



C_{enet} Biplot: a new proposal of sparse and orthogonal biplots methods by means of elastic net CSVD

Nerea González-García^{1,2} · Ana Belén Nieto-Librero^{1,2} · Purificación Galindo-Villardón^{1,2,3}

Received: 28 May 2020 / Revised: 11 June 2021 / Accepted: 20 September 2021
© The Author(s) 2021

Abstract

In this work, a new mathematical algorithm for sparse and orthogonal constrained biplots, called C_{enet} Biplots, is proposed. Biplots provide a joint representation of observations and variables of a multidimensional matrix in the same reference system. In this subspace the relationships between them can be interpreted in terms of geometric elements. C_{enet} Biplots projects a matrix onto a low-dimensional space generated simultaneously by sparse and orthogonal principal components. Sparsity is desired to select variables automatically, and orthogonality is necessary to keep the geometrical properties that ensure the biplots graphical interpretation. To this purpose, the present study focuses on two different objectives: 1) the extension of constrained singular value decomposition to incorporate an elastic net sparse constraint (C_{enet} SVD), and 2) the implementation of C_{enet} Biplots using C_{enet} SVD. The usefulness of the proposed methodologies for analysing high-dimensional and low-dimensional matrices is shown. Our method is implemented in R software and available for download from <https://github.com/ananieto/SparseCenetMA>.

Keywords Sparse biplot · Constrained singular value decomposition · Orthogonality · Sparsity · Elastic net · HJ-Biplot

Mathematical Subject Classification 15A23 · 62J07 · 62Hxx · 62H25 · 62P10

✉ Nerea González-García
nerea_gonzalez_garcia@usal.es

Ana Belén Nieto-Librero
ananieto@usal.es

Purificación Galindo-Villardón
pgalindo@usal.es

- ¹ Department of Statistics, University of Salamanca, 37007 Salamanca, Spain
- ² Instituto de Investigación Biomédica de Salamanca (IBSAL), 37007 Salamanca, Spain
- ³ Centro de Investigación Institucional (CII), Bernardo O'Higgins University, Santiago, Chile

1 Introduction

Principal component analysis (PCA) is the most widely used multivariate statistical technique for projecting a data set onto a lower dimensional space, preserving as much variability as possible (Jolliffe et al. 2016). The basis vectors of this new subspace known as *principal components* (PCs) are obtained as a linear combination of the original variables. The coefficients of that combination are called *loadings*. Each PC is calculated in terms of all variables and the results can be difficult to interpret. Therefore, several alternatives have been proposed to produce modified PCs with some zero loadings (named *sparse loadings*). These alternatives are known as sparse PCA (Jolliffe et al. 2003; Zou et al. 2006; Shen and Huang 2008; Journée et al. 2010; Li et al. 2016). This purpose is achieved by adding sparse-promoting constraints in the optimization problem. Different constraint techniques are proposed in the literature but some of the most used are Ridge (Hoerl and Kennard 1988) and Lasso (Tibshirani 1996). Ridge shrinks the coefficients towards zero and encourages highly correlated variables to have similar coefficients. On the other hand, Lasso makes some of them zero, but tends to choose a single variable from a set of highly correlated variables, discarding the others. To overcome this, Zou and Hastie (2005) proposed the use of elastic net (enet), which combines Lasso and Ridge to preserve both favourable properties. In addition, enet is particularly useful when the number of variables is higher than the number of observations (Zou and Hastie, 2005).

It is important to emphasise that in some of these Sparse PCA techniques, the loading matrix orthogonality is lost at the expense of sparsity. Thus, some authors, such as Trendafilov (2014) and Genicot et al. (2015), provide sparse and orthogonal components simultaneously.

The coordinates of the observations and the variables in the first components are used to graphically represent them in the score and loading plots, respectively. To visualize them on the same reference system simultaneously, Gabriel (1971) and Galindo (1986) proposed the use of biplot methods. These techniques have been applied in several fields (Xavier et al. 2018; Amor-Esteban et al. 2019; Carrasco et al. 2019; Bernal et al. 2020). Biplots define a common reference system where the rows and the columns of a matrix can be jointly displayed. So, the relationships between them can be interpreted by means of geometric elements in a Euclidean space (distances, angles, projections, ...) (Gabriel 1971; Galindo 1986).

In the case of sparse biplots, there are only two techniques mentioned in the literature related to sparse loadings: CDBiplot (Nieto-Librero et al. 2017), based on the CDPCA of Vichi and Saporta (2009), and Elastic-net HJ-Biplot (Cubilla-Montilla et al. 2021), based on the SPCA of Zou et al. (2006). On the one hand, CDBiplot extracts disjoint PCs, in which each original variable only contributes to the construction of one dimension. On the other hand, Elastic-net HJ-Biplot does not provide orthogonal sparse PCs, even though orthogonality is necessary to keep the geometrical properties that allow biplots interpretation. Also, in this work biplot coordinates are estimated once the sparse loading matrix is obtained from the SPCA (Zou et al. 2006). Nevertheless, as some authors have pointed out it is important to obtain the results in the same optimization process and not in a tandem analysis (Vichi and Saporta 2009; Nieto-Librero et al. 2017).

All things considered, our main objective is to propose a new mathematical technique, called C_{enet}Biplots, that simultaneously incorporates the orthogonality of PCs and the selection of variables by means of the enet sparse constraint. Since the biplot solution is obtained from the singular value decomposition (SVD) (Gabriel 1971; Galindo 1986), our research is focused on the sparse and orthogonal SVD via Lasso proposed by Guillemot et al. (2019) but imposing the enet constraint to overcome the disadvantages mentioned above.

