Statistics and Its Interface Volume 17 (2024) 305–316

Model-based statistical depth for matrix data

Yue Mu, Guanyu Hu, and Wei Wu*

The field of matrix data learning has witnessed significant advancements in recent years, encompassing diverse datasets such as medical images, social networks, and personalized recommendation systems. These advancements have found widespread application in various domains, including medicine, biology, public health, engineering, finance, economics, sports analytics, and environmental sciences. While extensive research has been conducted on estimation, inference, prediction, and computation for matrix data, the ranking problem has not received adequate attention. Statistical depth, a measure providing a center-outward rank for different data types, has been introduced in the past few decades. However, its exploration has been limited due to the complexity of the second and higher order-statistics. In this paper, we propose an approach to rank matrix data by employing a model-based depth framework. Our methodology involves estimating the eigen-decomposition of a 4th-order covariance tensor. To enable this process using conventional matrix operations, we specify the tensor product operator between matrices and 4th-order tensors. Furthermore, we introduce a Kronecker product form on the covariance to enhance the robustness and efficiency of the estimation process, effectively reducing the number of parameters in the model. Based on this new framework, we develop an efficient algorithm to estimate the model-based statistical depth. To validate the effectiveness of our proposed method, we conduct simulations and apply it to two real-world applications: field goal attempts of NBA players and global temperature anomalies.

Keywords and phrases: Matrix data, Data depth, Covariance tensor, Eigen-decomposition.

1. INTRODUCTION

Matrix data has garnered significant interest in various applications, such as brain imaging studies [15], environmental science [11] and sports analytics [36]. The rapid development in this field has led to numerous challenges in estimation, inference, prediction, and computation when dealing with matrix data. One prominent example of matrix data can be found in environmental studies, where measurements of interest, such as temperature, humidity, or air quality, are recorded across a range of spatial locations. Additionally, heat maps illustrating the locations of field goal attempts on a basketball court or indicators of player effectiveness in different areas of a football pitch can also be represented as matrices, capturing the intensity of occurrences of the target events within a spatial context. While there is rich literature [4, 15, 18, 25, 35, 40] that extensively discusses regression models for matrix response data using parametric or nonparametric approaches, relatively less attention has been given to the ranking of such data.

Statistical depth is a concept used in data analysis to assess the centrality or outlyingness of a data point within a dataset or with respect to a given probability distribution. It quantifies how deep a point is relative to the other points in the dataset. Essentially, it provides a way to rank data points based on their relative centrality. In the context of matrix data, which involves complex structures with multiple dimensions, using statistical depth allows us to evaluate the significance or centrality of a particular matrix within a set of matrices. This can be crucial in applications like image processing, where we want to identify the most representative or characteristic images in a collection. Over the past few decades, numerous methods of data depth have been proposed for multivariate data, including the halfspace depth [34], simplicial depth [20], projection depth [41], Mahalanobis depth [21], and majority depth [31]. Based on the definitions of depth for data in Euclidean space, researchers have also explored depths in other spaces. or instance, the concept of halfspace depth has been extended to arbitrary metric spaces [7]. Functional data analysis is widely applied in fields like medicine, biology, and engineering, where the notion of statistical depth has been extended to functional data for tasks such as center-outward ranking, outlier detection, and classification. Related studies include band depth [22], half-region depth [23] and extremal depth [28]. However, these methods primarily focus on univariate functional data, which are curves in $\mathbb{R}^2$ space. While additional depth methods have been developed for curves in higher-dimensional spaces R^p, where $p \geq 3$, such as multivariate band depth [16], simplicial band depth [24] and multivariate functional halfspace depth [6]), the analysis of depth for matrix data or image data, which represent surfaces in R^3, remains insufficiently explored.

In this article, we consider the binary function space $\mathscr{L}^2\left([0,1] \times [0,1]\right)$, any $f \in \mathscr{F}$ is written as $f(t)$, $t \in [0,1] \times [0,1]$ or $f(t_1, t_2)$, $t_1, t_2 \in [0,1]$. This space is commonly used in biomedical studies and environmental science. Recently, a model-based statistical depth approach

*Corresponding author.

was proposed based on functional principal component analysis and reproducing kernel Hilbert space norm [39]. The key step in this approach involves conducting eigen-decomposition using a covariance operator. While the eigen-decomposition procedure is straightforward for univariate curve data, it becomes more challenging for matrix data, where the covariance operator is represented by a 4th-order tensor. To the best of our knowledge, there are no well-established methods for eigen-decomposition of a 4th-order tensor to estimate pairs of eigenvalues and eigen-matrices. The well-known higher-order singular value decomposition (HOSVD) for tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ [8, 38], computes the singular value decomposition (SVD) for each mode-k unfolding, resulting in a Tucker decomposition form. In this form, a core tensor is multiplied by a matrix for every mode, allowing for a compact representation of the tensor. However, it's important to note that the HOSVD does not generate pairs of eigenvalues and eigen-matrices. Another widely used tensor decomposition method is the Canonical Polyadic (CP) decomposition, also known as PARAFAC [2, 12]. This method expresses a tensor as the sum of rank-one tensors, where each rank-one tensor is formed by taking the outer product of vectors along each mode. The CP decomposition is a powerful tool for approximating high-order tensors and extracting interpretable patterns. However, similar to the HOSVD, the CP decomposition does not directly provide eigenvalues and eigenvectors for a 4th-order tensor. Other multivariate functional principal component analysis methods focus on the functions in the form of $(f_1, \ldots, f_p)$ [1, 5, 19], where each observation is a vector of p functions (p curves), and do not decompose 4th-order tensors. In this paper, we introduce a rigorous framework to address this issue by utilizing an isomorphic mapping to convert a 4th-order tensor into a matrix.

The second challenge lies in the robustness of the estimation. Matrix data often involves a large number of parameters in the 4th-order tensor, and a robust procedure is desirable to avoid overfitting problems. We propose using the Kronecker product covariance structure in our estimation procedure to reduce the number of parameters. Building upon this robust eigen-decomposition, we present an efficient algorithm for estimating the model-based statistical depth [39] for functional data in the space $\mathscr{L}^2\left([0,1]\times[0,1]\right)$. We demonstrate the effectiveness of the proposed method using simulations and real world data, comparing it with benchmark methods for bivariate data, such as modified band depth [22] and modified half-region (HR) depth [23].

The structure of this paper is as follows. In Section 2 we briefly review the definitions of tensor product operator and covariance operator [13]. We then specify the tensor product between matrices and tensors and use it to determine the discretized form of the covariance operator. In Section 3, we describe the estimation of such covariance operator using the Kronecker product [33]. We proceed to discuss the model-based depth [39] for bivariate functional data or matrix data. Section 4 illustrates simulation results and real-world data analysis. Finally, Section 5 provides related discussions and outlines avenues for future research.

2. THEORY ON TENSOR PRODUCT OPERATORS

In order to adopt the model-based depth to rank 2-dimensional objects, we need to have a proper estimation of eigenvalues and eigenfunctions from a 4th-order covariance tensor. In this section, we will provide a rigorous framework to define the tensor operator and robustly estimate the eigen-decomposition.

2.1 Basic definitions on tensor product operator

At first, we review the general notion of tensor product operators in the Hilbert space adopted from [13] and then illustrate its simple discrete form in Euclidean space $\mathbb{R}^p$

Definition 2.1. Let x_1, x_2 be elements of Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively. The tensor product operator $(x_1 \otimes_1 x_2) : \mathbb{H}_1 \mapsto \mathbb{H}_2$ is defined by

$$(x_1 \otimes_1 x_2)y = \langle x_1, y\rangle_1 x_2 \tag{1}$$

for $y \in \mathbb{H}_1$, where $\langle\cdot,\cdot\rangle_1$ is the inner-product operation in $\mathbb{H}_1$. If $\mathbb{H}_1 = \mathbb{H}_2$ we use $\otimes$ in lieu of $\otimes_1$.

Let $Z = \{Z(t) : t \in E\}$ be a stochastic process on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where E is a general compact metric space. We then adopt some definitions in [13]. The mean function of the process Z is $m(t) = \mathbb{E}[Z(t)]$, and the covariance function (or covariance kernel) is $K(s,t) = \mathrm{Cov}(Z(s), Z(t))$.

