Sparse Principal Component Analysis in Hyperspectral Change Detection

Allan A. Nielsen[a], Rasmus Larsen[b] and Jacob S. Vestergaard[b]

[a]Technical University of Denmark
DTU Space – National Space Institute
DK-2800 Kgs. Lyngby, Denmark

[b]Technical University of Denmark
DTU Informatics – Informatics and Mathematical Modelling
DK-2800 Kgs. Lyngby, Denmark

ABSTRACT

This contribution deals with change detection by means of sparse principal component analysis (PCA) of simple differences of calibrated, bi-temporal HyMap data. Results show that if we retain only 15 nonzero loadings (out of 126) in the sparse PCA the resulting change scores appear visually very similar although the loadings are very different from their usual non-sparse counterparts. The choice of three wavelength regions as being most important for change detection demonstrates the feature selection capability of sparse PCA.

Keywords: Airborne remote sensing, HyMap, Feature selection

1. INTRODUCTION

Principal component analysis (PCA) first described by Hotelling in 1933[3] is a very popular way of orthogonalizing and compressing multi- and hypervariate data. Because the principal components (PCs) are weighted linear combinations of all the original variables they are sometimes difficult to interpret. Sparse PCA carries out the transformation in a way such that some or even many weights (also known as loadings) are forced to zero. This facilitates interpretation of the resulting sparse principal component scores. Section 2 describes PCA including a sparse version. Section 3 describes an example and gives results, and Section 4 concludes.

2. PRINCIPAL COMPONENT ANALYSIS

Let us consider a data set with n observations and p variables organized as the usual data matrix X with n rows and p columns; each column contains measurements over all n observations from one variable and each row consists of a vector of measurements x_i^T from p variables for a particular observation, the superscript ^{T} denotes the transpose. Without loss of generality we assume that the variables in the columns of X have mean value zero.

In ordinary PCA we find projections $x^T u = u^T x$ of the rows of X which maximize the variance $\text{Var}\{u^T x\} = u^T S u$ with $u^T u = 1$ where S is the sample variance-covariance matrix, $S = X^T X/(n-1) = 1/(n-1)\sum_{i=1}^n x_i x_i^T$ which is p by p. This may be done by means of a Lagrange multiplier technique where we maximize $L = u^T S u - \lambda(u^T u - 1)$ without constraints by setting the partial derivatives $\partial L/\partial u = 2(Su - \lambda u)$ to zero. If $X^T X$ is rank $r \leq \min(n,p)$ this leads to r non-zero eigenvalues λ_i and r orthogonal or mutually conjugate unit length eigenvectors u_i ($u_i^T u_j = 0, i \neq j$; $u_i^T u_i = 1$) from the eigenvalue problem $Su_i = \lambda_i u_i$. We see that the sign of u_i is arbitrary. To find the principal component scores for an observation x we project x onto the eigenvectors, $x^T u_i = u_i^T x$. The variance of these scores is $u_i^T S u_i = \lambda_i u_i^T u_i = \lambda_i$.

Further author information:

