Optimal Class Separation in Hyperspectral Image Data: Iterated Canonical Discriminant Analysis

Allan Aasbjerg Nielsen and Andreas Müller

***Abstract*—This paper describes canonical discriminant analysis and sketches an iterative version which is then applied to obtain optimal separation between a region, here examplified by either "water" or "wood/trees" and the rest of a HyMap image. We show that the iterative version greatly enhances the separation between the regions.**

***Index Terms*—Two-class discrimination, HyMap.**

I. Introduction

THIS paper in Section II describes the established multivariate statistical technique canonical discriminant analysis [1], [2] and sketches an iterative improvement.

Section III shows the application of the iterative version to the separation between 1) "water" and "everything else", and 2) "wood/trees" and "everything else" to 126-band HyMap [3] data covering a small agricultural area in Germany. Section IV gives conclusions.

II. Theory

Here we describe canonical discriminant analysis (CDA) for optimal separation between k groups or classes of multi- or hypervariate observations. In our application we use CDA for two classes only, $k = 2$.

Also, we sketch an iterative extension of CDA and mention a method for automatic thresholding as a part of the iterative scheme.

A. Canonical Discriminant Analysis

The idea in CDA is to find projections in multi- or hypervariate feature space which give maximal separation between groups (or classes or populations) of the data.

Consider k groups with $n_1, \dots, n_k$ multivariate (p-dimensional) observations $\{\boldsymbol{x}_{ij}\}$, where i is the group index and the j is the (multivariate) observation number. The group means are denoted $\bar{\boldsymbol{x}}_1, \dots, \bar{\boldsymbol{x}}_k$ and the overall mean is denoted $\bar{\boldsymbol{x}}$, i.e.,

$$\bar{\boldsymbol{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}, \; i = 1, \dots, k$$

and

$$\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \text{ with } N = \sum_{i=1}^{k} n_i.$$

A. A. Nielsen is with DTU Space – National Space Institute, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, e-mail aa@space.dtu.dk, homepage http://www.imm.dtu.dk/∼aa. A. Müller is with the German Remote Sensing Data Center, DLR – German Aerospace Center, Wessling, D-82234, Germany, e-mail Andreas.Mueller@dlr.de.

As in a one-way analysis of variance the "total" sum of squares matrix is

$$\boldsymbol{T} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})^T.$$

We define the "among groups" (sometimes termed the "between groups") matrix as

$$\boldsymbol{A} = \sum_{i=1}^{k} n_i (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})^T$$

and the "within groups" matrix as

$$\boldsymbol{W} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)^T.$$

With these definitions we have

$$\boldsymbol{T} = \boldsymbol{A} + \boldsymbol{W}$$

or in words: the total variation can be written as a sum of the variation of the group means around the overall mean and the variation around the group means.

We are looking for projections of the original data $Y = \boldsymbol{d}^T(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})$ called canonical variates (CVs) which maximize the ratio between variation among groups and variation within groups; the latter can be considered as the natural level of variance of the variables $\boldsymbol{x}_{ij}$. The idea of maximizing this ratio is due to Fisher [1]. This ratio equals the Rayleigh quotient $\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d}/\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d} = \lambda_\ell$, i.e., the transformation is defined by the conjugate eigenvectors $\boldsymbol{d}_\ell$ of $\boldsymbol{A}$ with respect to $\boldsymbol{W}$

$$\boldsymbol{A}\boldsymbol{d}_\ell = \lambda_\ell \boldsymbol{W}\boldsymbol{d}_\ell.$$

We norm $\boldsymbol{d}_\ell$ to unit length, $\boldsymbol{d}_\ell^T\boldsymbol{d}_\ell = 1$. The higher values we obtain for λ_ℓ the higher the discriminatory power of the canonical variates. The new variates are

$$Y_\ell = \boldsymbol{d}_\ell^T(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}).$$

The first CV defined by $\boldsymbol{d}_1$ is the affine transformation of the original variables that gives the best discrimination between the k groups. A higher order CV is the affine transformation of the original variables that gives the best discrimination between the k groups subject to the constraint that it is orthogonal (with respect to $\boldsymbol{A}$ and $\boldsymbol{W}$) to the lower order CVs. Note, that the number of CVs is given by rank considerations for $\boldsymbol{A}$ and $\boldsymbol{W}$. If $\boldsymbol{A}$ and $\boldsymbol{W}$ have full rank this number equals $\min(k-1, p)$.

