Some improvements of the SIR method for the estimation of Mars physical properties from hyperspectral images

Stéphane Girard

Mistis team, INRIA Grenoble Rhône-Alpes.
http ://mistis.inrialpes.fr/~girard

Joint work with CQFD team, INRIA Bordeaux Sud-Ouest and Laboratoire de Planétologie de Grenoble.

Outline

Outline

Multivariate regression

Let $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$. The goal is to estimate $G : \mathbb{R}^p \to \mathbb{R}$ such that

$$Y = G(X) + \xi \quad \text{where } \xi \text{ is independent of } X.$$

- Unrealistic when p is large (*curse of dimensionality*).
- **Dimension reduction** : Replace X by its projection on a subspace of lower dimension without loss of information on the distribution of Y given X.
- **Central subspace** : smallest subspace S such that, conditionally on the projection of X on S, Y and X are independent.

- Assume (for the sake of simplicity) that $\dim(S) = 1$ *i.e.* $S =$span(b), with $b \in \mathbb{R}^p \Longrightarrow$ **Single index model :**

$$Y = g(b^t X) + \xi$$

where ξ is independent of X.
- The estimation of the p-variate function G is replaced by the estimation of the univariate function g and of the direction b.
- **Goal of SIR** [Li, 1991] : Estimate a basis of the central subspace. (*i.e.* b *in this particular case.*)

SIR

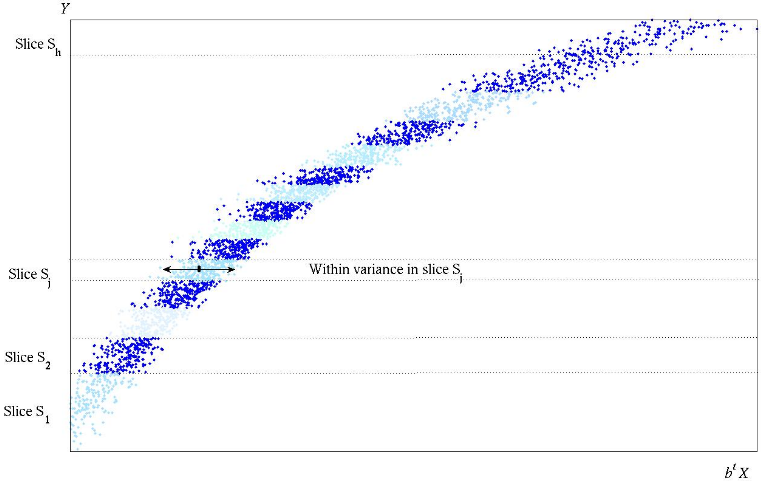
Idea :

- Find the direction b such that $b^t X$ best explains Y.
- Conversely, when Y is fixed, $b^t X$ should not vary.
- Find the direction b minimizing the variations of $b^t X$ given Y.

In practice :

- The support of Y is divided into h slices S_j.
- Minimization of the within-slice variance of $b^t X$ under the constraint $\text{var}(b^t X) = 1$.
- Equivalent to maximizing the between-slice variance under the same constraint.

Illustration



Estimation procedure

Given a sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, the direction b is estimated by

$$\hat{b} = \underset{b}{\operatorname{argmax}}\, b^t \hat{\Gamma} b \text{ such that } b^t \hat{\Sigma} b = 1. \qquad (1)$$

where $\hat{\Sigma}$ is the empirical covariance matrix and $\hat{\Gamma}$ is the between-slice covariance matrix defined by

$$\hat{\Gamma} = \sum_{j=1}^{h} \frac{n_j}{n} (\bar{X}_j - \bar{X})(\bar{X}_j - \bar{X})^t, \quad \bar{X}_j = \frac{1}{n_j} \sum_{Y_i \in S_j} X_i,$$

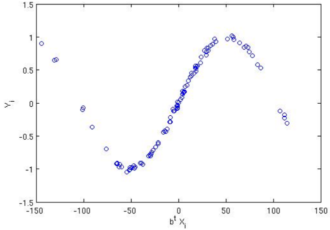
where n_j is the number of observations in the slice S_j.
The optimization problem (1) has a closed-form solution : $\hat{b}$ is the eigenvector of $\hat{\Sigma}^{-1}\hat{\Gamma}$ associated to the largest eigenvalue.

