How to Sparsify the Singular Value Decomposition with Orthogonal Components

Vincent Guillemot [1], Aida Eslami [2], Arnaud Gloaguen [3], Arthur Tenenhaus [3], & Hervé Abdi [4]

[1] *Institut Pasteur C3BI, USR 3756 IP CNRS Paris, France*
[2] *University of British Columbia, Vancouver, BC, Canada*
[3] *Laboratoire des Signaux et Systèmes, CentraleSupelec, Gif-Sur-Yvette, France*
[4] *The University of Texas at Dallas, Richardson, TX, USA*

Résumé. La décomposition en valeurs singulières (SVD) est au cœur de la plupart des méthodes multivariées. Pour extraire l'information contenue dans des tableaux de données, la SVD calcule des composantes (pour les lignes) et poids (pour les colonnes) orthogonaux . Les poids sont utilisés pour interpréter la variabilité des individus le long des composantes, et cette interprétation est grandement facilitée si la plupart de ces poids sont nuls et ce d'autant plus que les variables sont nombreuses. Il existe des méthodes qui permettent de générer des poids parcimonieux, mais ces méthodes le font, en général, au détriment de l'orthogonalité. Ici, nous proposons une nouvelle méthode, nommée CSVD, qui respecte l'orthogonalité, et l'appliquons à des données psychométriques.

Mots-clés. Décomposition en valeurs singulières (SVD), Parcimonie, LASSO, ACP

Abstract. The Singular Value Decomposition (SVD)—the core of most popular multivariate methods—analyzes a data table by generating orthogonal components (for the rows) and loadings (for the columns) that, together, extract the important information of a data table. loadings are used to interpret the corresponding components and this interpretation is greatly facilitated when only few variables have large loadings. When this pattern does not hold, several techniques can generate sparse components and loadings but, in most methods, this sparsification is obtained at the cost of orthogonality. Here we propose a new approach for the SVD that includes sparsity constraints on the columns and rows of a rectangular matrix while keeping the pseudo-singular vectors orthogonal. We illustrate this new approach with a psychometric application.

Keywords. Singular Value Decomposition (SVD), Sparsification, LASSO, PCA

1 Introduction

The singular value decomposition (SVD) underlies most popular multivariate statistical methods. To analyze data sets, the SVD generates pairwise orthogonal optimal linear combinations of the original variables called *components* or *factor scores* that extract the important information in the original data tables. The coefficients of these optimal linear combinations—called *loadings*—are used to interpret the corresponding components. Because both loadings and components are pairwise orthogonal, different sets of loadings or components do not share information and so the interpretation of the loadings and the components can be performed one set of loadings or components at a time. This interpretation is facilitated when only few variables have large loadings. If this sparse pattern does not naturally hold, several procedures can be used to select the variables important for a component. For example, the early psychometric school used rotation in the loading space. Recent approaches, by contrast, select important variables with an explicit optimization procedure such as the LASSO. Unfortunately, LASSO based sparsification methods create sparse components and loadings that are not pairwise orthogonal and this, in turn, makes the interpretation of the results more difficult because of the correlation between factors. Here we present a new sparsification based method for the SVD that incorporates orthogonality constraints on both loadings and components. First we present the standard SVD, then our new algorithm (CSVD), and finally an example on psychometric data illustrating how sparsification increases the interpretability of the components and decomposes items into meaningful groups.

2 Unconstrained Singular Value Decomposition

The SVD (see [1] whose notations we follow here) of a data matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$ of rank $L \leq \min(I, J)$ gives the solution to the following problem: How to find an optimal rank R (with $R \leq L$) approximation of $\mathbf{X}$, denoted $\widehat{\mathbf{X}}_{[R]}$. Specifically, the SVD solves the following optimization problem

$$\underset{\widehat{\mathbf{X}}_{[R]} \in \mathcal{M}(R)}{\arg\min} \ \frac{1}{2} \left\| \mathbf{X} - \widehat{\mathbf{X}}_{[R]} \right\|_F^2 = \underset{\widehat{\mathbf{X}} \in \mathcal{M}(R)}{\arg\min} \left\{ \operatorname{trace} \left(\left(\mathbf{X} - \widehat{\mathbf{X}}_{[R]} \right)^\top \left(\mathbf{X} - \widehat{\mathbf{X}}_{[R]} \right) \right) \right\}, \qquad (1)$$

