Dimerison Reduction and Data Visualization for Fréchet Regression

Qi Zhang, Lingzhou Xue, and Bing Li
Department of Statistics, Pennsylvania State University

Abstract

With the rapid development of data collection techniques, complex data objects that are not in the Euclidean space are frequently encountered in new statistical applications. Fréchet regression model (Peterson & Müller 2019) provides a promising framework for regression analysis with metric space-valued responses. In this paper, we introduce a flexible sufficient dimension reduction (SDR) method for Fréchet regression to achieve two purposes: to mitigate the curse of dimensionality caused by high-dimensional predictors, and to provide a tool for data visualization for Fréchet regression. Our approach is flexible enough to turn any existing SDR method for Euclidean $(X, Y)$ into one for Euclidean $X$ and metric space-valued $Y$. The basic idea is to first map the metric-space valued random object $Y$ to a real-valued random variable $f(Y)$ using a class of functions, and then perform classical SDR to the transformed data. If the class of functions is sufficiently rich, then we are guaranteed to uncover the Fréchet SDR space. We showed that such a class, which we call an ensemble, can be generated by a universal kernel. We established the consistency and asymptotic convergence rate of the proposed methods. The finite-sample performance of the proposed methods is illustrated through simulation studies for several commonly encountered metric spaces that include Wasserstein space, the space of symmetric positive definite matrices, and the sphere. We illustrated the data visualization aspect of our method by exploring the human mortality distribution data across countries and by studying the distribution of hematoma density.

Keywords: Ensembled sufficient dimension reduction, Inverse regression, Statistical objects, Universal kernel, Wasserstein space.
1 Introduction

With the rapid development of data collection techniques, complex data objects that are not in the Euclidean space are frequently encountered in new statistical applications, such as the graph Laplacians of networks, the covariance or correlation matrices for the brain functional connectivity in neuroscience (Ferreira and Busatto 2013), and probability distributions in CT hematoma density data (Petersen and Müller 2019). These data objects, also known as random objects, do not obey the operation rules of a vector space with an inner product or a norm, but instead reside in a general metric space. In a prescient paper, Fréchet (1948) proposed the Fréchet mean as a natural generalization of the expectation of a random vector. By extending the Fréchet mean to the conditional Fréchet mean, Petersen and Müller (2019) introduced the Fréchet regression model with random objects as response and Euclidean vectors as predictors, which provides a promising framework for regression analysis with metric space-valued responses. Dubey and Müller (2019) showed the consistency of the sample Fréchet mean using the results of Petersen and Müller (2019), derived a central limit theorem for the sample Fréchet variance that quantifies the variation around the Fréchet mean, and further developed the Fréchet analysis of variance for random objects. Dubey and Müller (2020) designed a method for change-point detection and inference in a sequence of metric-space-valued data objects.

The Fréchet regression of Petersen and Müller (2019) employs the global least squares and the local linear or polynomial regression to fit the conditional Fréchet mean. It is well known that the global least squares is based on a restrictive assumption of the regression relation. Although the local regression is more flexible, it is effective only when the dimension of predictor is relatively low. As this dimension gets higher, its accuracy drops significantly—a phenomenon known as the curse of dimensionality. To address this issue, it is essential to reduce the dimension of the predictor without losing the information about response. For classical regression, this task is performed by sufficient dimension reduction (SDR; see Li 1991, Cook and Weisberg 1991, Cook 1996 and Li 2018 among others). SDR works by projecting the high-dimensional predictor onto a low-dimensional subspace that preserves the information about response, through the use of sufficiency.

Besides assisting regression to overcome curse of dimensionality, another important function of SDR for classical regression is to provide a data visualization tool to gain insights into how the regression surface looks like in high-dimensional space before even fitting a model. This function is also needed in Fréchet regression. In fact, it can be argued that data visualization is even more important for regression of random objects, as the regression
relation may be even more difficult to discern among the complex details of the objects.

To fulfill these demands, in this paper, we systematically develop the theories and methodologies of sufficient dimension reduction for Fréchet regression. To set the stage, we first give an outline of SDR for classical regression. Let $X$ be a $p$-dimensional random vector in $\mathbb{R}^p$ and $Y$ a random variable in $\mathbb{R}$. The classical SDR aims to find a dimension reduction subspace $S$ of $\mathbb{R}^p$ such that $Y$ and $X$ are independent conditioning on $P_S X$, that is, $Y \perp X | P_S X$, where $P_S$ is the projection on to $S$ with respect to the usual inner product in $\mathbb{R}^p$. In this way, $P_S X$ can be used as the synthetic predictor without loss of regression information about the response $Y$. Under mild conditions, the intersection of all such dimension reduction subspaces is also a dimension reduction subspace, and the intersection is called the central subspace denoted by $S_{Y|X}$ (Cook, 1996; Yin et al., 2008). For the situation where the primary interest is in estimating the regression function, Cook and Li (2002) introduced a weaker form of SDR, the mean dimension reduction subspace. A subspace $S$ of $\mathbb{R}^p$ is a mean SDR subspace if satisfies $E(Y|X) = E(Y|P_S X)$, and the intersection of all such spaces, if it is still a mean SDR subspace, is the central mean subspace, denoted by $S_{E(Y|X)}$. The central mean subspace $S_{E(Y|X)}$ is always contained in central subspace $S_{Y|X}$ when they exist. Many estimating methods for the central subspace and the central mean subspace have been developed over the past decades. For example, for the central subspace, we have the sliced inverse regression (SIR; Li 1991), the sliced average variance estimate (SAVE; Cook and Weisberg 1991), the contour regression (CR; Li et al. 2005), and the directional regression (DR; Li and Wang 2007). For the central mean subspace, we have the ordinary least squares (OLS; Li and Duan 1989), the principal Hessian directions (PHD; Li 1992), the iterative Hessian transformation (IHT; Cook and Li 2002), the outer product of gradients (OPG) and the minimum average variance estimator (MAVE) of Xia et al. (2002).

SDR has been extended to accommodate some complex data structures in the past, for example, to functional data (Ferré and Yao 2003; Hsing and Ren 2009; Li and Song 2017), to tensorial data (Li et al. 2010; Ding and Cook 2015), and to panel data (Fan et al. 2017; Yu et al. 2020; Luo et al. 2021). Most recently, Ying and Yu (2020) extended SIR to the case where the response takes values in a metric space. Taking a substantial step forward, in this paper we introduce a comprehensive and flexible method that can adapt any existing SDR estimators to metric-space-valued responses.

The basic idea of our method stems from the ensemble SDR for Euclidean $X$ and $Y$ of Yin and Li (2011), which recovers the central subspace $S_{Y|X}$ by repeatedly estimating the central mean subspace $S_{E(f(Y)|X)}$ for a family $F$ of functions $f$ that is rich enough to
determine the conditional distribution of $Y|X$. Such a family $\mathcal{F}$ is called an ensemble, and satisfies $\mathcal{S}_{Y|X} = \cup \{ \mathcal{S}_{E[f(Y)|X]} : f \in \mathcal{F} \}$. Using this relation, we can turn any method for estimating the central mean space into one that estimate the central subspace.

While borrowing the idea of ensemble, our goal is different from Yin and Li (2011): we are not interested in turning an estimator for the central mean subspace into one for the central subspace. Instead, we are interested in turning any existing SDR method for Euclidean $(X,Y)$ into one for Euclidean $X$ and metric space-valued $Y$. Let $X$ be a random vector in $\mathbb{R}^p$ and $Y$ a random object that takes values in a metric space $(\Omega_Y, d)$. Still use the symbol $S_{Y|X}$ to represent the intersection of all subspaces of $\mathbb{R}^p$ satisfying $Y \parallel X | P_X$. We call $S_{Y|X}$ the central subspace for Fréchet SDR, or simply the Fréchet central subspace. Let $\mathcal{F}$ be a family of functions $f : \Omega_Y \rightarrow \mathbb{R}$ that are measurable with respect to the Borel $\sigma$-field on the metric space. We use two types of ensembles to connect classical SDR with Fréchet SDR:

- **Central Mean Space ensemble (CMS-ensemble)** is a family $\mathcal{F}$ that is rich enough so that $\mathcal{S}_{Y|X} = \cup \{ \mathcal{S}_{E[f(Y)|X]} : f \in \mathcal{F} \}$. Note that we know how to estimate the spaces $\mathcal{S}_{E[f(Y)|X]}$ using the existing SDR methods, since $f(Y)$ is a number. We use this ensemble to turn an SDR method that targets the central mean subspace into one that targets the Fréchet central subspace. We will focus on two forward regression methods: OPG and MAVE and three moment estimators of the CMS.

- **Central Space ensemble (CS-ensemble)** is a family $\mathcal{F}$ that is rich enough so that $\mathcal{S}_{Y|X} = \cup \{ \mathcal{S}_{f(Y)|X} : f \in \mathcal{F} \}$. We use this ensemble to turn an SDR method that targets the central subspace for real-valued response into one that targets the Fréchet central subspace. We will focus on three inverse regression methods: SIR, SAVE, and DR.

A key step in implementing both of the above schemes is to construct an ensemble $\mathcal{F}$ in each case. For this purpose we assume that the metric space $(\Omega_Y, d)$ is continuously embeddable into a Hilbert space. Under this assumption one can construct a universal reproducing kernel, which leads to an $\mathcal{F}$ that satisfies the required characterizing property.

As with classical SDR, the Fréchet SDR can also be used to assist data visualization. To illustrate this aspect, we consider an application involving factors that influence the mortality distributions of 162 countries (see Section 8 for details). For each country, the response is a histogram with numbers of deaths for each five-year period from age 0 to age 100, which is smoothed to produce a density estimate, as shown in panel (a) of Figure 1. We considered nine predictors characterizing each country’s demography, economy, labor market, health care, and environment. Using our ensemble method we obtained a set of sufficient
predictors. In panel (b) of Figure 1 we show the mortality densities plotted against the first sufficient predictor. A clear pattern is shown in the plot: for countries with low values of the first sufficient predictor, the modes of the mortality distributions are at lower ages, and there are upticks at age 0, indicating high infant mortality rates; for countries with high values of the first sufficient predictor, the modes of the mortality distributions are significantly higher, and there are no upticks at age 0, indicating very low infant mortality rates. The information provided by the plot is clearly useful, and much further insights can be gained about what affects the mortality distribution by taking a careful look at the loadings of the first sufficient predictor, as will be detailed in Section 8.

Figure 1: Data visualization in Fréchet regression for mortality distributions of 162 countries. Panel (a) plots mortality densities that are placed in the random order, and Panel (b) plots mortality densities versus the first sufficient predictor estimated by our ensemble method.

The rest of this paper is organized as follows. Section 2 defines the Fréchet SDR problem and provides sufficient conditions for a family $\mathfrak{F}$ to characterize the central subspace. Section 3 then constructs ensemble $\mathfrak{F}$ for the Wasserstein space of univariate distributions, the space of covariance matrix, and a special Riemannian manifold, the sphere. Section 4 proposes the CMS-ensembles by extending five existing SDR methods that target the central mean space for real-valued response: OLS, PHD, IHT, OPG and MAVE. Section 5 proposes the CS-ensemble for three existing SDR methods that target the central subspace for real-valued response: SIR, SAVE, and DR. Section 6 establishes the convergence rate of the proposed methods. Section 7 uses simulation studies to examine the numerical performances of different ensemble estimators in different settings including distributional responses and covariance matrix responses. In Section 8 we explore two real distributions to demonstrate the usefulness of our methods. All the proofs are presented in the Supplementary Material.
2 Characterization of the Fréchet Central Subspace

Let \((\Omega, \mathcal{F}, P)\) be a probability space. Let \((\Omega_Y, d)\) be a metric space with metric \(d\) and \(\mathcal{B}_Y\) the Borel \(\sigma\)-field generated by the open sets in \(\Omega_Y\). Let \(\Omega_X\) be a subset of \(\mathbb{R}^p\) and \(\mathcal{B}_X\) the Borel \(\sigma\)-field generated by the open sets in \(\Omega_X\). Let \((X, Y)\) be a random element mapping from \(\Omega\) to \(\Omega_X \times \Omega_Y\) measurable with respect to the product \(\sigma\)-field \(\mathcal{B}_X \times \mathcal{B}_Y\). We denote the marginal distributions of \(X\) and \(Y\) by \(P_X\) and \(P_Y\), respectively, and the conditional distributions of \(Y|X\) and \(X|Y\) by \(P_{Y|X}\) and \(P_{X|Y}\), respectively. We formulate the Fréchet SDR problem as finding a subspace \(S\) of \(\mathbb{R}^p\) such that \(Y\) and \(X\) are independent conditioning on \(P_S X\):

\[ Y \perp\!\!\!\!\perp X | P_S X, \]  

where \(P_S\) is the projection on to \(S\) with respect to the inner product in \(\mathbb{R}^p\). As in the classical definition of SDR, under mild conditions (Cook and Li, 2002) the intersection of all such subspaces \(S\) still satisfies (1). We call this subspace the Fréchet central subspace and denote it by \(S_{Y|X}\). Similar to Cook (1996), it can be shown that, if the support of \(X\) is open and convex, the Fréchet central subspace \(S_{Y|X}\) satisfies (1).