A.A.N.: Located at DTU Informatics – Informatics and Mathematical Modelling, Asmussens Allé, Building 305, Tel +45 4525 3425, Fax +45 4588 1397, E-mail aa@space.dtu.dk, http://www.imm.dtu.dk/~aa. R.L.: E-mail rl@imm.dtu.dk, http://www.imm.dtu.dk/~rl. J.S.V.: E-mail jacob.vestergaard@gmail.com.
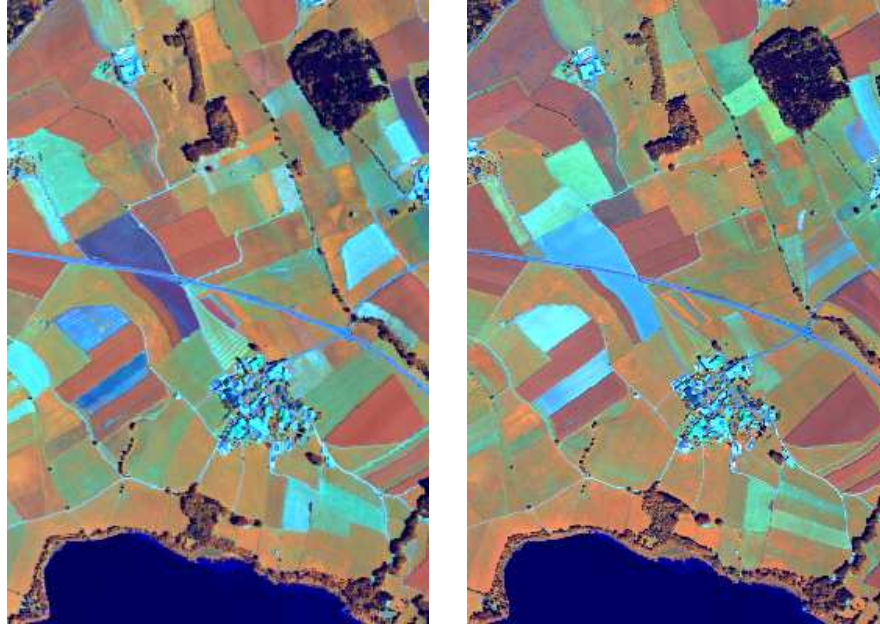
Figure 1. Left: HyMap bands 27 (828 nm), 81 (1,648 nm) and 16 (662 nm) as RGB, 30 June 2003 8:43 UTC. Right: HyMap bands 27 (828 nm), 81 (1,648 nm) and 16 (662 nm) as RGB, 4 August 2003 10:23 UTC.

2.1 Sparse PCA

Sparsity, i.e., the fact that several weights or loadings (u above) are zero is enforced by adding the constraint $\mathrm{card}(u) < m$, $m \leq p$ to the usual PCA formulation: maximize $u^T S u$ with $u^T u = 1$. Here card is the cardinality meaning the number of nonzero elements in the eigenvector. In principle this requires maximization under an L_0-norm constraint. In[2, 6–8] this is rewritten to a sparse regression problem with so-called elastic net constraints (here for the leading sparse principal component only): minimize $\sum_{i=1}^{n} \|x_i - \theta u^T x_i\|_2^2 + \lambda_2 \|u\|_2^2 + \lambda_1 \|u\|_1$ with $\|\theta\|_2^2 = 1$ and where the L_0-norm is replaced by the L_1-norm. x_i is the ith row of X.

Sparse PCA is calculated on data normalized to unit variance.

3. BI-TEMPORAL HYMAP DATA

Here we use all 126 spectral bands of 400 rows by 270 columns 5 m pixels HyMap[1] data covering a small agricultural area near Lake Waging-Taching in Bavaria, Germany. HyMap is an airborne, hyperspectral instrument which records 126 spectral bands covering most of the wavelength region from 438 to 2,483 nm with 15-20 nm spacing. Figure 1 shows HyMap bands 27 (828 nm), 81 (1,648 nm) and 16 (662 nm) as RGB acquired at 30 June 2003 8:43 UTC (left) and 4 August 2003 10:23 UTC (right). The data at the two time points were radiometrically calibrated and orthorectified using GPS/IMU measurements, a DEM and ground control points. One pixel accuracy was obtained. These data are dealt with in[4, 5] also.

Figure 2 shows the simple difference image between the August and the June images.

Figure 3 shows ordinary leading principal component loadings (the leading eigenvector, left) and scores (right).

Figure 4 shows the percentage of explained variance (PEV) of the leading sparse principal component as a function of the number of nonzero elements in the eigenvector. We see that if we choose 100 nonzero elements we sacrifice very little explained variance.

