Detecting galaxies from the Multi Unit Spectroscopic Explorer (MUSE) by means of robust hyperspectral anomaly detectors

Rencontres d'Astrostatistique 2014

Miguel A. Veganzones

gipsa-lab, CNRS, UMR 5216, F-38420, Grenoble, France
miguel-angel.veganzones@gipsa-lab.fr

UNIVERSITÉ DE GRENOBLE

gipsa-lab

Grenoble | images | parole | signal | automatique | laboratoire

Motivation (I)

> Study of robust M-estimators for hyperspectral applications.

> One year posdoc in GIPSA-lab with Prof. Jocelyn Chanussot (DGA contract).

> Collaboration with :

- Prof. Jean-Philippe Ovarlez (ONERA).
- Prof. Frederic Pascal (SONDRA/Supelec).
- Ph.D. Student Joana Frontera-Pons (SONDRA/Supelec).

Motivation (I)

- Study of robust M-estimators for hyperspectral applications.
- One year posdoc in GIPSA-lab with Prof. Jocelyn Chanussot (DGA contract).
- Collaboration with :
  - Prof. Jean-Philippe Ovarlez (ONERA).
  - Prof. Frederic Pascal (SONDRA/Supelec).
  - Ph.D. Student Joana Frontera-Pons (SONDRA/Supelec).

Motivation (I)

- Study of robust M-estimators for hyperspectral applications.
- One year posdoc in GIPSA-lab with Prof. Jocelyn Chanussot (DGA contract).
- Collaboration with :
  - Prof. Jean-Philippe Ovarlez (ONERA).
  - Prof. Frederic Pascal (SONDRA/Supelec).
  - Ph.D. Student Joana Frontera-Pons (SONDRA/Supelec).

Motivation (II)

- Multi Unit Spectroscopic Explorer (MUSE) synthetic data $\rightarrow$ http ://muse.univ-lyon1.fr.
- Goal : detection of galaxies by means of robust anomaly detectors.
- Proposals :
  - **Statistical : robust M-estimators.**
  - Computational : binary partition trees.

Motivation (II)

- Multi Unit Spectroscopic Explorer (MUSE) synthetic data $\rightarrow$ http ://muse.univ-lyon1.fr.
- Goal : detection of galaxies by means of robust anomaly detectors.
- Proposals :
  - **Statistical : robust M-estimators.**
  - Computational : binary partition trees.

Motivation (II)

- Multi Unit Spectroscopic Explorer (MUSE) synthetic data $\rightarrow$ http ://muse.univ-lyon1.fr.
- Goal : detection of galaxies by means of robust anomaly detectors.
- Proposals :
  - **Statistical : robust M-estimators**.
  - Computational : binary partition trees.

Outline

Outline

Hyperspectral imagery

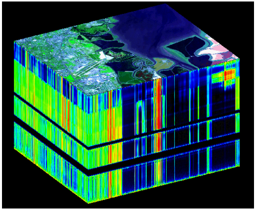


FIGURE : Hyperspectral cube.

- > Optical data.
- > Hundreds of contiguous high-resolution spectral bands.
- > Physical quantities : radiance, reflectance.
- > **High-dimensionality and high between-bands correlation.**

Hyperspectral imagery

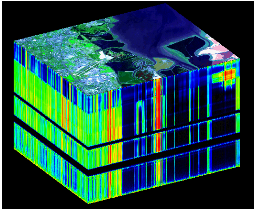


FIGURE : Hyperspectral cube.

- Optical data.
- Hundreds of contiguous high-resolution spectral bands.
- Physical quantities : radiance, reflectance.
- **High-dimensionality and high between-bands correlation.**

Hyperspectral imagery

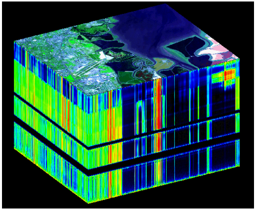


FIGURE : Hyperspectral cube.

