An aggregator view of NL-Means

E. Le Pennec and J. Salmon

LPMA - Université Paris Diderot (Paris 7)
SELECT - INRIA Saclay

Aug 2009 - SPIE

NL-Means and aggregation

Setting

- Estimate an image I from a noisy observation Y

$$Y = I + \sigma W \qquad (W \text{ Gaussian white noise})$$

NL-Means and aggregation

Setting

- Estimate an image I from a noisy observation Y
$$Y = I + \sigma W \qquad (W \text{ Gaussian white noise})$$

State of the art

- Classical solution: replace the pixel values by a local average...
- "Patch" based approach: use pixel neighborhoods instead of pixel values.
- NL-Means: Gaussian smoothing in a patch space.

NL-Means and aggregation

Setting

- Estimate an image I from a noisy observation Y
  $$Y = I + \sigma W \qquad (W \text{ Gaussian white noise})$$

State of the art

- Classical solution: replace the pixel values by a local average...
- "Patch" based approach: use pixel neighborhoods instead of pixel values.
- NL-Means: Gaussian smoothing in a patch space.

An aggregator point of view

- Look at the NL-Means approach as a quest for an optimal local kernel, an optimal patch combination.
- Statistical aggregation setting.
- New point of view and new results...

NL-Means and aggregation

Setting

- Estimate an image I from a noisy observation Y
$$Y = I + \sigma W \qquad (W \text{ Gaussian white noise})$$

State of the art

- Classical solution: replace the pixel values by a local average...
- "Patch" based approach: use pixel neighborhoods instead of pixel values.
- NL-Means: Gaussian smoothing in a patch space.

An aggregator point of view

- Look at the NL-Means approach as a quest for an optimal local kernel, an optimal patch combination.
- Statistical aggregation setting.
- New point of view and new results...

Outline

1. Kernel methods and NL-Means
2. Aggregation

Outline

Outline

1. Kernel methods and NL-Means
2. Aggregation
3. Patch based aggregation

Outline

Images, noise and estimate

Image $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$ with $(i_1, i_2) \in [1, N]^2$.
- L_2 (quadratic) norm.

Images, noise and estimate

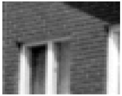


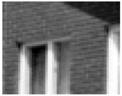
Image $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$ with $(i_1, i_2) \in [1, N]^2$.
- L_2 (quadratic) norm.

Noisy observation

- $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2)$.
- W standard Gaussian i.i.d. noise and σ^2 known variance.
- Other noise possible...

Images, noise and estimate



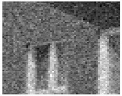


Image $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$ with $(i_1, i_2) \in [1, N]^2$.
- L_2 (quadratic) norm.

Noisy observation

- $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2) \quad .$
- W standard Gaussian i.i.d. noise and σ^2 known variance.
- Other noise possible...

Estimation

- Estimate $I(i_1, i_2)$ by $\widehat{I}(i_1, i_2)$ from Y.
- Non local behavior possible...

Images, noise and estimate

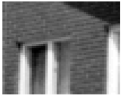
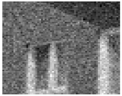
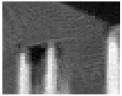




Image $N \times N$

- $I(i_1, i_2) \in \mathbb{R}$ with $(i_1, i_2) \in [1, N]^2$.
- L_2 (quadratic) norm.

Noisy observation

- $Y(i_1, i_2) = f(i_1, i_2) + \sigma W(i_1, i_2) \quad .$
- W standard Gaussian i.i.d. noise and σ^2 known variance.
- Other noise possible...

Estimation

- Estimate $I(i_1, i_2)$ by $\widehat{I}(i_1, i_2)$ from Y.
- Non local behavior possible...

Kernel methods

General kernel method

- Estimate $I(i_1, i_2)$ through a local average
  $\widehat{I}(i_1, i_2) = \sum_{(k_1,k_2)\in[1,N]^2} \lambda_{i_1,i_2,k_1,k_2} Y_{k_1,k_2}$
- The weights $\lambda_{i_1,i_2,k_1,k_2}$ may (will) depend on Y.

Kernel methods

General kernel method

- Estimate $I(i_1, i_2)$ through a local average
  $$\widehat{I}(i_1, i_2) = \sum_{(k_1,k_2)\in[1,N]^2} \lambda_{i_1,i_2,k_1,k_2} Y_{k_1,k_2}$$
- The weights $\lambda_{i_1,i_2,k_1,k_2}$ may (will) depend on Y.

