NL-Means et Procédures d'Agrégation

Joseph Salmon [1]

[1]LPMA-UMR 7599
Université Paris Diderot - Paris 7

10 Décembre 2009 - Séminaire Image du Greyc

NL-Means et agrégation

Cadre

- Estimer une image I à partir d'une observation bruitée Y

$$Y = f + \sigma\varepsilon \qquad (\varepsilon \text{ bruit blanc gaussien})$$

NL-Means et agrégation

Cadre

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

État de l'art

- Solution classique : moyenne locale des valeurs des pixels
- Approche « Patch » : utiliser des voisinages de pixels plutôt que les pixels seuls
- NL-Means : Lissage gaussien dans l'espace des patchs

NL-Means et agrégation

Cadre

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

État de l'art

- Solution classique : moyenne locale des valeurs des pixels
- Approche « Patch » : utiliser des voisinages de pixels plutôt que les pixels seuls
- NL-Means : Lissage gaussien dans l'espace des patchs

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

$$Y = f + \sigma\varepsilon \qquad (\varepsilon \text{ bruit blanc gaussien})$$

État de l'art

- Solution classique : moyenne locale des valeurs des pixels
- Approche « Patch » : utiliser des voisinages de pixels plutôt que les pixels seuls
- NL-Means : Lissage gaussien dans l'espace des patchs

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

Plan

Méthodes à noyaux et NL-Means

- Image, bruit et méthodes à noyaux
- Patchs
- NL-Means et interprétations

Agrégation d'estimateurs

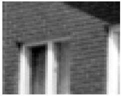
Agrégation de patchs

Images, bruit et estimation

f : Image idéale $N \times N$

- Pour un pixel $i = (i_1, i_2) \in [\![1, N]\!]^2, f(i) \in \mathbb{R}$
- Perte L_2 (perte quadratique), norme associée : $\| \cdot \|$

Images, bruit et estimation



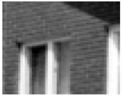
f : Image idéale $N \times N$

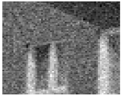
- Pour un pixel $i = (i_1, i_2) \in [\![1, N]\!]^2, f(i) \in \mathbb{R}$
- Perte L_2 (perte quadratique), norme associée : $\| \cdot \|$

Observation bruitée

- $Y(i) = f(i) + \sigma\varepsilon(i)$
- ε bruit gaussien standard i.i.d. et σ^2 connue
- Autres bruits possibles...

Images, bruit et estimation





f : Image idéale $N \times N$

- Pour un pixel $i = (i_1, i_2) \in [\![1, N]\!]^2, f(i) \in \mathbb{R}$
- Perte L_2 (perte quadratique), norme associée : $\| \cdot \|$

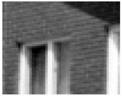
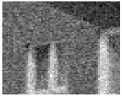
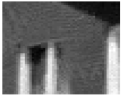
Observation bruitée

- $Y(i) = f(i) + \sigma\varepsilon(i)$
- ε bruit gaussien standard i.i.d. et <u>σ^2 connue</u>
- Autres bruits possibles...

Estimation

- Estimer $f(i)$ par $\widehat{f}(i)$ à partir Y
- Comportement non local possible...

Images, bruit et estimation

f : Image idéale $N \times N$

- Pour un pixel $i = (i_1, i_2) \in [\![1, N]\!]^2, f(i) \in \mathbb{R}$
- Perte L_2 (perte quadratique), norme associée : $\| \cdot \|$

Observation bruitée

- $Y(i) = f(i) + \sigma\varepsilon(i)$
- ε bruit gaussien standard i.i.d. et $\underline{\sigma^2 \text{ connue}}$
- Autres bruits possibles...

Estimation

- Estimer $f(i)$ par $\widehat{f}(i)$ à partir Y
- Comportement non local possible...

Méthodes à noyaux

Méthode à noyaux générique

- Estimer $f(i)$ par moyennage $\widehat{f}(i) = \sum_{k \in [\![1,N]\!]^2} \theta_{i,k} Y_k$
- Les poids $\theta_{i,k}$ peuvent (vont) dépendre de la position et de Y

Méthodes à noyaux

Méthode à noyaux générique

- Estimer $f(i)$ par moyennage $\widehat{f}(i) = \sum_{k \in [\![1,N]\!]^2} \theta_{i,k} Y_k$
- Les poids $\theta_{i,k}$ peuvent (vont) dépendre de la position et de Y

Noyau classique

- $\theta_{i,k} = \dfrac{K_h(i-k)}{\sum_{k'} K_h(i-k')}$ (pas de dépendance en Y)
  - Exemple gaussien $K_h(i) = e^{-(i_1^2+i_2^2)/2h^2}$
  - Utilise la proximité spatiale uniquement

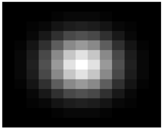
Méthodes à noyaux

Méthode à noyaux générique

- Estimer $f(i)$ par moyennage $\widehat{f}(i) = \sum_{k \in [\![1,N]\!]^2} \theta_{i,k} Y_k$
- Les poids $\theta_{i,k}$ peuvent (vont) dépendre de la position et de Y

Noyau classique

- $\theta_{i,k} = \frac{K_h(i-k)}{\sum_{k'} K_h(i-k')}$ (pas de dépendance en Y)



- Exemple gaussien $K_h(i) = e^{-(i_1^2+i_2^2)/2h^2}$
- Utilise la proximité spatiale uniquement

Noyaux dépendants des données

Exemples

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

Noyaux dépendants des données

Exemples

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

Filtrage bilatéral (Tomasi et al. [TM98])

- $\theta_{i,k} = \dfrac{K_h(i,k) \times K'_{h'}(Y(i) - Y(k))}{\sum_{k'} K_{h'}(i-k', i-k') \times K'_{h'}(Y(i) - Y(k'))}$
- Version gaussienne : $\theta_{i,k} = \dfrac{e^{-\frac{(i_1-k_1)^2+(i_2-k_2)^2}{2h^2}} \times e^{-\frac{(Y(i)-Y(k))^2}{2h'^2}}}{\sum_{k'} e^{-\frac{(i_1-k'_1)^2+(i_2-k'_2)^2}{2h^2}} \times e^{-\frac{(Y(i)-Y(k'))^2}{2h'^2}}}$
  - Intuition : moyenner des pixels proches à la fois en distance et en valeur
  - Problème : non robuste car trop local...