which is equivalent to decomposing $\mathbf{X}$ as $\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^\top$ with $\mathbf{P}^\top\mathbf{P} = \mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ and $\boldsymbol{\Delta} = \operatorname{diag}(\boldsymbol{\delta})$ with $\delta_1 \geq \delta_2 \dots \delta_L > 0$. The $I \times R$ matrix $\mathbf{P}$ (resp. the $J \times R$ matrix $\mathbf{Q}$) stores the left (resp. right) singular vectors of $\mathbf{X}$ and the diagonal $R \times R$ matrix $\boldsymbol{\Delta}$ stores the singular values of $\mathbf{X}$. If $\mathbf{p}_\ell$ (resp. $\mathbf{q}_\ell$) denotes the ℓ-th column of $\mathbf{P}$ (resp. $\mathbf{Q}$), δ_ℓ the ℓ-th element of $\boldsymbol{\delta}$, and $\mathcal{M}(R)$ the set of all real $I \times J$ matrices of rank R, then for $R \leq L$, the optimal matrix $\widehat{\mathbf{X}}_{[R]}$ is $\widehat{\mathbf{X}}_{[R]} = \sum_{\ell=1}^{R} \delta_\ell \mathbf{p}_\ell \mathbf{q}_\ell^\top$ with $\mathbf{p}_\ell^\top \mathbf{p}_\ell = \mathbf{q}_\ell^\top \mathbf{q}_\ell = 1$ and $\mathbf{q}_\ell^\top \mathbf{q}_{\ell'} = \mathbf{p}_\ell^\top \mathbf{p}_{\ell'} = 0$, for all $\ell \neq \ell'$.

A classic (non-optimal) algorithm for the SVD of $\mathbf{X}$ is based on the "power method" (originally developed for the eigen-decomposition) which provides the first singular triplet (i.e., the first singular value and first left and right singular vectors). To ensure orthogonality between singular vectors, the first rank-1 approximation of $\mathbf{X}$, computed as $\widehat{\mathbf{X}}_{[1]} = \delta_1 \mathbf{p}_1 \mathbf{q}_1^\top$, is subtracted from $\mathbf{X}$. This procedure—called deflation—gives the new matrix $\mathbf{X}^{(1)} = \mathbf{X} - \delta_1 \mathbf{p}_1 \mathbf{q}_1^\top$, orthogonal to $\widehat{\mathbf{X}}_{[1]}$. The power method is then applied to the deflated matrix $\mathbf{X}^{(1)}$, giving a second rank-1 approximation denoted $\delta_2 \mathbf{p}_2 \mathbf{q}_2^\top$. The deflation is then applied to $\mathbf{X}^{(1)}$ to give a new residual matrix $\mathbf{X}^{(2)}$ orthogonal to $\mathbf{X}^{(1)}$, and so on, until $\mathbf{X}$ is completely decomposed. This way, the problem of Eq. 1 becomes:

$$\underset{\delta_\ell, \mathbf{p}_\ell, \mathbf{q}_\ell}{\arg\min} \frac{1}{2} \left\| \mathbf{X} - \sum_{\ell=1}^{R} \delta_\ell \mathbf{p}_\ell \mathbf{q}_\ell^\top \right\|_F^2 \text{ subject to } \begin{cases} \mathbf{p}_\ell^\top \mathbf{p}_\ell = \mathbf{q}_\ell^\top \mathbf{q}_\ell = 1 \\ \mathbf{p}_\ell^\top \mathbf{p}_{\ell'} = \mathbf{q}_\ell^\top \mathbf{q}_{\ell'} = 0, \end{cases} \quad \forall\ \ell' \neq \ell\ . \tag{2}$$

