

# Optimisation rapide basée sur la méthode de sous-espace pour les problèmes Bayésiens Variationnels

Yuling ZHENG

*Encadrants : Thomas RODET, Aurélia FRAYSSE*

Laboratoire des Signaux et Systèmes (L2S)  
Université de Paris-Sud



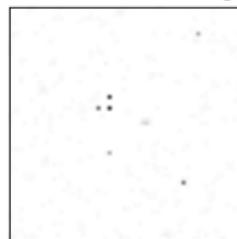
# Introduction

Problèmes abordés :

- Déconvolution

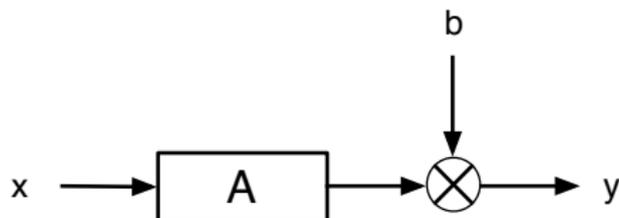


- Tomographie



- Super-résolution





## ■ Modèle direct :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad (1)$$

où l'opérateur  $\mathbf{A} \in \mathbb{R}^{M \times N}$  est connu et  $\mathbf{b}$  est i.i.d. Gaussien.

## ■ Objectif : Retrouver $\mathbf{x}$ à partir des données $\mathbf{y} \Leftrightarrow$ Problème Inverse, généralement **mal posé**

## ■ Méthodes

- Régularisation
- Approches Bayésiennes

# Approches bayésiennes

- Vraisemblance

$$p(\mathbf{y}|\mathbf{x}, \gamma_{\mathbf{b}}) = (2\pi/\gamma_{\mathbf{b}})^{-M/2} \exp \left[ -\frac{\gamma_{\mathbf{b}} \|\mathbf{y} - \mathbf{Ax}\|^2}{2} \right] \quad (2)$$

- Loi *a priori* :

$$p(\mathbf{x}|\gamma_p) = C(\gamma_p) \exp \left[ -\frac{1}{2} \gamma_p \mathcal{R}(\mathbf{x}) \right] \quad (3)$$

- Loi *a posteriori* (en utilisant la règle de Bayes)

$$p(\mathbf{x}|\mathbf{y}, \gamma_{\mathbf{b}}, \gamma_p) \propto p(\mathbf{x}, \mathbf{y}|\gamma_{\mathbf{b}}, \gamma_p) = p(\mathbf{y}|\mathbf{x}, \gamma_{\mathbf{b}})p(\mathbf{x}|\gamma_p) \quad (4)$$

$$\propto \exp \left[ -\frac{\gamma_{\mathbf{b}} \|\mathbf{y} - \mathbf{Ax}\|^2}{2} - \frac{1}{2} \gamma_p \mathcal{R}(\mathbf{x}) \right] \quad (5)$$

- Lien avec la régularisation :

Estimation maximum *a posteriori* (MAP)

$$\Rightarrow \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{\gamma_{\mathbf{b}} \|\mathbf{y} - \mathbf{Ax}\|^2}{2} + \frac{1}{2} \gamma_p \mathcal{R}(\mathbf{x}) \Leftrightarrow \text{Régularisation avec } \lambda = \frac{\gamma_p}{\gamma_{\mathbf{b}}}$$

## Estimer conjointement $\mathbf{x}$ et les hyperparamètres

- Loi *a priori* sur les hyperparamètres : Jeffreys

$$p(\gamma_{\mathbf{b}}) = \gamma_{\mathbf{b}}^{-1} \quad p(\gamma_p) = \gamma_p^{-1} \quad (6)$$

- Loi *a posteriori* (en utilisant la règle de Bayes)

$$p(\mathbf{x}, \gamma_{\mathbf{b}}, \gamma_p | \mathbf{y}) \propto p(\mathbf{x}, \gamma_{\mathbf{b}}, \gamma_p, \mathbf{y}) = p(\mathbf{y} | \mathbf{x}, \gamma_{\mathbf{b}}) p(\mathbf{x} | \gamma_p) p(\gamma_{\mathbf{b}}) p(\gamma_p) \quad (7)$$

$$\propto \gamma_{\mathbf{b}}^{M/2-1} C(\gamma_p) \gamma_p^{-1} \exp \left[ -\frac{\gamma_{\mathbf{b}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2} - \frac{1}{2} \gamma_p \mathcal{R}(\mathbf{x}) \right] \quad (8)$$

