

KERNEL METHODS IN ORTHOGONALIZATION OF MULTI- AND HYPERVARIATE DATA

Allan Aasbjerg Nielsen

Technical University of Denmark
DTU Space – National Space Institute
Richard Petersens Plads, Building 321
DK-2800 Kgs. Lyngby, Denmark
Tel +45 4525 3425, Fax +45 4588 1397
E-mail aa@space.dtu.dk, <http://www.imm.dtu.dk/~aa>

ABSTRACT

A kernel version of maximum autocorrelation factor (MAF) analysis is described very briefly, and applied to change detection in remotely sensed hyperspectral image (HyMap) data. The kernel version is based on a dual formulation also termed Q-mode analysis in which the data enter into the analysis via inner products in the Gram matrix only. In the kernel version the inner products are replaced by inner products between nonlinear mappings into higher dimensional feature space of the original data. Via kernel substitution also known as the kernel trick these inner products between the mappings are in turn replaced by a kernel function and all quantities needed in the analysis are expressed in terms of this kernel function. This means that we need not know the nonlinear mappings explicitly. Kernel PCA and MAF analyses handle nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space. An example shows the successful application of kernel MAF analysis to change detection in HyMap data covering a small agricultural area near Lake Waging-Taching, Bavaria, Germany.

Index Terms— Orthogonal transformations, dual formulation, Q-mode analysis, kernel trick, kernel MAF.

1. INTRODUCTION

Based on work by Pearson [1] in 1901, Hotelling [2] in 1933 introduced principal component analysis (PCA). PCA is often used for linear orthogonalization or compression by dimensionality reduction of correlated multivariate data, see [3] for a comprehensive description of PCA and related techniques. An interesting dilemma in reduction of dimensionality of data is the desire to obtain simplicity for better understanding, visualization and interpretation of the data on the one hand, and the desire to retain sufficient detail for adequate representation on the other hand.

[4] introduced maximum autocorrelation factor (MAF) analysis. [5] used MAF analysis to detect change in images consisting of simple differences between corresponding spectral bands acquired at two points in time. [6] introduced the minimum noise fraction (MNF) transformation. Both the MAF and the MNF transformations contain spatial elements and they are therefore (conceptually) better suited for spatial data than PCA. Several other orthogonalization techniques including canonical correlation analysis (CCA) exist; these will not be dealt with in this paper.

[7] introduced kernel PCA. [8] described kernel CCA, and [9] described kernel independent component analysis (ICA) based on

kernel CCA. [10, 11] are good references for kernel methods in general. [12, 13] describe kernel methods among many other subjects. [14] used kernel PCA for change detection in univariate image data.

In this paper we sketch a kernel version of MAF analysis (Section 2). All orthogonalization techniques including PCA and MAF can be used for different types of feature extraction in general. In this paper we apply kernel PCA and MAF analysis to detect change over time in remotely sensed images. This is done by finding the projections along the eigenvectors for data consisting of simple band-by-band differences of coregistered, calibrated variables which represent the same spectral bands covering the same geographical region acquired at two different time points (Sections 3 and 4).

In the dual formulation of PCA and MAF analysis the data enter into the problem as inner products between the observations. These inner products may be replaced by inner products between mappings of the measured variables into higher order feature space. The idea in kernel orthogonalization is to express the inner products between the mappings in terms of a kernel function to avoid the explicit use of the mappings. Both the eigenvalue problem, the centering to zero mean and the projections onto eigenvectors to find kernel scores may be expressed by means of the kernel function. Kernel orthogonalization methods handle nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space.

2. MAXIMUM AUTOCORRELATION FACTOR ANALYSIS

In maximum autocorrelation factor (MAF) analysis first suggested in [4], we maximize the autocorrelation of linear combinations, $a^T x(r)$, of zero-mean original (spatial) variables, $x(r)$, see also [5, 15, 16]. $x(r)$ is a multivariate observation at location r and $x(r + \Delta)$ is an observation of the same variables at location $r + \Delta$; Δ is a spatial displacement vector.

