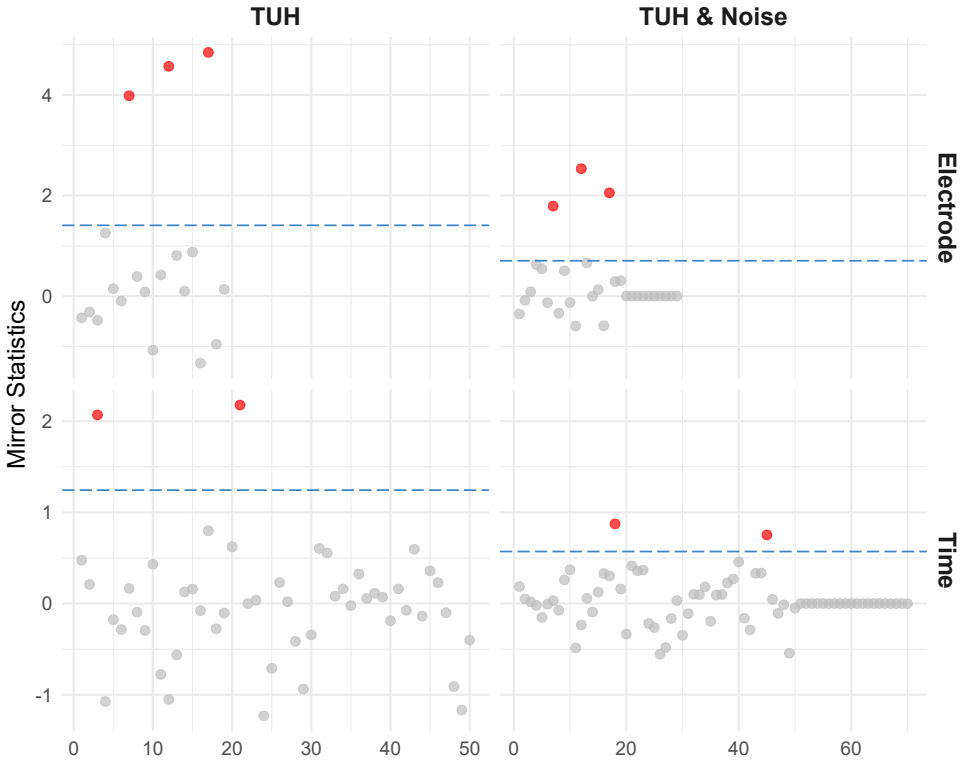

Figure 4: Scatterplot of mirror statistics for electrodes and time points, with the red points and blue line denoting selected predictors and the threshold corresponding to FDR level 0.2.

6 Discussions

This paper introduces a novel procedure for model-free variable selection with matrix-valued predictors that achieves FDR control. Our approach utilizes folding selection subspaces for robust inference without specifying the model and applies data splitting to construct mirror statistics. We establish theoretical guarantees, proving asymptotic FDR control for row and column selection via data-driven thresholds, even as dimensions p and q diverge. Furthermore, we show element-wise FDR control is achieved under a Kronecker product assumption on the variable importance pattern.

Our framework can be extended to tensor-valued predictors in a mode-wise fashion. For an M-way tensor predictor $\mathbf{X} \in \mathbb{R}^{p_1 \times \cdots \times p_M}$, let $\mathbf{X}_{(m)} \in \mathbb{R}^{p_m \times \prod_{\ell \neq m} p_\ell}$ denote the mode-m matricization. Similar to the row/column folding selection subspaces in the matrix case, one can define a mode-m folding selection subspace that captures the important indices along mode m. Under a tensor normal assumption for $\mathbf{X}$, applying our procedure to the matrix $\mathbf{X}_{(m)}$ achieves mode-wise FDR control by row-wise FDR control. Repeating this for $m = 1, \ldots, M$ yields a conceptually straightforward extension of our framework for identifying important slices along each tensor mode with FDR control.

There are potential avenues for future research. The data splitting strategy, key to the mirror statistics construction, reduces the effective sample size. Additionally, rigorous element-wise FDR control relies on the Kronecker structure assumption. Therefore, developing high-dimensional inference methods without data splitting and establishing direct element-wise FDR control under more general structural conditions, are important and challenging directions for future exploration.