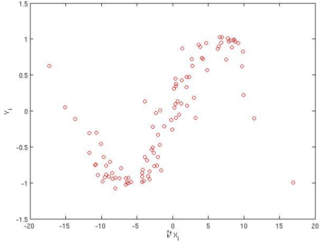
Illustration

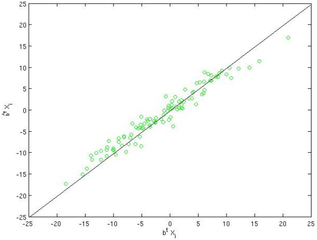
Simulated data.

- Sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ of size $n = 100$ with $X_i \in \mathbb{R}^p$ and $Y_i \in \mathbb{R}$, $i = 1, \ldots, n$.
- $X_i \sim \mathcal{N}_p(0, \Sigma)$ where $\Sigma = Q \Delta Q^t$ with
  - $\Delta =$diag$(p^\theta, \ldots, 2^\theta, 1^\theta)$,
  - θ controls the decreasing rate of the eigenvalue screeplot,
  - Q is an orientation matrix drawn from the uniform distribution on the set of orthogonal matrices.
- $Y_i = g(b^t X_i) + \xi$ where
  - g is the link function $g(t) = \sin(\pi t / 2)$,
  - b is the true direction $b = 5^{-1/2} Q(1, 1, 1, 1, 1, 0, \ldots, 0)^t$,
  - $\xi \sim \mathcal{N}_1(0, 9.10^{-4})$

Results with $\theta = 2$, dimension $p = 10$





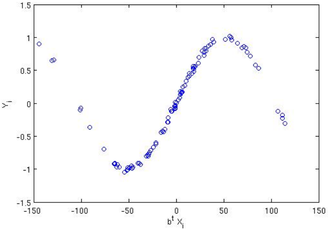


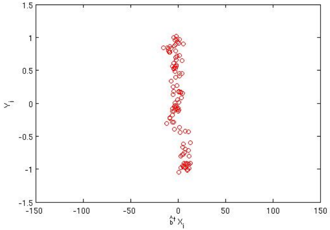
Blue : Y_i versus the projections $b^t X_i$ on the true direction b,

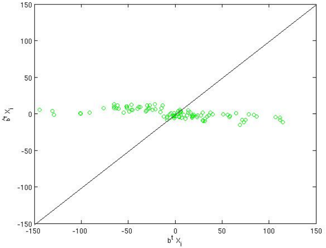
Red : Y_i versus the projections $\hat{b}^t X_i$ on the estimated direction $\hat{b}$,

Green : $\hat{b}^t X_i$ versus $b^t X_i$.

Results with $\theta = 2$, dimension $p = 50$







Blue : Y_i versus the projections $b^t X_i$ on the true direction b,

Red : Y_i versus the projections $\hat{b}^t X_i$ on the estimated direction $\hat{b}$,

Green : $\hat{b}^t X_i$ versus $b^t X_i$.

Explanation

Problem : $\hat{\Sigma}$ may be singular or at least ill-conditioned in several situations.

- Since rank$(\hat{\Sigma}) \leq \min(n-1, p)$, if $n \leq p$ then $\hat{\Sigma}$ is singular.
- Even if n and p are of the same order, $\hat{\Sigma}$ is ill-conditioned, and its inversion yields numerical problems in the estimation of the central subspace.
- The same phenomenon occurs if the coordinates of X are strongly correlated.

In the previous example, the condition number of Σ was p^{θ}.

Outline

Inverse regression model

Model introduced in [Cook, 2007].

$$X = \mu + c(Y)Vb + \varepsilon, \tag{2}$$

where

- μ and b are vectors of $\mathbb{R}^p$,
- $\varepsilon \sim \mathcal{N}_p(0, V)$, independent of Y,
- $c : \mathbb{R} \to \mathbb{R}$ the coordinate function.