2.1. Primal Formulation

The autocovariance R of a linear combination $a^T x(r)$ of zero-mean $x(r)$ is

$$R = \text{Cov}\{a^T x(r), a^T x(r + \Delta)\} \quad (1)$$

$$= a^T \text{Cov}\{x(r), x(r + \Delta)\} a \quad (2)$$

$$= a^T C_{\Delta} a \quad (3)$$

where C_{Δ} is the covariance between $x(r)$ and $x(r + \Delta)$. Assuming or imposing second order stationarity of $x(r)$, C_{Δ} is independent of

location, r . Introduce the multivariate difference $x_{\Delta}(r) = x(r) - x(r + \Delta)$ with variance-covariance matrix $S_{\Delta} = 2S - (C_{\Delta} + C_{\Delta}^T)$ where S is the variance-covariance matrix of x . Since

$$a^T C_{\Delta} a = (a^T C_{\Delta} a)^T \quad (4)$$

$$= a^T C_{\Delta}^T a \quad (5)$$

$$= a^T (C_{\Delta} + C_{\Delta}^T) a / 2 \quad (6)$$

we obtain

$$R = a^T (S - S_{\Delta} / 2) a. \quad (7)$$

To get the autocorrelation ρ of the linear combination we divide the covariance by its variance $a^T S a$

$$\rho = 1 - \frac{1}{2} \frac{a^T S_{\Delta} a}{a^T S a} \quad (8)$$

$$= 1 - \frac{1}{2} \frac{a^T X_{\Delta}^T X_{\Delta} a}{a^T X^T X a} \quad (9)$$

where X is the n by p data matrix with columns x_i^T and X_{Δ} is a similarly defined matrix for x_{Δ} with zero-mean columns. C_{Δ} above equals $X^T X_{\Delta} / (n - 1)$. To maximize ρ we must minimize the Rayleigh coefficient $a^T X_{\Delta}^T X_{\Delta} a / (a^T X^T X a)$ or maximize its inverse. This is done by solving a symmetric generalized eigenvalue problem.

Unlike linear PCA, the result from linear MAF analysis is scale invariant: if x_i is replaced by some matrix transformation $T x_i$ corresponding to replacing X by $X T$, the result is the same.

As with kernel principal component analysis we use a reparameterisation $a \propto X^T b$ and the kernel trick to obtain an implicit nonlinear mapping for the MAF transform. A detailed account of this is given in [20].

3. DATA

To illustrate the techniques we use all spectral bands of 400 rows by 270 columns 5 m pixels HyMap [21] 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, 30 June 2003 8:43 UTC (top) and 4 August 2003 10:23 UTC (bottom). The data at the two time points are geometrically coregistered and radiometrically calibrated. These data are dealt with in [22, 23] also.

4. RESULTS AND DISCUSSION

To be able to carry out kernel PCA and MAF/MNF analysis on the large amounts of pixels typically present in Earth observation data, we sub-sample the image and use a small portion termed the training data only. We use typically in the order 10^3 randomly sampled training pixels to find the eigenvectors onto which we then project the entire image termed the test data kernelized with the training data. This sub-sampling potentially avoids problems that may arise from the spatial autocorrelation inherent to image data. A Gaussian kernel $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$ with σ equal to the mean distance between the observations in feature space is used.

In the change detection analysis all band-wise differences of 126 spectral bands of the HyMap are included.

