

# Un point de vue agrégé sur les NL-Means

E. Le Pennec et J. Salmon

LPMA - Université Paris Diderot (Paris 7)  
SELECT - INRIA Saclay, Université Paris Sud (Orsay)

Oct 2009 - Journées STAR

# NL-Means et agrégation

## Cadre

- Estimer une image  $I$  à partir d'une observation bruitée  $Y$   
 $Y = I + \sigma W$  ( $W$  bruit blanc gaussien)

# NL-Means et agrégation

## Cadre

- Estimer une image  $I$  à partir d'une observation bruitée  $Y$   
$$Y = I + \sigma W \quad (W \text{ bruit blanc gaussien})$$

## État de l'art

- Solution classique : moyennage local des valeurs des pixels.
- Approche « Patch » : utiliser des voisinages de pixels plutôt que les pixels seuls.
- NL-Means : Lissage gaussien dans l'espace des patches.

# NL-Means et agrégation

## Cadre

- Estimer une image  $I$  à partir d'une observation bruitée  $Y$   
$$Y = I + \sigma W \quad (W \text{ bruit blanc gaussien})$$

## État de l'art

- Solution classique : moyennage local des valeurs des pixels.
- Approche « Patch » : utiliser des voisinages de pixels plutôt que les pixels seuls.
- NL-Means : Lissage gaussien dans l'espace des patches.

## Le point de vue agrégé

- Relire les NL-Means comme la recherche d'un noyau local optimal, une combinaison optimale de pixels.
- Cadre de l'agrégation statistique.
- Point de vue différent et des résultats nouveaux.

# NL-Means et agrégation

## Cadre

- Estimer une image  $I$  à partir d'une observation bruitée  $Y$   
$$Y = I + \sigma W \quad (W \text{ bruit blanc gaussien})$$

## État de l'art

- Solution classique : moyennage local des valeurs des pixels.
- Approche « Patch » : utiliser des voisinages de pixels plutôt que les pixels seuls.
- NL-Means : Lissage gaussien dans l'espace des patches.

## Le point de vue agrégé

- Relire les NL-Means comme la recherche d'un noyau local optimal, une combinaison optimale de pixels.
- Cadre de l'agrégation statistique.
- Point de vue différent et des résultats nouveaux.

# Plan

- 1 Méthodes à noyaux et NL-Means
- 2 Agrégation d'estimateurs
- 3 Agrégation de patches

# Plan

- 1 Méthodes à noyaux et NL-Means
- 2 Agrégation d'estimateurs

# Plan

- 1 Méthodes à noyaux et NL-Means
- 2 Agrégation d'estimateurs
- 3 Agrégation de patches

# Plan

- 1 Méthodes à noyaux et NL-Means
- 2 Agrégation d'estimateurs
- 3 Agrégation de patches

# Plan

- 1 Méthodes à noyaux et NL-Means
  - Image, bruit et méthodes à noyaux
  - Patches
  - NL-Means et interprétations
- 2 Agrégation d'estimateurs
- 3 Agrégation de patches

# Images, bruit et estimée

Image  $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$  avec  $(i_1, i_2) \in [1, N]^2$ .
- Perte : norme  $L_2$  (perte quadratique).

# Images, bruit et estimée



Image  $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$  avec  $(i_1, i_2) \in [1, N]^2$ .
- Perte : norme  $L_2$  (perte quadratique).

Observation bruitée

- $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2)$  .
- $W$  bruit gaussien standard i.i.d. et  $\sigma^2$  variance connue.
- Autres bruits possibles...

# Images, bruit et estimée



## Image $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$  avec  $(i_1, i_2) \in [1, N]^2$ .
- Perte : norme  $L_2$  (perte quadratique).

## Observation bruitée

- $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2)$  .
- $W$  bruit gaussien standard i.i.d. et  $\sigma^2$  variance connue.
- Autres bruits possibles...

## Estimation

- Estimer  $I(i_1, i_2)$  par  $\hat{I}(i_1, i_2)$  à partir  $Y$ .
- Comportement non local possible...

# Images, bruit et estimée



## Image $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$  avec  $(i_1, i_2) \in [1, N]^2$ .
- Perte : norme  $L_2$  (perte quadratique).

## Observation bruitée

- $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2)$  .
- $W$  bruit gaussien standard i.i.d. et  $\sigma^2$  variance connue.
- Autres bruits possibles...

## Estimation

- Estimer  $I(i_1, i_2)$  par  $\hat{I}(i_1, i_2)$  à partir  $Y$ .
- Comportement non local possible...

# Méthodes à noyaux

## Méthode à noyaux générique

- Estimer  $l(i_1, i_2)$  par moyennage  $\hat{l}(i_1, i_2) = \sum_{(k_1, k_2) \in [1, M]^2} \lambda_{i_1, i_2, k_1, k_2} Y_{k_1, k_2}$
- Les poids  $\lambda_{i_1, i_2, k_1, k_2}$  peuvent (vont) dépendre de la position et de  $Y$ .

# Méthodes à noyaux

## Méthode à noyaux générique

- Estimer  $I(i_1, i_2)$  par moyennage  $\hat{I}(i_1, i_2) = \sum_{(k_1, k_2) \in [1, M]^2} \lambda_{i_1, i_2, k_1, k_2} Y_{k_1, k_2}$
- Les poids  $\lambda_{i_1, i_2, k_1, k_2}$  peuvent (vont) dépendre de la position et de  $Y$ .

## Noyau classique

- $\lambda_{i_1, i_2, k_1, k_2} = \frac{K(i_1 - k_1, i_2 - k_2)}{\sum_{k'_1, k'_2} K(i_1 - k'_1, i_2 - k'_2)}$  (pas de dépendance en  $Y$ ).
- Exemple : noyau gaussien  $K(i_1, i_2) = e^{-(i_1^2 + i_2^2)/2h^2}$ .
- Adaptation locale (taille, forme, ...) du noyau  $K$  (dépendance en  $Y$ ).

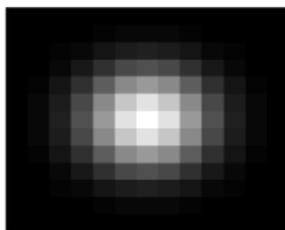
# Méthodes à noyaux

## Méthode à noyaux générique

- Estimer  $I(i_1, i_2)$  par moyennage  $\hat{I}(i_1, i_2) = \sum_{(k_1, k_2) \in [1, M]^2} \lambda_{i_1, i_2, k_1, k_2} Y_{k_1, k_2}$
- Les poids  $\lambda_{i_1, i_2, k_1, k_2}$  peuvent (vont) dépendre de la position et de  $Y$ .

## Noyau classique

- $\lambda_{i_1, i_2, k_1, k_2} = \frac{K(i_1 - k_1, i_2 - k_2)}{\sum_{k'_1, k'_2} K(i_1 - k'_1, i_2 - k'_2)}$  (pas de dépendance en  $Y$ ).



- Exemple : noyau gaussien  $K(i_1, i_2) = e^{-(i_1^2 + i_2^2)/2h^2}$ .
- Adaptation locale (taille, forme,...) du noyau  $K$  (dépendance en  $Y$ ).

# Noyaux dépendants des données

## Exemples

- Noyaux adaptatifs.
- Seuillage en  $\star$ -let (dépendance complexe des poids)
- Filtrage bilatéral (dépendant de différences pixel à pixel)

# Noyaux dépendants des données

## Exemples

- Noyaux adaptatifs.
- Seuillage en  $\star$ -let (dépendance complexe des poids)
- Filtrage bilatéral (dépendant de différences pixel à pixel)

## Filtrage bilatéral (Tomasi et al.)

- $$\lambda_{i_1, i_2, k_1, k_2} = \frac{K(i_1 - k_1, i_2 - k_2) \times K'(Y(i_1, i_2) - Y(k_1, k_2))}{\sum_{k'_1, k'_2} K(i_1 - k'_1, i_2 - k'_2) \times K'(Y(i_1, i_2) - Y(k'_1, k'_2))}$$
- Version gaussienne : 
$$\lambda_{i_1, i_2, k_1, k_2} = \frac{e^{-\frac{(i_1 - k_1)^2 + (i_2 - k_2)^2}{2h^2}} \times e^{-\frac{(Y(i_1, i_2) - Y(k_1, k_2))^2}{2h'^2}}}{\sum_{k'_1, k'_2} e^{-\frac{(i_1 - k'_1)^2 + (i_2 - k'_2)^2}{2h^2}} \times e^{-\frac{(Y(i_1, i_2) - Y(k'_1, k'_2))^2}{2h'^2}}}$$
  - Intuition : moyenner des pixels proches à la fois en distance et en valeur.
  - Problème : valeur des pixels = trop locale...

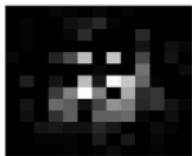
# Noyaux dépendants des données

## Exemples

- Noyaux adaptatifs.
- Seuillage en  $\star$ -let (dépendance complexe des poids)
- Filtrage bilatéral (dépendant de différences pixel à pixel)

## Filtrage bilatéral (Tomasi et al.)

- $$\lambda_{i_1, i_2, k_1, k_2} = \frac{K(i_1 - k_1, i_2 - k_2) \times K'(Y(i_1, i_2) - Y(k_1, k_2))}{\sum_{k'_1, k'_2} K(i_1 - k'_1, i_2 - k'_2) \times K'(Y(i_1, i_2) - Y(k'_1, k'_2))}$$
- Version gaussienne : 
$$\lambda_{i_1, i_2, k_1, k_2} = \frac{e^{-\frac{(i_1 - k_1)^2 + (i_2 - k_2)^2}{2h^2}} \times e^{-\frac{(Y(i_1, i_2) - Y(k_1, k_2))^2}{2h'^2}}}{\sum_{k'_1, k'_2} e^{-\frac{(i_1 - k'_1)^2 + (i_2 - k'_2)^2}{2h^2}} \times e^{-\frac{(Y(i_1, i_2) - Y(k'_1, k'_2))^2}{2h'^2}}}$$



- Intuition : moyenner des pixels proches à la fois en distance et en valeur.
- Problème : valeur des pixels = trop locale...

# Patches

## Patch

- Patch : version moins locale des valeurs de pixel.
- Patch centré  $P(I)(i_1, i_2)$  de largeur  $W$  :

$$P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2) \text{ avec } -\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$$

# Patches

## Patch

- Patch : version moins locale des valeurs de pixel.
- Patch centré  $P(I)(i_1, i_2)$  de largeur  $W$  :

$$P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2) \text{ avec } -\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$$

## Patch et images

- Opérateur  $I \mapsto P(I)$  envoie l'image  $I$  sur la collection de patches  $P(I)$ .
- Relèvement d'une image à valeur 1-D en une image à valeur  $W^2$ -D.
- Reprojection facile d'une collection de patches  $P(I)$  à une image  $I$ ...