- Difficulté : loi *a posteriori* trop compliquée
  - Maximum *a posteriori* : problème d'optimisation non-convexe
  - Moyenne *a posteriori* : infaisable car la constante de normalisation est inconnue
- Solutions :
  - Markov chain Monte Carlo (MCMC) : coûteuse en temps de calcul
  - Méthode Bayésienne variationnelle

## Principe

Approcher la loi *a posteriori*  $p(\mathbf{x}, \gamma_{\mathbf{b}}, \gamma_p | \mathbf{y})$  par des **lois séparables**

$$q(\mathbf{x}, \gamma_{\mathbf{b}}, \gamma_p) = \prod_i q_i(\mathbf{x}_i) q_{\mathbf{b}}(\gamma_{\mathbf{b}}) q_p(\gamma_p) \quad (9)$$

en minimisant la divergence de Kullback-Leibler entre  $q(\cdot)$  et  $p(\cdot | \mathbf{y})$  –  $\mathcal{KL}[q(\cdot) \| p(\cdot | \mathbf{y})]$ .

Notons  $\mathbf{w} = (\mathbf{x}, \gamma_{\mathbf{b}}, \gamma_p)$ , problème d'optimisation convexe à résoudre :

$$q^{opt} = \arg \min_{q \text{ p.d.f. separable}} \mathcal{KL}[q(\mathbf{w}) \| p(\mathbf{w} | \mathbf{y})] \quad (10)$$

où  $\mathcal{KL}[q(\mathbf{w}) \| p(\mathbf{w} | \mathbf{y})] = \int_{\mathbb{R}^N} q(\mathbf{w}) \log \left( \frac{q(\mathbf{w})}{p(\mathbf{w} | \mathbf{y})} \right) d\mathbf{w}$

## Optimisation dans un espace de densité de probabilité

# Approche Bayésienne variationnelle classique

- La solution de (10) est donnée par  $q(\mathbf{w}) = \prod_i q_i(w_i)$ , où

$$q_i(w_i) = \frac{1}{K_i} \exp \left( \langle \log p(\mathbf{y}, \mathbf{w}) \rangle_{\prod_{j \neq i} q_j(w_j)} \right). \quad (11)$$

- Algorithme bayésien variationnel classique

- 1 Initialisation
- 2 Mettre à jour  $q_1^{k+1}(x_1)$
- 3 ...
- 4 Mettre à jour  $q_N^{k+1}(x_N)$  avec

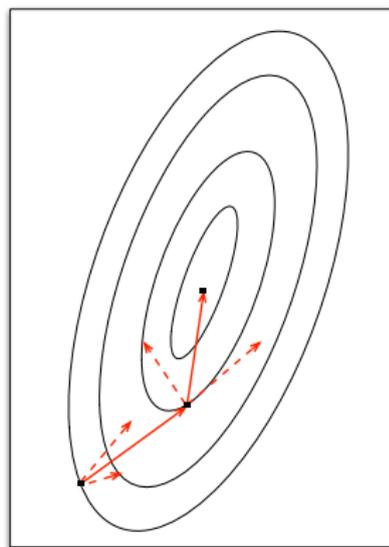
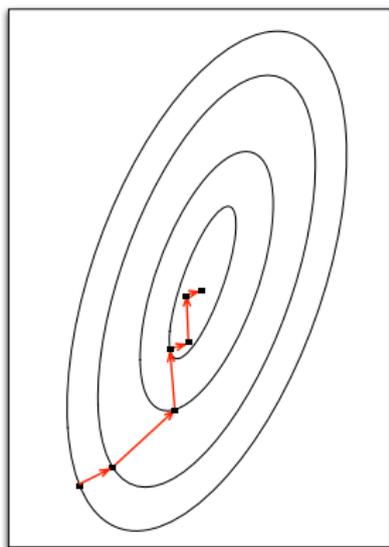
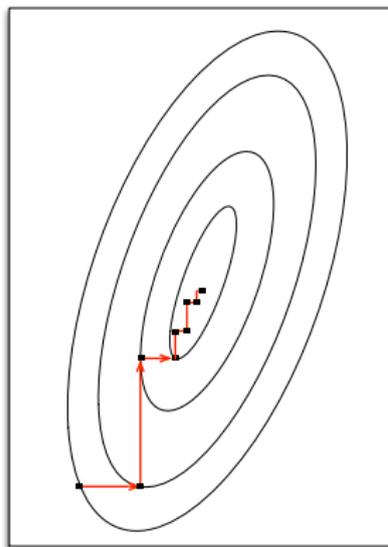
$$q_i^{k+1}(x_i) \propto \exp \left( \langle \log p(\mathbf{y}, \mathbf{x}, \gamma_{\mathbf{b}}, \gamma_{\mathbf{p}}) \rangle_{\prod_{j < i} q_j^{k+1}(x_j) \prod_{j > i} q_j^k(x_j) q_{\mathbf{b}}^k(\gamma_{\mathbf{b}}) q_{\mathbf{p}}^k(\gamma_{\mathbf{p}})} \right) \quad (12)$$