Classic kernel

- $\lambda_{i_1,i_2,k_1,k_2} = \dfrac{K(i_1 - k_1, i_2 - k_2)}{\sum_{k_1',k_2'} K(i_1 - k_1', i_2 - k_2')}$ (no dependency on Y).
- Example: Gaussian kernel $K(i_1, i_2) = e^{-(i_1^2+i_2^2)/2h^2}$.
- Adaptation of the local kernel K (dependency on Y).

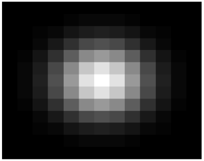
Kernel methods

General kernel method

- Estimate $I(i_1, i_2)$ through a local average
  $\widehat{I}(i_1, i_2) = \sum_{(k_1,k_2)\in[1,N]^2} \lambda_{i_1,i_2,k_1,k_2} Y_{k_1,k_2}$
- The weights $\lambda_{i_1,i_2,k_1,k_2}$ may (will) depend on Y.

Classic kernel

- $\lambda_{i_1,i_2,k_1,k_2} = \dfrac{K(i_1 - k_1, i_2 - k_2)}{\sum_{k_1',k_2'} K(i_1 - k_1', i_2 - k_2')}$ (no dependency on Y).



- Example: Gaussian kernel $K(i_1, i_2) = e^{-(i_1^2+i_2^2)/2h^2}$.
- Adaptation of the local kernel K (dependency on Y).

Data dependent methods

Example of data dependent methods

- $\star$-let thresholding (complex dependency of the weights).
- Bilateral filtering (dependency on pixelwise difference).

Data dependent methods

Example of data dependent methods

- $\star$-let thresholding (complex dependency of the weights).
- Bilateral filtering (dependency on pixelwise difference).

Bilateral filtering

- $\lambda_{i_1,i_2,k_1,k_2} = \dfrac{K(i_1-k_1, i_2-k_2) \times K'(Y(i_1,i_2) - Y(k_1,k_2))}{\sum_{k'_1,k'_2} K(i_1-k'_1, i_2-k'_2) \times K'(Y(i_1,i_2) - Y(k'_1,k'_2))}$
- Gaussian version:

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

- Intuition: average values that are close in both distance and values.
- Issue: pixel value is a too local feature...

Data dependent methods

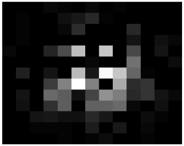
Example of data dependent methods

- $\star$-let thresholding (complex dependency of the weights).
- Bilateral filtering (dependency on pixelwise difference).

Bilateral filtering

- $$\lambda_{i_1,i_2,k_1,k_2} = \frac{K(i_1 - k_1, i_2 - k_2) \times K'(Y(i_1,i_2) - Y(k_1,k_2))}{\sum_{k'_1,k'_2} K(i_1 - k'_1, i_2 - k'_2) \times K'(Y(i_1,i_2) - Y(k'_1,k'_2))}$$
- Gaussian version:

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



- Intuition: average values that are close in both distance and values.
- Issue: pixel value is a too local feature...

Patch based method

Patch

- Patch: less localized version of pixel values.
- Centered patch $P(I)(i_1, i_2)$ of width W:
  $P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2)$ with $-\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$
- Easy reprojection from patch collection $P(I)$ to an image I...

Patch based method

Patch

- Patch: less localized version of pixel values.
- Centered patch $P(I)(i_1, i_2)$ of width W:
  $$P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2) \text{ with } -\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$$
- Easy reprojection from patch collection $P(I)$ to an image I...

Intuition

- Use weights that take into account the patch similarity:

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

- Patch $P(Y)(i_1, i_2)$ to denoise,
- Similar patches, useful: large weights,
- Less similar patches, less useful: small weights,
- Very different patches, useless: no weights.

Patch based method

Patch

- Patch: less localized version of pixel values.
- Centered patch $P(I)(i_1, i_2)$ of width W:
  $P(I)(i_1, i_2)(j_1, j_2) = I(i_1 + j_1, i_2 + j_2)$ with $-\frac{W-1}{2} \leq j_1, j_2 \leq \frac{W-1}{2}$
- Easy reprojection from patch collection $P(I)$ to an image I...

Intuition

- Use weights that take into account the patch similarity:

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

- Patch $P(Y)(i_1, i_2)$ to denoise,
- Similar patches, useful: large weights,
- Less similar patches, less useful: small weights,
- Very different patches, useless: no weights.

