

# Agrégation d'estimateurs pour le débruitage d'images

Troisièmes Rencontres des Jeunes Statisticiens -Aussois

Joseph Salmon <sup>1</sup> et Erwan Le Pennec <sup>1,2</sup>

<sup>1</sup>LPMA

Université Paris Diderot - Paris 7

<sup>2</sup>Projet SELECT  
INRIA - Saclay

31 Août - 4 septembre 2009

# Sommaire

## Méthodes à noyaux et NL-Means

Image, bruit et méthodes à noyaux

Patchs et NL-Means

Liens avec les Noyaux

## Agrégation en statistique

Pré-estimateurs et agrégation

Agrégation PAC-Bayésienne

## Agrégation de Patchs

Agrégation de Patchs et résultats théoriques

Implémentation de l'estimateur PAC-Bayésien

Résultats numériques

## Conclusion

# NL-Means et Agrégation

## Cadre

- ▶ Estimer une image  $f$  à partir d'une observation bruitée  $Y$   
$$Y = f + \sigma W \quad (W \text{ Normale, centrée, réduite})$$

## État de l'art

- ▶ Lissage selon les valeurs d'un pixel (pb : floutage)
- ▶ Approche orientée "Patch" : utiliser des voisinages de pixels plutôt que juste les valeurs des pixels
- ▶ NL-Means : Lissage gaussien dans un espace de Patches

## Point de vue de l'agrégation d'estimateurs

- ▶ Regarder les NL-means comme une recherche de la meilleur combinaison de patches
- ▶ Cadre statistique de l'agrégation
- ▶ Nouveau point de vue et nouveaux résultats...

# Image, bruit et estimation



## Image $N \times N$

- ▶  $f(i_1, i_2) \in \mathbb{R}$  avec  $i = (i_1, i_2) \in \llbracket 1, N \rrbracket^2$ .
- ▶ Norme euclidienne

## Observation bruitée

- ▶  $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2)$  .
- ▶  $W(i_1, i_2)$  bruits gaussiens i.i.d. centrés, réduits.  $\sigma^2$  connue
- ▶ Autres bruits possibles

## Estimation

- ▶ Estimer  $f(i_1, i_2)$  à partir de  $Y$
- ▶ Comportement non local possible...

# Méthodes à Noyaux

## Principe

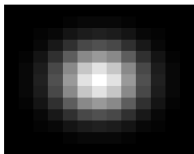
- ▶ Estimer  $f(i_1, i_2)$  par une moyenne locale

$$\hat{f}(i_1, i_2) = \sum_{(k_1, k_2) \in \llbracket 1, N \rrbracket^2} \theta_{i_1, i_2, k_1, k_2} Y_{k_1, k_2}$$

- ▶ Les poids  $\theta_{i_1, i_2, k_1, k_2} = \theta_{i, k}$  peuvent (vont) dependre de  $Y$

## Noyau Classique

- ▶  $\theta_{i_1, i_2, k_1, k_2} = \frac{K(i_1 - k_1, i_2 - k_2)}{\sum_{k'_1, k'_2} K(i_1 - k'_1, i_2 - k'_2)}$  (ne dépend pas de  $Y$ )



- ▶ Exemple : Noyau gaussien  $K(i_1, i_2) = e^{-(i_1^2 + i_2^2)/2h^2}$

# Noyaux dépendants des données

## Bilateral filtering

$$\theta_{i,k} = \frac{K(i_1 - k_1, i_2 - k_2) \times K'(Y(i_1, i_2) - Y(k_1, k_2))}{\sum_{k'_1, k'_2} K(i_1 - k'_1, i_2 - k'_2) \times K'(Y(i_1, i_2) - Y(k'_1, k'_2))}$$

- Version Gaussienne :

$$\theta_{i,k} = \frac{e^{-\frac{(i_1 - k_1)^2 + (i_2 - k_2)^2}{2h^2}} \times e^{-\frac{(Y(i_1, i_2) - Y(k_1, k_2))^2}{2h'^2}}}{\sum_{k'_1, k'_2} e^{-\frac{(i_1 - k'_1)^2 + (i_2 - k'_2)^2}{2h^2}} \times e^{-\frac{(Y(i_1, i_2) - Y(k'_1, k'_2))^2}{2h'^2}}}$$