2.1 Two types of ensembles and their sufficient conditions

Let \(\mathcal{F}\) be a family of measurable functions \(f : \Omega_Y \rightarrow \mathbb{R}\), and for an \(f \in \mathcal{F}\), let \(S_{E[f(Y)|X]}\) be the central mean subspace of \(f(Y)\) versus \(X\). As mentioned in the Introduction, we use two types of ensembles to recover the Fréchet central subspace. The first type is defined as any \(\mathcal{F}\) that satisfies

\[ \text{span}\{S_{E[f(Y)|X]} : f \in \mathcal{F}\} = S_{Y|X}. \]  

This is the same ensemble as that in Yin and Li (2011), except that, here, the right-hand side is the Fréchet central subspace. Relation (2) allows us to recover the Fréchet central subspace \(S_{Y|X}\) by a collection of classical central mean subspaces. We call a class \(\mathcal{F}\) that satisfies (2) a central mean space ensemble (or CMS-ensemble). The second type of ensembles is any family \(\mathcal{F}\) that satisfies

\[ \text{span}\{S_{f(Y)|X} : f \in \mathcal{F}\} = S_{Y|X}, \]  

which we call a central space ensemble (or CS-ensemble). The next proposition shows that a CMS ensemble is always a CS-ensemble.

**Proposition 1.** If \(\mathcal{F}\) is a CMS-ensemble, then it is a CS-ensemble.
We next develop a sufficient condition for an $\mathcal{F}$ to be a CMS-ensemble, and hence also a CS-ensemble. Let $\mathcal{B} = \{ I_B : B \text{ is Borel set in } \Omega_Y \}$ be the family of measurable indicator functions on $\Omega_Y$. Yin and Li (2011) showed that if $\mathcal{F}$ is a subset of $L_2(P_Y)$ that is dense in $\mathcal{B}$, then (2) holds for the classical $\mathcal{S}_{|X}$. Here, we generalize that result to our setting by requiring only $\text{span}(\mathcal{F})$ to be dense in $\mathcal{B}$, where $\text{span}(\mathcal{F})$ is the linear span

$$\text{span}(\mathcal{F}) = \left\{ \sum_{i=1}^{k} \alpha_i f_i : k \in \mathbb{N}, \alpha_1, \ldots, \alpha_k \in \mathbb{R}, f_1, \ldots, f_k \in \mathcal{F} \right\}.$$ 

**Lemma 1.** If $\mathcal{F}$ is a subset of $L_2(P_Y)$ and $\text{span}(\mathcal{F})$ is dense in $\mathcal{B}$ with respect to the $L_2(P_Y)$-metric, then $\mathcal{F}$ is a CMS-ensemble and hence also a CS-ensemble.

### 2.2 Construction of the CMS-ensemble

To construct a CMS-ensemble, we resort to the notion of the universal kernel. Suppose $(\Omega_Y, d)$ is compact and let $C(\Omega_Y)$ be the family of continuous real-valued functions on $\Omega_Y$. The universal kernel on $\Omega_Y$ is defined as the continuous kernel on $\Omega_Y$ such that the RKHS generated by this kernel is dense in $C(\Omega_Y)$ with respect to the uniform norm (Steinwart and Christmann, 2008). The next theorem connects the universal kernel on $(\Omega_Y, d)$ with the CMS-ensemble, which is the theoretical foundation of our method.

**Theorem 1.** If $(\Omega_Y, d)$ is a compact metric space and $\kappa : \Omega_Y \times \Omega_Y \to \mathbb{R}$ is a universal kernel, then the family $\mathcal{F} = \{ \kappa(\cdot, y) : y \in \Omega_Y \}$ is a CMS-ensemble.

We next look into the conditions under which a metric space has a universal kernel and how to construct such a kernel when it does. It is well known that if $\Omega_Y$ is a compact subset of $\mathbb{R}^d$, many standard kernels, including Laplacian kernels and Gaussian RBF kernels, are universal. Unfortunately, when $\Omega_Y$ is a general metric space, direct extension of these types of kernels, for example $k(y, y') = \exp(-\gamma d(y, y')^2)$, are no longer guaranteed to be universal. Christmann and Steinwart (2010) showed that if there exists a separable Hilbert space $\mathcal{H}$ and a continuous injection $\rho : \Omega_Y \to \mathcal{H}$, then for any analytic function $K : \mathbb{R} \to \mathbb{R}$ whose Taylor series at zero has strictly positive coefficients, the function $\kappa(y, y') = K(\langle \rho(y), \rho(y') \rangle_{\mathcal{H}})$ defines a universal kernel on $\Omega_Y$. They also provide an analogous definition of the Gaussian-type kernel in the above case. To make the paper self-contained, we reproduce their result below as a proposition and attach a proof in the Supplementary Material.

**Proposition 2.** Suppose $(\Omega_Y, d)$ is a compact metric space, and there exists a separable Hilbert space $\mathcal{H}$ and a continuous injection $\rho : \Omega_Y \to \mathcal{H}$. If $K : \mathbb{R} \to \mathbb{R}$ is an analytic
function of the form
\[ K(t) = \sum_{n=0}^{\infty} a_n t^n, \quad t \in [-r, r], \text{ where } 0 < r \leq +\infty, \quad a_n \geq 0 \text{ for all } n \geq 1 \] (4)
then the function \( \kappa : \Omega_Y \times \Omega_Y \rightarrow \mathbb{R} \) defined by \( \kappa(y, y') = K(\langle \rho(y), \rho(y') \rangle_{\mathcal{H}}) \) is a positive definite kernel. Furthermore, if \( a_n > 0 \) for all \( n \geq 1 \), then \( \kappa \) is a universal kernel.

As an example, the Gaussian radial basis function is universal on \( \Omega_Y \) and produces a CMS-ensemble.

**Corollary 1.** Suppose the conditions in Proposition 2 are satisfied and \( \kappa_\gamma \) is the Gaussian-type kernel \( \kappa_\gamma(y, y') = \exp(-\gamma \|\rho(y) - \rho(y')\|_{\mathcal{H}}^2) \) where \( \gamma > 0 \). Consequently, \( \mathcal{F} = \{\kappa_\gamma(\cdot, y), y \in \Omega_Y\} \) is a CMS-ensemble.

**Corollary 2.** If, in Proposition 2, the continuous function \( \rho : \Omega_Y \rightarrow \mathcal{H} \) is isometric, that is, \( d(y, y') = \|\rho(y) - \rho(y')\|_{\mathcal{H}} \), then Gaussian-type kernel \( \kappa_\gamma(y, y') = \exp(-\gamma d^2(y, y')) \) is universal and \( \mathcal{F} = \{\kappa_\gamma(\cdot, y), y \in \Omega_Y\} \) is a CMS-ensemble.

This is a direct consequence of Corollary 1 since an isometry is an injection. The similar results to Corollaries 1 and 2 can also be established for Laplacian type kernel \( \kappa_\gamma(y, y') = \exp(-\gamma \|\rho(y) - \rho(y')\|_{\mathcal{H}}) \). Proposition 2 provides a general mechanism to construct the CMS-ensemble over a metric space without a linear structure, provided it can be continuously embedded in a separable Hilbert space. Although the continuous embedding condition looks somewhat restrictive at first sight, it indeed covers several scenarios of particularly interest in statistics. In the next section, we apply Proposition 2 to construct the CMS-ensemble on the space of univariate distributions endowed with the Wasserstein-2 distance, the space of correlation matrix endowed with Frobenius distance, and the sphere endowed with geodesic distance. These metric spaces are often encountered in statistical applications.

### 3 Important Metric Spaces and their CMS Ensembles

In this section, we illustrate the construction of CMS-ensembles for three metric spaces commonly encountered in statistics.

#### 3.1 Wasserstein space

Let \( I \) be \( \mathbb{R} \) or a closed interval of \( \mathbb{R} \), \( \mathcal{B}(I) \) the \( \sigma \)-field of Borel subsets of \( I \) and \( \mathcal{P}(I) \) the collection of all probability measures on \((I, \mathcal{B}(I))\). The Wasserstein space \( \mathcal{W}_2(I) \) is defined as the
subset of \( \mathcal{P}(I) \) with finite second moment, that is, \( \mathcal{W}_2(I) = \{ \mu \in \mathcal{P}(I) : \int_I t^2 \, d\mu(t) < \infty \} \), endowed with the quadratic Wasserstein distance \( d_W(\mu_1, \mu_2) = \left( \int_0^1 [F_{\mu_1}^{-1}(s) - F_{\mu_2}^{-1}(s)]^2 \, ds \right)^{1/2} \), where \( \mu_1 \) and \( \mu_2 \) are members of \( \mathcal{W}_2(I) \) and \( F_{\mu_1}^{-1} \) and \( F_{\mu_2}^{-1} \) are the quantile functions of \( \mu_1 \) and \( \mu_2 \), which we assume to be well defined. This distance can be equivalently written as
\[
d_W(\mu_1, \mu_2) = \left( \int_I [F_{\mu_1}^{-1}(t) - F_{\mu_2}^{-1}(t)]^2 \, d\mu(t) \right)^{1/2}.
\]
The set \( \mathcal{W}_2(I) \) endowed with \( d_W \) is a metric space with a formal Riemannian structure (Ambrosio et al., 2004).

Here, we present some basic results that characterize \( \mathcal{W}_2(I) \), whose proofs can be found, for example, in (Ambrosio et al., 2004) and (Bigot et al., 2017). For \( \mu_1, \mu_2 \in \mathcal{W}_2(I) \), we say that a \( \mathcal{B}(I) \)-measurable map \( r : I \to I \) transports \( \mu_1 \) to \( \mu_2 \) if \( \mu_2 = \mu_1 \circ r^{-1} \). This relation is often written as \( \mu_2 = r_{\#} \mu_1 \). Let \( \mu_0 \in \mathcal{W}_2(I) \) be a reference measure with a continuous \( F_{\mu_0} \). The tangent space at \( \mu_0 \) is \( T_{\mu_0} = \text{cl}_{L_2(\mu_0)} \{ \lambda (F_{\mu_0}^{-1} - \text{id}) : \mu \in \mathcal{W}_2(I), \lambda > 0 \} \), where, for a set \( A \subseteq L_2(\mu_0) \), \( \text{cl}_{L_2(\mu_0)}(A) \) denotes the \( L_2(\mu_0) \)-closure of \( A \). The exponential map \( \exp_{\mu_0} \) from \( T_{\mu_0} \) to \( \mathcal{W}_2(I) \), defined by \( \exp_{\mu_0}(r) = (r + \text{id})_{\#} \mu_0 \), is surjective. Therefore its inverse, \( \log_{\mu_0} : \mathcal{W}_2(I) \to T_{\mu_0} \), defined by \( \log_{\mu_0}(\mu) = F_{\mu_0}^{-1} \circ F_{\mu_0} - \text{id} \), is well defined on \( \mathcal{W}_2(I) \). It is well known that the exponential map restricted to the image of log map, denoted as \( \exp_{\mu_0} \big| \log_{\mu_0}(\mu)_{(\mathcal{W}_2(I))} \), is an isometric homeomorphism (Bigot et al., 2017). Therefore, \( \log_{\mu_0} \) is a continuous injection from \( \mathcal{W}_2(I) \) to \( L_2(\mu_0) \). We can then construct CMS-ensembles using the general constructive method provided by Theorem 1 and Proposition 2. The next proposition gives two such constructions.

**Proposition 3.** If \( I \) is a closed interval of \( \mathbb{R} \), then both
\[
\kappa_G(y, y') = \exp(-\gamma \| \log_{\mu_0}(y) - \log_{\mu_0}(y') \|^2_{L_2(\mu_0)}) = \exp(-\gamma d_W(y, y')) \quad \text{and}
\]
\[
\kappa_L(y, y') = \exp(-\gamma \| \log_{\mu_0}(y) - \log_{\mu_0}(y') \|^2_{L_2(\mu_0)}) = \exp(-\gamma d_W(y, y'))
\]
are universal kernels on \( \mathcal{W}_2(I) \). Consequently, the families \( \mathcal{H}_G = \{ \exp(-\gamma d_W(\cdot, t)^2) : t \in \mathcal{W}_2(I) \} \) and \( \mathcal{H}_L = \{ \exp(-\gamma d_W(\cdot, t)) : t \in \mathcal{W}_2(I) \} \) are CMS-ensembles.

The subscripts “G” and “L” for the two kernels in Proposition 3 refer to “Gaussian” and “Laplacian”, respectively.

### 3.2 Space of symmetric positive definite matrices

We first introduce some notations. Let \( \text{Sym}(r) \) be the set of \( r \times r \) invertible symmetric matrices with real entries and \( \text{Sym}^+(r) \) the set of \( r \times r \) symmetric positive definite (SPD) matrices. For any \( Y \in \mathbb{R}^{r \times r} \), the *matrix exponential* of \( Y \) is defined as the infinite power
series \( \exp(Y) = \sum_{k=0}^{\infty} Y^k/k! \). For any \( X \in \text{Sym}^+(r) \), the \textit{matrix logarithm} of \( X \) is defined as any \( r \times r \) matrix \( Y \) such that \( \exp(Y) = X \) and denoted by \( \log(X) \).