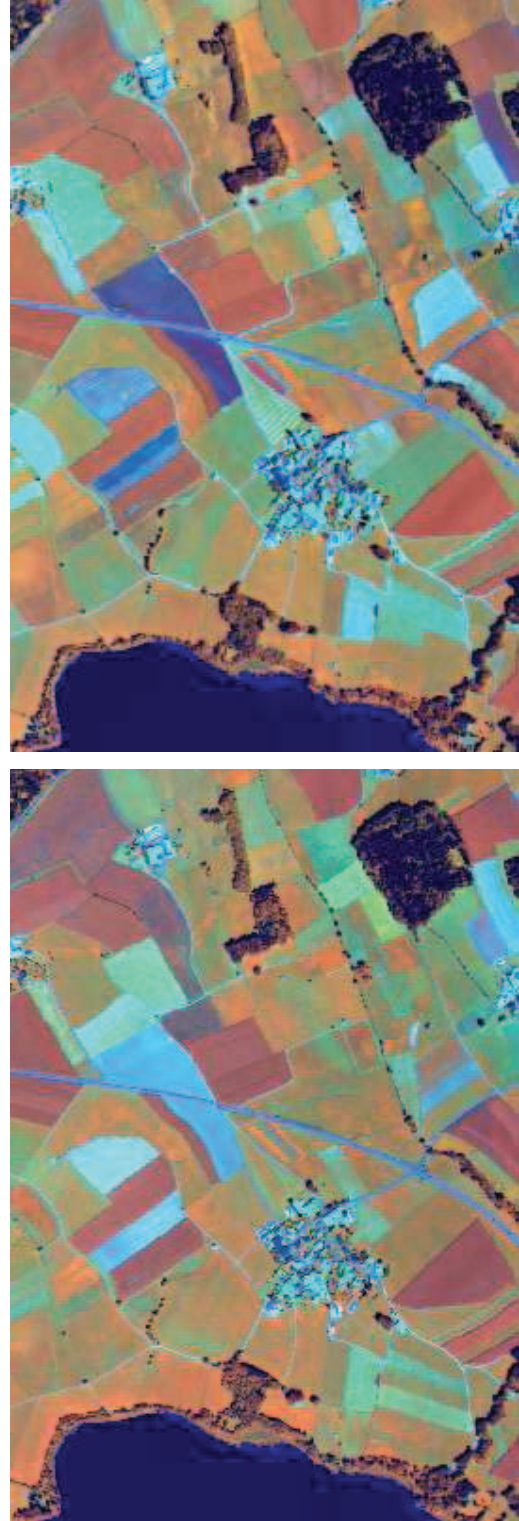


Fig. 1. HyMap bands 27 (828 nm), 81 (1,648 nm) and 16 (662 nm) as RGB, 30 June 2003 8:43 UTC (top) and 4 August 2003 10:23 UTC (bottom).

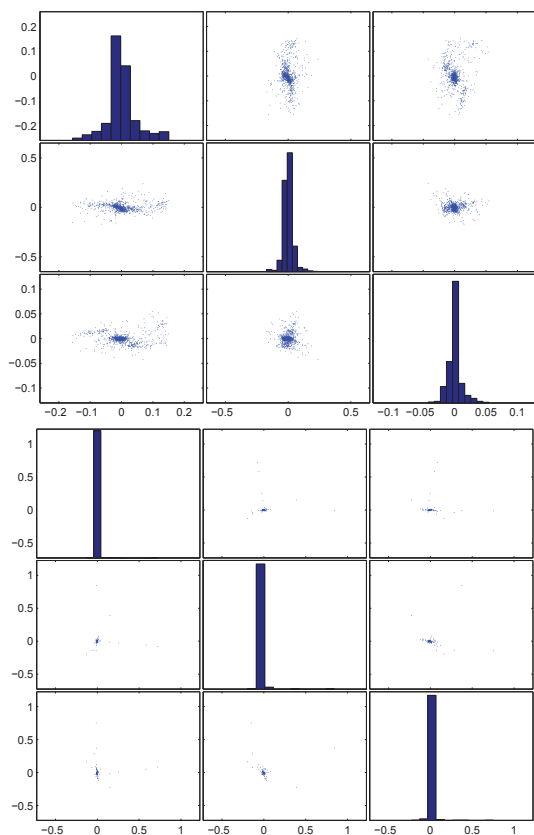


Fig. 2. Scatterplots and histograms of the first three kernel PCs (top) and kernel MAFs (bottom).

Figure 2 shows scatterplots and histograms of the first three kernel PCs (top) and kernel MAFs (bottom) for the $\sim 1,000$ training samples. We see that the histograms for the kernel MAFs are very narrow and that many more samples are concentrated in the center of the scatterplots for the kernel MAFs, i.e., we have a better isolation of the no-change observations.

Figure 3 shows kernel principal components 1–3 (top) and kernel maximum autocorrelation factors 1–3 (bottom) of simple band-by-band difference images as RGB. All bands are stretched linearly between mean minus and plus three standard deviations. In this representation no-change areas will appear as grayish and change regions will appear in saturated colours. The 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, solar effects give rise to edge effects where height differences occur (both solar elevation and azimuth have changed). We see that both types of kernel analysis emphasize change and that unlike kernel PCA, kernel MAF analysis seems to focus on the most conspicuous changes and that it gives a very strong discrimination between change and no-change regions.

