Journal of Multivariate Analysis 133 (2015) 27–37

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Self-consistency and a generalized principal subspace theorem

Thaddeus Tarpey[a,*], Nicola Loperfido[b]

[a] Department of Mathematics and Statistics, Wright State University, 120 MM Building, Dayton, OH, 45435, USA
[b] Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino, "Carlo Bo", Via Saffi 2, 61029 Urbino (PU), Italy

HIGHLIGHTS

- A principal subspace theorem for multivariate mixture distributions is proved based on the notion of self-consistency.
- The results are used to characterize principal points for multivariate skew-normal distributions.
- The results are used in projection pursuit to find projections of multivariate data that deviate from normality.

ARTICLE INFO

Article history:
Received 21 October 2013
Available online 19 September 2014

AMS subject classifications:
primary 62H30
secondary 62H25

Keywords:
Location mixtures
Principal points
Projection pursuit
Skewness
Skew-normal distribution
Spherical distribution

ABSTRACT

Principal subspace theorems deal with the problem of finding subspaces supporting optimal approximations of multivariate distributions. The optimality criterion considered in this paper is the minimization of the mean squared distance between the given distribution and an approximating distribution, subject to some constraints. Statistical applications include, but are not limited to, cluster analysis, principal components analysis and projection pursuit. Most principal subspace theorems deal with elliptical distributions or with mixtures of spherical distributions. We generalize these results using the notion of self-consistency. We also show their connections with the skew-normal distribution and projection pursuit techniques. We also discuss their implications, with special focus on principal points and self-consistent points. Finally, we access the practical relevance of the theoretical results by means of several simulation studies.

1. Introduction

One of the overarching goals of statistics is to obtain a useful summary of data by means of well-chosen statistics. In multivariate data analysis, this goal is often achieved via dimension reduction, using tools such as principal components and projection pursuit. Just as the mean is frequently used to summarize an entire distribution by a single point (i.e. a measure of location), cluster analysis generalizes the mean from one to several points. The cluster means serve as prototypical points that represent the heterogeneity within a probability distribution. Cluster analysis is related to the problem of optimally stratifying a probability distribution whereby the cluster means determine a partition or stratification of the population. These well-known statistical methods, as well as other methods, are related to the notion of self-consistency [45]. In the context of this paper, given a random vector $\boldsymbol{X}$ of interest, the notion of self-consistency basically refers to a random vector, say $\boldsymbol{Y}$, that provides a summarization of $\boldsymbol{X}$ (e.g. via dimension reduction or an optimal partitioning of the support of $\boldsymbol{X}$).

[*] Corresponding author.
E-mail addresses: thaddeus.tarpey@wright.edu (T. Tarpey), nicola.loperfido@uniurb.it (N. Loperfido).

http://dx.doi.org/10.1016/j.jmva.2014.08.012
0047-259X/© 2014 Elsevier Inc. All rights reserved.

First, we begin with some definitions, starting with the notion of self-consistency. A very broad definition is to say a random vector $\boldsymbol{Y}$ is self-consistent for a random vector $\boldsymbol{X}$ if $E[\boldsymbol{X}|\boldsymbol{Y}] = \boldsymbol{Y}$ almost surely. The following definition is more restrictive but encompasses most applications of self-consistency:

Definition 1.1. For two jointly distributed p-variate random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we say that $\boldsymbol{Y}$ is self-consistent for $\boldsymbol{X}$ if $E[\boldsymbol{X}|\boldsymbol{Y}] = \boldsymbol{Y}$ almost surely where the support of $\boldsymbol{Y}$ spans a linear subspace, say $\mathcal{S}$, of dimension $q \leq p$ and that $\boldsymbol{Y}$ is measurable with respect to the projection of $\boldsymbol{X}$ on $\mathcal{S}$.

This notion of self-consistency includes not only cluster means and principal components, but also principal curves [24], principal variables [38], and others.

Applications of self-consistency that we shall focus on primarily are principal points [14] and self-consistent points [15]:

Definition 1.2. Let $\boldsymbol{X}$ denote a p-variate random vector. Consider a set of k distinct points $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$ in $\Re^p$. Define $\boldsymbol{Y} = \boldsymbol{y}_j$ if $\|\boldsymbol{X} - \boldsymbol{y}_j\| < \|\boldsymbol{X} - \boldsymbol{y}_h\|,\ h \neq j$. If $\boldsymbol{Y}$ is self-consistent for $\boldsymbol{X}$, then the points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ are called *k self-consistent points* for $\boldsymbol{X}$.