- Optical data.
- Hundreds of contiguous high-resolution spectral bands.
- Physical quantities : radiance, reflectance.
- **High-dimensionality and high between-bands correlation.**

Hyperspectral imagery

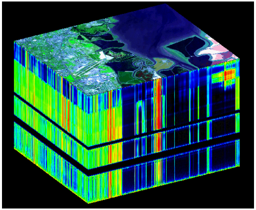
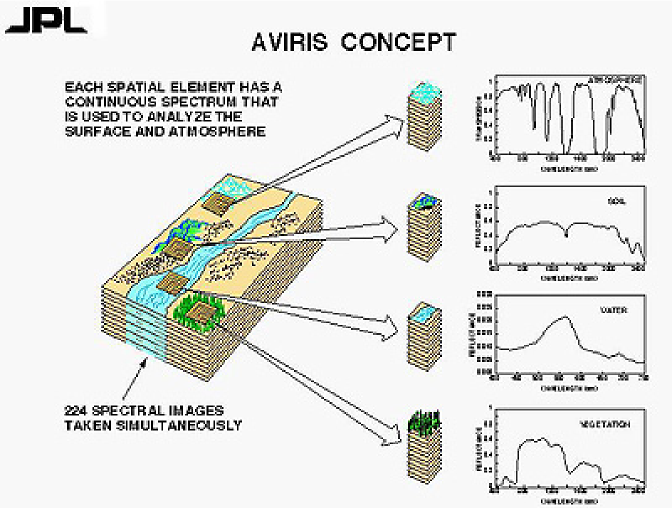


FIGURE : Hyperspectral cube.

- Optical data.
- Hundreds of contiguous high-resolution spectral bands.
- Physical quantities : radiance, reflectance.
- **High-dimensionality and high between-bands correlation.**


JPL
AVIRIS CONCEPT
EACH SPATIAL ELEMENT HAS A CONTINUOUS SPECTRUM THAT IS USED TO ANALYZE THE SURFACE AND ATMOSPHERE
224 SPECTRAL IMAGES TAKEN SIMULTANEOUSLY
SOIL
WATER
VEGETATION

Anomaly detectors (I)

- **Goal : locate objects in the image that are anomalous with respect to the background.**
- Statistical target detection is based on the Neyman-Pearson (NP) criterion → maximize the probability of detection for a given probability of false alarm.
- Very arbitrary definition → they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest.
- This fact makes extremely difficult to define a false alarm rate for the anomaly detectors.

Anomaly detectors (I)

- **Goal : locate objects in the image that are anomalous with respect to the background.**
- Statistical target detection is based on the Neyman-Pearson (NP) criterion $\rightarrow$ maximize the probability of detection for a given probability of false alarm.
- Very arbitrary definition $\rightarrow$ they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest.
- This fact makes extremely difficult to define a false alarm rate for the anomaly detectors.

Anomaly detectors (I)

- **Goal : locate objects in the image that are anomalous with respect to the background.**
- Statistical target detection is based on the Neyman-Pearson (NP) criterion $\rightarrow$ maximize the probability of detection for a given probability of false alarm.
- Very arbitrary definition $\rightarrow$ they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest.
- This fact makes extremely difficult to define a false alarm rate for the anomaly detectors.

Anomaly detectors (I)

- **Goal : locate objects in the image that are anomalous with respect to the background.**
- Statistical target detection is based on the Neyman-Pearson (NP) criterion $\rightarrow$ maximize the probability of detection for a given probability of false alarm.
- Very arbitrary definition $\rightarrow$ they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest.
- This fact makes extremely difficult to define a false alarm rate for the anomaly detectors.

Anomaly detectors (II)

- **Anomalies are defined with reference to a model of the background.**
- Most of AD methods rely on the classical Gaussian distribution assumption and need for the statistical characterization of the background.
- Adaptive AD $\rightarrow$ estimate the background statistics using reference (a.k.a. secondary) data :
  - Using all pixels in the image.
  - Using a local neighbourhood around the observation vector.

Anomaly detectors (II)

- **Anomalies are defined with reference to a model of the background.**
- Most of AD methods rely on the classical Gaussian distribution assumption and need for the statistical characterization of the background.
- Adaptive AD $\rightarrow$ estimate the background statistics using reference (a.k.a. secondary) data :
  - Using all pixels in the image.
  - Using a local neighbourhood around the observation vector.

Anomaly detectors (II)

> **Anomalies are defined with reference to a model of the background.**

> Most of AD methods rely on the classical Gaussian distribution assumption and need for the statistical characterization of the background.