Suppose the mean function and the covariance function of Z are well defined and continuous, we define an integral operator in $\mathscr{L}^2(E, \mathscr{B}(E), \mu)$

$$(\mathscr{K} f)(s) = \int_E K(s,t) f(t) d\mu(t), \tag{2}$$

where μ is a finite measure. Then $\mathscr{K}$ is the covariance operator of Z. If we furthermore assume $Z = \{Z(t) : t \in E\}$ is jointly measurable with mean zero, then Z is viewed as a random element in some Hilbert space $\mathbb{H}_Z$. The covariance operator $\mathbb{E}(Z \otimes Z)$ is well defined and coincides with the operator $\mathscr{K}$ in Equation (2). Suppose $\lambda \in \mathbb{R}$ is a eigenvalue of $\mathscr{K}$ and $f \in \mathscr{L}^2(E, \mathscr{B}(E), \mu)$ is the corresponding eigenfunction; that is,

$$\mathscr{K} f = \lambda f,$$

which is equivalent to

$$\int_E K(s,t) f(t) d\mu(t) = \lambda f(s) \tag{3}$$

When the two Hilbert spaces are conventional finite-dimensional Euclidean spaces $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$, respectively, the

inner product is the classical operation between two vectors. Then $(x_1 \otimes_1 x_2)y = \langle x_1, y\rangle_1 x_2 = x_2(x_1^\top y) = (x_2 x_1^\top)y$, and hence, $(x_1 \otimes_1 x_2) = x_2 x_1^\top$ for $x_i \in \mathbb{R}^{P_i}$. If $x \in \mathbb{R}^p$ is a random variable, The covariance operator $\mathbb{E}(x \otimes x)$ is actually $\mathrm{Cov}(x)$. Consider $E = [0,1]$ and $Z = \{Z(t) : t \in E\}$ is a stochastic process with good properties as we mentioned before. Suppose z is the discretized form of $Z(t)$, then the sample covariance $\hat{S}$ can be used to estimate the covariance operator of Z. If μ is the counting measure in Equation (2), both $\mathscr{K}$ and $K(s,t)$ are discretized to $\hat{S}$, so the eigenvalue of $\hat{S}$ coincides λ in (3) and the correponding eigenvector of z is an estimation of the eigenfunction of $\mathscr{K}$.

2.2 Tensor product operator for matrix data

We now consider $E = [0,1] \times [0,1]$ and study tensor operator for matrices. In this case, the two Hilbert spaces, $\mathbb{H}_1 = \mathbb{R}^{p\times q}$ and $\mathbb{H}_2 = \mathbb{R}^{r\times s}$ for some finite positive integers $p, q, r, s \geq 2$, are equipped with Frobenius norms $\|\cdot\|_1$, $\|\cdot\|_2$ and Frobenius inner products $\langle\cdot,\cdot\rangle_1$, $\langle\cdot,\cdot\rangle_2$, respectively. For $A, B \in \mathbb{H}_k$, $\|A\|_k = \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{\mathrm{tr}(A^\top A)}$, $\langle A, B\rangle_k = \mathrm{tr}(A^\top B) = \sum_{i,j} A_{ij}B_{ij} = \mathrm{vec}(A)^\top \mathrm{vec}(B)$, $k = 1, 2$. The covariance for random matrix is a tensor of order 4. To understand tensor operation, we should define a multiplication between a 4th-order tensor and a matrix with proper dimensions.

Definition 2.2. Let $\underline{W} \in \mathbb{R}^{r\times s\times p\times q}$ and $A \in \mathbb{R}^{p\times q}$. The tensor-times-matrix multiplication $\cdot$ is defined by

$$(\underline{W} \cdot A)_{ij} = \langle \underline{W}_{ij\cdot\cdot}, A\rangle_F, \tag{4}$$

where $\langle\cdot,\cdot\rangle_F$ is Frobenius inner product. The product $\underline{W} \cdot A$ is in $\mathbb{R}^{r\times s}$.

This multiplication is extended to the tensor-times-tensor case. Let $\underline{A} \in \mathbb{R}^{p\times q\times r\times s}$ and $\underline{B} \in \mathbb{R}^{r\times s\times u\times v}$. Then

$$(\underline{A} \cdot \underline{B})_{ijkl} = \langle \underline{A}_{ij\cdot\cdot}, \underline{B}_{\cdot\cdot kl}\rangle_F. \tag{5}$$

Remark 2.1.
1. These two multiplication rules imitate the common matrix-times-vector and matrix-times-matrix multiplications, respectively.
2. For $\underline{I} \in \mathbb{R}^{p\times q\times p\times q}$, let $\underline{I}_{ijkl} = 1$ if $i = k$, $j = l$, and $I_{ijkl} = 0$ otherwise. Then $\underline{I}$ is the **identity tensor**. It is easy to show that $\underline{I} \cdot A = A$ for $A \in \mathbb{R}^{p\times q}$, $\underline{I} \cdot \underline{A} = \underline{A}$ for $\underline{A} \in \mathbb{R}^{p\times q\times r\times s}$ and $\underline{A} \cdot \underline{I} = \underline{A}$ for $\underline{A} \in \mathbb{R}^{r\times s\times p\times q}$.

$\underline{W} \cdot A$ can be regarded as an operator $\underline{W}$ acts on a matrix A. Suppose that $X \in \mathbb{R}^{p\times q}$, $Z \in \mathbb{R}^{r\times s}$ are two matrices. The tensor product operator $(X \otimes_1 Z)$ is a 4th-order tensor $\underline{W} \in \mathbb{R}^{r\times s\times p\times q}$ such that $\underline{W}_{ijkl} = Z_{ij}X_{kl}$. With the specific form of tensor product operator, the covariance tensor of a random matrix $X \in \mathbb{R}^{p\times q}$ with mean zero is $\underline{\Sigma} = \mathrm{Cov}(X) := \mathbb{E}(X \otimes X) \in \mathbb{R}^{p\times q\times p\times q}$.

The practical eigen-decomposition has only been defined on a matrix form (i.e., the classical singular value decomposition). To apply eigen-decomposition to a tensor, we need to develop methods to convert a 4th-order tensor into a 2nd-order, yet much larger matrix. Matricization is a commonly used technique to analyze tensor data, which is summarized in [17]. Now we consider the metric space $E = [0,1] \times [0,1]$. A stochastic process $Z = Z(t_1, t_2), (t_1, t_2) \in [0,1] \times [0,1]$ has a discrete matrix form $X \in \mathbb{R}^{p\times q}$. The covariance tensor of X is defined as follows.

Definition 2.3. Suppose $X \in \mathbb{R}^{p\times q}$ is a random matrix with mean zero. Then the covariance tensor of X is $\underline{\Sigma} = \mathrm{Cov}(X) := \mathbb{E}(X \otimes X) \in \mathbb{R}^{p\times q\times p\times q}$.

Since the covariance operator $\mathbb{E}(Z \otimes Z)$ coincides with the operator $\mathscr{K}$ in Equation (2), the covariance tensor of X, $\underline{\Sigma} = \mathbb{E}(X \otimes X)$ in Definition 2.3, is the discrete form of the operator $\mathscr{K}$. If we consider the covariance function $K(s,t) = \mathrm{Cov}(Z(s), Z(t))$, $\underline{\Sigma} = \mathrm{Cov}(X)$ is also the discrete form of $K(s,t)$. If μ is the counting measure in Equation (2), both $\mathscr{K}$ and $K(s,t)$ are discretized to $\underline{\Sigma}$ and f is discretized to a matrix A, we see both side of Equation (2) are $\underline{\Sigma} \cdot A$, where the multiplication $\cdot$ is defined in Definition 2.2. Hence, $\mathscr{K}$ and $K(s,t)$ have the same discrete representation.

If we further discretize $K(s,t)$ and $f(t)$ in Equation (3) and let μ be the counting measure, we have $\underline{\Sigma} \cdot A = \lambda A$ for $\lambda \in \mathbb{R}$. That is, we use the eigenvector (or eigenmatrix) of the covariance tensor $\underline{\Sigma}$ (with respect to X) to estimate the eigenfunction of the covariance operator $\mathscr{K}$ or the covariance function $K(s,t)$ (with respect to $Z(t)$). In addition, the eigenvalue of $\underline{\Sigma}$ is the same as the eigenvalue of $\mathscr{K}$ or $K(s,t)$. In the following content, we discuss how to estimate the covariance tensor $\underline{\Sigma}$ given bivariate discretized functional data $\{X_i(t), t \in [0,1] \times [0,1], i \in \{1, \ldots, n\}\}$ and to calculate the eigenvalue and the eigenmatrix of $\underline{\Sigma}$.