Therefore, this work is structured as follows. Section 2 includes the notation paragraph and Sect. 3 defines the extension of constrained singular value decomposition as the solution of a convex-optimization problem with enet and orthogonality restrictions. Section 3 also shows the algorithm used to solve C_{enet}SVD, extending the projection onto convex sets (POCS) algorithm in the sense of a divide and conquer algorithm. Section 4 presents the implementation of the sparse and orthogonal biplot methods, known as the C_{enet}Biplots. The selection of the sparsity parameters is proposed in Sect. 5. Section 6 shows the usefulness of these methodologies analysing high-dimensional real genomic data and low-dimensional psychometric data. Finally, Sect. 7 includes a discussion and the main conclusions of the study.

2 Notation

We present below the notation and terminology used in this manuscript. X_{IJ} denotes a matrix with the information of I observations in the rows and J variables in the columns. The elements of a matrix X are denoted as x_{ij} . The transpose of a matrix X is denoted by X^T , and its inverse, as X^{-1} . The ℓ_2 norm of a matrix X is defined by $\|X\|_F^2$. The ℓ_2 norm of a vector $\mathbf{x} = \{x_j\}_{j=1}^J$ is computed by $\sqrt{\sum_j x_j^2}$, and the ℓ_1 norm is calculated by $\sum_j |x_j|$. A vector is normalized when it is divided by its ℓ_2 norm. Constraint balls are defined as the regions $\mathfrak{B}_\tau^{\ell_2}(\mathbf{x}) = \{\mathbf{x}/\|\mathbf{x}\|_2^2 \leq \tau\}$, $\mathfrak{B}_\tau^{\ell_1}(\mathbf{x}) = \{\mathbf{x}/\|\mathbf{x}\|_1 \leq \tau\}$ and $\mathfrak{B}_\tau^{\ell_1+\ell_2}(\mathbf{x}) = \{\mathbf{x}/(1-\alpha)\|\mathbf{x}\|_1 + \alpha\|\mathbf{x}\|_2^2 \leq \tau\}$ for some $\alpha \in [0, 1]$.

3 Singular value decomposition

Given a matrix X_{IJ} of rank $R \leq \min(I, J)$, the SVD of X is defined as the product:

$$X_{IJ} = U_{IR} \mathbf{D}_R \mathbf{V}^T_{RJ} \quad (1)$$

where $U = [u_1, \dots, u_I]$ and $V = [v_1, \dots, v_J]$ are orthonormal matrices, $U^T U = I$ and $V^T V = I$. U contains the left-singular vectors of the SVD in columns, V contains the right-singular vectors and \mathbf{D} is a diagonal matrix containing the d_r singular values of X ($r = 1, \dots, R$), conveniently expressed so that $d_1 \geq d_2 \geq \dots \geq d_R \geq 0$. For optimal $Q \leq R$, SVD provides the best low Q -rank approximation \widehat{X}_Q of X in the sense of least squares by minimizing the ℓ_2 norm of the difference between the initial

and the reconstructed matrices (Eckart and Young 1936; Shen and Huang 2008). \widehat{X}_Q , is defined as:

$$\widehat{X}_Q = U_{IQ} D_Q V^T_{QJ} = \sum_{q=1}^Q d_q \mathbf{u}_q \mathbf{v}_q^T \tag{2}$$

with $\mathbf{u}_q^T \mathbf{u}_q = \mathbf{v}_q^T \mathbf{v}_q = 1$ and $\mathbf{u}_q^T \mathbf{u}_{q'} = \mathbf{v}_q^T \mathbf{v}_{q'} = 0 \forall q \neq q' (q = 1, \dots, Q)$.

Frequently, the orthogonal singular vectors of SVD are computed using the power iteration algorithm together with a deflation approach. Instead, Guillemot et al. (2019) suggest obtaining the singular vectors by the projection onto the convex sets (POCS) algorithm (Bauschke and Combettes 2017).

3.1 Extension of constrained singular value decomposition to elastic net (C_{enet}SVD)

C_{enet}SVD provides a factorization of X_{IJ} by means of sparse and orthogonal singular vectors (called pseudo-singular vectors) and pseudo-singular values. The key point of C_{enet}SVD is the calculation of sparse and orthogonal vectors simultaneously. The C_{enet}SVD formulation is based on the constrained optimization problem of CSVD proposed by Guillemot et al. (2019), replacing lasso by enet restriction:

$$\begin{aligned} & \underset{d, \mathbf{u}, \mathbf{v}}{\operatorname{argmin}} \frac{1}{2} \left\| X - \sum_{q=1}^Q d_q \mathbf{u}_q \mathbf{v}_q^T \right\|_F^2 \\ & \text{s.t.} \left\{ \begin{array}{l} \mathbf{u}_q^T \mathbf{u}_q = \mathbf{v}_q^T \mathbf{v}_q = 1, \mathbf{u}_q^T \mathbf{u}_{q'} = \mathbf{v}_q^T \mathbf{v}_{q'} = 0 \quad \forall q \neq q' \\ (1 - \alpha) \|\mathbf{u}_q\|_1 + \alpha \|\mathbf{u}_q\|_2^2 \leq \tau_{1,q}; (1 - \alpha) \|\mathbf{v}_q\|_1 + \alpha \|\mathbf{v}_q\|_2^2 \leq \tau_{2,q} \end{array} \right. \end{aligned} \tag{3}$$

where $\tau_{1,q}, \tau_{2,q} > 0$ are the shrinkage parameters that control the sparsity degree included in the constrained model. The higher τ_1 or τ_2 is, the fewer sparse coefficients there are. It is important to remark that only some values for $\tau_{1,q}$ and $\tau_{2,q}$ lead to possible solutions (see Sect. 5.2 and Online Resource 2A). The parameter $\alpha \in [0, 1)$ defines the amount of the Lasso or the Ridge constraint included in the enet restriction.