We define the canonical correlation coefficients R_ℓ by their squares, $R_\ell^2 = \boldsymbol{d}_\ell^T\boldsymbol{A}\boldsymbol{d}_\ell/\boldsymbol{d}_\ell^T\boldsymbol{T}\boldsymbol{d}_\ell$ which is a measure in the interval from 0 to 1 of the discriminatory power of the canonical variates. The relation between the eigenvalues λ_ℓ and the squared canonical correlations R^2 is

$$R_\ell^2 = \frac{\lambda_\ell}{\lambda_\ell + 1}.$$

In general, scatter plots of the first few CVs give a good visual impression of the separability of the observations.

In our application we use CDA for two classes only, $k = 2$; this means that we have one CV only (since the rank of $\boldsymbol{A}$ is 1). In this case we need not solve the eigenvalue problem, we can find the desired projection direction $\boldsymbol{d}_1$ from

$$\boldsymbol{d}_1 = \boldsymbol{T}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$$

and norming to unit variance.

It turns out, that in this case with two classes only, the canonical method gives the same solution as ordinary linear discriminant analysis.

B. *Iterative Extension and Automatic Thresholding*

In our application we first use CDA based on a manually selected training area which then constitutes one of the two classes; the rest of the image is the other class. This gives rise to a potential problem: the rest of the image may contain regions which actually belong to the first class. To identify such regions and to update the training area, in a series of iterations new training areas for the CDA are selected by automatically thresholding the CV calculated in the previous iteration. Iterations stop when the canonical correlation R stops improving.

The method used to automatically threshold the canonical variate is based on [4]. This method is in itself a univariate version of CDA.

III. CASE: HYMAP DATA

We show two examples based on HyMap data, one where we separate "water" from "everything else" and one slightly more difficult example where we separate "wood/trees" from "everything else".

HyMap is an airborne, hyperspectral instrument which records 126 spectral bands covering most of the wavelength region from 438 to 2483 nm with 15-20 nm spacing. In this case study we use data acquired on 30 June 2003 at 8:43 UTC covering an agricultural area near Lake Waging-Taching in Bavaria, Germany. The image consists of 400 rows by 270 columns with 5 meter pixels. The same data have been used previously in change detection studies by means of related methods in for example [5], [6], see also [7].

Figure 1 shows HyMap bands 27 (828 nm), 81 (1648 nm) and 16 (662 nm) as RGB.

A. *Water Mask*

Figure 2 shows the (16.) iterated canonical variate stretched linearly excluding the 2% extreme observations. This CV gives the optimal separation between the two classes "water" and "everything else"; there is a good separation between the very bright water pixels and land.

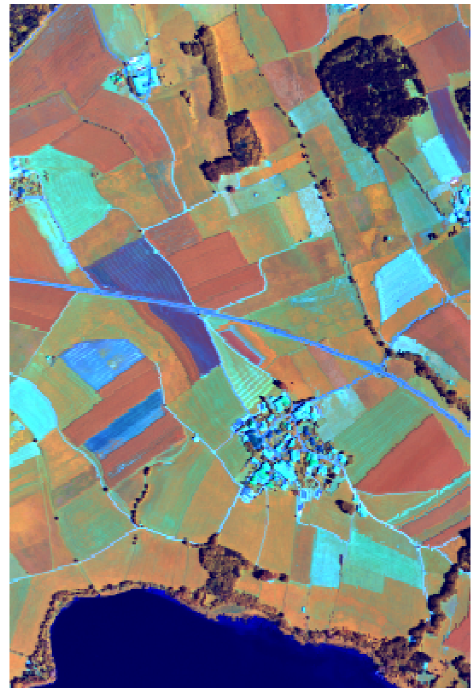

Fig. 1. HyMap bands 27 (828 nm), 81 (1648 nm) and 16 (662 nm) as RGB.

