

AGRÉGATION D'ESTIMATEURS POUR LE DÉBRUITAGE D'IMAGE

Joseph Salmon & Erwan Le Pennec

*Laboratoire de Probabilité et Modèles Aléatoires, CNRS-UMR 7599,
Université Paris 7-Diderot
175 rue du Chevaleret 75013 Paris*

MOTS CLEFS: Modèles semi et non paramétriques, Statistique mathématique, Agrégation, NL-Means, Diffusion.

RÉSUMÉ: *Dans ce travail sur le débruitage d'image, nous présentons un nouveau type d'estimateur par patches inspiré de la méthode NL-Means proposée par Buades, Coll et Morel (2005) et des techniques PAC-Bayésienne étudiées par Dalalyan et Tsybakov (2007). Nous présentons le cadre théorique adapté pour ces estimateurs, leur implémentation ainsi que leurs performances théoriques et pratiques.*

ABSTRACT: *In this work on image denoising, we present a novel type of patch base estimator inspired by the Non Local Means proposed by Buades, Coll et Morel and the PAC-Bayesian techniques studied by Dalalyan and Tsybakov. We present the theoretical framework adapted to these estimators and deal with both theoretical and practical performances of these estimators.*

Le débruitage d'images numériques, l'estimation d'images corrompues par un bruit, est un thème classique à la frontière du traitement du signal et des statistiques. Inspirés par les travaux de Buades, Coll et Morel (2005) sur les Non Local Means (NL-means) et les techniques PAC-Bayésienne étudié par Dalalyan et Tsybakov (2007), nous proposons un nouveau type d'estimateur utilisant des patches qui repose sur des techniques d'agrégation d'estimateurs.

Le modèle considéré est le modèle classique de régression sur une grille fixe: pour chaque pixel i d'une image de taille $n \times n$, on observe

$$Y_i = f(i) + \varepsilon_i \quad ,$$

où $f(i)$ est la vraie valeur et ε_i , une suite i.i.d. de gaussienne centrée de variance σ^2 connue. Vectoriellement, $Y = f + \varepsilon$. On recherche alors des bons estimateurs de f en tout point de la grille à partir de l'observation de Y , la perte étant mesurée par la norme ℓ^2 . De nombreux estimateurs ont été proposés dans ce cadre: estimateurs à noyaux, seuillage dans des bases d'ondelettes ou des représentations géométriques... Nous étudions ici une classe d'estimateurs différente : celle des estimateurs basés sur des moyennes de patches voisins.

Le premier estimateur de ce type a été proposé par Buades, Coll et Morel en 2005 sous le nom de NL-Means. Il est construit de la manière suivante. On place autour d'un pixel i_0 une petite fenêtre (que l'on choisira en pratique de taille 5×5), et l'on appelle patch bruité P_{i_0} la restriction de Y à celle-ci. On considère également dans un voisinage de ce pixel, M patches $P(i_0, 1), \dots, P(i_0, M)$ de même forme. On cherche alors à estimer la restriction de f à la fenêtre centrée autour de i_0 par une combinaison linéaire $\hat{P}(i_0)$ des patches voisins,

$$\hat{P}(i_0) = \sum_{j=1}^M \lambda(i_0, j) P(i_0, j) \quad ,$$

où les poids $\lambda(i_0, j)$ dépendent de la similarité du patch correspondant à $P(i_0, j)$ avec le patch central. De manière plus précise, les poids associés à la méthode des NL-means sont proportionnels à des exponentielles en la distance ℓ^2 des patches et somment à 1:

$$\lambda(i_0, j) = \frac{\exp(-\beta^{-1} \|P_{i_0} - P(i_0, j)\|^2)}{\sum_{k=1}^M \exp(-\beta^{-1} \|P_{i_0} - P(i_0, k)\|^2)} \quad .$$

Le paramètre β , dit paramètre de température, permet de régler la mesure de similarité et joue le rôle de la fenêtre dans les méthodes à noyaux. Il contrôle donc le niveau de lissage. Cette méthode simple donne de très bon résultats pratiques mais il n'existe pas de preuve théorique de son efficacité.

Dans le travail de Dalalyan et Tsybakov (2007) sur l'agrégation d'estimateurs on voit apparaître une construction similaire pour laquelle ils obtiennent des résultats théoriques. Dans le même modèle statistique, ils se donnent une famille $P(1), \dots, P(M)$ de pré-estimateurs et cherche à estimer f à partir d'une combinaison linéaire P_λ de ceux-ci

$$P_\lambda = \sum_{i=1}^M \lambda(i) P(i) \quad .$$

Pour cela, ils fixent une loi a priori π sur \mathbb{R}^M et définissent l'estimateur PAC-Bayésien associé \hat{f}_π par

$$\hat{f}_\pi = \frac{\int_{\mathbb{R}^M} P_\lambda \exp(-\beta^{-1} \|Y - P_\lambda\|^2) \pi(d\lambda)}{\int_{\mathbb{R}^M} \exp(-\beta^{-1} \|Y - P_\lambda\|^2) \pi(d\lambda)} \quad .$$