To find the solution of (3), an iterative process is defined (Guillemot et al. 2019). First, it is necessary to establish an equivalent form of the previous minimization problem. Equation (3) is equivalent to:

$$\begin{aligned} & \underset{\mathbf{u}, \mathbf{v}}{\operatorname{argmax}} \mathbf{u}^T X \mathbf{v} \\ & \text{s.t.} \left\{ \begin{array}{l} \mathbf{u}^T \mathbf{u} \leq 1, \mathbf{v}^T \mathbf{v} \leq 1, \mathbf{u}^T \mathbf{u}_{q'} = \mathbf{v}^T \mathbf{v}_{q'} = 0 \quad \forall q' < q \\ (1 - \alpha_1) \|\mathbf{u}\|_1 + \alpha_1 \|\mathbf{u}\|_2^2 \leq \tau_{1,q}; (1 - \alpha_2) \|\mathbf{v}\|_1 + \alpha_2 \|\mathbf{v}\|_2^2 \leq \tau_{2,q} \end{array} \right. \end{aligned} \tag{4}$$

With $\mathbf{v}_{q'}$ and $\mathbf{u}_{q'}$ previously calculated, $0 \leq q' < q$ and $q \geq 1$. Equation (4) is solved by block relaxation, an iterative process that resolves two alternating parts:

(1) Find the vector solution \mathbf{u} optimizing the function with \mathbf{v} fixed:

$$\begin{aligned} & \arg \min_u \frac{1}{2} \|\mathbf{u} - \mathbf{X}\mathbf{v}\|_F^2 \\ & s.t. \{ \mathbf{u} \in \mathfrak{B}_{\ell_1+\ell_2}(\tau), \mathbf{u} \in \mathfrak{B}_{\ell_2}(1), \mathbf{u} \in \mathbf{U}^\perp \leftrightarrow \mathbf{u} \in \mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1) \end{aligned} \tag{5}$$

where \mathbf{U}^\perp is the orthogonal complement of the space spanned by the columns of a matrix \mathbf{U} . These constraints involved projections of a vector onto a convex space. In addition, it should be noted that the intersection of two convex spaces is also convex. Thus, the problem can be solved by using the POCS and the PL1L2 algorithms (Gloaguen et al. 2017; Guillemot et al. 2019). The projection of a vector onto the enet ball ($\mathfrak{B}_{\ell_1+\ell_2}$) proposed here follows the line of the algorithms in linear time to the projection onto the Lasso ball \mathfrak{B}_{ℓ_1} (Berg et al. 2008; Duchi et al. 2008; Guillemot et al. 2019) and the enet ball $\mathfrak{B}_{\ell_1+\ell_2}$ (Mairal et al. 2010) (see Online Resource 2B). In our case, the method for projecting a vector onto $\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}$ is an extension of the fast and exact algorithm for the projection onto the intersection of \mathfrak{B}_{ℓ_1} and \mathfrak{B}_{ℓ_2} proposed by Guillemot et al. (2019) (see Online Resource 2C).

(2) Maximize (4) to find the vector \mathbf{v} solution with \mathbf{u} fixed:

$$\begin{aligned} & \operatorname{argmin}_u \frac{1}{2} \|\mathbf{v} - \mathbf{X}^T \mathbf{u}\|_F^2 \\ & s.t. \{ \mathbf{v} \in \mathfrak{B}_{\ell_1+\ell_2}(\tau), \mathbf{v} \in \mathfrak{B}_{\ell_2}(1), \mathbf{v} \in \mathbf{V}^\perp \leftrightarrow \mathbf{v} \in \mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1) \end{aligned} \tag{6}$$

where \mathbf{V}^\perp is the orthogonal complement of the space spanned by the columns of a matrix \mathbf{V} . The projection $\mathbf{v}_{t+1} = \operatorname{proj}^{\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1) \cap \mathbf{V}^\perp}(\mathbf{X}^T \mathbf{u}_{t+1})$ is carried out in the same manner as for (1).

The global optimization problem of C_{enet}SVD is handled using the POCS algorithm (Table 1). Lines 6 and 7 are modified from (Guillemot et al. 2019) to address the problem of projection onto the $\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}$ space.

4 Extension of C_{enet}SVD to the biplot methods

The results of traditional PCA and Biplot methods are often calculated from the SVD of a matrix. Consequently, sparse and orthogonal constrained PCA (C_{enet}PCA) and sparse and orthogonal constrained biplots (C_{enet}Biplots) are obtained from the results of the C_{enet}SVD. Set the C_{enet}SVD of \mathbf{X} as the low Q -rank approximation of the original matrix $\mathbf{X} \approx \mathbf{U}_{enet} \mathbf{D} \mathbf{V}_{enet}^T$, where \mathbf{U}_{enet} and \mathbf{V}_{enet} are sparse and orthonormal matrices.

4.1 Sparse and orthogonal PCA

The main objective of C_{enet}PCA is to project the original matrix onto a subspace determined by a new set of $Q < J$ sparse and orthogonal PCs. Given \mathbf{X}_{IJ} , the sparse

Table 1 Algorithm for the implementation of $C_{enet}SVD$ based on the POCS algorithm

Algorithm 1: POCS - $C_{enet}SVD$. Projection of a vector onto $\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1) \cap \mathbf{M}^\perp$	
Input:	X_{IJ} , rank Q , $\varepsilon \approx 0, \tau \in [1, (1 - \alpha)\sqrt{J} + \alpha]$
Output:	$\mathbf{U}_{IQ}, \mathbf{D}_{QQ}, \mathbf{V}_{JQ}$
Initialization:	$\mathbf{U} = \mathbf{0}, \mathbf{V} = \mathbf{0}, d_0 = 0$
1:	For q in 1: q do:
2:	Initialize randomly: $\mathbf{u}_0, \mathbf{v}_0$
3:	$d_1 = \mathbf{u}_0^T \mathbf{X} \mathbf{v}_0$
4:	$t = 0$
5:	While $ d_{t+1} - d_t \geq \varepsilon$ do:
6:	$\mathbf{u}_{t+1} = \text{proj}^{\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1) \cap \mathbf{U}^\perp}(\mathbf{X} \mathbf{v}_t)$
7:	$\mathbf{v}_{t+1} = \text{proj}^{\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1) \cap \mathbf{V}^\perp}(\mathbf{X}^T \mathbf{u}_{t+1})$
8:	$d_{t+1} = \mathbf{u}_{t+1}^T \mathbf{X} \mathbf{v}_{t+1}$
9:	$t = t + 1$
10:	$\mathbf{d} = [\mathbf{d}, d_{t+1}]$
11:	$\mathbf{U} = [\mathbf{U}, \mathbf{u}_{t+1}]$
12:	$\mathbf{V} = [\mathbf{V}, \mathbf{v}_{t+1}]$
13:	end
14:	$\mathbf{D} = \text{diag}(\mathbf{d})$
15:	end