# Patches

## Patch

- Patch : version moins locale des valeurs de pixel.
- Patch centré  $P(I)(i_1, i_2)$  de largeur  $W$  :

$$P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2) \text{ avec } -\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$$

## Patch et images

- Opérateur  $I \mapsto P(I)$  envoie l'image  $I$  sur la collection de patches  $P(I)$ .
- Relèvement d'une image à valeur 1-D en une image à valeur  $W^2$ -D.
- Reprojection facile d'une collection de patches  $P(I)$  à une image  $I$ ...

## Patch et estimation

- Problème d'estimation : Estimation de la collection de patches  $P(I)$  à partir de la collection de patches  $P(Y)$ .
- Problème très différent de l'estimation de  $P(I)$  à partir  $P(I) + \sigma W$ .

# Patches

## Patch

- Patch : version moins locale des valeurs de pixel.
- Patch centré  $P(I)(i_1, i_2)$  de largeur  $W$  :

$$P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2) \text{ avec } -\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$$

## Patch et images

- Opérateur  $I \mapsto P(I)$  envoie l'image  $I$  sur la collection de patches  $P(I)$ .
- Relèvement d'une image à valeur 1-D en une image à valeur  $W^2$ -D.
- Reprojection facile d'une collection de patches  $P(I)$  à une image  $I$ ...

## Patch et estimation

- Problème d'estimation : Estimation de la collection de patches  $P(I)$  à partir de la collection de patches  $P(Y)$ .
- Problème très différent de l'estimation de  $P(I)$  à partir  $P(I) + \sigma W$ .

# Méthodes à patches

## Méthodes à noyaux et à patches

- Estimation par moyennage de patches :

$$\widehat{P}(I)(i_1, i_2)(j_1, j_2) = \sum_k \lambda_{i_1, i_2, k_1, k_2} P(Y)(k_1, k_2)(j_1, j_2).$$

- Si poids indépendants de  $P(Y)$ , le passage par les patches est inutile...

# Méthodes à patches

## Méthodes à noyaux et à patches

- Estimation par moyennage de patches :

$$\widehat{P(I)}(i_1, i_2)(j_1, j_2) = \sum_k \lambda_{i_1, i_2, k_1, k_2} P(Y)(k_1, k_2)(j_1, j_2).$$

- Si poids indépendants de  $P(Y)$ , le passage par les patches est inutile...

## Intuition

- Utiliser des poids qui prennent en compte la similarité des patches :

Patches  $P(Y)(i_1, i_2) = P_{(i_1, i_2)}$  :

- Patch  $P(Y)(i_1, i_2)$  à débruiter,
- Patches similaires, utiles : poids importants,
- Patches moins similaires, moins utiles : poids faibles,
- Patches très différents, inutiles : poids quasi nuls.

# Méthodes à patches

## Méthodes à noyaux et à patches

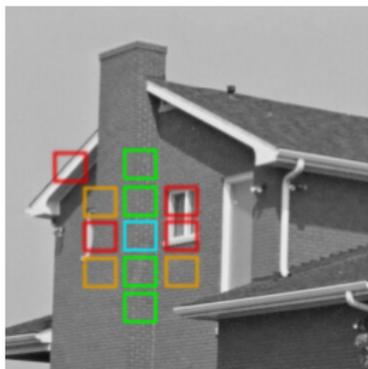
- Estimation par moyennage de patches :

$$\widehat{P}(I)(i_1, i_2)(j_1, j_2) = \sum_k \lambda_{i_1, i_2, k_1, k_2} P(Y)(k_1, k_2)(j_1, j_2).$$

- Si poids indépendants de  $P(Y)$ , le passage par les patches est inutile...

## Intuition

- Utiliser des poids qui prennent en compte la similarité des patches :



Patches  $P(Y)(i_1, i_2) = P_{(i_1, i_2)}$  :

- Patch  $P(Y)(i_1, i_2)$  à débruiter,
- Patches similaires, utiles : poids importants,
- Patches moins similaires, moins utiles : poids faibles,
- Patches très différents, inutiles : poids quasi nuls.

# NL-Means

## NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité  $D$  entre patches.
- Utiliser des poids  $\lambda_{i_1, i_2, k_1, k_2} = \frac{K'(D(P_{(i_1, i_2)}, P_{(k_1, k_2)}))}{\sum_{k'_1, k'_2} K'(D(P_{(i_1, i_2)}, P_{(k'_1, k'_2)}))}$
- Choisir  $D(P_{(i_1, i_2)}, P_{(k_1, k_2)}) = \|P_{(i_1, i_2)} - P_{(k_1, k_2)}\|$  pour mesurer la dissimilarité, un noyau gaussien  $K'(x) = \exp(-x^2/\beta)$  et une température  $\beta = \gamma\sigma^2$ .

# NL-Means

## NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité  $D$  entre patches.
- Utiliser des poids  $\lambda_{i_1, i_2, k_1, k_2} = \frac{K'(D(P_{(i_1, i_2)}, P_{(k_1, k_2)}))}{\sum_{k'_1, k'_2} K'(D(P_{(i_1, i_2)}, P_{(k'_1, k'_2)}))}$
- Choisir  $D(P_{(i_1, i_2)}, P_{(k_1, k_2)}) = \|P_{(i_1, i_2)} - P_{(k_1, k_2)}\|$  pour mesurer la dissimilarité, un noyau gaussien  $K'(x) = \exp(-x^2/\beta)$  et une température  $\beta = \gamma\sigma^2$ .

## Résultats

- Méthode rapide et efficace.
- Performance proche de l'état de l'art.

# NL-Means

## NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité  $D$  entre patches.
- Utiliser des poids  $\lambda_{i_1, i_2, k_1, k_2} = \frac{K'(D(P_{(i_1, i_2)}, P_{(k_1, k_2)}))}{\sum_{k'_1, k'_2} K'(D(P_{(i_1, i_2)}, P_{(k'_1, k'_2)}))}$
- Choisir  $D(P_{(i_1, i_2)}, P_{(k_1, k_2)}) = \|P_{(i_1, i_2)} - P_{(k_1, k_2)}\|$  pour mesurer la dissimilarité, un noyau gaussien  $K'(x) = \exp(-x^2/\beta)$  et une température  $\beta = \gamma\sigma^2$ .

## Résultats

- Méthode rapide et efficace.
- Performance proche de l'état de l'art.

## Variations

- Adaptation automatique la zone de recherche. (Kervrann et al.)
- Approximations locales d'ordre plus élevé. (Buadès et al.)
- Utilisation de mesures de dissimilarité différentes. (Guichard et al.)

# NL-Means

## NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité  $D$  entre patches.
- Utiliser des poids  $\lambda_{i_1, i_2, k_1, k_2} = \frac{K'(D(P_{(i_1, i_2)}, P_{(k_1, k_2)}))}{\sum_{k'_1, k'_2} K'(D(P_{(i_1, i_2)}, P_{(k'_1, k'_2)}))}$
- Choisir  $D(P_{(i_1, i_2)}, P_{(k_1, k_2)}) = \|P_{(i_1, i_2)} - P_{(k_1, k_2)}\|$  pour mesurer la dissimilarité, un noyau gaussien  $K'(x) = \exp(-x^2/\beta)$  et une température  $\beta = \gamma\sigma^2$ .

## Résultats

- Méthode rapide et efficace.
- Performance proche de l'état de l'art.

## Variations

- Adaptation automatique la zone de recherche. (Kervrann et al.)
- Approximations locales d'ordre plus élevé. (Buadès et al.)
- Utilisation de mesures de dissimilarité différentes. (Guichard et al.)

# Interprétation(s) des NL-Means

## NL-Means

- $\widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k'_1, k'_2)\|^2}} P_{(k_1, k_2)}(i_1, i_2).$

# Interprétation(s) des NL-Means

## NL-Means

- $$\widehat{P(I)}(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k'_1, k'_2)\|^2}} P_{(k_1, k_2)}(i_1, i_2).$$

## Processus stationnaire

- Résultat de consistance pour des processus  $\beta$ -mélangeants.

# Interprétation(s) des NL-Means

## NL-Means

$$\bullet \widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k'_1, k'_2)\|^2}} P_{(k_1, k_2)}(i_1, i_2).$$

## Processus stationnaire

- Résultat de consistance pour des processus  $\beta$ -mélangeants.

## Diffusion / Lissage sur des variétés

- Explication intuitive mais les preuves demandent des hypothèses très fortes...
- Variante sur des graphes.

# Interprétation(s) des NL-Means

## NL-Means

- $$\widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k'_1, k'_2)\|^2}} P_{(k_1, k_2)}(i_1, i_2).$$

## Processus stationnaire

- Résultat de consistance pour des processus  $\beta$ -mélangeants.

## Diffusion / Lissage sur des variétés

- Explication intuitive mais les preuves demandent des hypothèses très fortes...
- Variante sur des graphes.

## Modèle pseudo bayésien

- Modèle d'observation (faux) :  $P(Y)(i_1, i_2) = P(I)(i_1, i_2) + \sqrt{\beta/2}W.$
- Loi a priori uniforme sur les  $P(Y)(k_1, k_2)$ ...

# Interprétation(s) des NL-Means

## NL-Means

$$\bullet \widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \|P(Y)(i_1, i_2) - P(Y)(k'_1, k'_2)\|^2}} P_{(k_1, k_2)}(i_1, i_2).$$

## Processus stationnaire

- Résultat de consistance pour des processus  $\beta$ -mélangeants.

## Diffusion / Lissage sur des variétés

- Explication intuitive mais les preuves demandent des hypothèses très fortes...
- Variante sur des graphes.

## Modèle pseudo bayésien

- Modèle d'observation (faux) :  $P(Y)(i_1, i_2) = P(I)(i_1, i_2) + \sqrt{\beta/2}W$ .
- Loi a priori uniforme sur les  $P(Y)(k_1, k_2)$ ...

# NL-Means et meilleur noyau

Un noyau local optimisé

- NL-Means produit un noyau local qui s'adapte à la géométrie locale.

# NL-Means et meilleur noyau

## Un noyau local optimisé



- NL-Means produit un noyau local qui s'adapte à la géométrie locale.

## Un noyau local optimal ?

- Peut-on comparer les NL-Means au meilleur noyau local :

$$E(\|I - \hat{I}\|^2) \leq C \arg \min_{\lambda} \underbrace{\sum_{i_1, i_2} |I(i_1, i_2) - \sum_{k_1, k_2} \lambda_{i_1 - k_1, i_2 - k_2} I(k_1, k_2)|^2}_{\text{biais}} + \underbrace{N^2 \sigma^2 \|\lambda\|^2}_{\text{variance}} ?$$

# NL-Means et meilleur noyau

## Un noyau local optimisé



- NL-Means produit un noyau local qui s'adapte à la géométrie locale.