- 5 Mettre à jour  $q_{\mathbf{b}}^{k+1}(\gamma_{\mathbf{b}})$  et  $q_{\mathbf{p}}^{k+1}(\gamma_{\mathbf{p}})$
- 6 Retourner à 2 jusqu'à convergence

**Désavantage : Utilisation de l'algorithme alterné – inefficace**

# Méthodes d'optimisation dans $\mathbb{R}^N$

- Algorithme alterné
- Algorithme du gradient
- Méthode des sous-espaces  $\Rightarrow$  adapté pour nos algorithmes bayésiens variationnels accélérés.



# Méthode des sous-espaces dans $\mathbb{R}^N$

[Chouzenoux *et al.* , 2011]

- L'équation d'itération :

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \delta^k = \mathbf{x}^k + \mathbf{D}^k \mathbf{s}^k \quad (13)$$

où  $\mathbf{D}^k = [\mathbf{d}_1^k, \dots, \mathbf{d}_M^k]$  : directions définissant le sous-espace,  
 $\mathbf{s}^k = [s_1^k, \dots, s_M^k]$  : pas suivant chaque direction.

$$M = 2$$

Pour le compromis entre la rapidité et la complexité

- Construction des sous-espaces :

$$SG : \quad \mathbf{d}_1^k = \mathbf{g}^k, \mathbf{d}_2^k = \mathbf{g}^{k-1}, \quad (14)$$

$$GM : \quad \mathbf{d}_1^k = \mathbf{g}^k, \mathbf{d}_2^k = \delta^{k-1}. \quad (15)$$

où  $\mathbf{g}^k$  et  $\mathbf{g}^{k-1}$  : gradient à l'itération  $k$  et  $k - 1$ .  
 $\delta^{k-1}$  : direction précédente.

# Méthode des sous-espaces pour le problème Bayésien variationnel

- L'équation d'itération :

$$q^{k+1}(\mathbf{w}) = K^k q^k(\mathbf{w}) \exp(\delta^k) \quad (16)$$

$$\text{où } \delta^k = s_1 d_1^k(\mathbf{w}) + s_2 d_2^k(\mathbf{w}).$$

- Sous-espace Gradient à mémoire (GM) :

$$d_1^k(\mathbf{w}) = df(q^k, \mathbf{w}), \quad (17)$$

$$d_2^k(\mathbf{w}) = \delta^{k-1}(\mathbf{w}) = \ln\left(\frac{q^k}{q^{k-1}}\right). \quad (18)$$

- Distribution qui dépend de  $\mathbf{s} = [s_1, s_2]$ , le pas de l'algorithme

$$q^{\mathbf{s}}(\mathbf{w}) = K^k \prod_i q_i^k \left( \frac{\exp(\langle \log p(\mathbf{y}, \mathbf{w}) \rangle_{\prod_{j \neq i} q_j^k})}{q_i^k} \right)^{s_1} \left( \frac{q_i^k}{q_i^{k-1}} \right)^{s_2} \quad (19)$$

# Algorithme bayésien variationnel (BV) proposé

## ■ Notre approche bayésienne variationnelle

1 Initialisation

2 Calculer la fonction intermédiaire :  $q_i^r(x_i) = \exp(\langle \log p(\mathbf{y}, \mathbf{w}) \rangle_{\prod_{j \neq i} q_j^k})$

3 Déterminer l'exponentiel du sous-espace :  $\left[ \frac{q_i^r}{q_i^k}, \frac{q_i^k}{q_i^{k-1}} \right]$

4 Déterminer le pas de l'algorithme  $\mathbf{s}^{subopt} = [s_1^{subopt}, s_2^{subopt}]$

5 Mettre à jour  $(q_i^{k+1}(x_i))_{i=1, \dots, N}$

$$q_i^{k+1}(x_i) \propto q_i^k(x_i) \left( \frac{\exp(\langle \log p(\mathbf{y}, \mathbf{x}, \gamma_{\mathbf{b}}, \gamma_{\mathbf{p}}) \rangle_{\prod_{j \neq i} q_j^k(x_j) q_{\mathbf{b}}^k(\gamma_{\mathbf{b}}) q_{\mathbf{p}}^k(\gamma_{\mathbf{p}})})}{q_i^k(x_i)} \right)^{s_1^{subopt}} \\ \times \left( \frac{q_i^k(x_i)}{q_i^{k-1}(x_i)} \right)^{s_2^{subopt}} \quad (20)$$

