Optimized Dimensionality Reduction Methods for Interval-Valued Variables and Their Application to Facial Recognition

Jorge Arce Garro [1] and Oldemar Rodríguez Rojas [2,*

[1]   National Bank of Costa Rica, 10101 San José, Costa Rica; jarceg@bncr.fi.cr
[2]   School of Mathematics, Research Center in Pure and Applied Mathematics (CIMPA), University of Costa Rica, 10101 San José, Costa Rica; oldemar.rodriguez@ucr.ac.cr
*   Correspondence: oldemar.rodriguez@ucr.ac.cr; Tel.: +506-2511-6606

Abstract: The center method, which was first proposed in *Rev. Stat. Appl.* 1997 by Cazes et al. and *Stat. Anal. Data Mining* 2011 by Douzal-Chouakria et al., extends the well-known Principal Component Analysis (PCA) method to particular types of symbolic objects that are characterized by multivalued interval-type variables. In contrast to classical data, symbolic data have internal variation. The authors who originally proposed the center method used the center of a hyper-rectangle in $\mathbb{R}^m$ as a base point to carry out PCA, followed by the projection of all vertices of the hyper-rectangles as supplementary elements. Since these publications, the center point of the hyper-rectangle has typically been assumed to be the best point for the initial PCA. However, in this paper, we show that this is not always the case, if the aim is to maximize the variance of projections or minimize the squared distance between the vertices and their respective projections. Instead, we propose the use of an optimization algorithm that maximizes the variance of the projections (or that minimizes the distances between the squares of the vertices and their respective projections) and finds the optimal point for the initial PCA. The vertices of the hyper-rectangles are, then, projected as supplementary variables to this optimal point, which we call the "Best Point" for projection. For this purpose, we propose four new algorithms and two new theorems. The proposed methods and algorithms are illustrated using a data set comprised of measurements of facial characteristics from a study on facial recognition patterns for use in surveillance. The performance of our approach is compared with that of another procedure in the literature, and the results show that our symbolic analyses provide more accurate information. Our approach can be regarded as an optimization method, as it maximizes the explained variance or minimizes the squared distance between projections and the original points. In addition, the symbolic analyses generate more informative conclusions, compared with the classical analysis in which classical surrogates replace intervals. All the methods proposed in this paper can be executed in the RSDA package developed in R.

Keywords: interval-valued variables; principal component analysis; symbolic data analysis; Best Point method

1. The Center Method

Symbolic data were introduced by Diday in [1]. In contrast to classical data analysis, in which a variable takes a single value, a variable in symbolic data can take a finite or infinite set of values: For example, an interval variable can take an infinite set of numerical values that range from low to high. As Principal Component Analysis (PCA) is one of the most popular multivariate methods for dimension reduction, its extension to symbolic data is important. Many generalizations of PCA have been developed and several studies have contributed to its extension to interval-valued data.

Among the methods for this in the literature, two are the vertex method and the center method [2–4]. In [5], the authors introduced new PCA techniques in order to visualize and compare the structures of interval data. Then, the authors of [6] proposed an approach that extended the classical PCA method to interval-valued data by using symbolic covariance to determine the principal component space to reflect the total variation in the interval-valued data. PCA has also been extended to histogram data in a number of studies (see [7–11]).

Most of these methods were developed for interval matrices, where an interval matrix X is defined as

$$
X = \begin{bmatrix}
[a_{11}, b_{11}] & [a_{12}, b_{12}] & \cdots & [a_{1m}, b_{1m}] \\
[a_{21}, b_{21}] & [a_{22}, b_{22}] & \cdots & [a_{2m}, b_{2m}] \\
\vdots & \vdots & \ddots & \vdots \\
[a_{n1}, b_{n1}] & [a_{n2}, b_{n2}] & \cdots & [a_{nm}, b_{nm}]
\end{bmatrix},
\tag{1}
$$

where $a_{ij} \leq b_{ij}$ for all $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$ (others authors denote the interval $\left[a_{ij}, b_{ij}\right]$ by $\left[x_{ij}^{lo}, x_{ij}^{up}\right]$ or $\left[\underline{x_{ij}}, \overline{x_{ij}}\right]$ for all $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$). An interval matrix can be considered a subset of a matrix $M_{n \times m}$, which we denote by $\mathbb{X}$, such that $\mathbb{X} = \left\{ Z \in M_{n \times m} \mid \forall i \in \{1, 2, ..., n\}, \forall j \in \{1, 2, ..., m\}, Z_{ij} \in \left[a_{ij}, b_{ij}\right] \right\}$. In this case, $Z \in \mathbb{X}$.

The center matrix of X is defined as

$$
X^c = \begin{bmatrix}
X_{11}^c & X_{12}^c & \cdots & X_{1m}^c \\
X_{21}^c & X_{22}^c & \cdots & X_{2m}^c \\
\vdots & \vdots & \ddots & \vdots \\
X_{n1}^c & X_{n2}^c & \cdots & X_{nm}^c
\end{bmatrix},
\tag{2}
$$

where

$$
X_{ij}^c = \frac{a_{ij} + b_{ij}}{2}.
\tag{3}
$$

Here, $X^c \in \mathbb{X}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$; X_{ij}^c is a real number and not an interval. Thus, the center matrix (2) is a classical matrix. The center principal component method starts from the center matrix; in other words, classical PCA is applied to the center matrix X^c. Then, the kth principal components of the centers are

$$
Y_{(k)}^c = X^c v_k^c,
\tag{4}
$$

where v_k^c is the kth eigenvector of the variance–covariance matrix of X^c, defined in Equation (2). For cases (rows) $i = 1, \ldots, n$, the kth principal component for an interval variable is constructed as follows. Let $Y_{ik}^c = \left[y_{ik}^{lo}, y_{ik}^{up}\right]$ be the interval principal component for an interval variable. Then,

$$
y_{ik}^{lo} = \sum_{j \in J_c^-} (b_{ij} - \bar{X}_{(j)}) v_{kj}^c + \sum_{j \in J_c^+} (a_{ij} - \bar{X}_{(j)}) v_{kj}^c,
\tag{5}
$$

$$
y_{ik}^{up} = \sum_{j \in J_c^-} (a_{ij} - \bar{X}_{(j)}) v_{kj}^c + \sum_{j \in J_c^+} (b_{ij} - \bar{X}_{(j)}) v_{kj}^c,
\tag{6}
$$

where $J_c^- = \{j | v_{kj}^c < 0\}$ and $J_c^+ = \{j | v_{kj}^c \geq 0\}$. More details can be found in [2].