Consequence : The expectation of $X - \mu$ given Y is collinear to the direction Vb.

Maximum likelihood estimation (1/3)

- $c(.)$ is expanded as a linear combination of h basis functions $s_j(.)$,

$$c(.) = \sum_{j=1}^{h} c_j s_j(.) = s^t(.)c,$$

where $c = (c_1, \dots, c_h)^t$ is unknown and $s(.) = (s_1(.), \dots, s_h(.))^t$.

- Model (2) can be rewritten as

$$X = \mu + s^t(Y)cVb + \varepsilon, \;\; \varepsilon \sim \mathcal{N}_p(0, V),$$

Maximum likelihood estimation (2/3)

Notations

- W : The $h \times h$ empirical covariance matrix of $s(Y)$ defined by

$$W = \frac{1}{n}\sum_{i=1}^{n}(s(Y_i) - \bar{s})(s(Y_i) - \bar{s})^t \;\; \text{with} \;\; \bar{s} = \frac{1}{n}\sum_{i=1}^{n} s(Y_i).$$

- M : the $h \times p$ matrix defined by

$$M = \frac{1}{n}\sum_{i=1}^{n}(s(Y_i) - \bar{s})(X_i - \bar{X})^t,$$

Maximum likelihood estimation (3/3)

- If W and $\hat{\Sigma}$ are regular, then the maximum likelihood estimator of b is $\hat{b}$ the eigenvector associated to the largest eigenvalue of $\hat{\Sigma}^{-1}M^tW^{-1}M$.
  $\Longrightarrow$ The inversion $\hat{\Sigma}$ is necessary.
- In the particular case of **piecewise constant basis functions**

$$s_j(.) = \mathbb{I}\{. \in S_j\}, \;\; j = 1, \dots, h,$$

  it can be shown that $M^tW^{-1}M = \hat{\Gamma}$ and thus $\hat{b}$ is the eigenvector associated to the largest eigenvalue of $\hat{\Sigma}^{-1}\hat{\Gamma}$.
  $\Longrightarrow$ SIR method.

Regularized SIR

- Introduction of a Gaussian prior $\mathcal{N}(0, \Omega)$ on the unknown vector b. Ω describes which directions in $\mathbb{R}^p$ are more likely to contain b.
- If W and $\Omega\hat{\Sigma} + I_p$ are regular, then $\hat{b}$ is the eigenvector associated to the largest eigenvalue of $(\Omega\hat{\Sigma} + I_p)^{-1}\Omega M^t W^{-1} M$.
- In the particular case where the basis functions are piecewise constant, $\hat{b}$ is the eigenvector associated to the largest eigenvalue of $(\Omega\hat{\Sigma} + I_p)^{-1}\Omega\hat{\Gamma}$.

$\Longrightarrow$ The inversion of $\hat{\Sigma}$ is replaced by the inversion of $\Omega\hat{\Sigma} + I_p$.
$\Longrightarrow$ For a well-chosen *a priori* matrix Ω, numerical problems disappear.

Links with existing methods

- Ridge [Zhong et al, 2005] : $\Omega = \tau^{-1} I_p$. No privileged direction for b in $\mathbb{R}^p$. $\tau > 0$ is a regularization parameter.
- PCA+SIR [Chiaromonte et al, 2002] :

$$\Omega = \sum_{j=1}^{d} \frac{1}{\hat{\delta}_j} \hat{q}_j \hat{q}_j^t,$$

where $d \in \{1, \dots, p\}$ is fixed, $\hat{\delta}_1 \geq \cdots \geq \hat{\delta}_d$ are the d largest eigenvalues of $\hat{\Sigma}$ and $\hat{q}_1, \dots, \hat{q}_d$ are the associated eigenvectors.