## Un noyau local optimal ?

- Peut-on comparer les NL-Means au meilleur noyau local :

$$E(\|I - \hat{I}\|^2) \leq C \arg \min_{\lambda} \underbrace{\sum_{i_1, i_2} |I(i_1, i_2) - \sum_{k_1, k_2} \lambda_{i_1 - k_1, i_2 - k_2} I(k_1, k_2)|^2}_{\text{biais}} + \underbrace{N^2 \sigma^2 \|\lambda\|^2}_{\text{variance}} ?$$

# Plan

- 1 Méthodes à noyaux et NL-Means
- 2 Agrégation d'estimateurs
  - Estimateurs initiaux et agrégation
  - Agrégation PAC-Bayésienne
  - Esquisse de preuve
- 3 Agrégation de patches

# Estimateurs initiaux et agrégation

## Modèle et estimateurs initiaux

- $Y = I + \sigma W$  de taille  $W \times W$ .
- $\{P_k\}$  collections de  $M$  estimateur initiaux de  $I$ .

# Estimateurs initiaux et agrégation

## Modèle et estimateurs initiaux

- $Y = I + \sigma W$  de taille  $W \times W$ .
- $\{P_k\}$  collections de  $M$  estimateur initiaux de  $I$ .

## Agrégation

- Estimer  $I$  comme une combinaison linéaire :  $\hat{I} = P_\lambda = \sum_k \lambda_k P_k$ .
- Procédure d'agrégation : manière de choisir  $\lambda$  à partir de  $Y$ .

# Estimateurs initiaux et agrégation

## Modèle et estimateurs initiaux

- $Y = I + \sigma W$  de taille  $W \times W$ .
- $\{P_k\}$  collections de  $M$  estimateur initiaux de  $I$ .

## Agrégation

- Estimer  $I$  comme une combinaison linéaire :  $\hat{I} = P_\lambda = \sum_k \lambda_k P_k$ .
- Procédure d'agrégation : manière de choisir  $\lambda$  à partir de  $Y$ .

## Inégalités de type oracle

- Résultat typique : agrégation « optimal » sur une classe  $\Lambda$ ,

$$\mathbb{E} \left( \|I - \hat{I}\|^2 \right) \leq C \inf_{\lambda \in \Lambda} \mathbb{E} \left( \|I - P_\lambda\|^2 \right) + \sigma^2 \text{pen}(\lambda)$$

- $C$ ,  $\Lambda$  et  $\text{pen}$  dépendent de la procédure.

# Estimateurs initiaux et agrégation

## Modèle et estimateurs initiaux

- $Y = I + \sigma W$  de taille  $W \times W$ .
- $\{P_k\}$  collections de  $M$  estimateur initiaux de  $I$ .

## Agrégation

- Estimer  $I$  comme une combinaison linéaire :  $\hat{I} = P_\lambda = \sum_k \lambda_k P_k$ .
- Procédure d'agrégation : manière de choisir  $\lambda$  à partir de  $Y$ .

## Inégalités de type oracle

- Résultat typique : agrégation « optimal » sur une classe  $\Lambda$ ,

$$\mathbb{E} \left( \|I - \hat{I}\|^2 \right) \leq C \inf_{\lambda \in \Lambda} \mathbb{E} \left( \|I - P_\lambda\|^2 \right) + \sigma^2 \text{pen}(\lambda)$$

- $C$ ,  $\Lambda$  et  $\text{pen}$  dépendent de la procédure.

# Procédures d'agrégation

## Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_0.$$

- Inégalité oracle si les  $P_k$  sont fixes ou obtenus par projection.
- Optimisation numérique très difficile...

# Procédures d'agrégation

## Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_0.$$

- Inégalité oracle si les  $P_k$  sont fixes ou obtenus par projection.
- Optimisation numérique très difficile...

## Pénalisation $\ell^1$

- Sélection par pénalisation proportionnelle à la norme  $\ell^1$  :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_1.$$

- Résultats pour des  $P_k$  fixes sous des hypothèses de structure fortes .

# Procédures d'agrégation

## Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_0.$$

- Inégalité oracle si les  $P_k$  sont fixes ou obtenus par projection.
- Optimisation numérique très difficile...

## Pénalisation $\ell^1$

- Sélection par pénalisation proportionnelle à la norme  $\ell^1$  :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_1.$$

- Résultats pour des  $P_k$  fixes sous des hypothèses de structure fortes .

## Poids exponentiels

- Mélange de solutions pondérées par des poids exponentiels :

$$\hat{\lambda} = \frac{1}{Z} \sum_k e^{-\frac{1}{\beta} \|Y - P_k\|^2} P_k \dots$$

# Procédures d'agrégation

## Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_0.$$

- Inégalité oracle si les  $P_k$  sont fixes ou obtenus par projection.
- Optimisation numérique très difficile...

## Pénalisation $\ell^1$

- Sélection par pénalisation proportionnelle à la norme  $\ell^1$  :

$$\hat{\lambda} = \arg \min_{\lambda} \|Y - P_{\lambda}\|^2 + \beta \|\lambda\|_1.$$

- Résultats pour des  $P_k$  fixes sous des hypothèses de structure fortes .

## Poids exponentiels

- Mélange de solutions pondérées par des poids exponentiels :

$$\hat{\lambda} = \frac{1}{Z} \sum_k e^{-\frac{1}{\beta} \|Y - P_k\|^2} P_k \dots$$

# Agrégation PAC-Bayésienne

## Prérequis

- Estimateurs sans biais  $\hat{r}_\lambda$  de  $\mathbb{E}_W (\|I - P_\lambda\|^2)$  obtenus par Stein.
- Loi a priori  $\pi$  sur  $\Lambda \subset \mathbb{R}^M$ .

# Agrégation PAC-Bayésienne

## Prérequis

- Estimateurs sans biais  $\hat{r}_\lambda$  de  $\mathbb{E}_W (\|I - P_\lambda\|^2)$  obtenus par Stein.
- Loi a priori  $\pi$  sur  $\Lambda \subset \mathbb{R}^M$ .

## Agrégé PAC-Bayésien

- Procédure spécifique reposant sur des poids exponentiels.
- Estimateur dépend de  $\pi$  et d'une température  $\beta$  :  $\hat{I} = P_{\lambda_\pi}$  avec

$$\lambda_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} \lambda d\pi(\lambda) \quad .$$

- Si la loi a priori est de la forme  $\pi = \sum_k \pi_k$  :

$$\hat{I} = \sum_k \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda} \pi_k}{\sum_{k'} e^{-\frac{1}{\beta} \hat{r}_{k'}} \pi_{k'}} P_k \quad .$$

# Agrégation PAC-Bayésienne

## Prérequis

- Estimateurs sans biais  $\hat{r}_\lambda$  de  $\mathbb{E}_W (\|I - P_\lambda\|^2)$  obtenus par Stein.
- Loi a priori  $\pi$  sur  $\Lambda \subset \mathbb{R}^M$ .

## Agrégé PAC-Bayésien

- Procédure spécifique reposant sur des poids exponentiels.
- Estimateur dépend de  $\pi$  et d'une température  $\beta$  :  $\hat{I} = P_{\lambda_\pi}$  avec

$$\lambda_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} \lambda d\pi(\lambda) \quad .$$

- Si la loi a priori est de la forme  $\pi = \sum_k \pi_k$  :

$$\hat{I} = \sum_k \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda} \pi_k}{\sum_{k'} e^{-\frac{1}{\beta} \hat{r}_{k'}} \pi_{k'}} P_k \quad .$$

# Résultats théoriques

## Hypothèses sur $P_k$

- Pas de résultats généraux...
- Estimateurs initiaux obtenues par projections (Barron, Leung).
- Estimateurs initiaux indépendants de  $Y$  (Dalalyan, Tsybakov).

# Résultats théoriques

## Hypothèses sur $P_k$

- Pas de résultats généraux...
- Estimateurs initiaux obtenues par projections (Barron, Leung).
- Estimateurs initiaux indépendants de  $Y$  (Dalalyan, Tsybakov).

## Inégalité oracle

- Inégalité oracle « précise » : si  $\beta \geq 4\sigma^2$ ,

$$\mathbb{E} \left( \|I - \hat{T}\|^2 \right) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \mathbb{E} \left( \|I - P_\lambda\|^2 \right) dp + \beta \mathcal{K}(p, \pi)$$

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

$$\mathcal{K}(p, \pi) = \begin{cases} \mathbb{E}_p \left( \log \left( \frac{dp}{d\pi} \right) \right) & \text{si } p \ll \pi \\ +\infty & \text{sinon} \end{cases}$$

# Résultats théoriques

## Hypothèses sur $P_k$

- Pas de résultats généraux...
- Estimateurs initiaux obtenues par projections (Barron, Leung).
- Estimateurs initiaux indépendants de  $Y$  (Dalalyan, Tsybakov).

## Inégalité oracle

- Inégalité oracle « précise » : si  $\beta \geq 4\sigma^2$ ,

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

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

$$\mathcal{K}(p, \pi) = \begin{cases} \mathbb{E}_p \left( \log \left( \frac{dp}{d\pi} \right) \right) & \text{si } p \ll \pi \\ +\infty & \text{sinon} \end{cases} .$$

# Choix de la loi a priori

## Borne d'erreur et a priori

- $\mathbb{E} \left( \|I - \widehat{I}\|^2 \right) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \mathbb{E} \left( \|I - P_\lambda\|^2 \right) dp + \beta \mathcal{K}(p, \pi)$
- Compromis entre un  $p$  concentré autour du  $\lambda$  du meilleur agrégé  $P_\lambda$  et un  $p$  proche de  $\pi$ .
- Choix de  $\pi$  de sorte que cette quantité soit « uniformément » petite...

# Choix de la loi a priori

## Borne d'erreur et a priori

- $\mathbb{E} (\|I - \hat{T}\|^2) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \mathbb{E} (\|I - P_\lambda\|^2) dp + \beta \mathcal{K}(p, \pi)$
- Compromis entre un  $p$  concentré autour du  $\lambda$  du meilleur agrégé  $P_\lambda$  et un  $p$  proche de  $\pi$ .
- Choix de  $\pi$  de sorte que cette quantité soit « uniformément » petite...

## Loi a priori discrète

- $\pi = \sum_k \delta_k : \mathbb{E} (\|I - \hat{T}\|^2) \leq \inf_k \mathbb{E} (\|I - P_k\|^2) - \beta \log \delta_k$  .
- Mieux que le meilleur estimateur initial pénalisé : sélection de modèle.