The dual problem in the center PCA method was introduced by Rodriguez in [12]. To generalize duality relations, we let D be an interval matrix, defined as

$$D_{ij} = \left[\frac{a_{ij} - \bar{X}^c_{(j)}}{\sqrt{n}\sigma_{(j)}}, \frac{b_{ij} - \bar{X}^c_{(j)}}{\sqrt{n}\sigma_{(j)}} \right],$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, with $\bar{X}^c_{(j)}$ and $\sigma_{(j)}$ denoting the average and standard deviation of column j, respectively. The formulas shown in Theorem 1 are thus obtained and can be used to calculate the projections of interval variables.

Theorem 1. *If the hyper-rectangle defined by the jth column of D in the ith principal component is projected in the direction of v_i, then the minimum and maximum values can be computed by Equations (7) and (8), respectively.*

$$\underline{r_{ij}} = \sum_{k=1, v_{kj}<0}^{m} \bar{d}^c_{ki} v_{kj} + \sum_{k=1, v_{kj}>0}^{m} \underline{d}^c_{ki} v_{kj}, \tag{7}$$

$$\overline{r_{ij}} = \sum_{k=1, v_{kj}<0}^{m} \underline{d}^c_{ki} v_{kj} + \sum_{k=1, v_{kj}>0}^{m} \bar{d}^c_{ki} v_{kj}. \tag{8}$$

The proof of this theorem can be found in [12,13].

2. The Best Point Method

Let X be an $n \times m$ matrix of interval variables and let $Z \in \mathbb{X}$. If we apply PCA to a matrix Z, then the kth principal component of Z for an observation ξ_u, with $k = 1, \ldots, s < m$ and $i = 1, \ldots, m$, is

$$y^Z_{ik} = \sum_{j=1}^{m} (Z_{jk} - \bar{Z}_{(j)}) w^Z_{k_j}, \tag{9}$$

where $\bar{Z}_{(j)}$ is the average of the variable $Z_{(j)}$ (i.e., $\bar{Z}_{(j)} = \frac{1}{n} \sum_{i=1}^{n} Z_{ij}$), and $w^Z_k = (w^Z_{k_1}, \ldots, w^Z_{k_m})$ is the kth eigenvector associated with the variance–covariance matrix of Z. It is clear that $\beta(Z) = \{w^Z_1, \ldots, w^Z_m\}$ is an orthonormal basis of $\mathbb{R}^m$.

For the matrix X defined in (1), we let $X_i = ([a_{i1}, b_{i1}], \ldots, [a_{im}, b_{im}])$ for $i = 1, \ldots, n$. Then, we define the vertex matrix for an observation i as

$$X^v_i = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{im} \\ a_{i1} & a_{i2} & \ldots & b_{im} \\ \vdots & \vdots & \ddots & \vdots \\ b_{i1} & b_{i2} & \ldots & a_{im} \\ b_{i1} & b_{i2} & \ldots & b_{im} \end{bmatrix}. \tag{10}$$

Thus, the vertex matrix of X is

$$
X^v = \begin{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{11} & b_{12} & \dots & b_{1m} \end{bmatrix} \\ \vdots \;\; \vdots \;\; \vdots \;\; \vdots \\ \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{im} \\ \vdots & \vdots & \ddots & \vdots \\ b_{i1} & b_{i2} & \dots & b_{im} \end{bmatrix} \\ \vdots \;\; \vdots \;\; \vdots \;\; \vdots \\ \begin{bmatrix} a_{n1} & a_{n2} & \dots & a_{nm} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix} \end{bmatrix}.
\tag{11}
$$

Next, the rows of the vertex matrix of X are projected as supplementary elements in the PCA of Z. We define the supplementary vertex matrix as

$$
\widetilde{\mathbf{X}}_i^v(Z) = \begin{bmatrix} \frac{a_{i1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{a_{i2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} & \dots & \frac{a_{im}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \\ \frac{a_{i1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{a_{i2}-\bar{Z}_2}{\sqrt{n}\sigma_{(2)}} & \dots & \frac{a_{im}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b_{i1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{b_{i2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} & \dots & \frac{b_{im}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \end{bmatrix},
\tag{12}
$$

where $\sigma_{(j)}$ is the standard deviation of $Z_{(j)}$. To simplify this approach, we denote each row of the matrix $\widetilde{\mathbf{X}}_i^v(Z)$ by $\widetilde{x}_{i_{tj}}^v(Z)$, with $t = 1, \dots, 2^{m_i}$, in which m_i is the number of nontrivial intervals, and $j = 1, \dots, m$. Then, the co-ordinates are obtained:

$$
C^k(x_{i_j}^v) = \sum_{h=1}^{m} \widetilde{x}_{i_{jh}}^v(Z) w_{k_h},
\tag{13}
$$

with $j = 1, \dots, 2^{m_i}$, in which m_i is the number of nontrivial intervals. Then, the minimum and maximum of the interval can be calculated:

$$
\widetilde{Y}_{ik}^v = \widetilde{y}_{ik} = [\widetilde{y}_{ik}^{a_Z}, \widetilde{y}_{ik}^{b_Z}] \text{ with } k = 1, \dots, s < m,
\tag{14}
$$

and

$$
\widetilde{y}_{ik}^{a_Z} = \min_{j=1,\dots,2^{m_i}} C^k(x_{i_j}^v),
\tag{15}
$$

$$
\widetilde{y}_{ik}^{b_Z} = \max_{j=1,\dots,2^{m_i}} C^k(x_{i_j}^v).
\tag{16}
$$

The formulas in the following theorem allow us to compute Equations (15) and (16) much more quickly.

Theorem 2. *The co-ordinates of $\widetilde{Y}_{ik}^{v_Z}$ can be found as follows:*

$$
\widetilde{y}_{ik}^{a_Z} = \sum_{j \in J_Z^-} (b_{ij} - \overline{Z}_{(j)}) w_{k_j}^Z + \sum_{j \in J_Z^+} (a_{ij} - \overline{Z}_{(j)}) w_{k_j}^Z,
$$

$$\hat{y}_{ik}^{b_Z} = \sum_{j \in J_Z^-} (a_{ij} - \overline{Z}_{(j)}) w_{k_j}^Z + \sum_{j \in J_Z^+} (b_{ij} - \overline{Z}_{(j)}) w_{k_j}^Z,$$

where $J_Z^- = \{j | w_{kj}^v < 0\}$, $J_Z^+ = \{j | w_{kj}^v \geq 0\}$, and $\overline{Z}_{(j)}$ is the mean of the jth column.

