

# Déconvolution distribuée pour les très grandes images hyperspectrales en interférométrie radio

Céline Meillier, Pascal Bianchi, Walid Hachem

12 mai 2016

4<sup>ème</sup> colloque du Groupe Hyperspectral de la SFPT

Travaux financés par l'ANR Magellan (ANR-14-CE23-0004-01)



LTCI – CNRS – Telecom Paristech  
Université Paris Saclay

## Contexte : grands interféromètres pour la radio astronomie

**Nouvelle génération d'interféromètres : *software telescopes*.**

### LOFAR :

- Gamme de fréquence : 10MHz à 240 MHz.
- 48 stations en Europe.
- 25 000 antennes.

### Square Kilometer Array :

- Gamme de fréquence : 10 MHz à 10/20 GHz.
- Stations d'antennes : Australie et Afrique du Sud.
- Stations basses fréquences :
  - 130 000 antennes,
  - 157 To/s.





# Déconvolution en grande dimension

## Problématique de déconvolution :

- Grand réseau d'antennes : PSF complexe.
- PSF connue (liées à la géométrie du réseau).
- Problème inverse à résoudre.

## Problématique "grandes dimensions" :

- Dimension des données (1 cube hyperspectral SKA  $\simeq$  80 To).
- Données stockées sur un cluster dédié.
- Proposer une architecture distribuée pour la résolution du problème inverse.

# Plan de la présentation

**Introduction**

**Modélisation et architecture distribuée**

**Algorithmes distribués**

**Résultats sur données quasi-réelles**

**Conclusion et perspectives**

# Plan de la présentation

Introduction

**Modélisation et architecture distribuée**

Algorithmes distribués

Résultats sur données quasi-réelles

Conclusion et perspectives

## Notations et modèle

Modèle des images radio-interférométriques :

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{R}^M$$

où :

→  $M = N \times L$  avec  $N$  le nombre de pixels dans une image et  $L$  le nombre de fréquence.

→  $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_L \end{pmatrix}$  et  $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_L \end{pmatrix}$ , avec  $\mathbf{y}_l$  et  $\mathbf{x}_l \in \mathbb{R}^N$  pour tout  $l = 1 \dots L$ .

→  $\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{H}_L \end{pmatrix} \in \mathbb{R}^{M \times M}$

## Formulation du problème

Problème d'optimisation sous contraintes considéré :

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_\epsilon}{2} \|\mathbf{x}\|_2^2 + \mathbf{v}_{\mathbb{R}^+}(\mathbf{x}) + \mu_s \|\mathbf{W}_s \mathbf{x}\|_1 + \mu_\nu \|\mathbf{W}_\nu \mathbf{x}\|_1 \quad (1)$$

Contraintes de parcimonie :

- sur la décomposition en ondelettes des images :  $\mu_s \|\mathbf{W}_s \mathbf{x}\|_1$ ,
- sur la décomposition en cosinus discret des spectres :  $\mu_\nu \|\mathbf{W}_\nu \mathbf{x}\|_1$ ,

Contraintes de positivité :

- image = luminosité = énergie,
- remarque :  $\mathbf{v}_{\mathbb{R}^+}(\mathbf{x})$  et  $\mu_s \|\mathbf{W}_s \mathbf{x}\|_1$  peuvent avoir des comportements antagonistes dans la minimisation.

## Résolution dans le primal

Reformulation du problème primal :

$$\min_{\mathbf{x} \in \mathbb{R}^M} f(\mathbf{x}) + h(\mathbf{W}\mathbf{x}), \quad (2)$$

avec :

$$\rightarrow f(\mathbf{x}) = (1/2)\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + (\mu_\epsilon/2)\|\mathbf{x}\|_2^2 + \mathbf{1}_{\mathbb{R}^+}(\mathbf{x}),$$

$$\rightarrow h(\mathbf{u}) = \|\mathbf{u}\|_1$$

$$\rightarrow \mathbf{W} = \begin{pmatrix} \mu_s \tilde{\mathbf{W}}_s \\ - - - \\ \mu_\nu \tilde{\mathbf{W}}_\nu \end{pmatrix}$$

Résolution du problème :

→ Algorithme primal-dual (Lagrangien augmenté)

→ Résolution du problème dual.