3 Constrained SVD (CSVD)

The constrained SVD (CSVD) still decomposes $\mathbf{X}$ into "pseudo"-singular vectors (and values), but with additional constraints that induce sparsity of the weights. Although the theory of sparsity-inducing constraints is well documented, we present a general formulation that could also be applied to other types of sparsification as well as more sophisticated constraints. We consider the following optimization problem:

$$\underset{\delta_\ell, \mathbf{p}_\ell, \mathbf{q}_\ell}{\arg\min} \frac{1}{2} \left\| \mathbf{X} - \sum_{\ell}^{R} \delta_\ell \mathbf{p}_\ell \mathbf{q}_\ell^\top \right\|_F^2 \text{ subject to } \begin{cases} \mathbf{p}_\ell^\top \mathbf{p}_\ell \leq 1 \\ \mathbf{p}_\ell^\top \mathbf{p}_{\ell'} = 0 \\ \mathbf{q}_\ell^\top \mathbf{q}_\ell \leq 1 \\ \mathbf{q}_\ell^\top \mathbf{q}_{\ell'} = 0 \end{cases} \quad \forall \ell' \neq \ell \text{ and to } \begin{cases} C_1(\mathbf{p}_\ell) \leq c_{1,\ell} \\ C_2(\mathbf{q}_\ell) \leq c_{2,\ell} \end{cases} \tag{3}$$

where C_1 and C_2 are convex penalty functions from $\mathbb{R}^I$ (resp. $\mathbb{R}^J$) to $\mathbb{R}^+$, (which could be, e.g., the LASSO or the group-LASSO), and with $c_{1,\ell}$ and $c_{2,\ell}$ being positive constants. Note that for all the constraints to be active, parameter $c_{1,\ell}$ (resp. $c_{2,\ell}$) has to take its value between 1 and $\sqrt{I}$ (resp. $\sqrt{J}$).

We can show that Eq. 3 defines a biconcave maximization problem with convex constraints. This problem can be solved using Block Relaxation, an efficient alternating procedure. This iterative algorithm consists in a series of two-part iterations in which (Part 1) the expression in Eq. 3 is maximized for $\mathbf{p}$ with $\mathbf{q}$ being fixed, and is then (Part 2) maximized for $\mathbf{q}$ with $\mathbf{p}$ being fixed. Part 1 of the iteration can be re-expressed as the following optimization problem:

$$\underset{\mathbf{p}}{\arg\min} \left\{ \left\| \mathbf{p} - \mathbf{X}\mathbf{q} \right\|_2^2 \right\} \text{ subject to } \mathbf{p} \in \mathscr{B}_{L_2}(1) \cap \mathscr{B}_{L_1}(c_1) \cap \mathbf{P}^\perp, \tag{4}$$

with $\mathbf{P}^\perp$ the space orthogonal to the previously estimated left vectors, the L_1-ball (respectively L_2) of radius ρ is denoted $\mathscr{B}_{L_1}(\rho) = \left\{ \mathbf{x} \,\middle|\, \|x\|_1 \leq \rho \right\}$ (respectively $\mathscr{B}_{L_2}(\rho) = \left\{ \mathbf{x} \,\middle|\, \|x\|_2 \leq \rho \right\}$). Eq. 4 shows that finding the optimal value for $\mathbf{p}$ (i.e., Part 1 of the alternating procedure) is equivalent to finding the projection of the vector $\mathbf{X}\mathbf{q}$ onto the subspace of $\mathbb{R}^I$ defined by the intersection of all the convex sets involved by the constraints. During Part 2, $\mathbf{p}$ is fixed and therefore Part 2 can be expressed as:

$$\underset{\mathbf{q}}{\arg\min} \left\{ \frac{1}{2} \left\| \mathbf{q} - \mathbf{X}^\top \mathbf{p} \right\|_2^2 \right\} \text{ subject to } \mathbf{q} \in \mathscr{B}_{L_2}(\mathbf{1}) \cap \mathscr{B}_{L_1}(c_2) \cap \mathbf{Q}^\perp\ . \tag{5}$$

Solving Eq. 5 requires the projection of the vector $\mathbf{X}^\top \mathbf{p}$ onto the intersection of the convex sets representing the constraints. Finally, because the intersection of several convex sets is also a convex set [3], the block relaxation algorithm is essentially composed of sequential series applied until convergence of the two projections onto their respective convex sets. It is important to note that, because of the non-linearity introduced by the L_1 constraint, it is not possible anymore to impose the orthogonality constraint with deflation.