We have shown that the tensor product operator of two matrices is a 4th-order tensor. However, it is still not ready to apply eigen-decomposition to a tensor, because the practical eigen-decomposition has only been defined on a matrix form (i.e., the classical singular value decomposition). To achieve that goal, we need to develop methods to convert a 4th-order tensor into a 2nd-order, yet much larger matrix. The next definition forms a matrix row by row. We will show that it has many good properties to obtain eigen-decomposition in a 4th-order tensor.

Definition 2.4. Let $\underline{W} \in \mathbb{R}^{r\times s\times p\times q}$. $\mathrm{mat}(\underline{W})$ is a matricization of $\underline{W}$ in the following row-by-row form:

$$\mathrm{mat}(\underline{W}) = \begin{pmatrix} \mathrm{vec}(\underline{W}_{11\cdot\cdot})^\top \\ \mathrm{vec}(\underline{W}_{21\cdot\cdot})^\top \\ \vdots \\ \mathrm{vec}(\underline{W}_{r1\cdot\cdot})^\top \\ \mathrm{vec}(\underline{W}_{12\cdot\cdot})^\top \\ \mathrm{vec}(\underline{W}_{22\cdot\cdot})^\top \\ \vdots \\ \mathrm{vec}(\underline{W}_{r2\cdot\cdot})^\top \\ \vdots \\ \mathrm{vec}(\underline{W}_{1s\cdot\cdot})^\top \\ \vdots \\ \mathrm{vec}(\underline{W}_{rs\cdot\cdot})^\top \end{pmatrix}.$$

Based on Definition 2.4, we will show several important properties of the matricization $\mathrm{mat}(\cdot)$. At first, we point out the identity term:

Proposition 2.1. *Let $\underline{I}_{p\times q\times p\times q}$ be the identity tensor in $\mathbb{R}^{p\times q\times p\times q}$. Then $\mathrm{mat}(\underline{I}_{p\times q\times p\times q}) = I_{pq}$, i.e., $\mathrm{mat}(\underline{I}_{p\times q\times p\times q})$ is the identity matrix in $\mathbb{R}^{pq\times pq}$.*

Now we focus on vectorization operation on the product of a tensor and a matrix, which is crucial for the eigen-decomposition. Suppose $\underline{W} \in \mathbb{R}^{r\times s\times p\times q}$ and $A \in \mathbb{R}^{p\times q}$. Then it is easy to see

$$\mathrm{vec}(\underline{W} \cdot A) = \mathrm{mat}(\underline{W})\,\mathrm{vec}(A). \tag{6}$$

As we mentioned before, the eigenvalue and the eigenfunction of the covariance operator $\mathscr{K}$ of bivariate function $Z(t), t \in [0,1]\times[0,1]$ can be estimated by the eigenvalue and the eigenmatrix of the covariance tensor of the corresponding random matrix X, respectively. Based on (6), we can easily obtain eigenvalue and eigenvector in the matricization form of a 4th-order tensor. This is given in the following proposition:

Proposition 2.2. *Suppose $\underline{W} \in \mathbb{R}^{p\times q\times p\times q}$ and $A \in \mathbb{R}^{p\times q}$. Then for $\lambda \in \mathbb{R}$, $\underline{W}\cdot A = \lambda A$ if and only if* $\mathrm{mat}(\underline{W})\,\mathrm{vec}(A) = \lambda\,\mathrm{vec}(A)$.

We can easily prove that $\mathrm{mat}(\cdot)$ is a linear isomorphism from $\mathbb{R}^{r\times s\times p\times q}$ to $\mathbb{R}^{rs\times pq}$, where r, s, p, q are any positive integers. We focus back on the covariance of a random matrix. The following proposition gives the matricization of a covariance tensor.

We then examine the algebraic properties of the matricization operator mat. mat is a linear isomorphism between 4th-order tensors and 2-d order matrices. It is easy to verify that $\mathrm{mat}(W_1 + W_2) = \mathrm{mat}(W_1) + \mathrm{mat}(W_2)$ and $\mathrm{mat}(cW_1) = c\,\mathrm{mat}(W_1)$ for any $W_1, W_2 \in \mathbb{R}^{r\times s\times p\times q}$ and $c \in \mathbb{R}$, so mat is a linear transformation between two vector spaces $R^{r\times s\times p\times q}$ and $R^{rs\times pq}$. Next, we focus on group isomorphism between $G(p,q) = \{\underline{A} \in \mathbb{R}^{p\times q\times p\times q} | \,\mathrm{mat}(\underline{A})$ is nonsingular$\}$ and the general linear group $GL(pq) = \{A \in \mathbb{R}^{pq\times pq} | A$ is nonsingular$\}$. The following proposition show that *mat* maps products in $G(p,q)$ to products in $GL(pq)$.

Proposition 2.3. $\mathrm{mat}(\underline{A}\cdot\underline{B}) = \mathrm{mat}(\underline{A})\,\mathrm{mat}(\underline{B})$ *for any $\underline{A} \in \mathbb{R}^{p\times q\times r\times s}$, $\underline{B} \in \mathbb{R}^{r\times s\times u\times v}$, where p, q r, s, u, v are any positive integers.*

Therefore, $(G(p,q),\cdot)$ is a group with identity element $\underline{I}_{p\times q\times p\times q}$, and the inverse element of $\underline{A}$, denoted as $\underline{A}^{-1}$, satisfies $\mathrm{mat}(\underline{A}^{-1}) = \mathrm{mat}(\underline{A})^{-1}$. As a result, we obtain that $(G(p,q),\cdot)$ isomorphic to the general linear group with degree pq.

Based on the definition of the covariance tensor of random matrix (Definition 2.3) as well as the developed theory on tensor matricization (Proposition 2.2), we are ready to show the calculation procedure for the eigenvalues and the eigenvectors of a covariance tensor. The following proposition gives the matricization of a covariance tensor.

Proposition 2.4. *Suppose $X \in \mathbb{R}^{p\times q}$ is a random matrix, then* $\mathrm{Cov}(\mathrm{vec}(X)) = \mathrm{mat}(\mathrm{Cov}(X))$.

Given a $p\times q$ random matrix X with covariance tensor $\underline{\Sigma}$. Then by Proposition 2.4, $\mathrm{Cov}(\mathrm{vec}(X)) = \mathrm{mat}(\underline{\Sigma})$; that is, we estimate the covariance matrix of $\mathrm{vec}(X)$ instead of $\underline{\Sigma}$. Let $\hat{S}$ be the estimated $\mathrm{Cov}(\mathrm{vec}(X))$, since mat is a linear isomorphism between 4th-order tensors and matrices, there is an unique $\hat{\underline{\Sigma}}$ such that $\hat{\underline{\Sigma}} = \mathrm{mat}_2^{-1}(\hat{S})$. According to Proposition 2.2, we can conduct eigen-decomposition to $\hat{S}$ directly. If $\hat{S}a = \lambda a$ with $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}^{pq}$, then λ is also the eigenvalue of $\hat{\underline{\Sigma}}$. Let A be the inverse of the vectorization of a; that is, $A = \mathrm{vec}^{-1}(a) \in \mathbb{R}^{p\times q}$. Then A is the eigenmatrix of $\hat{\underline{\Sigma}}$, i.e., $\hat{\underline{\Sigma}}\cdot A = \lambda A$.

In summary, this section establishes the foundation for the estimation and the matricization of covariance tensor, it also shows the eigen-decomposition approach for a 4th-order tensor. Actually, Definition 2.2 (the tensor-times-matrix multiplication) can be extended to higher order case. For example, we can define the multiplication between a 6th-order tensor and a 3rd-order tensor in the similar way. The covariance can also be specified for higher order random tensors. Using the above theoretical construction, we are ready to adopt the model-based method in [39] to estimate the depth value of matrix data. In the next section, we will propose a robust and efficient procedure to conduct the eigen-decomposition.

3. COVARIANCE-BASED DEPTH ESTIMATION

In this section, we address the calculation issue for covariance estimation and then provide an algorithm to compute the model-based depth for matrix data.

3.1 Estimation of Kronecker product covariance

Suppose we have $X_1, X_2, \ldots, X_n \in \mathbb{R}^{p\times q}$ after discretization. Our goal is to estimate the covariance matrix. However, as the size of the covariance matrix is $pq \times pq$, it usually contains too many parameters to estimate. Hence, we propose to adopt the Kronecker product to simplify the covariance, so that the estimated covariance has the form $\hat{S} = P \otimes_{\mathrm{Kron}} Q$ where $P \in \mathbb{R}^p$ and $Q \in \mathbb{R}^q$. Recent studies have introduced the notion of matrix normal distribution and proposed likelihood-based algorithms to estimate Kronecker product covariance [9, 33, 32]. Moreover, the tensor normal distribution for a J-th order random tensor of dimensions $p_1 \times \cdots \times p_J$ and its maximum likelihood estimations are presented in [26].