6 Mettre à jour  $q_{\mathbf{b}}^{k+1}(\gamma_{\mathbf{b}})$  et  $q_{\mathbf{p}}^{k+1}(\gamma_{\mathbf{p}})$

7 Retourner à 2 jusqu'à convergence

# Application à un problème de tomographie (1/2)

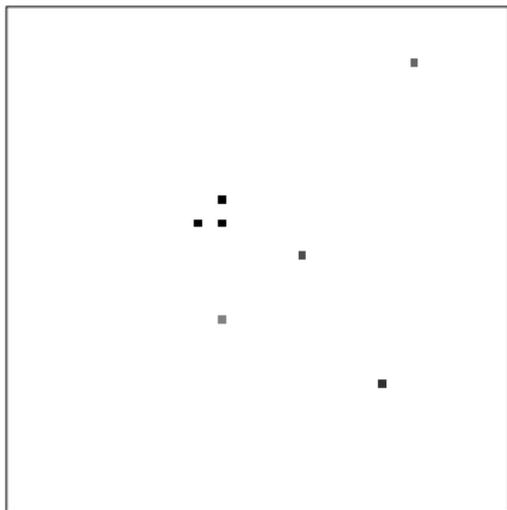
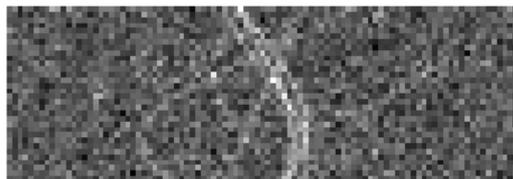


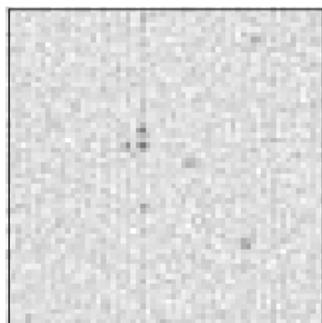
Image vraie (Taille :  
 $64 \times 64 = 4096$ )



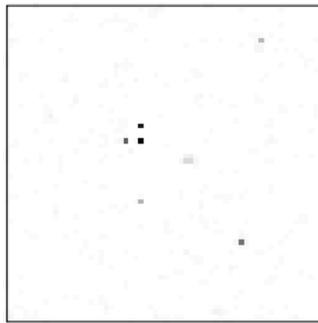
Données (Taille :  
 $32 \times 95 = 3040$ )

# Résultats – Application à un problème de tomographie (2/2)

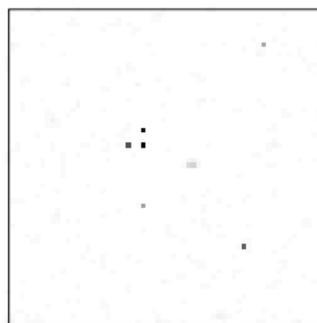
- Reconstructions avec quatre différentes méthodes :



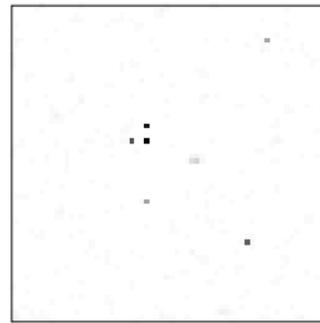
FBP



BV classique



Gradient



Algo BV proposé

- Temps de Calcul :

Méthode	BV classique	Gradient	Notre approche
PSNR(dB)	35.1	35.9	35.9
Temps(s)	723.5	23.4	3.2

# Problème de super-résolution



- Étant donné un certain nombre d'images de basse résolution, estimer une image de haute-résolution  $\implies$

- Opérateur  $\mathbf{A}$  :

**Déformation** (translation, rotation...),  
**Convolution**, **Décimation**



# Application à un problème Super-résolution

- La distribution *a priori* de  $\mathbf{x}$

$$p(\mathbf{x}|\gamma_p) = \frac{1}{Z_{TV}(\gamma_p)} \exp[-\gamma_p TV(\mathbf{x})], \quad (21)$$

où

$$TV(\mathbf{x}) = \sum_{i=1}^N \sqrt{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2}. \quad (22)$$

Ici,  $\Delta_i^h$ ,  $\Delta_i^v$  : les différences d'ordre un (horizontale, verticale).

