Estimation of Mars surface physical properties from hyperspectral images using the SIR method

Caroline Bernard-Michel, Sylvain Douté, Laurent Gardes and Stéphane Girard
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Introduction

Spectrometer

Hyperspectral cube

Spectral dimension

Spatial dimension

Spectrum at one pixel

Source: Nasa
Radiative transfer model

**RADIATIVE TRANSFER MODEL**: evaluates direct link between parameters and spectra. Allows the construction of a training data
INVERSE PROBLEM: evaluates the properties of atmospheric and surface materials from the spectra
Usual methods

- Nearest neighbor
- Weighted Nearest neighbor
Aim

- To establish functional relationships between:
  - Spectra $x \in \mathbb{R}^p$ (p=184) from Mars Express mission
  - Physical parameter $y \in \mathbb{R}$ : proportion of water, proportion of dust, grain size...
  - Construct $f$ in order to estimate parameters:

$$f : \mathbb{R}^p \rightarrow \mathbb{R}$$

$$x \rightarrow y$$
Difficulties

- Curse of dimensionality (184 wavelengths): dimension of $x$ has to be reduced
- Find projection axis $a \in \mathbb{R}^p$ (here, only the first axis will be retained)
- Instead of estimating $f$ such as $y = f(x)$, we will suppose there exists $g : \mathbb{R} \to \mathbb{R}$ exists such that:

\[ y = g(<a,x>, \varepsilon) \]
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Principal component analysis

- **Maximizes the variance** of the projections of the observations $x$
- **Does not take into account** $y$
Sliced inverse regression

• Proposed by Li (1991)
• Maximizes the between-slice variance of projections
• PCA of $E(X/Y)$ with $Z = \Sigma^{-1/2} X$
• Eigenvectors of $\Sigma^{-1} \Gamma$ with $\Gamma = \text{var}(E(X/Y))$
  $\Sigma = \text{var}(X)$
Application of SIR

**TRAINING DATA**

**OBSERVED DATA**
Problem

- Covariance matrix is ill-conditioned
  - Bad estimations of the directions
  - Sensitivity to noise
- Can be solved using regularization
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Regularization (1)

• Usual SIR:
  – eigenvectors of $\Sigma^{-1}\Gamma$

• Regularized SIR:
  – Zhong et al., 2005: eigenvectors of
    \[
    (\Sigma + \lambda I_d)^{-1} \Gamma
    \]
  – Tikhonov regularization: eigenvectors of
    \[
    (\Sigma^2 + \lambda I_d)^{-1} \Sigma \Gamma
    \]
Regularization (2)

- **Usual SIR**
- **Regularized SIR (Tikhonov)**

- Depends on the regularization parameter $\lambda$
- The condition number of the matrix decreases when $\lambda$ increases
- The estimation bias increases when $\lambda$ increases
Estimation

- Nearest neighbors (quite long!)
- Spline functions (choice of new parameters, boundaries)
- **Linear interpolation**
Choice of the regularization parameter

- By minimization of “Normalized RMSE” criterion

\[
\frac{\|\hat{y} - y\|}{\|y - \bar{y}\|} = \frac{\text{Residuals sum of square}}{\text{Total sum of square}}
\]
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Validation (1)

Application of Regularized Sliced Inverse Regression:
Determination of an axis $a(\lambda)$ depending on a regularization parameter $\lambda$.

Training data → Training data + noise → Estimation of parameters by linear interpolation → Minimization of Normalized RMSE → Optimal axis $a(\lambda)$ → Test data

ESTIMATION VALIDATION
Validation (2)

Regularized Sliced inverse regression (Tikhonov)

Nearest neighbors

Weighted nearest neighbors
Validation (3)

Normalized RMSE criterion:
\[
\frac{\|\hat{y} - y\|}{\|y - \bar{y}\|} = \frac{\text{Residuals sum of square}}{\text{Total sum of square}}
\]
Better when close to 0

SIR criterion:
\[
\frac{t^a \Gamma a}{t^a \sum a} = \frac{\text{between-slice variance}}{\text{total variance}}
\]
Better when close to 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variation interval</th>
<th>Tikhonov</th>
<th>Zhong</th>
<th>Nearest neighbor</th>
<th>Weighted nearest neighbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of dust</td>
<td>[0.0006 0.002]</td>
<td>0.33</td>
<td>0.92</td>
<td>0.33</td>
<td>0.92</td>
</tr>
<tr>
<td>Proportion of CO2</td>
<td>[0.9960 0.9988]</td>
<td>0.27</td>
<td>0.96</td>
<td>0.23</td>
<td>0.94</td>
</tr>
<tr>
<td>Proportion of water</td>
<td>[0.0006 0.002]</td>
<td>0.13</td>
<td>1.00</td>
<td>0.12</td>
<td>1.00</td>
</tr>
<tr>
<td>Grain size of water</td>
<td>[100 400]</td>
<td>0.37</td>
<td>0.92</td>
<td>0.38</td>
<td>0.87</td>
</tr>
<tr>
<td>Grain size of CO2</td>
<td>[400000 1050000]</td>
<td>0.19</td>
<td>0.99</td>
<td>0.18</td>
<td>0.98</td>
</tr>
</tbody>
</table>