Definition 1.3. Given a set of k distinct points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$, define $\boldsymbol{Y} = \boldsymbol{\xi}_j$ if $\|\boldsymbol{X} - \boldsymbol{\xi}_j\| < \|\boldsymbol{X} - \boldsymbol{\xi}_h\|,\ h \neq j$. Then the points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ are called *k principal points* for $\boldsymbol{X}$ if $E\|\boldsymbol{X} - \boldsymbol{Y}\|^2 \leq E\|\boldsymbol{X} - \boldsymbol{Y}^*\|^2$ where $\boldsymbol{Y}^*$ is any other random vector with support consisting of at most k points.

For a given value of k, a distribution can have several different sets of k self-consistent points [e.g. see [42]] and it is also possible to have more than one set of k principal points. [47] proved that the mean of a distribution must lie in the convex hull of any set of k self-consistent points. Without loss of generality, we shall assume throughout that the mean of the distribution under consideration is zero. We shall also assume that the underlying distribution $\boldsymbol{X}$ is continuous and has finite second moments. Thus, the inequality sign in Definition 1.2 can be changed to less-than-or-equal without effecting the definition. [15] showed that principal points must be self-consistent points. Basically, principal points are cluster means for theoretical distributions. The cluster means from the standard k-means algorithm [e.g. [22,23,33]] are self-consistent points for the empirical distribution and represent nonparametric estimators of the principal points of the underlying distribution.

Typically, analytical solutions for the principal points (and other self-consistent objects) are not available, particularly for multivariate distributions. The search for self-consistent objects, such as principal points, becomes much easier if the search can be confined to a smaller dimensional subspace. Principal subspace theorems stipulate that the support of a self-consistent approximation lies in some particular subspace, and they are the primary theoretical motivation for this paper.

The rest of the paper is organized as follows. Section 2 provides reviews of the main results on principal subspaces, with special emphasis on principal points. Section 3 contains the main results, dealing with self-consistency and with mixtures spherical distributions. Section 4 shows their connections with the skew-normal distribution and skewness-based projection pursuit. The simulation study in Section 5 assesses the usefulness of the latter method for estimating principal subspaces. Section 6 provides a simulation example that illustrates the connection between a principal component axis and self-consistency using the theoretical results in Section 3. Section 7 discusses the paper's results, their limitations and gives some hints for future research.

2. Prior results

[14] was the first to conjecture a principal subspace theorem that provided a connection between principal points and principal component subspaces. In particular, it was conjectured that for elliptical distributions, if k principal points span a space of dimension $q < p$, then this linear space coincides with the space spanned by the q eigenvectors of the covariance matrix associated with the q largest eigenvalues. [14] proved the conjecture for $k = 2$ principal points, and the conjecture was proved for any value of k by [47] who were the first to use the term *principal subspace theorem*. In particular, they proved that if k self-consistent points of an elliptical distribution span a space of dimension $q < p$, then this space must be spanned by q eigenvectors of the covariance matrix (i.e. be a principal component space) and for principal points, this space must be spanned by the q eigenvectors of the covariance matrix associated with the largest eigenvalues. The principal subspace theorem was extended to the infinite dimensional realm by [46], who proved the theorem for Gaussian random functions.

A general form of the principal subspace theorem was proved by [45, Theorem 4.1] for distributions with a linear conditional expectation (which includes elliptical distributions) that states that if $\boldsymbol{Y}$ is self-consistent for $\boldsymbol{X}$ and the support of $\boldsymbol{Y}$ spans a linear subspace $\mathcal{S}$ of dimension $q < p$, then $\mathcal{S}$ is spanned by q eigenvectors of the covariance matrix of $\boldsymbol{X}$.

Since these publications, other principal subspace theorems have appeared in the literature for non-elliptical distributions, primarily mixture distributions. First, [49] proved a principal subspace result for $k = 2$ principal points for multivariate location mixtures of spherically symmetric distributions. This result for $k = 2$ was extended to a broader class of mixture distributions by [28]. [35] provided another extension for $k = 2$ for general location mixtures of spherically symmetric distributions. Following up on this, in [36], a principal subspace result was proved for an arbitrary number of principal points for mixtures of spherically symmetric distributions defined as

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{V} + \boldsymbol{U}, \tag{1}$$

where $\boldsymbol{V}$ is a spherically symmetric p-dimensional random vector with covariance matrix $\sigma^2\boldsymbol{I}$, and $\boldsymbol{U}$ is an arbitrary p-dimensional random vector with mean zero and finite second moments, independent of $\boldsymbol{V}$. They assume the support