Journal of Statistical Planning and Inference 167 (2015) 58–68

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Functional sufficient dimension reduction: Convergence rates and multiple functional case

Heng Lian

School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, Australia

ARTICLE INFO

Article history:
Received 24 December 2014
Received in revised form 24 May 2015
Accepted 24 May 2015
Available online 1 June 2015

Keywords:
Convergence rates
Functional data
Regularization
Sliced average variance estimation
Sliced inverse regression

ABSTRACT

Although sufficient dimension reduction for functional data has received some attention in the literature, its theoretical properties are less understood. Besides, the current literature only focused on sliced inverse regression (SIR). In this paper we consider functional version of SIR and SAVE (sliced average variance estimation) via a Tikhonov regularization approach. Besides consistency, we show that their convergence rates are the same as the minimax rates for functional linear regression, which we think is an interesting theoretical result given that sufficient dimension reduction is much more flexible than functional linear regression. In sufficient dimension reduction, it is well known that estimation of multiple directions requires extraction of multiple eigenfunctions. We also consider multiple functional dimension reduction, in which *one eigenfunction* surprisingly recovers multiple index functions at once, despite its similarity with single functional case. The numerical properties are illustrated using several simulation examples as well as a Japanese weather dataset.

1. Introduction

In functional regression, we observe i.i.d. copies of (X, Y), denoted by $(X_i, Y_i), i = 1, \ldots, n$, where $X \in L_2[0, 1]$ is a stochastic process used as the functional predictor and $Y \in R$ is the response. In parametric or nonparametric functional regression, we assume

$$Y = f(X) + \epsilon,$$

where ϵ is an additive mean zero noise, f either assumes a parametric form $f(X) = \langle \beta, X \rangle$ in functional linear regression, or is an entirely unknown function (except for some smoothness assumption) in functional nonparametric regression (Cardot et al., 1999; Ferraty and Vieu, 2002; Cardot et al., 2003; Ferraty and Vieu, 2004; Yao et al., 2005; Cai and Hall, 2006; Preda, 2007; Preda et al., 2007; Ferraty et al., 2010; Delaigle and Hall, 2012). Here we use $\langle \beta, X \rangle$ for the usual inner product $\int_0^1 \beta(t)X(t)dt$.

In this paper, we consider the semiparametric sufficient dimension reduction approach for functional data. The basic assumption for sufficient dimension reduction is that the *distribution* of Y given X depends on K linear projections of X, say $\langle \beta_1, X \rangle, \ldots, \langle \beta_K, X \rangle$ (that is, these projections are "sufficient" to explain the responses given the covariates). This is a quite general formulation and enables us to write the true model as

$$Y = g(\langle \beta_1, X \rangle, \ldots, \langle \beta_K, X \rangle, \epsilon). \quad (1)$$

E-mail address: heng.lian@unsw.edu.au.

http://dx.doi.org/10.1016/j.jspi.2015.05.008
0378-3758/

In particular, the linear projection can appear inside the variance component of the model, g is not necessarily additive, and the noise does not need to be assumed to be additive on the mean function. Although the directions themselves are not identifiable, the space spanned by these directions is identifiable and is called the effective dimension reduction (e.d.r.) space. Two estimation methods, sliced inverse regression (SIR) and sliced average variance estimation (SAVE), have received much attention (Li, 1991; Cook and Weisberg, 1991). The former is based on the estimation of the conditional mean of X given Y (hence the name "inverse regression") and the latter on the estimation of conditional variance. The estimation procedure does not involve explicit estimation of the link function, which can be regarded as an advantage when only the β_j are of interest.