Abrupt changes in PEV occur at around 80, 55 and (to a lesser extent) 15 nonzero elements. Hence we calculate and inspect leading sparse principal component loadings (eigenvectors) and scores for 100, 80, 55 and 15 nonzero elements, see Figures 5 to 8. No-change regions have values close to zero in the simple differences
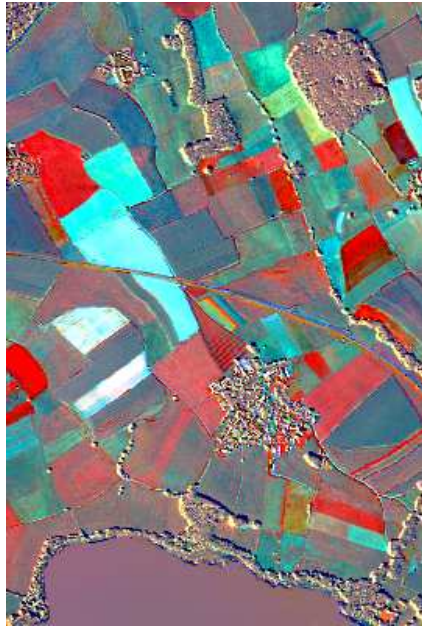
Figure 2. Simple differences: August minus June data, HyMap bands 27 (828 nm), 81 (1,648 nm) and 16 (662 nm) as RGB.
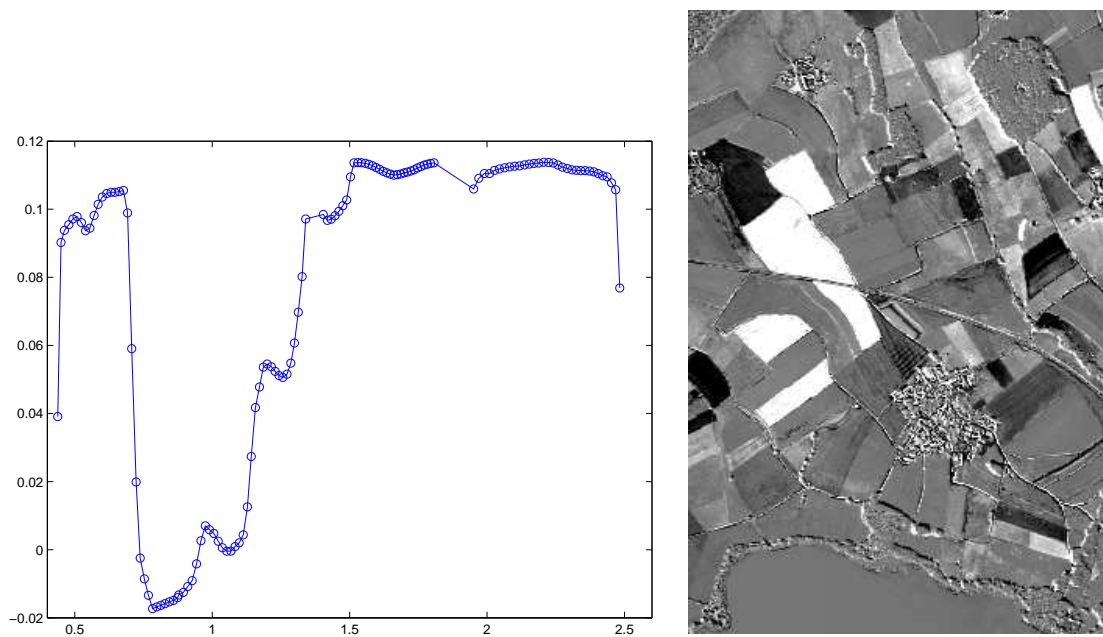

Figure 3. Left: Leading eigenvector. Right: Leading scores.

and also in the (sparse) principal component scores and thus appear middle gray. Change regions have either low or high values and appear as either dark or bright regions.