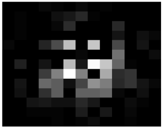
Noyaux dépendants des données

Exemples

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

Filtrage bilatéral (Tomasi et al. [TM98])

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



- Intuition : moyenner des pixels proches à la fois en distance et en valeur
- Problème : non robuste car trop local...

Patchs

Patch

- Patch : version moins locale que la valeur d'un pixel
- $P(f)(i)$: Patch (=sous-image) centré en i, de largeur W :
  $P(f)(i)(j) = f(i_1+j_1, i_2+j_2)$ avec $-\dfrac{W-1}{2} \leq j_1, j_2 \leq \dfrac{W-1}{2}$

Patchs

Patch

- Patch : version moins locale que la valeur d'un pixel
- $P(f)(i)$: Patch (=sous-image) centré en i, de largeur W :
  $P(f)(i)(j) = f(i_1+j_1, i_2+j_2)$ avec $-\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$

Patch et images

- $f \mapsto P(f)$ à une image f associe la collection de patchs $P(f)$
- Relèvement : f est 1-D et $P(f)$ est W^2-D
- Reprojection d'une collection de patchs $P(f)$ vers une image $\tilde{f}$

Patchs

Patch

- Patch : version moins locale que la valeur d'un pixel
- $P(f)(i)$: Patch (=sous-image) centré en i, de largeur W :
  $P(f)(i)(j) = f(i_1+j_1, i_2+j_2)$ avec $-\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$

Patch et images

- $f \mapsto P(f)$ à une image f associe la collection de patchs $P(f)$
- Relèvement : f est 1-D et $P(f)$ est W^2-D
- Reprojection d'une collection de patchs $P(f)$ vers une image $\tilde{f}$

Patch et estimation

- Problème d'estimation : Estimation de la collection de patchs $P(f)$ à partir de la collection de patchs $P(Y)$
- Problème très différent de l'estimation de $P(f)$ à partir de $P(f) + \sigma\varepsilon$ (structure du bruit différente)

Patchs

Patch

- Patch : version moins locale que la valeur d'un pixel
- $P(f)(i)$: Patch (=sous-image) centré en i, de largeur W :
  $P(f)(i)(j) = f(i_1+j_1, i_2+j_2)$ avec $-\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$

Patch et images

- $f \mapsto P(f)$ à une image f associe la collection de patchs $P(f)$
- Relèvement : f est 1-D et $P(f)$ est W^2-D
- Reprojection d'une collection de patchs $P(f)$ vers une image $\tilde{f}$

Patch et estimation

- Problème d'estimation : Estimation de la collection de patchs $P(f)$ à partir de la collection de patchs $P(Y)$
- Problème très différent de l'estimation de $P(f)$ à partir de $P(f) + \sigma\varepsilon$ (structure du bruit différente)

Méthodes à patchs

Méthodes à noyaux et à patchs

- Estimation par moyennage de patchs :

$$\widehat{P(f)}(i)(j) = \sum_k \theta_{i,k} P(Y)(k)(j)$$

- Patchs = inutiles si les $\theta_{i,k}$ ne dépendent pas de $P(Y)$!

Méthodes à patchs

Méthodes à noyaux et à patchs

- Estimation par moyennage de patchs :

$$\widehat{P(f)}(i)(j) = \sum_k \theta_{i,k} P(Y)(k)(j)$$

- Patchs = inutiles si les $\theta_{i,k}$ ne dépendent pas de $P(Y)$!

Intuition

- Utiliser des poids tenant compte de la similarité des patchs :
  - Patch $P_{i_0} = P(Y)(i_0)$ à débruiter
  - Patchs similaires, utiles : poids importants
  - Patchs moins similaires, moins utiles : poids faibles
  - Patchs très différents, inutiles : poids quasi nuls

Méthodes à patchs

Méthodes à noyaux et à patchs

- Estimation par moyennage de patchs :

$$\widehat{P(f)}(i)(j) = \sum_k \theta_{i,k} P(Y)(k)(j)$$

- Patchs = inutiles si les $\theta_{i,k}$ ne dépendent pas de $P(Y)$!

Intuition

- Utiliser des poids tenant compte de la similarité des patchs :

- Patch $P_{i_0} = P(Y)(i_0)$ à débruiter
- Patchs similaires, utiles : poids importants
- Patchs moins similaires, moins utiles : poids faibles
- Patchs très différents, inutiles : poids quasi nuls

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité D entre patchs
- Utiliser des poids $\theta_{i,k} = \frac{K(D(P_i,P_k))}{\sum_{k'} K(D(P_i,P_{k'}))}$
- Choisir $D(P_i, P_k) = \|P_i - P_k\|$ pour mesurer la dissimilarité, un noyau gaussien $K(x) = \exp(-x^2/\beta)$ et une température β

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité D entre patchs
- Utiliser des poids $\theta_{i,k} = \frac{K(D(P_i,P_k))}{\sum_{k'} K(D(P_i,P_{k'}))}$
- Choisir $D(P_i, P_k) = \|P_i - P_k\|$ pour mesurer la dissimilarité, un noyau gaussien $K(x) = \exp(-x^2/\beta)$ et une température β

Résultats

- Méthode intuitivement simple, efficace et pas trop lente
- Performance proche de l'état de l'art

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité D entre patchs
- Utiliser des poids $\theta_{i,k} = \frac{K(D(P_i,P_k))}{\sum_{k'} K(D(P_i,P_{k'}))}$
- Choisir $D(P_i, P_k) = \|P_i - P_k\|$ pour mesurer la dissimilarité, un noyau gaussien $K(x) = \exp(-x^2/\beta)$ et une température β

Résultats

- Méthode intuitivement simple, efficace et pas trop lente
- Performance proche de l'état de l'art

Variations

- Adapter automatiquement la zone de recherche (Kervrann et al. [KB06])
- Polynôme local d'ordre plus élevé (Takeda et al. [TFM07])
- Autres mesures de similarité (Guichard et al. [APG07])

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choisir une mesure de dissimilarité D entre patchs
- Utiliser des poids $\theta_{i,k} = \frac{K(D(P_i,P_k))}{\sum_{k'} K(D(P_i,P_{k'}))}$
- Choisir $D(P_i, P_k) = \|P_i - P_k\|$ pour mesurer la dissimilarité, un noyau gaussien $K(x) = \exp(-x^2/\beta)$ et une température β

