On the use of the Hotelling's T2 statistic for the hierarchical clustering of hyperspectral data

Miguel Angel Veganzones, Joana Frontera-Pons, Jocelyn Chanussot,
Jean-Philippe Ovarlez

▶ **To cite this version:**

HAL Id: hal-01010382
https://hal.science/hal-01010382

Submitted on 19 Jun 2014

ON THE USE OF THE HOTELLING'S T^2 STATISTIC FOR THE HIERARCHICAL CLUSTERING OF HYPERSPECTRAL DATA

M.A. Veganzones[1], J. Frontera-Pons[2], J. Chanussot[1], J.P. Ovarlez[3]

[1]GIPSA-lab, Grenoble-INP, Saint Martin d'Hères, France
[2] SONDRA Research Alliance, Supélec, France
[3]French Aerospace Lab, ONERA DEMR/TSI, France

ABSTRACT

In this work we propose a hierarchical clustering methodology for hyperspectral data based on the Hotelling's T^2 statistic. For each hypespectral sample data, the statistical sample mean is calculated using a window-based neighborhood. Then, the pairwise similarities between any two hyperspectral samples are computed based on the Hotelling's T^2 statistic. This statistic assumes a Gaussian distribution of the data while hyperspectral data have been observed to be long tailed distributed. In order to improve the statistic robustness we use the Fixed Point estimates, and compare them to the classical sample mean estimator. The similarities are then used to hierarchically cluster the hyperspectral data. We give some preliminary qualitative results of the proposed approach over the Indian Pines hyperspectral scene. Results show that the use of the Fixed Point estimator does not significantly affect the clustering results. Further work will be focused on the use of the robust Hotelling statistic.

Index Terms— hypespectral imaging, hierarchical clustering, Fixed Point estimates

1. INTRODUCTION

Hyperspectral data have been observed not to be multivariate normal but long tailed distributed [1]. In order to take into account these features, the class of elliptical contoured distributions (ECD) is considered to describe clutter statistical behavior. It provides a multivariate location-scatter family of distributions that primarily serves as heavy tailed alternative to the multivariate normal model. A m-dimensional random vector $\mathbf{y} = [y_1 y_2 ... y_m]^T$ with mean $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ has an elliptical distribution if its probability density function (PDF) has the following form:

$$f_y(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1} h_m((\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})), \quad (1)$$

where T denotes the transpose operator and $h_m(.)$ is any function such as (1) defines a PDF in $\mathbb{R}^m$.

The ECD class includes a large number of distributions, notably the Gaussian distribution, multivariate t distribution, K-distribution or multivariate Cauchy. Thus it allows for heterogeneity of the background power with the texture. Although non-Gaussian models are assumed for background modelling and test design, the parameters estimation is still performed using classical Gaussian based estimators; as for the covariance matrix, it is generally determined according to the Sample Covariance Matrix (SCM), $\hat{\mathbf{M}}_{SCM} = \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^T$. We use here the Fixed Point (FP) estimators [2] as a robust alternative for background statistical characterization.

In [3], authors propose a general approach for high-resolution polarimetric SAR (POLSAR) data classification in heterogeneous clutter, based on a statistical test of equality of covariance matrices. In hyperspectral domain the information provided by the mean is of high relevance. Thus, we propose here a similar approach for hyperspectral data using the Hotelling's T^2 statistic as a statistical test of equality of means [4]. We define a pairwise similarity between two pixels by using the output of the Hotelling's test for the means of the pixels calculated using a sliding window. As a preliminary test, we segment the hyperspectral scene by means of a hierarchical clustering algorithm [5] whose input is given by the previously calculated pairwise distances. We test the proposed approach in the Indian Pines hyperspectral scene.