Let \( d_F \) be the Frobenius metric. Then \((\text{Sym}(r), d_F)\) is a metric space continuously embedded by identity mapping in \( \text{Sym}(r) \), which is a Hilbert space with the Frobenius inner product \( \langle A, B \rangle = \text{tr}(A^tB) \). Also, the identity mapping \( \text{id} : \text{Sym}^+(r) \to \text{Sym}(r) \) is obviously isometric. Therefore, by Corollary 2, the two types of radial basis function kernels for Wasserstein space can be similarly extended to \( \text{Sym}^+(r) \). That is, let \( \kappa_G(y, y') = \exp(-\gamma d_F(y, y')^2) \) and \( \kappa_L(y, y') = \exp(-\gamma d_F(y, y')) \), then \( \mathcal{F}_G = \{\kappa_G(y, y'), y' \in \text{Sym}^+(r)\} \) and \( \mathcal{F}_L = \{\kappa_L(y, y'), y' \in \text{Sym}^+(r)\} \) are CMS-ensembles.

Another widely used metric over \( \text{Sym}^+(r) \) is the log-Euclidean distance that is defined as \( d_{\log}(Y_1, Y_2) = \| \log(Y_1) - \log(Y_2) \|_F \). Basically, it pulls the Frobenius metric on \( \text{Sym}(r) \) back to \( \text{Sym}^+(r) \) by the matrix logarithm map. The matrix logarithm \( \log(\cdot) \) is a continuous injection to Hilbert \( \text{Sym}(r) \). By Corollary 2, the two types of radial basis function kernels \( \kappa_{G, \log}(y, y') = \exp(-\gamma d_{\log}(y, y')^2) \) and \( \kappa_{L, \log}(y, y') = \exp(-\gamma d_{\log}(y, y')) \) are universal. Then, \( \mathcal{F}_{G, \log} = \{\kappa_{G, \log}(y, y'), y' \in W_2(I)\} \) and \( \mathcal{F}_{L, \log} = \{\kappa_{L, \log}(y, y'), y' \in W_2(I)\} \) are CMS-ensembles.

3.3 The sphere

Consider the random vector taking values in the sphere \( \mathbb{S}^n = \{x \in \mathbb{R}^n : \|x\| = 1\} \). To respect the nonzero curvature of \( \mathbb{S}^n \), the geodesic distance \( d_g(Y_1, Y_2) = \arccos(Y_1^tY_2) \), which is derived from its Riemannian geometry, is often used rather than the Euclidean distance. However, the popular Gaussian-type RBF kernel \( \kappa_G(y, y') = \exp(-\gamma d_g(y, y')^2) \) is not positive definite on \( \mathbb{S}^n \) \cite{Jayasumana13}. In fact, \cite{Feragen15} proved that for complete Riemannian manifold \( M \) with its associated geodesic distance \( d_g \), \( \kappa_G(y, y') = \exp(-\gamma d_g(y, y')^2) \) is positive semidefinite only if \( M \) is isometric to a Euclidean space. \cite{Honeine10} and \cite{Jayasumana13} proved that the Laplacian-type kernel \( \kappa_L(y, y') = \exp(-\gamma d_g(y, y')) \) is positive definite on the sphere \( \mathbb{S}^n \). We show in the following proposition that \( \kappa_L(y, y') \) is universal.

**Proposition 4.** The Laplacian-type kernel \( \kappa_L(y, y') : \mathbb{S}^n \times \mathbb{S}^n \to \mathbb{R} \), defined by \( \kappa_L(y, y') = \exp(-\gamma d_g(y, y')) \), where \( d_g \) is the geodesic distance on \( \mathbb{S}^n \), is a universal kernel for any \( \gamma > 0 \). Consequently, \( \mathcal{F}_L = \{\exp(-\gamma d_g(\cdot, t)), t \in \mathbb{S}^n\} \) is a CMS-ensemble.
4 Fréchet SDR via CMS-ensemble

In this section we develop the Fréchet SDR estimators based on the CMS-ensembles and establish their Fisher consistency.

4.1 Ensembled moment estimators

We first develop a general class of Fréchet SDR estimators based on the ensembled moment estimators of the CMS, such as the OLS, PHD and IHT. Let $\mathcal{P}_{XY}$ be the collection of all distributions of $(X,Y)$, and let $M : \mathcal{P}_{XY} \to \mathbb{R}^{p \times p}$ be a measurable function to be used as an estimator of the Fréchet central subspace $S_{Y|X}$. A function defined on $\mathcal{P}_{XY}$ is called statistical functional; see, for example, Chapter 9 of Li (2018). In the SDR literature, such a function is also called a candidate matrix (Ye and Weiss, 2003). Let $F_{XY}$ be a generic member of $\mathcal{P}_{XY}$, $F_{0}^{(0)}_{XY}$ the true distribution of $(X,Y)$, and $F_{n}^{(n)}_{XY}$ the empirical distribution of $(X,Y)$ based on an i.i.d. sample $(X_{1},Y_{1}),\ldots,(X_{n},Y_{n})$. Extending the terminology of classical SDR (see, for example, Li 2018, Chapter 2), we say that the estimate $M_{0}(F_{XY})$ is unbiased if $\text{span}\{M_{0}(F_{XY}, \kappa(\cdot, y))\} \subseteq S_{Y|X}$, exhaustive if $\text{span}\{M_{0}(F_{XY}, \kappa(\cdot, y))\} \supseteq S_{Y|X}$, and Fisher consistent if $\text{span}\{M_{0}(F_{XY}, \kappa(\cdot, y))\} = S_{Y|X}$. We refer to $M_{0}$ as the Fréchet candidate matrix.

Suppose we are given a CMS-ensemble $\mathcal{F}$. Let $M_{0} : \mathcal{P}_{XY} \times \mathcal{F} \to \mathbb{R}^{p \times p}$ be a function to be used as an estimator of $S_{E_{f}(Y)|X}$ for each $f$. This is not a statistical functional in the classical sense, as it involves an additional set $\mathcal{F}$. So, we redefine unbiasedness, exhaustiveness, and Fisher consistency for this type of augmented statistical functional.

**Definition 1.** We say that $M_{0}$ is unbiased for estimating $\{S_{E_{f}(Y)|X} : f \in \mathcal{F}\}$ if, for each $f \in \mathcal{F}$, $\text{span}\{M_{0}(F_{XY}, f)\} \subseteq S_{E_{f}(Y)|X}$. Exhaustiveness and Fisher consistency of $M_{0}$ are defined by replacing $\subseteq$ in the above by $\supseteq$ and $=$, respectively.

Note that $M_{0}(\cdot, f)$ is an estimator of the classical central mean subspace $S_{E_{f}(Y)|X}$, as $f(Y)$ is a random number rather than a random object. We refer to $M_{0}$ as the ensemble candidate matrix, or, when confusion is possible, CMS-ensemble candidate matrix. Our goal is to construct a Fréchet candidate matrix $M : \mathcal{P}_{XY} \to \mathbb{R}^{p \times p}$ from the ensemble candidate matrix $M_{0} : \mathcal{P}_{XY} \times \mathcal{F} \to \mathbb{R}^{p \times p}$. To do so, we assume $\mathcal{F}$ is of the form $\{\kappa(\cdot, y) : y \in \Omega_{Y}\}$, where $\kappa : \Omega_{Y} \times \Omega_{Y} \to \mathbb{R}$ is a universal kernel. Given such an $\mathcal{F}$ and $M_{0}$, we define $M$ as follows

$$M(F_{XY}) = \int_{\Omega_{Y}} M_{0}(F_{XY}, \kappa(\cdot, y)) dF_{Y}(y),$$

where $F_{Y}$ is the distribution of $Y$ derived from $F_{XY}$. 

11
It is interesting to note the peculiar way in which $M_0$ depends on $F_{XY}$ and $f = \kappa(\cdot, y')$: if we let $T_{y'}$ denote the mapping $(x, y) \mapsto (x, \kappa(y, y'))$, then $M_0(F_{XY}, \kappa(\cdot, y'))$ is actually a function of the induced measure $F_{XY} \circ T_{y'}^{-1}$; that is, $M_0(F_{XY}, \kappa(\cdot, y')) = \tilde{M}_0(F_{XY} \circ T_{y'}^{-1})$ for some function $\tilde{M}_0$. Thus, an alternative way to write the Fréchet candidate matrix is $E[\tilde{M}_0(F_{XY} \circ T_{y'}^{-1})]$. In the random measure $F_{XY} \circ T_{y'}^{-1}$, the only random element is $Y'$.

We now adapt several estimates for the classical central mean subspace to the estimation of Fréchet SDR: the ordinary least squares (OLS; Li and Duan 1989), the principal Hessian directions (PHD; Li 1992), and the Iterative Hessian Transformation (IHT; Cook and Li 2002). These estimates are based on sample moments, and require additional conditions on the predictor $X$ for their unbiasedness. Specifically, we make the following assumptions:

1. $E(X|\beta^TX)$ is a linear function of $\beta^TX$, where $\beta$ is a basis matrix of the Fréchet central subspace $S_{Y|X}$;

2. $\text{var}(X|\beta^TX)$ is a nonrandom matrix.

Under the first assumption, the ensemble OLS and IHT are unbiased for estimating the Fréchet central subspace; under both assumptions, the ensemble PHD is unbiased for estimating $S_{Y|X}$. It is most convenient to construct these ensemble estimators using standardized directions (PHD; Li 1992), and the Iterative Hessian Transformation (IHT; Cook and Li 2002). The theoretical basis for doing so is an equivariant property of the Fréchet central subspace, as stated in the next proposition.

**Proposition 5.** If $S_{Y|X}$ is the Fréchet central subspace, $A \in \mathbb{R}^{p \times p}$ is a non-singular matrix, and $b$ is a vector in $\mathbb{R}^r$, then $S_{Y|AX+b} = A'S_{Y|X}$.

The proof is essentially the same as that for the classical central subspace (see, for example, Li 2018, page 24), and is omitted. Using this property, we first transform $X$ to $Z = \text{var}(X)^{-1/2}(X - EX)$, estimate the Fréchet central subspace $S_{Y|Z}$, and then transform it by $\text{var}(X)^{-1/2}S_{Y|Z}$, which is the same as $S_{Y|X}$.

For the ensemble OLS, let $C(y) = \text{cov}[Z, \kappa(Y, y)]$. Then the CMS-ensemble candidate matrix $M_o(F_{XY}, \kappa(\cdot, y))$ is $C(y)C(y)^T$. The Fréchet candidate matrix is $M(F_{XY}) = E[C(Y)C(Y)^T]$. For Fréchet PHD, let $H(y) = E\{[\kappa(Y, y) - EX(Y, y)]ZZ^T\}$. The candidate matrices $M_o$ and $M$ are, respectively, $H(y)$ and $EH(Y)$. For the ensemble IHT, let $r$ be a positive number, and $W(y) = (C(y), H(y)C(y), \ldots, H(y)^rC(y))$. The ensemble and Fréchet candidate matrices for Fréchet IHT are, respectively, $M_o(F_{XY}, \kappa(\cdot, y)) = W(y)W(y)^T$, $M(F_{XY}) = E[W(Y)W(Y)^T]$.

The sample estimates for ensemble OLS, PHD, and IHT can then be constructed by replacing the expectations in $M_o$ and $M$ with sample moments whenever possible. The
A detailed algorithm is given in Algorithm 1. In the algorithm, $\kappa_c(y, y')$ stands for the centered kernel $\kappa(y, y') - E_n \kappa(Y, y')$.

**Algorithm 1: Fréchet OLS, Fréchet PHD and Fréchet IHT**

**Step 1.** Standardize predictors. Compute sample mean $\hat{\mu} = E_n(X)$ and sample variance $\hat{\Sigma} = \text{var}_n(X)$. Then let $Z_i = \hat{\Sigma}^{-1/2}(X_i - \hat{\mu})$.

**Step 2.**

1. For Fréchet OLS: Compute $\hat{C}(y) = \text{cov}_n[Z, \kappa(Y, y)]$, $\hat{M}_0(y) = \hat{C}(y)\hat{C}(y)^T$ for $y = Y_1, \ldots, Y_n$;
2. For Fréchet PHD: Compute $\hat{M}_0(y) = E_n[ZZ^T \kappa_{c}(Y, y)]$ for $y = Y_1, \ldots, Y_n$;
3. For Fréchet IHT: Compute
   $$\hat{H}(y) = E_n[ZZ^T \kappa_{c}(Y, y)], \quad \hat{W}(y) = (\hat{C}(y), \hat{H}(y)\hat{C}(y), \ldots, \hat{H}(y)^r\hat{C}(y)), \quad \hat{M}_0(y) = \hat{W}(y)\hat{W}(y)^T, \quad \text{for } y = Y_1, \ldots, Y_n.$$

**Step 3.** Compute $\hat{M} = \frac{1}{n} \sum_{i=1}^n \hat{M}_0(Y_i)$.

**Step 4.** Let $\hat{v}_1, \ldots, \hat{v}_{d_0}$ be the leading $d_0$ eigenvectors of $\hat{M}$, and let $u_k = \hat{\Sigma}^{-1/2}\hat{v}_k$, for $k = 1, \ldots, d_0$. Then use $\{u_1, \ldots, u_{d_0}\}$ to estimate a basis of the Fréchet central subspace $S_{Y|X}$.