Résultats

- Méthode intuitivement simple, efficace et pas trop lente
- Performance proche de l'état de l'art

Variations

- Adapter automatiquement la zone de recherche (Kervrann et al. [KB06])
- Polynôme local d'ordre plus élevé (Takeda et al. [TFM07])
- Autres mesures de similarité (Guichard et al. [APG07])

Interprétation(s) des NL-Means

NL-Means au patch i

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k)\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k')\|^2}} P(Y)(k)$$

Interprétation(s) des NL-Means

NL-Means au patch i

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k)\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k')\|^2}} P(Y)(k)$$

Diffusion / Lissage sur des Variétés

- Explication intuitive mais les preuves demandent des hypothèses fortes [Bua06]
- Variante sur des graphes [Pey08], des variétés [TB09]

Interprétation(s) des NL-Means

NL-Means au patch i

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k)\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k')\|^2}} P(Y)(k)$$

Diffusion / Lissage sur des Variétés

- Explication intuitive mais les preuves demandent des hypothèses fortes [Bua06]
- Variante sur des graphes [Pey08], des variétés [TB09]

Modèle pseudo bayésien

- Modèle d'observation (faux) : $P(Y)(i) = P(Y)(k) + \sqrt{\beta/2}\varepsilon$
- Loi a priori uniforme sur les $P(Y)(k)$

Interprétation(s) des NL-Means

NL-Means au patch i

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k)\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|P(Y)(i)-P(Y)(k')\|^2}} P(Y)(k)$$

Diffusion / Lissage sur des Variétés

- Explication intuitive mais les preuves demandent des hypothèses fortes [Bua06]
- Variante sur des graphes [Pey08], des variétés [TB09]

Modèle pseudo bayésien

- Modèle d'observation (faux) : $P(Y)(i) = P(Y)(k) + \sqrt{\beta/2}\varepsilon$
- Loi a priori uniforme sur les $P(Y)(k)$

Zone de recherche et limites de la méthodes

Hypothèse sur l'image et zone de recherche

- Hypothèses fortes : stationnaires et β-mélangeants (vraies pour les textures . . .) [Bua06]

Zone de recherche et limites de la méthodes

Hypothèse sur l'image et zone de recherche

- Hypothèses fortes : stationnaires et β-mélangeants (vraies pour les textures ...) [Bua06]
- $\beta \longrightarrow 0$ (théorie) : [BCM05]
  $\beta = 12\sigma^2$? est-ce un bon choix ?

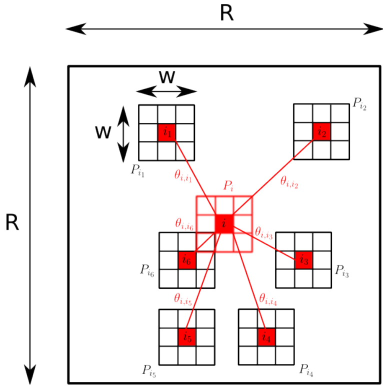
Zone de recherche et limites de la méthodes

Hypothèse sur l'image et zone de recherche

- Hypothèses fortes : stationnaires et β-mélangeants (vraies pour les textures . . .) [Bua06]
- $\beta \longrightarrow 0$ (théorie) : [BCM05]
  $\beta = 12\sigma^2$? est-ce un bon choix ?
- Zone de recherche = image entière (en théorie) : mais trop long, on choisit une petite zone de recherche (ex : $R = 21$)

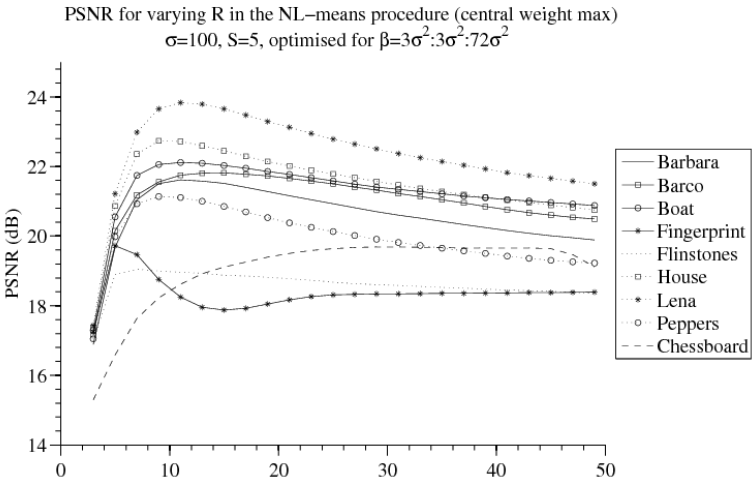
Zone de recherche et limites de la méthodes

Hypothèse sur l'image et zone de recherche



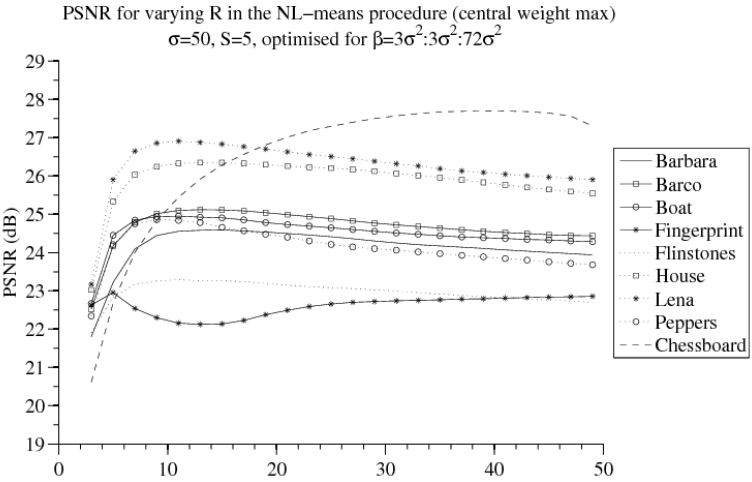
- Hypothèses fortes : stationnaires et β-mélangeants (vraies pour les textures ...) [Bua06]
- $\beta \longrightarrow 0$ (théorie) : [BCM05]
  $\beta = 12\sigma^2$? est-ce un bon choix ?
- Zone de recherche = image entière (en théorie) : mais trop long, on choisit une petite zone de recherche (ex : $R = 21$)