- Intuition : moyenne des pixels proches en distance et en valeurs
- Limite : la valeur d'un pixel est un paramètre trop local

# Méthode à Patches

## Patch

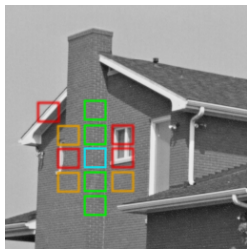
- ▶ Patch : version moins localisée que la valeur d'un pixel
- ▶ Patch centré en  $P(f)(i)$  de largeur  $w$  :

$$P(f)(i)(k_1, k_2) = f(i_1 + k_1, i_1 + k_2), \text{ with } 0 \leq |k_1|, |k_2| \leq \frac{w-1}{2}$$

- ▶ On retrouve facilement une image  $f$  à partir de patches  $P(f)$

## Intuition

- ▶ Utiliser des poids qui rendent compte de la ressemblance des patches

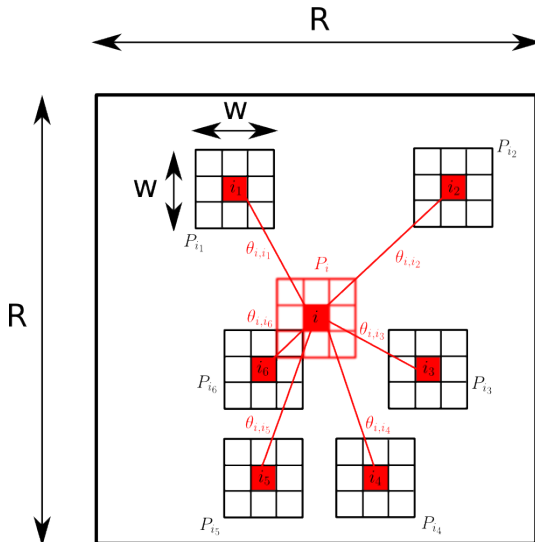


Patch  $P(Y)(i) = P_i$  :

- ▶ Patch  $P$  à débruiter
- ▶ Patches similaires, utiles : fort poids
- ▶ Patches moins similaires, moins utiles : petits poids
- ▶ Patches très différents, inutiles : poids à zéro

# Zone de recherche, poids et patches

Largeur du patch :  $w=3$ , largeur de la zone de recherche :  $R$





# NL-Means

## NL-Means [BCM05]

- ▶ Choisir une mesure de similarité  $D$  entre patches

- ▶ Utiliser des poids  $\theta_{i,k} = \frac{K'(D(P_i, P_k))}{\sum_k K'(D(P_i, P_k))}$  avec

$D(P_i, P_k) = \|P_i - P_k\|$  pour mesurer la similarité, un noyau gaussien  $K'(x) = \exp(-x^2/\beta)$  et une température  $\beta = \gamma\sigma^2$ .

## Résultats

- ▶ Méthode simple et efficace
- ▶ Performance proche des meilleurs méthodes de débruitage

## Variations

- ▶ Choix automatique de la taille de la zone de recherche (Kervrann et al. [KB06])
- ▶ Polynômes locaux d'ordre plus élevé (Takeda et al. [TFM07])
- ▶ Changer la mesure de similarité (Azzabou et al. [APG07])

# Interprétation des NL-Means

## Noyau local optimal ?



- ▶ les NL-Means trouve un noyau local adapté à la géométrie !

## Le meilleur noyau local ?

- ▶ Peut-on comparer les NL-means au meilleur noyau local

$$\mathbb{E}(\|f - \hat{f}\|^2) \leq C \arg \min_{\theta} \sum_i |f(i) - \sum_k \theta_{i,k} f(k)|^2 + N^2 \sigma^2 \|\theta\|^2 ?$$

# Consistance de la méthode [BCM05]

## Limites :

- ▶ besoin d'hypothèses fortes : stationnarité,  $\beta$ -mélange (vraies pour les textures ...)
- ▶ zone de recherche = image entière : algorithme trop lent, déraisonnable en pratique
- ▶ fenêtre  $\rightarrow 0$  : choix [BCM05]  $12\sigma^2$  ?  
Solution partielle [KB06] méthode Lepskii.

## Morale :

Le cadre asymptotique explique mal les bonnes performances pratiques  $\Rightarrow$  Changer de point de vue !