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choose a dissimilarity measure D between patches.
- Use a weight $\lambda_{i_1,i_2,k_1,k_2} = \dfrac{K'(D(P_{(i_1,i_2)}, P_{(k_1,k_2)}))}{\sum_{k'_1,k'_2} K'(D(P_{(i_1,i_2)}, P_{(k'_1,k'_2)}))}$
- Use $D(P_{(i_1,i_2)}, P_{(k_1,k_2)}) = \|P_{(i_1,i_2)} - P_{(k_1,k_2)}\|$ to measure the dissimilarity, a Gaussian kernel $K'(x) = \exp(-x^2/\beta)$ and a temperature $\beta = \gamma\sigma^2$.

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choose a dissimilarity measure D between patches.
- Use a weight $\lambda_{i_1,i_2,k_1,k_2} = \dfrac{K'(D(P_{(i_1,i_2)}, P_{(k_1,k_2)}))}{\sum_{k'_1,k'_2} K'(D(P_{(i_1,i_2)}, P_{(k'_1,k'_2)}))}$
- Use $D(P_{(i_1,i_2)}, P_{(k_1,k_2)}) = \|P_{(i_1,i_2)} - P_{(k_1,k_2)}\|$ to measure the dissimilarity, a Gaussian kernel $K'(x) = \exp(-x^2/\beta)$ and a temperature $\beta = \gamma\sigma^2$.

Results

- Fast and efficient method.
- Performance very close to the best denoising method.

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choose a dissimilarity measure D between patches.
- Use a weight $\lambda_{i_1,i_2,k_1,k_2} = \frac{K'(D(P_{(i_1,i_2)}, P_{(k_1,k_2)}))}{\sum_{k'_1,k'_2} K'(D(P_{(i_1,i_2)}, P_{(k'_1,k'_2)}))}$
- Use $D(P_{(i_1,i_2)}, P_{(k_1,k_2)}) = \|P_{(i_1,i_2)} - P_{(k_1,k_2)}\|$ to measure the dissimilarity, a Gaussian kernel $K'(x) = \exp(-x^2/\beta)$ and a temperature $\beta = \gamma\sigma^2$.

Results

- Fast and efficient method.
- Performance very close to the best denoising method.

Variations

- Adapt automatically the search zone. (Kervrann et al.)
- Use a higher order local approximation. (Buades et al.)
- Use a different dissimilarity measure. (Guichard et al.)

NL-Means

NL-Means (Buadès, Coll and Morel)

- Choose a dissimilarity measure D between patches.
- Use a weight $\lambda_{i_1,i_2,k_1,k_2} = \dfrac{K'(D(P_{(i_1,i_2)}, P_{(k_1,k_2)}))}{\sum_{k'_1,k'_2} K'(D(P_{(i_1,i_2)}, P_{(k'_1,k'_2)}))}$
- Use $D(P_{(i_1,i_2)}, P_{(k_1,k_2)}) = \|P_{(i_1,i_2)} - P_{(k_1,k_2)}\|$ to measure the dissimilarity, a Gaussian kernel $K'(x) = \exp(-x^2/\beta)$ and a temperature $\beta = \gamma\sigma^2$.

Results

- Fast and efficient method.
- Performance very close to the best denoising method.

Variations

- Adapt automatically the search zone. (Kervrann et al.)
- Use a higher order local approximation. (Buades et al.)
- Use a different dissimilarity measure. (Guichard et al.)

NL-Means interpretation

Diffusion / Smoothing on the patch manifold

- Intuitive explanation but proof requires some strong assumptions.

NL-Means interpretation

Diffusion / Smoothing on the patch manifold

- Intuitive explanation but proof requires some strong assumptions.

Optimized local kernel

- NL-Means induces a local kernel adapted to the local geometry.

NL-Means interpretation

Diffusion / Smoothing on the patch manifold

- Intuitive explanation but proof requires some strong assumptions.

Optimized local kernel

- NL-Means induces a local kernel adapted to the local geometry.

A best local kernel?

- Can we compare the NL-Means to the best local kernel:

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

NL-Means interpretation

Diffusion / Smoothing on the patch manifold

- Intuitive explanation but proof requires some strong assumptions.

Optimized local kernel

- NL-Means induces a local kernel adapted to the local geometry.

A best local kernel?