# Choix de la loi a priori

## Borne d'erreur et a priori

- $\mathbb{E} (\|I - \hat{T}\|^2) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \mathbb{E} (\|I - P_\lambda\|^2) dp + \beta \mathcal{K}(p, \pi)$
- Compromis entre un  $p$  concentré autour du  $\lambda$  du meilleur agrégé  $P_\lambda$  et un  $p$  proche de  $\pi$ .
- Choix de  $\pi$  de sorte que cette quantité soit « uniformément » petite...

## Loi a priori discrète

- $\pi = \sum_k \delta_k : \mathbb{E} (\|I - \hat{T}\|^2) \leq \inf_k \mathbb{E} (\|I - P_k\|^2) - \beta \log \delta_k$  .
- Mieux que le meilleur estimateur initial pénalisé : sélection de modèle.

## Loi a priori parcimonieuse

- $\pi$  : i.i.d. Student ou mélange de gaussienne (Dalalyan et al.).
- Borne :  $\mathbb{E} (\|I - \hat{T}\|^2) \leq \inf_\lambda \mathbb{E} (\|I - P_\lambda\|^2) + C\beta \|\lambda\|_0 \log M$  .
- Mieux que la meilleure agrégation « parcimonieuse »...

# Choix de la loi a priori

## Borne d'erreur et a priori

- $\mathbb{E}(\|I - \hat{T}\|^2) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \mathbb{E}(\|I - P_\lambda\|^2) dp + \beta \mathcal{K}(p, \pi)$
- Compromis entre un  $p$  concentré autour du  $\lambda$  du meilleur agrégé  $P_\lambda$  et un  $p$  proche de  $\pi$ .
- Choix de  $\pi$  de sorte que cette quantité soit « uniformément » petite...

## Loi a priori discrète

- $\pi = \sum_k \delta_k : \mathbb{E}(\|I - \hat{T}\|^2) \leq \inf_k \mathbb{E}(\|I - P_k\|^2) - \beta \log \delta_k$  .
- Mieux que le meilleur estimateur initial pénalisé : sélection de modèle.

## Loi a priori parcimonieuse

- $\pi$  : i.i.d. Student ou mélange de gaussienne (Dalalyan et al.).
- Borne :  $\mathbb{E}(\|I - \hat{T}\|^2) \leq \inf_\lambda \mathbb{E}(\|I - P_\lambda\|^2) + C\beta \|\lambda\|_0 \log M$  .
- Mieux que la meilleure agrégation « parcimonieuse »...

# Esquisse de preuve

## Estimateur PAC-Bayésien

- $\hat{I} = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} P_\lambda d\pi(\lambda)$  .
- Théorème : Si  $\beta \geq 4\sigma^2$ ,

$$\mathbb{E} \left( \|I - \hat{I}\|^2 \right) \leq \inf_P \int_{\lambda \in \mathbb{R}^M} \mathbb{E} \left( \|I - P_\lambda\|^2 \right) dP + \beta \mathcal{K}(P, \pi)$$

# Esquisse de preuve

## Estimateur PAC-Bayésien

- $\hat{I} = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} P_\lambda d\pi(\lambda) \quad .$
- Théorème : Si  $\beta \geq 4\sigma^2$ ,

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

## Outils

- Estimateur sans biais du risque de Stein (SURE).
- Propriétés des mesures de Gibbs (exponentielles).
- Forme spécifique des poids.

# Esquisse de preuve

## Estimateur PAC-Bayésien

- $\hat{T} = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} P_\lambda d\pi(\lambda) \quad .$
- Théorème : Si  $\beta \geq 4\sigma^2$ ,

$$\mathbb{E} \left( \|I - \hat{T}\|^2 \right) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \mathbb{E} \left( \|I - P_\lambda\|^2 \right) dp + \beta \mathcal{K}(p, \pi)$$

## Outils

- Estimateur sans biais du risque de Stein (SURE).
- Propriétés des mesures de Gibbs (exponentielles).
- Forme spécifique des poids.

# Esquisse de preuve

## Stein

- Estimateur pondéré :  $\tilde{I} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} P_{\lambda} d\pi(\lambda)$

- Formule de Stein : estimateur sans biais  $\hat{r}$  du risque quadratique de  $\tilde{I}$ .

$$\hat{r} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} \left( \hat{r}_{\lambda} - \|P_{\lambda} - \tilde{I}\|^2 - 2\sigma^2 \sum_{i=1}^{w^2} \nabla_i \rho(\lambda) (P_{\lambda} - \tilde{I})_i \right) d\pi(\lambda)$$

# Esquisse de preuve

## Stein

- Estimateur pondéré :  $\tilde{T} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} P_{\lambda} d\pi(\lambda)$
- Formule de Stein : estimateur sans biais  $\hat{r}$  du risque quadratique de  $\tilde{T}$ .

$$\hat{r} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} \left( \hat{r}_{\lambda} - \|P_{\lambda} - \tilde{T}\|^2 - 2\sigma^2 \sum_{i=1}^W \nabla_i \rho(\lambda) (P_{\lambda} - \tilde{T})_i \right) d\pi(\lambda)$$

## Forme spécifique des poids

- Pour  $\rho(\lambda) = \frac{1}{\beta} \hat{r}_{\lambda}$  et  $\nabla \hat{r}_{\lambda} = -2P_{\lambda} + C$  :

$$\hat{r} = \int_{\lambda} \frac{e^{-\frac{1}{\beta} \hat{r}_{\lambda}}}{\int_{\lambda} e^{-\frac{1}{\beta} \hat{r}_{\lambda}} d\pi(\lambda)} \left( \hat{r}_{\lambda} - \left(1 - \frac{4\sigma^2}{\beta}\right) \|P_{\lambda} - \tilde{T}\|^2 \right) d\pi(\lambda)$$

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \int_{\lambda} \frac{e^{-\frac{1}{\beta} \hat{r}_{\lambda}}}{\int_{\lambda} e^{-\frac{1}{\beta} \hat{r}_{\lambda}} d\pi(\lambda)} \hat{r}_{\lambda} d\pi(\lambda)$ .

# Esquisse de preuve

## Stein

- Estimateur pondéré :  $\tilde{T} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} P_{\lambda} d\pi(\lambda)$
- Formule de Stein : estimateur sans biais  $\hat{r}$  du risque quadratique de  $\tilde{T}$ .

$$\hat{r} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} \left( \hat{r}_{\lambda} - \|P_{\lambda} - \tilde{T}\|^2 - 2\sigma^2 \sum_{i=1}^W \nabla_i \rho(\lambda) (P_{\lambda} - \tilde{T})_i \right) d\pi(\lambda)$$

## Forme spécifique des poids

- Pour  $\rho(\lambda) = \frac{1}{\beta} \hat{r}_{\lambda}$  et  $\nabla \hat{r}_{\lambda} = -2P_{\lambda} + C$  :

$$\hat{r} = \int_{\lambda} \frac{e^{-\frac{1}{\beta} \hat{r}_{\lambda}}}{\int_{\lambda} e^{-\frac{1}{\beta} \hat{r}_{\lambda}} d\pi(\lambda)} \left( \hat{r}_{\lambda} - \left(1 - \frac{4\sigma^2}{\beta}\right) \|P_{\lambda} - \tilde{T}\|^2 \right) d\pi(\lambda)$$

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \int_{\lambda} \frac{e^{-\frac{1}{\beta} \hat{r}_{\lambda}}}{\int_{\lambda} e^{-\frac{1}{\beta} \hat{r}_{\lambda}} d\pi(\lambda)} \hat{r}_{\lambda} d\pi(\lambda)$ .

# Esquisse de preuve

Forme spécifique des poids

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \int_{\lambda} \frac{e^{-\frac{1}{\beta}\hat{r}_{\lambda}}}{\int_{\lambda'} e^{-\frac{1}{\beta}\hat{r}_{\lambda'}} d\pi(\lambda')} \hat{r}_{\lambda} d\pi(\lambda)$ .
- Passage à l'espérance impossible !

# Esquisse de preuve

## Forme spécifique des poids

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \int_{\lambda} \frac{e^{-\frac{1}{\beta}\hat{r}_{\lambda}}}{\int_{\lambda'} e^{-\frac{1}{\beta}\hat{r}_{\lambda'}} d\pi(\lambda')} \hat{r}_{\lambda} d\pi(\lambda)$ .
- Passage à l'espérance impossible !

## Mesure de Gibbs

- Inégalité :  
$$\int_{\lambda} \rho(\lambda) \frac{e^{-\frac{1}{\beta}\rho(\lambda)}}{\int_{\lambda'} e^{-\frac{1}{\beta}\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda) \leq \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \beta\mathcal{K}(p, \pi).$$

# Esquisse de preuve

## Forme spécifique des poids

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \int_{\lambda} \frac{e^{-\frac{1}{\beta}\hat{r}_{\lambda}}}{\int_{\lambda'} e^{-\frac{1}{\beta}\hat{r}_{\lambda'}} d\pi(\lambda')} \hat{r}_{\lambda} d\pi(\lambda)$ .
- Passage à l'espérance impossible !

## Mesure de Gibbs

- Inégalité :  
$$\int_{\lambda} \rho(\lambda) \frac{e^{-\frac{1}{\beta}\rho(\lambda)}}{\int_{\lambda'} e^{-\frac{1}{\beta}\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda) \leq \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \beta\mathcal{K}(p, \pi).$$

## Passage à l'espérance

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \min_p \int_{\lambda} \hat{r}_{\lambda} dp(\lambda) + \beta\mathcal{K}(p, \pi)$ .
- Passage à l'espérance possible !

$$\mathbb{E}(\hat{r}) \leq \min_p \int_{\lambda} \mathbb{E}(\hat{r}_{\lambda}) dp(\lambda) + \beta\mathcal{K}(p, \pi).$$

# Esquisse de preuve

## Forme spécifique des poids

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \int_{\lambda} \frac{e^{-\frac{1}{\beta}\hat{r}_{\lambda}}}{\int_{\lambda'} e^{-\frac{1}{\beta}\hat{r}_{\lambda'}} d\pi(\lambda')} \hat{r}_{\lambda} d\pi(\lambda)$ .
- Passage à l'espérance impossible !

