

Débruitage d'images, a-t-on atteint l'optimum ?

SIMON ABELARD, SIMON AMAR-NIVAUT, EDOUARD STRICKLER
ENCADRANTS : JEAN-MICHEL MOREL, MAURICIO DELBRACIO

2 juillet 2012

Introduction

L'observation d'un signal par un appareil de mesure s'accompagne souvent d'une information non désirée qui perturbe cette première : c'est ce qu'on appelle le bruit. Il se traduit de différentes manières, par des "scratches" dans un enregistrement sonore, ou par une sorte de grain sur une image, par exemple. De façon générique on notera $\tilde{u} = u + n$ l'observation bruitée \tilde{u} de l'image originale u , et notre but sera évidemment de retrouver u en ne connaissant que \tilde{u} .

L'essentiel du bruit n est dû au capteur : celui-ci doit récolter un maximum de photons pour obtenir la meilleure image possible, or le nombre de ces particules de lumière qu'il reçoit peut être limité, notamment à cause d'un éclairage trop faible ou d'un temps d'exposition trop court. Le bruit est une variable aléatoire, dont il faudra se donner un modèle afin de pouvoir comparer les différents algorithmes de débruitage qui seront proposés tout au long de cet article.

La première partie consiste en un tour d'horizon des philosophies menant aux premières méthodes de débruitage. Nous proposons ensuite le critère du minimax, qui nous donne une optimalité à laquelle nous confrontons quelques algorithmes. Enfin, nous exposons un algorithme récent et vérifions qu'il est bien optimal au sens du minimax.

1 Premières méthodes de débruitage

1.1 Quel modèle pour le bruit ?

En réalité, pour chaque pixel i de l'image, la valeur bruitée $\tilde{u}(i)$ est une variable aléatoire qui suit une loi de Poisson de paramètre $u(i)$. Cependant, pour des valeurs suffisamment grandes de $\tilde{u}(i)$, la loi normale $\mathcal{N}(u(i), u(i))$ est une très bonne approximation de la loi de Poisson évoquée. Ceci légitime un modèle gaussien pour le bruit, que nous supposons de plus centré et de variance σ^2 constante sur l'ensemble de l'image. Autrement dit, nous aurons pour chaque pixel $n(i) \sim \mathcal{N}(0, \sigma^2)$. Certaines astuces qui dépassent le cadre du présent papier permettent de se ramener à quelque chose de très proche de ce modèle.

1.2 Méthodes basées sur des patches

La plupart des algorithmes modernes de débruitage suivent une méthode qui utilisent des patches, c'est-à-dire des petits morceaux de l'image. Ainsi, au lieu d'appliquer des méthodes pixel par pixel, elles seront appliquées patch par patch, et la valeur débruitée d'un pixel sera obtenue en faisant des moyennes des valeurs de chaque patch qui le contient. Dans la majorité des algorithmes, ces patches seront carrés et parallèles aux frontières de l'image. Un algorithme récent, dont nous montrerons l'optimalité théorique, propose cependant d'utiliser des patches rectangulaires et orientés différemment pour tenter d'améliorer ses performances.

Dans la suite, on notera P le patch de l'image à estimer et \tilde{P} sa version bruitée. Le modèle considéré du bruit, et l'hypothèse de l'indépendance de celui-ci en chaque pixel, nous donne :

$$\mathbb{P}(\tilde{P} | P) = c \exp\left(-\frac{\|P - \tilde{P}\|^2}{2\sigma^2}\right),$$

formule que nous allons mettre à profit dans le prochain paragraphe.

1.3 Distribution gaussienne et analyse en composantes principales

L'idée de cette première méthode est de maximiser $\mathbb{P}(P | \tilde{P})$. Nous connaissons l'autre probabilité conditionnelle, et donc, grâce à la formule de Bayes :

$$\mathbb{P}(P | \tilde{P}) = \frac{\mathbb{P}(\tilde{P} | P)\mathbb{P}(P)}{\mathbb{P}(\tilde{P})}.$$

Nous ne possédons malheureusement pas de modèle de probabilité pour P . On va cependant se ramener à un échantillonnage gaussien de P . C'est un méthode habituelle lorsque l'on dispose de plusieurs échantillons mais pas de la loi de probabilité. Dans le cas des patches Q similaires à un patch P donné, ce modèle est pertinent, eu égard à tous les facteurs aléatoires pouvant expliquer les différences entre P et Q : les détails, la texture, la luminosité... On justifie ce choix par le théorème central limite. Bref, on suppose que les patches similaires à P suivent une loi gaussienne, de moyenne \bar{P} et de matrice de covariance C_P . Ceci peut se traduire de la façon suivante :

$$\mathbb{P}(Q) = c \exp\left(-\frac{(Q - \bar{P})^t C_P^{-1} (Q - \bar{P})}{2}\right).$$