## Résolution dans le dual

Problème dual associé :

$$\min_{\lambda \in \mathbb{R}^{(m_s+1)M}} f^*(-W^T \lambda) + h^*(\lambda) \equiv \min_{\lambda: \|\lambda\|_\infty \leq 1} f^*(-W^T \lambda). \quad (3)$$

où  $f^*$  et  $h^*$  sont les transformées de Fenchel-Legendre de  $f$  et  $g$  :

$$\rightarrow f^*(\phi) = \frac{1}{2} \phi^T \Delta^{-1} \phi + \phi^T \Delta^{-1} H y + \frac{1}{2} y^T (H \Delta^{-1} H^T - I) y, \text{ avec } \phi = W^T \lambda,$$

$$\rightarrow h^*(\lambda) = \mathfrak{z}_{\mathcal{B}_\infty}(\lambda).$$

et  $\lambda$  est le vecteur dual :

$$\rightarrow \lambda = \begin{pmatrix} \lambda_s \\ - \\ \lambda_\nu \end{pmatrix}, \text{ avec } \lambda_s \in \mathbb{R}^{m_s M} \text{ et } \lambda_\nu \in \mathbb{R}^M$$

# Architecture distribuée

**Chargement et manipulation centralisés des données impossible :**

- ☹ Mémoire
- ☹ Puissance de calcul

# Architecture distribuée

Chargement et manipulation centralisés des données impossible :

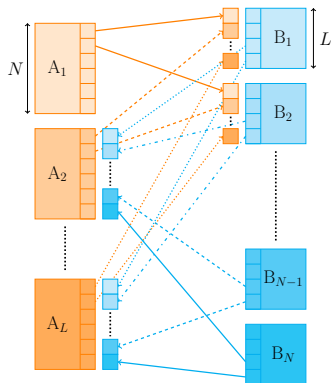


Mémoire



Puissance de calcul

Architecture proposée :



# Architecture distribuée

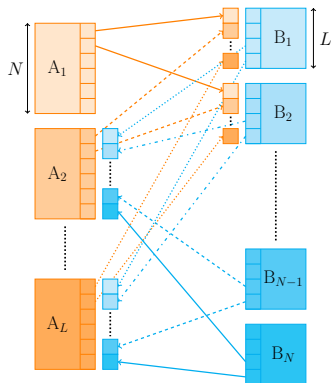
**Chargement et manipulation centralisés des données impossible :**

- ☹ Mémoire
- ☹ Puissance de calcul

**Avantage de l'architecture proposée :**

- 😊 Un noeud = quelques images ou quelques spectres
- 😊 Limitation des échanges de données

**Architecture proposée :**



# Plan de la présentation

Introduction

Modélisation et architecture distribuée

**Algorithmes distribués**

Résultats sur données quasi-réelles

Conclusion et perspectives

## Algorithmes étudiés

Algorithme	Domaine	Complexité		Mémoire	
		Noeuds A	Noeuds B	Noeuds A	Noeuds B
ADMM	primal-dual	$\mathcal{O}(N_+ \log N_+)$	$\mathcal{O}(L \log L)$	$N \times L : 7$ $m_s N \times L : 5$ $N_+ \times L : 2$	$N \times L : 8$
Gradient projeté	dual	$\mathcal{O}(N_+ \log N_+)$	$\mathcal{O}(L \log L)$	$N \times L : 6$ $m_s N \times L : 4$ $N_+ \times L : 2$	$N \times L : 7$
FISTA	dual	$\mathcal{O}(N_+ \log N_+)$	$\mathcal{O}(L \log L)$	$N \times L : 8$ $m_s N \times L : 6$ $N_+ \times L : 2$	$N \times L : 9$

où  $N_+ = N + PSF$  : taille de l'image après convolution par la PSF.