> Adaptive AD $\rightarrow$ estimate the background statistics using reference (a.k.a. secondary) data :

- Using all pixels in the image.
- Using a local neighbourhood around the observation vector.

Adaptive RX AD (I)

- Considered the baseline AD for hyperspectral data.
- The RX AD was derived from the Generalized Likelihood Ratio Test (GLRT) assuming Gaussian hypothesis [1].

$$\begin{cases} \mathcal{H}_0 : \mathbf{y} = \mathbf{b} \\ \mathcal{H}_1 : \mathbf{y} = \mathbf{s} + \mathbf{b} \end{cases}, \tag{1}$$

where $\mathbf{b}$ represents the background and $\mathbf{s}$ denotes the presence of an anomalous signal.

[1] I. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 10, pp. 1760-1770, 1990.

Adaptive RX AD (I)

> **Considered the baseline AD for hyperspectral data.**

> The RX AD was derived from the Generalized Likelihood Ratio Test (GLRT) assuming Gaussian hypothesis [1].

$$\begin{cases} \mathcal{H}_0 : \mathbf{y} = \mathbf{b} \\ \mathcal{H}_1 : \mathbf{y} = \mathbf{s} + \mathbf{b} \end{cases} , \tag{1}$$

where $\mathbf{b}$ represents the background and $\mathbf{s}$ denotes the presence of an anomalous signal.

[1] I. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 10, pp. 1760-1770, 1990.

Adaptive RX AD (II)

> Statistical characterization of the background :

$$\mathbf{b} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{2}$$

> Sample estimation of the statistical parameters using secondary data, $\mathbf{y}_1, \ldots, \mathbf{y}_L$:

$$\hat{\boldsymbol{\mu}}_{\mathrm{SMV}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l, \tag{3}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{SCM}} = \frac{1}{L}\sum_{l=1}^{L}\left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\mathrm{SMV}}\right)\left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\mathrm{SMV}}\right)^T. \tag{4}$$

Adaptive RX AD (II)

- Statistical characterization of the background :

$$\mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\,. \tag{2}$$

- Sample estimation of the statistical parameters using secondary data, $\mathbf{y}_1, \ldots, \mathbf{y}_L$:

$$\hat{\boldsymbol{\mu}}_{\text{SMV}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}_l, \tag{3}$$

$$\hat{\boldsymbol{\Sigma}}_{\text{SCM}} = \frac{1}{L}\sum_{l=1}^{L} \left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}\right)\left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}\right)^T. \tag{4}$$

Adaptive RX AD (III)

> GLRT solution to the Adaptive RX AD :

$$\Lambda_{\text{ARX}} = \left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}\right)^T \hat{\boldsymbol{\Sigma}}_{\text{SCM}}^{-1} \left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (5)$$

> Assuming the null hypothesis is correct :

$$\frac{L-m+1}{mL}\Lambda_{\text{ARX}} \sim F_{m,L-m+1}. \quad (6)$$

> For high values of L, $(L > 10m)$, it can be approximated by a χ^2-distribution.

Adaptive RX AD (III)

> GLRT solution to the Adaptive RX AD :

$$\Lambda_{\text{ARX}} = (\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}})^T \, \hat{\boldsymbol{\Sigma}}_{\text{SCM}}^{-1} \, (\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \tag{5}$$

> Assuming the null hypothesis is correct :

$$\frac{L - m + 1}{mL} \Lambda_{\text{ARX}} \sim F_{m, L-m+1}. \tag{6}$$

> For high values of L, $(L > 10m)$, it can be approximated by a χ^2-distribution.

> GLRT solution to the Adaptive RX AD :

$$\Lambda_{\text{ARX}} = \left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}\right)^T \hat{\boldsymbol{\Sigma}}_{\text{SCM}}^{-1} \left(\mathbf{y}_l - \hat{\boldsymbol{\mu}}_{\text{SMV}}\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \tag{5}$$

> Assuming the null hypothesis is correct :

$$\frac{L-m+1}{mL}\Lambda_{\text{ARX}} \sim F_{m,L-m+1}. \tag{6}$$

> For high values of L, $(L > 10m)$, it can be approximated by a χ^2-distribution.

