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

Advances in Data Analysis and Classification

Nerea González-García[1,2], Ana Belén Nieto-Librero[1,2], Purificación Galindo-Villardón[1,2]

[1]Department of Statistics, University of Salamanca, Salamanca, Spain; 37007
[2]Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, Spain; 37007

Corresponding author
1. González-García, Nerea
[1]Department of Statistics, University of Salamanca, Salamanca, Spain; 37007: nerea_gonzalez_garcia@usal.es
[2]Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, Spain; 37007
Main Number: (+34) 923294500 ext. 6988
e-mail: nerea_gonzalez_garcia@usal.es

ONLINE RESOURCE 2

A. Geometrical intuition of the possible values for tuning parameter $\boldsymbol{\tau}$ to generate workable solutions

The tuning parameter τ must be defined in $[1,(1-\alpha)\sqrt{J}+\alpha]$ to obtain a feasible solution of the constrained elastic net optimization problem. The explanation and geometrical intuition for this fact are shown below.

Lemma 1. Set $\boldsymbol{x} \in \mathbb{R}^J$; then:

$$\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_1 \leq \sqrt{J}\|\boldsymbol{x}\|_2 \quad \text{(A1)}$$

See (Guillemot et al., 2019) to find the Lemma 1 demonstration. Consequently, it is easy to observe that:

$$(1-\alpha)\|\boldsymbol{x}\|_2 + \alpha\|\boldsymbol{x}\|_2^2 \leq (1-\alpha)\|\boldsymbol{x}\|_1 + \alpha\|\boldsymbol{x}\|_2^2 \leq (1-\alpha)\sqrt{J}\|\boldsymbol{x}\|_2 + \alpha\|\boldsymbol{x}\|_2^2 \quad \text{(A2)}$$

$$\frac{(1-\alpha)\|\boldsymbol{x}\|_2 + \alpha\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2} \leq \frac{(1-\alpha)\|\boldsymbol{x}\|_1 + \alpha\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2} \leq \frac{(1-\alpha)\sqrt{J}\|\boldsymbol{x}\|_2 + \alpha\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2}$$

$$(1-\alpha) + \alpha\|\boldsymbol{x}\|_2 \leq \tau \leq (1-\alpha)\sqrt{J} + \alpha\|\boldsymbol{x}\|_2$$

Due to the approach of the constrained problem of optimization, the solution $\boldsymbol{x} \in \mathfrak{B}_{\ell_2}(1)$, and therefore, $\|\boldsymbol{x}\|_2 = 1$. Thus:

$$(1-\alpha) + \alpha \leq \tau \leq (1-\alpha)\sqrt{J} + \alpha \quad \text{(A3)}$$

As shown in Fig. 5, we restrict τ_1 and τ_2 to the range $1 \leq \tau \leq (1-\alpha)\sqrt{J}+\alpha$ for any $\alpha \in [0,1]$. Obviously, if $\alpha = 1$, $\tau = 1$, and the solution leads to $\mathfrak{B}_{\ell_2}(1)$. If $\alpha = 0$, the tuning parameter will be restricted to $1 \leq \tau \leq \sqrt{J}$, as evidenced in (Guillemot et al. 2019; Witten et al. 2009).

