

GRSIR 2.0 : Inversion multi-axes et régression non-linéaire

M. Fauvel - S. Girard - S. Douté - L. Gardes

INRIA - MISTIS team & LPG

15 juin 2010

Introduction 1/2

- Inversion d'images hyperspectrales : reconstruire y à partir de $\mathbf{x} \in \mathbb{R}^d$
- Deux « problèmes » :
 - ★ La grande dimension, d très grand !
 - ★ Le passage aux données réelles, $\mathbf{x}_{\text{simu}} \approx \mathbf{x}_{\text{obs}}$
- Modèle d'inversion :

$$y = g(\mathbf{B}^t \mathbf{x}) + \varepsilon$$

- ★ $\mathbf{B} = [\beta_1, \dots, \beta_p]$ génère \mathcal{S} , avec $\dim(\mathcal{S}) = p \ll d$
- ★ $g : \mathcal{S} \rightarrow \mathbb{R}$, fonction de régression multidimensionnelle
- \mathbf{B} estimé avec **GRSIR**
- g estimée avec les **moindres carrés non linéaires**

Introduction 2/2

- Pourquoi GRSIR :
 - ★ Utilisation des variables explicatives (\mathbf{x}) et à expliquer (y)
 - ★ Solution explicite
 - ★ Effective en grande dimension (**régularisation**)

- Pourquoi MCNL :
 - ★ Pas de modèle a priori sur la projection GRSIR
 - ★ Flexibilité lié au noyau
 - ★ Peu sensible à la dimension
 - ★ Solution explicite

Plan de l'exposé

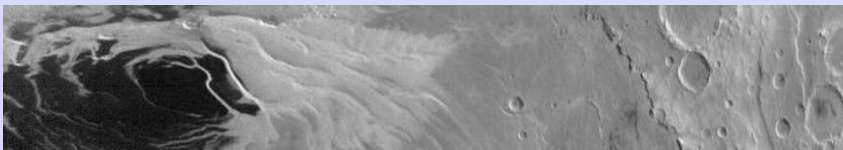
- 1 Image hyperspectrale
- 2 Gaussian Regularized Sliced Inverse Regression
- 3 Moindres carrés non linéaires
- 4 Données réelles
- 5 Conclusion

- 1 Image hyperspectrale
- 2 Gaussian Regularized Sliced Inverse Regression
- 3 Moindres carrés non linéaires
- 4 Données réelles
- 5 Conclusion

Espace image :



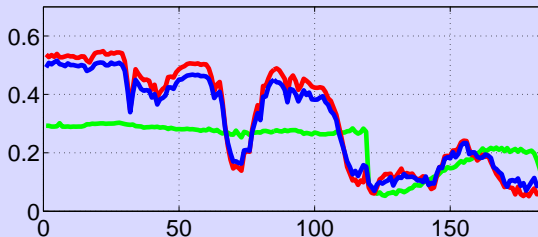
Bande 100



Bande 170

Les données 1/2

Espace spectral :



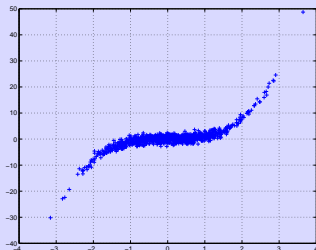
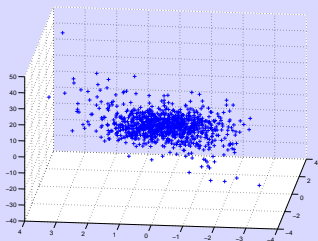
Paramètres physique :

- ① Proportion d'eau
- ② Proportion de CO_2
- ③ Proportion de poussière
- ④ Taille des graines d'eau
- ⑤ Taille des graines de CO_2

- 1 Image hyperspectrale
- 2 Gaussian Regularized Sliced Inverse Regression**
- 3 Moindres carrés non linéaires
- 4 Données réelles
- 5 Conclusion

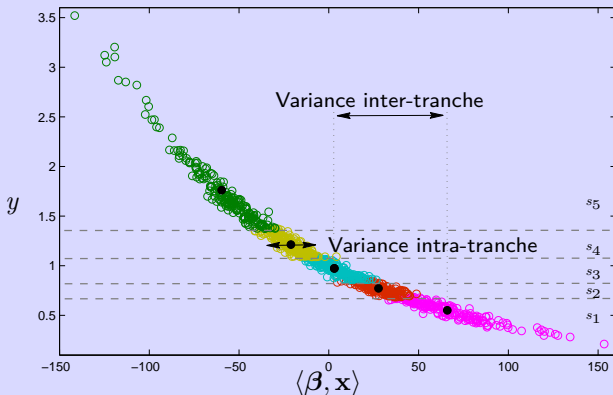
Principe 1/2

- Trouver le sous-espace contenant toute l'information de \mathbf{x} sur y
- $\beta = [0.44; 0.89]$, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ et $g(\cdot) = (\cdot)^3$:



Principe 2/2

- ① Découpe le support des y en H tranches s_j
- ② Maximise la variance inter-tranche sous $\text{var}(\beta^t \mathbf{x}) = 1$



Mise en œuvre

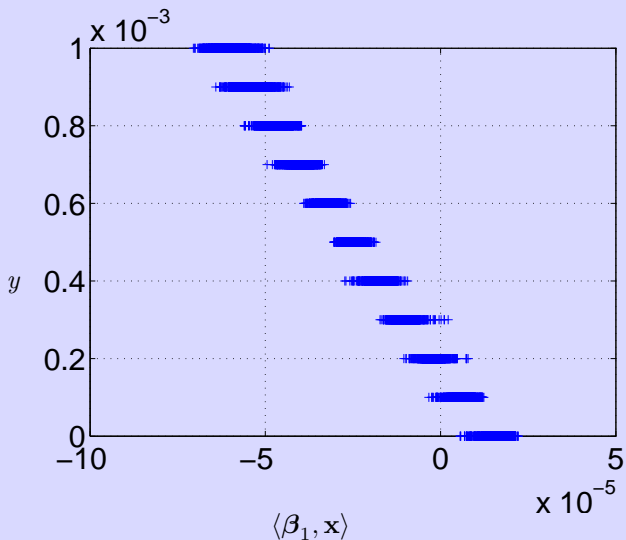
- Ensemble d'entraînement $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$
- $\hat{\beta}_l = \arg \max_{\beta} [\beta^t \hat{\Gamma} \beta]$ avec $\beta^t \hat{\Sigma} \beta_i = \delta_{li}$, $i = 1, \dots, l$
- $\hat{\Gamma}$ estimée de la matrice de covariance intra-tranche

$$\hat{\Gamma} = \sum_{j=1}^H \frac{n_j}{n} (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^t, \quad \bar{\mathbf{x}}_j = \frac{1}{n_j} \sum_{\mathbf{x}_i | y_i \in s_j} \mathbf{x}_i$$

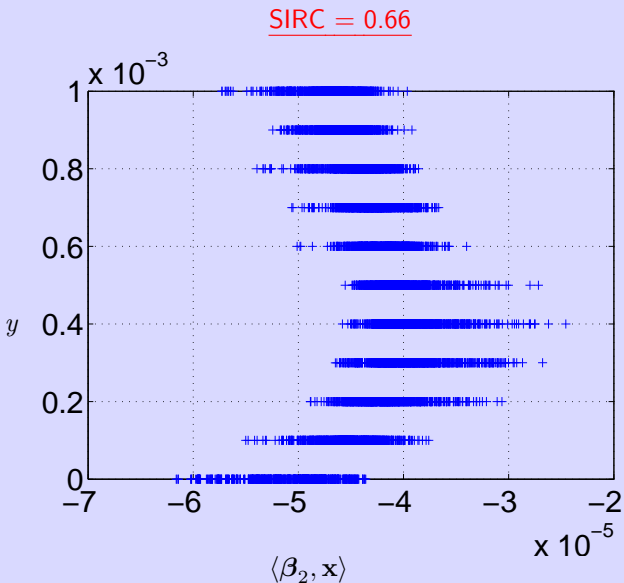
- Solution explicite : $\hat{\beta}_l$ est le vecteur propre correspondant à la $l^{\text{ième}}$ plus grande valeur propre de $\hat{\Sigma}^{-1} \hat{\Gamma}$
- Régularisation de $\hat{\Sigma}^{-1} \rightarrow (\hat{\Sigma} + \tau \mathbf{I})^{-1}$
- Qualité de la projection : $\text{SIRC}(\hat{\beta}_l) = \frac{\hat{\beta}_l^t \hat{\Gamma} \hat{\beta}_l}{\hat{\beta}_l^t \hat{\Sigma} \hat{\beta}_l}$