- Can we compare the NL-Means to the best local kernel:

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

Outline

1. Kernel methods and NL-Means
2. Aggregation
   - Preliminary estimators and aggregation
   - PAC-Bayesian aggregation
3. Patch based aggregation

Preliminary estimators and aggregation

Model and preliminary estimators

- $Y = I + \sigma W$ of size $N \times N$.
- $\{P_k\}$ set of M preliminary estimators of I (obtained independently).

Preliminary estimators and aggregation

Model and preliminary estimators

- $Y = I + \sigma W$ of size $N \times N$.
- $\{P_k\}$ set of M preliminary estimators of I (obtained independently).

Aggregation

- Estimate I as a weighted average $\widehat{I} = P_\lambda = \sum_k \lambda_k P_k$.
- Aggregation procedure: way to choose λ_k from Y.

Preliminary estimators and aggregation

Model and preliminary estimators

- $Y = I + \sigma W$ of size $N \times N$.
- $\{P_k\}$ set of M preliminary estimators of I (obtained independently).

Aggregation

- Estimate I as a weighted average $\widehat{I} = P_\lambda = \sum_k \lambda_k P_k$.
- Aggregation procedure: way to choose λ_k from Y.

Oracle type inequalities

- Typical results: "Optimal" aggregation amongst a class Λ,

$$E(\|I - \widehat{I}\|^2) \leq C \inf_{\lambda \in \Lambda} \|I - P_\lambda\|^2 + \sigma^2 \mathrm{pen}(\lambda)$$

- C, Λ and pen depend on the procedure.

Preliminary estimators and aggregation

Model and preliminary estimators

- $Y = I + \sigma W$ of size $N \times N$.
- $\{P_k\}$ set of M preliminary estimators of I (obtained independently).

Aggregation

- Estimate I as a weighted average $\widehat{I} = P_\lambda = \sum_k \lambda_k P_k$.
- Aggregation procedure: way to choose λ_k from Y.

Oracle type inequalities

- Typical results: "Optimal" aggregation amongst a class Λ,

$$E(\|I - \widehat{I}\|^2) \leq C \inf_{\lambda \in \Lambda} \|I - P_\lambda\|^2 + \sigma^2 \mathrm{pen}(\lambda)$$

- C, Λ and pen depend on the procedure.

PAC-Bayesian aggregation

PAC-Bayesian aggregation

- Specific aggregation procedure based on exponential weights.
- Defined from a prior π on λ by $\widehat{I} = P_{\lambda_\pi}$ with

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

- For the prior $\pi = \sum_k \delta_k$: $\widehat{I} = \sum_k \frac{e^{-\frac{1}{\beta}\|Y-P_k\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|Y-P_{k'}\|^2}} P_k \quad .$

PAC-Bayesian aggregation

PAC-Bayesian aggregation

- Specific aggregation procedure based on exponential weights.
- Defined from a prior π on λ by $\widehat{l} = P_{\lambda_\pi}$ with

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

- For the prior $\pi = \sum_k \delta_k$: $\widehat{l} = \sum_k \frac{e^{-\frac{1}{\beta}\|Y-P_k\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|Y-P_{k'}\|^2}} P_k \quad .$

Oracle inequality

- Sharp oracle inequality: If $\beta \geq 4\sigma^2$,

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

with $\mathcal{K}(p,\pi)$ the Kullback-Leibler divergence.

PAC-Bayesian aggregation

PAC-Bayesian aggregation

- Specific aggregation procedure based on exponential weights.
- Defined from a prior π on λ by $\widehat{I} = P_{\lambda_\pi}$ with

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

- For the prior $\pi = \sum_k \delta_k$: $\widehat{I} = \sum_k \frac{e^{-\frac{1}{\beta}\|Y - P_k\|^2}}{\sum_{k'} e^{-\frac{1}{\beta}\|Y - P_{k'}\|^2}} P_k \quad .$

Oracle inequality

- Sharp oracle inequality: If $\beta \geq 4\sigma^2$,

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

with $\mathcal{K}(p, \pi)$ the Kullback-Leibler divergence.

Prior choice

Error bound and prior

- $E(\|I-\widehat{I}\|^2) \leq \inf_p \int_{\lambda \in \mathbf{R}^M} \|I-P_\lambda\|^2 dp + \beta \mathcal{K}(p,\pi)$
- Trade-off between a localization of p close to the best "oracle" aggregation P_λ and a proximity with the prior π.
- Prior π should be chosen so that this quantity is small "uniformly"...