Three new methods

- PCA+ridge :

$$\Omega = \frac{1}{\tau} \sum_{j=1}^{d} \hat{q}_j \hat{q}_j^t.$$

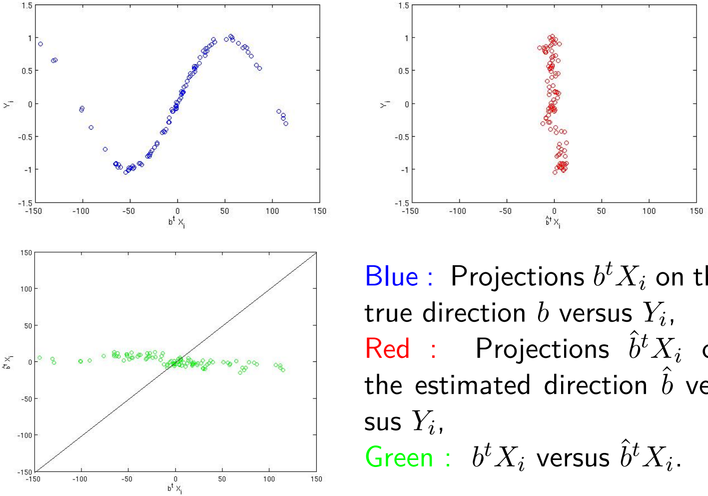
In the eigenspace of dimension d, all the directions are *a priori* equivalent.

- Tikhonov : $\Omega = \tau^{-1}\hat{\Sigma}$. The directions with large variance are the most likely to contain b.
- PCA+Tikhonov :

$$\Omega = \frac{1}{\tau} \sum_{j=1}^{d} \hat{\delta}_j \hat{q}_j \hat{q}_j^t.$$

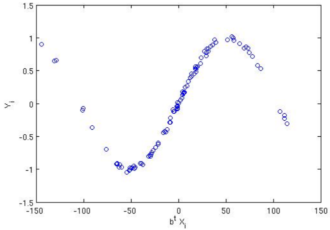
In the eigenspace of dimension d, the directions with large variance are the most likely to contain b.

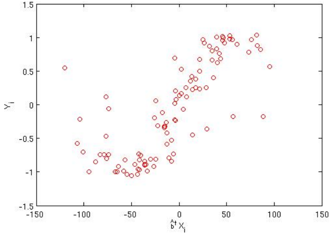
Recall of SIR results with $\theta = 2$ and $p = 50$

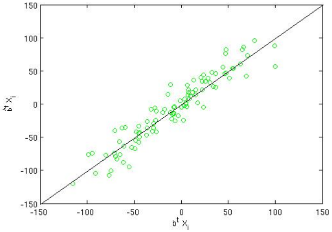


Blue : Projections $b^t X_i$ on the true direction b versus Y_i,
Red : Projections $\hat{b}^t X_i$ on the estimated direction $\hat{b}$ versus Y_i,
Green : $b^t X_i$ versus $\hat{b}^t X_i$.

Regularized SIR results (PCA+Ridge)







Blue : Projections $b^t X_i$ on the true direction b versus Y_i,
Red : Projections $\hat{b}^t X_i$ on the estimated direction $\hat{b}$ versus Y_i,
Green : $b^t X_i$ versus $\hat{b}^t X_i$.

Validation on simulations

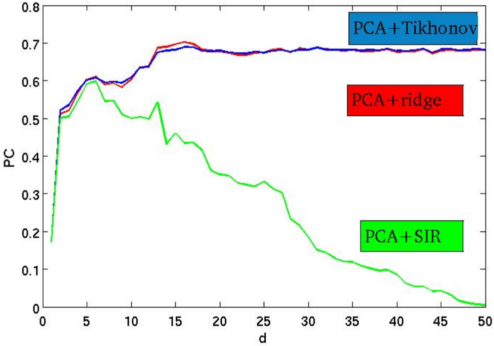
Proximity criterion between the true direction b and the estimated ones $\hat{b}^{(r)}$ on $N = 100$ replications :

$$\text{PC} = \frac{1}{N}\sum_{r=1}^{N}\cos^2(b, \hat{b}^{(r)})$$

- $0 \leq$ PC≤ 1,
- a value close to 0 implies a low proximity : The $\hat{b}^{(r)}$ are nearly orthogonal to b,
- a value close to 1 implies a high proximity : The $\hat{b}^{(r)}$ are approximately collinear with b.