Basically, a random matrix $X \in \mathbb{R}^{p\times q}$ follows the matrix normal distribution $\mathcal{MN}_{p\times q}(M,P,Q)$ if it has the probability density function

$$p(X|M,P,Q) = \frac{\exp(-\frac{1}{2}\mathrm{Tr}[Q^{-1}(X-M)^\top P^{-1}(X-M)])}{(2\pi)^{pq/2}|Q|^{p/2}|P|^{q/2}},$$

where p, q are positive integers, M is $p \times q$, P is $p \times p$ and Q is $q \times q$. M is expectation, P is among-row variance and Q is among-column variance, P and Q are positive definite matrices. An important property of matrix normal distribution is that $X \sim \mathcal{MN}_{p\times q}(M, P, Q)$ if and only if $\text{vec}(X)$ has the multivariate normal distribution $\mathcal{N}_{pq}(\text{vec}(M), Q \otimes_{\text{Kron}} P)$.

Let $X_1, \ldots, X_n$ be independent and identically distributed from $\mathcal{MN}_{p\times q}(M, P, Q)$. The maximum likelihood estimate (MLE) of M is simply the average $\hat{M} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. A flip-flop algorithm is proposed to calculate the MLEs $\hat{P}$ and $\hat{Q}$ [33]. Let $Y_i = X_i - \hat{M}, i \in \{1, \ldots, n\}$. At first give the initial guesses such that $\hat{P}^{(0)}, \hat{Q}^{(0)}$ are positive definite and $Q_{qq} = 1$, then update $\hat{P}$ and $\hat{Q}$ as follows

$$\begin{cases} \hat{P}^{(k+1)} = \dfrac{1}{qn}\displaystyle\sum_{i=1}^{n} Y_i (\hat{Q}^{(k)})^{-1} Y_i^\top, \\ \hat{R}^{(k+1)} = \dfrac{1}{pn}\displaystyle\sum_{i=1}^{n} Y_i^\top (\hat{P}^{(k)})^{-1} Y_i, \\ \hat{Q}^{(k+1)} = \dfrac{1}{\hat{R}^{(k+1)}_{qq}} \hat{R}^{(k+1)}. \end{cases} \tag{7}$$

In practice we may choose identity matrices to be the initial values $\hat{P}^{(0)}$ and $\hat{Q}^{(0)}$, and the estimated $\hat{P}$ and $\hat{Q}$ are always positive definite. Note that this flip-flop reduces the number of parameters in covariance estimation, but it only works well when the matrix size is not too large. If $n > \max\{p, q\}$, the maximum likelihood estimate (7) has a unique solution [33].

The estimated covariance matrix of $\{\text{vec}(X_i)\}_{i=1}^n$ is $\hat{S} = \hat{Q} \otimes_{\text{Kron}} \hat{P}$. Hence, we conduct SVD to Kronecker product covariance. Suppose $\hat{Q} = U_Q D_Q V_Q^\top$ and $\hat{P} = U_P D_P V_P^\top$, then $\hat{S} = \hat{Q} \otimes_{\text{Kron}} \hat{P} = (U_Q \otimes_{\text{Kron}} U_P)(D_Q \otimes_{\text{Kron}} D_P)(V_Q \otimes_{\text{Kron}} V_P)^\top$. Given matrix data $X_1, X_2, \ldots, X_n \in \mathbb{R}^{p\times q}$, every function f_i is discretely estimated by a $p \times q$ matrix. The algorithm for applying principal analysis is shown in Algorithm 1.

Algorithm 1 Principal Component Analysis for Matrix Data

1. Apply the flip-flop algorithm (Equation (7)) to estimate the matricized covariance tensor S for f_i's, find the Kronecker product structure $\hat{S} = \hat{Q} \otimes_{\text{Kron}} \hat{P}$, where $\hat{Q}$ is $q \times q$ and $\hat{P}$ is $p \times p$;
2. Conduct SVD to $\hat{Q}$ and $\hat{P}$ separately, $\hat{Q} = U_Q D_Q V_Q^\top$, $\hat{P} = U_P D_P V_P^\top$;
3. Calculate $D_0 = D_P \otimes_{\text{Kron}} D_Q$ and $U_0 = U_P \otimes_{\text{Kron}} U_Q$. Order the diagonal entries of D_0, denotes as d_{0i}, $i \in \{1, \ldots, n\}$, from high to low, and arrange the columns of U_0 in the same order of d_{0i}. Let $D = \text{reorder}(D_0)$, $U = \text{reorder}(U_0)$;
4. Reshape every columns of U, denoted as U_i, $i \in \{1, \ldots, n\}$, into $p \times q$ matrices, $\phi_i = \text{vec}^{-1}(U_i)$, $i \in \{1, \ldots, n\}$;
5. The covariance tensor is estimated by $\hat{K}(s_1, s_1, t_1, t_2) = \sum_{w=1}^{W} d_w \phi_w(s_1, s_2)\phi_w(t_1, t_2)$ for some $W < n$.

The Kronecker product covariance estimation greatly reduces the number of parameters; we estimate a $p \times p$ matrix and a $q \times q$ matrix instead of estimating a $pq \times pq$ matrix. On the other hand, applying SVD to $\hat{Q}$ and $\hat{P}$ separately is more efficient than applying SVD to $\hat{Q} \otimes_{\text{Kron}} \hat{P}$ directly, especially when p and q are large.

3.2 Model-based depth algorithm

In this subsection, we propose to use principal component analysis to calculate statistical depths for matrix data, where the covariance-based algorithms are developed to estimate norm-based statistical depth and inner-product-based statistical depths for functional data [39]. Given n independent samples from finite-dimensional zero-mean Gaussian process $Z_1, \ldots, Z_n$ on $t \in [0, 1]$, and an observed function Z_{obs}. We first compute sample mean function $\bar{Z}(t) = \frac{1}{n}\sum_{i=1}^{n} Z_i(t)$, empirical covariance function $\tilde{K}(s, t) = \frac{1}{n}\sum_{i=1}^{n} \big[Z_i(s) - \bar{Z}(s)\big]\big[Z_i(t) - \bar{Z}(t)\big]$, and apply eigen-decomposition to covariance: $\tilde{K}(s, t) = \sum_{w=1}^{n} \hat{\lambda}_{w,n}\hat{\phi}_{w,n}(s)\hat{\phi}_{p,n}(t)$. Then we choose a number $W < n$ and estimate $\hat{K}(s, t) = \sum_{w=1}^{W} \hat{\lambda}_{w,n}\hat{\phi}_{w,n}(s)\hat{\phi}_{w,n}(t)$. Compute $\hat{Z}_w = \int_0^1 Z_{obs}(t)\phi_{w,n}(t)dt$ for $w = 1, \ldots, W$, and the induced reproducing kernel Hilbert space norm $\|Z_{obs}\|^2_{H_{\hat{K}}} = \sum_{z=1}^{W} \frac{\hat{Z}_w^2}{\hat{\lambda}_{w,n}}$. The inner-product-based depth is calculated as $D_{ip}(Z_{obs}) = 1 - \Phi(\|Z_{obs}\|_{H_{\hat{K}}})$, where $\Phi(x)$ is the cumulative distribution function standard normal distribution. The reproducing kernel Hilbert space norm induced depth is calculated as $D = 1 - F(\|Z_{obs}\|^2_{H_{\hat{K}}})$ where $F(x)$ denotes the cumulative distribution function of $\chi^2(P)$.

For matrix data, we propose to use Algorithm 1 to conduct eigen-decomposition, which implies that we use Kronecker product covariance estimation by Equation (7) instead of sample covariance matrix. Different from the algorithms in [39], discretization is a necessary step in our algorithms for the bivariate estimation. This is because the Kronecker product is only defined on the finite space and the flip-flop algorithm can only be applied to matrices but not functional data. Hence, we should first discretize functional data into matrices in our algorithms. In practical use, this is not a problem as observed data are already in discrete forms.

Suppose we have n independent samples from finite-dimensional zero-mean Gaussian process $Z_1, \ldots, Z_n$ on $t \in [0, 1] \times [0, 1]$, and an observed function Z_{obs}. Suppose every function is estimated by a $p \times q$ matrix, the algorithms for estimating model-based statistical depths are shown as follows.

Remark 3.1. For simplification, Algorithm 2 is only for observations from Gaussian processes. In fact, the depth estimation procedure for any second-order stochastic process can be extended to bivariate data by using the flip-flop algorithm. We only need to replace Step 4 or 4′ by a bootstrapping procedure. The computational details on the depth estimation part are found in [39].