Proof. Let $Z \in \mathbb{X}$; then,

$$\forall j, \forall i, Z_{ij} \in X_{ij} = \left[a_{ij}, b_{ij}\right]. \tag{17}$$

As a_{ij} and b_{ij} are supplementary elements in the PCA of Z, they must first be centered with respect to the columns (variables) of Z. Thus,

$$\forall i, \forall j, a_{ij} - \overline{Z}_{(j)}, b_{ij} - \overline{Z}_{(j)}. \tag{18}$$

Then, from Equations (17) and (18),

$$\forall i, \forall j, z_{ij} - \overline{Z}_{(j)} \in X_{ij} - \overline{Z}_{(j)} = \left[a_{ij} - \overline{Z}_{(j)}, b_{ij} - \overline{Z}_{(j)}\right]. \tag{19}$$

- **Case 1:** $\forall j, w_{k_j}^Z > 0$,

$$\forall j, \forall i, w_{k_j}^Z \left(z_{ij} - \overline{Z}_{(j)}\right) \in w_{k_j}^Z \left[a_{ij} - \overline{Z}_{(j)}, b_{ij} - \overline{Z}_{(j)}\right]$$

$$\Rightarrow \quad \sum_{j=1}^p \left(a_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j=1}^p \left(z_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j=1}^p \left(b_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z.$$

- **Case 2:** $\forall j, w_{kj}^Z < 0$,

$$\forall j, \forall i, w_{k_j}^Z \left(z_{ij} - \overline{Z}_{(j)}\right) \in w_{k_j}^Z \left[b_{ij} - \overline{Z}_{(j)}, a_{ij} - \overline{Z}_{(j)}\right]$$

$$\Rightarrow \quad \sum_{j=1}^p \left(b_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j=1}^p \left(z_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j=1}^p \left(a_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z.$$

- **Case 3:** Let $J_Z^- = \{j | w_{k_j} < 0\}$ and $J_Z^+ = \{j | w_{k_j} \geq 0\}$.
  For **Case 1** applied to J_Z^+,

$$\sum_{j \in J_Z^+} \left(a_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j \in J_Z^+} \left(z_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j \in J_Z^+} \left(b_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z. \tag{20}$$

For **Case 2** applied to J_v^-,

$$\sum_{j \in J_Z^+} \left(b_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j \in J_Z^+} \left(z_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z \leq \sum_{j \in J_Z^+} \left(a_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z. \tag{21}$$

Therefore, from Equations (20) and (21), we obtain

$$\sum_{j \in J_Z^-} \left(b_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z + \sum_{j \in J_Z^+} \left(a_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z$$

$$\leq \quad \sum_{j=1}^n \left(z_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z$$

$$\leq \quad \sum_{j \in J_Z^-} \left(a_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z + \sum_{j \in J_Z^+} \left(b_{ij} - \overline{Z}_{(j)}\right) w_{k_j}^Z.$$

Therefore,

$$\widehat{y}_{ik}^{a_Z} = \sum_{j \in J_Z^-} \left(b_{ij} - \overline{Z}_{(j)} \right) w_{k_j}^Z + \sum_{j \in J_Z^+} \left(a_{ij} - \overline{Z}_{(j)} \right) w_{k_j}^Z,$$

$$\widehat{y}_{ik}^{b_Z} = \sum_{j \in J_Z^-} \left(a_{ij} - \overline{Z}_{(j)} \right) w_{k_j}^Z + \sum_{j \in J_Z^+} \left(b_{ij} - \overline{Z}_{(j)} \right) w_{k_j}^Z.$$

□

The following theorem provides the co-ordinates in the variable space; this is a dual relationship. We need to center and standardize the matrix Z.

$$\tilde{Z}_{ij} = \frac{Z_{ij} - \bar{Z}_{(j)}}{\sqrt{n}\sigma_{(j)}}.$$

Next, we next focus on the matrix $\tilde{Z} = \tilde{Z}_{ij} \ \forall i, \forall j$. Let $\tilde{z}^j$ be the jth column of $\tilde{Z}$, with $(\tilde{z}^j)^t \cdot \tilde{z}^i = R(i, j) \leq 1$. Then, the interval matrix is centered and standardized with respect to Z:

$$\tilde{\mathbf{X}}(Z) = \begin{bmatrix} \left[\frac{a_{11}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}}, \frac{b_{11}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} \right] & \left[\frac{a_{12}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}}, \frac{b_{12}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} \right] & \cdots & \left[\frac{a_{1m}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}}, \frac{b_{1m}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \right] \\ \vdots & \vdots & \ddots & \vdots \\ \left[\frac{a_{i1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}}, \frac{b_{i1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} \right] & \left[\frac{a_{i2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}}, \frac{b_{i2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} \right] & \cdots & \left[\frac{a_{im}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}}, \frac{b_{im}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \right] \\ \vdots & \vdots & \ddots & \vdots \\ \left[\frac{a_{n1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}}, \frac{b_{n1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} \right] & \left[\frac{a_{n2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}}, \frac{b_{n2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} \right] & \cdots & \left[\frac{a_{nm}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}}, \frac{b_{nm}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \right] \end{bmatrix}. \tag{22}$$

To facilitate the analysis, we define

$$(\tilde{\mathbf{X}}(Z))_{ij} = \left[\frac{a_{ij} - \bar{Z}_{(j)}}{\sqrt{n}\sigma_{(j)}}, \frac{b_{ij} - \bar{Z}_{(j)}}{\sqrt{n}\sigma_{(j)}} \right] = [a_{ij}^Z, b_{ij}^Z].$$

The inertia matrix $\tilde{Z}\tilde{Z}^t$ is symmetric and positive semidefinite, so all its eigenvectors are orthogonal and its eigenvalues are real and nonnegative. We let $v_1^Z, v_2^Z, \ldots, v_s^Z$ denote the s eigenvectors of $\tilde{Z}\tilde{Z}^t$ associated with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s \geq 0$. Then, $V(Z) = \left[v_1^Z | v_2^Z | \ldots | v_s^Z \right]$ is defined as a matrix of the size $n \times s$ whose columns are the eigenvectors of $\tilde{Z}\tilde{Z}^t$. We can compute the co-ordinates of the variables in the correlation circle as $\tilde{Z}^t V$, and we can then compute the ith column of Z in the jth principal component (in the v_j^Z direction) using Equation (23).

$$r_{ij}^Z = \sum_{k=1}^m \tilde{Z}_{ki} v_{kj}^Z. \tag{23}$$

The next theorem proves the duality relation of any matrix that belongs to an interval matrix.