We see that the wavelength region to first get zero weights is 738–1,128 nm (Figure 5), followed by 707–1,327 nm (Figure 6), and 438–1,489 nm (Figure 7). With only 15 nonzero weights wavelengths around 1,516, 1,807 and 2,206 nm (bands 71, 94 and 109) are highlighted (Figure 8) as being important for change detection in this
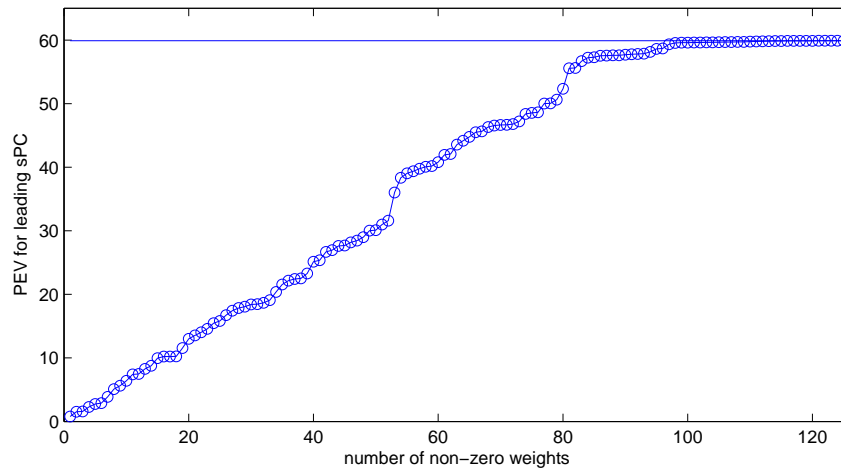
Figure 4. Percentage explained variance of leading sparse principal component vs number of nonzero elements in eigenvector.
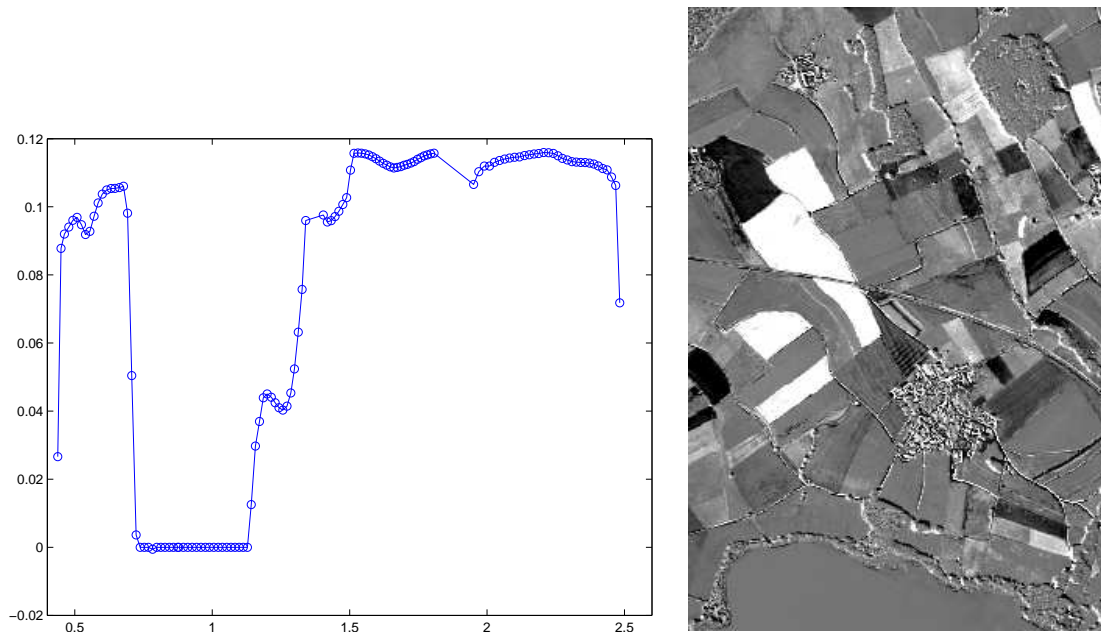

Figure 5. Left: Leading sparse eigenvector with 100 nonzero elements. Right: Leading sparse scores, 100 nonzero elements.

case. This is an example of the feature selection capability of sparse PCA.