- SIR gives better results than nearest neighbor classification
- Tikhonov and Zhong regularizations are equivalent
- With Tikhonov regularization, minimal normalized RMSE is reached on a larger interval than with Zhong’s.
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Application to south polar cap of Mars

- Model determined by physicists (water + CO2 + dust)
- 17753 spectra
- 184 wavelengths
- Training data simulated by radiative transfer model
- 5 parameters to study: proportions of water, dust and CO2, grain sizes of CO2 and water.
Proportion of CO2

Regularized Sliced Inverse Regression (Tikhonov)

Nearest neighbors

Weighted nearest neighbors
Proportion of water

Regularized Sliced Inverse Regression (Tikhonov)

Nearest neighbors

Weighted nearest neighbors
Outline

I. Context
   • Hyperspectral image data
   • Inverse problem

II. Dimension reduction
   • PCA
   • SIR

III. Regularization and estimation
   • Zhong et al., 2005
   • Tikhonov

IV. Validation on simulations

V. Application to the south polar cap of Mars

VI. Conclusion and future work
Conclusion and further work

- Good results on simulations.
- Realistic results on real data.
- Validation is difficult because of the lack of ground measurements.
- SIR for data streams.
We consider data arriving sequentially by blocks in a stream. Each data block $j = 1, \ldots, J$ is a $n$-sample from the regression model $y = g(\langle a, x \rangle, \varepsilon)$. 

**Goal**: Update the estimation of the direction $a$ at each arrival of a new block of observations.

*Joint work with team CQFD, INRIA Bordeaux Sud-Ouest*
Method

- Compute the **individual directions** $\hat{a}_j$ on each block $j = 1, \ldots, J$ using regularized SIR.
- Compute a **common direction** as

$$\hat{a} = \arg\max_{||a||=1} \sum_{j=1}^{J} \cos^2(\hat{a}_j, a) \cos^2(\hat{a}_j, \hat{a}_J).$$

**Idea** : If $\hat{a}_j$ is close to $\hat{a}_J$ then $\hat{a}$ should be close to $\hat{a}_j$.

**Explicit solution** : $\hat{a}$ is the eigenvector associated to the largest eigenvalue of

$$M_J = \sum_{j=1}^{J} \hat{a}_j\hat{a}_j^t \cos^2(\hat{a}_j, \hat{a}_J).$$
Advantages of SIRdatastream

- Computational complexity $O(Jnp^2)$ v.s. $O(J^2np^2)$ for the brute-force method which would consist in applying regularized SIR on the union of the $j$ first blocks for $j = 1, \ldots, J$.
- Data storage $O(np)$ v.s. $O(Jnp)$ for the brute-force method.

(under the assumption $n \gg \max(J, p)$).
- Interpretation of the weights $\cos^2(\hat{a}_j, \hat{a}_J)$. 
Scenario 1: A common direction in all the 60 blocks.

Left: $\cos^2(\hat{a}, a)$ for SIRdatastream, SIR brute-force and SIR estimators at each time $t$. Right: $\cos^2(\hat{a}_j, \hat{a}_J)$. The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.
Scenario 2: The 10th block is an outlier.

Left: $\cos^2(\hat{a}, a)$ for SIRdatastream, SIR brute-force and SIR estimators at each time $t$. Right: $\cos^2(\hat{a}_j, \hat{a}_J)$. The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.
Scenario 3: A drift occurs from the 10th block \((a \text{ to } a')\)

*Left*: \(\cos^2(\hat{a}, a)\) for SIRdatastream, SIR brute-force and SIR estimators at each time \(t\). *Right*: \(\cos^2(\hat{a}_j, \hat{a}_J)\). The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.
Illustration on simulations

Scenario 3 (cont’d) : A drift occurs from the 10th block \((a \text{ to } a')\)

Left : \(\cos^2(\hat{a}, a')\) for SIRdatastream and SIR brute-force. Right : \(\cos^2(\hat{a}, a')\)
Scenario 4: From the 10th block to the last one, there is no common direction.

*Left*: $\cos^2(\hat{a}, a)$ for SIRdatastream, SIR brute-force and SIR estimators at each time $t$. *Right*: $\cos^2(\hat{a}_j, \hat{a}_J)$. The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.
Related research themes in **MISTIS**

**Dimension reduction**
- Dimension reduction for regression (this talk),
- Unsupervised dimension reduction (nonlinear PCA),
- Dimension reduction for classification and clustering.

**Classification**
- Robust classification and outlier detection,
- Classification with missing data,
- Classification of spatial data,
- Classification of heterogeneous data.


References on this work