## Mesure de Gibbs

- Inégalité :  
$$\int_{\lambda} \rho(\lambda) \frac{e^{-\frac{1}{\beta}\rho(\lambda)}}{\int_{\lambda'} e^{-\frac{1}{\beta}\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda) \leq \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \beta\mathcal{K}(p, \pi).$$

## Passage à l'espérance

- Si  $\beta \geq 4\sigma^2$ ,  $\hat{r} \leq \min_p \int_{\lambda} \hat{r}_{\lambda} dp(\lambda) + \beta\mathcal{K}(p, \pi)$ .
- Passage à l'espérance possible !

$$\mathbb{E}(\hat{r}) \leq \min_p \int_{\lambda} \mathbb{E}(\hat{r}_{\lambda}) dp(\lambda) + \beta\mathcal{K}(p, \pi).$$

# Stein

## Formule de Stein

- Estimateur  $h(Y)$  : estimateur  $\hat{\tau}$  sans biais du risque donné par

$$\hat{\tau} = \|h(Y) - Y\|^2 + \sigma^2 \sum_{i=1}^{W^2} (2\nabla_i h_i(Y) - 1).$$

# Stein

## Formule de Stein

- Estimateur  $h(Y)$  : estimateur  $\hat{r}$  sans biais du risque donné par

$$\hat{r} = \|h(Y) - Y\|^2 + \sigma^2 \sum_{i=1}^{W^2} (2\nabla_i h_i(Y) - 1).$$

## Application

- Application à  $\tilde{T} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda} e^{-\rho(\lambda)} d\pi(\lambda)} P_{\lambda} d\pi(\lambda) = \int_{\lambda} w(\lambda) P_{\lambda} d\pi(\lambda)$ 

$$\hat{r} = \|\tilde{T} - Y\|^2 + \sigma^2 \sum_{i=1}^{W^2} \left( 2 \int (w(\lambda) \nabla_i P_{\lambda,i} + \nabla_i w(\lambda) P_{\lambda,i}) d\pi(\lambda) - 1 \right)$$
- $\|\tilde{T} - Y\|^2 = \int w(\lambda) \left( \|P_{\lambda} - Y\|^2 - \|P_{\lambda} - \tilde{T}\|^2 \right) d\pi(\lambda)$  et expression de  $\hat{r}_{\lambda}$  :
 
$$\hat{r} = \int \left( w(\lambda) (\hat{r}_{\lambda} - \|P_{\lambda} - \tilde{T}\|^2) + 2\sigma^2 \sum_{i=1}^{W^2} \nabla_i w(\lambda) P_{\lambda,i} \right) d\pi(\lambda)$$
- $\int \nabla_i w(\lambda) d\pi(\lambda) = 0$  et  $\nabla_i w(\lambda) = w(\lambda) \nabla_i \log w(\lambda)$  :
 
$$\hat{r} = \int w(\lambda) \left( \hat{r}_{\lambda} - \|P_{\lambda} - \tilde{T}\|^2 - 2\sigma^2 \sum_{i=1}^{W^2} \nabla_i \rho(\lambda) (P_{\lambda} - \tilde{T})_i \right) d\pi(\lambda)$$

# Stein

## Formule de Stein

- Estimateur  $h(Y)$  : estimateur  $\hat{r}$  sans biais du risque donné par

$$\hat{r} = \|h(Y) - Y\|^2 + \sigma^2 \sum_{i=1}^{W^2} (2\nabla_i h_i(Y) - 1).$$

## Application

- Application à  $\tilde{T} = \int_{\lambda} \frac{e^{-\rho(\lambda)}}{\int_{\lambda} e^{-\rho(\lambda)} d\pi(\lambda)} P_{\lambda} d\pi(\lambda) = \int_{\lambda} w(\lambda) P_{\lambda} d\pi(\lambda)$   
$$\hat{r} = \|\tilde{T} - Y\|^2 + \sigma^2 \sum_{i=1}^{W^2} \left( 2 \int (w(\lambda) \nabla_i P_{\lambda,i} + \nabla_i w(\lambda) P_{\lambda,i}) d\pi(\lambda) - 1 \right)$$
- $\|\tilde{T} - Y\|^2 = \int w(\lambda) \left( \|P_{\lambda} - Y\|^2 - \|P_{\lambda} - \tilde{T}\|^2 \right) d\pi(\lambda)$  et expression de  $\hat{r}_{\lambda}$  :  
$$\hat{r} = \int \left( w(\lambda) (\hat{r}_{\lambda} - \|P_{\lambda} - \tilde{T}\|^2) + 2\sigma^2 \sum_{i=1}^{W^2} \nabla_i w(\lambda) P_{\lambda,i} \right) d\pi(\lambda)$$
- $\int \nabla_i w(\lambda) d\pi(\lambda) = 0$  et  $\nabla_i w(\lambda) = w(\lambda) \nabla_i \log w(\lambda)$  :  
$$\hat{r} = \int w(\lambda) (\hat{r}_{\lambda} - \|P_{\lambda} - \tilde{T}\|^2 - 2\sigma^2 \sum_{i=1}^{W^2} \nabla_i \rho(\lambda) (P_{\lambda} - \tilde{T})_i) d\pi(\lambda)$$

# Gibbs

## Mesure de Gibbs

- Loi a priori  $\pi$ , fonction  $\rho$  :  $d\pi_{\exp(-\rho)}(\lambda) = \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda)$ .
- Propriété :  $\pi_{\exp(-\rho)} = \arg \min_{\rho} \int_{\lambda} \rho(\lambda) d\rho(\lambda) + \mathcal{K}(\rho, \pi)$ .
- Cor. :  $\int_{\lambda} \rho(\lambda) \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda) \leq \min_{\rho} \int_{\lambda} \rho(\lambda) d\rho(\lambda) + \mathcal{K}(\rho, \pi)$ .

# Gibbs

## Mesure de Gibbs

- Loi a priori  $\pi$ , fonction  $\rho$  :  $d\pi_{\exp(-\rho)}(\lambda) = \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda)$ .
- Propriété :  $\pi_{\exp(-\rho)} = \arg \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \mathcal{K}(p, \pi)$ .
- Cor. :  $\int_{\lambda} \rho(\lambda) \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda) \leq \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \mathcal{K}(p, \pi)$ .

## Preuve

- Soit  $dp = f_p d\pi$ ,

$$\begin{aligned}\mathcal{K}(p, \pi_{\exp(-\rho)}) &= \int_{\lambda} \log \left( \frac{f_p(\lambda) \int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')}{e^{-\rho(\lambda)}} \right) dp(\lambda) \\ &= \int_{\lambda} \left( \rho(\lambda) + \log(f_p(\lambda)) + \log \left( \int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda') \right) \right) dp(\lambda) \\ &= \int_{\lambda} \rho(\lambda) dp(\lambda) + \mathcal{K}(p, \pi) + \log \left( \int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda') \right)\end{aligned}$$

# Gibbs

## Mesure de Gibbs

- Loi a priori  $\pi$ , fonction  $\rho$  :  $d\pi_{\exp(-\rho)}(\lambda) = \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda)$ .
- Propriété :  $\pi_{\exp(-\rho)} = \arg \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \mathcal{K}(p, \pi)$ .
- Cor. :  $\int_{\lambda} \rho(\lambda) \frac{e^{-\rho(\lambda)}}{\int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')} d\pi(\lambda) \leq \min_p \int_{\lambda} \rho(\lambda) dp(\lambda) + \mathcal{K}(p, \pi)$ .

## Preuve

- Soit  $dp = f_p d\pi$ ,

$$\begin{aligned}\mathcal{K}(p, \pi_{\exp(-\rho)}) &= \int_{\lambda} \log \left( \frac{f_p(\lambda) \int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda')}{e^{-\rho(\lambda)}} \right) dp(\lambda) \\ &= \int_{\lambda} \left( \rho(\lambda) + \log(f_p(\lambda)) + \log \left( \int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda') \right) \right) dp(\lambda) \\ &= \int_{\lambda} \rho(\lambda) dp(\lambda) + \mathcal{K}(p, \pi) + \log \left( \int_{\lambda'} e^{-\rho(\lambda')} d\pi(\lambda') \right)\end{aligned}$$

# Forme spécifique

## Estimateurs préliminaires indépendants de $Y$

- $P_k$  indépendant de  $Y$  et  $P_\lambda = \sum_k \lambda_k P_k$ .
- Stein :  $\hat{r}_\lambda = \|Y - P_\lambda\|^2 - \sigma^2 W^2$ .
- $\nabla \hat{r}_\lambda = 2(Y - P_\lambda) = -2P_\lambda + 2Y$ .
- Aucune restriction sur la loi a priori  $\pi$ .

# Forme spécifique

## Estimateurs préliminaires indépendants de $Y$

- $P_k$  indépendant de  $Y$  et  $P_\lambda = \sum_k \lambda_k P_k$ .
- Stein :  $\hat{r}_\lambda = \|Y - P_\lambda\|^2 - \sigma^2 W^2$ .
- $\nabla \hat{r}_\lambda = 2(Y - P_\lambda) = -2P_\lambda + 2Y$ .
- Aucune restriction sur la loi a priori  $\pi$ .

## Estimateurs initiaux par projections

- $P_k = P_{V_k}^\perp(Y)$  avec  $V_k$  espace de dimension  $D_k$  de b.o.n.  $X^1, \dots, X^{D_k}$ .
- $P_{V_k}^\perp(Y) = \sum_{d=1}^{D_k} \langle Y, X^d \rangle X^d \implies \sum_i \nabla_i (P_{V_k}^\perp(Y))_i = \sum_i \sum_{d=1}^{D_k} |X_i^d|^2 = D_k$ .
- $P_\lambda = P_\lambda(Y) = \sum_k \lambda_k P_{V_k}^\perp(Y)$ .
- Stein :  $\hat{r}_\lambda = \|Y - P_\lambda\|^2 + \sigma^2 (2 \sum_k \lambda_k D_k - W^2)$ .
- Gradient :  $\nabla \hat{r}_\lambda = 2(Y - 2P_\lambda(Y) + (P_\lambda^* P_\lambda)Y)$ .
- $\nabla \hat{r}_\lambda = -2P_\lambda + C$  si  $\lambda_k \lambda_{k'} = \pm \delta_{k,k'} \delta_{V_k \perp V_{k'}} \dots$
- Hypothèse sur le support de  $\pi$  satisfaite par la loi a priori discrète...
- Extension :  $\exists C, \delta \ (-\nabla \hat{r}_\lambda + C)^* (P_\lambda - \hat{T}) \leq 2\delta \|P_\lambda - \hat{T}\|^2 \implies \text{OK si } \beta \geq 4\delta\sigma^2 \dots$

# Forme spécifique

## Estimateurs préliminaires indépendants de $Y$

- $P_k$  indépendant de  $Y$  et  $P_\lambda = \sum_k \lambda_k P_k$ .
- Stein :  $\hat{r}_\lambda = \|Y - P_\lambda\|^2 - \sigma^2 W^2$ .
- $\nabla \hat{r}_\lambda = 2(Y - P_\lambda) = -2P_\lambda + 2Y$ .
- Aucune restriction sur la loi a priori  $\pi$ .