2. THE FIXED POINT ESTIMATES

The Fixed Point estimates proposed by Tyler in [2] belong to the wider class of robust M-estimates [6]. According to the FP approach, the joint estimation of the scatter matrix, $\mathbf{M}$, and the mean vector, $\boldsymbol{\mu}$, leads to [2]:

$$\hat{\mathbf{M}}_{\text{FP}} = \frac{m}{N} \sum_{n=1}^N \frac{(\mathbf{y}_n - \hat{\boldsymbol{\mu}})(\mathbf{y}_n - \hat{\boldsymbol{\mu}})^H}{(\mathbf{y}_n - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{\text{FP}}^{-1} (\mathbf{y}_n - \hat{\boldsymbol{\mu}})} \quad (2)$$

and

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{n=1}^{N} \dfrac{\mathbf{y}_n}{\left((\mathbf{y}_n - \hat{\boldsymbol{\mu}})^T \hat{\mathbf{M}}_{\mathrm{FP}}^{-1}(\mathbf{y}_n - \hat{\boldsymbol{\mu}})\right)^{1/2}}}{\sum_{n=1}^{N} \dfrac{1}{\left((\mathbf{y}_n - \hat{\boldsymbol{\mu}})^T \hat{\mathbf{M}}_{\mathrm{FP}}^{-1}(\mathbf{y}_n - \hat{\boldsymbol{\mu}})\right)^{1/2}}} \qquad (3)$$

where N denotes the number of secondary data and $\mathbf{y}_n$ the vector under observation. For the matrix estimate, existence and uniqueness have been established in [7]. Although the proof for simultaneous scatter and location estimates is still an open question, they have been found to be useful and reliable for the estimation of ECD parameters because of its easy implementation. They are specified by implicit equations and can be easily computed using a recursive algorithm. We refer to [8] for a detailed performance analysis of the FP covariance matrix estimate. The main results of the statistical properties of the $\hat{\mathbf{M}}_{\mathrm{FP}}$ are summarized: $\hat{\mathbf{M}}_{\mathrm{FP}}$ is a consistent and unbiased estimate of $\mathbf{M}$; its asymptotic distribution is Gaussian and is the same as the asymptotic distribution of a Wishart matrix with $mN/(m+1)$ degrees of freedom. They provide a robust alternative to the classical estimators and its definition on hyperspectral detection framework was introduced in [9].

3. HOTELLING'S T^2 HYPOTHESIS TESTS

Let $\mathbf{X} \sim \mathrm{ECD}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, h_{m,1})$ and $\mathbf{Y} \sim \mathrm{ECD}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, h_{m,2})$ be two independent random vectors, elliptically distributed. We intend to decide if they belong to the same class comparing their mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The classification problem can be formulated as a binary hypothesis test. Symbolically, we aim to distinguish between:

$$\begin{cases} H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \\ H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \end{cases}$$

where the decision on the hypothesis test is made according to some statistical based criteria. We propose a similarity function based on the parameters estimation.

Hotelling's T^2 statistic. Assuming $\mathbf{x}_1, ..., \mathbf{x}_{N_1}$ and $\mathbf{y}_1, ..., \mathbf{y}_{N_2}$ are i.i.d. samples from $\mathbf{X} \sim \mathcal{N}_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathbf{Y} \sim \mathcal{N}_m(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ respectively. And the mean vector computed as, $\hat{\boldsymbol{\mu}}_x = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbf{x}_i$ and $\hat{\boldsymbol{\mu}}_y = \frac{1}{N_2}\sum_{i=1}^{N_2} \mathbf{y}_i$ and the covariance matrix estimated according to:

$$\mathbf{W} = \frac{\sum_{i=1}^{N_1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_x)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_x)^T + \sum_{i=1}^{N_2}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_y)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_y)^T}{N_1 + N_2 - 2}$$

which correspond to the unbiased pool estimate. Then the *Hotelling T^2* statistic test is defined as:

$$t^2 = \frac{N_1 N_2}{N_1 + N_2}(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \mathbf{W}^{-1}(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y). \qquad (4)$$

This expression is easily derived from the likelihood function of a multivariate normal model when H_0 is assumed, see e.g.

[4]. The distribution of the test under the null hypothesis can be related to the F-distribution according to

$$\frac{N_1 + N_2 - m - 1}{(N_1 + N_2 - 2)m}t^2 \sim F(m, N_1 + N_2 - 1 - m).$$

A different situation arises when samples do not fit on a multivariate gaussian distribution. We can improve the robustness of the test replacing the mean and matrix estimates for some robust alternative (e.g. the FP) if its asymptotic distribution is gaussian. The equality of the covariances matrices is a simplification that will certainly modify the test, as the resulting matrix estimate will not correspond to a Wishart distribution $(1/N_1\boldsymbol{\Sigma}_1 + 1/N_2\boldsymbol{\Sigma}_2)$. The results found on statistics literature for robust Hotelling and for both matrix and mean comparison will be further investigated.