### 4.2 Ensembled forward regression

In this subsection we adapt the OPG and MAVE (Xia et al. 2002), two popular methods for estimating the classical central mean subspace based on nonparametric forward regression, to the estimation of the Fréchet central subspace. The framework of the statistical functional $M_0(F_{XY}, f)$ is no longer sufficient to cover the present case, because we now have a tuning parameter here. So, we adopt the notion of tuned statistical functional in (Li, 2018, Section 11.2) to accommodate a tuning parameter.

Let $\mathcal{P}_{XY}$, $F_{XY}$, $F^{(0)}_{XY}$ and $F^{(u)}_{XY}$ be as defined in Section 4.1. For simplicity, we assume the tuning parameter $h$ to be a scalar, but it could also be a vector. Given a CMS-ensemble $\mathfrak{F}$, let $T_0 : \mathcal{P}_{XY} \times \mathfrak{F} \times \mathbb{R} \to \mathbb{R}^{p \times p}$ be a tuned functional to be used as an estimator of $S_{E_{[f(Y)|X]}}$ for each $f$. We refer to $T_0$ as the ensemble tuned candidate matrix. The unbiasedness, exhaustiveness, and Fisher consistency of this type of augmented tuned statistical functional are defined as follows.
**Definition 2.** We say that $T_0$ is unbiased for estimating $\{S_{E[f_{(Y)}]} : f \in \mathcal{F}\}$ if, for each $f \in \mathcal{F}$, span$\{\lim_{h \to 0} T_0(f_{XY}, f, h)\} \subseteq S_{E[f_{(Y)}]}$. Exhaustiveness and Fisher consistency of $T_0$ are defined by replacing $\subseteq$ in the above by $\supseteq$ and $\equiv$, respectively.

Given $\mathcal{F} = \{\kappa(\cdot, y) : y \in \Omega_Y\}$ and $T_0$, we define the tuned Fréchet candidate matrix $T : \mathcal{P}_{XY} \times \mathbb{R} \to \mathbb{R}^{p \times p}$ as $T(F_{XY}, h) = \int_{\Omega_Y} T_0(F_{XY}, \kappa(\cdot, y), h)dF_Y(y)$. We say that the estimate $T(T_{XY}^{(0)}, h)$ is unbiased if span$\{\lim_{h \to 0} T(F_{XY}^{(0)}, h)\} \subseteq S_{Y|X}$, exhaustive if span$\{\lim_{h \to 0} T(F_{XY}^{(0)}, h)\} \supseteq S_{Y|X}$, and Fisher consistent if span$\{\lim_{h \to 0} T(F_{XY}^{(0)}, h)\} = S_{Y|X}$.

Using this general framework, we first adapt the OPG to Fréchet SDR. In the following, for a function $h(x)$, we use $\partial h(X)/\partial X$ to denote $\partial h(x)/\partial x$ evaluated at $x = X$. The OPG aims to estimate central mean subspace $S_{E[f_{(Y)}|X]}$ by $E\left[\frac{\partial E[\kappa(Y,Y)|X]}{\partial X} \cdot \frac{\partial E[\kappa(Y,Y)|X]}{\partial X}\right]$ where the gradient $\partial E(\kappa(Y,Y)|X)/\partial X$ is estimated by local linear approximation as follows. Let $K_0 : \mathbb{R} \to [0, \infty)$ be a kernel function as used in kernel estimation. For any $v \in \mathbb{R}^p$ and bandwidth $h > 0$, let $K_h(v) = h^{-p}K_0(||v||)/h$. At the population level, for fixed $x \in \Omega_X$ and $y \in \Omega_Y$, we minimize the objective function

$$E\{[\kappa(Y,y) - a - b^T(X-x)]^2 K_h(X-x)\} / EK_h(X-x)$$

over all $a \in \mathbb{R}$ and $b \in \mathbb{R}^{d_0}$. Here, the presence of $E K_h(X-x)$ has no effect on optimization, but we keep it there to be consistent with the later development of MAVE. The minimizer depends on $x, y$ and we write it as $(a_h(x, y), b_h(x, y))$. The ensemble tuned candidate matrix for estimating the central mean subspace $S_{E[\kappa(Y,y)|X]}$ is $T_0(F_{XY}, \kappa(\cdot, y), h) = E[b_h(X, y)b_h(X, y)^T]$ and the tuned Fréchet candidate matrix is $T(F_{XY}, h) = E[b_h(X, Y)b_h(X, Y)^T]$.

At the sample level, we use $T(F_{XY}, h_n)$ to estimate $S_{Y|X}$ where $h_n \downarrow 0$ is the bandwidth parameter. This is realized through the following steps. For simplicity, we abbreviate $h_n$ as $h$. Suppose we are given an i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $(X, Y)$. We minimize, for each $j, k = 1, \ldots, n$, the empirical objective function

$$\sum_{i=1}^n w_h(X_i, X_j) \left[\kappa_h(Y_i, Y_k) - a_{jk} - b^T_{jk}(X_i - X_j)\right]^2$$

over $a_{jk} \in \mathbb{R}$ and $b_{jk} \in \mathbb{R}^p$, where $w_h(X_i, X_j) = K_h(X_i - X_j)/\sum_{i=1}^n K_h(X_i - X_j)$. Following Xia et al. [2002], we take the bandwidth to be $h = c_0 n^{1/(p_0+6)}$ where $p_0 = \max\{p, 3\}$ and $c_0 = 2.34$, which is slightly larger than the optimal $n^{-1/(p+4)}$ in terms of the mean integrated squared errors. As proposed in Li [2018] Lemma 11.6, instead of solving $b_{jk}$ from $n^2$ times, we solve the following multivariate weighted least squares problems to obtain $b_{11}, \ldots, b_{in}$ simultaneously. Let $\Gamma \in \mathbb{R}^{p \times m}$, where $m \leq p$, be a matrix to be specified later in [8] and
Algorithm 1. Let
\[
\kappa_. = \{\kappa_j(Y_j, Y_k)\}_{j,k=1}^n, \quad A_j = \begin{pmatrix} a_{j1}, \ldots, a_{jm} \\ b_{j1}, \ldots, b_{jm} \end{pmatrix},
\]
\[
W_j(h, \Gamma) = \text{diag}(w_h(\Gamma^T X_j, \Gamma^T X_1), \ldots, w_h(\Gamma^T X_j, \Gamma^T X_n)), \quad \Delta_i = (X_1 - X_i, \ldots, X_n - X_i)^T.
\]

Then, the \(n^2\) weighted least squares optimization problems in (6) can be rewritten as \(n\) multivariate least squares problems with matrix-valued objective function
\[
[\kappa_. - (1_n, \Delta_. j)A_j]^T W_j(h, I_p)[\kappa_. - (1_n, \Delta_. j)A_j],
\]
which has the following minimizer over \(A_j \in \mathbb{R}^{(p+1) \times n}\) in terms of Louwner’s ordering:
\[
\hat{A}_j = [(1_n, \Delta_. j)^T W_j(h, I_p)(1_n, \Delta_. j)]^{-1}(1_n, \Delta_. j)^T W_j(h, I_p)\kappa_.
\]
The estimates \(\hat{b}_{j1}, \ldots, \hat{b}_{jm}\) are simply the bottom \(p\) rows of \(\hat{A}_j\). We perform eigen-decomposition on \(\sum_{j,k} \hat{b}_{jk} \hat{b}^T_{jk}\), and use the first \(d_0\) eigenvectors, \(\hat{v}_1, \ldots, \hat{v}_{d_0}\), to estimate the Fréchet central subspace. Dimension \(d_0\) of the Fréchet central subspace is assumed to be known in this paper. We also need to standardize \(X_i\) marginally so as to set the same baseline for determining the bandwidth for different components of \(X\). We call this procedure the Fréchet Outer Product of Gradients, or FOPG.

We can further enhance the performance of FOPG by projecting the original predictors onto the directions produced by the FOPG to re-estimate \(S_{Y|X}\). Specifically, after the first round of FOPG, we form the matrix \(\hat{B} = (\hat{v}_1, \ldots, \hat{v}_{d_0})\) and replace the kernel \(K_h(X_j - X_i)\) by \(K_h(\hat{B}^T(X_j - X_i))\) with an updated bandwidth \(h\), and complete the next round of iteration, which leads to an updated \(\hat{B}\). We then iterate this process until convergence. In this way, we reduce the dimension of the kernel from \(p\) to \(d_0\), and mitigates the “curse of dimensionality”.

We call this procedure refined Fréchet OPG, or rFOPG. The algorithms for FOPG and rFOPG are summarized as Algorithm 2, where the FOPG algorithm is simply the steps 1, 2, 3 of the first iteration of Algorithm 2.

We then adapt MAVE [Xia et al. 2002] to Fréchet SDR. Recall that, for FOPG, at the population level, we minimize the objective function (5) with respect to \(a, b\) for each \((x, y)\), which gives the minimizer \(b(x, y)\). Since, for a small \(h\), \(b(x, y)\) roughly belongs to the central mean space \(S_{E(\kappa(Y,y)|X)} \subseteq S_{Y|X}\) for any \(x \in \Omega_x\) and \(y \in \Omega_y\), we proactively model \(b(x, y)\) by \(Bc(x, y)\), where \(B\) is a \(p \times d_0\) matrix, resulting in the following population-level objective function
\[
\int_{\Omega_y} \int_{\Omega_x} E \left\{ [\kappa(Y, y) - a(x, y) - c(x, y)^T B^T (X - x)]^2 K_h(X - x) \right\} dF_X(x) dF_Y(y),
\]
The optimization can be broken down into iterations between two steps. We can also refine FMAVE as we did for FOPG; that is, we use an existing consistent estimate of a basis of the Fréchet central subspace. This estimator is called the Fréchet MAVE, or FMAVE.

At the sample level, we minimize the empirical objective function

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} w_h(X_i, X_j) \left[ \kappa(Y_i, Y_k) - a_{jk} - c_{jk}^T B(X_i - X_j) \right]^2
\]

over \(a_{jk} \in \mathbb{R}, c_{jk} \in \mathbb{R}^{d_0}, \text{ and } B \in \mathbb{R}^{p \times d_0}\), where the weights \(w_h(X_i, X_j)\) and the bandwidth \(h\) are the same as in the FOPG algorithm. The optimal \(\hat{B}\) is then taken as the estimate of a basis of the Fréchet central subspace. This estimator is called the Fréchet MAVE, or FMAVE.

We can also refine FMAVE as we did for FOPG; that is, we use an existing consistent estimate of \(B\) to reduce the dimension of the kernel function, so that smoothing is carried out over a \(d_0\)-dimensional, rather than a \(p\)-dimensional subspace. We call this estimate the refined FMAVE, or rFMAVE. The optimization can be broken down into iterations between two steps.
steps, each of which is a quadratic optimization problem with explicit solution. The detailed algorithm for FMAVE and rFMAVE are given in Algorithm 3. The FMAVE is just rFMAVE with \( w_j(h, \hat{B}^{(t-1)}) \) replaced by \( W_j(h, I_p) \).

**Algorithm 3: FMAVE and rFMAVE**

**Step 1.** Marginally standardize \( X_1, \ldots, X_n \) as in Algorithm 1 and set the initial bandwidth \( h_0 \) as in step 1 of Algorithm 1. Use FOPG to compute the initial \( \hat{B}^{(0)} \). Set iteration time \( t = 1 \).

**Step 2.** Compute \( \kappa_-, \ W_j(h, \hat{B}^{(t-1)}) \) and \( \Delta_j(\hat{B}^{(t-1)}) \) according to (7) for each \( j = 1, \ldots, n \). Then compute

\[
\hat{G}^{(t)}_j = [(1, \Delta_j(\hat{B}^{(t-1)})) W_j(h_{t-1}, \hat{B}^{(t-1)})(1, \Delta_j(\hat{B}^{(t-1)}))]^{-1}
\times (1, \Delta_j(\hat{B}^{(t-1)})) W_j(h_{t-1}, \hat{B}^{(t-1)}) \kappa_-
\]

Read off \( \hat{a}_{j1}^{(t)}, \ldots, \hat{a}_{jn}^{(t)} \) from the first row of the \((p + 1) \times n\) matrix \( \hat{G}^{(t)}_j \), and \( \hat{c}_{j1}^{(t)}, \ldots, \hat{c}_{jn}^{(t)} \) from the bottom \( p \) rows of this matrix.

**Step 3.** Compute

\[
L_{jk}(h_{t-1}, \hat{B}^{(t-1)}) = \left( w_{h_{t-1}}((\hat{B}^{(t-1)})^T X_j, (\hat{B}^{(t-1)})^T X_k)[\kappa(Y_1, Y_k) - \hat{a}_{jk}^{(t)}], \ldots, w_{h_{t-1}}((\hat{B}^{(t-1)})^T X_j, (\hat{B}^{(t-1)})^T X_k)[\kappa(Y_n, Y_k) - \hat{a}_{jk}^{(t)}]\right)^T.
\]

Then compute \( \hat{B}^{(t)} \) by

\[
\text{vec}(\hat{B}^{(t)}) = \left[ \sum_{j,k=1}^n (\Delta_j^T W_j(h_{t-1}, \hat{B}^{(t-1)}) \Delta_j) \otimes c_{jk}^{(t)} c_{jk}^{(t)T} \right]^{-1}
\times \left[ \sum_{j,k=1}^n (\Delta_j^T \otimes c_{jk}^{(t)}) L_{jk}(h_{t-1}, \hat{B}^{(t-1)}) \right]
\]

and set \( t = t + 1 \).