Prior choice

Error bound and prior

- $E(\|I-\widehat{I}\|^2) \leq \inf_p \int_{\lambda \in \mathbf{R}^M} \|I - P_\lambda\|^2 dp + \beta \mathcal{K}(p, \pi)$
- Trade-off between a localization of p close to the best "oracle" aggregation P_λ and a proximity with the prior π.
- Prior π should be chosen so that this quantity is small "uniformly"...

Discrete prior

- Prior $\pi = \sum_k \delta_k$: $E(\|I-\widehat{I}\|^2) \leq \inf_k \|I - P_k\|^2 + \beta \log M$.
- As good as the best preliminary estimator...

Prior choice

Error bound and prior

- $E(\|I-\widehat{I}\|^2) \leq \inf_p \int_{\lambda \in \mathbf{R}^M} \|I - P_\lambda\|^2 dp + \beta \mathcal{K}(p, \pi)$
- Trade-off between a localization of p close to the best "oracle" aggregation P_λ and a proximity with the prior π.
- Prior π should be chosen so that this quantity is small "uniformly"...

Discrete prior

- Prior $\pi = \sum_k \delta_k$: $E(\|I-\widehat{I}\|^2) \leq \inf_k \|I - P_k\|^2 + \beta \log M$.
- As good as the best preliminary estimator...

Sparsifying prior

- Prior π: i.i.d. Student or Gaussian mixture (Dalalyan et al.).
- Bound: $E(\|I-\widehat{I}\|^2) \leq \inf_\lambda \|I - P_\lambda\|^2 + C\beta \|\lambda\|_0 \log M$.
- As good as the best "sparse" aggregation...

Prior choice

Error bound and prior

- $E(\|I-\widehat{I}\|^2) \leq \inf_p \int_{\lambda \in \mathbf{R}^M} \|I - P_\lambda\|^2 dp + \beta \mathcal{K}(p, \pi)$
- Trade-off between a localization of p close to the best "oracle" aggregation P_λ and a proximity with the prior π.
- Prior π should be chosen so that this quantity is small "uniformly"...

Discrete prior

- Prior $\pi = \sum_k \delta_k$: $E(\|I-\widehat{I}\|^2) \leq \inf_k \|I - P_k\|^2 + \beta \log M$.
- As good as the best preliminary estimator...

Sparsifying prior

- Prior π: i.i.d. Student or Gaussian mixture (Dalalyan et al.).
- Bound: $E(\|I-\widehat{I}\|^2) \leq \inf_\lambda \|I - P_\lambda\|^2 + C\beta \|\lambda\|_0 \log M$.
- As good as the best "sparse" aggregation...

Outline

1. Kernel methods and NL-Means
2. Aggregation
3. Patch based aggregation
   - Patch based aggregation and theoretical results
   - How to compute the PAC-Bayesian estimate?
   - Numerical results

Patch based aggregation

Localization to patches

- Consider patch $P(Y)(i_1, i_2)$ as observation and patches $P(Y)(k_1, k_2)$ as preliminary estimators.
- Only issue: non independency with the observation $P(Y)(i_1, i_2)$.

Patch based aggregation

Localization to patches

- Consider patch $P(Y)(i_1, i_2)$ as observation and patches $P(Y)(k_1, k_2)$ as preliminary estimators.
- Only issue: non independency with the observation $P(Y)(i_1, i_2)$.

Theorem

- Same flavor than for regular aggregation:

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

Patch based aggregation

Localization to patches

- Consider patch $P(Y)(i_1, i_2)$ as observation and patches $P(Y)(k_1, k_2)$ as preliminary estimators.
- Only issue: non independency with the observation $P(Y)(i_1, i_2)$.

Theorem?

- Same flavor than for regular aggregation:
$$E(\|P(I)(i_1,i_2) - \widehat{P(I)}(i_1,i_2)\|^2) \leq \inf_p \int_{\lambda \in \mathbb{R}^M} \left(\|P(I)(i_1,i_2) - P_\lambda\|^2 + W^2\sigma^2\|\lambda\|^2 \right) dp + \beta\mathcal{K}(p,\pi)$$
- Proof require either some splitting or some more homework...

Patch based priors

- Discrete (NL-Means): selection...
- Sparsifying (Student, Gaussian mixture): sparse kernel optimization!