Theorem 3. *If the hyper-rectangle defined by the jth column of $\tilde{\mathbf{X}}(Z)$ in the ith principal component is projected in the direction of v_i, then the maximum and minimum values can be obtained by Equations (24) and (25), respectively.*

$$\underline{r_{ij}} = \sum_{k=1, v_{kj}<0}^m b_{ki}^Z v_{kj} + \sum_{k=1, v_{kj}>0}^m a_{ki}^Z v_{kj}, \tag{24}$$

$$\overline{r_{ij}} = \sum_{k=1, v_{kj}<0}^m a_{ki}^Z v_{kj} + \sum_{k=1, v_{kj}>0}^m b_{ki}^Z v_{kj}. \tag{25}$$

Proof. The proof of this theorem is similar to the proof of Theorem 2. □

In the above, we prove that $p\widehat{z}_{ij} \in [\underline{r_{ij}}, \overline{r_{ij}}]$ and that $\underline{r_{ij}}$ and $\overline{r_{ij}}$ are a combination of the projections of the vertices of the hyper-rectangle $\mathbb{R}^m$. We can form duality relations between the eigenvectors of $\tilde{Z}\tilde{Z}^t$ and $\tilde{Z}^t\tilde{Z}$: Both matrices have the same s positive eigenvalues $\lambda_1^Z, \lambda_2^Z, \dots, \lambda_s^Z$, and if $u_1^Z, u_2^Z, \dots, u_s^Z$ are the first s eigenvectors of $\tilde{Z}^t\tilde{Z}$, then the relations between the eigenvectors of $\tilde{Z}\tilde{Z}^t$ and $\tilde{Z}^t\tilde{Z}$ can be computed by Equations (26) and (27):

$$u_\ell^Z = \frac{\tilde{Z}^t v_\ell^Z}{\sqrt{\lambda_\ell}} \text{ for } \ell = 1, 2, \dots, s. \tag{26}$$

$$v_\ell^Z = \frac{\tilde{Z} u_\ell^Z}{\sqrt{\lambda_\ell}} \text{ for } \ell = 1, 2, \dots, s. \tag{27}$$

Above, we provide the theory to apply PCA to all matrices $Z \in \mathbb{X}$. Now, we aim to find a matrix $Z^* \in \mathbb{X}$ that is optimal for one of two criteria: (1) The minimization of the square of the distance from the vertices of the hypercubes to the principal axes of Z, or (2) the maximization of the variance of the first components of Z. We develop these concepts in the following two sections.

2.1. Minimizing the Square of the Distance from the Hypercube Vertices to the Principal Axes of Z

Let X be an interval of an $n \times m$ matrix, $Z \in \mathbb{X}$, and $\beta(Z) = \{w_1^Z, \dots, w_s^Z\}$, with $s \leq m$ and w_i^Z eigenvectors of the variance–covariance matrix of Z. We let X^v denote the vertex matrix of X and $N = \sum_{i=1}^n 2^{m_i}$, in which m_i is the number of nontrivial intervals for case ξ_i. Then, the centered and standardized vertex matrix with respect to Z has the following form:

$$\tilde{X}^v(Z) = \begin{bmatrix} \begin{bmatrix} \frac{a_{11}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{a_{12}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} & \cdots & \frac{a_{1m}-\bar{Z}_{(m)}}{\sigma_{(m)}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{b_{11}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{b_{12}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} & \cdots & \frac{b_{1m}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \end{bmatrix} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ \begin{bmatrix} \frac{a_{n1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{a_{n2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} & \cdots & \frac{a_{nm}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{b_{n1}-\bar{Z}_{(1)}}{\sqrt{n}\sigma_{(1)}} & \frac{b_{n2}-\bar{Z}_{(2)}}{\sqrt{n}\sigma_{(2)}} & \cdots & \frac{b_{nm}-\bar{Z}_{(m)}}{\sqrt{n}\sigma_{(m)}} \end{bmatrix} \end{bmatrix}. \tag{28}$$

Let $\varphi(Z) : \mathbb{X} \to \mathbb{R}^+ \cup \{0\}$ be the function defined by

$$\varphi(Z) = \sum_{i=1}^N \left\| \tilde{X}_i^v(Z) - Pr_{\beta(Z)}(\tilde{X}_i^v(Z)) \right\|^2, \tag{29}$$

where $\|.\|$ is the Euclidean norm. To compute $\varphi(Z)$, we propose Algorithm 1:

Algorithm 1 The computation of φ.

Require: X an $n \times m$ matrix of intervals, $Z \in \mathbb{X}$, s number of principal components.
Ensure: $\varphi(Z)$.
 1: Apply PCA to Z.
 2: $\beta = \{w_1, \ldots, w_s\}$, with $s \le m$ and w_i eigenvectors of the variance-covariance matrix of Z.
 3: Compute the vertex matrix of X (X^v).
 4: Compute the vertex matrix of the centered and standardized X with respect to Z $(\tilde{X}^v(Z))$.
 5: $\varphi(Z) = \sum_{i=1}^{N} \left\| \tilde{X}_i^v(Z) - Pr_{\beta(Z)}(\tilde{X}_i^v(Z)) \right\|^2$.
 6: return $\varphi(Z)$.

As $Z \in \mathbb{X}$, $\mathbb{X}$ is a finite union of compact sets and $\varphi(Z)$ is a continuous function, φ always reaches the minimum and the maximum. In this case, the aim is to obtain the matrix Z that minimizes the distance to the vertex matrix X^v. The problem that we aim to solve is

$$
\begin{aligned}
\text{Minimize} \quad & \varphi(Z) = \sum_{i=1}^{N} \left\| \tilde{X}_i^v(Z) - Pr_{\beta(Z)}(\tilde{X}_i^v(Z)) \right\|^2 \\
\text{Subject to} \quad & Z \in \mathbb{X}.
\end{aligned}
\tag{30}
$$

Definition 1. *The matrix $Z \in \mathbb{X}$ that solves Problem 30 is the optimal matrix with respect to distance, which is denoted by Z^φ.*

To perform the optimization that computes Z^φ, we propose Algorithm 2:

Algorithm 2 Computation of the Best Matrix with respect to the distances of the vertices.

Require: X a symbolic matrix of intervals of dimension $n \times m$, $Z \in \mathbb{X}$, s number of principal components, TOL is the variation tolerance between iterations, and N is the maximum number of iterations.
Ensure: $Y^{\tilde{V}_{Z^\varphi}}$.
 1: Consider $Z = X^c$, the center matrix 2, to be the initial value.
 2: Get Z^φ by means of optimization algorithm $\left(\text{initialvalue} = Z, \text{function} = \varphi(Z), TOL, N \right)$.
 3: Get $Y^{\tilde{V}_{Z^\varphi}}$ Use Theorem 2.
 4: return $Y^{\tilde{V}_{Z^\varphi}}$.

2.2. Maximizing the Variance of the First Components

Let X be an interval matrix of dimension $n \times m$, $Z \in \mathbb{X}$, and $\beta(Z) = \{w_1^Z, \ldots, w_s^Z\}$, with $s \le m$ and w_i^Z eigenvectors of the variance–covariance matrix of Z and $\lambda(Z) = \lambda_1^Z, \ldots, \lambda_s^Z$ denoting the set of associated eigenvalues of the variance–covariance matrix of Z. We define the function $\Lambda(Z, s)$: $\mathbb{X} \times \mathbb{N} \to \mathbb{R}^+$ as $\Lambda(Z, s) = \sum_{i=1}^{s} \lambda_i^Z$. To compute $\Lambda(Z, s)$, we propose Algorithm 3:

Algorithm 3 The computation of Λ.