Attention au choix de R



$\sigma = 100$

Attention au choix de R (II)



$\sigma = 50$

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 :

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

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 :

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

Plan

Méthodes à noyaux et NL-Means

Agrégation d'estimateurs

- Estimateurs initiaux et agrégation
- Agrégation PAC-Bayésienne
- Outils de preuve

Agrégation de patchs

Estimateurs initiaux et agrégation

Modèle et estimateurs initiaux

- $Y = f + \sigma\varepsilon$ de taille $W \times W$
- $\{P_k\}$ collection de M estimateur initiaux de f

Estimateurs initiaux et agrégation

Modèle et estimateurs initiaux

- $Y = f + \sigma\varepsilon$ de taille $W \times W$
- $\{P_k\}$ collection de M estimateur initiaux de f

Agrégation

- Estimer f par combinaisons linéaires : $\widehat{f} = P_\theta = \sum_k \theta_k P_k$
- Procédure d'agrégation : manière de choisir θ à partir de Y
- Notations : $e_k = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^M$ et δ_k mesure de Dirac associée sur $\mathbb{R}^M$

Estimateurs initiaux et agrégation

Modèle et estimateurs initiaux

- $Y = f + \sigma\varepsilon$ de taille $W \times W$
- $\{P_k\}$ collection de M estimateur initiaux de f

Agrégation

- Estimer f par combinaisons linéaires : $\widehat{f} = P_\theta = \sum_k \theta_k P_k$
- Procédure d'agrégation : manière de choisir θ à partir de Y
- Notations : $e_k = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^M$ et δ_k mesure de Dirac associée sur $\mathbb{R}^M$

Inégalités de type oracle

- Résultat typique : agrégation « optimale » sur une classe Θ,

$$\mathbb{E}\left(\|f - \widehat{f}\|^2\right) \leq C \inf_{\theta \in \Theta} \left[\|f - P_\theta\|^2 + \sigma^2 \mathsf{V}(\theta)\right]$$

- C, Θ et V dépendent de la procédure

Estimateurs initiaux et agrégation

Modèle et estimateurs initiaux

- $Y = f + \sigma\varepsilon$ de taille $W \times W$
- $\{P_k\}$ collection de M estimateur initiaux de f

Agrégation

- Estimer f par combinaisons linéaires : $\widehat{f} = P_\theta = \sum_k \theta_k P_k$
- Procédure d'agrégation : manière de choisir θ à partir de Y
- Notations : $e_k = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^M$ et δ_k mesure de Dirac associée sur $\mathbb{R}^M$

Inégalités de type oracle

- Résultat typique : agrégation « optimale » sur une classe Θ,

$$\mathbb{E}\left(\|f - \widehat{f}\|^2\right) \leq C \inf_{\theta \in \Theta} \left[\|f - P_\theta\|^2 + \sigma^2 \mathsf{V}(\theta)\right]$$

- C, Θ et V dépendent de la procédure

Procédures d'agrégation

Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

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

- Optimisation numérique très difficile...

Procédures d'agrégation

Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

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

- Optimisation numérique très difficile...

Pénalisation ℓ^1

- Sélection par pénalisation proportionnelle à la norme ℓ^1 :

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

- 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 :

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

- Optimisation numérique très difficile...

Pénalisation ℓ^1

- Sélection par pénalisation proportionnelle à la norme ℓ^1 :

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

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

Poids exponentiels

- Mélange de solutions pondérées par des poids exponentiels :

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

Procédures d'agrégation

Pénalisation AIC/BIC

- Sélection par pénalisation proportionnelle à la dimension :

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

- Optimisation numérique très difficile...

Pénalisation ℓ^1

- Sélection par pénalisation proportionnelle à la norme ℓ^1 :

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

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

Poids exponentiels

- Mélange de solutions pondérées par des poids exponentiels :

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

Agrégation PAC-Bayésienne

Pré-requis

- Estimateurs sans biais $\hat{r}_\theta$ de $\mathbb{E}\|f - P_\theta\|^2$ obtenus par Stein
- Loi a priori π sur $\theta \subset \mathbb{R}^M$

Agrégation PAC-Bayésienne

Pré-requis

- Estimateurs sans biais $\hat{r}_\theta$ de $\mathbb{E}\|f - P_\theta\|^2$ obtenus par Stein
- Loi a priori π sur $\theta \subset \mathbb{R}^M$

Agrégé PAC-Bayésien

- Procédure spécifique reposant sur des poids exponentiels
- Estimateur dépend de π et d'une température β : $\widehat{f} = P_{\theta_\pi}$ avec

$$\theta_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\hat{r}_\theta}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\hat{r}_{\theta'}} d\pi(\theta')} \theta \, d\pi(\theta) \in \mathbb{R}^M$$

- Si l'a priori ne charge que les estimateurs initiaux, $\pi = \sum_k \pi_k \delta_k$, notant $\hat{r}_k = \hat{r}_\theta$ si $\theta = e_k$, alors on retrouve

$$\widehat{f} = \sum_k \frac{e^{-\frac{1}{\beta}\hat{r}_k}\pi_k}{\sum_{k'} e^{-\frac{1}{\beta}\hat{r}_{k'}}\pi_{k'}} P_k$$

Agrégation PAC-Bayésienne

Pré-requis

- Estimateurs sans biais $\hat{r}_\theta$ de $\mathbb{E}\|f - P_\theta\|^2$ obtenus par Stein
- Loi a priori π sur $\theta \subset \mathbb{R}^M$

Agrégé PAC-Bayésien

- Procédure spécifique reposant sur des poids exponentiels
- Estimateur dépend de π et d'une température β : $\widehat{f} = P_{\theta_\pi}$ avec
$$\theta_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\hat{r}_\theta}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\hat{r}_{\theta'}}\, d\pi(\theta')} \theta \, d\pi(\theta) \in \mathbb{R}^M$$
- Si l'a priori ne charge que les estimateurs initiaux, $\pi = \sum_k \pi_k \delta_k$, notant $\hat{r}_k = \hat{r}_\theta$ si $\theta = e_k$, alors on retrouve
$$\widehat{f} = \sum_k \frac{e^{-\frac{1}{\beta}\hat{r}_k}\pi_k}{\sum_{k'} e^{-\frac{1}{\beta}\hat{r}_{k'}}\pi_{k'}} P_k$$