Données réelles

SIRC = 0.98

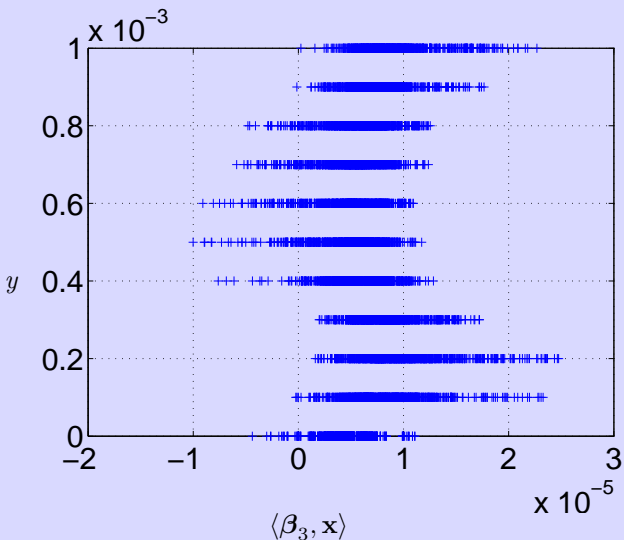


Données réelles

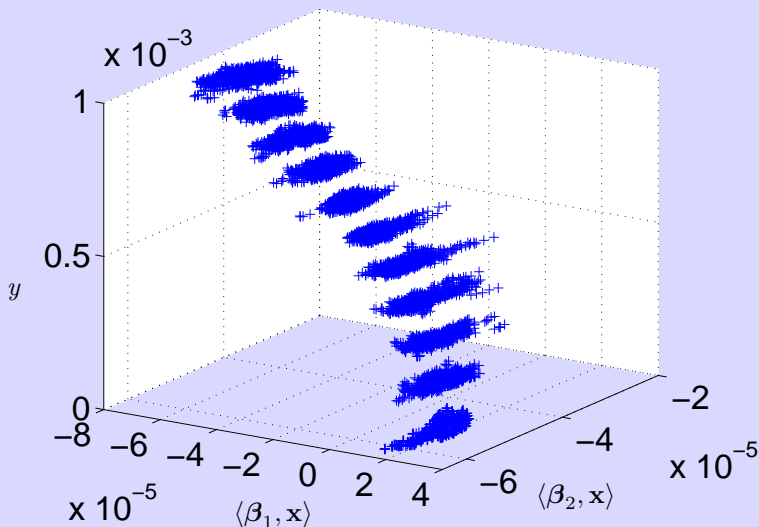


Données réelles

SIRC = 0.27



Données réelles



- 1 Image hyperspectrale
- 2 Gaussian Regularized Sliced Inverse Regression
- 3 Moindres carrés non linéaires**
- 4 Données réelles
- 5 Conclusion

Principe

- $\min_g \left[\frac{1}{n} \sum_{i=1}^n (y_i - g(\mathbf{x}_i))^2 + \lambda \|g\|^2 \right]$
- $g(\mathbf{x}) = \sum_{j=1}^n \alpha_j k(\mathbf{x}, \mathbf{x}_j) + b$
- $\|g\|^2 = \sum_{i,j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$
- $k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(- \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right)$
- Solution explicite :

$$\alpha_0 = [\mathbf{K}_0 + \lambda \mathbf{I}]^{-1} \mathbf{Y}_0$$

Principe

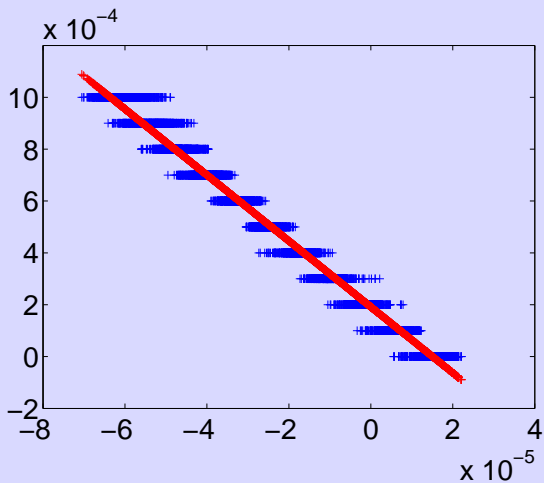
- $\min_g \left[\frac{1}{n} \sum_{i=1}^n (y_i - g(\mathbf{x}_i))^2 + \lambda \|g\|^2 \right]$
- $g(\mathbf{x}) = \sum_{j=1}^n \alpha_j k(\mathbf{x}, \mathbf{x}_j) + b$
- $\|g\|^2 = \sum_{i,j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$
- $k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(- \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right)$
- Solution explicite :

$$\frac{1}{n} \left(\left[\begin{array}{c|c} \mathbf{y} & \mathbf{1} \\ \hline 0 & 0 \end{array} \right] - \left[\begin{array}{c|c} \mathbf{K} & \mathbf{1} \\ \hline \mathbf{1}' & 0 \end{array} \right] \left[\begin{array}{c} \alpha \\ b \end{array} \right] \right)^2$$