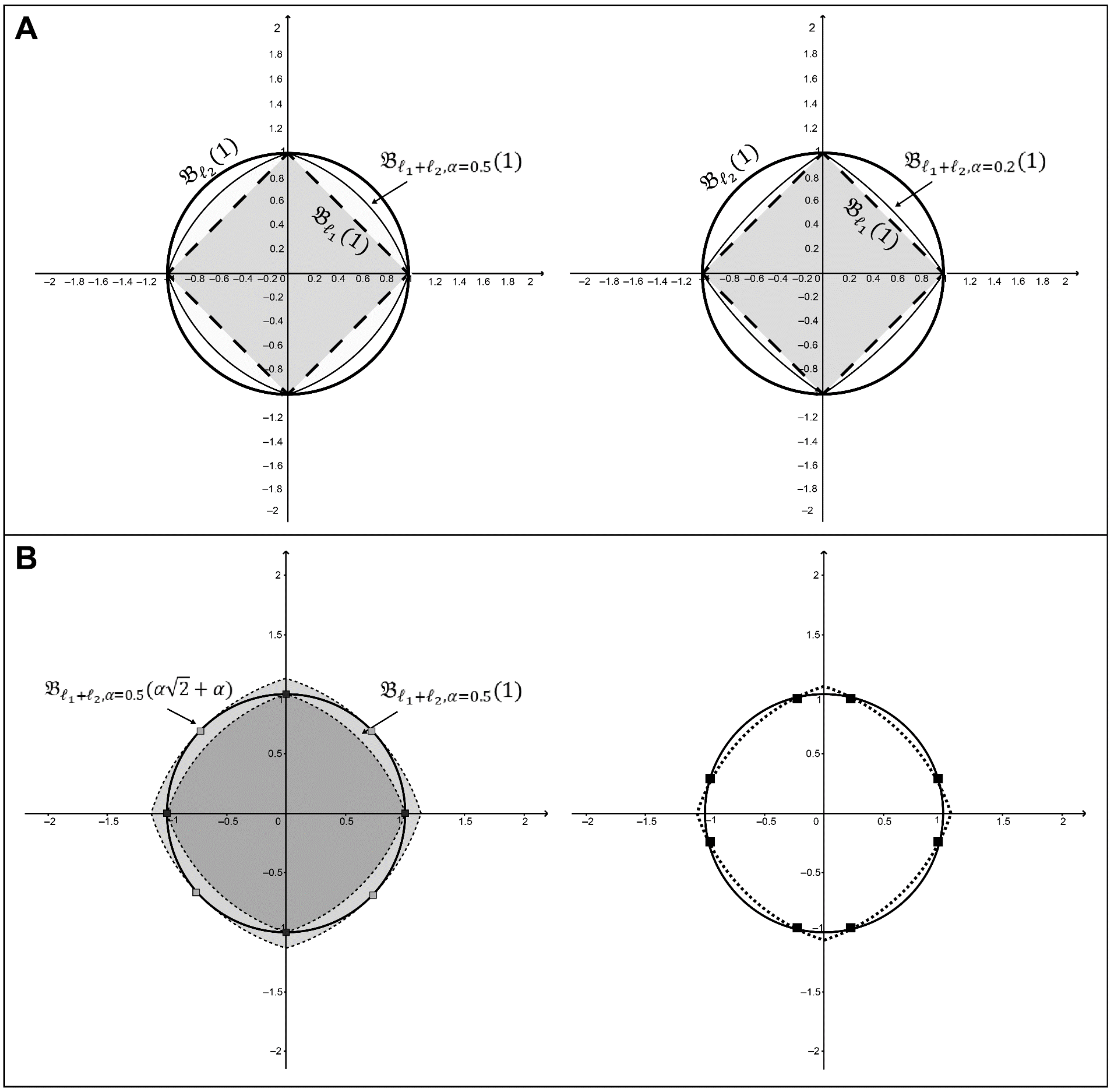

Fig. **1** A graphical display in $\mathbb{R}^2$ of the ℓ_1, ℓ_2 and $\ell_1 + \ell_2$ constraints on a vector $\boldsymbol{x}$ for the projection onto the $\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1)$ ball of the $\mathrm{C_{enet}SVD}$ procedure. The ℓ_2 constraints restrict $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_2^2 \leq 1$, and the ℓ_1 region, $\|\boldsymbol{x}\|_1 \leq \tau$. In the case of the elastic net ball constraint, $\boldsymbol{x}$ is required to be $(1-\alpha)\|\boldsymbol{x}\|_1 + \alpha\|\boldsymbol{x}\|_2^2 \leq \tau$. Panel **A** shows the $\mathfrak{B}_{\ell_2}(1), \mathfrak{B}_{\ell_1}(1)$ and $\mathfrak{B}_{\ell_1+\ell_2}(1)$ ball constraints on different values of α in the enet ball (left, $\alpha = 0.5$; right, $\alpha = 0.2$). It is shown that the higher the values of α are, the more similar an enet restriction ball is to the $\mathfrak{B}_{\ell_2}$ constraint. Panel B (left) reflects the constraints $(1-\alpha)\|\boldsymbol{x}\|_1 + \alpha\|\boldsymbol{x}\|_2^2 \leq 1$ and $(1-\alpha)\|\boldsymbol{x}\|_1 + \alpha\|\boldsymbol{x}\|_2^2 \leq (0.5\sqrt{2} + 0.5)$ using dashed lines and $\|\boldsymbol{x}\|_2^2 \leq 1$ using a solid circle. The rectangles refer to the possible point solution at the intersection of both restrictions. In panel B (right), feasible solutions for $1 \leq \tau \leq 0.5\sqrt{J} + 0.5$ are shown ($J = 2$), where the Ridge and enet ball constraints are active.

B. General solution of the projection onto the $\mathfrak{B}_{\ell_1+\ell_2}(\tau)$ ball

Given $\boldsymbol{x} \in \mathbb{R}^J$, the projection of a vector $\boldsymbol{y} = \mathrm{P}_{\tau}^{\mathfrak{B}_{\ell_1+\ell_2}}(\boldsymbol{x}) \in \mathbb{R}^J$ is defined as the solution of the constrained optimization problem:

$$\boldsymbol{y} := \mathrm{P}_{\tau}^{\mathfrak{B}_{\ell_1+\ell_2}}(\boldsymbol{x}) = \left\{ \underset{\boldsymbol{y} \in \mathbb{R}^J}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \,; s.t{:}\ (1-\alpha)\|\boldsymbol{y}\|_1 + \alpha\|\boldsymbol{y}\|_2^2 \leq \tau \right\} \qquad \text{(A4)}$$

where $\alpha \in [0,1]$. The function $(1-\alpha)\|\boldsymbol{y}\|_1 + \alpha\|\boldsymbol{y}\|_2^2$ defines the elastic net penalty as a convex combination of the Lasso and Ridge convex penalties, possessing the optimal properties of both operators. Assume that $\boldsymbol{x} \geq 0$ is necessary to construct the dual optimization problem from the Lagrange function associated with $\lambda \geq 0$:

$$\text{Argmax } \mathrm{L}(\boldsymbol{y}, \lambda) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \lambda(1-\alpha)\|\boldsymbol{y}\|_1 + \alpha\lambda\|\boldsymbol{y}\|_2^2 - \lambda\tau \qquad \text{(A5)}$$

The constrained sparse vector solution $\boldsymbol{y} := \mathrm{P}_{\tau,1}^{\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}}(\boldsymbol{x})$ of the previous optimization problem is determined by**:**

$$\boldsymbol{y} = \mathrm{Proj}_{\lambda(\alpha\|\cdot\|_1 + (1-\alpha)\|\cdot\|_2^2)}(\boldsymbol{x}) = \left(\mathrm{Proj}_{\lambda\alpha\|\cdot\|_2^2} {}^\circ \mathrm{Proj}_{\lambda(1-\alpha)\|\cdot\|_1}\right)(\boldsymbol{x})$$

$$= \frac{\mathrm{sign}(\boldsymbol{x}_j)(|\boldsymbol{x}_j| - \lambda(1-\alpha))_+}{1 + 2\lambda\alpha} = \frac{\mathrm{S}_{\lambda(1-\alpha)}(\boldsymbol{x})}{1 + 2\lambda\alpha} \qquad \text{(A6)}$$

with $\mathrm{S}_{\lambda(1-\alpha)}(\boldsymbol{x})$ being the soft-thresholding operator.