q - $C_{enet}PC$ is defined as a linear combination of the initial variables as:

$$\mathbf{Y}_{IQ} = \mathbf{X}_{IJ} \mathbf{V}_{enetJQ} \tag{7}$$

being \mathbf{Y}_{IQ} the score matrix, which contains the coordinates of the observations in the new subspace, and \mathbf{V}_{enet} the sparse loading matrix. A flowchart of $C_{enet}PCA$ is shown in Fig. 1a.

4.2 Sparse and orthogonal Biplots

Biplot methods are optimal tools to visualize multivariate data in a low-dimensional space. The relationship between the observations and the variables can be interpreted in a scatterplot because of the inner product properties. Gabriel (1971) proposed the JK and GH biplots, which assign an optimal quality of representation to rows (GH-Biplot) and columns (JK-Biplot) in the same Euclidean space. In order to provide the highest-quality representation for both rows and columns in the same reference

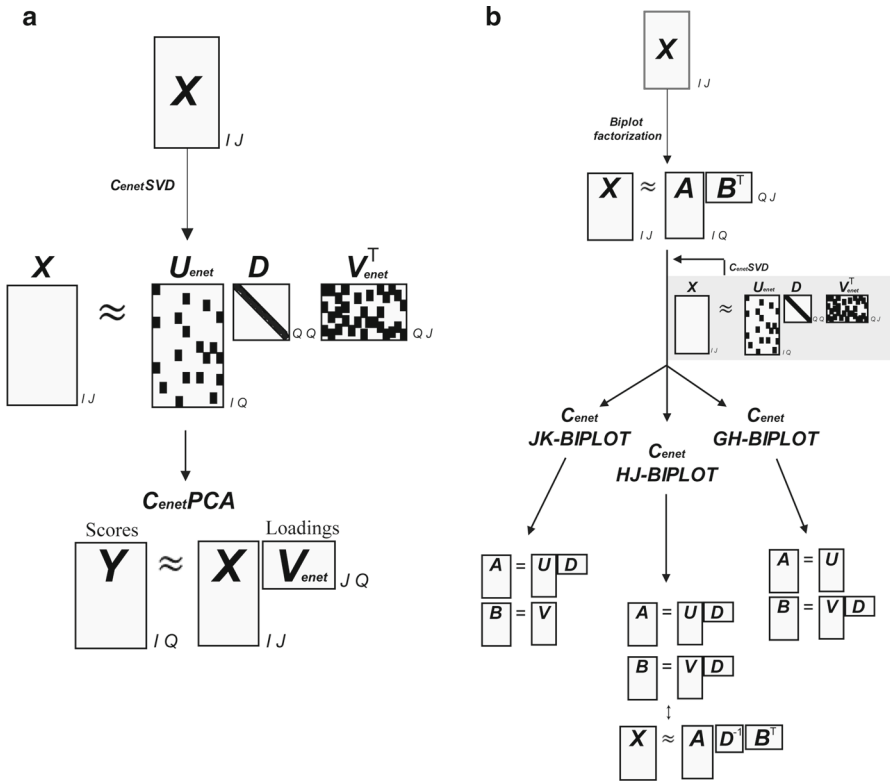


Fig. 1 C_{enet}PCA and C_{enet}Biplots formulation. Sparse and orthogonal PCA model (a) and sparse and orthogonal JK, GH and HJ biplots defined via C_{enet}SVD (b)

system, Galindo (1986) proposed the HJ-Biplot. To this end, the original matrix is factorized as $X \approx AB^T$, so that the inner product $a_i^T b_j$ approximates the element x_{ij} as closely as possible. A_{IQ} and B_{JQ} are the row and the column markers matrices, respectively. The pseudo-singular vectors and values obtained from the C_{enet}SVD of X are used to implement sparse and orthogonal biplots. C_{enet}JK-Biplot sets the row and column markers as $A = U_{enet}D$ and $B = V_{enet}$, C_{enet}GH-Biplot sets $A = U_{enet}$ and $B = V_{enet}D$, and finally C_{enet}HJ-Biplot sets $A = U_{enet}D$ and $B = V_{enet}D$, respectively (Fig. 1b).

To interpret the C_{enet}Biplots graphical representation, it is important to note that i) observations are represented by dots and variables by vectors in the graph; ii) distances between points show the dissimilarities between observations; iii) lengths of the vectors refer to the variability of the variables; iv) relationships between variables are interpreted from the cosine of the angles between the corresponding vectors (obtuse: inverse relationship; acute: direct relationship; right angle: linear independence); and v) projections of the points in the direction of a vector approximate the values of the variables for those observations.