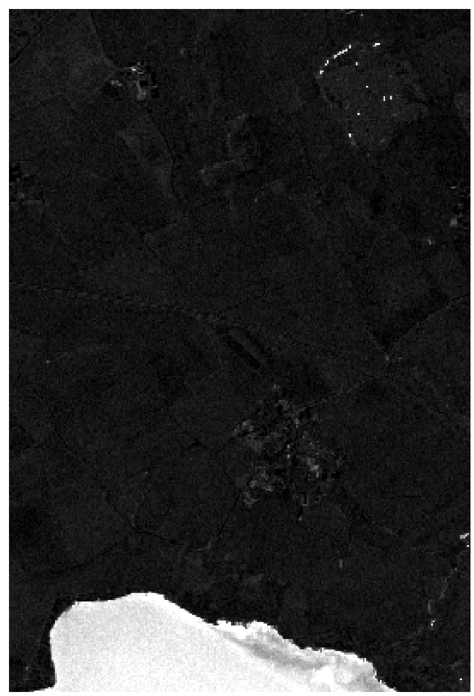

Fig. 2. Iterated canonical variate, optimal separation.

Figure 3 shows the initial hand drawn water mask and the corresponding CV stretched linearly between minimum and maximum. Figure 4 shows the water mask after 16 iterations and the corresponding CV, again stretched linearly between minimum and maximum. Note, that we seem to get all the water at the cost of a few erroneous pixels on land.

Figure 5 shows the histogram of the CV based on the hand drawn mask (and the corresponding squared canonical

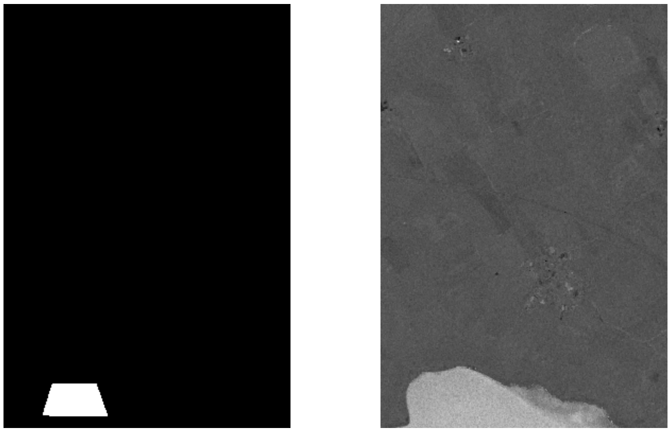

Fig. 3. Hand drawn water mask and corresponding CV.

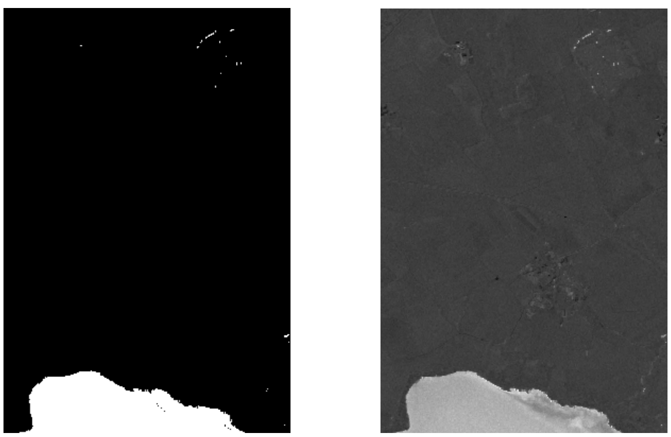

Fig. 4. Water mask, 16. iteration and corresponding iterated CV.

correlation). Figure 6 shows the histogram of the 16. iterated CV and the 16 corresponding squared canonical correlations. From the two histograms we see that a much better separation between water and land is obtained after iterations. From the plot of the canonical correlations we see that the first iteration is by far the most important one.

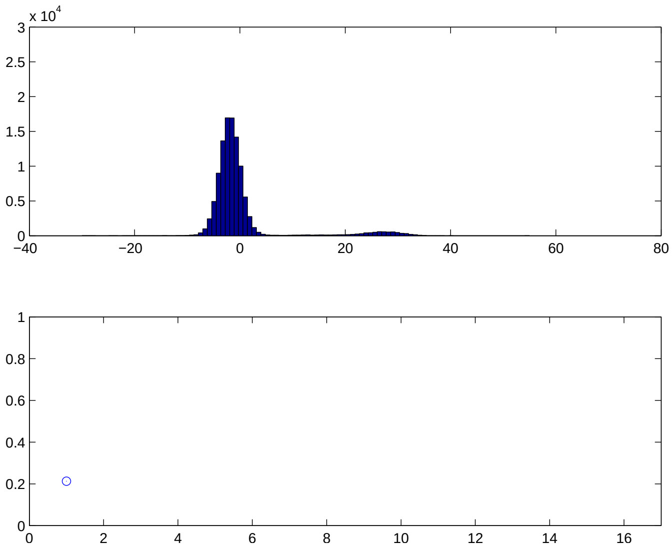

Fig. 5. Histogram for CV and squared canonical correlation, hand drawn water mask.