Background distribution

- The actual distribution of the background pixels differs from the theoretically predicted under Gaussian hypothesis.
- The empirical distribution usually has heavier tails compared to the Gaussian distribution [2].
- These tails strongly influence the observed false-alarm rate of the detector.
- **Proposal : characterize the background statistics by the class of Elliptical distributions.**

[2] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Geoscience and Remote Sensing Symposium*, 2002. IGARSS'02. 2002 IEEE International, vol. 3. IEEE, 2002, pp. 1664-1666.

- The actual distribution of the background pixels differs from the theoretically predicted under Gaussian hypothesis.
- The empirical distribution usually has heavier tails compared to the Gaussian distribution [2].
- These tails strongly influence the observed false-alarm rate of the detector.
- **Proposal : characterize the background statistics by the class of Elliptical distributions.**

[2] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Geoscience and Remote Sensing Symposium*, 2002. IGARSS'02. 2002 IEEE International, vol. 3. IEEE, 2002, pp. 1664-1666.

> The actual distribution of the background pixels differs from the theoretically predicted under Gaussian hypothesis.
> The empirical distribution usually has heavier tails compared to the Gaussian distribution [2].
> These tails strongly influence the observed false-alarm rate of the detector.
> **Proposal : characterize the background statistics by the class of Elliptical distributions.**

[2] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Geoscience and Remote Sensing Symposium*, 2002. IGARSS'02. 2002 IEEE International, vol. 3. IEEE, 2002, pp. 1664-1666.

Background distribution

- The actual distribution of the background pixels differs from the theoretically predicted under Gaussian hypothesis.
- The empirical distribution usually has heavier tails compared to the Gaussian distribution [2].
- These tails strongly influence the observed false-alarm rate of the detector.
- **Proposal : characterize the background statistics by the class of Elliptical distributions.**

[2] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Geoscience and Remote Sensing Symposium*, 2002. IGARSS'02. 2002 IEEE International, vol. 3. IEEE, 2002, pp. 1664-1666.

Statistical estimators

- the classical Gaussian-based estimators do not provide optimal performance due to heavy tails.
- **Proposal : Fixed Point (FP) robust estimators (also known as Tyler's estimators [3]).**
- FP estimates can be used as plug-in estimators in place of the unknowns mean vector and covariance matrix in the detection scheme.
- Simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption.

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

Statistical estimators

- the classical Gaussian-based estimators do not provide optimal performance due to heavy tails.
- **Proposal : Fixed Point (FP) robust estimators (also known as Tyler's estimators [3]).**
- FP estimates can be used as plug-in estimators in place of the unknowns mean vector and covariance matrix in the detection scheme.
- Simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption.

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

Statistical estimators

- the classical Gaussian-based estimators do not provide optimal performance due to heavy tails.
- **Proposal : Fixed Point (FP) robust estimators (also known as Tyler's estimators [3]).**
- FP estimates can be used as plug-in estimators in place of the unknowns mean vector and covariance matrix in the detection scheme.
- Simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption.

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

Statistical estimators

- the classical Gaussian-based estimators do not provide optimal performance due to heavy tails.
- **Proposal : Fixed Point (FP) robust estimators (also known as Tyler's estimators [3]).**
- FP estimates can be used as plug-in estimators in place of the unknowns mean vector and covariance matrix in the detection scheme.
- Simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption.

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

Secondary data (I)

> Conventional approach : using sliding windows.

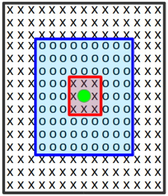


FIGURE : Selection of the secondary data by means of sliding windows.

> Outer window (blue) : delimits the pixels used as secondary data.
> Guard window (red) : prevents possible anomalous pixels to be selected as secondary data.

Secondary data (II)

- Local strategy provides more realistic scenario for the background characterization, but :
  - It can be sensitive to the presence of false alarms due to isolated anomalies.
  - Background should be uni-modal.
- Need of a size trade-off :
  - Increasing size : higher number of secondary data.
  - Decreasing size : less risk of including isolated anomalies and/or mixed distributions.
- **Proposal : define the secondary data using binary partition trees.**

Secondary data (II)

> Local strategy provides more realistic scenario for the background characterization, but :
  - It can be sensitive to the presence of false alarms due to isolated anomalies.
  - Background should be uni-modal.