5 Selection of the sparsity parameters

5.1 Selection of α and $(1 - \alpha)$

The parameter $\alpha \in [0, 1)$ defines the amount of Lasso or Ridge constraint included in the enet restriction. Usually, α is set to 0.5 to aggregate the same Lasso and Ridge constraint to the model. In practice, the parameter α could be selected manually. But also, we suggest to select the α that minimizes the cross-validation error (James et al. 2014) through an iterative process. This process starts defining an increasing sequence of possible values for α and segmenting the matrix into $n = 10$ folds. For each of the n -folds, the training matrix is conformed by the observations of the submatrices formed in the $(n - 1)$ -folds. The remaining observations constitute the test matrix. Then, the iterative process for each of the $\alpha = (\alpha_1, \dots, \alpha_t)$ values is initiated. $C_{enet}SVD$ of the training matrix is carried out, and the matrix V_{enet} is used to calculate the mean reconstruction error $MSE_{\alpha_1,1}$ of the test matrix:

$$MSE_{\alpha_1,1} = \|X_{TEST} - \widehat{X}_{TEST}\|^2 = \|X_{TEST} - X_{TEST}V_{enet}V_{enet}^T\|^2 \quad (8)$$

Reconstruction errors $MSE_{\alpha_1,1}, \dots, MSE_{\alpha_1,n}$ are obtained for each of the folds; hence, the final MSE_{α_1} is computed as the mean of the errors $MSE_{\alpha_1,1:n}$.

This step is repeated for the whole sequence of α_t values. The optimum α is the one that provides the minimum $MSE = \min\{MSE_{\alpha_1}, \dots, MSE_{\alpha_t}\}$.

5.2 Selection of τ

The shrinkage parameter τ inversely controls the degree of sparsity; that is, the larger its value is, the fewer zero loadings. In this work, we propose to select τ using the Bayesian information criterion (BIC) as in (Guo and James 2010; Croux et al. 2013):

$$BIC(\tau) = \frac{\|X - XV_{enet}V_{enet}^T\|^2}{\|X - XVV^T\|^2} + df(\tau) \frac{\log(I)}{I} \quad (9)$$

where X_{IJ} is the original matrix, V_{enet} is the right-pseudo-singular vector matrix obtained from $C_{enet}SVD$, V is the right-singular vector matrix obtained from unconstrained SVD and $df(\tau)$ is the number of non-zero elements in V_{enet} . The parameter τ that minimizes $BIC(\tau)$ is selected from a sequence of possible $\tau \in [1, (1 - \alpha)\sqrt{J} + \alpha]$. As stated by Guillemot et al. (2019) and Witten et al. (2009), only some values of the constraints lead to solutions (see Online Resource 2).

6 Analysing data with C_{enet} methods

In this section, we illustrate the performance of $C_{enet}PCA$ and $C_{enet}HJ$ -Biplot to analyse two matrices in different contexts: 1) the Object-Spatial Imagery Questionnaire (OSIQ) data ($J < I$) and 2) the MILE gene expression data ($J \gg I$). All data

analyses were performed using the R software (R Core Team 2020). All figures were illustrated using the ggplot2 library (Wickham 2016).

6.1 Analysing the mental imagery questionnaire data: a psychometric example

The OSIQ dataset is publicly available on GitHub (<https://github.com/HerveAbdi/data4PCCAR>).

To show the usefulness of C_{enet}HJ-Biplot in a toy example in which the number of variables is less than the number of observations, a random selection of 100 simulated participants who responded to the items of the OSIQ scale (Blajenkova et al. 2006) was carried out. This scale consists of 30 items rated on a 5-category Likert scale, which are structured around two latent dimensions: the spatial and object imagery scales. The correlation matrix between items is shown in Online Resource 1.

Two-dimensional solutions of the traditional HJ-Biplot and C_{enet}HJ-Biplot are shown in Fig. 2. To obtain a low sparsity degree, C_{enet}HJ-Biplot was employed with a shrinkage parameter $\tau_2 = (0.5\sqrt{J} + 0.5) \cdot (3/4)$ (low level of sparsity). There is no sparsity restriction imposed on individuals. In both plots, participants are represented by dots, and questionnaire items, by arrows. The cosine of the angles between the vectors reflects the relationships between items. Thus, items of the same theoretical dimension are strongly and directly correlated (acute angles). As can be seen, if the C_{enet}HJ-Biplot is applied (Fig. 2b), the relationships of the items in the two constructed dimensions, even with low constraints, reflect the theoretical psychometric structure more clearly than in the case of the HJ-Biplot (Fig. 2a). Additionally, the item o15 correlates directly and weakly with the items of its own scale and correlates inversely and strongly with the items of the other scale in the results of HJ-Biplot (Fig. 1a). The C_{enet}HJ-Biplot tends to maintain these relationships (Fig. 2b). The same can be observed with items s06, s11 and o10, which do not correlate with their dimension items but whose relationships to the items of another subscale are preserved.

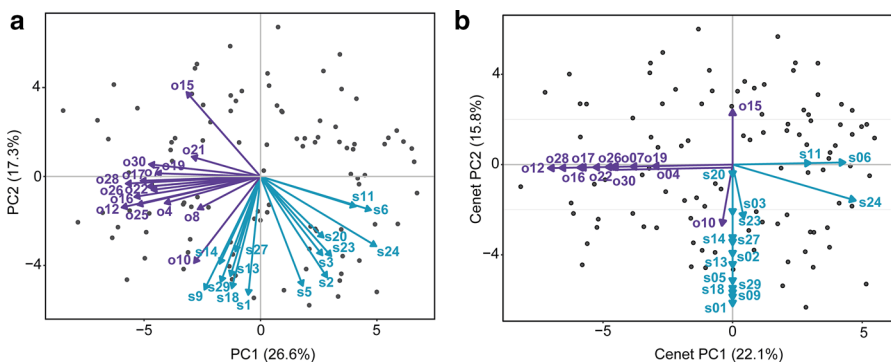


Fig. 2 2D plots of components 1–2 of the HJ-Biplot (a) and sparse and orthogonal C_{enet}HJ-Biplot (b). Labels starting with “o” refer to object imagery scale items, and labels starting with “s” indicate spatial imagery scale items

6.2 Analysing leukaemia gene expression data: a genomic example

The use of statistical methods to analyse microarray data involving variable selection has become increasingly important in tumours classification context (Algamil and Lee 2015). In the case of gene expression data, it is common that the number of genes exceeds the number of samples. Frequently only a few genes have relevant information and therefore the use of sparse constraints for the automatic selection of genes is necessary.