C. Projection onto the $\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1)$ ball

Given $\boldsymbol{x} \in \mathbb{R}^J$, the projection of a vector $\boldsymbol{y} = \mathrm{P}_{\tau,1}^{\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}}(\boldsymbol{x}) \in \mathbb{R}^J$ is defined as the solution of the constrained optimization problem:

$$\boldsymbol{y} := \mathrm{P}_{\tau,1}^{\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}}(\boldsymbol{x}) = \left\{ \underset{\boldsymbol{y} \in \mathbb{R}^J}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \,; s.t{:}\ (1-\alpha)\|\boldsymbol{y}\|_1 + \alpha\|\boldsymbol{y}\|_2^2 \leq \tau, \|\boldsymbol{y}\|_2 \leq 1 \right\} \qquad \text{(A7)}$$

We extend the algorithm proposed for the projection onto the intersection of the L1 and L2 balls (Guillemot et al., 2019) to the function:

$$\Omega(\lambda) = \frac{(1-\alpha)\left\|\mathrm{S}^*_{\lambda(1-\alpha)}(\widetilde{\boldsymbol{x}})\right\|_1 + \alpha\left\|\mathrm{S}^*_{\lambda(1-\alpha)}(\widetilde{\boldsymbol{x}})\right\|_2^2}{\left\|\mathrm{S}^*_{\lambda(1-\alpha)}(\widetilde{\boldsymbol{x}})\right\|_2} \tag{A8}$$

where $\widetilde{\boldsymbol{x}} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_J)$ is a vector composed of the absolute values of $\boldsymbol{x}$, ordered in a decreasing way, and

$$\mathrm{S}^*_{\lambda(1-\alpha)}(\widetilde{\boldsymbol{x}}) = \frac{\mathrm{S}_{\lambda(1-\alpha)}(\widetilde{\boldsymbol{x}})}{1+2\lambda\alpha} \tag{A9}$$

$\Omega(\lambda)$ is a continuous and decreasing function from $\Omega(0) = ((1-\alpha)\|\widetilde{\boldsymbol{x}}\|_1 + \alpha\|\widetilde{\boldsymbol{x}}\|_2^2)/\|\widetilde{\boldsymbol{x}}\|_2$ to $\Omega(\tilde{x}_1) = 0$. Due to the properties of Ω, there is a positive scalar $\lambda > 0$ such that $\Omega(\lambda) = \tau$, with $\lambda \in [\tilde{x}_1, \tilde{x}_J]$ and $\tau \in [\Omega(\tilde{x}_1), \Omega(0)]$. Thus, $\forall \tau \in [\Omega(\tilde{x}_1), \Omega(0)] = [1, (1-\alpha)\sqrt{J} + \alpha\|\widetilde{\boldsymbol{x}}\|_2]$ exists with $k \in \mathbb{Z}$, $k \leq J$, such that

$$\Omega(\widetilde{\boldsymbol{x}}_k) \leq \tau < \Omega(\widetilde{\boldsymbol{x}}_{k+1}) \tag{A10}$$

Due to the ℓ_1, ℓ_2 and $\ell_1 + \ell_2$ norm formulation and basic algebraic operations, determining the value of λ that verifies $\Omega(\lambda) = \tau$ is equivalent to finding $\lambda \in [\tilde{x}_1, \tilde{x}_J]$ as the solution to a fourth-degree polynomial equation:

$$\begin{aligned}(\tau^2 l_2 - \alpha^2 l_2^2 - 2\alpha l_1 l_2(1-\alpha) - l_1^2(1-\alpha)^2) + \lambda \\ \cdot (4\alpha l_2\tau^2 - 2\tau^2 l_1(1-\alpha) + 2k l_2\alpha(1-\alpha)^2 + 2k l_1(1-\alpha)^3) + \lambda^2 \\ \cdot (k\tau^2(1-\alpha)^2 + 4\alpha^2 l_2\tau^2 - 8\alpha\tau^2(1-\alpha)l_1 + 2l_1 k\alpha(1-\alpha)^3 \\ + 2l_2 k\alpha^2(1-\alpha)^2 - k^2(1-\alpha)^4) + \lambda^3 \\ \cdot (4k\alpha\tau^2(1-\alpha)^2 - 2k^2\alpha(1-\alpha)^4 - 8\alpha^2\tau^2(1-\alpha)l_1) + \lambda^4 \\ \cdot (4k\alpha^2\tau^2(1-\alpha)^2 - k^2\alpha^2(1-\alpha)^4) = 0\end{aligned} \tag{A11}$$

with $l_1 = \sum_{i=1}^{k} \widetilde{\boldsymbol{x}}_i$ and $l_2 = \sum_{i=1}^{k} \widetilde{\boldsymbol{x}}_i^{\,2}$. Once λ is found, the vector solution $\boldsymbol{y} = \mathrm{P}_{\tau,1}^{\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}}(\boldsymbol{x}) \in \mathbb{R}^J$ corresponds to:

$$\boldsymbol{y} = \frac{\mathrm{S}_{\lambda(1-\alpha)}(\widetilde{\boldsymbol{x}})}{1+2\lambda\alpha} \tag{A12}$$

The whole process is summarized as a four-step method:

1. Compute the vector $\widetilde{\boldsymbol{x}}$ of $\boldsymbol{x}$ coefficients ordered as absolute values.
2. Find k that verifies (A5).
3. Calculate λ as the solution to (A6).
4. Determine the solution (A12).

Table S1 displays the pseudocode for an efficient projection of $\boldsymbol{x}$ onto the $\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1)$ constraint. Lines 13 to 20 and 28-29 are modified from Guillemot et al (2019) to address the problem of projection onto the intersection of the elastic net and ℓ_2 norm balls.

Table S1. Projection onto the $\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}$ space algorithm

Projection of a vector onto $\mathfrak{B}_{\ell_1+\ell_2}(\tau) \cap \mathfrak{B}_{\ell_2}(1)$		
Input:		$\boldsymbol{x} \in \mathbb{R}^J, \tau \in \mathbb{R}, \tau \in [1, \sqrt{J}]$
Output		$\boldsymbol{y} = P_\tau^{\mathfrak{B}_{\ell_1+\ell_2} \cap \mathfrak{B}_{\ell_2}}(\boldsymbol{x})$
Initialization:		$s_1 = 0, s_2 = 0, nb = 0, p = \lvert \boldsymbol{x}^* \rvert$
		being $\boldsymbol{x}^* = \{x_i \in \boldsymbol{x}, i = 1, \ldots, J / x_i \neq 0\}$
1:	**If** $\|\boldsymbol{x}\|_2 = 0$ **then** $\boldsymbol{x} = \boldsymbol{v}$	
2:	**If** $\|\boldsymbol{x}\|_1 / \|\boldsymbol{x}\|_2 \leq \tau$ **then** $\boldsymbol{y} = \boldsymbol{x}$	
3:	**Else**	
4:	**While** TRUE do:	
5:		$N = \text{length}(p)$
6:		Randomly select $r \in \{1, \ldots, N\}$ with $\tilde{x}_r \neq \max(p)$
7:		$\tilde{x}_k = p[r]$
8:		# Split p into two disjoint sets:
		$H = \{p_j \,/\, p_j < \tilde{x}_k\}$
		$L = \{p_j \,/\, p_j > \tilde{x}_k\}$
9:		# Compute k as:
		$nb_{\tilde{x}_k} = \text{length}(\{p_j \,/\text{e}\, p_j == \tilde{x}_k\})$
10:		$k = nb + \text{length}(L) + nb_{\tilde{x}_k}$
11:		$s_{low_1} = \sum_{i \in L} p_i + nb_{\tilde{x}_k} \cdot \tilde{x}_k$
12:		$s_{low_2} = \sum_{i \in L} {p_i}^2 + nb_{\tilde{x}_k} \cdot {\tilde{x}_k}^2$