Require: X an $n \times m$ symbolic matrix of intervals of dimension $Z \in \mathbb{X}$, s number of principal components.
Ensure: $\Lambda(Z, s)$.
 1: Apply PCA to Z.
 2: $\lambda(Z) = \lambda_1^Z, \ldots, \lambda_s^Z$ set of associated eigenvalues of the variance-covariance matrix of Z.
 3: $\Lambda(Z, s) = \sum_{i=1}^{s} \lambda_i^Z$.
 4: return $\Lambda(Z, s)$.

As above, since $Z \in \mathbb{X}$ and $\mathbb{X}$ is the finite union of compact sets with s number of principal components, $\Lambda(Z,s)$ is a continuous function and, thus, always reaches the minimum and the maximum. In this case, the aim is to obtain the matrix Z that maximizes the accumulated inertia in the first s principal components. The problem that we want to solve is

$$
\begin{aligned}
\text{Maximize} \quad & \Lambda(Z,s) = \sum_{i=1}^{s} \lambda_i^Z \\
\text{Subject to} \quad & Z \in \mathbb{X}.
\end{aligned}
\tag{31}
$$

Definition 2. *The matrix $Z \in \mathbb{X}$ that solves Problem 31 is the optimal matrix with respect to inertia, denoted by Z^Λ.*

To perform the optimization that computes Z^Λ, we propose Algorithm 4:

Algorithm 4 The computation of the Best Matrix with respect to inertia.

Require: X an $n \times m$ `symbolic matrix of intervals of dimension,` $Z \in \mathbb{X}$, s `number of`
`   principal components.`
`Ensure:` $\tilde{Y}^{V}_{Z^\Lambda}$.
 1: `Consider` $Z = X^c$, `center matrix` 2`, as the initial value.`
 2: `Get` Z^Λ `by means of the optimization algorithm` $\left(\text{initialvalue} = Z, \text{function} = \Lambda(Z,s)\right)$.
 3: `Get` $\tilde{Y}^{V}_{Z^\Lambda}$ `using Theorem` 2`.`
 4: `return` $Y^{\tilde{V}}_{Z^\Lambda}$.

---

3. Experimental Evaluation: The Application to Facial Recognition

Automatic facial recognition has recently gained momentum, especially in the context of security issues such as access to buildings, and in the context of monitoring and continued surveillance. A well-known application of facial recognition is its incorporation in the iPhone X. According to Apple's support website, the technology that enables facial ID is some of the most advanced hardware and software that has ever been created. The TrueDepth camera captures accurate facial data by projecting and analyzing over 30,000 invisible dots to create a facial depth map, while also capturing an infrared image. These images are transformed into a mathematical representation, which is compared with registered facial data. In both R and Python, a significant number of libraries, such as the `videoplayR` package, have also been developed for facial recognition. The link below contains more details: http://www.stoltzmaniac.com/facial-recognition-in-r/.

As described in this section, we applied all the proposed methods using $p = 6$ interval-valued variables in a data set of $m = 27$ faces for a total of 27,000 photos. The data set was taken from [14], in which the authors investigated facial characteristics for detection purposes in a surveillance study. In this study, the center PCA method in [4] was applied, as shown in Figure 1. The data are provided in Table 1.

The data set contains measurements of $p = 6$ random variables designed to identify each face: The length spanned by the eyes X_1 (distance AD in Figure 1), the length between the eyes X_2 (distance BC), the length from the outer right eye to the upper-middle lip at point H between the nose and mouth X_3 (distance AH), the corresponding length for the left eye X_4 (DH), the length from point H to the outside of the mouth on the right side X_5 (EH), and the corresponding distance to the left side of the mouth X_6 (GH). For each facial image in this facial recognition process, salient features, such as the nose, mouth, and eyes, are located using morphological operators. The boundaries of the located elements are extracted by using a specific active contour method based on Fourier descriptors, which incorporates information about the global shape of each object. Finally, the specific points delimiting each extracted boundary are located, and the distance is measured between a specific pair of points, as represented by the random variables in Figure 1. This distance measure is expressed as

the number of pixels in a facial image. As there is a sequence of such images, the actual measured distances are interval-valued variables. Thus, for example, the eye span distance X_1 for case HUS1 is $X_1 = [168.86, 172.84]$ for this series of images. Notably, different conditions of alignment, illumination, pose, and occlusion cause variation in the distances extracted from different images of the same person. The study that generated the data set involved nine men and three sequences for each subject for a total of $m = 27$ cases. The complete data set is provided in Table 1.
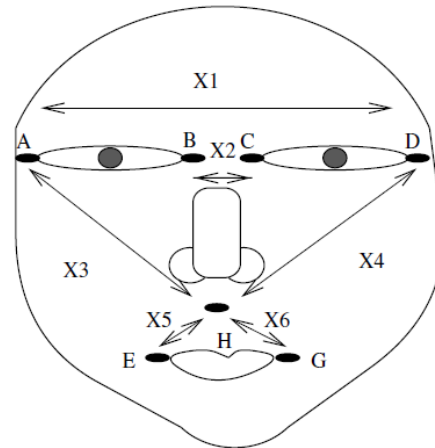

Figure 1. Random variables for facial description.

Table 1. Faces data set.