> Need of a size trade-off :
  - Increasing size : higher number of secondary data.
  - Decreasing size : less risk of including isolated anomalies and/or mixed distributions.

> **Proposal : define the secondary data using binary partition trees.**

Secondary data (II)

- Local strategy provides more realistic scenario for the background characterization, but :
  - It can be sensitive to the presence of false alarms due to isolated anomalies.
  - Background should be uni-modal.
- Need of a size trade-off :
  - Increasing size : higher number of secondary data.
  - Decreasing size : less risk of including isolated anomalies and/or mixed distributions.
- **Proposal : define the secondary data using binary partition trees.**

Outline

Elliptical distributions (I)

- **Hyperspectral data have been proven not to be multivariate normal but long tailed distributed.**
- The class of elliptical distributions is considered to describe clutter statistical behavior.
- The family of elliptical distributions includes a large number of distributions, notably the Gaussian distribution, multivariate t-distribution, K-distribution or multivariate Cauchy.

Elliptical distributions (I)

- **Hyperspectral data have been proven not to be multivariate normal but long tailed distributed.**
- The class of elliptical distributions is considered to describe clutter statistical behavior.
- The family of elliptical distributions includes a large number of distributions, notably the Gaussian distribution, multivariate t-distribution, K-distribution or multivariate Cauchy.

Elliptical distributions (I)

- **Hyperspectral data have been proven not to be multivariate normal but long tailed distributed.**
- The class of elliptical distributions is considered to describe clutter statistical behavior.
- The family of elliptical distributions includes a large number of distributions, notably the Gaussian distribution, multivariate t-distribution, K-distribution or multivariate Cauchy.

Elliptical distributions (II)

> Formalization of the elliptical distribution :

$$f_{\mathbf{X}}(\mathbf{x}) = c_{m,h} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} h_m \left(\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \qquad (7)$$

> $c_{m,h}$ is a normalization constant.

> $h_m(\cdot)$ is any function (*density generator*) such that $f_{\mathbf{X}}(\mathbf{x})$ defines a p.d.f. $\rightarrow$ assumed to be only approximately known.

> Σ is a positive semi-definite matrix called scatter matrix.

Elliptical distributions (II)

> Formalization of the elliptical distribution :

$$f_{\mathbf{X}}(\mathbf{x}) = c_{m,h} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} h_m \left(\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \quad (7)$$

> $c_{m,h}$ is a normalization constant.

> $h_m(\cdot)$ is any function (*density generator*) such that $f_{\mathbf{X}}(\mathbf{x})$ defines a p.d.f. $\rightarrow$ assumed to be only approximately known.

> Σ is a positive semi-definite matrix called scatter matrix.

Elliptical distributions (II)

> Formalization of the elliptical distribution :

$$f_{\mathbf{X}}(\mathbf{x}) = c_{m,h} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} h_m \left(\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \qquad (7)$$

> $c_{m,h}$ is a normalization constant.

> $h_m(\cdot)$ is any function (*density generator*) such that $f_{\mathbf{X}}(\mathbf{x})$ defines a p.d.f. $\rightarrow$ assumed to be only approximately known.

> Σ is a positive semi-definite matrix called scatter matrix.

Elliptical distributions (II)

> Formalization of the elliptical distribution :

$$f_{\mathbf{X}}(\mathbf{x}) = c_{m,h} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} h_m \left(\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \quad (7)$$

> $c_{m,h}$ is a normalization constant.

> $h_m(\cdot)$ is any function (*density generator*) such that $f_{\mathbf{X}}(\mathbf{x})$ defines a p.d.f. $\rightarrow$ assumed to be only approximately known.

> $\boldsymbol{\Sigma}$ is a positive semi-definite matrix called scatter matrix.

Elliptical distributions (III)

- Remark that $f_{\mathbf{x}}(\mathbf{x})$ depends on $\mathbf{x}$ only through the quadratic form $(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})$.
- Thus, the level sets of the density are ellipsoids in the Euclidean m-space.
- **If the second-order moment exists, then Σ reflects the structure of the covariance matrix of the elliptically distributed random vector x, i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant.**
- Then, it serves to characterize the correlation structure existing within the spectral bands.