## Estimateurs initiaux par projections

- $P_k = P_{V_k}^\perp(Y)$  avec  $V_k$  espace de dimension  $D_k$  de b.o.n.  $X^1, \dots, X^{D_k}$ .
- $P_{V_k}^\perp(Y) = \sum_{d=1}^{D_k} \langle Y, X^d \rangle X^d \implies \sum_i \nabla_i (P_{V_k}^\perp(Y))_i = \sum_i \sum_{d=1}^{D_k} |X_i^d|^2 = D_k$ .
- $P_\lambda = P_\lambda(Y) = \sum_k \lambda_k P_{V_k}^\perp(Y)$ .
- Stein :  $\hat{r}_\lambda = \|Y - P_\lambda\|^2 + \sigma^2 (2 \sum_k \lambda_k D_k - W^2)$ .
- Gradient :  $\nabla \hat{r}_\lambda = 2(Y - 2P_\lambda(Y) + (P_\lambda^* P_\lambda)Y)$ .
- $\nabla \hat{r}_\lambda = -2P_\lambda + C$  si  $\lambda_k \lambda_{k'} = \pm \delta_{k,k'} \delta_{V_k \perp V_{k'}}$ ...
- Hypothèse sur le support de  $\pi$  satisfaite par la loi a priori discrète...
- Extension :  $\exists C, \delta (-\nabla \hat{r}_\lambda + C)^*(P_\lambda - \hat{T}) \leq 2\delta \|P_\lambda - \hat{T}\|^2 \implies \text{OK si } \beta \geq 4\delta\sigma^2 \dots$

# Plan

- 1 Méthodes à noyaux et NL-Means
- 2 Agrégation d'estimateurs
- 3 Agrégation de patches
  - Principe et résultats théoriques
  - Estimation PAC-Bayésienne numérique

# Agrégation de patches

## Patches, estimateurs agrégés, SURE et a priori

- Utiliser le patch  $P(Y)(i_1, i_2)$  comme observation et  $M$  patches  $P(Y)(k_1, k_2)$  comme estimateurs primaires.
  - Estimateurs agrégés :  $P_\lambda(i_1, i_2) = \sum_{k_1, k_2} \lambda_{i_1, i_2, k_1, k_2} P(Y)(k_1, k_2)$
  - Estimateur sans biais  $\hat{r}_\lambda(i_1, i_2)$  du risque (SURE)  
( $\mathbb{E}(\hat{r}_\lambda(i_1, i_2)) = \mathbb{E}(\|P(I)(i_1, i_2) - P_\lambda(i_1, i_2)\|^2)$ ) :
- $$\hat{r}_\lambda(i_1, i_2) = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_{(i_1, i_2)})\sigma^2.$$
- Choix d'une loi a priori  $\pi$  et d'une température  $\beta$ .

# Agrégation de patches

## Patches, estimateurs agrégés, SURE et a priori

- Utiliser le patch  $P(Y)(i_1, i_2)$  comme observation et  $M$  patches  $P(Y)(k_1, k_2)$  comme estimateurs primaires.
  - Estimateurs agrégés :  $P_\lambda(i_1, i_2) = \sum_{k_1, k_2} \lambda_{i_1, i_2, k_1, k_2} P(Y)(k_1, k_2)$
  - Estimateur sans biais  $\hat{r}_\lambda(i_1, i_2)$  du risque (SURE)  
( $\mathbb{E}(\hat{r}_\lambda(i_1, i_2)) = \mathbb{E}(\|P(I)(i_1, i_2) - P_\lambda(i_1, i_2)\|^2)$ ) :
- $$\hat{r}_\lambda(i_1, i_2) = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_{(i_1, i_2)})\sigma^2.$$
- Choix d'une loi a priori  $\pi$  et d'une température  $\beta$ .

## Estimateur PAC-Bayésien à patches

- $\widehat{P(I)}(i_1, i_2) = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda(i_1, i_2)}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}(i_1, i_2)} d\pi(\lambda')} P_\lambda(i_1, i_2) d\pi(\lambda)$  .

# Agrégation de patches

## Patches, estimateurs agrégés, SURE et a priori

- Utiliser le patch  $P(Y)(i_1, i_2)$  comme observation et  $M$  patches  $P(Y)(k_1, k_2)$  comme estimateurs primaires.
  - Estimateurs agrégés :  $P_\lambda(i_1, i_2) = \sum_{k_1, k_2} \lambda_{i_1, i_2, k_1, k_2} P(Y)(k_1, k_2)$
  - Estimateur sans biais  $\hat{r}_\lambda(i_1, i_2)$  du risque (SURE)  
( $\mathbb{E}(\hat{r}_\lambda(i_1, i_2)) = \mathbb{E}(\|P(I)(i_1, i_2) - P_\lambda(i_1, i_2)\|^2)$ ) :
- $$\hat{r}_\lambda(i_1, i_2) = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_{(i_1, i_2)})\sigma^2.$$
- Choix d'une loi a priori  $\pi$  et d'une température  $\beta$ .

## Estimateur PAC-Bayésien à patches

- $\widehat{P(I)}(i_1, i_2) = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda(i_1, i_2)}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}(i_1, i_2)} d\pi(\lambda')} P_\lambda(i_1, i_2) d\pi(\lambda)$  .

# Résultats théoriques

## Théorème

- Pour  $\beta \geq 4\sigma^2$ ,

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P}(I)(i_1, i_2)\|^2 \right) \\ & \leq \inf_p \int_{\lambda} \left( \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 \right) dp + \beta \mathcal{K}(p, \pi) \end{aligned}$$

# Résultats théoriques

## Théorème

- Pour  $\beta \geq 4\sigma^2$ ,

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P}(I)(i_1, i_2)\|^2 \right) \\ & \leq \inf_{\rho} \int_{\lambda} \left( \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2\sigma^2\|\lambda\|^2 \right) d\rho + \beta\mathcal{K}(\rho, \pi) \end{aligned}$$

## Travail en cours

- Difficultés techniques sur la partie spécifique aux poids...
- Résultat obtenu si on observe deux images bruitées ou avec un « splitting » des pixels...
- Travail en cours pour obtenir le résultat souhaité avec éventuellement une perte dans les constantes (Catoni, Audibert et Alquier) et/ou une modification des poids (Leung et Barron).

# Résultats théoriques

## Théorème ?

- Pour  $\beta \geq 4\sigma^2$ ,

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P}(I)(i_1, i_2)\|^2 \right) \\ & \leq \inf_{\rho} \int_{\lambda} \left( \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 \right) d\rho + \beta \mathcal{K}(\rho, \pi) \end{aligned}$$

## Travail en cours

- Difficultés techniques sur la partie spécifique aux poids...
- Résultat obtenu si on observe deux images bruitées ou avec un « splitting » des pixels...
- Travail en cours pour obtenir le résultat souhaité avec éventuellement une perte dans les constantes (Catoni, Audibert et Alquier) et/ou une modification des poids (Leung et Barron).

# Loi a priori « discrète »

## NL-Means

- Choix  $\pi = \sum_{(k_1, k_2)} \pi_{(k_1, k_2)} \delta_{(k_1, k_2)}$ .
- Agrégation NL-Means pondérées :

$$\widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \hat{r}_{(k_1, k_2)}(i_1, i_2)} \pi_{(k_1, k_2)}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \hat{r}_{(k'_1, k'_2)}(i_1, i_2)} \pi_{(k_1, k'_2)}} P_{(k_1, k_2)}(i_1, i_2).$$

# Loi a priori « discrète »

## NL-Means

- Choix  $\pi = \sum_{(k_1, k_2)} \pi_{(k_1, k_2)} \delta_{(k_1, k_2)}$ .
- Agrégation NL-Means pondérées :

$$\widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \hat{r}_{(k_1, k_2)}(i_1, i_2)} \pi_{(k_1, k_2)}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \hat{r}_{(k'_1, k'_2)}(i_1, i_2)} \pi_{(k_1, k'_2)}} P_{(k_1, k_2)}(i_1, i_2).$$

## Optimalité

- Résultat de type sélection :

$$\begin{aligned} \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P}(I)(i_1, i_2)\|^2 \right) \\ \leq \min_{(k_1, k_2)} \|P(I)(i_1, i_2) - P(I)(k_1, k_2)\|^2 + W^2 \sigma^2 - \beta \log \pi_{(k_1, k_2)} \\ \leq W^2 \sigma^2 - \beta \log \pi_{(i_1, i_2)} \end{aligned}$$

- (Presque) Mieux que rien...

# Loi a priori « discrète »

## NL-Means

- Choix  $\pi = \sum_{(k_1, k_2)} \pi_{(k_1, k_2)} \delta_{(k_1, k_2)}$ .
- Agrégation NL-Means pondérées :

$$\widehat{P}(I)(i_1, i_2) = \sum_{(k_1, k_2)} \frac{e^{-\frac{1}{\beta} \hat{r}_{(k_1, k_2)}(i_1, i_2)} \pi_{(k_1, k_2)}}{\sum_{(k'_1, k'_2)} e^{-\frac{1}{\beta} \hat{r}_{(k'_1, k'_2)}(i_1, i_2)} \pi_{(k_1, k'_2)}} P_{(k_1, k_2)}(i_1, i_2).$$

## Optimalité

- Résultat de type sélection :

$$\begin{aligned} \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P}(I)(i_1, i_2)\|^2 \right) \\ \leq \min_{(k_1, k_2)} \|P(I)(i_1, i_2) - P(I)(k_1, k_2)\|^2 + W^2 \sigma^2 - \beta \log \pi_{(k_1, k_2)} \\ \leq W^2 \sigma^2 - \beta \log \pi_{(i_1, i_2)} \end{aligned}$$

- (Presque) Mieux que rien...

# Le problème du poids central

## Rôle(s) du patch central

- $P(Y)(i_1, i_2)$  : patch de référence et patch utilisé pour l'estimateur.
- Quel poids pour ce patch central ?

$$\frac{\exp^{-\frac{1}{\beta} \rho_{(i_1, i_2)}(i_1, i_2)} \pi_{(i_1, i_2)}}{\sum_{(k'_1, k'_2)} \exp^{-\frac{1}{\beta} \rho_{(k'_1, k'_2)}(i_1, i_2)} \pi_{(k'_1, k'_2)}}$$

# Le problème du poids central

## Rôle(s) du patch central

- $P(Y)(i_1, i_2)$  : patch de référence et patch utilisé pour l'estimateur.
- Quel poids pour ce patch central ?

$$\frac{\exp^{-\frac{1}{\beta}\rho(i_1, i_2)} \pi(i_1, i_2)}{\sum_{(k'_1, k'_2)} \exp^{-\frac{1}{\beta}\rho(k'_1, k'_2)} \pi(k'_1, k'_2)}$$

## 4 propositions

- PAC-Bayésien / Stein : utiliser  $\hat{r}_\lambda$ ,  
 $\rho_{(k_1, k_2)}(i_1, i_2) = \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2 - W^2 \sigma^2$  et  
 $\rho_{(i_1, i_2)}(i_1, i_2) = W^2 \sigma^2$
- Buadès : remplacer  $\rho_{(i_1, i_2)}(i_1, i_2)$  par  $\max_{(k_1, k_2) \neq (i_1, i_2)} \rho_{(k_1, k_2)}(i_1, i_2)$ ,
- NL-Means « pur » :  $\rho_{(k_1, k_2)}(i_1, i_2) = \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2$ ,
- Zéro : ne pas utiliser le patch central...