Case	$X_1 = AD$	$X_2 = BC$	$X_3 = AH$	$X_4 = DH$	$X_5 = EH$	$X_6 = GH$
FRA1	[155.00, 157.00]	[58.00, 61.01]	[100.45, 103.28]	[105.00, 107.30]	[61.40, 65.73]	[64.20, 67.80]
FRA2	[154.00, 160.01]	[57.00, 64.00]	[101.98, 105.55]	[104.35, 107.30]	[60.88, 63.03]	[62.94, 66.47]
FRA3	[154.01, 161.00]	[57.00, 63.00]	[99.36, 105.65]	[101.04, 109.04]	[60.95, 65.60]	[60.42, 66.40]
HUS1	[168.9,172.84]	[58.55,63.39]	[102.83,106.53]	[122.38,124.52]	[56.73,61.07]	[60.44,64.54]
HUS2	[169.8,175.03]	[60.21,64.38]	[102.94,108.71]	[120.24,124.52]	[56.73,62.37]	[60.44,66.84]
HUS3	[168.8,175.15]	[61.4,63.51]	[104.35,107.45]	[120.93,125.18]	[57.2,61.72]	[58.14,67.08]
INC1	[155.3,160.45]	[53.15,60.21]	[95.88,98.49]	[91.68,94.37]	[62.48,66.22]	[58.9,63.13]
INC2	[156.3,161.31]	[51.09,60.07]	[95.77,99.36]	[91.21,96.83]	[54.92,64.2]	[54.41,61.55]
INC3	[154.5,160.31]	[55.08,59.03]	[93.54,98.98]	[90.43,96.43]	[59.03,65.86]	[55.97,65.8]
ISA1	[164,168]	[55.01,60.03]	[120.28,123.04]	[117.52,121.02]	[54.38,57.45]	[50.8,53.25]
ISA2	[163,170]	[54.04,59]	[118.8,123.04]	[116.67,120.24]	[55.47,58.67]	[52.43,55.23]
ISA3	[164,169.01]	[55,59.01]	[117.38,123.11]	[116.67,122.43]	[52.8,58.31]	[52.2,55.47]
JPL1	[167.1,171.19]	[61.03,65.01]	[118.23,121.82]	[108.3,111.2]	[63.89,67.88]	[57.28,60.83]
JPL2	[169.1,173.18]	[60.07,65.07]	[118.85,120.88]	[108.98,113.17]	[62.63,69.07]	[57.38,61.62]
JPL3	[169,170.11]	[59.01,65.01]	[115.88,121.38]	[110.34,112.49]	[61.72,68.25]	[59.46,62.94]
KHA1	[149.3,155.54]	[54.15,59.14]	[111.95,115.75]	[105.36,111.07]	[54.2,58.14]	[48.27,50.61]
KHA2	[149.3,155.32]	[52.04,58.22]	[111.2,113.22]	[105.36,111.07]	[53.71,58.14]	[49.41,52.8]
KHA3	[150.3,157.26]	[52.09,60.21]	[109.04,112.7]	[104.74,111.07]	[55.47,60.03]	[49.2,53.41]
LOT1	[152.6,157.62]	[51.35,56.22]	[116.73,119.67]	[114.62,117.41]	[55.44,59.55]	[53.01,56.6]
LOT2	[154.6,157.62]	[52.24,56.32]	[117.52,119.67]	[114.28,117.41]	[57.63,60.61]	[54.41,57.98]
LOT3	[154.8,157.81]	[50.36,55.23]	[117.59,119.75]	[114.04,116.83]	[56.64,61.07]	[55.23,57.8]
PHI1	[163.1,167.07]	[66.03,68.07]	[115.26,119.6]	[116.1,121.02]	[60.96,65.3]	[57.01,59.82]
PHI2	[164,168.03]	[65.03,68.12]	[114.55,119.6]	[115.26,120.97]	[60.96,67.27]	[55.32,61.52]
PHI3	[161,167]	[64.07,69.01]	[116.67,118.79]	[114.59,118.83]	[61.52,68.68]	[56.57,60.11]
ROM1	[167.2,171.24]	[64.07,68.07]	[123.75,126.59]	[122.92,126.37]	[51.22,54.64]	[49.65,53.71]
ROM2	[168.2,172.14]	[63.13,68.07]	[122.33,127.29]	[124.08,127.14]	[50.22,57.14]	[49.93,56.94]
ROM3	[167.1,171.19]	[63.13,68.03]	[121.62,126.57]	[122.58,127.78]	[49.41,57.28]	[50.99,60.46]

It is important to note that the data in Table 1 are aggregated. There are 27 interval-valued cases; if each case is drawn from a sequence of 1000 images, then there are 27,000 classical point observations in $\mathbb{R}^6$. An underlying assumption of the standard classical analysis is that all 27,000 observations are independent. However, this is not the case in this data set. The data values for each face form a set of 1000 dependent observations. Therefore, if we were to use each image as a statistical unit by performing classical analysis, then we would lose information about the dependence contained in the 27,000 observations. The resulting principal component analysis would look for axes that maximize the variability across all 27,000 images, regardless of whether some images belong to the same sequence. In contrast, as interval-valued observations are obtained from each sequence, the Best Point method extracts the principal component axes that maximize the variability in each interval (i.e., those that maximize the internal variability), thereby retaining information on the dependency among the 1000 images in each sequence.

*Comparison between the Center, Vertex, and Best Point Methods*

We applied the vertex, center, and best point principal component methods to the data in Table 1. The Best Point principal component method was run with two different goals: (1) To minimize the squared distance and (2) to maximize the variance. From this point, the Best Point principal component method that minimizes the squared distance was designated as the Best Point Distance, and the Best Point principal component method that maximizes the variance was designated as the Best Point Variance. Table 2 shows the first two principal components generated by the four methods.

Table 2. Comparison of the first two principal components from the four methods.

