Sufficient dimension reduction via distance covariance

Wenhui Sheng       Xiangrong Yin

Department of Statistics, University of Georgia, Athens, GA 30602

EXTENDED ABSTRACT

In this article, we introduce a novel approach to sufficient dimension reduction problems through using distance covariance. To be precise, we consider the model \( Y \perp X | \eta^T X \), where \( \eta \) is a \( p \times d \) matrix with \( d \leq p \). The column space of \( \eta \), denoted by \( \text{Span}(\eta) \), is called a dimension reduction subspace. When the intersection of all dimension reduction subspaces is itself a dimension reduction subspace, it is called a central subspace, denoted by \( S_{Y|X} \).

The target of this article is to estimate a basis for the central subspace \( S_{Y|X} \). The basic idea is to maximize sample version of squared distance covariance under a certain constraint to estimate the basis. Distance covariance is a new distance measure, analogous to product-moment covariance. To introduce the sample version of squared distance covariance briefly, let \( \beta \) to be a \( p \times d \) matrix and \((X, Y) = \{(X_k, Y_k) : k = 1, \cdots, n\} \) to be a random sample from the joint distribution of random vector \( X \) in \( \mathbb{R}^p \) and a random variable \( Y \) in \( \mathbb{R} \). The sample version of squared distance covariance between \( \beta^T X \) and \( Y \) is defined as

\[
V_n^2(\beta^T X, Y) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl}(\beta)B_{kl},
\]

where, for \( k, l = 1, \cdots, n \),

\[
A_{kl}(\beta) = a_{kl}(\beta) - \bar{a}_k(\beta) - \bar{a}_l(\beta) + \bar{a}.(\beta)
\]

\[
a_{kl}(\beta) = |\beta^T X_k - \beta^T X_l|, \quad \bar{a}_k(\beta) = \frac{1}{n} \sum_{l=1}^n a_{kl}(\beta),
\]

\[
\bar{a}_l(\beta) = \frac{1}{n} \sum_{k=1}^n a_{kl}(\beta), \quad \bar{a}.(\beta) = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}(\beta).
\]
Similarly, define $b_{kl} = |Y_k - Y_l|$ and $B_{kl} = b_{kl} - \bar{b}_k - \bar{b}_l + \bar{b}_n$. Here $| \cdot |$ is the Euclidean norm in the respective dimension. We proved that under a mild condition, the estimate of a basis for $S_Y|X$ maximized $V_n^2(\beta^T X, Y)$ under a certain constraint, that is,

$$\eta_n = \arg \max_{\beta^T \Sigma_X \beta = I_d} V_n^2(\beta^T X, Y),$$

where $\Sigma_X$ is the sample version of the covariance matrix of $X$ and $I_d$ is an identity matrix. This extends the recent work of Sheng and Yin [Direction estimation in single-index models via distance covariance, resubmitted] who proposed the method and used it to estimate the direction in single-index models. Through this extension, the method is more flexible, so it can deal with more complicated regression problems and we find new aspect of the extended methodology, which is given in the following proposition.

**Proposition 1** Let $\beta$ be a $p \times d_1$ matrix. Suppose $Y \perp X \mid \eta^T X$, where $\eta$ is a $p \times d$ matrix. If $\text{Span}(\beta) \subseteq \text{Span}(\eta) = S_Y|X$, then $V^2(\beta^T X, Y) \leq V^2(\eta^T X, Y)$. The equality holds if and only if $\text{Span}(\beta) = \text{Span}(\eta)$. On the other hand, if $\text{Span}(\beta) \nsubseteq \text{Span}(\beta)$ and further assume $P_{\eta^T \Sigma_X \eta}^T X \perp Q_{\eta^T \Sigma_X \eta}^T X$, then $V^2(\beta^T X, Y) < V^2(\eta^T X, Y)$. Here $\beta^T \Sigma_X \beta = I_{d_1}$ and $\eta^T \Sigma_X \eta = I_d$.

Proposition 1 presents an interesting pattern of the squared distance covariance $V_n^2(\beta^T X, Y)$, that is the value of $V_n^2(\beta^T X, Y)$ is getting bigger when the distance between the subspaces $\text{Span}(\beta)$ and $\text{Span}(\eta)$ is getting closer and the maximum value is achieved if and only if the subspaces $\text{Span}(\beta)$ and $\text{Span}(\eta)$ match each other. This proposition indicates that we can always recover the central subspace by maximizing $V_n^2(\beta^T X, Y)$ with respect to $\beta$. As for the independence condition $P_{\eta^T \Sigma_X \eta}^T X \perp Q_{\eta^T \Sigma_X \eta}^T X$, it is not as strong as it seems to be,
and we showed that it could be satisfied when \( p \) was reasonably large. Here \( P_{\eta(\Sigma_X)} \) is a projection operator and \( Q_{\eta(\Sigma_X)} = I - P_{\eta(\Sigma_X)} \).

In addition, we used a permutation test to estimate the dimension of the central subspace. We also investigated the asymptotic properties of our estimate and showed that under regular conditions, it was root-n consistent and asymptotically normal. We compared the performance of our method with some other dimension reduction methods by simulations and found that our method was very competitive and robust across a number of models. We also analyzed a real data set concerning the identification of different sounds to demonstrate the efficacy of our method.

KEY WORDS: Brownian Distance Covariance; Central Subspace; Sufficient Dimension Reduction; Multiple-index Regression