7 Disclosure statement

The authors report there are no competing interests to declare.

References

Barber, R. F. & Candès, E. J. (2015), 'Controlling the false discovery rate via knockoffs', *The Annals of Statistics* **43**(5), 2055–2085.

Baumer, F. M., Pfeifer, K., Fogarty, A., Pena-Solorzano, D., Rolle, C. E., Wallace, J. L., Rotenberg, A. & Fisher, R. S. (2020), 'Cortical excitability, synaptic plasticity, and cognition in benign epilepsy with centrotemporal spikes: A pilot tms-emg-eeg study', *Journal of Clinical Neurophysiology* **37**(2), 170–180.

Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **57**(1), 289–300.

Benjamini, Y. & Yekutieli, D. (2001), 'The control of the false discovery rate in multiple testing under dependency', *Annals of statistics* pp. 1165–1188.

Candès, E., Fan, Y., Janson, L. & Lv, J. (2018), 'Panning for gold: 'model-x' knockoffs for high dimensional controlled variable selection', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **80**(3), 551–577.

Chen, X., Cook, R. D. & Zou, C. (2015), 'Diagnostic studies in sufficient dimension reduction', *Biometrika* **102**(3), 545–558.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2013), 'Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors', *The Annals*

of Statistics **41**(6), 2786 – 2819.

URL: *https://doi.org/10.1214/13-AOS1161*

Chernozhukov, V., Chetverikov, D. & Kato, K. (2017), 'Central limit theorems and bootstrap in high dimensions', *The Annals of Probability* **45**(4), 2309 – 2352.

Cook, R. D. (2004), 'Testing predictor contributions in sufficient dimension reduction', *The Annals of Statistics* **32**(3), 1062 – 1092.

Cook, R. D. & Ni, L. (2006), 'Using intraslice covariances for improved estimation of the central subspace in regression', *Biometrika* **93**(1), 65–74.

Dai, C., Lin, B., Xing, X. & Liu, J. S. (2023*a*), 'False discovery rate control via data splitting', *Journal of the American Statistical Association* **118**(544), 2503–2520.

Dai, C., Lin, B., Xing, X. & Liu, J. S. (2023*b*), 'A scale-free approach for false discovery rate control in generalized linear models', *Journal of the American Statistical Association* **118**(543), 1551–1565.

Dutilleul, P. (1999), 'The mle algorithm for the matrix normal distribution', *Journal of statistical computation and simulation* **64**(2), 105–123.

Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B. & Smola, A. (2007), A kernel statistical test of independence, *in* J. Platt, D. Koller, Y. Singer & S. Roweis, eds, 'Advances in Neural Information Processing Systems', Vol. 20, Curran Associates, Inc.

Inoue, T., Uda, T., Kuki, I., Yamamoto, N., Nagase, S., Nukui, M., Okazaki, S., Kawashima, T., Nakanishi, Y., Kunihiro, N., Matsuzaka, Y., Kawawaki, H. & Otsubo, H. (2021), 'Distinct dual cortico-cortical networks successfully identified between supplemental and primary motor areas during intracranial eeg for drug-resistant frontal lobe epilepsy', *Epilepsy & Behavior Reports* **15**, 100429.

Iwama, S., Morishige, M., Kodama, M., Takahashi, Y., Hirose, R. & Ushiba, J. (2023), 'High-density scalp electroencephalogram dataset during sensorimotor rhythm-based brain-computer interfacing', *Scientific Data* **10**(1), 385.

Javanmard, A. & Javadi, H. (2019), 'False discovery rate control via debiased lasso', *Electronic Journal of Statistics* **13**(1), 1212 – 1253.
URL: *https://doi.org/10.1214/19-EJS1554*

Landsman, Z. & Nešlehová, J. (2008), 'Stein's lemma for elliptical random vectors', *Journal of Multivariate Analysis* **99**(5), 912–927.

Li, B., Kim, M. K. & Altman, N. (2010), 'On dimension folding of matrix-or array-valued statistical objects', *Annals of Statistics* **38**(2), 1094–1121.

Li, B. & Wang, S. (2007), 'On directional regression for dimension reduction', *Journal of the American Statistical Association* **102**(479), 997–1008.