Cases	Vertex		Center		Best Point Distance		Best Point Variance	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
FRA1	[1.61,2.66]	[0.27,1.57]	[-2.97,-1.75]	[0.24,1.72]	[−2.21,−1.39]	[0.24,1.25]	[−2.83,−1.57]	[0.45,1.87]
FRA2	[1.03,2.49]	[−0.11,1.61]	[−2.73,−1.11]	[−0.2,1.79]	[−2.09,−0.93]	[−0.06,1.29]	[−2.57,−0.92]	[0.01,1.95]
FRA3	[0.81,2.99]	[−0.4,1.88]	[−3.29,−0.86]	[−0.52,2.08]	[−2.49,−0.73]	[−0.29,1.48]	[−3.15,−0.73]	[−0.31,2.21]
HUS1	[−1.1,0.24]	[0.39,2.05]	[−0.37,1.16]	[0.38,2.28]	[−0.19,0.86]	[0.28,1.58]	[−0.24,1.34]	[0.47,2.31]
HUS2	[−1.41,0.4]	[0.56,2.65]	[−0.58,1.48]	[0.59,2.98]	[−0.33,1.11]	[0.42,2.04]	[−0.42,1.66]	[0.66,2.97]
HUS3	[−1.42,0.24]	[0.43,2.52]	[−0.4,1.51]	[0.46,2.82]	[−0.21,1.13]	[0.33,1.94]	[−0.22,1.63]	[0.53,2.8]
INC1	[2.29,3.77]	[−0.67,1.23]	[−4.07,−2.38]	[−0.9,1.29]	[−3.11,−1.95]	[−0.53,0.95]	[−3.97,−2.23]	[−0.55,1.56]
INC2	[1.35,3.66]	[−2.05,0.92]	[−3.91,−1.21]	[−2.51,0.93]	[−3.02,−1.22]	[−1.63,0.7]	[−3.8,−0.97]	[−2.02,1.22]
INC3	[1.86,4.02]	[−1.2,1.41]	[−4.33,−1.84]	[−1.5,1.47]	[−3.34,−1.61]	[−0.94,1.08]	[−4.14,−1.67]	[−1.08,1.77]
ISA1	[−2.01,−0.8]	[−1.83,−0.46]	[0.88,2.24]	[−2.06,−0.48]	[0.66,1.61]	[−1.42,−0.34]	[0.75,2.15]	[−2.11,−0.59]
ISA2	[−1.86,−0.37]	[−1.71,−0.08]	[0.39,2.04]	[−1.93,−0.07]	[0.32,1.51]	[−1.32,−0.07]	[0.26,1.95]	[−1.99,−0.19]
ISA3	[−2.11,−0.41]	[−1.84,−0.12]	[0.44,2.36]	[−2.09,−0.11]	[0.34,1.7]	[−1.43,−0.09]	[0.35,2.32]	[−2.1,−0.21]
JPL1	[−0.92,0.36]	[0.54,2.03]	[−0.57,0.89]	[0.67,2.38]	[−0.29,0.73]	[0.43,1.59]	[−0.7,0.79]	[0.49,2.14]
JPL2	[−1.17,0.34]	[0.48,2.37]	[−0.59,1.17]	[0.58,2.76]	[−0.25,0.93]	[0.37,1.84]	[−0.76,1.1]	[0.43,2.47]
JPL3	[−0.93,0.52]	[0.5,2.28]	[−0.78,0.92]	[0.59,2.64]	[−0.4,0.73]	[0.38,1.78]	[−0.89,0.91]	[0.45,2.43]
KHA1	[−0.39,1.18]	[−3.08,−1.46]	[−1.1,0.64]	[−3.51,−1.65]	[−1.01,0.25]	[−2.34,−1.09]	[−1.21,0.58]	[−3.32,−1.56]
KHA2	[−0.15,1.46]	[−3.17,−1.32]	[−1.43,0.39]	[−3.65,−1.5]	[−1.23,0.05]	[−2.42,−0.98]	[−1.52,0.37]	[−3.41,−1.39]
KHA3	[−0.25,1.71]	[−2.95,−0.72]	[−1.73,0.47]	[−3.39,−0.82]	[−1.42,0.13]	[−2.26,−0.52]	[−1.83,0.43]	[−3.18,−0.72]
LOT1	[−0.61,0.74]	[−2.51,−0.87]	[−0.79,0.74]	[−2.86,−0.98]	[−0.6,0.47]	[−1.91,−0.64]	[−0.94,0.66]	[−2.82,−1.01]
LOT2	[−0.4,0.69]	[−1.94,−0.62]	[−0.77,0.47]	[−2.2,−0.44]	[−0.55,0.31]	[−1.47,−0.44]	[−0.91,0.36]	[−2.21,−0.75]
LOT3	[−0.34,0.82]	[−2.12,−0.7]	[−0.93,0.41]	[−2.44,−0.78]	[−0.64,0.26]	[−1.63,−0.51]	[−1.09,0.33]	[−2.41,−0.85]
PHI1	[−1.51,−0.22]	[0.56,1.84]	[0.11,1.57]	[0.72,2.18]	[0.14,1.19]	[0.49,1.47]	[0.05,1.53]	[0.59,1.97]
PHI2	[−1.66,0.09]	[0.33,2.29]	[−0.27,1.74]	[0.45,2.69]	[−0.1,1.3]	[0.3,1.82]	[−0.35,1.67]	[0.32,2.46]
PHI3	[−1.38,0.25]	[0.25,2.25]	[−0.45,1.44]	[0.36,2.67]	[−0.22,1.07]	[0.25,1.8]	[−0.62,1.4]	[0.24,2.39]
ROM1	[−3.45,−2.19]	[−1.2,0.29]	[2.41,3.84]	[−1.27,0.44]	[1.74,2.74]	[−0.89,0.26]	[2.39,3.84]	[−1.38,0.26]
ROM2	[−3.63,−1.85]	[−1.3,0.97]	[1.96,4.04]	[−1.39,1.2]	[1.49,2.89]	[−0.98,0.79]	[1.94,4.06]	[−1.5,0.99]
ROM3	[−3.57,−1.48]	[−1.33,1.31]	[1.53,3.98]	[−1.44,1.58]	[1.17,2.83]	[−1,1.06]	[1.57,4.04]	[−1.5,1.39]

Figure 2 compares the data in Table 1 with the principal planes generated by the vertex, center, Best Point Distance, and Best Point Variance principal component methods. The plots of these, along with the first principal component (PC1) and second principal component (PC2) axes, are shown in Figure 2. The proximity of the three sequences for the three faces for each individual can be readily observed, which validates their within-subject coherence. Furthermore, with all four methods, the same four classes of faces were distinguished. Faces {INC, FRA} can be regarded as one class, faces {HUS, PHI, JPL} and {ISA, ROM} constitute two other classes, and faces {LOT, KHA} form the fourth class. Of the four principal planes in the graph, those corresponding to the Best Point Distance and

Best Point Variance methods show much better separation of the face classes, which results in superior facial classification. In other words, the proposed methods more accurately predicted the individual in the photo.
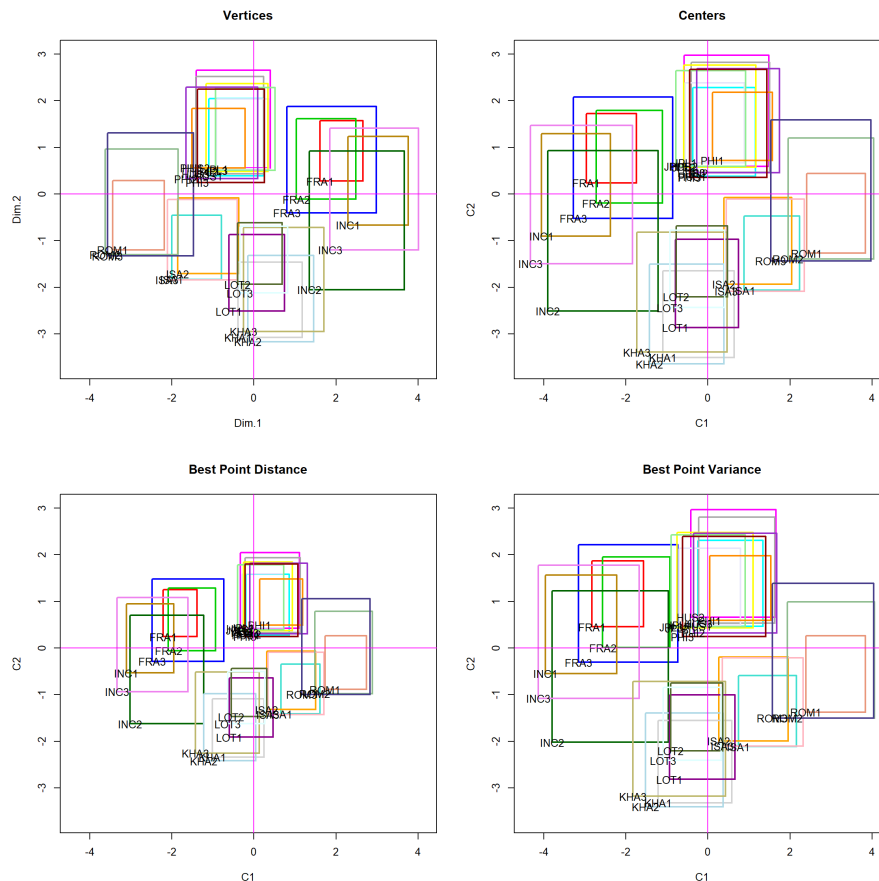
Figure 2. Principal component analysis (PCA) comparison.