Résultats théoriques

Hypothèses sur P_k

- Estimateurs initiaux obtenues par projections (Leung et Barron [LB06])
- Estimateurs initiaux indépendants de Y (Dalalyan et Tsybakov [DT09])

Résultats théoriques

Hypothèses sur P_k

- Estimateurs initiaux obtenues par projections (Leung et Barron [LB06])
- Estimateurs initiaux indépendants de Y (Dalalyan et Tsybakov [DT09])

Inégalité oracle

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

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

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

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

Résultats théoriques

Hypothèses sur P_k

- Estimateurs initiaux obtenues par projections (Leung et Barron [LB06])
- Estimateurs initiaux indépendants de Y (Dalalyan et Tsybakov [DT09])

Inégalité oracle

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

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

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

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

Choix de la loi a priori

Borne d'erreur et a priori

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

Choix de la loi a priori

Borne d'erreur et a priori

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

Loi a priori discrète (sélection de modèle)

- $\pi = \frac{1}{M}\sum_k \delta_k$: $\mathbb{E}\left(\|f-\widehat{f}\|^2\right) \leq \inf_k \|f-P_k\|^2 + \beta \log M$
- « Mieux » que le meilleur estimateur initial

Choix de la loi a priori

Borne d'erreur et a priori

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

Loi a priori discrète (sélection de modèle)

- $\pi = \frac{1}{M}\sum_k \delta_k$: $\mathbb{E}\left(\|f-\widehat{f}\|^2\right) \leq \inf_k \|f-P_k\|^2 + \beta \log M$
- « Mieux » que le meilleur estimateur initial

Loi a priori parcimonieuse

- π : i.i.d. Student ou mélange de gaussienne [DT09]
- Borne : $\mathbb{E}(\|f-\widehat{f}\|^2) \leq \inf_\theta (\|f-P_\theta\|^2 + C\beta\|\theta\|_0 \log M)$
- Mieux que la meilleure agrégation « parcimonieuse »

Choix de la loi a priori

Borne d'erreur et a priori

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

Loi a priori discrète (sélection de modèle)

- $\pi = \frac{1}{M}\sum_k \delta_k$: $\mathbb{E}\left(\|f-\widehat{f}\|^2\right) \leq \inf_k \|f-P_k\|^2 + \beta\log M$
- « Mieux » que le meilleur estimateur initial

Loi a priori parcimonieuse

- π : i.i.d. Student ou mélange de gaussienne [DT09]
- Borne : $\mathbb{E}(\|f-\widehat{f}\|^2) \leq \inf_\theta(\|f-P_\theta\|^2 + C\beta\|\theta\|_0\log M)$
- Mieux que la meilleure agrégation « parcimonieuse »

Outils de preuve

Estimateur PAC-Bayésien

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

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

Outils de preuve

Estimateur PAC-Bayésien

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

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

Outils

- Estimateur sans biais du risque de Stein (SURE)
- Propriétés des mesures de Gibbs (exponentielles), et de la divergence de Kullback-Leibler
- Forme spécifique des poids

Outils de preuve

Estimateur PAC-Bayésien

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

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

Outils

- Estimateur sans biais du risque de Stein (SURE)
- Propriétés des mesures de Gibbs (exponentielles), et de la divergence de Kullback-Leibler
- Forme spécifique des poids

Plan

Agrégation de patchs

Patchs, estimateurs agrégés, SURE et a priori

- Utiliser le patch $P_i = P(Y)(i)$ comme observation et M patchs $P_k = P(Y)(k)$ comme estimateurs primaires
- Estimateurs agrégés : $P_\theta(i) = \sum_k \theta_{i,k} P(Y)(k)$
- Estimateur sans biais $\hat{r}_\theta(i)$ du risque (SURE), ie : $\mathbb{E}\left(\hat{r}_\theta(i)\right) = \mathbb{E}\left(\|P(f)(i) - P_\theta(i)\|^2\right)$ avec

$$\hat{r}_\theta(i) = \|P(Y)(i) - P_\theta(i)\|^2 - W^2(1 - 2\theta_{i,i})\sigma^2$$

- Choix d'une loi a priori π et d'une température β

Agrégation de patchs

Patchs, estimateurs agrégés, SURE et a priori

- Utiliser le patch $P_i = P(Y)(i)$ comme observation et M patchs $P_k = P(Y)(k)$ comme estimateurs primaires
- Estimateurs agrégés : $P_\theta(i) = \sum_k \theta_{i,k} P(Y)(k)$
- Estimateur sans biais $\hat{r}_\theta(i)$ du risque (SURE), ie : $\mathbb{E}\left(\hat{r}_\theta(i)\right) = \mathbb{E}\left(\|P(f)(i) - P_\theta(i)\|^2\right)$ avec

$$\hat{r}_\theta(i) = \|P(Y)(i) - P_\theta(i)\|^2 - W^2(1 - 2\theta_{i,i})\sigma^2$$

- Choix d'une loi a priori π et d'une température β

Estimateur PAC-Bayésien à patchs

- $\widehat{P(f)}(i) = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\hat{r}_\theta(i)}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\hat{r}_{\theta'}(i)} d\pi(\theta')} P_\theta(i) d\pi(\theta)$

Agrégation de patchs

Patchs, estimateurs agrégés, SURE et a priori

- Utiliser le patch $P_i = P(Y)(i)$ comme observation et M patchs $P_k = P(Y)(k)$ comme estimateurs primaires
- Estimateurs agrégés : $P_\theta(i) = \sum_k \theta_{i,k} P(Y)(k)$
- Estimateur sans biais $\hat{r}_\theta(i)$ du risque (SURE), ie : $\mathbb{E}\left(\hat{r}_\theta(i)\right) = \mathbb{E}\left(\|P(f)(i) - P_\theta(i)\|^2\right)$ avec

$$\hat{r}_\theta(i) = \|P(Y)(i) - P_\theta(i)\|^2 - W^2(1 - 2\theta_{i,i})\sigma^2$$

- Choix d'une loi a priori π et d'une température β

Estimateur PAC-Bayésien à patchs

- $\widehat{P(f)}(i) = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\hat{r}_\theta(i)}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\hat{r}_{\theta'}(i)} d\pi(\theta')} P_\theta(i) d\pi(\theta)$

Résultats théoriques

Théorème

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

$$\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_p \int_\theta \left(\|P(f)(i) - P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2\right) dp + \beta\mathcal{K}(p,\pi)$$

Résultats théoriques

Théorème

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

$$\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right) \\ \leq \inf_p \int_\theta \left(\|P(f)(i) - P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2\right) dp + \beta\mathcal{K}(p,\pi)$$

Travail en cours

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

Résultats théoriques

Théorème ?