We note that the visual differences between the usual and the sparse scores are very small for all examples shown.

Change detected over the five weeks is due to growth of the main crop types such as maize, barley and wheat. On pastures, which are constantly being grazed, in forest stands and in the lake to the south, no change is observed. Furthermore, both solar elevation and azimuth have changed which gives rise to edge effects where abrupt height differences on the ground occur.
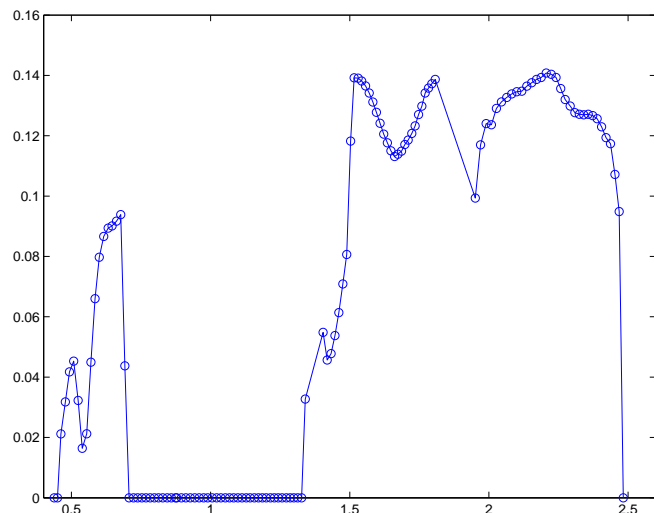
Figure 6. Left: Leading sparse eigenvector with 80 nonzero elements. Right: Leading sparse scores, 80 nonzero elements.
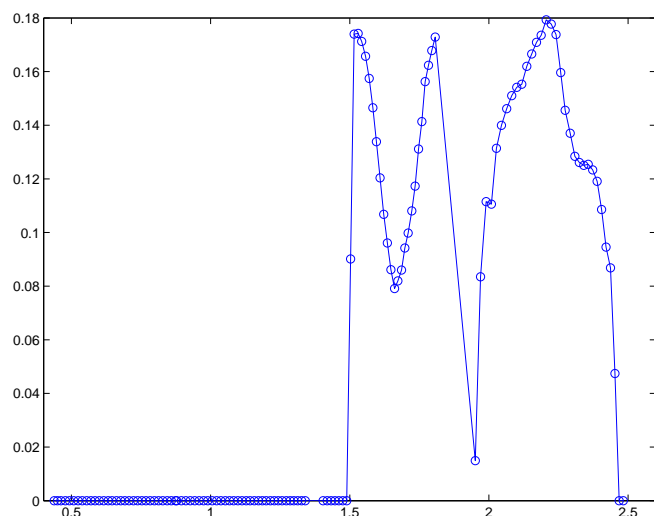

Figure 7. Left: Leading sparse eigenvector with 55 nonzero elements. Right: Leading sparse scores, 55 nonzero elements.

4. CONCLUSIONS

Because it forces the eigenvectors (or loadings) to have (several) zero elements sparse PCA facilitates the interpretation of the resulting principal components. This is demontrated in a case study with hyperspectral HyMap data from South Germany with 126 spectral bands and 400 rows by 270 columns 5 m pixels where visual differences between different scores were small in spite of very different loadings.