**Problème** :  $Z_{TV}(\gamma_p)$  inconnue

**Solution** : une approximation analytique [Babacan *et al.*, 2011]

$$p(\mathbf{x}|\gamma_p) \approx \tilde{p}(\mathbf{x}|\gamma_p) = c\gamma_p^{N/2} \exp[-\gamma_p TV(\mathbf{x})], \quad (23)$$

où  $c$  est une constante

- **Difficulté liée à l'a priori TV**

**Critère incalculable**

# Application à un problème Super-résolution

## ■ Majoration-Minimisation (MM)

$$\begin{aligned} \forall a \geq 0, b > 0 \quad \sqrt{a} &\leq \frac{a+b}{2\sqrt{b}} && \implies \\ &\tilde{p}(\mathbf{x}|\gamma_p) \geq M(\mathbf{x}, \gamma_p|\boldsymbol{\lambda}) \\ &= c\gamma_p^{N/2} \exp \left[ -\gamma_p \sum_{i=1}^N \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + \lambda_i}{2\sqrt{\lambda_i}} \right], \end{aligned} \quad (24)$$

où  $(\lambda_i)_{i=1,\dots,N}$  : variables auxiliaires positives.

$\implies$  Une majoration de la divergence de Kullback-Leibler  $\mathcal{KL}^M$

## ■ Algorithme alterné pour minimiser la divergence $\mathcal{KL}$

- 1 Initialisation
- 2 Minimiser la majoration  $\mathcal{KL}^M$  par rapport à  $q$
- 3 Minimiser la majoration  $\mathcal{KL}^M$  par rapport à  $(\lambda_i)_{i=1,\dots,N}$
- 4 Retourner à 2 jusqu'à convergence

# Résultats – Application à un problème Super-résolution

Taille de données :

$$64 \times 64 \times 12 = 49152$$

Augmentation de la  
résolution d'un facteur 16  $\Rightarrow$

Nb d'inconnues :

$$256 \times 256 = 65536$$



Babacan (14.9s)



Gradient (6.1s)



Algo proposé (4.0s)

# Résultats – Application à un problème Super-résolution

**TABLE:** PERFORMANCE DE BABACAN [BABACAN *et al.* , 2011], DE GRADIENT ET DE NOTRE APPROCHE EN TERME DE NOMBRE D'ITÉRATIONS ET TEMPS CPU (EN SECONDES).

Data		PSNR	Babacan	Gradient	Notre approche
Testpat	5dB	9.13	27/9.4	157/8.2	<b>67/4.9</b>
	25dB	24.14	13/11.5	237/13.9	<b>92/7.7</b>
	45dB	30.92	38/38.3	311/19.6	<b>110/8.9</b>
Camera -man	5dB	11.64	30/15.7	169/10.3	<b>64/5.5</b>
	25dB	30.59	15/14.9	103/6.1	<b>49/4.0</b>
	45dB	40.62	31/86.8	205/11.8	<b>91/7.3</b>
Lena	5dB	14.81	26/56.6	210/47.9	<b>68/21.7</b>
	25dB	33.42	12/46.8	106/25.2	<b>60/19.4</b>
	45dB	38.72	29/296.4	298/72.6	<b>104/33.7</b>

## ■ Conclusions :

- Les nouveaux algorithmes bayésiens variationnels basés sur l'optimisation des sous-espaces ont été développés.
- L'application à un problème de tomographie en utilisant un *a priori* favorisant les images parcimonieuses. [ZHENG *et al.*, ICASSP 2013].
- L'application à un problème de super-résolution en utilisant un *a priori* basé sur la variation totale (TV). [ZHENG *et al.*, ICIP 2013].

## ■ Perspectives :

- Mise en œuvre de nos algorithmes en utilisant un nouvel *a priori* [Giovannelli, 2008, Geman & Yang, 1995] pour s'affranchir du problème de sous-estimer les hyperparamètres au cas TV.

# Merci !

# Références I

-  Babacan, S. D., Molina, R., & Katsaggelos, A. K. 2011.  
Variational Bayesian super resolution.  
*IEEE Trans. Image Process.*, **20**(4), 984–999.
-  Chouzenoux, E., Idier, J., & Moussaoui, S. 2011.  
A Majorize-Minimize strategy for subspace optimization applied to image restoration.  
*IEEE Trans. Image Process.*, **20**(18), 1517–1528.
-  Geman, D., & Yang, C. 1995.  
Nonlinear Image Recovery with Half-Quadratic Regularization.  
*IEEE Transactions on Image Processing*, **4**(7), 932–946.



Giovannelli, J-F. 2008.

Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function.

*IEEE Trans. Image Process.*, **17**(1), 16–26.