La similarité entre cette formule et celle des NL-Means est frappante. De manière plus précise, on retrouve exactement les NL-Means en restreignant l'estimation au voisinage autour de i_0 , en choisissant pour pré-estimateurs les patches voisins $P(i) = P(i_0, i)$ et pour π la loi discrète uniforme sur les pré-estimateurs, soit $\pi = 1/M \sum_{i=1}^M \delta_{e_i}$ où e_1, \dots, e_M est la base canonique de \mathbb{R}^M , et δ_{e_i} est la mesure de Dirac associée. Lorsque les pré-estimateurs sont indépendants de Y , ils démontrent que si $\beta \geq 4\sigma^2$ alors

$$E \|f - \hat{f}_\pi\|^2 \leq \inf_p \int \|f - P_\lambda\|^2 p(d\lambda) + \beta \mathcal{K}(p, \pi) \quad ,$$

où p parcourt l'ensemble des probabilités sur \mathbb{R}^M et $\mathcal{K}(p, \pi)$ est la divergence de Kullback-Leibler entre p et π . Le risque de l'estimateur est plus petit que le risque de toute combinaison des estimateurs P_λ à un terme près mesurant la distance entre ces combinaisons et l'a priori utilisé. L'hypothèse d'indépendance n'est pas vérifiée dans le cas d'utilisation des patchs mais des travaux en cours suggèrent qu'une inégalité de la forme

$$E\|f - \hat{f}_\pi\|^2 \leq \inf_p \int \|f - f_\lambda\|^2 + \sigma^2|\lambda|^2 p(d\lambda) + \beta\mathcal{K}(p, \pi) \quad ,$$

est vérifiée. Celle-ci signifie que le risque de l'estimateur est majorée par toute le risque de toute combinaison d'estimateur à noyaux à un terme mesurant la distance entre l'a priori et cette combinaison près; on fait "aussi bien" que le meilleur des noyaux possibles au terme de divergence près.

Deux questions se posent alors: Comment choisir l'a priori π pour assurer que le membre de droite de l'inégalité précédente soit proche du minimum sans le terme de divergence? Comment calculer en pratique cet estimateur?

L'efficacité de la méthode repose en grande partie sur le choix de l'a priori sur les pré-estimateurs. L'avantage des NL-means est la simplicité de leur calculs. Le choix d'un a priori uniforme sur les valeurs $P(i_0, 1), \dots, P(i_0, M)$ permet de ne pas avoir d'intégrale à calculer. Dalalyan et Tsybakov ont montré l'intérêt d'un a priori symétrique à queue lourde (par exemple une loi 3-Student) lorsque le meilleur estimateur est une combinaison parcimonieuse des pré-estimateurs. Dans le cadre du débruitage, une approche "noyau" conduit à regarder un a priori gaussien, ou un mélange gaussien.

Dalalyan et Tsybakov (2009) suggèrent une méthode de type Langevin-Monte Carlo, analogue continu des méthodes MCMC classiques, pour le calcul de l'estimateur. Cette méthode est basée sur le fait que si V est une fonction suffisamment régulière (et sous quelques conditions techniques), la solution L de la diffusion régie par l'équation de Langevin,

$$dL_t = \nabla V(L_t)dt + \sqrt{2}dW_t, \quad L_0 = \lambda_0, \quad t \geq 0, \quad ,$$

avec $\lambda_0 \in \mathbb{R}^M$, W_t un mouvement Brownien M -dimensionnel, a pour distribution stationnaire $p_V(\lambda) \propto e^{V(\lambda)}$, $\lambda \in \mathbb{R}^M$. En prenant

$$V(\lambda) = -\beta^{-1}\|Y - f_\lambda\|_n^2 - \log(\pi(\lambda)) \quad ,$$

la diffusion converge vers la distribution apparaissant dans les estimateurs PAC-Bayésien. L'intégrale $\int_{\mathbb{R}^M} \lambda p_V(d\lambda)$ s'obtient alors comme limite des intégrales le long d'une trajectoire $\bar{L}_T = \frac{1}{T} \int_0^T L_t dt$.

Cette intégrale est approchée numériquement grâce à la résolution d'un schéma d'Euler discret à pas constant associé à la diffusion. On définit $L_0^E = \lambda_0$ et pour $k = 1, \dots, [T/h] - 1$,

$$L_{k+1}^E = L_k^E + h\nabla V(L_k^E) + \sqrt{2h}W_k$$

où W_1, W_2, \dots sont des gaussiennes standardisées et i.i.d de \mathbb{R}^M . On approche alors naturellement \bar{L}_T par:

$$\hat{L}_{T,h}^E = \frac{1}{\lfloor T/h \rfloor} \sum_{k=0}^{\lfloor T/h \rfloor - 1} L_k^E.$$

Cette méthode permet de calculer un estimateur proche de l'estimateur théorique.

Nous avons comparé les différents a priori sur des images usuelles et nous montrerons les variations de performances selon le type d'a priori choisi.

Bibliographie

- [1] A. Buades, B. Coll, and J-M. Morel (2005) "A review of image denoising algorithms, with a new one", *Multiscale Model.Simul.*, vol. 4, no. 2, pp. 490-530.
- [2] A. Dalalyan and A. Tsybakov (2007) "Aggregation by exponential weighting, sharp oracle inequalities and sparsity" in *COLT*, pp. 97-111.
- [3] A. Dalalyan and A. Tsybakov (2009) "PAC-Bayesian bounds for the expected error of aggregation by exponential weights" to appear.