# Pré-estimateurs et agrégation

## Modèle et Pré-estimateurs

- ▶  $Y = f + \sigma W$  de taille  $N \times N$ .
- ▶  $\{P_k\}$  ensemble de  $M$  de pré-estimateurs de  $f$  (obtenus indépendamment)

## Agrégation

- ▶ Estimer  $f$  comme moyenne pondérée  $\hat{f} = P_\theta = \sum_k \theta_k P_k$
- ▶ Procédure d'agrégation : manière de choisir  $\theta_k$  à partir de  $Y$

## Inégalité Oracle

- ▶ Résultat typique : agrégation "Optimale" parmi une classe  $\Theta$ ,

$$\mathbb{E}(\|f - \hat{f}\|^2) \leq C \inf_{\theta \in \Theta} \|f - P_\theta\|^2 + \text{pen}(\theta, \sigma)$$

- ▶  $C$ ,  $\Theta$  et  $\text{pen}$  dépendant de la procédure.

## Poids Exponentiels

$$\hat{\theta}_j^{EW} \triangleq \frac{\exp(-\|Y - P_j\|^2/\beta)}{\sum_{k=1}^M \exp(-\|Y - P_k\|^2/\beta)}$$

### Inégalité Oracle (Sélection de Modèle) [Lec07]

Soit  $P_1, \dots, P_M$  des fonctions bornées sur  $[0, 1]$ . Alors il existe une constante  $K > 0$ , telle que pour tout  $\varepsilon > 0$ , on a :

$$\mathbb{E}\|f_{\hat{\theta}^{EW}} - f\|^2 \leq (1 + \varepsilon) \min_{j=1, \dots, M} (\|f - P_j\|^2) + \frac{K \log M}{\varepsilon}$$

# Agrégation PAC-Bayésienne

## Agrégation PAC-Bayésienne

- ▶ Procédure d'agrégation basée sur des poids exponentiels
- ▶ Définie pour un a priori  $\pi$  sur  $\mathbb{R}^M$  par  $\hat{f} = P_{\theta_\pi}$ , avec

$$\theta_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \|Y - P_\theta\|^2}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \|Y - P_{\theta'}\|^2} d\pi(\theta')} \theta d\pi(\theta).$$

- ▶ Pour l'a priori  $\pi = \sum_k \delta_k$  :  $\hat{f} = \sum_k \frac{e^{-\frac{1}{\beta} \|Y - P_k\|^2}}{\sum_{k'} e^{-\frac{1}{\beta} \|Y - P_{k'}\|^2}} P_k$ .

## Inégalité Oracle

- ▶ Inégalité Oracle : si la température  $\beta \geq 4\sigma^2$ ,

$$\mathbb{E}(\|f - \hat{f}\|^2) \leq \inf_p \int_{\theta \in \mathbb{R}^M} \|f - P_\theta\|^2 dp + \beta \mathcal{K}(p, \pi).$$

avec  $\mathcal{K}(p, \pi)$  la divergence de Kullback-Leibler

# Choix de l'a priori

## Borne de l'erreur et a priori

- ▶  $\mathbb{E}(\|f - \hat{f}\|^2) \leq \inf_p \int_{\theta \in \mathbb{R}^M} \|f - P_\theta\|^2 dp + \beta \mathcal{K}(p, \pi).$
- ▶ Compromis entre la localisation de  $p$  proche du meilleur agrégat "Oracle"  $P_\theta$  et une proximité avec l'a priori  $\pi$ .
- ▶ Choisir  $\pi$  afin que cette quantité soit petite "uniformément"...

## A priori discret

- ▶ A priori  $\pi = \sum_k \delta_k : \mathbb{E}(\|f - \hat{f}\|^2) \leq \inf_k \|f - P_k\|^2 + \beta \log M.$
- ▶ Aussi bon que le meilleur pré-estimateur...

## A priori sparsifiant

- ▶  $\pi$  : i.i.d. loi à queues épaisses : Student [DT07]
- ▶ Borne :  $\mathbb{E}(\|f - \hat{f}\|^2) \leq \inf_{\theta \in \mathbb{R}^M} \|f - P_\theta\|^2 + C\beta \|\theta\|_0 \log M.$
- ▶ Aussi bien que le meilleur agrégat "sparse"...