**Step 4.** If \( t \leq 10 \), reset \( h_t = \max(r_n h_{t-1}; c_0 n^{-1/(d+4)}) \), where \( r_n = n^{-1/2(p_0 + 6)} \), increase \( t \) by 1 and return to step 2. Otherwise, Let \( \hat{D} \) be the diagonal matrix \( \text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_p) \).

A basis of the central subspace \( S_{Y|X} \) is estimated by \( \hat{D}^{-1/2} \hat{B}^{(t)} \).

### 4.3 Fisher consistency

In this subsection, we establish the unbiasedness and Fisher consistency of the tuned Fréchet candidate matrix. As a special case, the Fréchet candidate matrix constructed by any moment based methods in Section 4.1 can be considered as tuned Fréchet candidate matrix.
with the tuning parameter $h$ taken to be 0. The next theorem shows that if $T_o$ is unbiased (or Fisher consistent), then $T$ is unbiased (or Fisher consistent). In the following, we say that a measure $\mu$ on $\Omega_Y$ is strictly positive if and only if for any nonempty open set $U \subseteq \Omega_Y$, $\mu(U) > 0$. For a matrix $A$, $\|A\|$ represents the operator norm.

**Theorem 2.** Suppose $\mathfrak{F} = \{\kappa(\cdot, y) : y \in \Omega_y\}$ is a CMS-ensemble, where $\kappa$ is a universal kernel. We have the following results regarding unbiasedness and Fisher consistency for $T$.

1. If $T_o$ is unbiased for $\{S_{E[\kappa(Y,y)|X]} : f \in \mathfrak{F}\}$ and $\|T_o(F_{XY}(0), \kappa(\cdot, Y'), h)\| \leq G(Y')$, where $G(Y')$ is a real-valued function with $E[G(Y')] < \infty$, then $T$ is unbiased for $S_{Y|X}$;

2. If (a) $T_o$ is Fisher consistent for $\{S_{E[\kappa(Y,y)|X]} : f \in \mathfrak{F}\}$, (b) $T_o(F_{XY}(0), \kappa(\cdot, y), h)$ is positive semidefinite for each $y \in \Omega_y$, $h \in \mathbb{R}$ and $F_{XY} \in \mathcal{P}_{XY}$, (c) $\|T_o(F_{XY}(0), \kappa(\cdot, Y'), h)\| \leq G(Y')$ with $E[G(Y')] < \infty$, (d) $F_Y$ is strictly positive on $\Omega_Y$, and (e) the mapping $y' \mapsto \lim_{h \to 0} T_o(F_{XY}(0), \kappa(\cdot, y'), h)$ is continuous, then $T$ is Fisher consistent for $S_{Y|X}$.

5 Fréchet SDR via CS-ensemble

In this section we adapt several well known estimators for the classical central subspace to Fréchet SDR, which include SIR (Li, 1991), SAVE (Cook and Weisberg, 1991), and DR (Li and Wang, 2007). We use the CS-ensemble to combine these classical estimates through (3).

5.1 Ensemble SIR, SAVE, and DR for Fréchet SDR

SIR is one of the most well-known SDR methods for estimating the classical central subspace. Again, we work with the standard predictor $Z$ as we adapt SIR to Fréchet SDR. The CS-ensemble candidate matrix $M_o$ is $M_o(F_{ZY}(0), \kappa(\cdot, y)) = \text{var}[E(Z|\kappa(Y,y))] \equiv M_{\text{SIR}}(y)$. The Fréchet candidate matrix is $M(F_{ZY}) = EM_{\text{SIR}}(Y)$. For SAVE, the CS-ensemble candidate matrix is $M_o(F_{ZY}(0), \kappa(\cdot, y)) = [I_p - \text{var}(Z|\kappa(Y,y))]^2 \equiv M_{\text{SAVE}}(y)$, and the Fréchet candidate matrix $M(F_{ZY}) = EM_{\text{SAVE}}(Y)$. For DR,

$$M_o(F_{ZY}(0), \kappa(\cdot, y)) = 2E\{E[Z^T|\kappa(Y,y)]\} + 2E^2\{E[Z|\kappa(Y,y)]E[Z^T|\kappa(Y,y)]\}$$

$$+ 2E\{E[Z^T|\kappa(Y,y)]E[Z|\kappa(Y,y)]\} E\{E[Z|\kappa(Y,y)]E[Z^T|\kappa(Y,y)]\} - 2I_p$$

$$\equiv M_{\text{DR}}(y),$$

and $M(F_{ZY}) = EM_{\text{DR}}(Y)$. At the sample level, we replace any unconditional moment $E$ by the sample average $E_n$, and replace any conditional moment, such as $E(Z|\kappa(Y,y))$, by the
slice mean. The detailed algorithm is given by Algorithm 4, which involves the following notation. For each $i = 1, \ldots, n$, we use $J_i$ to denote the interval $[\min(\kappa(Y_1, Y_i), \ldots, \kappa(Y_n, Y_i)), \ldots, \max(\kappa(Y_1, Y_i), \ldots, \kappa(Y_n, Y_i))]$. We use $\{J_{i\ell} : \ell = 1, \ldots, s\}$ to denote a set of intervals that partition $J_i$.

**Algorithm 4:** Fréchet SIR, Fréchet SAVE, and Fréchet DR

1. **Step 1.** Standardization: compute $\hat{\mu} = E_n(X)$, $\hat{\Sigma} = \text{var}_n(X)$, and let $Z_i = \hat{\Sigma}^{1/2}(X_i - \hat{\mu})$.

2. **Step 2.** Preliminary moments: for each $i = 1, \ldots, n$, $\ell = 1, \ldots, s$, compute
   
   
   $A_i = n^{-1} \sum_{j=1}^{n} I(\kappa(Y_j, Y_i) \in J_{i\ell})$,  
   $B_i = n^{-1} \sum_{j=1}^{n} Z_j I(\kappa(Y_j, Y_i) \in J_{i\ell})$,  
   $C_i = n^{-1} \sum_{j=1}^{n} Z_j Z_j^T I(\kappa(Y_j, Y_i) \in J_{i\ell})$.

3. **Step 3.** Conditional moments: compute
   
   $\hat{E}_i(Z) = B_i / A_i$,  
   $\hat{E}_i(Z Z^T) = C_i / A_i$,  
   $\hat{V}_i(Z) = \hat{E}_i(Z Z^T) - \hat{E}_i(Z) \hat{E}_i(Z^T)$.

4. **Step 4.** Fréchet candidate matrices:

   (1) For Fréchet SIR, compute $\hat{M}_{\text{SIR}}(Y_i) = s^{-1} \sum_{\ell=1}^{s} A_{i\ell} \hat{E}_i(Z) \hat{E}_i(Z^T)$ for $i = 1, \ldots, n$, and $\hat{M} = E_n \hat{M}_{\text{SIR}}(Y)$.

   (2) For Fréchet SAVE, compute $\hat{M}_{\text{SAVE}}(Y_i) = s^{-1} \sum_{\ell=1}^{s} A_{i\ell} [I_p - \hat{V}_i(Z)]^2$ for $i = 1, \ldots, n$, and $\hat{M} = E_n \hat{M}_{\text{SAVE}}(Y)$.

   (3) For Fréchet DR, compute

   
   $M_{\text{DR}}(Y_i) = 2s^{-1} \sum_{\ell=1}^{s} A_{i\ell} \hat{E}_i(Z Z^T) + 2 \left(s^{-1} \sum_{\ell=1}^{s} A_{i\ell} \hat{E}_i(Z) \hat{E}_i(Z^T)\right)^2$

   
   $+ 2 \left(s^{-1} \sum_{\ell=1}^{s} A_{i\ell} \hat{E}_i(Z^T) \hat{E}_i(Z)\right) \left(s^{-1} \sum_{\ell=1}^{s} A_{i\ell} \hat{E}_i(Z) \hat{E}_i(Z^T)\right) - 2 I_p$

   
   for $i = 1, \ldots, n$, and $\hat{M} = E_n \hat{M}_{\text{DR}}(Y)$.

5. **Step 5.** Fréchet SDR: compute the first $d_0$ eigenvectors $\hat{v}_1, \ldots, \hat{v}_{d_0}$ of $\hat{M}$, and use $\hat{\Sigma}^{-1/2} \hat{v}_1, \ldots, \hat{\Sigma}^{-1/2} \hat{v}_{d_0}$ as the estimate of a basis of the Fréchet central subspace $S_{Y|X}$. 
5.2 Fisher consistency

We next develop Fisher consistency for Fréchet SDR based the CS-ensemble, which is parallel to the development of Section 4.3. Let \( \mathcal{F} = \{ \kappa(\cdot, y) : y \in \Omega_Y \} \) be a CS ensemble, where \( \kappa \) is a universal kernel. Let \( M_0 : \mathcal{P}_{XY} \times \mathcal{F} \rightarrow \mathbb{R}^{p \times p} \) be a CS-ensemble candidate matrix. Let \( M(\mathcal{F}_{XY}) = \int M_0(\mathcal{F}_{XY}, \kappa(\cdot, y))dF_Y(y) \) be the Fréchet candidate matrix. The next corollary says that if \( M_0 \) is Fréchet consistent for \( \{ \mathcal{S}_{\kappa(\cdot, y)} : f \in \mathcal{F} \} \), then \( M \) is Fréchet consistent for \( \mathcal{S}_{Y|X} \). The proof is similar to that of Theorem 2 and is omitted.

**Corollary 3.** Suppose \( \mathcal{F} = \{ \kappa(\cdot, y) : y \in \Omega_Y \} \) is a CS-ensemble, where \( \kappa \) is a universal kernel. We have the following results regarding unbiasedness and Fisher consistency for \( M \).

1. If \( M_0 \) is unbiased for \( \{ \mathcal{S}_{\kappa(\cdot, y)} : f \in \mathcal{F} \} \), then \( M \) is unbiased for \( \mathcal{S}_{Y|X} \);
2. If \( M_0 \) is Fisher consistent for \( \{ \mathcal{S}_{\kappa(\cdot, y)} : f \in \mathcal{F} \} \), \( M_0(\mathcal{F}_{XY}, \kappa(\cdot, y)) \) is positive semidefinite for each \( y \in \Omega_X \) and \( F_{XY} \in \mathcal{P}_{XY} \), \( F_Y \) is strictly positive, and the mapping \( y' \mapsto M_0(\mathcal{F}_{XY}, \kappa(\cdot, y')) \) is continuous, then \( M \) is Fisher consistent for \( \mathcal{S}_{Y|X} \).

6 Convergence Rates of the Ensemble Estimates

In this section we develop the convergence rates of the ensemble estimates for Fréchet SDR. To save space we will only consider the CMS-ensemble; the results for the CS-ensemble are largely parallel. To simplify the asymptotic development we make a slight modification of the ensemble estimator, which does not result in any significant numerical difference from the original ensembles developed in the previous sections. For each \( i = 1, \ldots, n \), let \( F_{XY}^{(n)} \) be the empirical distribution based on the sample with \( i \)-th subject removed: \( \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \setminus \{(X_i, Y_i)\} \). Our modified ensemble estimate is of the form

\[
T(F_{XY}^{(n)}, h_n) = n^{-1} \sum_{i=1}^{n} T_0(F_{XY}^{(i)}, \kappa(\cdot, Y_i), h_n).
\]

The purpose of this modification is to break the dependence between the ensemble member \( \kappa(\cdot, Y_i) \) and the CMS estimate, which substantially simplifies the asymptotic argument. Here, we let the tuning parameter \( h_n \) depend on \( n \). Again, the Fréchet candidate matrix constructed by moment based methods can be considered as a special case with \( h_n = 0 \).

Rather than deriving the convergence rate of each individual ensemble estimate case by case, we will show that, under some mild conditions, the ensemble convergence rate is the
same as the corresponding CMS-estimate’s rate. Since the convergence rates of many CMS-estimates are well established, including all the forward regression and sample moment based estimates mentioned earlier, our general result covers all the CMS-ensemble estimates.

In this following, for a matrix $A$, $\|A\|$ represents the operator norm and $\|A\|_P$ the Frobenius norm. If $\{a_n\}$ and $\{b_n\}$ are sequences of positive numbers, we write $a_n \prec b_n$ if $\lim_{n \to \infty} a_n/b_n = 0$; we write $a_n \preceq b_n$ if $a_n/b_n$ is a bounded sequence. We write $b_n \succ a_n$ (or $b_n \succeq a_n$) if $a_n/b_n$ is a nonnegative sequence. We write $a_n \succ b_n$ if $a_n \preceq b_n$ and $b_n \preceq a_n$. Let $T_0^*(F^{(0)}_{XY}, \kappa(\cdot, y)) = \lim_{h \to 0} T_0(F^{(0)}_{XY}, \kappa(\cdot, y), h)$ and $T^*(F^{(0)}_{XY}) = \lim_{h \to 0} T(F^{(0)}_{XY}, h)$.