Sensitivity with respect to the "cut-off" dimension

d versus PC. The condition number is fixed ($\theta = 2$) The optimal regularization parameter is used for each value of d.



- PCA+SIR : very sensitive to d.
- PCA+ridge and PCA+Tikhonov : stable as d increases.

Outline

Context

- We consider data arriving sequentially by blocks in a stream.
- Each data block $j = 1, \ldots, J$ is an i.i.d. sample (X_i, Y_i), $i = 1, \ldots, n$ from the regression model (2).
- **Goal** : Update the estimation of the direction b at each arrival of a new block of observations.

Method

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

$$\hat{b} = \underset{||b||=1}{\operatorname{argmax}} \sum_{j=1}^{J} \cos^2(\hat{b}_j, b) \cos^2(\hat{b}_j, \hat{b}_J).$$

Idea : If $\hat{b}_j$ is close to $\hat{b}_J$ then $\hat{b}$ should be close to $\hat{b}_j$.
Explicit solution : $\hat{b}$ is the eigenvector associated to the largest eigenvalue of

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

Advantages of SIRdatastream

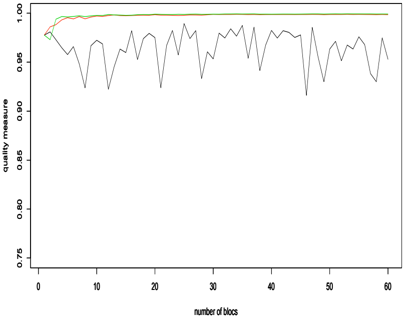
- Computational complexity $O(Jnp^2)$ v.s. $O(J^2np^2)$ for the brute-force method which would consist in applying regularized SIR on the union of the j first blocks for $j = 1, \ldots, J$.
- Data storage $O(np)$ v.s. $O(Jnp)$ for the brute-force method.

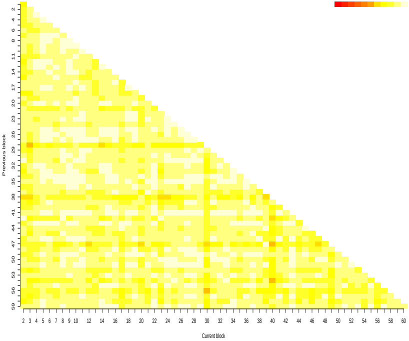
(under the assumption $n >> \max(J, p)$).

- Interpretation of the weights $\cos^2(\hat{b}_j, \hat{b}_J)$.

Illustration on simulations

Scenario 1 : A common direction in all the 60 blocks.

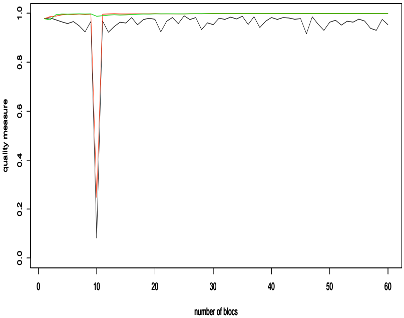


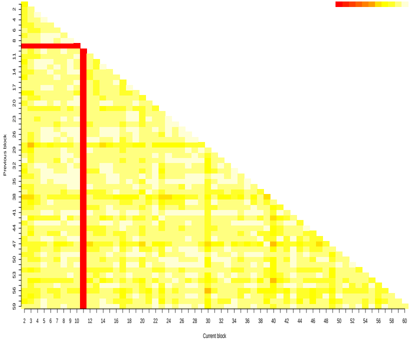


Left : $\cos^2(\hat{b}, b)$ for SIRdatastream, SIR brute-force and SIR estimators at each time t. *Right :* $\cos^2(\hat{b}_j, \hat{b}_J)$. The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.