4. STATISTICAL HIERARCHICAL CLUSTERING

A hierarchical clustering algorithm returns a hierarchy of clusters built by merging smaller components into bigger clusters (agglomerative clustering) or by splitting the whole image into smaller regions (divisive clustering). In order to do that, a pairwise similarity function is necessary, i.e. the angular distance, to compare each pair of components. Then, a linkage function indicates which of any possible pair of components is merged to (or split from) a bigger cluster.

A common linkage is the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method which consists in averaging all distances between the samples in each node, that is, the mean distance between elements of each cluster. Given two nodes A and B, the UPGMA linkage distance between the nodes, denoted $l(A, B)$ is given by:

$$l(A, B) = \frac{1}{|A| |B|} \sum_{\mathbf{x} \in A} \sum_{\mathbf{y} \in B} d(\mathbf{x}, \mathbf{y}) \qquad (5)$$

where $|\cdot|$ denotes the cardinality of a set and $d(\mathbf{x}, \mathbf{y})$ is any pairwise distance. In this work we propose to use the Hotelling's T^2 statistic (4) as the pairwise distance. The result is a hierarchy of clusters that can be represented as a dendrogram. A common criterion to stop merging (splitting) is to a priori set the number c of clusters one is looking for.

5. EXPERIMENTS AND RESULTS
5.1. Indian Pines scene

The Indian Pines scene[1] was gathered by airborne AVIRIS sensor over North-western Indiana and consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range 0.4-2.5 μm . We have reduced the number of bands to 200 by removing bands covering the region of water absorption: [104-108], [150-163], 220. Since the scene is taken in June some of the crops present, corn, soybeans, are in early stages of growth with less than 5% coverage. The ground
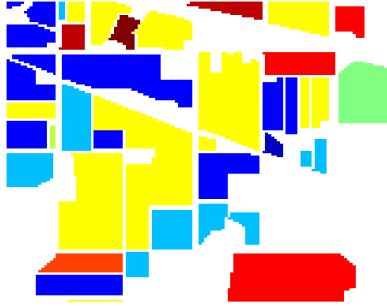
---

[1]https://engineering.purdue.edu/ biehl/MultiSpec/

Fig. 1. Reduced groundtruth for the Indian Pines scene

truth available is designated into sixteen classes with variable number of samples for each class. In order to simplify the problem in this preliminary stage, we have grouped some similar classes (see Fig.1). The grouped classes are Corn (corn, corn-notill and corn-mintill), Grass (grass-pasture, grass-pasture-mowed and grass-trees) and Soybean (soybean-notill, soybean-mintill and soybean-clean).

5.2. Experimental design and results

We compare the use of the classic estimator for the sample mean and SCM to the FP estimator for the clustering of the Indian Pines scene using the Hotelling's T^2 statistic (4) and the hierarchical clustering with UPGMA linkage function (5). For each pixel in the image we calculate its mean and covariance matrix by the classic estimator and the FP estimator using a 15×15 sliding window in order to ensure enough samples for the estimation stage ($N > 2m$). Those pixels on the borders of the image for which the window is incomplete are discarded resulting in a sample matrix of size 131×131.

We have performed two clustering experiments, a first one using the entire dataset (17161 samples) and a second one using only those data samples for which there is available groundtruth (9343 samples). For the latter, the statistics have been estimated using the entire dataset, but the linkage given by UPGMA algorithm have been built using only the samples from the groundtruth subset. Figure 2 shows the clustering results for the former and figure 3 does it for the latter. In order to visualize the results we have assigned to each cluster the RGB color resulting of averaging the groundtruth RGB color (see Figure 1) of the pixels of each cluster. Thus, a cluster with a color similar to any groundtruth color represents that the cluster is almost contained in a single groundtruth class. Clusters grouping several different groundtruth classes will present a mixed color.

The obtained results show that the use of the Hotelling's T^2 statistic yields to meaningful clusters, specially if we allow some overclustering (see figures 2 and 3 e-j). Moreover, it is possible to visually appreciate a better preformance of the FP estimator for a small number of clusters (see figures 2 and 3 a-d). The use of robust Hotelling's test will be further investigated to deal with the ECD assumption.