# Le problème du poids central

## Rôle(s) du patch central

- $P(Y)(i_1, i_2)$  : patch de référence et patch utilisé pour l'estimateur.
- Quel poids pour ce patch central ?

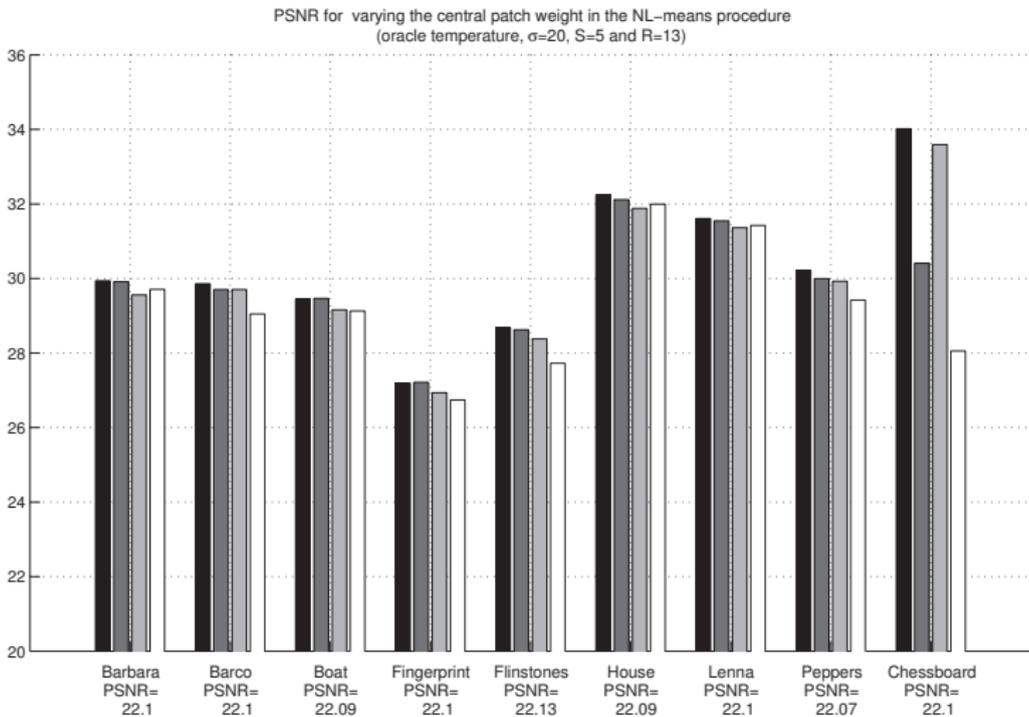
$$\frac{\exp^{-\frac{1}{\beta}\rho(i_1, i_2)} \pi(i_1, i_2)}{\sum_{(k'_1, k'_2)} \exp^{-\frac{1}{\beta}\rho(k'_1, k'_2)} \pi(k'_1, k'_2)}$$

## 4 propositions

- PAC-Bayésien / Stein : utiliser  $\hat{r}_\lambda$ ,  
 $\rho_{(k_1, k_2)}(i_1, i_2) = \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2 - W^2\sigma^2$  et  
 $\rho_{(i_1, i_2)}(i_1, i_2) = W^2\sigma^2$
- Buadès : remplacer  $\rho_{(i_1, i_2)}(i_1, i_2)$  par  $\max_{(k_1, k_2) \neq (i_1, i_2)} \rho_{(k_1, k_2)}(i_1, i_2)$ ,
- NL-Means « pur » :  $\rho_{(k_1, k_2)}(i_1, i_2) = \|P(Y)(i_1, i_2) - P(Y)(k_1, k_2)\|^2$ ,
- Zéro : ne pas utiliser le patch central...

# Résultats numériques

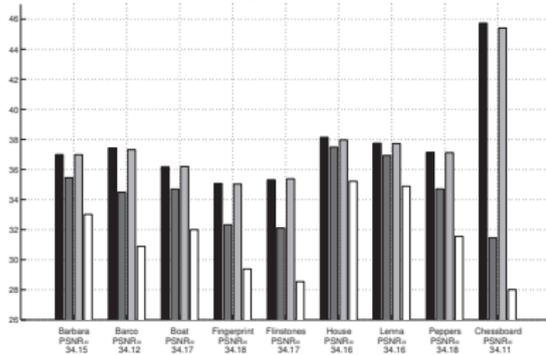
# Résultats numériques



$$\sigma = 20$$

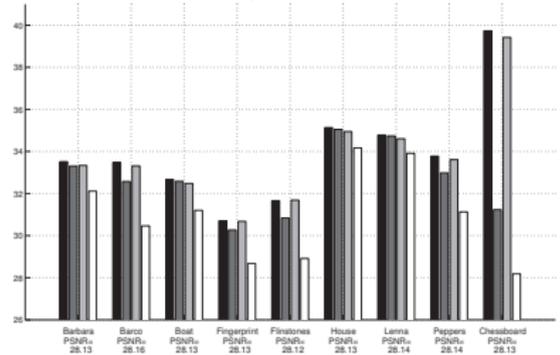
# Résultats numériques

PSNR for varying the central patch weight in the NL-means procedure  
(patch temperature,  $\sigma=5$ ,  $S=5$  and  $R=13$ )



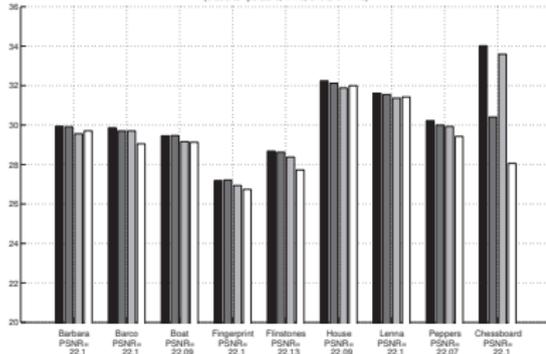
$\sigma = 5$

PSNR for varying the central patch weight in the NL-means procedure  
(patch temperature,  $\sigma=10$ ,  $S=5$  and  $R=13$ )



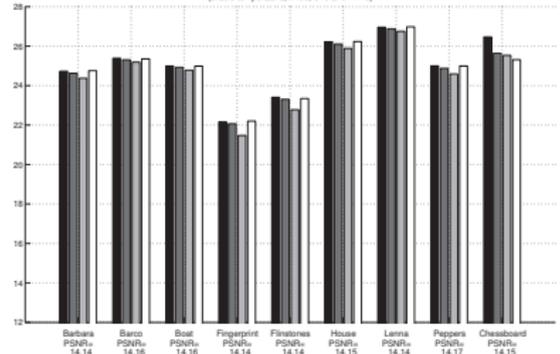
$\sigma = 10$

PSNR for varying the central patch weight in the NL-means procedure  
(patch temperature,  $\sigma=20$ ,  $S=5$  and  $R=13$ )



$\sigma = 20$

PSNR for varying the central patch weight in the NL-means procedure  
(patch temperature,  $\sigma=50$ ,  $S=5$  and  $R=13$ )



$\sigma = 50$

# Loi a priori parcimonieuse

## A priori continu

- Loi a priori  $\pi$  chargeant  $\mathbb{R}^M$  (ou le simplexe).
- Utilisation de l'estimée du risque des mélanges !
- Résultat d'optimalité parmi les mélanges :

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P}(I)(i_1, i_2)\|^2 \right) \\ & \leq \inf_p \left( \int_{\lambda} \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 + \beta \mathcal{K}(p, \pi) \right) \end{aligned}$$

# Loi a priori parcimonieuse

## A priori continu

- Loi a priori  $\pi$  chargeant  $\mathbb{R}^M$  (ou le simplexe).
- Utilisation de l'estimée du risque des mélanges !
- Résultat d'optimalité parmi les mélanges :

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P(I)}(i_1, i_2)\|^2 \right) \\ & \leq \inf_{\rho} \left( \int_{\lambda} \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 + \beta \mathcal{K}(\rho, \pi) \right) \end{aligned}$$

## A priori parcimonieux

- A priori  $\pi$  rendant le second membre « uniformément » petit.
- Student (ou mélange de gaussiennes)  $\rightarrow$  noyaux parcimonieux.
- Optimalité parmi les noyaux parcimonieux :

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P(I)}(i_1, i_2)\|^2 \right) \\ & \leq \inf_{\lambda} \left( \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 + c\beta \|\lambda\|_0 \log M \right) \end{aligned}$$

# Loi a priori parcimonieuse

## A priori continu

- Loi a priori  $\pi$  chargeant  $\mathbb{R}^M$  (ou le simplexe).
- Utilisation de l'estimée du risque des mélanges !
- Résultat d'optimalité parmi les mélanges :

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P(I)}(i_1, i_2)\|^2 \right) \\ & \leq \inf_{\lambda} \left( \int_{\mathcal{P}} \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 + \beta \mathcal{K}(p, \pi) \right) \end{aligned}$$

## A priori parcimonieux

- A priori  $\pi$  rendant le second membre « uniformément » petit.
- Student (ou mélange de gaussiennes)  $\rightarrow$  noyaux parcimonieux.
- Optimalité parmi les noyaux parcimonieux :

$$\begin{aligned} & \mathbb{E} \left( \|P(I)(i_1, i_2) - \widehat{P(I)}(i_1, i_2)\|^2 \right) \\ & \leq \inf_{\lambda} \left( \|P(I)(i_1, i_2) - P_{\lambda}(i_1, i_2)\|^2 + W^2 \sigma^2 \|\lambda\|^2 + c\beta \|\lambda\|_0 \log M \right) \end{aligned}$$

# Comment le calculer ?

Une intégrale en grande dimension

- $\hat{I} = P_{\lambda_\pi}(i_1, i_2)$  où

$$\lambda_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} \lambda d\pi(\lambda) \quad .$$

et  $\hat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda(i_1, i_2)\|^2 - W^2(1 - 2\lambda_{(i_1, i_2)})\sigma^2$ .

