NL-Means et Agrégation

Joseph Salmon Dominique Picard Erwan Le Pennec

LPMA
Paris 7-Diderot

Modélisation Statistique des Images, 2008

Sommaire

Introduction

- Cadre: débruitage d'images numériques
- Méthode étudiée: NL-Means
- Enjeu du travail: Pourquoi ça marche si bien?
- Résultats espérés: Inégalités Oracles (Non-asymptotiques)

Introduction

- Cadre: débruitage d'images numériques
- Méthode étudiée: NL-Means
- Enjeu du travail: Pourquoi ça marche si bien?
- Résultats espérés: Inégalités Oracles (Non-asymptotiques)

Introduction

- Cadre: débruitage d'images numériques
- Méthode étudiée: NL-Means
- Enjeu du travail: Pourquoi ça marche si bien?
- Résultats espérés: Inégalités Oracles (Non-asymptotiques)

Introduction

- Cadre: débruitage d'images numériques
- Méthode étudiée: NL-Means
- Enjeu du travail: Pourquoi ça marche si bien?
- Résultats espérés: Inégalités Oracles (Non-asymptotiques)

Plan de l'exposé

Modèle

"Fixed Design"

$$\text{Pixel:} \quad y_i = f(x_i) + \xi_i, \quad i = 1, \ldots, n$$

- $y(i)$ observation, $f(x_i)$ vraie valeur au pixel x_i (déterministes)
- ξ_i i.i.d, $\mathbb{E}(\xi_i) = 0$ et $\mathbb{E}(\xi_i^2) = \sigma^2$ (ex: $\xi_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$)
- Risque d'un estimateur $\widehat{f}_n$: $R(\widehat{f}_n, f) = \mathbb{E}\|\widehat{f}_n - f\|_n^2$

Notations: $Y = (y_1, \ldots, y_n)^\top, \quad X = (x_1, \ldots, x_n)^\top$

Distance: $\|Y - f\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$

Intuition

Couleur du "Patch":

- Rose: à débruiter
- Rouge: très ressemblant
- Orange: peu ressemblant
- Bleu: inutiles

Définition de la méthode selon [BCM05]

NL-Means

Estimateur: $$NL[y]_n(i) = \sum_{j=1}^{n} \theta(i,j) y(j)$$

Poids BCM: $$\theta(i,j) = \frac{\exp\left(-\|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2 / h^2\right)}{\sum_{j=1}^{n} \exp\left(-\|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2 / h^2\right)}$$

$y(\mathcal{N}_i)$ vecteur du "Patch" centré en i, a paramètre de lissage

Remarque: $0 \leq \theta(i,j) \leq 1$ et $\sum_{j=1}^{n} \theta(i,j) = 1$

Définition de la méthode selon [KB06]

NL-Means

Estimateur: $$NL[y]_n(i) = \sum_{j=1}^{n} \theta(i,j) y(j)$$

Poids KB: $$\theta(i,j) = \frac{L_a(y\left(\mathcal{N}_i\right) - y(\mathcal{N}_j)\big) K_h(d(x_i,x_j))}{\sum_{j \in I} L_a\left(y(\mathcal{N}_i) - y(\mathcal{N}_j)\right) K_h(d(x_i,x_j))}$$

$y(\mathcal{N}_i)$ vecteur du "Patch" centré en i
L, K noyaux réels et a, h fenêtres associées

Remarque: $0 \leq \theta(i,j) \leq 1$ (si les noyaux sont positifs) et $\sum_{j=1}^{n} \theta(i,j) = 1$

Définition générale

NL-Means

Estimateur: $NL[y]_n(i) = \sum_{j \in I} \theta(i,j) y(j)$

Poids : $\theta(i,j) = \underbrace{P(i,j|Y)}_{\text{fonction des patchs}} \times \underbrace{\pi(x_i, x_j)}_{\text{a priori}}$

Problème:

- choix des types de patch
- choix distance entre patch
- choix de la forme des poids

Plan de l'exposé

Consistance de la méthode

$Y(\mathcal{N}_i \setminus \{i\})$: Patch privé du point central

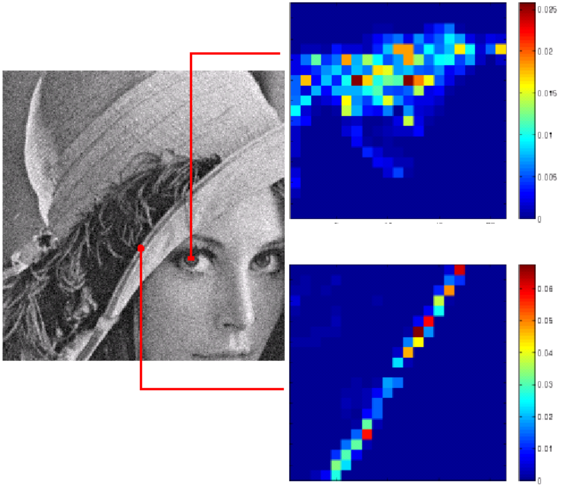
Théorème [BCM05]