Illustration on simulations

Scenario 2 : The 10th block is an outlier.

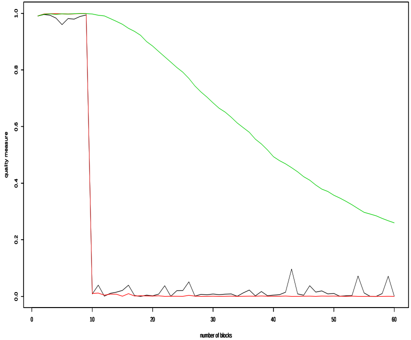


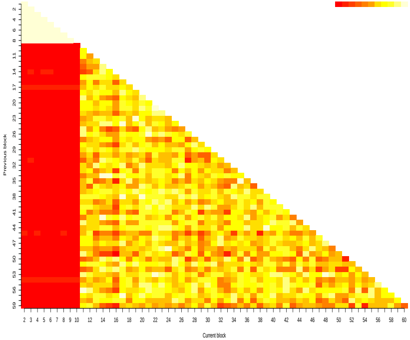


Left : $\cos^2(\hat{b}, b)$ for SIRdatastream, SIR brute-force and SIR estimators at each time t. *Right :* $\cos^2(\hat{b}_j, \hat{b}_J)$. The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.

Illustration on simulations

Scenario 3 : A drift occurs from the 10th block (b to b')

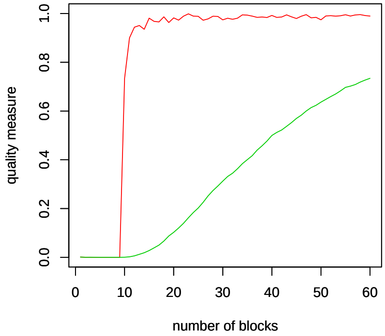


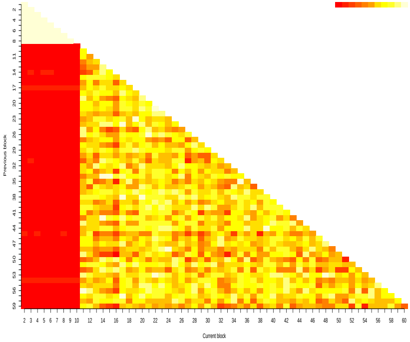


Left : $\cos^2(\hat{b}, b)$ for SIRdatastream, SIR brute-force and SIR estimators at each time t. *Right :* $\cos^2(\hat{b}_j, \hat{b}_J)$. The lighter (yellow) is the color, the larger is the weight. Red color stands for very small squared cosines.

Illustration on simulations

Scenario 3 (cont'd) :A drift occurs from the 10th block (b to b')

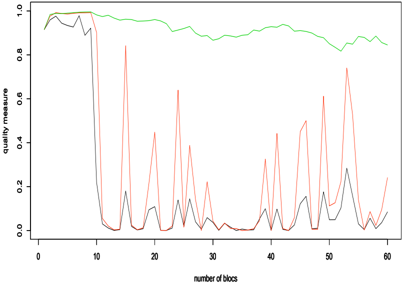


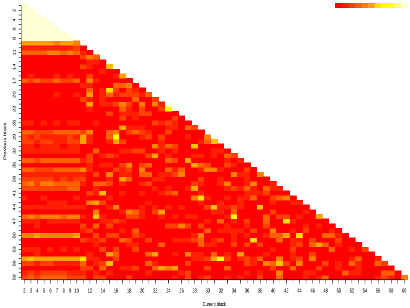


Left : $\cos^2(\hat{b}, b')$ for SIRdatastream and SIR brute-force. *Right :* $\cos^2(\hat{b}, b')$

Illustration on simulations

Scenario 4 : From the 10th block to the last one, there is no common direction.





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

Outline

1. Sliced Inverse Regression (SIR)
2. Regularization of SIR
3. SIR for data streams
4. Application to real data

Estimation of Mars surface physical properties from hyperspectral images

Context :

- Observation of the south pole of Mars at the end of summer, collected during orbit 61 by the French imaging spectrometer OMEGA on board Mars Express Mission.
- 3D image : On each pixel, a spectra containing $p = 184$ wavelengths is recorded.
- This portion of Mars mainly contains water ice, CO_2 and dust.