Algorithm 2 Model-Based Statistical Depth Estimation Algorithm.

Input: observations $\{Z_1, \ldots, Z_n\}$, any observation Z_{obs}, a threshold $\epsilon > 0$.

1. Use Algorithm 1 to conduct eigen-decomposition, the covariance function is estimated by

$$\hat{K}(s_1, s_1, t_1, t_2) = \sum_{w=1}^{W} d_w \phi_w(s_1, s_2)\phi_w(t_1, t_2);$$

2. For $w = 1, \ldots, W$, Compute

$$\hat{Z}_w = \int_0^1 \int_0^1 Z_{obs}(t_1, t_2)\phi_w(t_1, t_2)dt_1 dt_2;$$

3. Compute the induced reproducing kernel Hilbert space norm $\|Z_{obs}\|^2_{H_{\hat{K}}} = \sum_{w=1}^{W} \frac{\hat{Z}_w^2}{d_w}$;

4. (Inner-Product-Based Depth) Compute the depth as $D_{ip}(Z_{obs}) = 1 - \Phi(\|Z_{obs}\|_{H_{\hat{K}}})$,where $\Phi(x)$ denotes the cumulative distribution function of a standard normal random variable.

4′. (Reproducing Kernel Hilbert Space Norm-Based Depth) Compute the depth as $D = 1 - F(\|Z_{obs}\|^2_{H_{\hat{K}}})$, where $F(x)$ denotes the cumulative distribution function of $\chi^2(W)$. Hence, the proposed model-based depth can be extended in higher order spaces.

Remark 3.2. As we mentioned at the end of Section 2, the covariance for higher order random tensor is also specified. The maximum likelihood estimations for covariance under tensor normal distribution are given in [26]. Hence, the proposed depth method can be adapted for higher order data.

Step 4 and Step 4′ in Algorithm 2 present two different model-based depth methods. The norm-based depth is a generalization over various distance-based forms. In contrast, the inner-product depth is motivated with the classical halfspace depth by Tukey [34]. The key differences between these two methods are 1) norm-based depth depends on a center, but inner-product-based depth is independent of it, and 2) the inner-product-based depth is often meaningful for a finite dimensional space only [39]. In practice, they usually yield similar results. However, Step 4 outperforms in our numerical studies, and therefore only the results of reproducing kernel Hilbert space norm-based depth are shown in Section 4.

The value of W should be manually selected. One way is to find a significant gap among singular values or a threshold such as 95% for cumulative singular values in the classical functional PCA framework. Algorithm 2 assumes zero-mean Gaussian process. In practice, we need to centralize data (i.e., subtract the sample mean) if necessary. Then the depth values only depends on covariance.

4. NUMERICAL STUDIES

4.1 Simulations

We will use three simulated 2-dimensional image datasets to illustrate the proposed depth method.

1. The first dataset setting is a sequence of discretized images of the noise-contained bivariate Gaussian probability density functions on $[-3, 3] \times [-3, 3]$ with the same covariance but different means. Let $N(x, y \mid \mu, \Sigma)$ denote the bivariate Gaussian density function with mean μ and covariance Σ at $(x, y)^\top$. Each observation is a 31×31 image $I^n = \{I^n_{ij}\}_{i,j=1}^{31}$, $n = 1, \ldots, 50$, given as:

$$I^n_{ij} = 1000N(x_i, y_j \mid \mu_n, \Sigma) + \epsilon_{i,j,n},$$

where $x_i = -3 + 0.2 * (i-1), y_j = -3 + 0.2 * (j-1), \mu_n = [-0.5 + 0.02(n-1), 0]^\top, \Sigma = \left[\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right]$, and $\epsilon_{i,j,n}$'s are independent and identically distributed from $N(0, 1)$, $i, j = 1, \ldots, 100, n = 1, \ldots, 50$. Three of the observations (1st, 25th, and 50th) are shown in the left column of Figure 1(a).

2. The second dataset setting is very similar to the first one except that in this case the density functions have the same mean but different covariances. Each observation is given as:

$$I^n_{ij} = 1000N(x_i, y_j \mid \mu, \Sigma_n) + \epsilon_{i,j,n},$$

where $x_i = -3+0.06*(i-1), y_j = -3+0.06*(j-1), \mu = [0, 0]^\top$,

$$\Sigma_i = \begin{bmatrix} 1 & 0.25 - 0.01(i-1) \\ 0.25 - 0.01(i-1) & 1 \end{bmatrix},$$

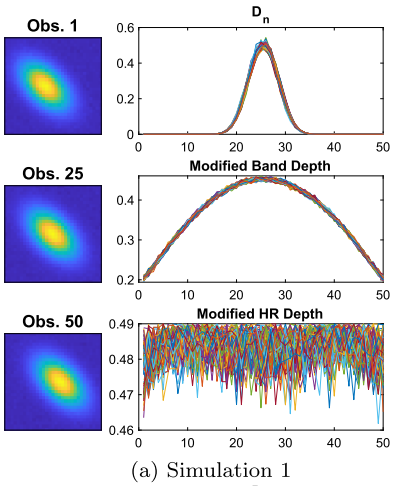
and $\epsilon_{i,j,n}$'s are independent and identically distributed from $N(0, 1)$, $i, j = 1, \ldots, 100, n = 1, \ldots, 50$. Three of the observations (1st, 25th, and 50th) are shown in the left column of Figure 1(b).

3. The third dataset setting is a mixture of two bivariate normal density functions. Each observation is given as:

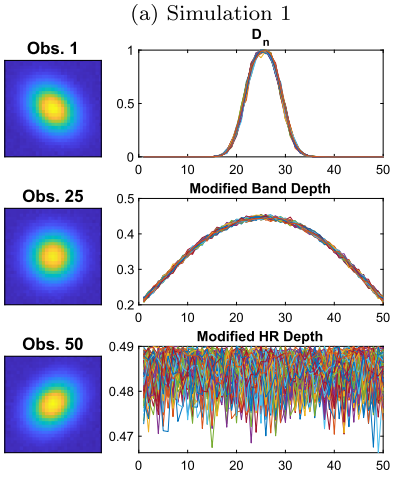
$$300N(x_i, y_j \mid \mu_1, \Sigma_{1n}) + 700N(x, y \mid \mu_2, \Sigma_{2n})) + \epsilon_{i,j,n},$$

where $x_i = -3 + 0.06 * (i-1)$, $y_j = -3 + 0.06 * (j-1)$, $\mu_1 = [-1.5, 0]^\top$, $\mu_2 = [1.5, 0]^\top, \Sigma_{1n} = (1 + 0.015(n-1))\left[\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right]$, $\Sigma_{2n} = (1 + 0.015(i-1))\left[\begin{smallmatrix} 2 & -1 \\ -1 & 2 \end{smallmatrix}\right]$, and $\epsilon_{i,j,n}$'s are independent and identically distributed from $N(0, 1)$, $i, j = 1, \ldots, 100, n = 1, \ldots, 50$. Three of the observations (1st, 25th, and 50th) are shown in the left column of Figure 1(c).

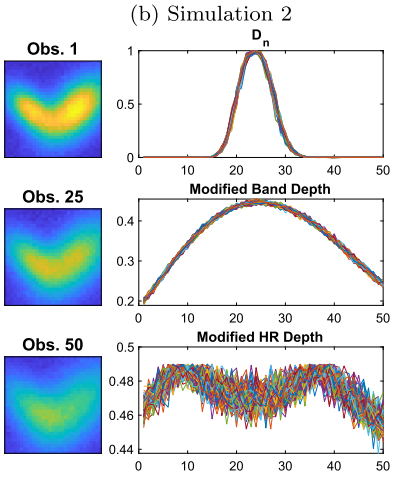
For each dataset setting, we repeat simulation 100 times. The dataset in every replica is the same except the noise $\epsilon_{i,j,n}$. In Simulation 1, the high density part moves along the x-axis from left to right. This is well captured by the proposed method with norm-based depth in Step 4′ of Algorithm 2. The result is shown in the upper-right panel in



(a) Simulation 1



(b) Simulation 2



(c) Simulation 3

Figure 1. Simulation result. (a), Three (1st, 25th, 50th) out of 50 observations in Simulation 1 are shown in the left column. The three panels on the right column show the depth values in the given 50 observations using the proposed norm-based method, modified band depth, and modified half-region depth, respectively. (b) and (c), Same as (a) except for Simulations 2 and 3, respectively.