To show the usefulness of $C_{\text{enet}}\text{PCA}$ and $C_{\text{enet}}\text{HJ-Biplot}$, in this section we analyse 216 leukaemia patients randomly selected from the GSE13204 series available at the Gene Expression Omnibus (GEO) repository. Among our samples, 32.9% had been classified as acute lymphoblastic leukaemia (ALL, $n = 71$), and 34.3%, as chronic lymphocytic leukaemia (CLL, $n = 74$); 32.9% of the samples were control patients ($n = 71$). RNA gene expression data were extracted from the Affymetrix Microarray Platform Human Genome U133 Plus 2.0 Array (HGU133Plus2). Data preprocessing was carried out using RMA normalization. For illustrative purposes, 2,000 genes showing the greatest variability from the CUR decomposition leverage scores were selected (more detailed information can be found in (Mahoney and Drineas 2009)).

Traditional PCA was performed on the centred matrix of 216 samples and 2,000 probe sets. The 45.5% of the data variability is explained by the two components retained. Figure 3a shows the score plot of components 1–2. The scores revealed three subgroups of samples clearly differentiated corresponding to control (rectangle), ALL (circle) and CLL (triangle) samples. The first PC discriminates between control and CLL samples, with the CLL samples located on the positive side of axis 1. The ALL samples were differentiated from the CLL and the control samples by their 2-axis coordinates. Figure 3a (middle, bottom) shows that each PC is a linear combination of a large number of genes, making them hardly interpretable despite their good discriminatory capacity.

To assess the effects of sparsification, $C_{\text{enet}}\text{PCA}$ is performed using $\alpha = 0.5$ (i.e., providing equal weight to the Lasso and Ridge constraints) and $\tau = 5.86$. The last one was selected according to the BIC criterion and the previously separation of the groups considered. The 25.8% of the data variability is explained by the two sparse components retained. Figure 3b reflects the score plot of the samples in the first two sparse PCs. The same classification is achieved, as shown in Fig. 3a (top). The loading plots of $C_{\text{enet}}\text{PCA}$ show the genes automatically selected with this technique (Fig. 3b, middle-bottom). Those genes with higher loadings in the PCs are those with higher loadings in the sparse $C_{\text{enet}}\text{PCs}$. Additionally, the variables with lower loadings in the PCs have a zero-value loading in the sparse $C_{\text{enet}}\text{PCs}$ (or close to 0). This fact simplifies the interpretation of the constrained components obtained. From the 2,000 gene probes considered, 1,752 present zero loadings in both $C_{\text{enet}}\text{PCs}$, and from the remaining 248 non-zero loadings, 43 gene probes present loadings lower than 0.01 along both constrained components.

Finally, to characterise the influence of the selected gene probes in the separation of the above-mentioned groups, the gene expression centred matrix of the

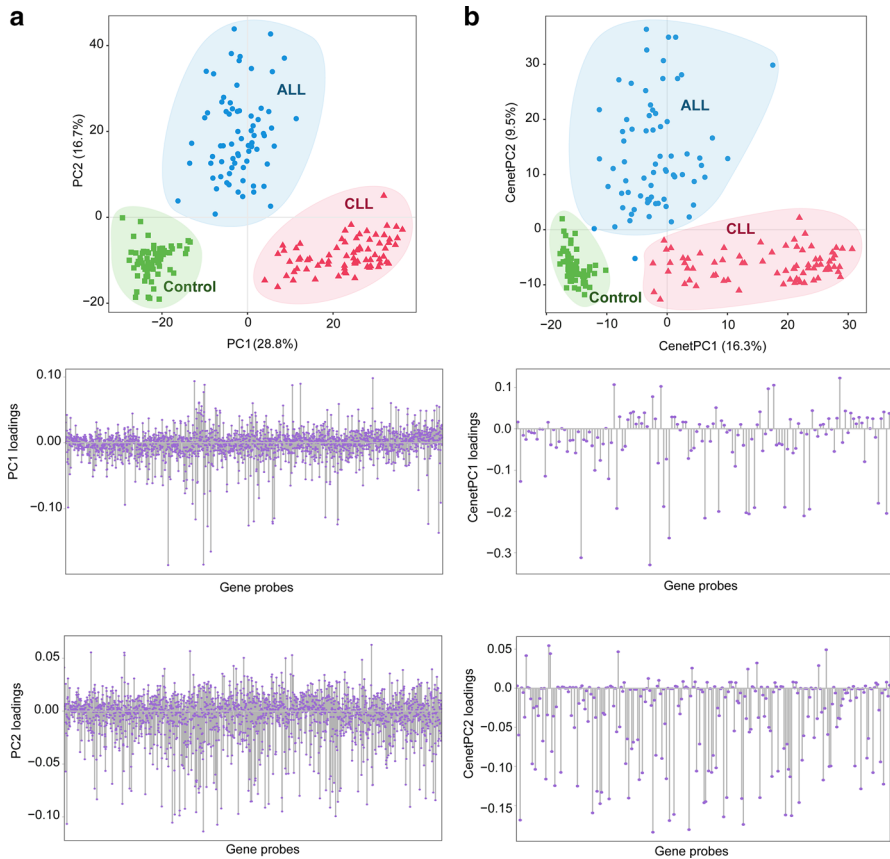


Fig. 3 2D plots of components 1–2 of PCA (**a**) and components 1–2 of sparse and orthogonal C_{enet}PCA (**b**). Each shape refers to one of the three groups of samples (rectangle: control; triangle: CLL; circle: ALL). Panel a contains the score plot of PCs 1–2 (top plot) and the loadings plots of PC1 (middle plot) and PC2 (bottom). Each PC is computed as a linear combination of the 2,000 gene probes measured. **b** contains the score plot (top) and the loadings plot of each of the constrained PCs (middle, bottom). Sparse and orthogonal PCs distinguish the three groups considered, although in this case many gene probes present zero loadings