$$\Omega(\tilde{x}_k) = \frac{\alpha\left(s_2 + s_{low_2}\right) + (1-\alpha)(s_1 + s_{low_1}) - k\tilde{x}_k(1-\alpha)^2 - k{\tilde{x}_k}^2\alpha(1-\alpha)^2}{(1+2\alpha\tilde{x}_k)\sqrt{(s_2 + s_{low_2}) - 2\tilde{x}_k(1-\alpha)\left(s_1 + s_{low_1}\right) + k{\tilde{x}_k}^2(1-\alpha)^2}}$$

If $\Omega(\tilde{x}_k) > \tau$ **then**

If $L = \emptyset$ **end**

Update p: $p = \{p_i\}_{i \in L}$

Else

$\tilde{x}_{k+1} = \max\left(\{p_i\}_{i \in H}\right)$

$$\Omega(\tilde{x}_{k+1}) = \frac{\alpha\left(s_2 + s_{low_2}\right) + (1-\alpha)(s_1 + s_{low_1}) - k\tilde{x}_{k+1}(1-\alpha)^2 - k{\tilde{x}_{k+1}}^2\alpha(1-\alpha)^2}{(1+2\alpha\tilde{x}_k)\sqrt{(s_2 + s_{low_2}) - 2\tilde{x}_{k+1}(1-\alpha)\left(s_1 + s_{low_1}\right) + k{\tilde{x}_{k+1}}^2(1-\alpha)^2}}$$

If $\Omega(\tilde{x}_{k+1}) > \tau$ **end**

Update p: $p = \{p_i\}_{i \in H}$

Update nb: $nb = k$

Update s_1: $s_1 = s_1 + s_{low_1}$

Update s_2: $s_2 = s_2 + s_{low_2}$

end

end

$l_1 = s_1 + s_{low_1}$

$l_2 = s_2 + s_{low_2}$

#Calculate λ:

$$\begin{aligned}&(\alpha^2 l_2^2 + 2\alpha l_2(1-\alpha)l_1 + (1-\alpha)^2 l_1^2 - \tau^2 l_2) + \lambda \\ &\cdot (4\alpha l_2\tau^2 - 2l_2k[\alpha - 2\alpha^2 + \alpha^3] - 2l_1k[1 - 3\alpha + 3\alpha^2 - \alpha^3] \\ &- 2\tau^2(1-\alpha)l_1) + \lambda^2 \\ &\cdot (k^2(1-\alpha)^4 + 4\alpha^2 l_2\tau^2 + k\tau^2(1-\alpha)^2 - 2l_2k[\alpha^2 - 2\alpha^3 + \alpha^4] \\ &- 2l_1k[\alpha - 3\alpha^2 + 3\alpha^3 - \alpha^4] - 8\alpha\tau^2(1-\alpha)l_1) + \lambda^3 \\ &\cdot (4k\alpha\tau^2(1-\alpha)^2 - 2k^2\alpha(1-\alpha)^4 - 8\alpha^2\tau^2(1-\alpha)l_1) + \lambda^4 \\ &\cdot (4k\alpha^2\tau^2(1-\alpha)^2 - k^2\alpha^2(1-\alpha)^4) = 0\end{aligned}$$

#Solution:

$$\boldsymbol{y} \coloneqq \frac{S_\lambda^*(\boldsymbol{x})}{\|\boldsymbol{x}\|_2} = \frac{\frac{S_\lambda(\boldsymbol{x})}{1+\lambda\gamma}}{\|\boldsymbol{x}\|_2} = \frac{\frac{\mathrm{sign}(\boldsymbol{x})(|\boldsymbol{x}| - \lambda(1-\alpha),0)_+}{1+2\lambda\alpha}}{\|\boldsymbol{x}\|_2}$$

end