Numerical analysis results confirm that the separation of classes from the Best Point Distance and Best Point Variance methods was much better than that from the other methods. Table 3 compares the accumulated variance of the vertex, center, Best Point Distance, and Best Point Variance principal component methods. The better methods are clearly Best Point Distance and Best Point Variance, which, in the third principal component, reached 91.25% and 99.72% of the accumulated variance, respectively; both of which were far superior to the results of the center and vertex methods.

As shown in Table 4, for the criterion of the minimum distance between the corners of the original hyper-rectangles in $\mathbb{R}^m$ and the principal components, the Best Point Distance method outperformed the other methods, with a minimum distance of 6676.43. This distance was significantly less than the distances obtained by the other methods.

In Table 5, we show the correlation of the original variables and the first principal component generated by the vertex, center, Best Point Distance, and Best Point Variance methods. For all variables except for X_6, the correlation was stronger with Best Point Distance or Best Point Variance. It can be inferred that the original variables were better represented by the PC1 component of the Best Point Distance and Best Point Variance methods. Interestingly, the correlation of the PC1 component from the Best Point Distance and Best Point Variance methods was stronger with the variables for the upper part of the face. The same result was generated if the correlations of the original variables were analyzed with respect to the other principal components of Best Point Distance and Best Point Variance.

Table 3. Comparison of the variance resulting from different methods using facial recognition data.

	Vertex	Center	Best Point Distance	Best Point Variance
PC1	42.67%	46.47%	45.49%	56.01%
PC2	72.64%	80.53%	81.05%	88.31%
PC3	83.35%	89.65%	91.25%	99.72%
PC4	91.28%	95.06%	95.80%	99.85%
PC5	96.86%	98.96%	99.28%	99.97%
PC6	100.00%	100.00%	100.00%	100.00%

Table 4. Comparison of the distances resulting from different methods in facial recognition data.

Vertex	Center	Best Point Distance	Best Point Variance
10368.00	12719.64	6676.43	12457.09

Table 5. Correlation between the first component of each method and the variables.

	$X_1 = AD$	$X_2 = BC$	$X_3 = AH$	$X_4 = DH$	$X_5 = EH$	$X_6 = GH$
Vertex-PC1	0.64	0.49	0.84	0.89	-0.47	-0.43
Center-PC1	0.61	0.47	0.83	0.88	-0.52	-0.47
BestPointDistance-PC1	0.65	0.48	0.84	0.90	-0.46	-0.43
BestPointVariance-PC1	0.63	0.49	0.80	0.88	-0.55	-0.43

4. Conclusions

This work focused on improving the center and vertex principal component methodology for interval-valued data. Compared with classical methods, symbolic methods based on interval-valued variables have important advantages, such as improved computational complexity due to reduced execution times, as small data tables are used. For example, for the facial recognition example, a table was passed from 27,000 cases to only 1 of 27 cases. In addition, symbolic methods allow for much better handling and interpretation of data variability. In the facial recognition scenario, the variation in the distances that measure different variables from one photo to another of the same person (such as the variation in the distance of the eyes X_1 from one photo to another) was due to the variation in the angle at which the photo was taken.

The Best Point methods proposed in this paper considerably improved both the center method and the vertex method. This is because Best Point Variance maximized the variance explained by the components and Best Point Distance minimized the squared distance between the vertices of the hyper-rectangles and their respective projections. As shown in the tables above, this led to a substantial improvement in all the quality indices used in principal component analysis. The result is better data clustering and, therefore, better prediction.

In future works, a consensus between the Best Point Variance and Best Point Distance methods could be constructed by applying a multiobjective optimization method to the functions φ and Λ. Finally, all the proposed algorithms for executing symbolic analyses of interval-valued data are available in the RSDA package in R (see [15]).

References

1. Diday, E. Introduction a lApproache Symbolique en Analyse des Données. *RAIRO Oper. Res.* **1989**, *23*, 193–236. [CrossRef]

2.   Billard, L.; Diday, E. *Symbolic Data Analysis: Conceptual Statistics and Data Mining*; John Wiley & Sons Ltd: Chichester, UK, 2006.

3.   Cazes, P.; Chouakria, A.; Diday, E.; Schektman, Y. Extension de l'analyse en com-posantes principales á des données de type intervalle. *Rev. Stat. Appl.* **1997**, *3*, 5–24.

4.   Douzal-Chouakria, A.; Billard, L.; Diday, E. Principal component analysis for interval-valued observations. *Stat. Anal. Data Min.* **2011**, *4*, 229–246. [CrossRef]

5.   Lauro, C.; Palumbo, F. Principal Component Analysis of Interval Data: A Symbolic Data Analysis Approach. *Comput. Stat.* **2000**, *15*, 73–87. [CrossRef]

6.   Le-Rademacher, J.; Billard, L. Symbolic Covariance Principal Component Analysis and Visualization for Interval-Valued Data. *J. Comput. Gr. Stat.* **2012**, *21*, 413–432. [CrossRef]

7.   Diday, E. Principal component analysis for bar charts and metabins tables. *Stat. Anal. Data Min.* **2013**, *6*, 403–430. [CrossRef]

8.   Diday, E.; Emilion, R.; Wang, C.; Wang, H.; Wang, S.; Sampling Based Histogram PCA and its Mapreduce Parallel Implementation on Multicore. *Symmetry* **2018**, *10*, 5.

9.   Ichino, M. Symbolic PCA for histogram-valued data. In Proceedings of the IASC, Yokohama, Japan, 5–8 December 2008.

10.   Kallyth, S.M. Analyse en Composantes Principales de Variables Symboliques de Type Histogramme. Ph.D. Thesis, Université Paris-Dauphine, Paris, France, 2010.

11.   Rodrıguez, O.; Diday, E.;Winsberg, S. Generalization of the principal components analysis to histogram data. In Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, Lyon, France, 13–16 September 2000.

12.   Rodríguez, O. Classification et Modèles Linéaires en Analyse des Données Symboliques. PhD Thesis, Paris IX-Dauphine University, France, 2000.

13.   Arce, J. Dimensionality Reduction Methods for Symbolic Interval Variables. Master's Thesis, University of Costa Rica, San José, Costa Rica, 2018.

14.   Leroy, B.; Chouakria, A.; Herlin, I.; Diday, E. Approche geometrique et classication pour la reconnaissance de visage. Reconnaissance des Forms et Intelligence Artificelle. In Proceedings of the Congrès de Reconnaissance des Formes et IntelligenceArtificielle, Rennes, France, January 1996; pp. 548–557. (In France)

15.   Rodríguez, O. RSDA—R to Symbolic Data Analysis. R Package Version 2.0.8. 2019. Available online: http://CRAN.R-project.org/package=RSDA (accessed on 16 October 2019)