216 observations and 205 resultant gene probes of C_{enet}PCA were analysed via C_{enet}HJ-Biplot. The sparsity constraint τ_2 was fixed at a medium level of sparsity ($\tau = ((1 - \alpha)\sqrt{J} + \alpha) \cdot (1/3)$) (Guillemot et al. 2019). The 27.6% of the data variability is explained by the two sparse components retained. The C_{enet}HJ-Biplot results are shown in Fig. 4. The first axis shows the separation of CLL samples from control and ALL samples. Axis 2 is a gradient direction differentiating the ALL samples from the rest. The genes that do not appear on the plane have null coordinates. In addition, the C_{enet}HJ-Biplot representation makes possible the recognition of a genetic characterization of each of the subgroups. In this sense, control samples are characterized by a high expression of *S100A12* and *S100A9*. These genes are responsible for discriminating between control and tumour samples. CLL samples are differentiated from

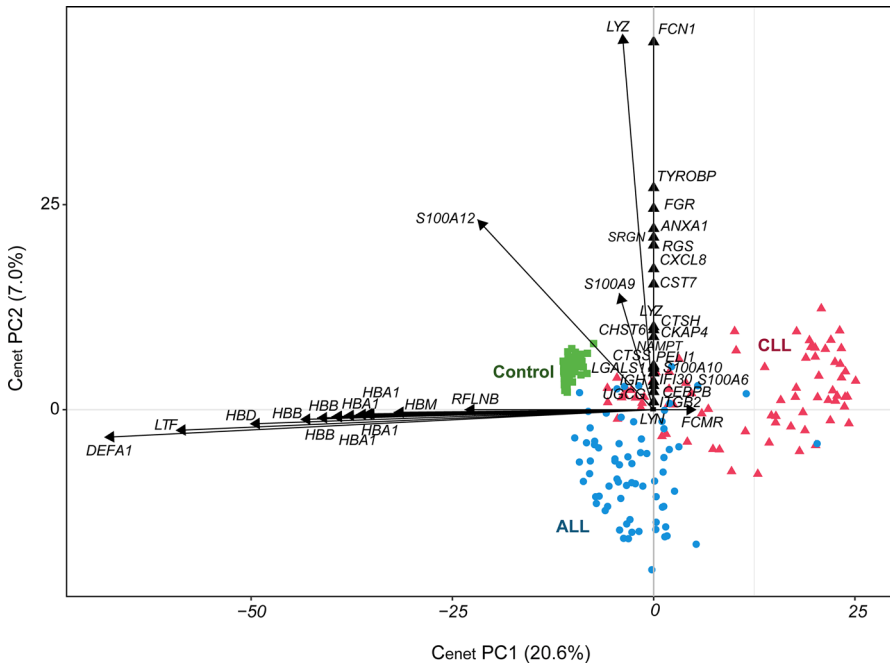


Fig. 4 2D plots of components 1–2 of sparse and orthogonal $C_{enet}HJ$ -Biplot. Each shape refers to one of the three groups of study samples (rectangle: control; triangle: CLL; circle: ALL)

control and ALL samples because it presents higher expression levels of *FCMR* and lower expression levels of *DEFA1*, *LTF*, *HDB*, *HBB*, *HBA1*, *HBM* and *RFLNB* genes.

7 Conclusions

In this work, the extension of CSVD (Guillemot et al. 2019) to the enet ball is proposed, integrating sparse and orthogonal vectors simultaneously. This method is based on the projection of a vector onto the convex intersection of enet and ℓ_2 balls. Here, the enet constraint is shown as a suitable constraint approach, restricting coefficients to zero while ensuring that correlated variables have similar coefficients, a desirable property in disciplines such as genomics or psychometry. Our approach using $C_{enet}SVD$ is useful for analysing large-scale problems with $J \gg I$ and datasets with $I > J$. Additionally, $C_{enet}SVD$ is extended to sparse and orthogonal constrained $C_{enet}PCA$ and sparse and orthogonal constrained $C_{enet}Biplots$. These techniques provide the possibility of recognizing groups with similar patterns and the causative variables associated with them. In addition, they are variable selection techniques that improve the interpretation of results due to the sparse components established. Furthermore, this work provides a sparsity parameter selection procedure based on the cross-validation and the BIC, as well as the possibility to manually establish distinct levels of sparsity.

Future lines of research may contemplate the possibility of applying other types of constraints within the CSVD framework or even the proposal of other algorithms for projecting a vector onto non-convex sets based on the correspondent mathematical theory. Additionally, statistical techniques of two-way and three-way data analysis could be developed through CenetSVD. We conclude that our proposed methods are promising tools for conducting multivariate analysis and are applicable to a wide range of research areas.

Our methods are available as R functions in the GitHub repository in SparseCenetMA (<https://github.com/ananieto/SparseCenetMA>).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11634-021-00468-1>.

Acknowledgements The authors would like to thank Guillemot et al. for the possibility of using the R public code of CSVD that can be found at <https://github.com/vguillemot/csvd>, with copyright:

“Copyright (c) 2018 Vincent Guillemot. Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE”.

Author Contributions Investigation: NG, ABN, PG; Conceptualization: NGG, ABN; Methodology: NG, ABN; Software: NG, ABN; Data curation: NG, ABN; Supervision: ABN, PG; Writing—Original draft preparation: NG, ABN, PGV. Writing—Revised draft preparation: NG, ABN, PG.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was not supported by any grant.

Data availability The datasets analysed during the current study are available in the GEO repository. [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13204>].

Code availability Both the R-code and data used in this work are available in the GitHub repository (<https://github.com/ananieto/SparseCenetMA>).