Elliptical distributions (III)

- Remark that $f_{\mathbf{X}}(\mathbf{x})$ depends on $\mathbf{x}$ only through the quadratic form $(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$.
- Thus, the level sets of the density are ellipsoids in the Euclidean m-space.
- **If the second-order moment exists, then Σ reflects the structure of the covariance matrix of the elliptically distributed random vector x, i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant.**
- Then, it serves to characterize the correlation structure existing within the spectral bands.

Elliptical distributions (III)

- Remark that $f_{\mathbf{X}}(\mathbf{x})$ depends on $\mathbf{x}$ only through the quadratic form $(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})$.
- Thus, the level sets of the density are ellipsoids in the Euclidean m-space.
- **If the second-order moment exists, then Σ reflects the structure of the covariance matrix of the elliptically distributed random vector $\mathbf{x}$, i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant.**
- Then, it serves to characterize the correlation structure existing within the spectral bands.

Elliptical distributions (III)

- Remark that $f_{\mathbf{X}}(\mathbf{x})$ depends on $\mathbf{x}$ only through the quadratic form $(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})$.
- Thus, the level sets of the density are ellipsoids in the Euclidean m-space.
- **If the second-order moment exists, then Σ reflects the structure of the covariance matrix of the elliptically distributed random vector $\mathbf{x}$, i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant**.
- Then, it serves to characterize the correlation structure existing within the spectral bands.

FP estimator (I)

- The FP estimates have been widely investigated in statistics and signal processing literature.
- These estimators belong to the wider class of robust M-estimators.
- **Σ_{FP} and Σ_{SCM} have the same asymptotic Gaussian distribution which differs on their second order moment by a factor of $\frac{m+1}{m}L$.**
- For L sufficiently large, Σ_{FP} behaves as a Wishart matrix with $\frac{m+1}{m}$ degrees of freedom.

FP estimator (I)

- The FP estimates have been widely investigated in statistics and signal processing literature.
- These estimators belong to the wider class of robust M-estimators.
- **Σ_{FP} and Σ_{SCM} have the same asymptotic Gaussian distribution which differs on their second order moment by a factor of $\frac{m+1}{m}L$.**
- For L sufficiently large, Σ_{FP} behaves as a Wishart matrix with $\frac{m+1}{m}$ degrees of freedom.

FP estimator (I)

- The FP estimates have been widely investigated in statistics and signal processing literature.
- These estimators belong to the wider class of robust M-estimators.
- **Σ_{FP} and Σ_{SCM} have the same asymptotic Gaussian distribution which differs on their second order moment by a factor of $\frac{m+1}{m}L$.**
- For L sufficiently large, Σ_{FP} behaves as a Wishart matrix with $\frac{m+1}{m}$ degrees of freedom.

FP estimator (I)

- The FP estimates have been widely investigated in statistics and signal processing literature.
- These estimators belong to the wider class of robust M-estimators.
- **$\Sigma_{\mathbf{FP}}$ and $\Sigma_{\mathbf{SCM}}$ have the same asymptotic Gaussian distribution which differs on their second order moment by a factor of $\frac{m+1}{m}L$.**
- For L sufficiently large, Σ_{FP} behaves as a Wishart matrix with $\frac{m+1}{m}$ degrees of freedom.

FP estimator (II)

> FP estimates :

$$\hat{\boldsymbol{\mu}}_{\mathrm{FP}} = \frac{\sum_{l=1}^{L} \frac{\mathbf{x}_i}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})^T \hat{\boldsymbol{\Sigma}}_{\mathrm{FP}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})\right)^{1/2}}}{\sum_{l=1}^{L} \frac{1}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})^T \hat{\boldsymbol{\Sigma}}_{\mathrm{FP}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})\right)^{1/2}}} \quad (8)$$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{FP}} = \frac{m}{L} \sum_{l=1}^{L} \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})^T}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})^T \hat{\boldsymbol{\Sigma}}_{\mathrm{FP}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{FP}})\right)} \quad (9)$$

> Alternate iterative process.

Outline

MUSE sensor

- The Multi Unit Spectroscopic Explorer (MUSE) project aims to provide astronomers with a new generation of optical instrument, capable of simultaneously imaging the sky (in 2D) and measuring the optical spectra of the light received at a given position on the sky.
- MUSE was installed on the VLT telescope and operational in 2013, and its performances are expected to allow observation of far galaxies up to 100 times fainter than those presently detectable.