**Theorem 3.** Let $C_n(y) = E\|T_0(F^{(n)}_{XY}, \kappa(\cdot, y), h_n) - T^*_0(F^{(0)}_{XY}, \kappa(\cdot, y))\|$ and $\{a_n\}$ be a positive sequence of numbers satisfying $a_{n+1}/a_n \succ 1$ and $a_n \succeq n^{-1/2}$. Suppose the entries of $T_0^*(F^{(0)}_{XY}, \kappa(\cdot, Y))$ have finite variances. If $E[C_n(Y)] = O(a_n)$, then $\|T(F^{(n)}_{XY}, h_n) - T^*(F^{(0)}_{XY})\| = O_P(a_n)$.

The above theorem says that, under some conditions, the convergence rate of an ensemble Fréchet SDR estimator is the same as the corresponding CMS estimator. This covers all the estimators developed in Section 4 and 5. Specifically:

1. For all moment-based ensemble methods, such as OLS, PHD, IHT, SIR, SAVE, DR, the ensemble candidate matrices can be written in the form $M_0(F^{(n)}_{XY}, \kappa(\cdot, y)) = \hat{\Lambda}(y)\hat{\Lambda}(y)^T$, where $\hat{\Lambda}(y)$ is a matrix possessing the second order von Mises expansion, implying $E[C_n(Y)] = O(n^{-1/2})$. See, for example, Li (2018).

2. For nonparametric forward regression ensemble methods, OPG and MAVE, the convergence rate of $C_n(y)$ was reported in Xia (2007) as $O(h_n^2 + h_n^{-1}\delta_n^2)$ where $\delta_n = \sqrt{(\log n)/nh_n^2}$. Although the convergence was established in terms of convergence in probability, under mild conditions such as uniformly integrability, we can obtain the same rate for $E[C_n(Y)]$.

### 7 Simulations

We evaluate the performance of the proposed Fréchet SDR methods with distributions, symmetric positive definite matrices, and spherical data as responses. For both the distributional response and symmetric positive definite matrices response, we use the Gaussian radial basis function $\kappa_\gamma(y, y') = \exp(-\gamma d(y, y')^2)$ to construct ensembles. For the spherical response, we use the Laplacian radial basis function $\kappa_\gamma(y, y') = \exp(-\gamma d(y, y'))$ to construct the ensemble. We use $K_0(x) = (2\pi)^{-1/2} \exp(-\|x\|_2^2/2)$ as the kernel for rFOPG and rFMAVE. We take the
7.1 Scenario I: Fréchet SDR for distributions

Let \((\Omega, d_W)\) be the metric space of univariate distributions endowed with Wasserstein metric \(d_W\), as described in Section 3. The construction of the ensembles requires computing the Wasserstein distances \(d_W(Y_i, Y_j)\) for \(i, j = 1, \ldots, n\). However, the distributions \(Y_1, \ldots, Y_n\) are usually not fully observed in practice, which means we need to estimate them in the implementation of the proposed methods. There are multiple ways to do so, such as by estimating the c.d.f.’s, the quantile functions (Parzen, 1979), or the p.d.f.’s (Petersen and Müller, 2016; Chen et al., 2020). For computation simplicity, we use the Wasserstein distances between the c.d.f.’s, the quantile functions (Parzen, 1979), or the p.d.f’s (Petersen and Müller, 2016; Chen et al., 2020). For computation simplicity, we use the Wasserstein distances between the empirical measures. Specifically, suppose we observe \((X_1, \{W_{ij}\}_{j=1}^{m_i}), \ldots, (X_n, \{W_{nj}\}_{j=1}^{m_n})\), where \(\{W_{ij}\}_{j=1}^{m_i}\) are independent samples from the distribution \(Y_i\). Let \(\hat{Y}_i\) be the empirical measure \(m_i^{-1} \sum_{j=1}^{m_i} \delta_{W_{ij}}\), where \(\delta_a\) is the Dirac measure at \(a\), then we estimate \(d_W(Y_i, Y_k)\) by \(d_W(\hat{Y}_i, \hat{Y}_k)\). For the theoretical justification, see Fournier and Guillin (2015) and Lei (2020). For simplicity, we assume the sample sizes \(m_1, \ldots, m_n\) to be the same and denote the common sample size by \(m\). Then the distance between empirical measures \(\hat{Y}_i\) and \(\hat{Y}_i\) is a simple function of the order statistics: \(d_W(\hat{Y}_i, \hat{Y}_k) = \left(\sum_{j=1}^{m}(W_{i(j)} - W_{k(j)})^2\right)^{1/2}\), where \(W_{i(j)}\) is the \(j\)-th order statistics of the sample \(W_{i1}, \ldots, W_{im}\).

To generate univariate distributional responses \(Y\), we add noise to the mean \(\mu_Y\) and variance \(\sigma_Y\) conditioning on \(X\). Specifically, we generate random normal distribution \(Y\) with mean and variance parameters as random variables dependent on \(X\), that is,

\[
Y = N(\mu_Y, \sigma_Y^2),
\]

where \(\sigma_Y > 0\) almost surely. Then, for each \(Y_i, i = 1, \ldots, n\), we further generate samples \(W_{i1}, \ldots, W_{im}\) from the normal distribution \(Y_i\).

Let \(\beta_1^T = (1, 1, 0, \ldots, 0) / \sqrt{2}, \beta_2^T = (0, \ldots, 0, 1, 1) / \sqrt{2}, \beta_3^T = (1, 2, 0, \ldots, 0, 2) / 3\) and \(\beta_4^T = (0, 0, 3, 4, 0, \ldots, 0) / 5\). We generate i.i.d. observations \(X_1, \ldots, X_n\) from a \(p\)-dimensional random vector whose components are independent \(U[0, 1]\) random variables. We then generate \(Y\) according to (10) with the following models:

**I-1:** \(\mu_Y|X \sim N(\exp(\beta_1^T X), 0.5^2)\) and \(\sigma_Y = 1\).

**I-2:** \(\mu_Y|X \sim N(\exp(\beta_1^T X), 0.5^2)\) and \(\sigma_Y|X = \exp(\beta_2^T X)\).

**I-3:** \(\mu_Y|X \sim N(0, |\beta_1^T X|)\) and \(\sigma_Y = |\mu_Y|\).
I-4: \( \mu_Y|X \sim N(0, (\beta_4^T X)^4) \) and \( \sigma_Y|X = (\beta_4^T X)^4 \).

Ying and Yu (2020) considered similar models to Model I-1 and Model I-2. In Model I-3 and Model I-4, the error depends on \( X \), which means the Fréchet central subspace contains direction out of the conditional Fréchet mean function. For Model I-1, \( B_0 = \beta_1 \) and \( d_0 = 1 \); for Model I-2, \( B_0 = (\beta_1, \beta_2) \) and \( d_0 = 2 \); and for Models I-3, I-4, \( B_0 = (\beta_3, \beta_4) \) and \( d_0 = 2 \).

We compare performances of the CMS ensemble methods and CS ensemble methods described in Sections 4 and 5, including rFOPG, rFMAVE, FOLS, FPHD, FIHT, FSIR, FSAVE and FDR, under Models I-1 to I-4. Here, we take \((n,p) = (200,10), (200,20), (400,30)\) and \( m = 100 \). For rFOPG and rFMAVE, the iteration time is set as 10, which is large enough to guarantee numerical convergence. For FSIR, FSAVE, the number of slices is chosen as \( \lfloor n/2p \rfloor \); for FDR, the number of slices is chosen as \( \lfloor n/50 \rfloor \). We also implement the weighted inverse regression ensemble (WIRE) method proposed by Ying and Yu (2020) for comparison. To facilitate the comparison, we compare the benchmark distance, which is set as the distance between two uniformly randomly generated space. We repeat the experiments 100 times and summarize the mean and standard deviation of estimation error in Table I. The benchmark distances are shown at the bottom of the table. A smaller distance indicates a more accurate estimate, and the estimate with the smallest distance is shown in boldface.

For Model I-1 and Model I-2, we note that the best performer is always rFMAVE. The moment based ensemble methods are slightly less accurate than rFOPG and rFMAVE, since they are affected by the marginal distribution of predictor \( X \). Compared with the benchmark, all methods listed in the table can successfully estimate the true central subspace except FPHD and FSAVE. In Model I-3 and Model I-4, there exist directions in central subspace but not contained in the conditional Fréchet mean. In these cases, all ensemble methods except FPHD and FSAVE achieve the successful estimation. The best performer is either FSIR, rFOPG or WIRE.

We also show the plots of \( Y \) versus the sufficient predictors obtained by rFOPG for Model I-2. From Figure 2 we see a strong relation between \( Y \) and the first two estimated sufficient predictors \( \hat{\beta}_1^T X \) and \( \hat{\beta}_2^T X \), compared with \( Y \) versus individual predictor \( X_3 \).

### 7.2 Scenario II: Fréchet SDR for positive definite matrices

Let \( \Omega_Y \) be the space Sym\(^+(r) \) endowed with Frobenius distance \( d_F(Y_1, Y_2) = \|Y_1 - Y_2\|_F \). To accommodate the anatomical intersubject variability, Schwartzman (2006) introduces the symmetric matrix variate Normal distributions. We adopt this distribution to construct the regression model with correlation matrices response. We say that \( Z \in \text{Sym}(r) \) has the
Table 1: Mean(± standard deviation) of estimation error as measured by \( \|P_{B_0} - P_B\|_F \) for different methods for Scenario I; the benchmark for Model I-1 with \( p = 10, 20, 30 \) are 1.337, 1.380, 1.394 respectively, for Model I-2,3,4 with \( p = 10, 20, 30 \) are 1.763, 1.892, 1.934, respectively. Bold-faced number indicates the best performer.

Standard symmetric matrix variate Normal distribution \( N_{rr}(0; I_r) \) if it has density \( \varphi(Z) = (2\pi)^{-q/2} \exp(-\text{tr}(Z)^2/2) \) with respect to Lebesgue measure on \( \mathbb{R}^{p(p+1)} \). As pointed out in \cite{Schwartzman2006}, this definition is equivalent to a symmetric matrix with independent \( N(0, 1) \) diagonal elements and \( N(0, 1/2) \) off-diagonal elements. We say \( Y \in \text{Sym}(r) \) has symmetric matrix variate Normal distribution \( N_{rr}(M; \Sigma) \) if \( Y = GZG^T + M \) where \( M \in \text{Sym}(r), \ G \in \mathbb{R}(r \times r) \) is a non-singular matrix, and \( \Sigma = G^TG \). As a special case, we say \( Y \in \text{Sym}(r) \sim N_{rr}(M; \sigma^2) \) if \( Y = \sigma Z + M \).

Let \( \beta_1^T = (1, 1, 0, \ldots, 0)/\sqrt{2}, \ \beta_2^T = (0, \ldots, 0, 1, 1)/\sqrt{2} \). We generate i.i.d observations \( X_1, \ldots, X_n \) from a \( p \)-dimensional random vector whose components are independent \( U[0, 1] \).
random variables. Let \( D(X) \) be matrices specified by the following models:

**II-1:**
\[
D(X) = \begin{pmatrix} 1 & \rho(X) \\ \rho(X) & 1 \end{pmatrix}, \text{ where } \rho(X) = [\exp(\beta_1^T X) - 1]/[\exp(\beta_1^T X) + 1].
\]

**II-2:**
\[
D(X) = \begin{pmatrix} 1 & \rho_1(X) & \rho_2(X) \\ \rho_1(X) & 1 & \rho_1(X) \\ \rho_2(X) & \rho_1(X) & 1 \end{pmatrix},
\]
where \( \rho_1(X) = 0.4[\exp(\beta_1^T X) - 1]/[\exp(\beta_1^T X) + 1] \) and \( \rho_2(X) = 0.4 \sin(\beta_3^T X) \).

We generate \( \log(Y) \) following \( N_{dd}(\log\{D(X)\}, 0.25) \), where \( \log(\cdot) \) is the matrix logarithm defined in Section 3. In Model II-1, \( B_0 = \beta_1 \) and \( d_0 = 1 \); in Model II-2, \( B_0 = (\beta_1, \beta_2) \) and \( d_0 = 2 \). We note that \( D(x) \) is not necessarily the Fréchet conditional mean of \( Y \) given \( X \), but still measures the central tendency of the conditional distribution \( Y|X \). We also compare performances of the CMS ensemble methods and CS ensemble methods, including rFOPG, rFMAVE, FOLS, FPHD, FIHT, FSIR, FSAVE and FDR, with \((n, p) = (100, 10), (200, 10), (100, 20), (200, 20)\). The experiments are repeated 100 times and the means and standard deviations of estimation errors are summarized in Table 2.