Declarations

Conflict of interest The authors declared that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Algamal ZY, Lee MH (2015) Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Comput Biol Med* 67:136–145. <https://doi.org/10.1016/j.combiomed.2015.10.008>
- Amor-Esteban V, Galindo-Villardón MP, García-Sánchez IM (2019) A multivariate proposal for a national corporate social responsibility practices index (NCSRPI) for international settings. *Soc Indic Res* 143:525–560. <https://doi.org/10.1007/s11205-018-1997-x>
- Bauschke HH, Combettes PL (2017) *Convex analysis and monotone operator theory in hilbert spaces*. Springer, Cham
- Berg E, Schmidt M, Friedlander M, Murphy K (2008) Group sparsity via linear-time projection, Tech. Rep. (TR-2008–09.) Dept. of Computer Science, UBC
- Bernal EF, del Rey AM, Villardón PG (2020) Analysis of madrid metro network: from structural to HJ-biplot perspective. *Appl Sci* 10:5689. <https://doi.org/10.3390/app10165689>
- Blajenkova O, Kozhevnikov M, Motes MA (2006) Object-spatial imagery: a new self-report imagery questionnaire. *Appl Cogn Psycho* 20:239–263. <https://doi.org/10.1002/acp.1182>
- Carrasco G, Molina JL, Patino-Alonso MC, Castillo DCM, Vicente-Galindo MP, Galindo-Villardón MC (2019) Water quality evaluation through a multivariate statistical HJ-Biplot approach. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2019.123993>
- Croux C, Filzmoser P, Fritz H (2013) Robust sparse principal component analysis. *Technometrics* 55:202–214. <https://doi.org/10.1080/00401706.2012.727746>
- Cubilla-Montilla M, Nieto-Librero AB, Galindo-Villardón MP, Torres-Cubilla CA (2021) Sparse HJ biplot: a new methodology via elastic net. *Mathematics* 9(11):1298. <https://doi.org/10.3390/math9111298>
- Duchi J, Shalev-Shwartz S, Singer Y, Chandra T (2008) Efficient projections onto the ℓ_1 -ball for learning in high dimensions. *Proc 25th Int Conf Mach Learn* 272–279
- Eckart C, Young G (1936) The approximation of one matrix by another of lower rank. *Psychometrika* 1:211–218
- Gabriel KR (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58:453–467
- Galindo MP (1986) An alternative for simultaneous representation: HJ-Biplot. *Questiio* 10:12–23
- Genicot M, Huang W, Trendafilov NT (2015) Weakly correlated sparse components with nearly orthonormal loadings. In *International Conference on Geometric Science of Information* (484–490). Springer, Cham
- Gloaguen A, Guillemot V, Tenenhaus A, et al (2017) An efficient algorithm to satisfy l_1 and l_2 constraints. *49èmes Journées de Statistique*. hal-01630744
- Guillemot V, Beaton D, Gloaguen A et al (2019) A constrained singular value decomposition method that integrates sparsity and orthogonality. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0211463>
- Guo J, James G (2010) Principal component analysis with sparse fused loadings. *J Comput Graph Stat* 19:930–946
- Hoerl A, Kennard R (1988) Ridge regression. *Encyclopedia of Statistical Sciences*, vol 8. Wiley, New York, pp 129–136
- James G, Witten D, Hastie T, Tibshirani R (2014) *An Introduction to Statistical Learning: with Applications in R*. An Introd to Stat Learn with Appl R. Springer, New York
- Jolliffe IT, Cadima J (2016) Principal component analysis : a review and recent developments. *Philos Trans Royal Soc A* 374:20150202
- Jolliffe IT, Trendafilov NT, Uddin M (2003) A modified principal component technique based on the LASSO. *J Comput Graph Stat* 12:531–547. <https://doi.org/10.1198/1061860032148>
- Journée M, Nesterov Y, Richtárik P, Sepulchre R (2010) Generalized power method for sparse principal component analysis. *J Mach Learn Res* 11:517–553
- Li B, Tian BB, Liu J (2016) A simple review of sparse principal component analysis. *Intell Comput Theor Technol* 7996:443–449. <https://doi.org/10.1007/978-3-642-39482-9>
- Mahoney MW, Drineas P (2009) CUR matrix decompositions for improved data analysis. *Proc Natl Acad Sci U S A* 106:697–702. <https://doi.org/10.1073/pnas.0803205106>
- Mairal J, Bach F, Ponce J, Sapiro G (2010) Online learning for matrix factorization and sparse coding. *J Mach Learn Res* 11:19–60. <https://doi.org/10.1145/1756006.1756008>
- Nieto-Librero AB, Sierra C, Vicente-Galindo MP et al (2017) Clustering disjoint HJ-Biplot: a new tool for identifying pollution patterns in geochemical studies. *Chemosphere* 176:389–396. <https://doi.org/10.1016/j.chemosphere.2017.02.125>

- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Shen H, Huang JZ (2008) Sparse principal component analysis via regularized low rank matrix approximation. *J Multivar Anal* 99:1015–1034. <https://doi.org/10.1016/j.jmva.2007.06.007>
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B* 58:267–288
- Trendafilov NT (2014) From simple structure to sparse components: a review. *Comput Stat* 29:431–454. <https://doi.org/10.1007/s00180-013-0434-5>
- Vichi M, Saporta G (2009) Clustering and disjoint principal component analysis. *Comput Stat Data Anal* 53:3194–3208. <https://doi.org/10.1016/j.csda.2008.05.028>
- Wickham H (2016) *ggplot2: elegant graphics for data analysis*. Springer, New York
- Witten DM, Tibshirani R, Hastie T (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10:515–534. <https://doi.org/10.1093/biostatistics/kxp008>
- Xavier A, Freitas MBC, Rosário MS, Fragoso R (2018) Disaggregating statistical data at the field level: an entropy approach. *Spat Stat* 23:91–108. <https://doi.org/10.1016/j.spasta.2017.11.005>
- Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* 67:301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- Zou H, Hastie T, Tibshirani R (2006) Sparse principal component analysis. *J Comput Graph Stat* 15:265–286. <https://doi.org/10.1198/106186006X113430>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.