The fact that the sparse PCA for few (here 15) nonzero weights points to three wavelength regions as being the most important ones for the change detection, may be considered as a form of feature selection.
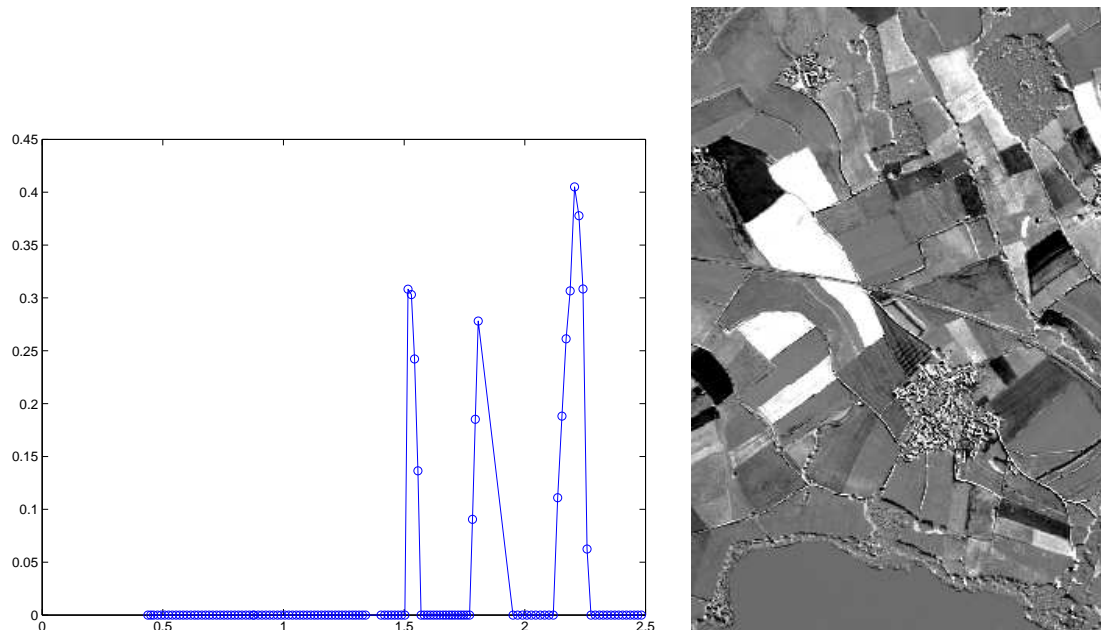
Figure 8. Left: Leading sparse eigenvector with 15 nonzero elements. Right: Leading sparse scores, 15 nonzero elements.

5. ACKNOWLEDGMENTS

REFERENCES

[1] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The HyMap airborne hyperspectral sensor: the system, calibration, and performance," in *Proceedings of 1st EARSeL Workshop on Imaging Spectroscopy,* pp. 37–42, Zürich, Switzerland, 6-8 October 1998. http://www.hyvista.com and http://www.intspec.com.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction,* second edition, Springer, 2009. http://www-stat.stanford.edu/∼hastie/Papers/ESLII.pdf.

[3] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441 and 498–520, 1933.

[4] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Transactions on Image Processing,* vol. 16, no. 2, pp. 463–468, 2007. http://www.imm.dtu.dk/pubdb/p.php?4695.

[5] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Transactions on Image Processing,* vol. 20, no. 3, pp. 612–624, 2011. http://www.imm.dtu.dk/pubdb/p.php?5925.

[6] K. Sjöstrand, "Regularized statistical analysis of anatomy," Ph.D. thesis, Technical University of Denmark, DTU Informatics, 2007. http://www.imm.dtu.dk/Om_IMM/Medarbejdere.aspx?lg=showcommon&id=202531.

[7] K. Sjöstrand, E. Rostrup, C. Ryberg, R. Larsen, C. Studholme, H. Baezner, J. Ferro, F. Fazekas, L. Pantoni, D. Inzitari, G. Waldemar, "Sparse Decomposition and Modeling of Anatomical Shape Variation," *IEEE Transactions on Medical Imaging,* vol. 26, no. 12, pp. 1625–1635, 2007. http://www.imm.dtu.dk/pubdb/p.php?5029.

[8] H. Zou, T. Hastie, and R. Tibshirani, "Sparse Principal Component Analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006. http://www-stat.stanford.edu/∼hastie/Papers/spc_jcgs.pdf.