Figure 1(a), where the depth value increases from left to the middle part, and then decreases from the middle to the right. The peak is at around 25th observations. Similar result can be obtained if instead we use the inner-product-based depth in Step 5 of Algorithm 2 and is omitted here. In comparison, we also apply the modified band depth [22] and modified half-region (HR) depth [23] to this dataset and calculate their depth values. The result is shown on the bottom two panels in the right column of Figure 1(a). Similar trend is observed in the modified band depth method, but the modified half-region depth totally misses the ranking in the given data set.

Consistent result is obtained in Simulation 2. In this case, the main axis in the elliptical contours rotates from passing through the 2nd–4th quadrants (centered at $(0, 0)$) to the y-axis, and then rotates to passing through the 1st–3rd quadrants (see Figure 1(b)). The proposed depth successfully captures this variation trend and generates a bell-shaped curve over the 50 observations (see the upper right panel). The modified band depth provides similar trend, whereas the modified half-region depth again misses the ranking (see the bottom two panels in the right column). In Simulation 3, the means in both Gaussian densities remain invariant, but their covariances get larger from observation 1 to 50. Consistent result is also obtained and the detail is shown in Figure 1(c).

4.2 NBA data

In this subsection, we will apply the proposed depth method to a real world dataset. Our data consists of field goal attempts' locations from the offensive half court of games in the 2019–2020 National Basketball Association (NBA) regular season. The data are available from the link https://www.basketball-reference.com/. We focus on players that made more than 500 field goal attempts (FTA). We model a player's shooting location choices on the offensive half court. The standard half court size for NBA is 47 ft $\times$ 50 ft rectangle. However, we only focus on 40 ft $\times$ 50 ft rectangle since most of players do not have shots beyond 40 ft, i.e., near the half court line, which is similar with Franks et al. [10]. The spatial domain for the basketball court in this paper is denoted as $D \in [0, 40] \times [0, 50]$. We partition the court to 2 ft $\times$ 2 ft blocks, which implies that there are in total $20 \times 25 = 500$ independent blocks in the basketball court. Each block is not overlapped with another. Our analysis includes two groups of players: one group consists of 79 forwards or centers, and the other group consists of 78 guards.

The figures represent log intensity maps which recover the shooting patterns of different players. For all the players of interest, we can model their shot charts through a log Gaussian cox process (LGCP) and estimate their associated intensity functions using integrated nested Laplace approximation [30]. The use of LGCP is very common in basketball shot chart analysis literature [27, 3, 14].

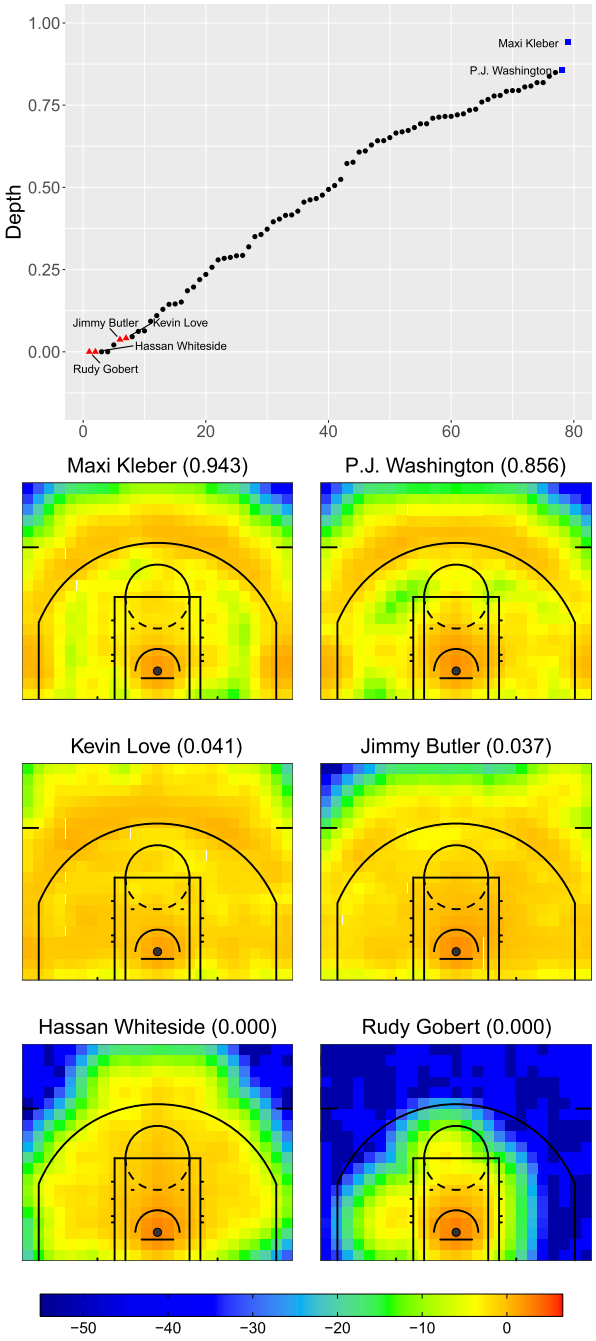


Figure 2. Depth Values for forwards and centers. Norm-based depth values of all 79 Forwards and Centers, denoted by black dots, are shown on the upper left panel, with two high valued players denoted in blue squares and four low valued players denoted by red triangles. The shooting heat maps of the two high valued ones are shown in the upper right two plots and the maps of the four low valued ones are shown in the four plots on the lower panel, where the value in parenthesis of each player's name indicates his depth value.

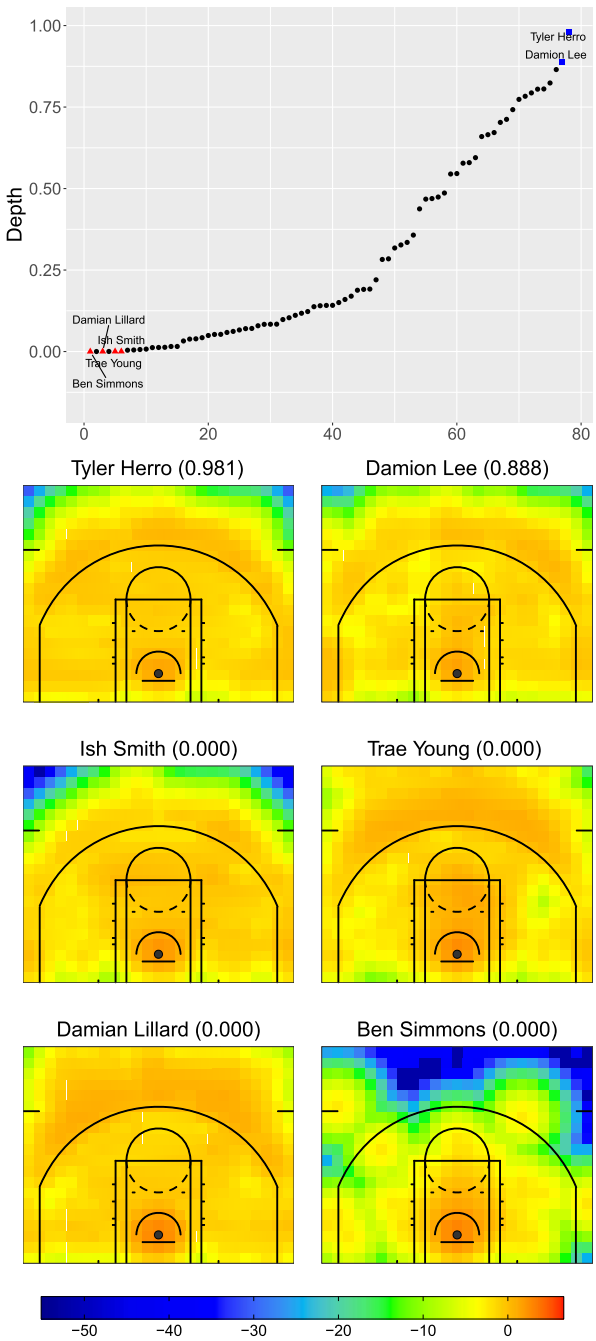


Figure 3. Depth Values for guard players. Same as Figure 2 except for the 78 guards.