Functional version of SIR has received interest recently (Ferré and Yao, 2003, 2005; Hsing and Ren, 2009; Li and Hsing, 2010). In SIR or SAVE, inverse of the covariance operator of the predictor plays a critical role in estimation. However, even when the covariance operator is injective, the inverse is generally not bounded and thus some form of regularization is necessary. In this article, we use a ridge-type regularization method to achieve boundedness in the inverse operator as in Hall and Horowitz (2005) and demonstrate their statistical properties in recovering the linear projections. Ferré and Yao (2005) avoid inverse operation using a different estimation approach but this was later shown to require much stronger assumptions for its validity (Cook et al., 2010). Alternatively, boundedness of the inverse covariance operator can be achieved by truncation of the Karhunen–Loève expansion of the random predictor as done in Ferré and Yao (2003). However, a continuous regularization parameter might provide a more flexible fitting to data as discussed in Ramsay and Silverman (2005). Compared to Ferré and Yao (2003) which only considered consistency of functional SIR, we also consider functional SAVE. It is well-known that SAVE can potentially recover more directions in the effective dimension reduction space, although it is more complicated for estimation. Furthermore, under more restrictive assumptions, we show that for functional SIR and SAVE the convergence rates of the estimator for $\beta_j, j = 1, \ldots, K$ are the same as the optimal rates for the estimator in functional linear regression. This result looks remarkable since the model structure assumed for dimension reduction is much more general than functional linear regression as discussed above. The fact that it can achieve the same convergence rate is a strong theoretical argument for its use in practice, especially at the exploratory data analysis stage when we hesitate to impose modelling constraints. Jiang et al. (2014) have recently considered functional sliced inverse regression for longitudinal data with some convergence rate, however they do not aim for optimal convergence rate, which is acknowledgably hard to obtain under their setup.

Recently, Lian and Li (2014) have considered series expansion approach to functional SIR and functional SAVE and showed their statistical consistency. Besides using a regularization approach here, we make two main contributions in this paper. First, we demonstrate minimax convergence rates for both functional SIR and functional SAVE. Second, we consider dimension reduction with multiple functions in which one eigenvector can surprisingly recover multiple index functions. The rest of the paper is organized as follows. We introduce the SIR and SAVE methods and present the estimation approach and asymptotic theory in Section 2. Dimension reduction involving multiple functional predictors is discussed in Section 3. Numerical examples and a real data application are presented in Section 4. Finally, we conclude the study with a discussion in Section 5. The proofs are contained in the supplementary Appendix.

2. Functional SIR and SAVE

2.1. Background

In this article, we assume the entire trajectory of the noise-free process X is observed. When the process is densely and accurately measured, this is a reasonable assumption. For simplicity, we assume $EX = 0$. We also assume the fourth moment of X exists, that is $E\|X\|^4 < \infty$. Using the well-known Karhunen–Loève expansion, we can write

$$X = \sum_{j=1}^{\infty} \xi_j \phi_j,$$

where $E\xi_j^2 = \lambda_j$ are the eigenvalues and ϕ_j are the eigenfunctions. The (population) covariance operator of X is defined by $\Gamma = Var(X) = E(X \otimes X)$, where for any $x, y \in L_2[0, 1]$, $x \otimes y$ denotes the linear operator $L_2[0, 1] \to L_2[0, 1]$ such that $(x \otimes y)(z) = \langle x, z\rangle y$. From now on, we assume all eigenvalues $\lambda_1 > \lambda_2 > \cdots > 0$ are positive and distinct as usually done in the functional data literature (Hall and Horowitz, 2007; Ferré and Yao, 2003). If some eigenvalues of Γ are zero, the components of β_k in the kernel space of Γ cannot be identified.

We focus on the estimation of the e.d.r. space spanned by K linearly independent directions $\beta_1, \ldots, \beta_K$, which is denoted by $\mathcal{S}_{Y|X}$. Let $\Gamma\mathcal{S}_{Y|X}$ be the space spanned by $\Gamma\beta_1, \ldots, \Gamma\beta_K$. Let $B_X = (\langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle)^T$. The principle of SIR and SAVE is based on the following result which was previously stated in Lian and Li (2014) without proof, since it is a straightforward extension of the multivariate case.

Theorem 1. (a) *(Ferré and Yao, 2003) Suppose for all $b \in L_2[0, 1]$, the conditional expectation $E(\langle b, X\rangle|B_X)$ is linear in $\langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle$. Then $E(X|Y) \in \Gamma\mathcal{S}_{Y|X}$;*

(b) *If in addition $Var(X|B_X)$ is nonrandom, $\Gamma - Var(X|Y) \in \Gamma\mathcal{S}_{Y|X}$.*