Patch based aggregation

Localization to patches

- Consider patch $P(Y)(i_1, i_2)$ as observation and patches $P(Y)(k_1, k_2)$ as preliminary estimators.
- Only issue: non independency with the observation $P(Y)(i_1, i_2)$.

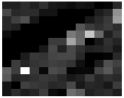
Theorem?

- Same flavor than for regular aggregation:

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

- Proof require either some splitting or some more homework...

Patch based priors



- Discrete (NL-Means): selection...
- Sparsifying (Student, Gaussian mixture): sparse kernel optimization!

SURE and its role

Stein Unbiased Risk Estimate

- $\widehat{r}_\lambda = \|Y - P_\lambda\|^2 - N^2\sigma^2$ is an unbiased estimate of $\|I - P_\lambda\|^2$.
- In the classical aggregation proof, use of $\exp(-\frac{1}{\beta}\widehat{r}_\lambda)$ instead of $\exp(-\frac{1}{\beta}\|Y - P_\lambda\|^2)$ + PAC-Bayesian machinery.
- No modification of the resulting estimate as the bias of $\|Y - P_\lambda\|^2$ does not depend on λ
- Key to generalization to non independent preliminary estimators (Barron and Leung).

SURE and its role

Stein Unbiased Risk Estimate

- $\widehat{r}_\lambda = \|Y - P_\lambda\|^2 - N^2\sigma^2$ is an unbiased estimate of $\|I - P_\lambda\|^2$.
- In the classical aggregation proof, use of $\exp(-\frac{1}{\beta}\widehat{r}_\lambda)$ instead of $\exp(-\frac{1}{\beta}\|Y - P_\lambda\|^2)$ + PAC-Bayesian machinery.
- No modification of the resulting estimate as the bias of $\|Y - P_\lambda\|^2$ does not depend on λ
- Key to generalization to non independent preliminary estimators (Barron and Leung).

Consequence for the patch based aggregation

- $\widehat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_0)\sigma^2$ should be used instead of $\|P(Y)(i_1, i_2) - P_\lambda\|^2$.
- NL-Means: use a weight $\propto \exp(-\frac{1}{\beta}W^2\sigma^2)$ for the central patch (numerical improvement)

SURE and its role

Stein Unbiased Risk Estimate

- $\widehat{r}_\lambda = \|Y - P_\lambda\|^2 - N^2\sigma^2$ is an unbiased estimate of $\|I - P_\lambda\|^2$.
- In the classical aggregation proof, use of $\exp(-\frac{1}{\beta}\widehat{r}_\lambda)$ instead of $\exp(-\frac{1}{\beta}\|Y - P_\lambda\|^2)$ + PAC-Bayesian machinery.
- No modification of the resulting estimate as the bias of $\|Y - P_\lambda\|^2$ does not depend on λ
- Key to generalization to non independent preliminary estimators (Barron and Leung).

Consequence for the patch based aggregation

- $\widehat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_0)\sigma^2$ should be used instead of $\|P(Y)(i_1, i_2) - P_\lambda\|^2$.
- NL-Means: use a weight $\propto \exp(-\frac{1}{\beta}W^2\sigma^2)$ for the central patch (numerical improvement)

PAC-Bayesian estimate and Monte Carlo method

The PAC-Bayesian estimate

- Explicit form: with $\widehat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_0)\sigma^2$,

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

- High dimensional integral similar to some integrals appearing in the Bayesian framework...

PAC-Bayesian estimate and Monte Carlo method

The PAC-Bayesian estimate

- Explicit form: with $\widehat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1-2\lambda_0)\sigma^2$,

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

- High dimensional integral similar to some integrals appearing in the Bayesian framework...

Computing the PAC-Bayesian estimate

- Important issue!
- Monte Carlo method based on a Langevin diffusion equation.
- Approximate values only... but sufficient precision.
- Some convergence issues still under investigation.
- Patch preselection seems to help...

PAC-Bayesian estimate and Monte Carlo method

The PAC-Bayesian estimate

- Explicit form: with $\widehat{r}_\lambda = \|P(Y)(i_1, i_2) - P_\lambda\|^2 - W^2(1 - 2\lambda_0)\sigma^2$,

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

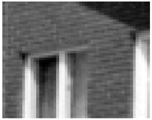
- High dimensional integral similar to some integrals appearing in the Bayesian framework...