Ordinary linear PCA or MAF analysis (not shown) does not give this beautiful discrimination between change and no-change regions.

The generation of three kernel MAFs for the entire image based on $\sim 1,000$ random samples calculated by Matlab code based on the `eigs` function takes around 80 seconds on a 32 bit, 2.00 GHz Intel

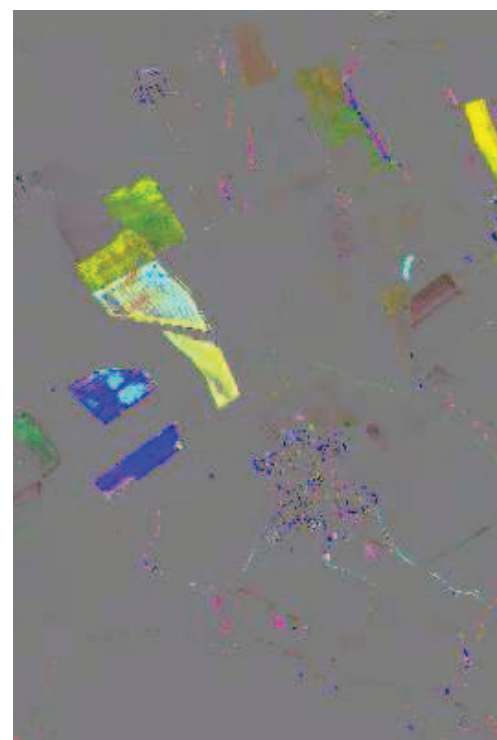
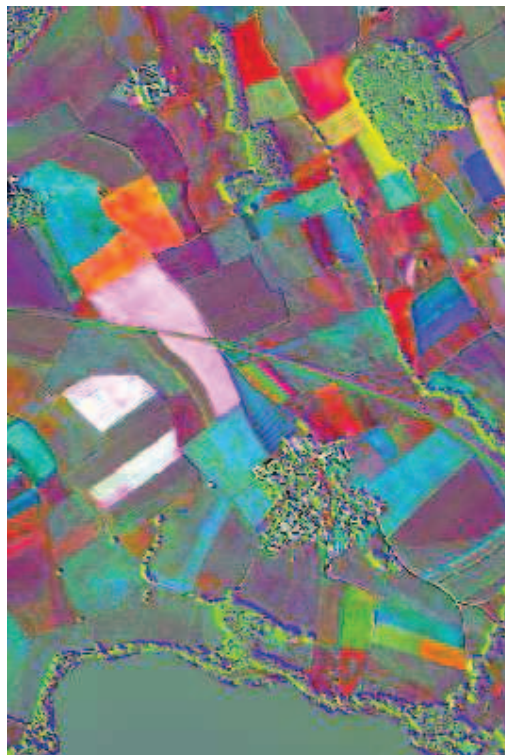


Fig. 3. Kernel principal components 1–3 (top) and kernel maximum autocorrelation factors 1–3 (bottom) of 126 simple difference images as RGB. All bands are stretched linearly between mean (which is zero) minus and plus three standard deviations.

Core 2 CPU laptop with 2.00 GB, 998 MHz memory.

5. CONCLUSIONS

Kernel orthogonalization with a Gaussian kernel $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$ is used for detecting change in coregistered, calibrated simple band-by-band difference HyMap images. Unlike ordinary linear PCA or MAF analysis, especially kernel MAF analysis gives a strong discrimination between change and no-change regions. This differencing is meaningful for calibrated or normalized data only. If the data available is not of this nature, generalized differences as described in [23, 24] may be applied.

Kernel PCA and kernel MAF analysis are so-called memory-based methods: where ordinary PCA and MAF analysis handle new observations by projecting them onto the eigenvectors found based on the training data, because of the kernelization of the new observations with the training observations, kernel PCA and kernel MAF analysis need the original data as well as the eigenvectors (and for PCA the eigenvalues) to handle new data.

It's important to realize that the information content in the original data is conveyed to a kernel method through the choice of kernel only (and possibly through a labeling of the data; this is not relevant for kernel PCA and kernel MAF analysis). For example, since kernel methods are implicitly based on inner products, any rotation by an orthogonal matrix Q of the original coordinate system will not influence the result of the analysis, $(Qx_i)^T Qx_j = x_i^T Q^T Qx_j = x_i^T x_j$.