6. CONCLUSIONS

We have presented an approach to statistically characterize the hyperspectral data and then use the Hotelling's T^2 statistic to define a pairwise similarity function. The definition of a pairwise similarity allows to use a hierarchical clustering algorithm to segment the hyperspectral image. We have shown that even when the Hotelling's statistic Gaussian assumption differs from the actual ECD, the clustering results are meaningful. The use of a robust FP estimator shows a better performance than the use of the classic estimator. The use of a robust Hotelling's estimator capable of dealing with non-Gaussian assumption will be further investigated.

Acknowledgements

7. REFERENCES

[1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 29 –43, jan 2002.

[2] D.E. Tyler, "A distribution-free m-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[3] P. Formont, F. Pascal, G. Vasile, J.-P. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric sar images," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 567 –576, june 2011.

[4] Martin Bilodeau and David Brenner, *Theory of multivariate statistics*, Springer, 1999.

[5] Joe H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[6] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[7] F. Pascal, Y. Chitour, J-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 34 –48, jan. 2008.

[8] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans.-SP*, vol. 56, no. 6, pp. 2206–2217, June 2008.

[9] J. Frontera-Pons, M. Mahot, JP Ovarlez, F. Pascal, SK Pang, and J. Chanussot, "A class of robust estimates for detection in hyperspectral images using elliptical distributions background," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International.* IEEE, 2012, pp. 4166–4169.
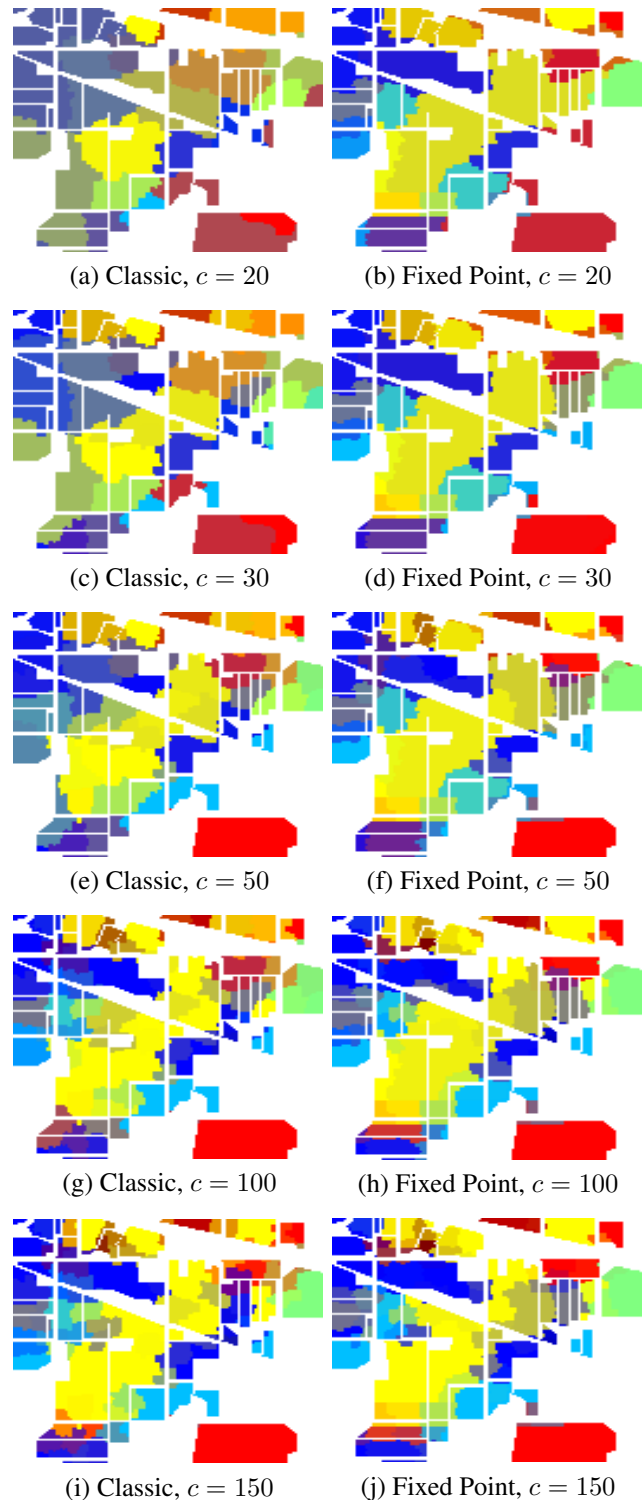
(a) Classic, $c = 20$ (b) Fixed Point, $c = 20$

(c) Classic, $c = 30$ (d) Fixed Point, $c = 30$

(e) Classic, $c = 50$ (f) Fixed Point, $c = 50$

(g) Classic, $c = 100$ (h) Fixed Point, $c = 100$

(i) Classic, $c = 150$ (j) Fixed Point, $c = 150$

Fig. 2. Results of the hierarchical clustering over the entire dataset using the Hotelling T^2 test with the classic estimator (left column) and the Fixed Point estimator (right column). Each row identifies with a different value of the number of clusters, c.