The resulting algorithm is presented in Algorithm 1. The projection step is performed with a procedure called POCS (Projection Onto Convex Sets) that is adapted to the projection onto the intersection of multiple convex sets: here, an L_2 ball, an L_1 ball, and the orthogonal subspace to the space defined by the previously computed "pseudo"-singular vectors. To reduce computational time, we used a simple and fast algorithm [4] for the projection onto the intersection of an L_2 ball and an L_1 ball based on the soft-thresholding operator.

```
Data: X, ε, R
Result: SVD of X
Define P = 0;
Define Q = 0;
for ℓ = 1, ..., R do
    p^(0) and q^(0) are randomly initialized;
    δ^(0) ← 0;
    δ^(1) ← p^(0)⊤ X q^(0);
    s ← 0;
    while |δ^(s+1) − δ^(s)| ≥ ε do
        p^(s+1) ← proj(X q^(s), 𝓑_1(c_{1,ℓ}) ∩ 𝓑_2(1) ∩ P^⊥);
        q^(s+1) ← proj(X^⊤ p^(s+1), 𝓑_1(c_{2,ℓ}) ∩ 𝓑_2(1) ∩ Q^⊥);
        δ^(s+1) ← p^(s+1)⊤ X q^(s+1);
        s ← s + 1 ;
    end
    δ_ℓ ← δ^(s+1);
    P ← vec(P, p^(s+1));
    Q ← vec(Q, q^(s+1));
end
```

Algorithm 1: General algorithm of the Constrained Singular Value Decomposition.

4 A Psychometric Example on Mental Imagery

The data set comes from a large project exploring components of human memory for which (self-selected) participants filled in several questionnaires on a web-based application in which participants rated their agreement to statements using a 5 point rating scale. This study was approved by Baycrests ethics board. Here, we analyze a psychometric instrument measuring mental imagery called the Object-Spatial-Verbal Imagery Questionnaire (OSVIQ) [2] which consists in three groups of 15 questions designed to evaluate three factors of mental imagery corresponding respectively to: 1) object 2) spatial, and 3) verbal imagery. Because OSVIQ was designed to evaluate three independent types of imagery, we expect to find three major dimensions in the data with the pattern of loadings on these dimensions reflecting their dissociation. Figures 1a and 1b show, however, that the loadings from a plain PCA did not match this expectation. By contrast, when we apply the CSVD, the pattern of loadings shows—on the first four dimensions—a clear dissociation of the three types of imagery (see Figures 1c, 1d and 1e) and identifies items that could be considered as "unpure". The scree plots of the analyses with and without sparsification (see Figure 1f), confirm that sparsity creates three components of almost equal "pseudo"-variance. This pattern was obtained by finely tuning the value of the sparsity parameter.

5 Conclusion and perspectives

The results obtained on this psychometric example indicated that the conjunction of the sparsification along with the orthogonality constraint was able to reveal theoretically meaningful patterns in the data. Interestingly, to achieve the goal obtained by the CSVD, the traditional psychometric approach would use rotation methods (e.g., VARIMAX) that, here, will require—in order to achieve an approaching result—to first estimate the "true" dimensionality of the data followed by a data driven step of variable pruning.

References

[1] Hervé Abdi. Singular value decomposition (svd) and generalized singular value decomposition (gsvd). In N.J. Salkind, editor, *Encyclopedia of Measurement and Statistic*, pages 907–912. Sage, Thousand Oaks (CA), 2007.

[2] Olessia Blajenkova, Maria Kozhevnikov, and Michael A. Motes. Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology*, 20(2):239–263, mar 2006.

[3] Stephen P. Boyd and Lieven. Vandenberghe. *Convex optimization.* Cambridge University Press, Cambridge, 1st edition, 2004.

[4] Arnaud Gloaguen, Vincent Guillemot, and Arthur Tenenhaus. An efficient algorithm to satisfy ℓ_1 and ℓ_2 constraints. In *49èmes Journées de statistique*, Avignon, France, 2017.