Etant donné \tilde{P} , maximiser $\mathbb{P}(P | \tilde{P})$ revient donc à maximiser $\mathbb{P}(\tilde{P} | P)\mathbb{P}(P)$, soit à maximiser

$$\exp\left(-\frac{\|P - \tilde{P}\|^2}{2\sigma^2}\right) \exp\left(-\frac{(P - \bar{P})^t C_P^{-1} (P - \bar{P})}{2}\right),$$

puis, à minimiser

$$\frac{\|P - \tilde{P}\|^2}{\sigma^2} + (P - \bar{P})^t C_P^{-1} (P - \bar{P}).$$

L'inconvénient est que nous n'avons pas d'observation parfaite de P , pas plus que des patchs similaires à P . Cependant, nous connaissons \tilde{P} , et on peut avoir une observation des patchs similaires à \tilde{P} . En rappelant que $\tilde{P} = P + n$, où n est gaussienne de matrice de covariance $\sigma^2 I$, on a

$$C_{\tilde{P}} = C_P + \sigma^2 I,$$

et

$$\mathbb{E}(\tilde{P}) = \tilde{P}.$$

Il faut chercher des patchs similaires à \tilde{P} à une distance suffisante pour inclure les patchs similaires à P . Cette distance cependant ne doit être trop grande, pour éviter les valeurs aberrantes. L'idée est donc de prendre une distance juste supérieure à la distance moyenne créée par le bruit. On remplace alors dans l'expression ci-dessus C_P par $C_{\tilde{P}}$ et \tilde{P} par $\bar{\tilde{P}}$.

$$\max_P \mathbb{P}(P | \tilde{P}) = \min_P \left[\frac{\|P - \tilde{P}\|^2}{\sigma^2} + (P - \bar{\tilde{P}})^t (C_{\tilde{P}}^{-1} - \sigma^2 I) (P - \bar{\tilde{P}}) \right].$$

L'argument du minimum \hat{P} nous donne donc un algorithme de débruitage, qui maximisera $\mathbb{P}(P | \tilde{P})$:

$$\hat{P}_1 = \bar{\tilde{P}} + [C_{\tilde{P}} - \sigma^2 I] C_{\tilde{P}}^{-1} (\tilde{P} - \bar{\tilde{P}}),$$

qui permet d'effectuer un débruitage si l'on connaît $\bar{\tilde{P}}$ et $C_{\tilde{P}}$.

Utilisation pratique On pose $\mathcal{P}(\tilde{P})$ l'ensemble des patchs \tilde{Q} similaire à \tilde{P} , obtenus avec un seuil de tolérance tel qu'on peut supposer qu'ils représentent des versions bruitées de P . En notant N le cardinal de cet ensemble, et en vertu de la loi forte des grands nombres (on rappelle que $\bar{\tilde{P}}$ représente la valeur moyenne des \tilde{Q} similaire à \tilde{P} , et que $C_{\tilde{P}}$ est la matrice de covariance de la loi gaussienne suivie par les \tilde{Q}) :

$$C_{\tilde{P}} \simeq \frac{1}{N-1} \sum_{\tilde{Q} \in \mathcal{P}(\tilde{P})} (\tilde{Q} - \bar{\tilde{P}})(\tilde{Q} - \bar{\tilde{P}})^t,$$

et

$$\bar{\tilde{P}} \simeq \frac{1}{N} \sum_{\tilde{Q} \in \mathcal{P}(\tilde{P})} \tilde{Q}.$$

Cependant, il est préférable de travailler en deux étapes :

- Dans une première, on utilisera l'algorithme tel quel pour effectuer un premier débruitage de \tilde{P}
- Dans une seconde, on utilise les patchs que l'on vient de débruiter pour donner de nouvelles approximations de $C_{\tilde{P}}^1$ et de $\bar{\tilde{P}}_1$ (autrement dit, on remplace $\mathcal{P}(\tilde{P})$ par un nouvelle ensemble de patchs similaires, qui auront déjà subis un débruitage.)

Ceci permet donc une deuxième approximation de P :

$$\hat{P}_2 = \bar{\tilde{P}}_1 + [C_{\tilde{P}}^1 - \sigma^2 I] (C_{\tilde{P}}^1)^{-1} (\tilde{P} - \bar{\tilde{P}}_1).$$

Cette méthode revient à effectuer une analyse en composante principale et à appliquer un filtre de Wiener.