(a) Classic, $c = 20$ (b) Fixed Point, $c = 20$

(c) Classic, $c = 30$ (d) Fixed Point, $c = 30$

(e) Classic, $c = 50$ (f) Fixed Point, $c = 50$

(g) Classic, $c = 100$ (h) Fixed Point, $c = 100$

(i) Classic, $c = 150$ (j) Fixed Point, $c = 150$

Fig. 3. Results of the hierarchical clustering over the reduced dataset using the Hotelling T^2 test with the classic estimator (left column) and the Fixed Point estimator (right column). Each row ide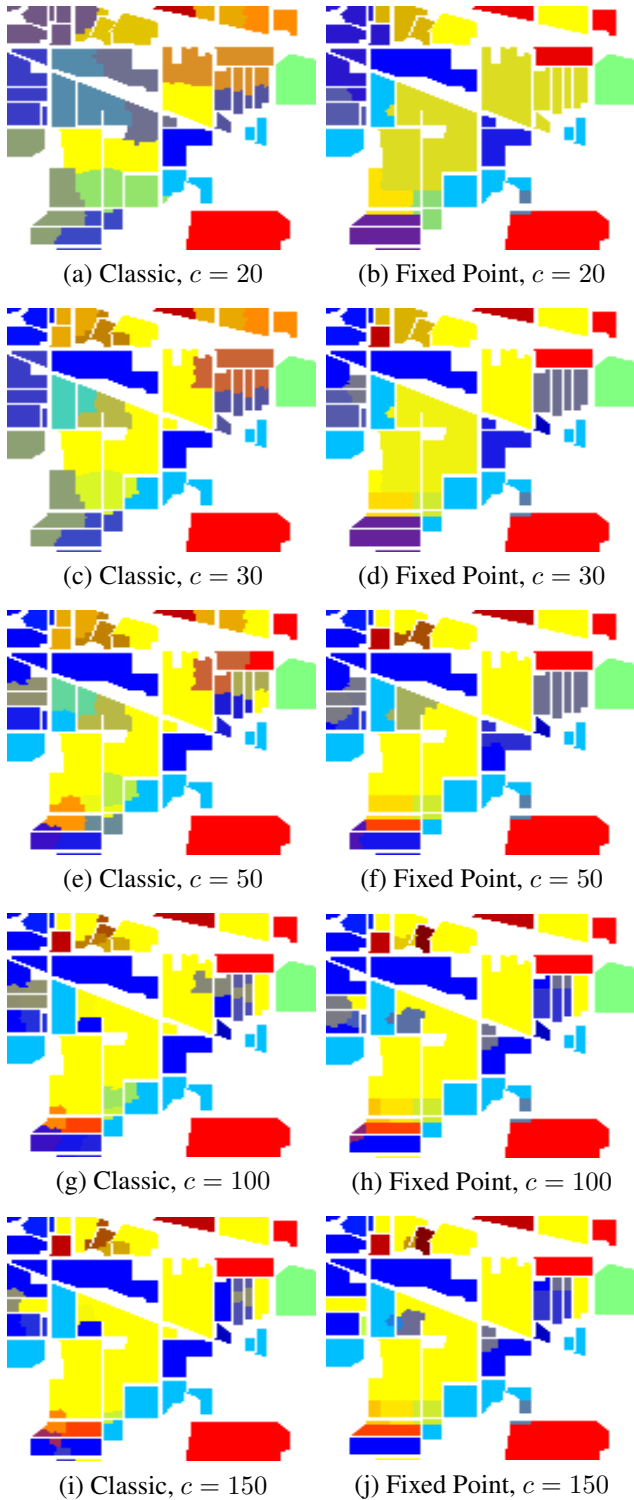ntifies with a different value of the number of clusters, c.