Computing the PAC-Bayesian estimate

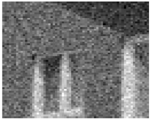
- Important issue!
- Monte Carlo method based on a Langevin diffusion equation.
- Approximate values only... but sufficient precision.
- Some convergence issues still under investigation.
- Patch preselection seems to help...

Numerical results

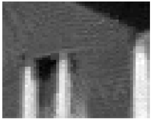
Numerical results



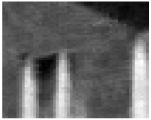
Original



Noisy (22.06 dB)



NL Means (29.69 dB)

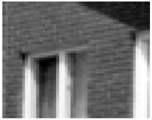


PAC-Bayesian (29.69 dB)

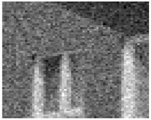
Experimental setting

- Comparison with classic NL-Means with $\gamma = 12$.
- PAC-Bayesian aggregation with Student prior.

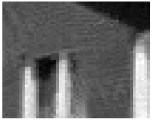
Numerical results



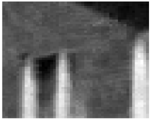
Original



Noisy (22.06 dB)



NL Means (29.69 dB)



PAC-Bayesian (29.69 dB)

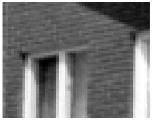
Experimental setting

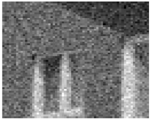
- Comparison with classic NL-Means with $\gamma = 12$.
- PAC-Bayesian aggregation with Student prior.

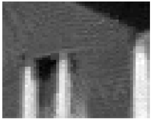
Results

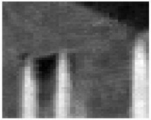
- Results similar to those obtained with NL-Means...
- with less hyperparameter dependency and room for improvement.

Numerical results


Original


Noisy (22.06 dB)


NL Means (29.69 dB)

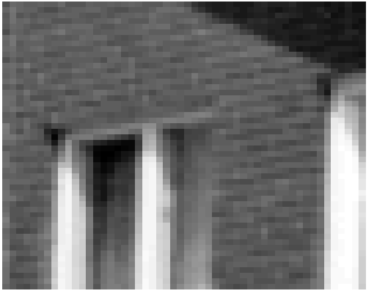

PAC-Bayesian (29.69 dB)

Experimental setting

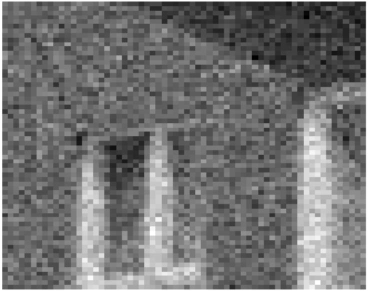
- Comparison with classic NL-Means with $\gamma = 12$.
- PAC-Bayesian aggregation with Student prior.

Results

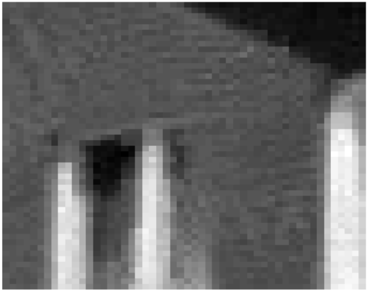
- Results similar to those obtained with NL-Means...
- with less hyperparameter dependency and room for improvement.



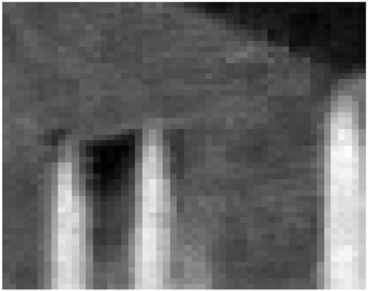
Original



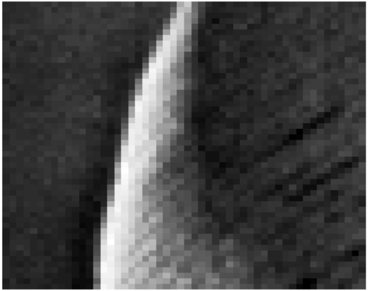
Noisy (22.06 dB)



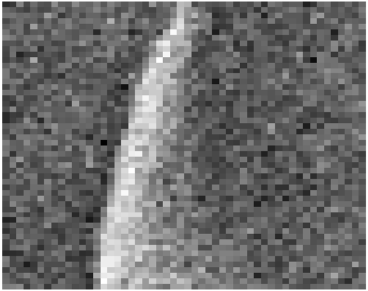
NL Means (29.69 dB)