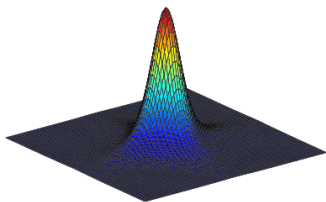
# Inégalité Oracle Parcimonieuse

Prenons :

$\Theta = \mathbb{R}^M$  et  $\Pi(d\theta) = \pi(\theta) d\theta$  :

$$\pi_0(t) = \frac{2}{\pi(1+t^2)^2}, \quad \forall t \in \mathbb{R}$$

$$\pi(\theta) = \prod_{j=1}^M \frac{1}{\tau} \pi_0\left(\frac{\theta_j}{\tau}\right), \quad \forall \theta \in \mathbb{R}^M$$



$\pi$  en dimension 2

## Inégalité Oracle "Sparse" [DT07]

Si  $W \sim \mathcal{N}(0, \sigma^2)$ ,  $\tau = \frac{\beta}{\sqrt{nM}}$  et  $\beta \geq 4\sigma^2$ , il existe une constante  $c$

$$\mathbb{E} \left( \|\hat{f} - f\|^2 \right) \leq \inf_{\theta \in \mathbb{R}^M} \left[ \|\mathbf{f}_\theta - f\|^2 + 4\beta \|\theta\|_0 \log \left( 1 + \frac{\|\theta\|_1}{\tau} \right) + c\beta \right]$$



# Agrégation de Patches

## Patches en tant que pré-estimateurs

- ▶ Utiliser les patches  $P(Y)(k, l)$  comme pré-estimateurs
- ▶ Seul problème : non indépendance avec l'observée  $P(Y)(i, j)$ .

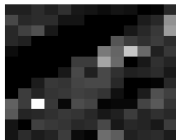
## Théorème ?

- ▶ Même type que pour l'agrégation :

$$\begin{aligned} & \mathbb{E}(\|P(f)(i, j) - P(\hat{f})(i, j)\|^2) \\ & \leq \inf_p \int_{\theta \in \mathbb{R}^M} \left( \|P(f)(i, j) - P_\theta\|^2 + N^2 \sigma^2 \|\theta\|^2 \right) dp + \beta \mathcal{K}(p, \pi) \end{aligned}$$

- ▶ preuve : besoin d'un découpage ou de plus de travail...

## A priori sur les patches



- ▶ Discrèt (NL-Means) : sélection...
- ▶ "Sparsifiant" (Student, mélange Gaussien) : noyau optimal sparse !

# Estimateur PAC-Bayésien et méthodes de Monte Carlo

## Implémentation de l'estimateur PAC-Bayésien

- ▶ Intégrale en grande dimension, similaire à des intégrales apparaissant dans le cadre bayésien
- ▶ Problème important !
- ▶ Méthode de Monte Carlo basée sur une diffusion de Langevin
- ▶ Valeur approchée seulement... mais précision suffisante
- ▶ Encore des problèmes de convergence
- ▶ Une pré-sélection des patches aide...

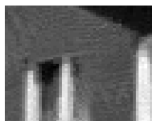
# Résultats numériques (PSNR)



Originale



Bruitée (22.06 dB)



NL Means (29.69 dB)



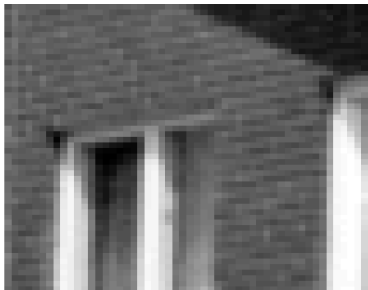
PAC-Bayésien (29.69 dB)

## Cadre expérimental

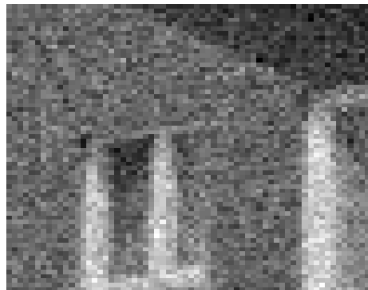
- ▶ Comparaison avec les NL-Means classiques  $\gamma = 12$
- ▶ Agrégation PAC-Bayésienne avec a priori de Student

## Resultats

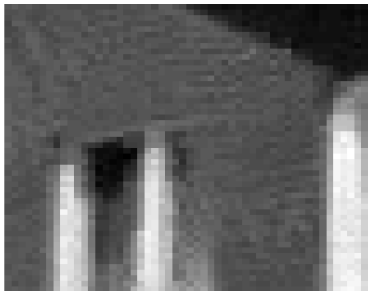
- ▶ Semblable aux NL-Means
- ▶ Améliorations possibles...