Si $(Y(i), Y(\mathcal{N}_i \setminus \{i\})$ est un processus stationnaire mélangeant, alors pour tout j de l'image:

$$NL[y]_n(j) \xrightarrow{n\to\infty} \mathbb{E}\left[Y_i | Y(\mathcal{N}_j \setminus \{j\}) = y(\mathcal{N}_j \setminus \{i\})\right] \ \text{ p.s}$$

Plan de l'exposé

Lissage et noyaux



Intensité des poids:

- Rouge: poids importants
- Intermediaire: poids faibles
- Bleu: poids nuls

Méthode à noyaux

Estimateur à Noyau: $\widehat{f}_n(x) = \sum_{i=1}^{n} \theta_i(x, X) y_i$

Nadaraya-Watson: $\theta_i(x, X) = \dfrac{K_h(x_i - x)}{\sum_{j=1}^{n} K_h(x_j - x)}$

Contrôle de l'erreur: minimisation biais/variance, connu
Limite: peu d'attache aux données et pb dépendance cas NL-means
Oracle: θ fonction de $\|f(\mathcal{N}_i) - f(\mathcal{N}_j)\| \simeq \|f(x_i) - f(x_j)\|$ (si régularité)

Pré-estimateurs et Oracle

Hypothsèse: on a une deuxième image $\widetilde{Y}$ tel que $\widetilde{Y} \overset{\perp\!\!\!\perp}{\sim} Y$
Soient $\widetilde{Y}_\theta(x_i) = \sum_{i=1}^n \theta(i,j)\widetilde{Y}_j$ (exemple: $\theta\left(y(\mathcal{N}_i), \widetilde{y}(\mathcal{N}_j), x_j\right)$) et $f_\theta(x_i) = \sum_{i=1}^n \theta(i,j) f_j$
But: trouver un $\hat{\theta}$ tel que

$$\mathbb{E}\|f - \widetilde{Y}_{\hat{\theta}}\|_n^2 \leq C \inf_\theta \|f - \widetilde{Y}_\theta\|_n^2 + C_n$$

$$\tilde{\mathbb{E}}\mathbb{E}\|f - \widetilde{Y}_{\hat{\theta}}\|_n^2 \leq C \inf_\theta \tilde{\mathbb{E}}\|f - \widetilde{f}_\theta\|_n^2 + C_n$$

$$\tilde{\mathbb{E}}\mathbb{E}\|f - \widetilde{Y}_{\hat{\theta}}\|_n^2 \leq C \inf_\theta \underbrace{\left(\|f - f_\theta\|_n^2 + \frac{\sigma^2}{n}\sum_{i=1}^n \theta_i^2\right)}_{\text{Biais/Variance estimateur à noyau}} + C_n$$

Plan de l'exposé

Définition de l'agrégation

But n°1: estimer f
M Pré-estimateurs de f: $\tilde{f}_1, \ldots, \tilde{f}_M$ (gelés)

Poids: $\theta = (\theta_1, \ldots, \theta_M)$

Agrégat: $\boxed{f_\theta \triangleq \sum_{i=1}^{M} \theta_i \tilde{f}_i = \sum_{i=1}^{M} \theta_i(X, Y) \tilde{f}_i}$

But n°2: trouver $\theta \in \Theta = \begin{cases} \mathbb{R}^M & (L) \\ \{\theta \in \mathbb{R}^M, \|\theta\|_0 = s\} & (S) \\ \Lambda^M = \left\{\theta \in \mathbb{R}^M : \theta_j \geq 0, \sum_{j=1}^M \theta_j \leq 1\right\} & (C) \\ \{e_1, \ldots, e_M\} & (MS) \end{cases}$

où $e_j = (0, \ldots, 1, \ldots, 0)$
$\|\theta\|_0 = \#\{j \,|\, 1 \leq j \leq M \text{ et } \theta_j \neq 0\}$

Résultat de type Oracle

Inégalité Oracle

Trouver un estimateur $f_{\widehat{\theta}}$ de f tel que:

$$\mathbb{E}\left(\|f - f_{\widehat{\theta}}\|_n^2\right) \leq C \inf_{\theta \in \Theta} \|f - f_\theta\|_n^2 + \psi_{n,M}^{\Theta}$$

$\psi_{n,M}^{\Theta}$ terme de reste, "prix de l'agrégation"(indépendant de f)

Contraintes:

- C à trouver proche de 1
- $\psi_{n,M}^{\Theta}$ à prendre petit.