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

$$\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_p \int_\theta \left(\|P(f)(i) - P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2\right) dp + \beta\mathcal{K}(p,\pi)$$

Travail en cours

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

Loi a priori « discrète »

NL-Means

- Choix $\pi = \sum_k \pi_k \delta_k$
- Agrégation NL-Means pondérées :

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\hat{r}_k(i)}\pi_k}{\sum_{k'} e^{-\frac{1}{\beta}\hat{r}_{k'}(i)}\pi_{k'}} P_k(i)$$

Loi a priori « discrète »

NL-Means

- Choix $\pi = \sum_k \pi_k \delta_k$
- Agrégation NL-Means pondérées :

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\hat{r}_k(i)}\pi_k}{\sum_{k'} e^{-\frac{1}{\beta}\hat{r}_{k'}(i)}\pi_{k'}} P_k(i)$$

Optimalité

- Résultat de type sélection :

$$\begin{aligned}\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right) &\leq \min_k \|P(f)(i) - P(f)(k)\|^2 + W^2\sigma^2 + \beta \log \frac{1}{\pi_i} \\ &\leq W^2\sigma^2 + \beta \log \frac{1}{\pi_i}\end{aligned}$$

- (Presque) Mieux que rien...

Loi a priori « discrète »

NL-Means

- Choix $\pi = \sum_k \pi_k \delta_k$
- Agrégation NL-Means pondérées :

$$\widehat{P(f)}(i) = \sum_k \frac{e^{-\frac{1}{\beta}\hat{r}_k(i)}\pi_k}{\sum_{k'} e^{-\frac{1}{\beta}\hat{r}_{k'}(i)}\pi_{k'}} P_k(i)$$

Optimalité

- Résultat de type sélection :

$$\begin{aligned}\mathbb{E}&\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right)\\ &\leq \min_k \|P(f)(i) - P(f)(k)\|^2 + W^2\sigma^2 + \beta \log \frac{1}{\pi_i}\\ &\leq W^2\sigma^2 + \beta \log \frac{1}{\pi_i}\end{aligned}$$

- (Presque) Mieux que rien...

Le problème du poids central [SL09a, Sal10]

Rôle(s) du patch central [BC07, ZDW08]

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

$$\frac{\exp^{-\frac{1}{\beta}\rho_{(k)}(i)}}{\sum_{k'} \exp^{-\frac{1}{\beta}\rho_{(k')}(i)}} P_k$$

Le problème du poids central [SL09a, Sal10]

Rôle(s) du patch central [BC07, ZDW08]

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

$$\frac{\exp^{-\frac{1}{\beta}\rho_{(k)}(i)}}{\sum_{k'} \exp^{-\frac{1}{\beta}\rho_{(k')}(i)}} P_k$$

4 propositions

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

Le problème du poids central [SL09a, Sal10]

Rôle(s) du patch central [BC07, ZDW08]

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

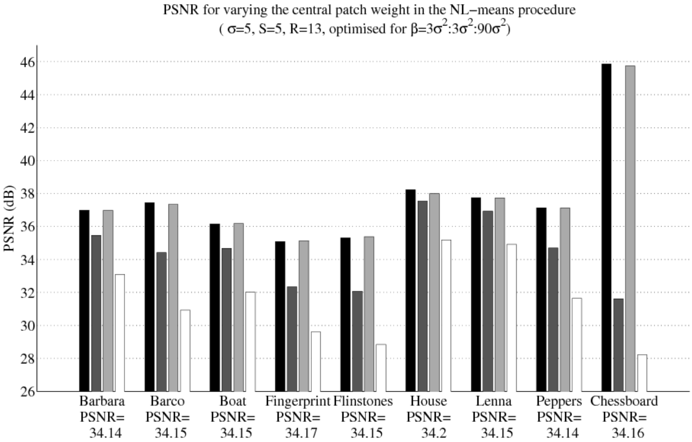
$$\frac{\exp^{-\frac{1}{\beta}\rho_{(k)}(i)}}{\sum_{k'} \exp^{-\frac{1}{\beta}\rho_{(k')}(i)}} P_k$$

4 propositions

- PAC-Bayésien / Stein : utiliser $\hat{r}_\theta$
  $\rho_{(k)}(i) = \|P(Y)(i) - P(Y)(k)\|^2 - W^2\sigma^2$ et $\rho_{(i)}(i) = W^2\sigma^2$
- Buadès et al. : remplacer $\rho_{(i)}(i)$ par $\max\limits_{k \neq i} \rho_{(k)}(i)$
- NL-Means « pur » : $\rho_{(k)}(i) = \|P(Y)(i) - P(Y)(k)\|^2$
- Zéro : ne pas utiliser le patch central...

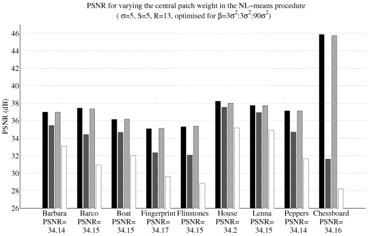
Résultats numériques [Sal10]

Résultats numériques [Sal10]

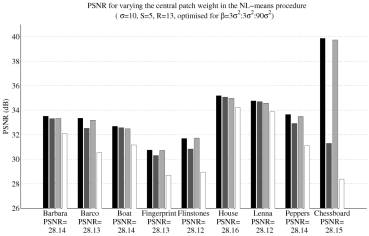


$$\sigma = 5$$

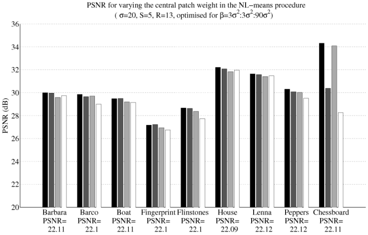
Résultats numériques [Sal10]