- Intégrale de grande dimension similaire à certaines intégrales apparaissant dans le cadre bayésien.

# Comment le calculer ?

## Une intégrale en grande dimension

- $\hat{I} = P_{\lambda_\pi}(i_1, i_2)$  où

$$\lambda_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} \lambda d\pi(\lambda) \quad .$$

et  $\hat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda(i_1, i_2)\|^2 - W^2(1 - 2\lambda_{(i_1, i_2)})\sigma^2$ .

- Intégrale de grande dimension similaire à certaines intégrales apparaissant dans le cadre bayésien.

## MCMC

- Approximation par des méthodes de Monte Carlo basées sur la diffusion de Langevin.
- Convergence de la chaîne assez lente.
- Accélération et amélioration avec une étape de présélection des patches.

# Comment le calculer ?

## Une intégrale en grande dimension

- $\hat{I} = P_{\lambda_\pi}(i_1, i_2)$  où

$$\lambda_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta} \hat{r}_\lambda}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta} \hat{r}_{\lambda'}} d\pi(\lambda')} \lambda d\pi(\lambda) \quad .$$

et  $\hat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda(i_1, i_2)\|^2 - W^2(1 - 2\lambda_{(i_1, i_2)})\sigma^2$ .

- Intégrale de grande dimension similaire à certaines intégrales apparaissant dans le cadre bayésien.

## MCMC

- Approximation par des méthodes de Monte Carlo basées sur la diffusion de Langevin.
- Convergence de la chaîne assez lente.
- Accélération et amélioration avec une étape de présélection des patches.

# Résultats numériques

# Résultats numériques



Originale



Bruitée (22.06 dB)



NL-Means (29.69 dB)



PAC-Bayésien (29.69 dB)

## Méthodologie

- Comparaison avec les NL-Means avec  $\gamma$  bien choisi.
- Agrégation PAC-Bayésienne avec un a priori de Student.

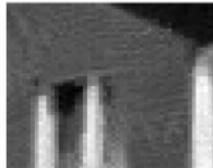
# Résultats numériques



Originale



Bruitée (22.06 dB)



NL-Means (29.69 dB)



PAC-Bayésien (29.69 dB)

## Méthodologie

- Comparaison avec les NL-Means avec  $\gamma$  bien choisi.
- Agrégation PAC-Bayésienne avec un a priori de Student.

## Résultats

- Résultats similaires à ceux obtenus avec les NL-Means...
- + stabilité pour les paramètres et marges d'améliorations...

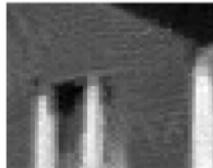
# Résultats numériques



Originale



Bruitée (22.06 dB)



NL-Means (29.69 dB)



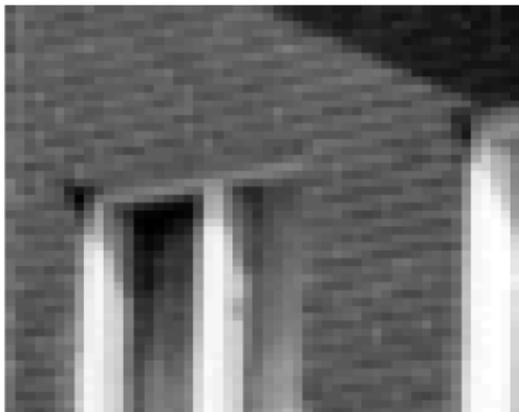
PAC-Bayésien (29.69 dB)

## Méthodologie

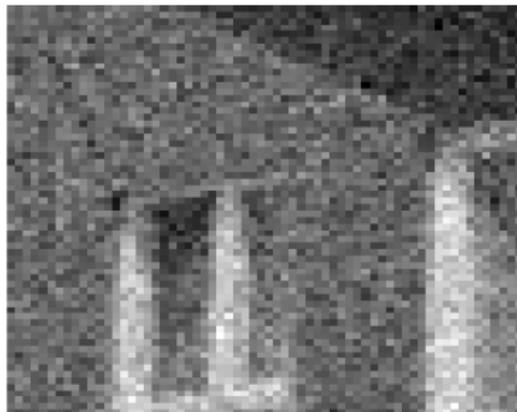
- Comparaison avec les NL-Means avec  $\gamma$  bien choisi.
- Agrégation PAC-Bayésienne avec un a priori de Student.

## Résultats

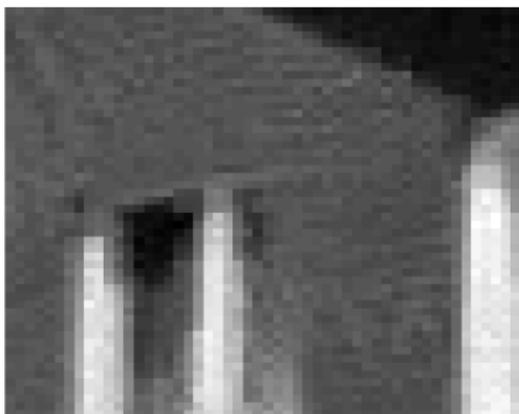
- Résultats similaires à ceux obtenus avec les NL-Means...
- + stabilité pour les paramètres et marges d'améliorations...



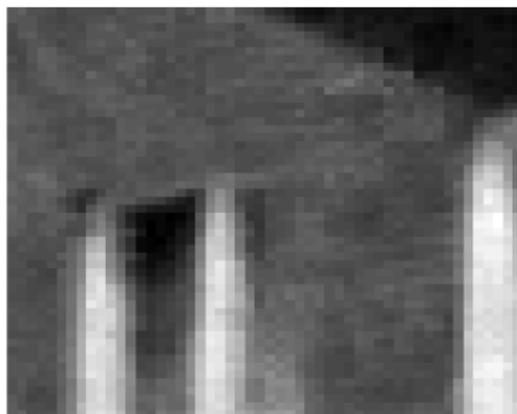
Originale



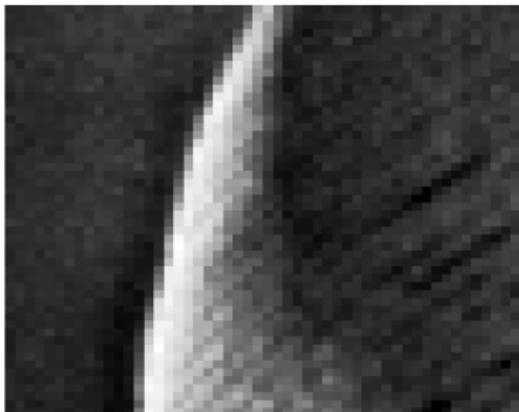
Bruitée (22.06 dB)



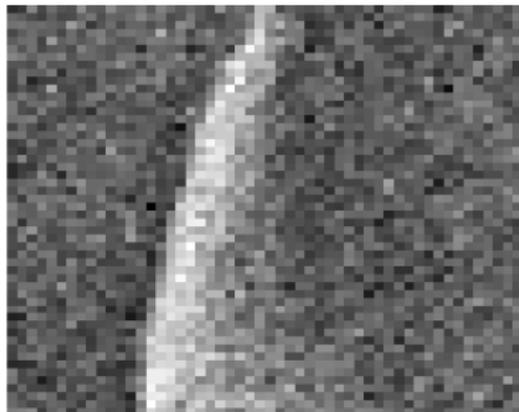
NL-Means (29.69 dB)



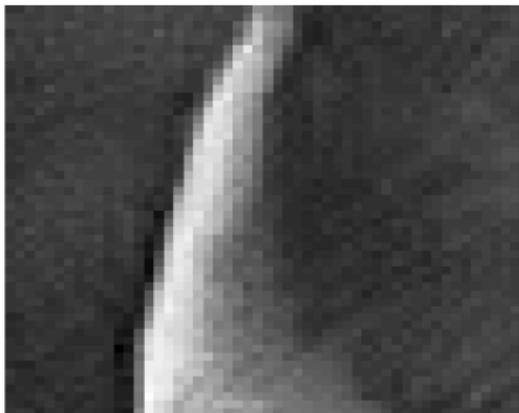
PAC-Bayésien (29.69 dB)



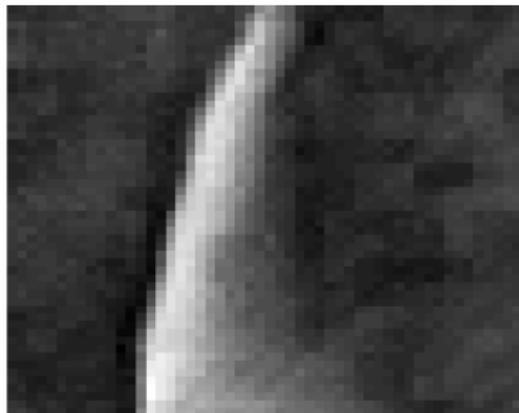
Originale



Bruitée (22.28 dB)



NL-Means (31.59 dB)



PAC-Bayésien (30.78 dB)



Originale



Bruitée (28.13 dB)



NL-Means (31.19 dB)



PAC-Bayésien (32.20 dB)

# Conclusion

## Un point de vue agrégé sur les NL-Means

- Relecture des poids exponentielles et de la distance entre patches.
- Lien avec l'agrégation PAC-Bayésienne et Stein.
- Proposition d'une modification raisonnable pour le poids central.
- Proposition d'une méthode d'agrégation de patch « imitant » le meilleur noyau.
- Outils méthodologiques pour des théorèmes.

# Conclusion

## Un point de vue agrégé sur les NL-Means

- Relecture des poids exponentielles et de la distance entre patchs.
- Lien avec l'agrégation PAC-Bayésienne et Stein.
- Proposition d'une modification raisonnable pour le poids central.
- Proposition d'une méthode d'agrégation de patch « imitant » le meilleur noyau.
- Outils méthodologiques pour des théorèmes.

## Work in progress...

- Résultats numériques pour le poids central.
- Résultats préliminaires pour l'agrégation.
- Vers des résultats théoriques avec splitting : « NL-Means » et régression avec random design.
- Cas plus général ???

# Conclusion

## Un point de vue agrégé sur les NL-Means

- Relecture des poids exponentielles et de la distance entre patches.
- Lien avec l'agrégation PAC-Bayésienne et Stein.
- Proposition d'une modification raisonnable pour le poids central.
- Proposition d'une méthode d'agrégation de patch « imitant » le meilleur noyau.
- Outils méthodologiques pour des théorèmes.

## Work in progress...

- Résultats numériques pour le poids central.
- Résultats préliminaires pour l'agrégation.
- Vers des résultats théoriques avec splitting : « NL-Means » et régression avec random design.
- Cas plus général ???