Plan de l'exposé

Pénalité L_1 et LASSO

Exemple bien connu:

$$\widehat{\theta}^{LASSO} = \underset{\theta \in R^M}{\arg\min} \left(\| Y - f_\theta \|_n^2 + \underbrace{2r\|\theta\|_1}_{\text{Pénalité sur } \theta} \right)$$

$$\widehat{f}_n^{LASSO} = f_{\widehat{\theta}_n}$$

où $$\|\theta\|_1 = \sum_{j=1}^{M} |\theta_j|$$

Inégalité oracle disponible [BTW07] (optimale au sens minimax)

Pénalité entropique et poids exponentiels [Lec07]

$$\widehat{\theta}^{EW} = \arg\min_{\theta \in \Lambda^M} \left(\sum_{j=1}^{M} \theta_j \|Y - \tilde{f}_j\|_n^2 + \underbrace{\frac{\beta}{n} \sum_{j=1}^{M} \theta_j \log \theta_j + \log M}_{\text{Pénalité sur } \theta} \right)$$

$$\widehat{f}^{EW} = f_{\widehat{\theta}^{EW}} = \sum_{j=1}^{M} \widehat{\theta}_j^{EW} \tilde{f}_j$$

Solution: $$\widehat{\theta}_j^{EW} = \frac{\exp\left(-n\|Y - \tilde{f}_j\|_n^2/\beta\right)}{\sum_{k=1}^{M} \exp\left(-n\|Y - \tilde{f}_k\|_n^2/\beta\right)}$$

Pénalité entropique et poids exponentiels (suite)

- $\mathbb{E}\|Y - \tilde{f}_j\|_n^2 - \sigma^2 = \|f - \tilde{f}_j\|_n^2$: poids inchangés par ajout de constante
- Optimisation fonction linéaire espace convexe:

$$\arg\min_{j=1,\dots,M} \|f - \tilde{f}_j\|_n^2 = \arg\min_{\theta\in\Lambda^M} \sum_{j=1}^{M} \theta_j \|f - \tilde{f}_j\|_n^2$$

ajouter une penalité sinon sélection modèle

Inégalité Oracle (Sélection de Modèle) [Lec07]

Soit $\tilde{f}_1, \dots, \tilde{f}_M$ des fonctions bornées sur $[0, 1]$. Alors il existe une constante K, telle que pour tout ε on a:

$$\mathbb{E}\|\widehat{f}^{EW} - f\|^2 \leq (1+\varepsilon) \min_{j=1,\dots,M} \left(\|f - \tilde{f}_j\|^2\right) + \frac{K \log M}{\varepsilon n}$$

Plan de l'exposé

Mélange à poids exponentiels

Pour une mesure a priori $\pi \in \mathcal{P}_\Theta$ (mesure sur θ) et $\beta > 0$, $f_\theta = \sum_{i=1}^{M} \theta_i \tilde{f}_i$:

$$\widehat{f}_n(x) \triangleq \int_\Theta \lambda_\theta(\mathbf{Y}) f_\theta(x) \pi(d\theta)$$

$$\text{et} \qquad \lambda_\theta(\mathbf{Y}) \triangleq \frac{\exp(-n\|\mathbf{Y} - f_\theta\|_n^2/\beta)}{\int_\Theta \exp(-n\|\mathbf{Y} - f_\omega\|_n^2/\beta)\pi(d\omega)}$$

où $\mathcal{P}_\Theta$ est l'ensemble des mesures de probabilité sur Θ
$\widehat{f}_n(x)$: espérance a posteriori de f_θ dans le model fantôme

$$Y_i = f_\theta(x_i) + \xi_i'$$

avec ξ_i' i.i.d $\mathcal{N}(0, \frac{\beta}{2})$

Plan de l'exposé

Majoration du risque

Théorème [DT07]

Supposons que $\forall j\ \|\tilde{f}_j\| \leq c_0$, $\xi_1 \sim \mathcal{N}(0, \sigma^2)$, $\forall \pi \in \mathcal{P}_\Theta$, $\beta \geq 4\sigma^2$ alors :

$$\mathbb{E}(\|\widehat{f}_n - f\|_n^2) \leq \inf_{p \in \mathcal{P}_\Theta} \left(\int_\Theta \|f_\theta - f\|_n^2 \, p(d\theta) + \frac{\beta \mathcal{K}(p, \pi)}{n} \right)$$

$$\mathcal{K}(p, \pi) = \int_\Theta p(\lambda) \log \left(\frac{p(\lambda)}{\pi(\lambda)} \right) d\lambda$$

Remarque:

- Généralisable aux bruits non gaussiens (Méthode de Stein)
- $p = \delta_{j_0}$:Inégalité Oracle de type Sélection de Modèle