## Exemple d'algorithme : Gradient projeté

Résolution du problème dual :

$$\min_{\lambda: \|\lambda\|_{\infty} \leq 1} f^*(-\mathbf{W}^T \lambda) \quad \text{avec} \quad \mathbf{W} = \begin{pmatrix} \mu_s \mathbf{W}_s \\ - \\ - \\ \mu_{\nu} \mathbf{W}_{\nu} \end{pmatrix} \quad \text{et} \quad \lambda = \begin{pmatrix} \lambda_s \\ - \\ - \\ \lambda_{\nu} \end{pmatrix}$$

Puisque  $f^*$  est lisse on peut utiliser l'algorithme du **gradient projeté** :

$$\begin{aligned} \lambda^{k+1} &= \mathcal{P}_{\infty} \left( \lambda^k + \rho \mathbf{W} \nabla f^*(-\mathbf{W}^T \lambda^k) \right) \\ &= \mathcal{P}_{\infty} \left( \lambda^k - \rho \mathbf{W} (\mathbf{H}^T \mathbf{H} + \mu_{\epsilon} \mathbf{I})^{-1} (\mathbf{W}^T \lambda^k - \mathbf{H} \mathbf{y}) \right) \end{aligned}$$

où  $\rho > 0$  et  $\mathcal{P}_{\infty}$  est le projecteur sur  $\mathcal{B}_{\infty}$ .

## Calcul distribué du gradient

Le terme de gradient  $\mathbf{W}(\mathbf{H}^T \mathbf{H} + \mu_\epsilon \mathbf{I})^{-1} \mathbf{W}^T \lambda^k$  est distribuable selon les images et selon les spectres :

$$\mathbf{W}(\mathbf{H}^T \mathbf{H} + \mu_\epsilon \mathbf{I})^{-1} \mathbf{W}^T \lambda = \begin{pmatrix} \mu_s \mathbf{W}_s (\mathbf{H}^T \mathbf{H} + \mu_\epsilon \mathbf{I})^{-1} (\mu_s \mathbf{W}_s^T \lambda_s + \mu_\nu \mathbf{W}_\nu^T \lambda_\nu) \\ \mu_\nu \mathbf{W}_\nu (\mathbf{H}^T \mathbf{H} + \mu_\epsilon \mathbf{I})^{-1} (\mu_s \mathbf{W}_s^T \lambda_s + \mu_\nu \mathbf{W}_\nu^T \lambda_\nu) \end{pmatrix} \cong \begin{pmatrix} \lambda_s \\ - \\ \lambda_\nu \end{pmatrix}$$

### Stratégie de distribution :

- Noeuds du groupe A = opérations parallélisables en fréquence + mise à jour et stockage de  $\lambda_s$ .
- Noeuds du groupe B = opérations parallélisables en pixels + mise à jour et stockage de  $\lambda_\nu$ .