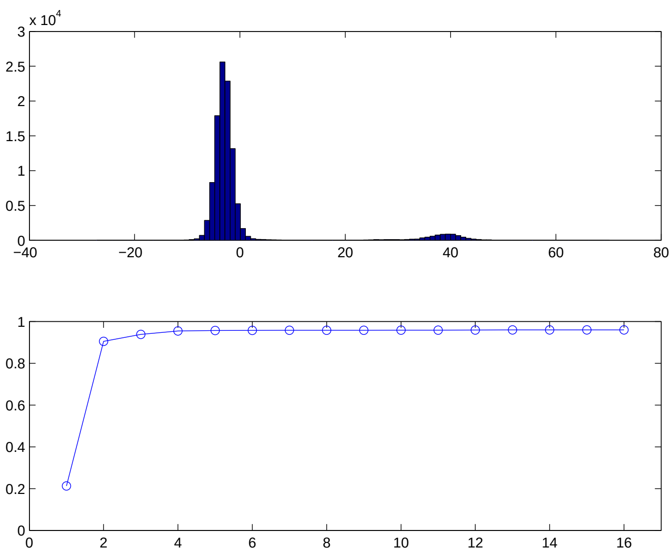

Fig. 6. Histogram for the 16. iteration CV and the 16 squared canonical correlations.

B. Wood/Trees Mask

Figure 7 shows the initial hand drawn wood/tree mask and the corresponding CV stretched linearly between minimum and maximum. Figure 8 shows the wood/tree mask after 20 iterations and the corresponding CV again stretched linearly between minimum and maximum. Note, that we seem to get all the wooded regions including individual trees.

Figure 9 shows the histogram of the CV based on the hand drawn mask (and the corresponding squared canonical correlation). Figure 10 shows the histogram of the 20. iterated CV and the 20 corresponding squared canonical correlations. From the two histograms we see that a much better separation between wood/trees and "everything else" is obtained after iterations. Again, we see that the first iteration is by far the most important one.

IV. CONCLUSIONS

An iterated version of canonical discriminant analysis is applied to perform optimal separation between two classes in 126-band HyMap data in two cases, 1) "water" vs "everything else" and 2) "wood/trees" vs "everything else". In the wood/trees case we get lower canonical correlations than in the water case showing that we obtain a better separation in the water case. We do however seem to pick up even individual trees in the more challenging wood/tree case. In these cases with two classes only, the canonical method gives the same solution as ordinary linear discriminant analysis.

A natural extension of the method is kernelization, see [6], [8], [9].

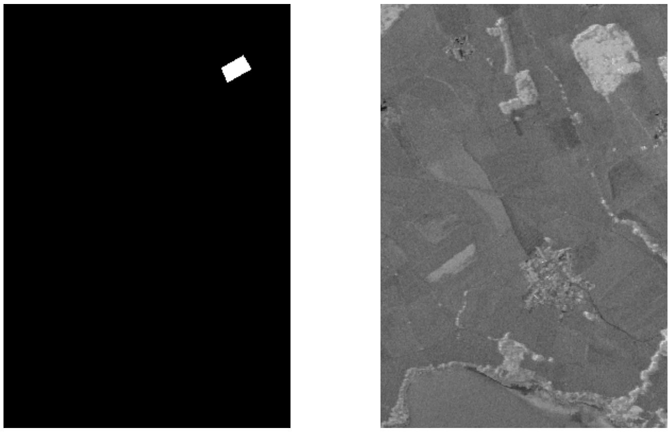

Fig. 7. Hand drawn wood mask and corresponding CV.