MUSE sensor

- The Multi Unit Spectroscopic Explorer (MUSE) project aims to provide astronomers with a new generation of optical instrument, capable of simultaneously imaging the sky (in 2D) and measuring the optical spectra of the light received at a given position on the sky.
- MUSE was installed on the VLT telescope and operational in 2013, and its performances are expected to allow observation of far galaxies up to 100 times fainter than those presently detectable.

MUSE datasets

- MUSE will deliver a 3D data-cube made of a stack of images recorded at 3578 different wavelengths over the range $465 - 930$ nm.
- Each monochromatic image represents a field of view of 60×60 arcsec, recorded with a spatial sampling of 0.2 arcsec.
- Each record results in a data cube of size 1570 MB encoding 3578 images of 300×300 pixels, possibly containing thousands of objects (galaxies) existing over different subsets of wavelengths.

MUSE datasets

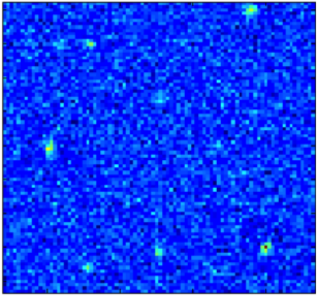
- MUSE will deliver a 3D data-cube made of a stack of images recorded at 3578 different wavelengths over the range $465 - 930$ nm.
- Each monochromatic image represents a field of view of 60×60 arcsec, recorded with a spatial sampling of 0.2 arcsec.
- Each record results in a data cube of size 1570 MB encoding 3578 images of 300×300 pixels, possibly containing thousands of objects (galaxies) existing over different subsets of wavelengths.

Available MUSE synthetic dataset

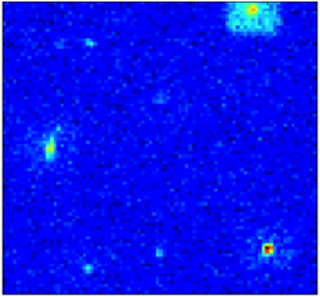
(a) MUSE data cube

Results

- From the 3578 available bands, we have chosen one band of each 100.
- The results for anomaly detection are presented for a fixed probability of false alarm, $P_{\mathrm{FA}} = 10^{-3}$.
- Note that detection with FP estimators provides results with lower false alarm rate than classical ones.



SMV-SCM



FP estimates

Outline

Summary

- Hyperspectral AD are usually based on Gaussian assumptions $\rightarrow$ not realistic (heavy tailed distributions).
- Conventional SMV and SCM estimators are not optimal with heavy tailed distributions.
- Proposal : use FP estimators $\rightarrow$ they work as plug-in estimators.
- Experimental results with MUSE synthetic data $\rightarrow$ galaxy detection.

Summary

- Hyperspectral AD are usually based on Gaussian assumptions → not realistic (heavy tailed distributions).
- Conventional SMV and SCM estimators are not optimal with heavy tailed distributions.
- Proposal : use FP estimators → they work as plug-in estimators.
- Experimental results with MUSE synthetic data → galaxy detection.

Summary

- Hyperspectral AD are usually based on Gaussian assumptions $\rightarrow$ not realistic (heavy tailed distributions).
- Conventional SMV and SCM estimators are not optimal with heavy tailed distributions.
- Proposal : use FP estimators $\rightarrow$ they work as plug-in estimators.
- Experimental results with MUSE synthetic data $\rightarrow$ galaxy detection.

Summary

- Hyperspectral AD are usually based on Gaussian assumptions $\rightarrow$ not realistic (heavy tailed distributions).
- Conventional SMV and SCM estimators are not optimal with heavy tailed distributions.
- Proposal : use FP estimators $\rightarrow$ they work as plug-in estimators.
- Experimental results with MUSE synthetic data $\rightarrow$ galaxy detection.

Detecting galaxies from the Multi Unit Spectroscopic Explorer (MUSE) by means of robust hyperspectral anomaly detectors

Rencontres d'Astrostatistique 2014

Miguel A. Veganzones

gipsa-lab, CNRS, UMR 5216, F-38420, Grenoble, France
miguel-angel.veganzones@gipsa-lab.fr