We conclude that all ensemble methods give reasonable estimation except FPHD and FSAVE. The best performer is rFOPG in all settings. To illustrate the relation between the response and estimated sufficient predictor \( \hat{\beta}_1^T X \), we adopt the ellipsoidal representation of SPD matrices. Each \( A \in \text{Sym}^+(d) \) can be associated with an ellipsoid centered at the origin \( E_A = \{x : x^T A^{-1} x \leq 1\} \). In Figure 3, we plot the responses ellipsoid versus the estimated sufficient predictor in panel (a), compared with the responses versus predictor \( X_{10} \). We can tell a clear pattern of change in shape and rotation of response ellipsoids versus \( \hat{\beta}_1^T X \).
Table 2: Mean(± standard deviation) of estimation error measured by $\|P_B - \hat{P}_B\|_F$ for different methods for Scenario II. Bold-faced number indicates the best performer.

| Model   | $(p, n)$ | rFOPG | rFMAVE | FOLS | FPHD | FIHT | FSIR | FSAVE | FDR  |
|---------|----------|-------|--------|------|------|------|------|-------|------|
| II-1    | (10, 100)| 0.128 | 0.132  | 0.173| 1.405| 0.173| 0.210| 0.656 | 0.212|
|         |          | ±0.035| ±0.038 | ±0.051| ±0.016| ±0.051| ±0.052| ±0.377| ±0.062|
|         | (10, 200)| 0.091 | 0.094  | 0.127| 1.410| 0.127| 0.130| 0.342 | 0.124|
|         |          | ±0.022| ±0.023 | ±0.030| ±0.006| ±0.030| ±0.032| ±0.198| ±0.031|
|         | (20, 100)| 0.212 | 0.232  | 0.282| 1.405| 0.282| 0.512| 0.720 | 0.399|
|         |          | ±0.037| ±0.042 | ±0.054| ±0.012| ±0.054| ±0.084| ±0.300| ±0.087|
|         | (20, 200)| 0.140 | 0.153  | 0.189| 1.410| 0.189| 0.217| 0.592 | 0.198|
|         |          | ±0.023| ±0.027 | ±0.035| ±0.005| ±0.035| ±0.040| ±0.319| ±0.035|

| II-2    | (10, 100)| 0.364 | 0.377  | 0.508| 1.885| 0.508| 0.545| 1.614 | 1.032|
|         |          | ±0.097| ±0.104 | ±0.155| ±0.084| ±0.155| ±0.164| ±0.207| ±0.323|
|         | (10, 200)| 0.233 | 0.238  | 0.343| 1.904| 0.343| 0.364| 1.554 | 0.495|
|         |          | ±0.050| ±0.052 | ±0.083| ±0.071| ±0.083| ±0.090| ±0.178| ±0.178|
|         | (20, 100)| 0.644 | 0.664  | 0.760| 1.951| 0.760| 0.831| 1.771 | 1.473|
|         |          | ±0.146| ±0.137 | ±0.151| ±0.036| ±0.151| ±0.157| ±0.109| ±0.162|
|         | (20, 200)| 0.393 | 0.435  | 0.559| 1.952| 0.559| 0.587| 1.777 | 0.953|
|         |          | ±0.076| ±0.091 | ±0.118| ±0.040| ±0.118| ±0.119| ±0.139| ±0.269|

Figure 3: Ellipsoidal plots of the SPD matrix response versus the rFOPG predictor $\hat{\beta}_1^T X$ and $X_{10}$ using Model II-1 with $(n = 100, p = 10)$. Each horizontal ellipse is a Ellipsoidal representation of a SPD matrix and the vertical axis is the value of (a) $\hat{\beta}_1^T X$; (b) $X_{10}$.

7.3 Scenario III: Fréchet SDR for Spherical Data

We implement the proposed CMS ensemble and CS ensemble methods when the responses lie in a sphere. Specifically, let $\Omega_Y = S^2$ or $S$, which is the unit sphere in $\mathbb{R}^3$ or $\mathbb{R}^2$, endowed with geodesic distance $d_g(Y_1, Y_2) = \arccos(Y_1^T Y_2)$. Let $\beta_1 = e_1 = (1, 0, \ldots, 0)$
and $\beta_2 = e_2 = (0, 1, 0, \ldots, 0)$. We still generate i.i.d. observations $X_1, \ldots, X_n$ from a $p$-dimensional random vector whose components are independent $U[0, 1]$ random variables. We generate $Y$ according to the following three models:

**III-1:** $\Omega_Y = \mathbb{S}$. Let the Fréchet regression function be

$$m(X_i) = E^{(d)}(Y_i|X_i) = (\cos(\pi \beta_1^TX_i), \sin(\pi \beta_1^TX_i)) = (\cos(\pi X_{i1}), \sin(\pi X_{i1})).$$

We generate Normally distributed error $\varepsilon_i$ on the tangent space $T_{m(x_i)}\Omega_Y$ with mean 0 and standard deviation 0.2. Specifically, we first generate $\delta_i \sim N(0, 0.2^2)$ on $\mathbb{R}$ and then let $\varepsilon_i = (-\delta_i \sin(\pi X_{i1}), \delta_i \cos(\pi X_{i1}))$ for $i = 1, \ldots, n$. We then generate $Y_i$ by

$$Y_i = \cos(\|\varepsilon_i\|)m(X_i) + \sin(\|\varepsilon_i\|)\frac{\varepsilon_i}{\|\varepsilon_i\|},$$

which is the exponential map $\text{Exp}_{m(X_i)}(\varepsilon_i)$.

**III-2:** $\Omega_Y = \mathbb{S}^2$. Let the Fréchet regression function be

$$m(X_i) = E^{(d)}(Y_i|X_i) = ((1 - X_{i2}^2)^{1/2} \cos(\pi X_{i1}), (1 - X_{i2}^2)^{1/2} \sin(\pi X_{i1}), X_{i2}).$$

We generate binary Normal error $\varepsilon_i$ on the tangent space $T_{m(x_i)}\Omega_Y$, with mean 0 and diagonal covariance matrix $\text{diag}(0.2^2, 0.2^2)$. Specifically, we first independently generate $\delta_{i1}, \delta_{i2} \overset{iid}{\sim} N(0, 0.2^2)$, then let $\varepsilon_i = \delta_{i1} v_1 + \delta_{i2} v_2$, where $\{v_1, v_2\}$ forms orthogonal basis of tangent space $T_{m(x_i)}\Omega_Y$. We then generate $Y_i$ by

$$Y_i = \text{Exp}_{m(X_i)}(\varepsilon_i) = \cos(\|\varepsilon_i\|)m(X_i) + \sin(\|\varepsilon_i\|)\frac{\varepsilon_i}{\|\varepsilon_i\|}, \quad i = 1, \ldots, n,$$

where $\| \cdot \|$ is the Euclidean norm.

**III-3:** $\Omega_Y = \mathbb{S}^2$. We generate independent errors $\varepsilon_{i1}, \varepsilon_{i2} \overset{iid}{\sim} N(0, 0.2^2)$. Then we generate $Y_i$ by

$$Y_i = (\sin(X_{i1} + \varepsilon_{i1}) \sin(X_{i2} + \varepsilon_{i2}), \sin(X_{i1} + \varepsilon_{i1}) \cos(X_{i2} + \varepsilon_{i2}), \cos(X_{i1} + \varepsilon_{i1})).$$

In this model, $B_0 = (e_1, e_2)$ and $d_0 = 2$.

For Model III-1, $B_0 = \beta_1$ and $d_0 = 1$; for Model III-2, $B_0 = (e_1, e_2)$ and $d_0 = 2$; for Model III-3, $B_0 = (e_1, e_2)$ and $d_0 = 2$. Models III-1, III-3 are the same as Models II, III in Ying and Yu (2020), respectively, and a similar model to Model III-2 was used in Petersen and Müller (2019, Section 8). In each setting, we consider two sample sizes, $n = 100, 200,$
and two dimension choices, \( p = 10, 20 \). For each combinations of \( n \) and \( p \), we repeat the experiments 100 times and summarize the estimation error in Table 3. We conclude that rFOPG, rFMAVE, FOLS, FIHT and FSIR can accurately estimate the central subspace in all settings. Also, as \( n \) increases, the performance of all methods except FPHD improve, especially for the more complex Models III-2 and III-3. We also show the scatter plots of

| Model | \((p,n)\) | rFOPG | rFMAVE | FOLS | FPHD | FIHT | FSIR | FSAVE | FDR |
|-------|----------|-------|--------|------|------|------|------|-------|-----|
| III-1 | (10,100) | 0.100 | 0.099  | 0.103 | 1.405 | 0.103 | 0.151 | 0.336 | 0.200|
|       |          | ±0.025| ±0.025 | ±0.026| ±0.011| ±0.026| ±0.037| ±0.220| ±0.052|
|       | (10,200) | 0.066 | 0.066  | 0.067 | 1.407 | 0.067 | 0.080 | 0.178 | 0.125|
|       |          | ±0.015| ±0.016 | ±0.015| ±0.009| ±0.015| ±0.021| ±0.115| ±0.029|
|       | (20,100) | 0.153 | 0.153  | 0.153 | 1.405 | 0.153 | 0.152 | 0.514 | 0.344|
|       |          | ±0.028| ±0.030 | ±0.025| ±0.013| ±0.026| ±0.064| ±0.188| ±0.062|
|       | (20,200) | 0.103 | 0.101  | 0.101 | 1.409 | 0.100 | 0.146 | 0.341 | 0.196|
|       |          | ±0.017| ±0.017 | ±0.019| ±0.007| ±0.019| ±0.027| ±0.152| ±0.034|
|       | (10,100) | 0.307 | 0.312  | 0.330 | 1.918 | 0.334 | 0.399 | 1.216 | 0.749|
|       |          | ±0.073| ±0.080 | ±0.078| ±0.053| ±0.079 |±0.105 |±0.226 |±0.299|
|       | (10,200) | 0.202 | 0.200  | 0.223 | 1.941 | 0.225 | 0.256 | 0.868 | 0.313|
|       |          | ±0.042| ±0.048 | ±0.044| ±0.046| ±0.045 |±0.054 |±0.295 |±0.067|
|       | (20,100) | 0.500 | 0.485  | 0.537 | 1.957 | 0.543 | 0.593 | 1.479 | 1.286|
|       |          | ±0.085| ±0.082 | ±0.104| ±0.030| ±0.107| ±0.094| ±0.086| ±0.173|
|       | (20,200) | 0.314 | 0.298  | 0.342 | 1.963 | 0.345 | 0.409 | 1.339 | 0.531|
|       |          | ±0.049| ±0.049 | ±0.050| ±0.031| ±0.052| ±0.066| ±0.137| ±0.089|
| III-2 | (10,100) | 0.198 | 0.163  | 0.256 | 1.959 | 0.256 | 0.292 | 1.054 | 0.664|
|       |          | ±0.050| ±0.031 | ±0.069| ±0.030| ±0.069 |±0.094 |±0.287 |±0.310|
|       | (10,200) | 0.129 | 0.110  | 0.180 | 1.977 | 0.180 | 0.170 | 0.620 | 0.218|
|       |          | ±0.025| ±0.021 | ±0.038| ±0.016| ±0.038 |±0.041 |±0.234 |±0.060|
|       | (20,100) | 0.342 | 0.258  | 0.408 | 1.970 | 0.409 | 0.517 | 1.422 | 1.202|
|       |          | ±0.058| ±0.036 | ±0.076| ±0.023| ±0.076 |±0.082 |±0.125 |±0.223|
|       | (20,200) | 0.205 | 0.163  | 0.279 | 1.983 | 0.279 | 0.285 | 1.193 | 0.416|
|       |          | ±0.030| ±0.019 | ±0.043| ±0.012| ±0.043 |±0.055 |±0.224 |±0.102|

Table 3: Mean (± standard deviation) of estimation error for different methods in Scenario III.

Y versus the sufficient predictors obtained by rFOPG. From Figure 4, we see a stronger relation between \( Y \) and \( \hat{\beta}^T X \) than between \( Y \) and an individual predictor \( X_{10} \).
Figure 4: Perspective plots of the spherical response versus the FOPG predictors and individual predictor $X_{10}$. (a), (b) are from Model III-1 with $n = 100, p = 10$; (c), (d), (e) are from Model III-2 with $n = 100, p = 10$, where the colors indicate the values of (c) first sufficient predictor $\hat{\beta}_1^T X$; (d) second sufficient predictor $\hat{\beta}_2^T X$ and (e) $X_{10}$.

8 Real Applications

This section presents two applications of the proposed methods to study human life spans across countries and to study intracerebral hemorrhage, respectively.

8.1 Application to the Human Mortality Data

Compared with a summary statistics such as the crude death rate, viewing the entire age-at-death distributions as data objects gives us more comprehensive understanding of human longevity and health conditions. Mortality distributions are affected by many factors such as economics, the health care system, as well as social and environmental factors. To investigate the potential factors that are related the mortality distributions across different countries, we collect 9 predictors listed below, covering demography, economics, labour market, nutrition, health and environment factors at 2015: (1) Population Density: population per square Kilometer; (2) Sex Ratio: number of males per 100 females in the population;
(3) Mean Childbearing Age: the average age of mothers at the birth of their children; (4) Gross Domestic Product (GDP) per Capita; (5) Gross Value Added (GVA) by Agriculture: the percentage of agriculture, hunting, forestry and fishing activities of gross value added, (6) Consumer price index: treat 2010 as the base year; (7) Unemployment Rate; (8) Expenditure on Health (percentage of GDP); (9) Arable Land (percentage of total land area). The data are collected from United Nation Databases [http://data.un.org/] and UN World Population Prospects 2019 Databases (https://population.un.org/wpp/Download). For each country and age, the life table contains the number of deaths $d(x, n)$ aggregated every 5 years. We treat these data as histograms of death at age, with bin widths equal to 5 years. We smooth the histograms for the 162 countries in 2015 using the ‘frechet’ package available at (https://cran.r-project.org/web/packages/frechet/index.html) to obtain smoothed probability density functions. We then calculate the Wasserstein distances between them.