## Plan de la présentation

Introduction

Modélisation et architecture distribuée

Algorithmes distribués

**Résultats sur données quasi-réelles**

Conclusion et perspectives

## Image quasi-réelle : galaxie m31

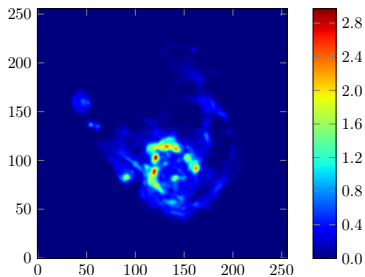


Image à reconstruire  $x$

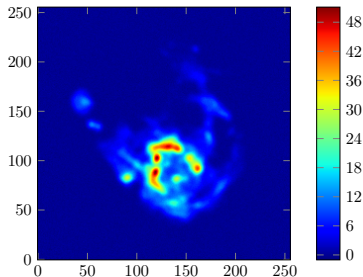
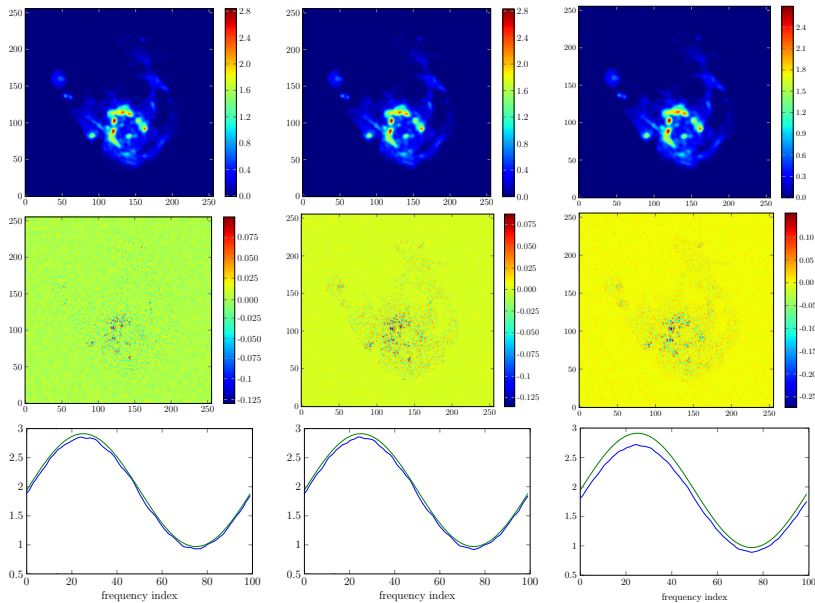


Image convoluée et bruitée  $y$

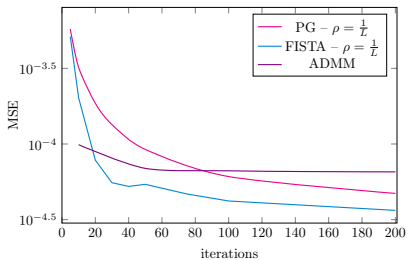
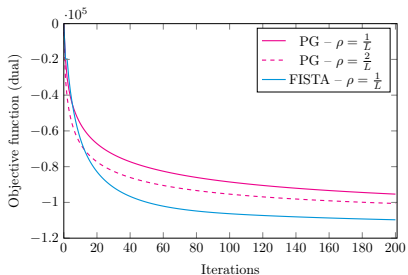
### Caractéristiques :

- Dimensions :  $256 \times 256 \times 100$  pixels.
- Bruit additif :  $15 \leq RSB \leq 25$  dB.
- PSF gaussienne de taille  $21 \times 21$  pixels.



**FIGURE:** Gauche : gradient projeté, centre : FISTA, droite : ADMM. De haut en bas : image reconstruite sur une bande spectrale (RSB = 25dB), erreur de reconstruction (biais) sur la même bande spectrale, spectre reconstruit en un point de l'image. 200 itérations.

# Performances



## Plan de la présentation

Introduction

Modélisation et architecture distribuée

Algorithmes distribués

Résultats sur données quasi-réelles

**Conclusion et perspectives**

## Conclusion et perspectives

### Algorithmes étudiés :

- L'algorithme du gradient projeté et son accélération FISTA : bon compromis complexité-gestion mémoire-convergence.
- ADMM : performances (déconvolution et complexité) satisfaisantes.
- L-BFGS-B : se distribue très difficilement.

### Perspectives :

- En cas de PSF longue mémoire : l'approximation par une matrice circulante pour l'inversion de la matrice de Gram  $\mathbf{H}^T \mathbf{H}$  devient inutilisable.
- Recherche d'algorithmes ne nécessitant pas d'inversion de matrice.