The depth analysis result for the Forward and Center group is given in Figure 2, where we show the norm-based depth values on all 79 players on the upper-left panel. We see that the depth values vary from 0 to 1, where high values indicate typical players and low values indicate outliers. Two players with the highest depths are Maxi Kleber and P.J. Washington (shown as blue squares in the panel). Their heat maps are shown in the two plots on the upper right

panel. Both players' shooting locations concentrate around the basket and out beyond the three-point arc, which is a typical pattern for a forward or center. In contrast, we also show four low depth values as red triangles in the upper-left panel. To understand the abnormal patterns, we also show the shooting heat maps of these four players in the four plots, respectively, on the lower panel. Specifically, Kevin Love is detected as an abnormal player in this group since he has exceedingly wider shooting area than other players. Jimmy Butler is recognized different with other players since his long shots are mainly distributed on the right side of the heat map. Hassan Whiteside and Rudy Gobert are abnormal since they mainly shoot in the paint.

Figure 3 shows the depth result for the Guard group, which has the same layout as Figure 2. The norm-based depth values are also throughout the interval $[0, 1]$ among this group, and Tyler Herro and Damion Lee have the highest depth values. As shown in the two plots on the upper right panel, both players shot evenly in and out of the three-point arc, which indicates the typical shooting mode for guard players. We also illustrate the heat maps of four players with very low depth values in the four plots on the lower panel. Trae Young and Damian Lillard are detected as abnormal players due to their extensive shooting area. Both of them have very long shooting ranges. On the other hand, Ben Simmons and Ish Smith are assigned with low depth values since their active area is relatively smaller than typical guards. In particular, Ben Simmons acts as a center among guards.

4.3 Global temperature data

We also apply the proposed model-based depth method to the monthly global temperature anomalies data which is previous analyzed in Gu and Shen [11], Zhang, Shen and Kong [37]. This dataset is accessed from the National Oceanic and Atmospheric Administration (NOAA, https://www.ncdc.noaa.gov/temp-and-precip/ghcn-gridded-products/maps), which contains the monthly temperature anomalies (departures from a long-term average) from 1999 to 2018. Since there are numbers of missing values on some areas, we only consider the temperature of colored part (red, blue and grey) as shown in Figure 4. We estimate the statistical depth of temperature data for Summer (northern temperate zone; June, July and August) and Winter (northern temperate zone; December, January and February), respectively. For both seasons, we illustrate three months with lowest depth values and the average temperature anomaly of the five months with highest depth values.

We demonstrate the depth-based result for Summer group in Figure 4, where the norm-based depth values of monthly temperature anomalies are shown in the upper-left panel. The norm-based depth values are approximately dispersed from 0.1 to 1, where high values point to the typical global temperature conditions and low values point to outliers. Several NOAA climate reports are provided to confirm

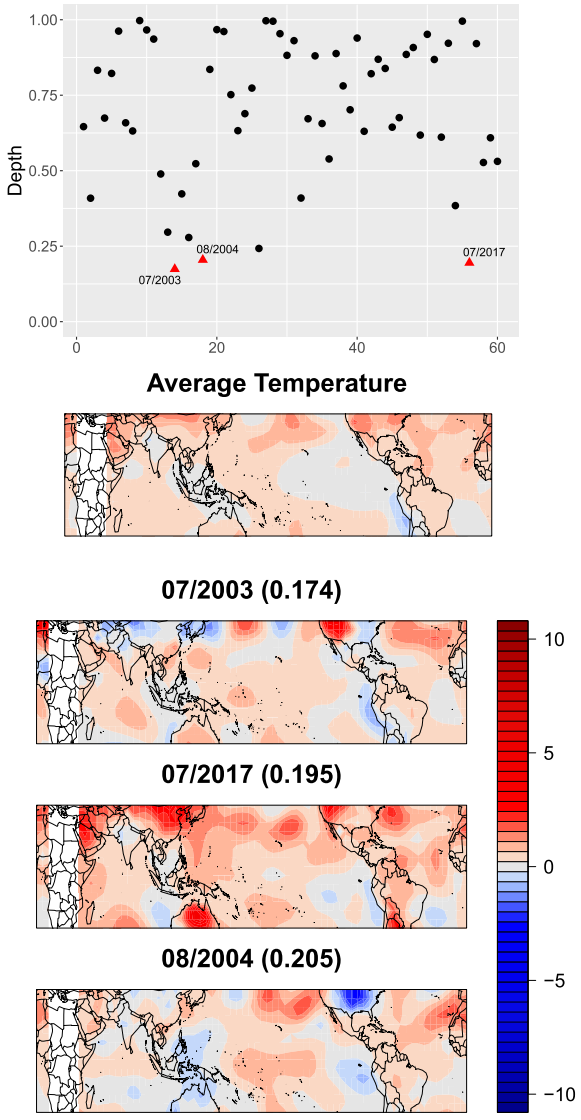


Figure 4. Depth Values for Summer Temperature Data. (a) Norm-based depth values of the Summer monthly temperature anomalies from 1999 to 2018, denoted by black dots, are shown on the upper left panel, with three low valued months denoted by red triangles. The average temperature anomaly graph of the five months with highest depth values is shown in the lower left plot and the temperature anomaly graph of three lowest valued months are shown in the three plots on the right panel, where the value in parenthesis of each month is its depth value.

the outliers detected by our depth analysis. For example, in July 2003 temperatures were much above the 1988–2002 average across Europe, the western U.S. and Southeast Asia though only a small part of Europe is considered in our analysis. The global temperature was also higher than average, which was the second warmest July since 1880. In July 2017, temperatures were above the average across much

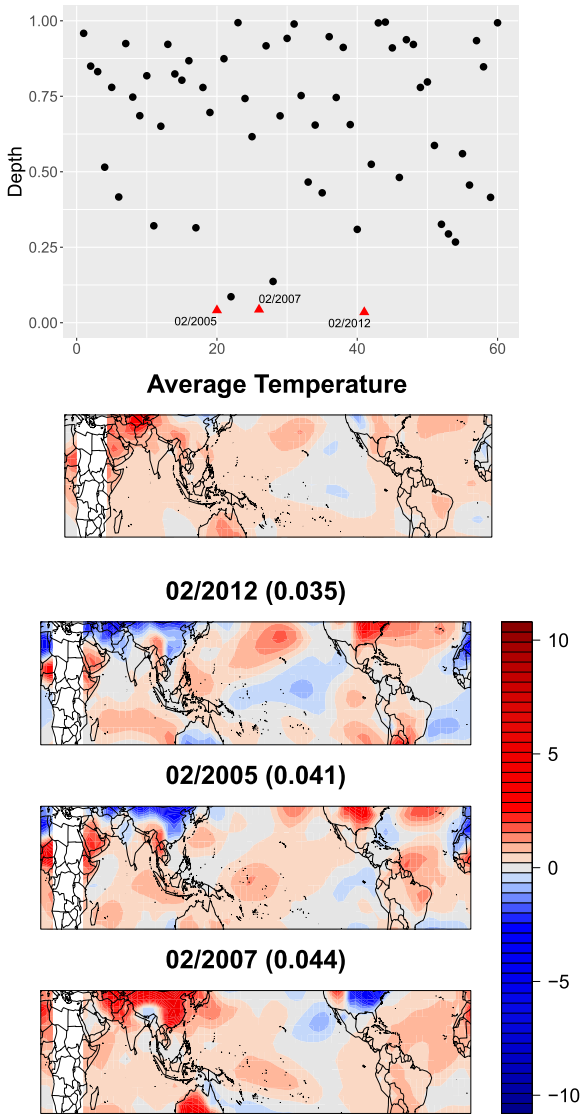


Figure 5. Depth Values for Winter Temperature Data. Same as Figure 4 except for the Winter group.

of the land and ocean surfaces, and were much warmer than the average across China, the western U.S, southern South America, the Middle East and Australia. In particular, on July 20, 2017, the temperature in Shanghai reached 40.9°C (105°F), which was the hottest day ever. On the contrary, August 2004 is detected as outlier since much of the U.S. were remarkably cooler than the average, and the cooler-than-average conditions also occurred across southern Southeast Asia and northern Australia.

The depth analysis for Winter group is displayed in Figure 5, whose layout is the same as Figure 4. We see the norm-based depth values are distributed from 0 to 1 for this group. The outliers in winter data are also well explained by NOAA reports. In February 2012, unusually cold weather occurred across Europe and Asia, whereas the eastern U.S. was much warmer than average. For February 2007, temperatures were much above the average in Europe and Asia and Australia, but were cooler than the average across the eastern U.S. In addition, There were sixth globally warmest February on record in 2007.