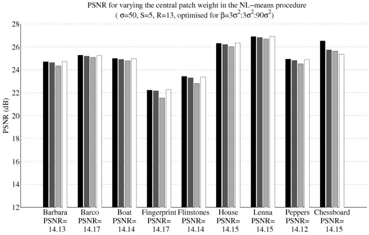
$\sigma = 5$



$\sigma = 10$



$\sigma = 20$



$\sigma = 50$

Loi a priori parcimonieuse

A priori continu

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

$$\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_p \left(\int_\theta \|P(f)(i) - P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2 + \beta\mathcal{K}(p,\pi) \right)$$

Loi a priori parcimonieuse

A priori continu

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

$$\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_p \left(\int_\theta \|P(f)(i) - P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2 + \beta\mathcal{K}(p,\pi) \right)$$

A priori parcimonieux

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

$$\mathbb{E}\left(\|P(f)(i) - \widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_\theta \left(\|P(f)(i) - P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2 + c\beta\|\theta\|_0 \log M \right)$$

Loi a priori parcimonieuse

A priori continu

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

$$\mathbb{E}\left(\|P(f)(i)-\widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_p \left(\int_\theta \|P(f)(i)-P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2 + \beta\mathcal{K}(p,\pi)\right)$$

A priori parcimonieux

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

$$\mathbb{E}\left(\|P(f)(i)-\widehat{P(f)}(i)\|^2\right)$$
$$\leq \inf_\theta \left(\|P(f)(i)-P_\theta(f)(i)\|^2 + W^2\sigma^2\|\theta\|^2 + c\beta\|\theta\|_0 \log M\right)$$

Comment le calculer ?

Une intégrale en grande dimension

- $\hat{f} = P_{\theta_\pi}(i)$ où
$$\theta_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\widehat{r}_\theta}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\widehat{r}_{\theta'}}\, d\pi(\theta')} \theta \, d\pi(\theta) \quad .$$
et $\widehat{r}_\theta = \|P(Y)(i) - P_\theta(i)\|^2 - W^2(1 - 2\theta_{i,i})\sigma^2$.
- Intégrale de grande dimension similaire à certaines intégrales apparaissant dans le cadre bayésien

Comment le calculer ?

Une intégrale en grande dimension

- $\hat{f} = P_{\theta_\pi}(i)$ où
$$\theta_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\widehat{r}_\theta}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\widehat{r}_{\theta'}}\, d\pi(\theta')} \theta \, d\pi(\theta) \quad .$$
et $\widehat{r}_\theta = \|P(Y)(i) - P_\theta(i)\|^2 - W^2(1 - 2\theta_{i,i})\sigma^2$.
- 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 patchs

Comment le calculer ?

Une intégrale en grande dimension

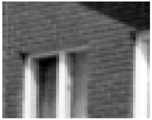
- $\hat{f} = P_{\theta_\pi}(i)$ où
$$\theta_\pi = \int_{\mathbb{R}^M} \frac{e^{-\frac{1}{\beta}\widehat{r}_\theta}}{\int_{\mathbb{R}^M} e^{-\frac{1}{\beta}\widehat{r}_{\theta'}} d\pi(\theta')} \theta \, d\pi(\theta) \quad .$$
et $\widehat{r}_\theta = \|P(Y)(i) - P_\theta(i)\|^2 - W^2(1-2\theta_{i,i})\sigma^2$.
- Intégrale de grande dimension similaire à certaines intégrales apparaissant dans le cadre bayésien

MCMC

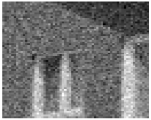
- Approximation par des méthodes de Monte Carlo basées sur la diffusion de Langevin
- Convergence de la chaîne assez lente
- Accélération et amélioration avec une étape de présélection des patchs

Résultats numériques [SL09b]

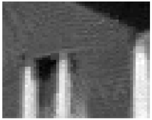
Résultats numériques [SL09b]



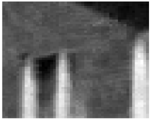
Originale



Bruitée (22.06 dB)



NL-Means (29.69 dB)

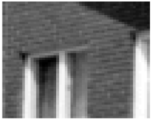


PAC-Bayésien (29.69 dB)

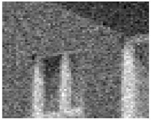
Méthodologie

- Comparaison avec les NL-Means avec β bien choisi
- Agrégation PAC-Bayésienne avec un a priori de Student

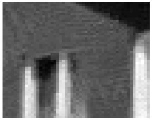
Résultats numériques [SL09b]



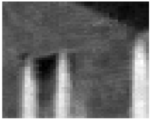
Originale



Bruitée (22.06 dB)



NL-Means (29.69 dB)



PAC-Bayésien (29.69 dB)

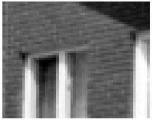
Méthodologie

- Comparaison avec les NL-Means avec β bien choisi
- Agrégation PAC-Bayésienne avec un a priori de Student

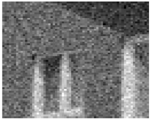
Résultats

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

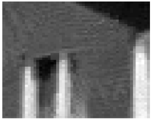
Résultats numériques [SL09b]



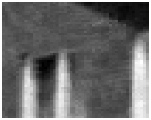
Originale



Bruitée (22.06 dB)



NL-Means (29.69 dB)



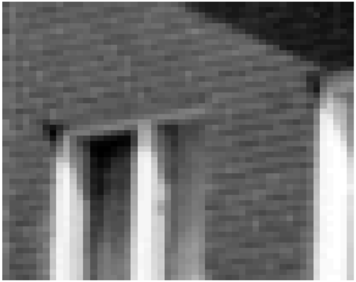
PAC-Bayésien (29.69 dB)

Méthodologie

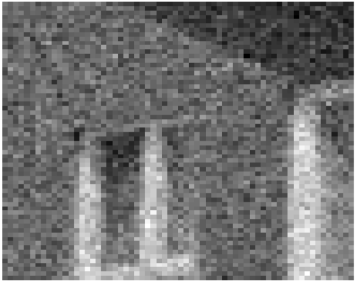
- Comparaison avec les NL-Means avec β bien choisi
- Agrégation PAC-Bayésienne avec un a priori de Student

Résultats

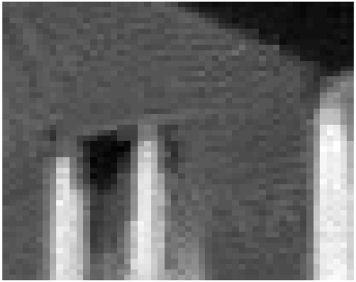
- Résultats similaires à ceux obtenus avec les NL-Means...
- $+$ stabilité pour les paramètres et marges d'améliorations...