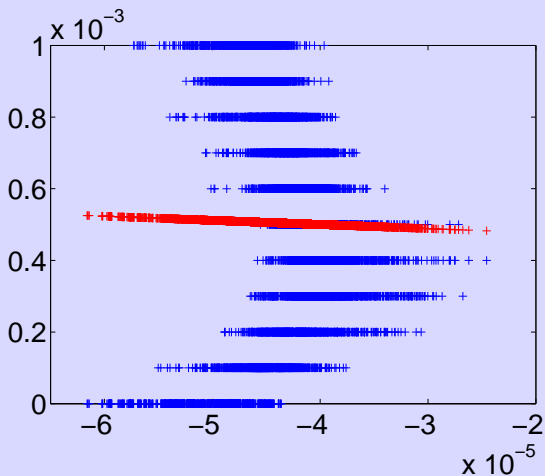
$$+ \lambda \left[\begin{array}{c|c} \alpha' & b \end{array} \right] \left[\begin{array}{c|c} \mathbf{K} & \mathbf{1} \\ \hline \mathbf{1}' & 0 \end{array} \right] \left[\begin{array}{c} \alpha \\ b \end{array} \right]$$

$$\alpha_0 = [\mathbf{K}_0 + \lambda \mathbf{I}]^{-1} \mathbf{Y}_0$$

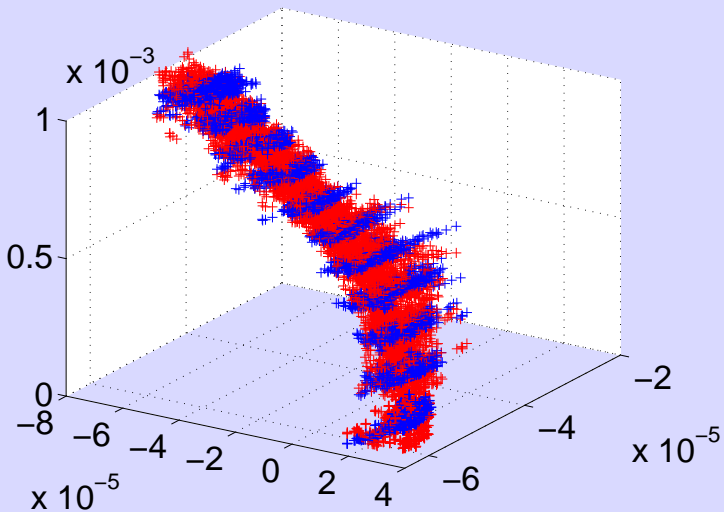
Exemples



Exemples



Exemples



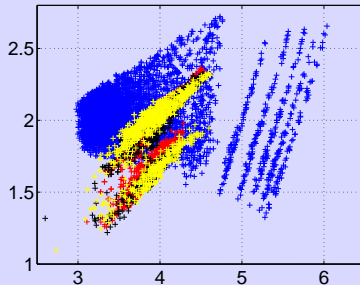
Résultats expérimentaux

Paramètre	NRMSE	
	1 axe	m axes
1	0.31	0.29 (4)
2	0.21	0.19 (5)
3	0.18	0.17 (5)
4	0.61	0.59 (5)
5	0.46	0.45 (5)
Moyenne	0.35	0.33

- 1 Image hyperspectrale
- 2 Gaussian Regularized Sliced Inverse Regression
- 3 Moindres carrés non linéaires
- 4 Données réelles**
- 5 Conclusion

Le problème !