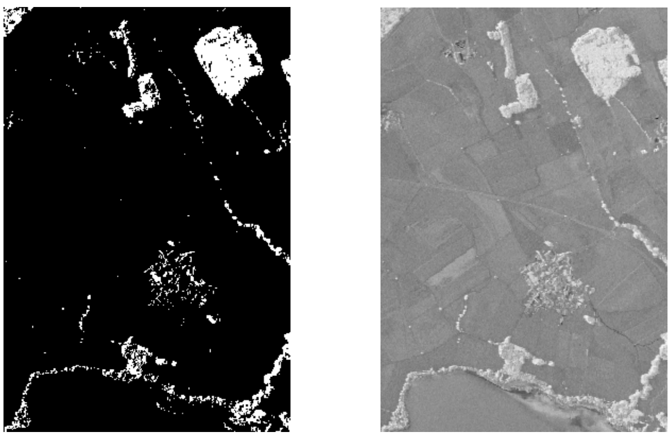

Fig. 8. Wood/tree mask, 20. iteration and corresponding iterated CV.

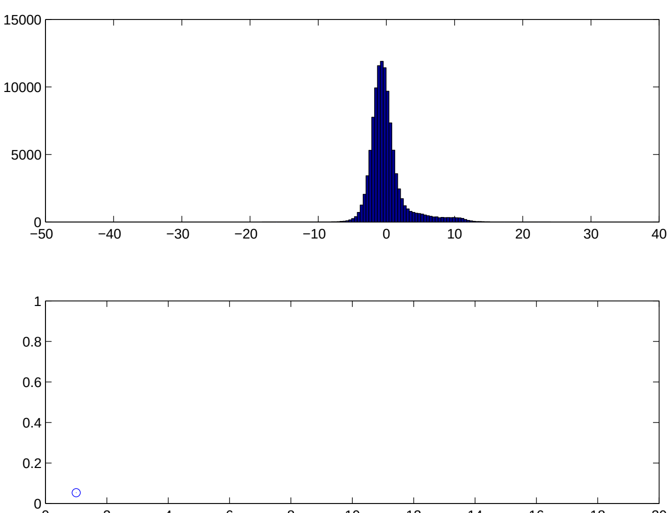

Fig. 9. Histogram for CV and squared canonical correlation, hand drawn wood mask.

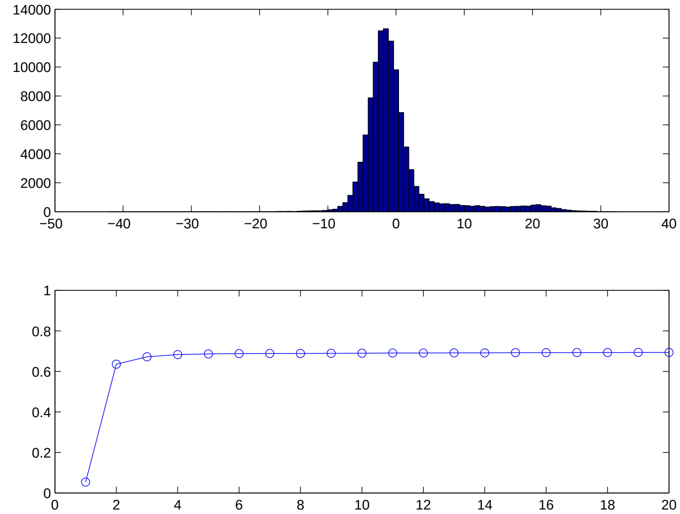

Fig. 10. Histogram for the 20. iteration CV and the 20 squared canonical correlations.

REFERENCES

[1] R. A. Fisher, "The utilization of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York, third edition, 2003.

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

[4] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, jan 1979.

[5] 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, feb 2007, Matlab code in zip file on http://www.imm.dtu.dk/pubdb/p.php?4695, ENVI/IDL and Python code on Dr. Morton J. Canty's home page.

[6] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 612–624, mar 2011, Matlab code in zip file on http://www.imm.dtu.dk/pubdb/p.php?5925, ENVI/IDL code on Dr. Morton J. Canty's home page.

[7] T. Schneider, S. Zimmermann, N. Pinnel, T. Heege, A. Müller, and W. Dorigo, "Lakes, eyes of the landscape. A long term concept for landscape monitoring on lake catchment area base," in *3rd EARSeL Workshop on Imaging Spectroscopy*, Herrsching, Germany, 13-16 May 2003, pp. 270–277.

[8] 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.

[9] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.