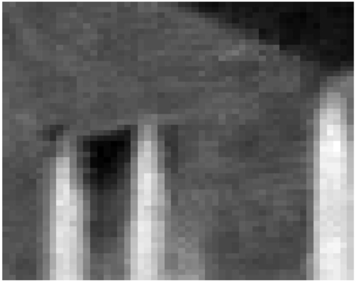
Originale



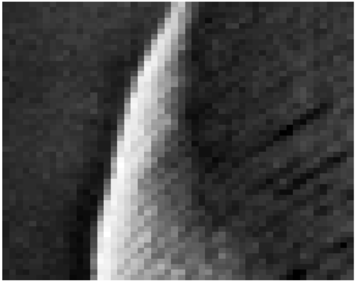
Bruitée (22.06 dB)



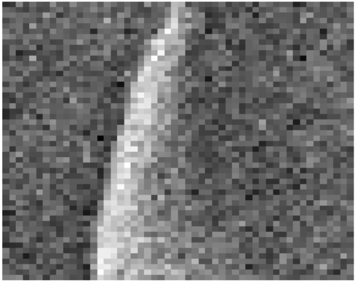
NL-Means (29.69 dB)



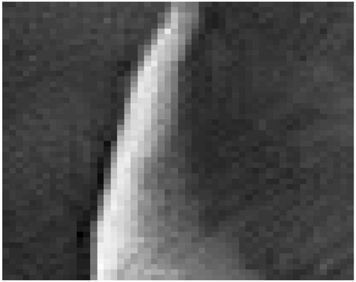
PAC-Bayésien (29.69 dB)



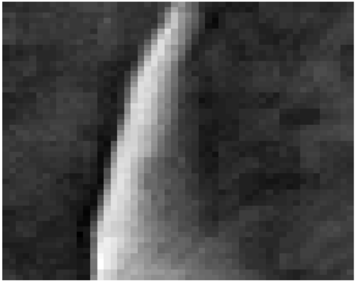
Originale



Bruitée (22.28 dB)

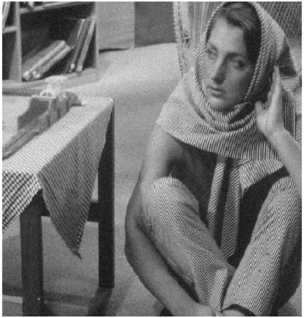


NL-Means (31.59 dB)



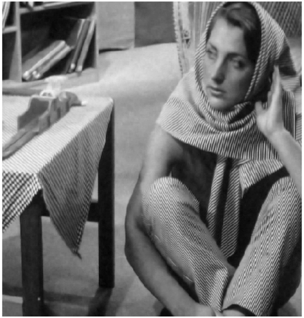
PAC-Bayésien (30.78 dB)

Originale



Bruitée (28.13 dB)

NL-Means (31.19 dB)



PAC-Bayesien (32.20 dB)

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
- Modification raisonnable pour le poids central
- Proposition d'une méthode d'agrégation de patchs « 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
- Modification raisonnable pour le poids central
- Proposition d'une méthode d'agrégation de patchs « 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 patchs
- Lien avec l'agrégation PAC-Bayésienne et Stein
- Modification raisonnable pour le poids central
- Proposition d'une méthode d'agrégation de patchs « 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 ? ? ?

References I

- N. Azzabou, N. Paragios, and F. Guichard.
  Image denoising based on adapted dictionary computation.
  In *ICIP*, pages 109–112, 2007.
- T. Brox and D. Cremers.
  Iterated nonlocal means for texture restoration.
  In *SSVM*, volume 4485 of *Lecture Notes in Computer Science*, pages 13–24, 2007.
- A. Buades, B. Coll, and J-M. Morel.
  A review of image denoising algorithms, with a new one.
  Multiscale Model. Simul., 4(2) :490–530 (electronic), 2005.
- A. Buades.
  Image and movie denoising by non local means.
  PhD thesis, Universitat de les Illes Balears, 2006.
- A. S. Dalalyan and A. B. Tsybakov.
  Sparse regression learning by aggregation and Langevin Monte-Carlo.
  In *22th Annual Conference on Learning Theory, COLT*, 2009.

References II

- C. Kervrann and J. Boulanger.
  Optimal spatial adaptation for patch-based image denoising.
  IEEE Trans. Image Process., 15(10) :'2866–2878, 2006.
- G. Leung and A.R. Barron.
  Information theory and mixing least-squares regressions.
  IEEE Trans. Inform. Theory, 52(8) :3396–3410, 2006.
- G. Peyré.
  Image processing with nonlocal spectral bases.
  Multiscale Model. Simul., 7(2) :703–730, 2008.
- J. Salmon.
  On two parameters for denoising with non-local means.
  IEEE Signal Process. Lett., 2010.
- J. Salmon and E. Le Pennec.
  An aggregator point of view on NL-Means.
  In *Proceedings of the SPIE Conference on Mathematical Imaging : Wavelet XIII*, volume 7446, page 74461E. SPIE, 2009.

References III

- J. Salmon and E. Le Pennec.
  Nl-means and aggregation procedures.
  In *ICIP*, 2009.
- D. Tschumperlé and L. Brun.
  Non-local image smoothing by applying anisotropic diffusion pde's in the space of patches.
  In *ICIP*, 2009.
- H. Takeda, S. Farsiu, and P. Milanfar.
  Kernel regression for image processing and reconstruction.
  IEEE Trans. Image Process., 16(2) :349–366, 2007.
- C. Tomasi and R. Manduchi.
  Bilateral filtering for gray and color images.
  In *ICCV*, page 839, Los Alamitos, CA, USA, 1998. IEEE Computer Society.
- S. Zimmer, S. Didas, and J. Weickert.
  A rotationally invariant block matching strategy improving image denoising with non-local means.
  In *LNLA*, 2008.