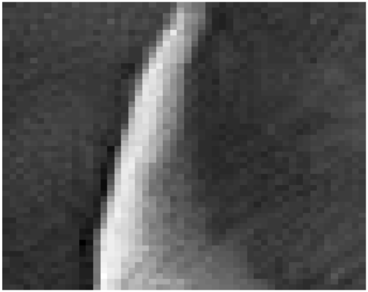
PAC-Bayesian (29.69 dB)



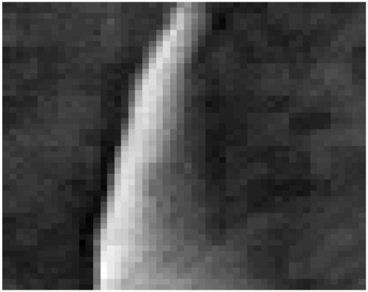
Original



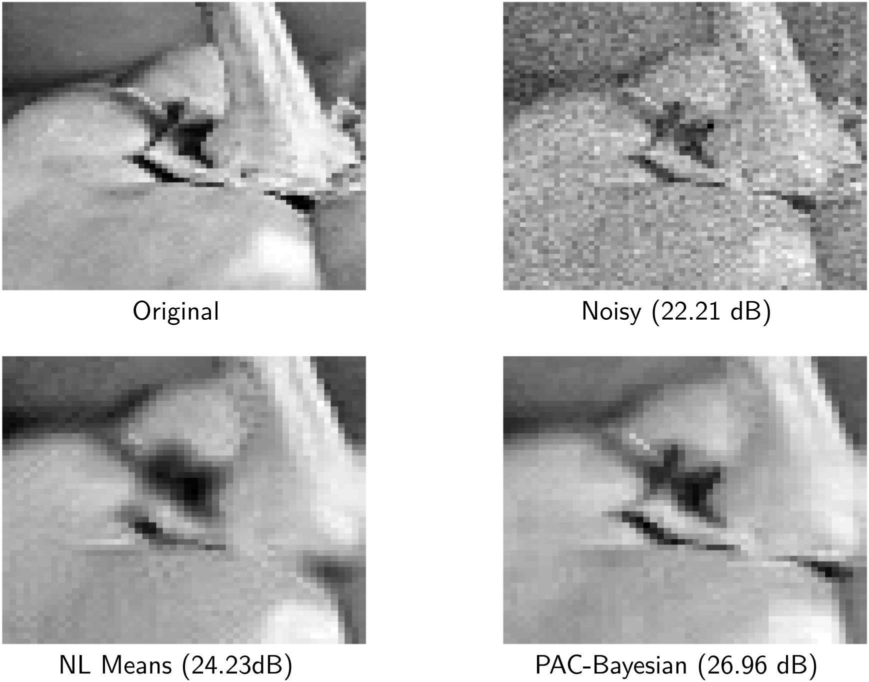
Noisy (22.28 dB)



NL Means (31.59 dB)



PAC-Bayesian (30.78 dB)


Original
Noisy (22.21 dB)
NL Means (24.23dB)
PAC-Bayesian (26.96 dB)

Conclusion

Statistical aggregation: a novel point of view on the NL-Means

- A new look on the exponential weights and the L_2 patch dissimilarity measure.
- A new procedure which performs as well as the NL-Means but with (some) theoretical control.
- A heuristic for the weight of the central patch in the classical NL-Means.

Conclusion

Statistical aggregation: a novel point of view on the NL-Means

- A new look on the exponential weights and the L_2 patch dissimilarity measure.
- A new procedure which performs as well as the NL-Means but with (some) theoretical control.
- A heuristic for the weight of the central patch in the classical NL-Means.

Work in progress...

- Extend the theorem to the fully dependent case,
- How to accelerate the Monte Carlo chain convergence?,
- Best choice for the prior,
- Use of sparse representation for the kernel,
- ...

Conclusion

Statistical aggregation: a novel point of view on the NL-Means

- A new look on the exponential weights and the L_2 patch dissimilarity measure.
- A new procedure which performs as well as the NL-Means but with (some) theoretical control.
- A heuristic for the weight of the central patch in the classical NL-Means.

Work in progress...

- Extend the theorem to the fully dependent case,
- How to accelerate the Monte Carlo chain convergence?,
- Best choice for the prior,
- Use of sparse representation for the kernel,
- ...