6. ACKNOWLEDGMENT

The geometrically coregistered and radiometrically calibrated HyMap data are kindly provided by Andreas Müller and co-workers, DLR German Aerospace Center, Oberpfaffenhofen, Germany.

7. REFERENCES

- [1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 6, no. 2, pp. 559–572, 1901.
- [2] 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.
- [3] I. T. Jolliffe, *Principal Component Analysis*, second edition, Springer, 2002.
- [4] P. Switzer and A. A. Green, "Min/max autocorrelation factors for multivariate spatial imagery," Tech. Rep. 6, Department of Statistics, Stanford University, 1984.
- [5] P. Switzer and S. E. Ingebritsen, "Ordering of time-difference data from multispectral imagery," *Remote Sensing of Environment*, vol. 20, pp. 85–94, 1986.
- [6] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
- [7] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [8] P. L. Lai and C. Fyfe, "Kernel and nonlinear canonical correlation analysis," *International Journal of Neural Systems*, vol. 10, no. 5, pp. 365–377, 2000.
- [9] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2002.
- [10] B. Schölkopf and A. Smola, *Learning with Kernels*, Massachusetts Institute of Technology Press, 2002.
- [11] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, third edition, Cambridge University Press, 2007.
- [14] A. A. Nielsen and M. J. Canty, "Kernel principal component analysis for change detection," SPIE Europe Remote Sensing Conference, vol. 7109A, Cardiff, Great Britain, 15–19 September 2008, Internet <http://www.imm.dtu.dk/pubdb/p.php?5667>.
- [15] A. A. Nielsen, "Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data," Ph.D. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 1994, Internet <http://www.imm.dtu.dk/pubdb/p.php?296>.
- [16] A. A. Nielsen, K. B. Hilger, O. B. Andersen and P. Knudsen, "A temporal extension to traditional empirical orthogonal function analysis," In L. Bruzzone and P. Smits (eds.) Proceedings of MultiTemp2001, Trento, Italy, 13–14 September 2001, Analysis of Multi-Temporal Remote Sensing Images, pp. 164–170, 2002. Internet <http://www.imm.dtu.dk/pubdb/p.php?289>.
- [17] A. A. Nielsen, "An extension to a filter implementation of a local quadratic surface for image noise estimation," In Proceedings from the 10th International Conference on Image Analysis and Processing, ICIAP'99, pp. 119–124, Venice, Italy, 27–29 September 1999, Internet <http://www.imm.dtu.dk/pubdb/p.php?3937>.
- [18] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steinfeldt, "Spatial factor analysis of stream sediment geochemistry data from South Greenland," In V. Pawlowsky-Glahn (ed.) Proceedings of the Third Annual Conference of the International Association for Mathematical Geology, IAMG'97, pp. 955–960, Barcelona, Spain, 22–27 September 1997, Internet <http://www.imm.dtu.dk/pubdb/p.php?5686>.
- [19] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steinfeldt, "Maximum autocorrelation factorial kriging," In W. J. Kleingeld and D. G. Krige (eds.) Proceedings of the 6th International Geostatistics Congress, Geostats 2000, pp. 538–547, Cape Town, South Africa, 10–14 April 2000, Internet <http://www.imm.dtu.dk/pubdb/p.php?3639>.
- [20] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," submitted.
- [21] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The HyMap airborne hyperspectral sensor: the system, calibration, and performance," 1st EARSeL Workshop on Imaging Spectroscopy, pp. 37–42, Zürich, Switzerland, 6–8 October 1998, Internet <http://www.hyvista.com> and <http://www.intspec.com>.
- [22] A. A. Nielsen, A. Müller, and W. Dorigo, "Hyperspectral data, change detection and the MAD transformation," 12th Australasian Remote Sensing and Photogrammetry Conference, pp. 683–688, Fremantle, Western Australia, 18–22 October 2004. *Session keynote address*. Internet <http://www.imm.dtu.dk/pubdb/p.php?3176>.
- [23] 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. 463–478, 2007, Internet <http://www.imm.dtu.dk/pubdb/p.php?24695>.
- [24] M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 1025–1036, 2008, Internet <http://www.imm.dtu.dk/pubdb/p.php?5362>.