We first standardize all covariates separately, then apply FOPG to the data. The first four largest eigenvalues of the FOPG matrix are 0.0554, 0.0216, 0.0042, 0.0022. From the scree plot, only the first two components lie to the left of the “elbow” of the graph and should be retained as significant. Thus we determine the dimension of central space is 2.

The first two directions obtained by rFOPG after 10 iterations are

$$\hat{\beta}_1 = (0.080, -0.302, 0.359, 0.856, 0.159, -0.005, 0.025, 0.114, -0.028)^T,$$

$$\hat{\beta}_2 = (-0.009, -0.265, -0.610, 0.284, -0.661, 0.016, -0.194, 0.054, -0.001)^T.$$

The plots of mortality densities versus the first estimated sufficient predictor $\hat{\beta}_1^T X$ are shown in Figure 5 (a) and 5 (b). In Figure 5 (c), we plot the means and standard deviations of the mortality distributions versus the first and second estimated sufficient predictor. From the plots we see that the first sufficient predictor captures the location and variation features of the mortality distributions. Specifically, with the increase of the first sufficient predictor, the mean of the mortality distribution increases non-linearly while the standard deviation decreases in a mirror pattern, which indicates the death age more concentrates between 70 and 80. By checking the estimated coefficients, this trend happens with the increase of GDP per capita, mean age of childbearing significantly. These two factors reflect the economic development level and population structure of a country, with large values for developed countries and small values for developing countries. This indicates that the mortality distributions of developed countries usually have larger means and less variation than the developing countries. Also there is an uptick at the right ends of the mortality distributions for developing countries with small $\hat{\beta}_1^T X$, indicating higher infant mortality rates among these countries.
8.2 Application to the Stroke Data

As a life-threatening type of stroke, intracerebral hemorrhage (ICH) is caused by bleeding inside the brain tissue itself. It is the second most common stroke sub-type and remains a significant cause of morbidity and mortality (Salazar et al., 2020). The density of the hematoma, detected by Computed tomography (CT) scan, is an important index to diagnose and study ICH (Barras et al., 2009; Salazar et al., 2020; Divani et al., 2020). Instead of using any particular summary index, one can study the distribution of hematoma density (i.e., a probability density function of hematoma density) throughout the entire hematoma as a functional object (Petersen et al., 2021). In WRI package (https://CRAN.R-project.org/package=WRI), the head CT hematoma densities of $n = 393$ ICH anonymous subjects were recorded as smoothed probability density functions. The dataset also contains 4 radiological variables and and 5 clinical variables as predictors. To better illustrate the rFOPG method, we only consider the quantitative predictors, including the logarithm of hematoma volume ($X_1$), a continuous index of hematoma shape ($X_2$), length of the time interval between stroke event and the CT scan ($X_3$), age ($X_4$) and weight ($X_5$). A complete description of the data source can be found in Hevesi et al. (2018).

When applying FOPG to the data, the first four largest eigenvalues of the ensemble OPG matrix are $1.607 \times 10^{-6}$, $2.814 \times 10^{-7}$, $2.095 \times 10^{-7}$ and $9.576 \times 10^{-8}$. Therefore, we consider a Fréchet central subspace with dimension 1. Then we implement rFOPG using 10 iterations and estimate the directions (coefficients of $X$) as $\hat{\beta}_1 = (0.839, 0.452, -0.106, 0.010, 0.282)$. Based on the estimated direction, the logarithm of hematoma volume and the hematoma shape are the most important factors that affect the hematoma density distributions, both of
which are significant under partial Wasserstein F-tests with p-values less than 0.001 (Petersen et al., 2021). The plots of $Y$, the hematoma density distributions, versus the first estimated sufficient predictor $\hat{\beta}_1 X$ is shown in Figure 6 (a)(b). In Figure 6 (c), we plot four summary statistics (mean, mode, standard deviation and skewness) of the hematoma density verses the first estimated sufficient predictor. From these two plots, we conclude that with the increase of the first sufficient predictor, the mean of the hematoma density increase linearly while the skewness decrease. However, the standard deviation of hematoma density shows only a weak increase pattern.

Figure 6: (a)(b) Plots of head CT hematoma density distributions versus the first sufficient predictor; (c) Plot of mean, mode, standard deviation and skewness of head CT hematoma density distributions versus the first sufficient predictor.

9 Conclusion

We provide a flexible SDR approach to mitigate the curse of dimensionality and provide a tool for data visualization for Fréchet regression. After mapping the metric-space valued random object $Y$ to a real-valued random variable $f(Y)$ using a class of functions, the proposed approach is flexible enough to turn any existing SDR method for Euclidean $(X,Y)$ into one for Euclidean $X$ and metric space-valued $Y$. If the class of functions is sufficiently rich, we provided the theoretical guarantee to uncover the Fréchet SDR space. We showed that such a class, which we call an ensemble, can be generated by a universal kernel. We also established the consistency and asymptotic convergence rate of the proposed methods. The numerical performance was demonstrated in both simulation studies and real applications.
References

Ambrosio, L., Gigli, N. and Savaré, G. (2004), ‘Gradient flows with metric and differentiable structures, and applications to the wasserstein space’, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl* 15(3-4), 327–343.

Barras, C. D., Tress, B. M., Christensen, S., MacGregor, L., Collins, M., Desmond, P. M., Skolnick, B. E., Mayer, S. A., Broderick, J. P. and Diringer, M. N. (2009), ‘Density and shape as ct predictors of intracerebral hemorrhage growth’, *Stroke* 40(4), 1325–1331.

Bigot, J., Gouet, R., Klein, T. and López, A. (2017), ‘Geodesic pca in the wasserstein space by convex pca’, *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 53, 1–26.

Chen, Y., Lin, Z. and Müller, H.-G. (2020), ‘Wasserstein regression’, *preprint arXiv:2006.09660*.

Christmann, A. and Steinwart, I. (2010), Universal kernels on non-standard input spaces, in ‘in Advances in Neural Information Processing Systems’, pp. 406–414.

Cook, R. D. (1996), ‘Graphics for regressions with a binary response’, *Journal of the American Statistical Association* 91(435), 983–992.

Cook, R. D. and Li, B. (2002), ‘Dimension reduction for conditional mean in regression’, *The Annals of Statistics* 30(2), 455–474.

Cook, R. D. and Weisberg, S. (1991), ‘Sliced inverse regression for dimension reduction: Comment’, *Journal of the American Statistical Association* 86(414), 328–332.

Ding, S. and Cook, R. D. (2015), ‘Tensor sliced inverse regression’, *Journal of Multivariate Analysis* 133, 216–231.

Divani, A. A., Liu, X., Petersen, A., Lattanzi, S., Anderson, C. S., Ziai, W., Torbey, M. T., Moullaali, T. J., James, M. L. and Jafarli, A. (2020), ‘The magnitude of blood pressure reduction predicts poor in-hospital outcome in acute intracerebral hemorrhage’, *Neurocritical Care* 33(2), 389–398.

Dubey, P. and Müller, H.-G. (2019), ‘Fréchet analysis of variance for random objects’, *Biometrika* 106(4), 803–821.

Dubey, P. and Müller, H.-G. (2020), ‘Fréchet change-point detection’, *The Annals of Statistics* 48(6), 3312–3335.
Fan, J., Xue, L. and Yao, J. (2017), ‘Sufficient forecasting using factor models’, *Journal of Econometrics* **201**(2), 292–306.

Feragen, A., Lauze, F. and Hauberg, S. (2015), Geodesic exponential kernels: when curvature and linearity conflict, *in* ‘Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition’, pp. 3032–3042.

Ferré, L. and Yao, A.-F. (2003), ‘Functional sliced inverse regression analysis’, *Statistics* **37**(6), 475–488.

Ferreira, L. K. and Busatto, G. F. (2013), ‘Resting-state functional connectivity in normal brain aging’, *Neuroscience & Biobehavioral Reviews* **37**(3), 384–400.

Fournier, N. and Guillin, A. (2015), ‘On the rate of convergence in wasserstein distance of the empirical measure’, *Probability Theory and Related Fields* **162**(3-4), 707–738.

Fréchet, M. (1948), ‘Les éléments aléatoires de nature quelconque dans un espace distancié’, *Annales de l’Institut Henri Poincaré* p. 215–310.

Hevesi, M., Bershad, E. M., Jafari, M., Mayer, S. A., Selim, M., Suarez, J. I. and Divani, A. A. (2018), ‘Untreated hypertension as predictor of in-hospital mortality in intracerebral hemorrhage: A multi-center study’, *Journal of Critical Care* **43**, 235–239.

Honeine, P. and Richard, C. (2010), The angular kernel in machine learning for hyperspectral data classification, *in* ‘2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing’, IEEE, pp. 1–4.

Hsing, T. and Ren, H. (2009), ‘An rkhs formulation of the inverse regression dimension-reduction problem’, *The Annals of Statistics* **37**(2), 726–755.

Jayasumana, S., Hartley, R., Salzmann, M., Li, H. and Harandi, M. (2013), Combining multiple manifold-valued descriptors for improved object recognition, *in* ‘2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)’, IEEE, pp. 1–6.

Lei, J. (2020), ‘Convergence and concentration of empirical measures under wasserstein distance in unbounded functional spaces’, *Bernoulli* **26**(1), 767–798.

Li, B. (2018), *Sufficient Dimension Reduction: Methods and Applications with R*, CRC Press.
Li, B., Kim, M. K. and Altman, N. (2010), ‘On dimension folding of matrix-or array-valued statistical objects’, *The Annals of Statistics* **38**(2), 1094–1121.

Li, B. and Song, J. (2017), ‘Nonlinear sufficient dimension reduction for functional data’, *The Annals of Statistics* **45**(3), 1059–1095.

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

Li, B., Zha, H. and Chiaromonte, F. (2005), ‘Contour regression: a general approach to dimension reduction’, *The Annals of Statistics* **33**(4), 1580–1616.

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, K.-C. and Duan, N. (1989), ‘Regression analysis under link violation’, *The Annals of Statistics* pp. 1009–1052.

Luo, W., Xue, L., Yao, J. and Yu, X. (2021), ‘Inverse moment methods for sufficient forecasting using high-dimensional predictors’, *Biometrika, in press*.

Parzen, E. (1979), ‘Nonparametric statistical data modeling’, *Journal of the American statistical Association* **74**(365), 105–121.

Petersen, A., Liu, X. and Divani, A. A. (2021), ‘Wasserstein $f$-tests and confidence bands for the fréchet regression of density response curves’, *The Annals of Statistics* **49**(1), 590–611.

Petersen, A. and Müller, H.-G. (2016), ‘Functional data analysis for density functions by transformation to a hilbert space’, *The Annals of Statistics* **44**(1), 183–218.

Petersen, A. and Müller, H.-G. (2019), ‘Fréchet regression for random objects with euclidean predictors’, *The Annals of Statistics* **47**(2), 691–719.

Salazar, P., Di Napoli, M., Jafari, M., Jafarli, A., Ziai, W., Petersen, A., Mayer, S. A., Bershad, E. M., Damani, R. and Divani, A. A. (2020), ‘Exploration of multiparameter hematoma 3d image analysis for predicting outcome after intracerebral hemorrhage’, *Neurocritical Care* **32**(2), 539–549.

35
Schwartzman, A. (2006), Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data, PhD thesis, Stanford University.

Steinwart, I. and Christmann, A. (2008), *Support Vector Machines*, Springer.

Xia, Y. (2007), ‘A constructive approach to the estimation of dimension reduction directions’, *The Annals of Statistics* **35**(6), 2654–2690.

Xia, Y., Tong, H., Li, W. K. and Zhu, L.-X. (2002), ‘An adaptive estimation of dimension reduction space (with discussion)’, *Journal of the Royal Statistical Society. Series B* **64**(3), 363–410.

Ye, Z. and Weiss, R. E. (2003), ‘Using the bootstrap to select one of a new class of dimension reduction methods’, *Journal of the American Statistical Association* **98**(464), 968–979.

Yin, X. and Li, B. (2011), ‘Sufficient dimension reduction based on an ensemble of minimum average variance estimators’, *The Annals of Statistics* **39**(6), 3392–3416.

Yin, X., Li, B. and Cook, R. D. (2008), ‘Successive direction extraction for estimating the central subspace in a multiple-index regression’, *Journal of Multivariate Analysis* **99**(8), 1733–1757.

Ying, C. and Yu, Z. (2020), ‘Fréchet sufficient dimension reduction for random objects’, *arXiv preprint arXiv:2007.00292*.

Yu, X., Yao, J. and Xue, L. (2020), ‘Nonparametric estimation and conformal inference of the sufficient forecasting with a diverging number of factors’, *Journal of Business & Economic Statistics, in press*. 