Li, H., Wei, F., Yan, K. & Zhang, Y. (2025), 'Assessing the exhaustiveness in sufficient dimension reduction', *Stat* **14**(2), e70069.

Li, K.-C. (1991), 'Sliced inverse regression for dimension reduction', *Journal of the American Statistical Association* **86**(414), 316–327.

Li, K.-C. (1992), 'On principal hessian directions for data visualization and dimension reduction: Another application of stein's lemma', *Journal of the American Statistical Association* **87**(420), 1025–1039.

Li, L., Zeng, J. & Zhang, X. (2023), 'Generalized liquid association analysis for multimodal data integration', *Journal of the American Statistical Association* **118**(543), 1984–1996.

Li, Z. & Dong, Y. (2021), 'Model-free variable selection with matrix-valued predictors', *Journal of Computational and Graphical Statistics* **30**(1), 171–181.

Liu, X., Niu, L. & Zhao, J. (2023), 'Statistical inference on the significance of rows and columns for matrix-valued data in an additive model', *TEST* **32**(3), 785–828.

Ma, R., Cai, T. T. & Li, H. (2021), 'Global and simultaneous hypothesis testing for high-dimensional logistic regression models', *Journal of the American Statistical Association* **116**(534), 984–998.

Mai, Q., Zhang, X., Pan, Y. & Deng, K. (2022), 'A doubly enhanced EM algorithm for model-based tensor clustering', *Journal of the American Statistical Association* **117**(540), 2120–2134.

Obeid, I. & Picone, J. (2018), 'The Temple university hospital EEG data corpus. augmentation of brain function: facts, fiction and controversy', *Brain-Machine Interfaces* **1**, 394–398.

Ra, J. S., Li, T. & Li, Y. (2021), 'A novel permutation entropy-based EEG channel selection for improving epileptic seizure prediction', *Sensors* **21**(23), 7972.

Ristić, A. J., Alexopoulos, A. V., So, N., Wong, C. & Najm, I. M. (2012), 'Parietal lobe epilepsy: the great imitator among focal epilepsies', *Epileptic Disorders* **14**(1), 22–31.

Shah, V., von Weltin, E., Lopez, S., McHugh, J. R., Veloso, L., Golmohammadi, M., Obeid, I. & Picone, J. (2018), 'The Temple university hospital seizure detection corpus', *Frontiers in Neuroinformatics* **12**.

Tang, C. Y., Fang, E. X. & Dong, Y. (2020), 'High-dimensional interactions detection with sparse principal hessian matrix', *Journal of Machine Learning Research* **21**(19), 1–25.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58**(1), 267–288.

Wang, N., Wang, W. & Zhang, X. (2024), 'Parsimonious tensor discriminant analysis', *Statistica Sinica* **34**, 157–180.

Wang, N., Zhang, X. & Li, B. (2022), 'Likelihood-based dimension folding on tensor data', *Statistica Sinica* **32**, 2405–2429.

Xing, X., Zhao, Z. & Liu, J. S. (2023), 'Controlling false discovery rate using gaussian mirrors', *Journal of the American Statistical Association* **118**(541), 222–241.

Yang, W., Shi, H., Guo, X. & Zou, C. (2024), Robust group and simultaneous inferences for high-dimensional single index model, *in* A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak & C. Zhang, eds, 'Advances in Neural Information Processing Systems', Vol. 37, Curran Associates, Inc., pp. 65217–65259.

Yin, X. & Hilafu, H. (2015), 'Sequential sufficient dimension reduction for large p, small n problems', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **77**(4), 879–892.

Zhang, X. & Li, L. (2017), 'Tensor envelope partial least-squares regression', *Technometrics* **59**(4), 426–436.

Zhao, J. & Leng, C. (2014), 'Structured lasso for regression with matrix covariates', *Statistica Sinica* **24**, 799–814.

Zhao, Z. & Xing, X. (2022), 'On the testing of multiple hypothesis in sliced inverse regression', *arXiv preprint arXiv:2210.05873* .

Zhong, W. & Suslick, K. S. (2015), 'Matrix discriminant analysis with application to colorimetric sensor array data', *Technometrics* **57**(4), 524–534.

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