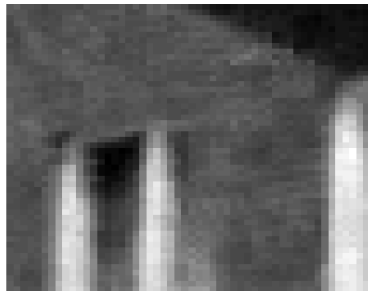
Originale



Bruitée (22.06 dB)



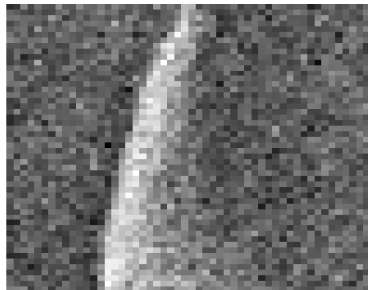
NL Means (29.69 dB)



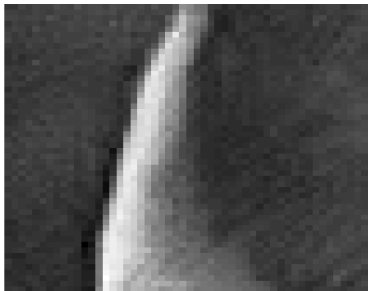
PAC-Bayésien (29.69 dB)



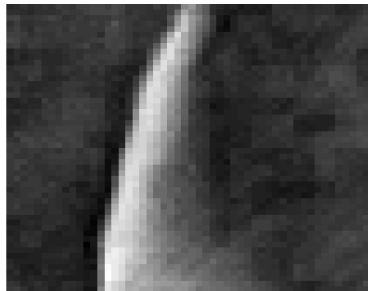
Originale



Bruitée (22.28 dB)



NL Means (31.59 dB)



PAC-Bayésien (30.78 dB)



Originale



Bruitée (22.21 dB)



NL Means (24.23dB)



PAC-Bayésien (26.96 dB)

# Conclusion

## Un nouveau regard sur les NL-means

- ▶ Un nouveau regard sur les poids exponentiels et sur la distance  $L_2$  pour la similarité entre patches
- ▶ Nouvelle procédure d'agrégation de performance comparable aux NL-Means avec un (peu de) contrôle théorique
- ▶ Adaptable à d'autres dictionnaires (pas seulement ceux provenant de l'image elle-même)

## Encore du pain sur la planche

- ▶ Étendre les théorèmes au cas dépendant,
- ▶ Choix du meilleur a priori
- ▶ Accélérer la convergence des chaînes de Monte Carlo
- ▶ ...

# Références bibliographiques I

- ▶ Noura Azzabou, Nikos Paragios, and Frederic Guichard.  
Image denoising based on adapted dictionary computation.  
In *ICIP (3)*, pages 109–112. IEEE, 2007.
- ▶ Antoni Buades, Bartomeu Coll, and Jean-Michel Morel.  
A review of image denoising algorithms, with a new one.  
*Multiscale Model. Simul.*, 4(2) :490–530 (electronic), 2005.
- ▶ A. S. Dalalyan and A. B. Tsybakov.  
Aggregation by exponential weighting, sharp oracle inequalities  
and sparsity.  
In *20th Annual Conference on Learning Theory, COLT*, pages  
97–111, 2007.
- ▶ Charles Kervrann and Jérôme Boulanger.  
Optimal spatial adaptation for patch-based image denoising.  
*IEEETIP*, 15(10) :2866–2878, 2006.



## Références bibliographiques II

- ▶ Guillaume Lécué.  
Optimal rates of aggregation in classification under low noise assumption.  
*Bernoulli*, 13(4) :1000–1022, 2007.
- ▶ H. Takeda, S. Farsiu, and P. Milanfar.  
Kernel regression for image processing and reconstruction.  
*IEEE Transactions on Image Processing*, 16(2) :349–366, 2007.