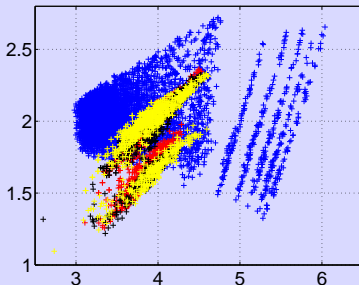
- Aux observations de la même zone géographique à des dates rapprochées doivent correspondre approximativement les mêmes estimations !
- En pratique ... pas évident :



- **Idée** : Trouver les hyperparamètres qui minimisent la variance des projections sur les axes SIR (nombre d'axes SIR et hyperparamètre de régularisation)

Le problème !

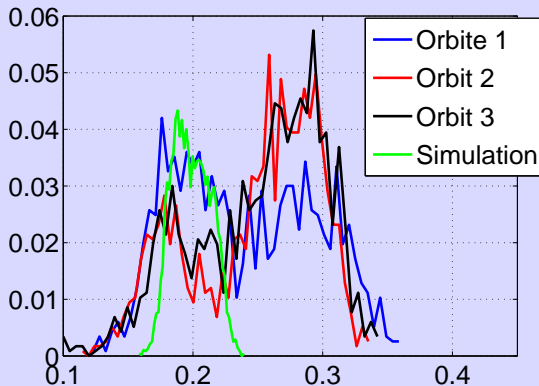
- Aux observations de la même zone géographique à des dates rapprochées doivent correspondre approximativement les mêmes estimations !
- En pratique ... pas évident :



- ~~Idée : Trouver les hyperparamètres qui minimisent la variance des projections sur les axes SIR (nombre d'axes SIR et hyperparamètre de régularisation)~~

Il reste un problème . . .

Les données simulées ne recouvrent pas les données observées

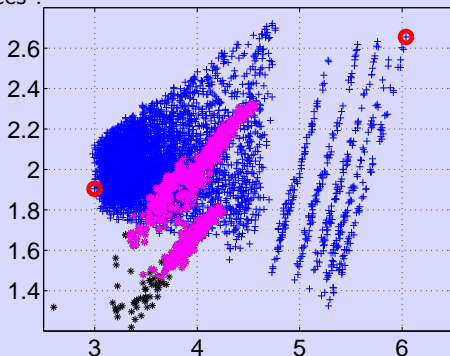


Regardons cela de plus près !

- Plus petite hyperbole enveloppante $\mathcal{B}(\mathcal{R}_B, \mathcal{C}_B,)$:

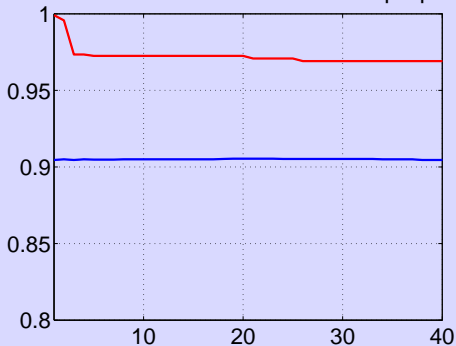
$$\min_{\mathcal{R}_B, \mathcal{C}_B} \mathcal{R}_B^2 \text{ contraint à } \|\mathbf{x}_i - \mathcal{C}_B\|^2 - \mathcal{R}_B^2 \leq 0, \forall \mathbf{x}_i$$

- Calculé dans \mathbb{R}^d puis projeté sur les deux premières axes propres des données simulées :



Intersection des \mathcal{B} sur données simulées et observées 1/4

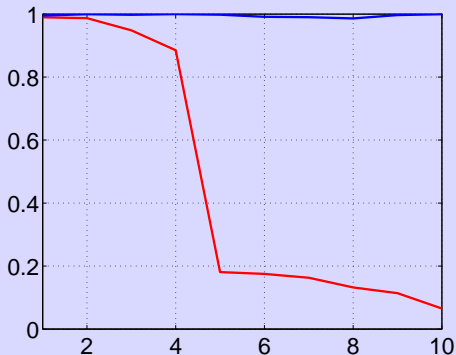
Proportion en fonction du nombres d'axes propres retenus



Proportion des observations dans $\mathcal{B}_{\text{simul}}$ et Proportion des simulations dans \mathcal{B}_{obs}

Intersection des \mathcal{B} sur données simulées et observées 2/4

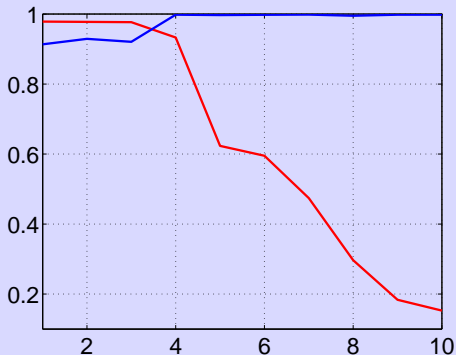
Proportion en fonction du nombres d'axes GRSIR retenus ($\tau = 10^{-8}$)