Plan de l’exposé

1. Non Local Means
   - Définition
   - Convergence
   - Noyaux et NL-Means
2. Agrégation d’Estimateurs
   - Définition
   - Exemples
3. Méthodes PAC/EAC-Bayesienne
   - Point de vue Bayesien
   - Théorème général
   - Résultat de type Oracle

Combinaisons linéaires et SOI

Ici $\Theta = \mathbb{R}^M$, $\forall j\ \|f_j\|_\infty \leq 1$ et $\pi(d\theta) = q(\theta)d\theta$:

$$q_0(t) = \frac{3}{2(1+|t|)^4} \quad \forall t \in \mathbb{R}$$

$$q(\theta) = \prod_{j=1}^{M} \frac{1}{\tau} q_0\left(\frac{\theta_j}{\tau}\right) \quad \forall \theta \in \mathbb{R}^M$$

Théorème [DT07]

Si $\xi_1 \sim \mathcal{N}(0,\sigma^2)$,$\tau = \frac{\sigma}{\sqrt{nM}}$ et $\beta > 4\sigma^2$, il existe une constante C

$$\mathbb{E}\left(\|\widehat{f}_n - f\|_n^2\right) \leq \inf_{\theta \in \mathbb{R}^M} \left[\|f_\theta - f\|_n^2 + \frac{C\|\theta\|_0}{n} \log\left(1 + \frac{\|\theta\|_1\sqrt{Mn}}{\sigma\|\theta\|_0}\right)\right]$$

Perspectives et limites

- Cadre "facile": plusieurs images disponibles (videos, bases de données,...); pas de problème de dépendance
- Cadre recherché: une image seulement.
- $f_j = Y_j$ non gelés! non indépendants! possible pré-découpage (Splitting)
- autre(s) solution(s) ???
- Problème pratique: poids du point central dans le patch ?
- Quid du type de poids? exponentiel = optimal?

Perspectives et limites

- Cadre "facile": plusieurs images disponibles (videos, bases de données,. . .); pas de problème de dépendance
- Cadre recherché: une image seulement.
- $f_j = Y_j$ non gelés! non indépendants! possible pré-découpage (Splitting)
- autre(s) solution(s) ???
- Problème pratique: poids du point central dans le patch ?
- Quid du type de poids? exponentiel = optimal?

Perspectives et limites

- Cadre "facile": plusieurs images disponibles (videos, bases de données,...); pas de problème de dépendance
- Cadre recherché: une image seulement.
- $f_j = Y_j$ non gelés! non indépendants! possible pré-découpage (Splitting)
- autre(s) solution(s) ???
- Problème pratique: poids du point central dans le patch ?
- Quid du type de poids? exponentiel = optimal?

Perspectives et limites

- Cadre "facile": plusieurs images disponibles (videos, bases de données,...); pas de problème de dépendance
- Cadre recherché: une image seulement.
- $f_j = Y_j$ non gelés! non indépendants! possible pré-découpage (Splitting)
- autre(s) solution(s) ???
- Problème pratique: poids du point central dans le patch ?
- Quid du type de poids? exponentiel = optimal?

Perspectives et limites

- Cadre "facile": plusieurs images disponibles (videos, bases de données,...); pas de problème de dépendance
- Cadre recherché: une image seulement.
- $f_j = Y_j$ non gelés! non indépendants! possible pré-découpage (Splitting)
- autre(s) solution(s) ???
- Problème pratique: poids du point central dans le patch ?
- Quid du type de poids? exponentiel = optimal?

Perspectives et limites

- Cadre "facile": plusieurs images disponibles (videos, bases de données,...); pas de problème de dépendance
- Cadre recherché: une image seulement.
- $f_j = Y_j$ non gelés! non indépendants! possible pré-découpage (Splitting)
- autre(s) solution(s) ???
- Problème pratique: poids du point central dans le patch ?
- Quid du type de poids? exponentiel = optimal?

Références bibliographiques I

- Antoni Buades, Bartomeu Coll, and Jean-Michel Morel.
  A review of image denoising algorithms, with a new one.
  Multiscale Model. Simul., 4(2):490–530 (electronic), 2005.
- Florentina Bunea, Alexandre B. Tsybakov, and Marten H. Wegkamp.
  Aggregation for Gaussian regression.
  Ann. Statist., 35(4):1674–1697, 2007.
- Arnak S. Dalalyan and Alexandre B. Tsybakov.
  Aggregation by exponential weighting, sharp oracle inequalities and sparsity.
  In *COLT*, pages 97–111, 2007.

Références bibliographiques II

- Charles Kervrann and Jérôme Boulanger.
  Optimal spatial adaptation for patch-based image denoising.
  IEEETIP, 15(10):'2866–2878, 2006.
- Guillaume Lecué.
  Optimal rates of aggregation in classification under low noise assumption.
  Bernoulli, 13(4):1000–1022, 2007.