5. SUMMARY AND FUTURE WORK

In this paper, we have proposed a new framework to conduct functional principal component analysis for matrix data and applied it to compute model-based statistical depth. We specify the covariance tensor of a random matrix and use it to estimate the covariance function of matrix data, and then solve the eigen-decomposition for a 4th-order tensor by matricization. In addition, the Kronecker product covariance structure has been introduced to reduce the number of parameters in covariance estimation, and such an approach improves robustness and efficiency and avoid overfitting. Based on this new framework, we have used the Kronecker product covariance and principal component analysis to complete the algorithm for computing the defined model-based statistical depth. In addition, the extensive simulation studies validate the usefulness of our proposed method. Compared with several benchmark methods, our proposed method provides more robust ranking results based on empirical studies. In the analysis of the NBA field goal attempts data and global temperature data, our proposed method successfully detect both *abnormal* and *typical* patterns of field goal attempts and temperature anomalies. The NBA teams and environmental scientists can be equipped with more *objective* and principled analysis in their domains, and hence make better data-informed decisions. Our contributions not only address the existing gap in ranking methodologies for matrix data but also offer practical insights into the applications of our proposed framework in diverse domains.

There are several important directions for further investigation in the field of depth measures for matrix or tensor data. The tensor-times-matrix multiplication (Definition 2.2) can be naturally extended to higher order case. That is we can define the multiplication between an nth-order tensor and a $2n$th-order tensor. The covariance can also be specified for higher order tensors or for multivariate functions. The estimation of covariance under the tensor normal distribution assumption [26] and the development of covariance-based statistical depth measures in higher-order spaces offer interesting directions. While there are existing depth measures for multivariate data, extending them to the matrix or tensor data requires careful consideration. This involves defining appropriate notions of depth that capture the structural characteristics of matrices and tensors. Investigating the theoretical properties of depth measures for matrix or tensor data is crucial for understanding their performance and usefulness. This includes studying properties like continuity, consistency, affine invariance, and robustness. Computational efficiency is another important aspect to consider when working with matrix or tensor data.

We will explore and develop efficient algorithms to reduce computational complexity in depth methods. Incorporating sparsity structure in covariance estimation in high dimensional settings will also be an interesting topic to investigate in the future. Finally, while the established body of research on statistical depth primarily focuses on the unsupervised setting, the concept of depth for regression has also been proposed [29, 42]. Statistical depth given covariates for matrix data presents an intriguing idea and promising direction in our future research.

ACKNOWLEDGEMENTS

Dr. Hu's research is partially supported by NSF awards SES-2243058 and DMS-2412923.

Received 1 March 2023

REFERENCES

[1] Berrendero, J. R., Justel, A. and Svarc, M. (2011). Principal components for multivariate functional data. *Comput. Stat. Data Anal.* **55** 2619–2634. MR2802340

[2] Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* **35** 283–319.

[3] Cervone, D., D'Amour, A., Bornn, L. and Goldsberry, K. (2014). POINTWISE: Predicting points and valuing decisions in real time with NBA optical tracking data. In *Proceedings of the 8th MIT Sloan Sports Analytics Conference, Boston, MA, USA* **28** 3.

[4] Chen, E. Y. and Fan, J. (2021). Statistical inference for high-dimensional matrix-variate factor models. *J. Am. Stat. Assoc.* 1–18. MR4595475

[5] Chiou, J.-M., Chen, Y.-T. and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Stat. Sin.* 1571–1596. MR3308652

[6] Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014). Multivariate functional halfspace depth. *J. Am. Stat. Assoc.* **109** 411–423. MR3180573

[7] Dai, X., Lopez-Pintado, S. and Alzheimer's Disease Neuroimaging Initiative (2021). Tukey's depth for object data. *J. Am. Stat. Assoc.* **just-accepted** 1–37. MR4646604

[8] De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278. MR1780272

[9] Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. *J. Stat. Comput. Simul.* **64** 105–123.

[10] Franks, A., Miller, A., Bornn, L. and Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *Ann. Appl. Stat.* **9** 94–121. MR3341109

[11] Gu, M. and Shen, W. (2020). Generalized probabilistic principal component analysis of correlated data. *J. Mach. Learn. Res.* **21** 13–1. MR4071196

[12] Harshman, R. A. et al. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis.

[13] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators* **997**. John Wiley & Sons. MR3379106

[14] Hu, G., Yang, H.-C. and Xue, Y. (2021). Bayesian group learning for shot selection of professional basketball players. *Stat* **10** e324. MR4276018

[15] Hu, W., Pan, T., Kong, D. and Shen, W. (2020). Nonparametric matrix response regression with application to brain imaging data analysis. *Biometrics*. MR4357833

[16] Ieva, F. and Paganoni, A. M. (2013). Depth measures for multivariate functional data. *Commun. Stat., Theory Methods* **42** 1265–1276. MR3031280

[17] Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056

[18] Kong, D., An, B., Zhang, J. and Zhu, H. (2019). L2rm: Low-rank linear regression models for high-dimensional matrix responses. *J. Am. Stat. Assoc.*. MR4078472

[19] Li, C., Xiao, L. and Luo, S. (2020). Fast covariance estimation for multivariate sparse functional data. *Stat* **9** e245. MR4116315

[20] Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Stat.* 405–414. MR1041400

[21] Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Am. Stat. Assoc.* **88** 252–260. MR1212489

[22] López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *J. Am. Stat. Assoc.* **104** 718–734. MR2541590

[23] López-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Comput. Stat. Data Anal.* **55** 1679–1695. MR2748671

[24] López-Pintado, S., Sun, Y., Lin, J. K. and Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Adv. Data Anal. Classif.* **8** 321–338. MR3253863

[25] Mai, Q. and Zhang, X. (2023). Statistical methods for tensor data analysis. In *Springer Handbook of Engineering Statistics* 817–829. Springer.

[26] Manceur, A. M. and Dutilleul, P. (2013). Maximum likelihood estimation for the tensor normal distribution: Algorithm, minimum sample size, and empirical bias and dispersion. *J. Comput. Appl. Math.* **239** 37–49. MR2991957

[27] Miller, A., Bornn, L., Adams, R. and Goldsberry, K. (2014). Factorized point process intensities: A spatial analysis of professional basketball. In *International Conference on Machine Learning* 235–243.

[28] Narisetty, N. N. and Nair, V. N. (2016). Extremal depth for functional data and applications. *J. Am. Stat. Assoc.* **111** 1705–1714. MR3601729

[29] Rousseeuw, P. J. and Hubert, M. (1999). Regression depth. *J. Am. Stat. Assoc.* **94** 388–402. MR1702314

[30] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **71** 319–392. MR2649602

[31] Singh, K. (1991). A notion of majority depth (Technical report). *Rutgers University*.

[32] Soloveychik, I. and Trushin, D. (2016). Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness. *J. Multivar. Anal.* **149** 92–113. MR3507317

[33] Srivastava, M. S., von Rosen, T. and Von Rosen, D. (2008). Models with a Kronecker product covariance structure: estimation and testing. *Math. Methods Stat.* **17** 357–370. MR2483463

[34] Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975* **2** 523–531. MR0426989

[35] Wang, X., Zhu, H. and Alzheimer's Disease Neuroimaging Initiative (2017). Generalized scalar-on-image regression models via total variation. *J. Am. Stat. Assoc.* **112** 1156–1168. MR3735367

[36] Yin, F., Hu, G. and Shen, W. (2023). Analysis of professional basketball field goal attempts via a Bayesian matrix clustering approach. *Journal of Computational and Graphical Statistics* **32:1** 49–60. MR4552936

[37] Zhang, Y., Shen, W. and Kong, D. (2020). Covariance estimation for matrix-valued data. *arXiv preprint arXiv:2004.05281*. MR4681608

[38] Zhang, A. and Xia, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445
[39] Zhao, W., Xu, Z., Mu, Y., Yang, Y. and Wu, W. (2023). Model-based statistical depth with applications to functional data. *J. Nonparametr. Stat.* 1–44.
[40] Zhou, H. and Li, L. (2014). Regularized matrix regression. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **76** 463–483. MR3164874
[41] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Ann. Stat.* **31** 1460–1490. MR2012822
[42] Zuo, Y. (2021). On general notions of depth for regression. *Stat. Sci.* **36** 142–157. MR4194208

Yue Mu
Department of Statistics
Florida State University
117 N. Woodward Ave
Tallahassee, FL
United States
E-mail address: ym18e@fsu.edu

Guanyu Hu
Center for Spatial Temporal Modeling for Applications in Population Sciences
Department of Biostatistics and Data Science
The University of Texas Health Science Center at Houston
7000 Fannin Street
Houston, TX
United States
E-mail address: Guanyu.Hu@uth.tmc.edu

Wei Wu
Department of Statistics
Florida State University
117 N. Woodward Ave
Tallahassee, FL
United States
E-mail address: wwu@fsu.edu