Proportion des observations dans $\mathcal{B}_{\text{simul}}$ et Proportion des simulations dans \mathcal{B}_{obs}

Intersection des \mathcal{B} sur données simulées et observées 3/4

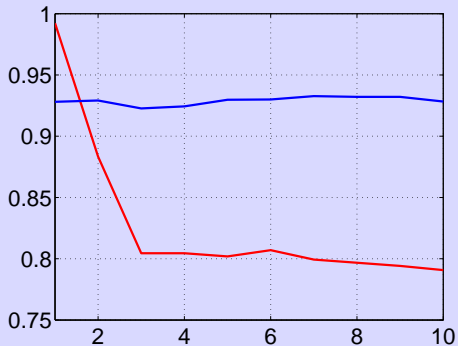
Proportion en fonction du nombres d'axes GRSIR retenus ($\tau = 10^{-5}$)



Proportion des observations dans $\mathcal{B}_{\text{simul}}$ et Proportion des simulations dans \mathcal{B}_{obs}

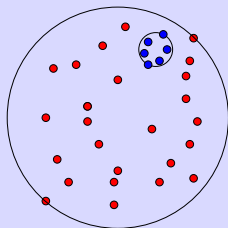
Intersection des \mathcal{B} sur données simulées et observées 4/4

Proportion en fonction du nombres d'axes GRSIR retenus ($\tau = 10^{-2}$)



Proportion des observations dans $\mathcal{B}_{\text{simul}}$ et Proportion des simulations dans \mathcal{B}_{obs}

Interprétation



- Géométrie : Simulations et observations ne sont pas projetées sur la même zone
- Statistique : Les moments d'ordre un et deux sont différents

Ajout d'une contrainte de régularisation sur la projection des observations

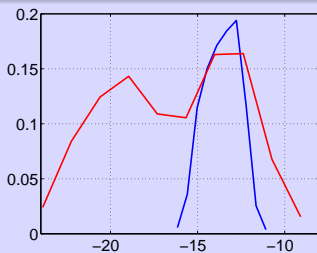
⇒ Algorithme semi-supervisé

Un algorithme

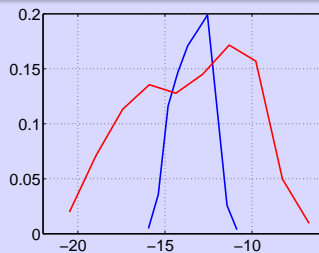
- Simulations $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ + observations $\{\mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+l}\}$
- $\mathcal{L} = \beta^t \hat{\Gamma} \beta - \nu \beta^t \Delta \beta + \lambda (1 - \beta^t \hat{\Sigma} \beta)$
- $\Delta = \frac{1}{n+l} \sum_{i,j=1}^{n+l} w_{ij} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^t$
- $\hat{\Sigma} = \frac{1}{n+l} \sum_{i=1}^{n+l} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}) (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^t$
- Finalement : $\nu > 0$ qui maximise la divergence de KL entre les projections des simul. et obs.

$$\hat{\Sigma}^{-1} [\hat{\Gamma} - \nu \Delta] \beta = \lambda \beta$$

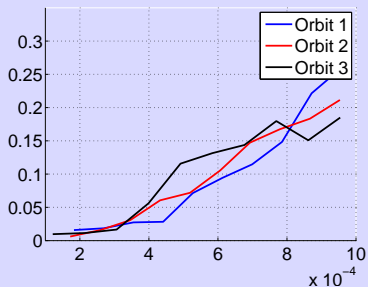
Premiers résultats



$\nu = 0$



$\nu = 0.13$



- 1 Image hyperspectrale
- 2 Gaussian Regularized Sliced Inverse Regression
- 3 Moindres carrés non linéaires
- 4 Données réelles
- 5 Conclusion**

- ☺ Plusieurs axes SIR peuvent être utiles pour l'inversion
- ☺ Il est plus intéressant d'utiliser un modèle non linéaire pour l'inversion
- ☹ Problèmes sur les données réelles : en dehors de la plage simulée, on ne sait pas ce qu'on fait !
- ☹ GRSIR 1.0 est plus facile à régler sur données réelles
- ☺ Approche semi-supervisée ...