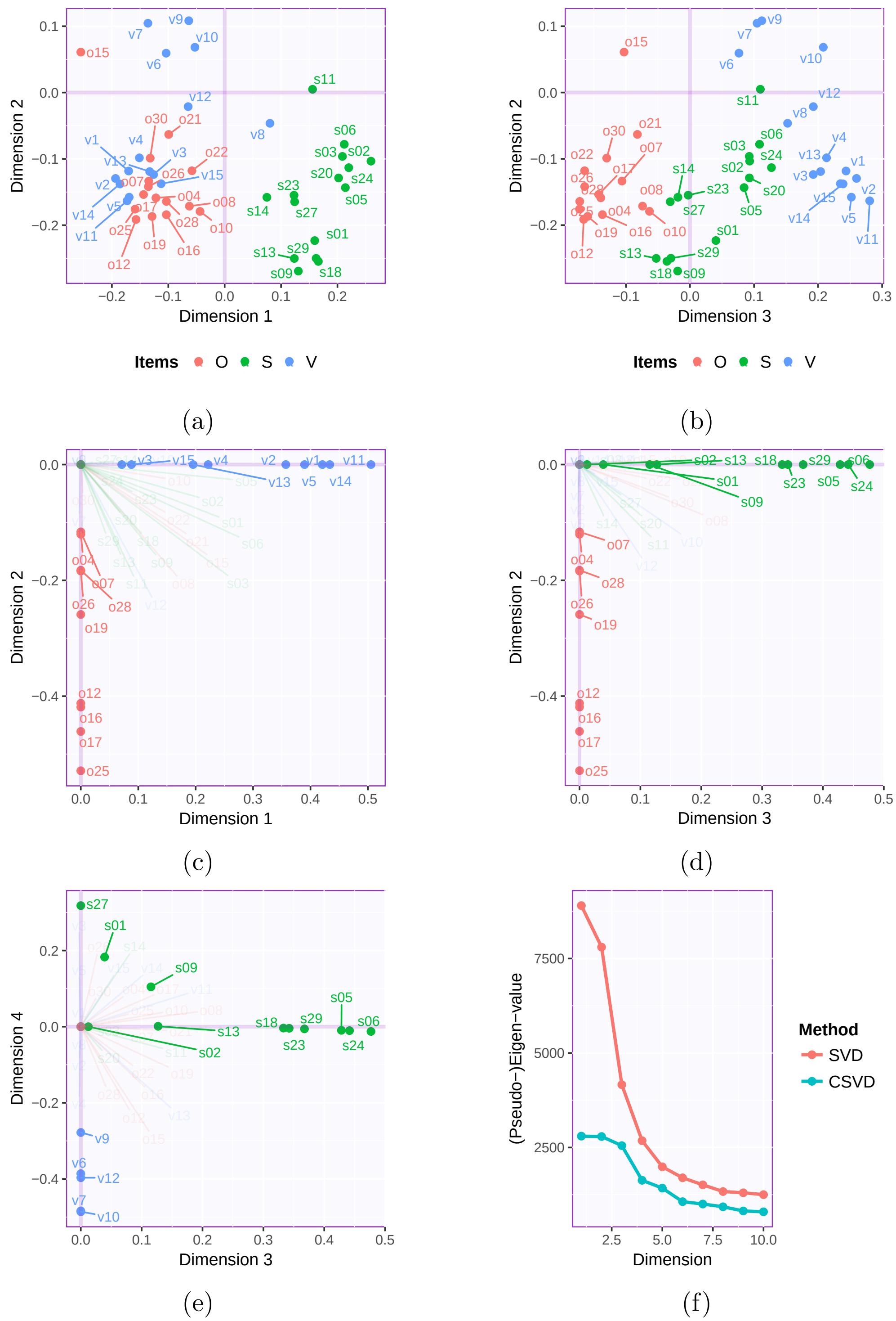

Figure 1: (a) SVD Dimensions 1 and 2 (b) SVD Dimensions 2 and 3 (c) CSVD Dimensions 1 and 2 (d) CSVD Dimensions 2 and 3 (e) CSVD Dimensions 3 and 4 (f) Scree and pseudo scree.