Goal : For each spectra $X \in \mathbb{R}^p$, estimate the corresponding physical parameter $Y \in \mathbb{R}$ (grain size of CO_2).

An inverse problem

Forward problem.

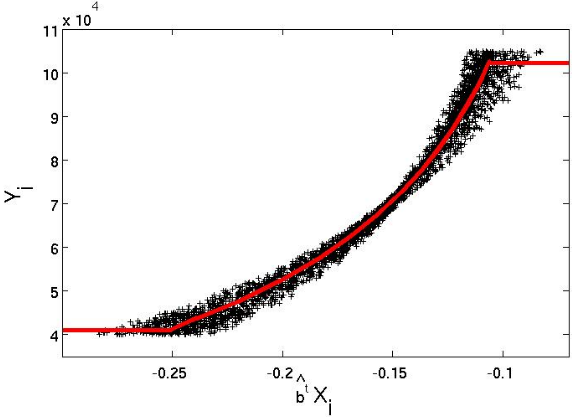
- Physical modeling of individual spectra with a surface reflectance model.
- Starting from a physical parameter Y, simulate $X = F(Y)$.
- Generation of $n = 12,000$ synthetic spectra with the corresponding parameters.

$\Longrightarrow$ Learning database.

Inverse problem.

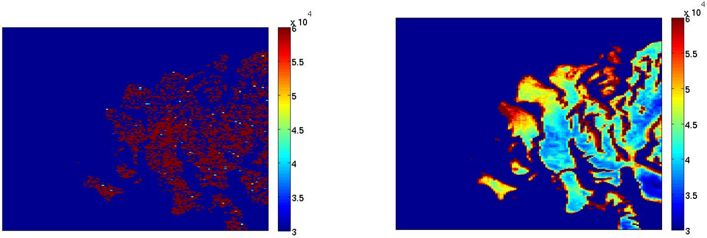
- Estimate the functional relationship $Y = G(X)$.
- Dimension reduction assumption $G(X) = g(b^t X)$.
- b is estimated by (regularized) SIR, g is estimated by a nonparametric one-dimensional regression.

Estimated function g



Estimated function g between the projected spectra $\hat{b}^t X$ on the first axis of regularized SIR (PCA+ridge) and Y, the grain size of CO_2.

Estimated CO_2 maps



Grain size of CO_2 estimated with SIR (left) and regularized SIR (right) on a hyperspectral image of Mars.

Dimension reduction

In this talk : dimension reduction for regression.
In the team Mistis :

- Unsupervised dimension reduction (nonlinear PCA),
- Dimension reduction for classification and clustering.

References on this work

- Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L. et Girard, S. (2009). Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression. *Journal of Geophysical Research - Planets*, **114**, E06005
- Bernard-Michel, C., Gardes, L. et Girard, S. (2009). Gaussian Regularized Sliced Inverse Regression, *Statistics and Computing*, **19**, 85–98.
- Bernard-Michel, C., Gardes, L. et Girard, S. (2008). A Note on Sliced Inverse Regression with Regularizations, *Biometrics*, **64**, 982–986.

References on SIR

- [Li, 1991] Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**, 316–327.
- [Cook, 2007]. Cook, R.D. (2007). Fisher lecture : Dimension reduction in regression. *Statistical Science*, **22**(1), 1–26.
- [Zhong et al, 2005] : Zhong, W., Zeng, P., Ma, P., Liu, J.S. and Zhu, Y. (2005). RSIR : Regularized Sliced Inverse Regression for motif discovery. *Bioinformatics*, **21**(22), 4169–4175.
- [Chiaromonte et al, 2002] : Chiaromonte, F. and Martinelli, J. (2002). Dimension reduction strategies for analyzing global gene expression data with a response. *Mathematical Biosciences*, **176**, 123–144.