Analyse en composantes principales Appliquer une analyse en composantes principales (ACP) à une image revient à en chercher les directions privilégiées, puis à projeter sur celles-ci. Plus exactement, pour une image u de taille $n \times n$, dont nous noterons P_i les patches mis sous forme de vecteurs, on considère la matrice de covariance, $M = \frac{1}{n} \sum_i P_i^t \cdot P_i$, que l'on va diagonaliser. Nous allons ensuite ordonner les valeurs propres par ordre décroissant d'importance : ce sont les composantes principales. Les plus faibles seront supprimées en appliquant un filtre de Wiener, détaillé en annexe : en effet, ce sont elles qui contiennent le plus de bruit, tandis que les plus importantes constituent la majorité du signal. Enfin, les patches sont reconstruits à partir de cette base, et une image débruitée sera ainsi obtenue.

1.4 Un premier résultat d'optimalité pour un premier critère

L'une des difficultés du débruitage est de pouvoir comparer l'efficacité de deux algorithmes différents, afin de choisir le meilleur, voire d'en trouver un optimal. Si le résultat visuel est évidemment prépondérant, il ne permet pas de comparer de manière systématique deux méthodes distinctes. Pour ce faire, il faudra se donner un critère, lequel sera souvent la mesure d'une quantité dont on cherchera un extremum.

Une idée naturelle consiste à évaluer l'erreur commise par l'algorithme de débruitage par rapport à l'image originale. Ainsi, on calculera l'erreur quadratique de moyenne (*MSE* pour *Mean Square Error*), qui pour un patch débruité \hat{P} est donnée par

$$MSE(\hat{P}) = \int \mathbb{P}(\tilde{P}) \left(\int \mathbb{P}(P | \tilde{P}) \| P - \hat{P} \|^2 dP \right) d\tilde{P}.$$

Le terme $\mathbb{P}(P | \tilde{P}) \| P - \hat{P} \|^2$ correspond à l'erreur commise en probabilité. On intègre alors sur tous les patches P possibles ; ce qui nous donne une première erreur qui dépend encore de \tilde{P} . On fait alors une moyenne sur toutes les valeurs de \tilde{P} , et on obtient la quantité désirée.

À présent, notre objectif est de minimiser cette erreur, et ainsi trouver l'algorithme qui, en moyenne, sera le plus proche de l'image recherchée. Pour ce faire, on minimise simplement chaque terme de l'intégrale en \tilde{P} , à savoir :

$$\mathbb{P}(\tilde{P}) \int \mathbb{P}(P | \tilde{P}) \| P - \hat{P} \|^2 dP.$$

On dérive cette expression par rapport à \hat{P} , et on égale à 0 :

$$\int \mathbb{P}(P | \tilde{P}) P dP = \hat{P} \int \mathbb{P}(P | \tilde{P}) dP,$$

ce qui donne :

$$\hat{P} = \frac{\int \mathbb{P}(P | \tilde{P}) P dP}{\int \mathbb{P}(P | \tilde{P}) dP},$$

soit

$$\hat{P} = \int \mathbb{P}(P | \tilde{P}) P dP,$$

puisque $\int \mathbb{P}(P | \tilde{P})dP = 1$. Une nouvelle utilisation de la formule de Bayes nous permet d'arriver à cette formule de l'estimateur :

$$\hat{P} = \int \frac{\mathbb{P}(\tilde{P} | P)\mathbb{P}(P)}{\mathbb{P}(\tilde{P})}PdP.$$

Dans l'état, la formule n'est pas applicable, l'intégrale n'étant absolument pas calculable, faute de connaître $\mathbb{P}(P)dP$ et $\mathbb{P}(\tilde{P})$. Si dans le paragraphe précédent, nous nous étions donné une distribution gaussienne pour pallier ce problème, nous allons maintenant effectuer le calcul par un échantillonnage des patches naturels P - autrement dit, on va remplacer l'intégrale par une somme, et les termes $\mathbb{P}(P)dP$ et $\mathbb{P}(\tilde{P})$ par $\frac{1}{N}$ et $\frac{1}{N} \sum \mathbb{P}(\tilde{P} | P_i)$ respectivement, où N est le nombre d'échantillons P_i tirés. D'où l'estimation suivante de \hat{P} :

$$\hat{P} = \frac{\frac{1}{N} \sum_{i=1}^N \mathbb{P}(\tilde{P} | P_i)P_i}{\frac{1}{N} \sum_{i=1}^N \mathbb{P}(\tilde{P} | P_i)}.$$

Cette méthode générale porte le nom de *Non-Local Means* (NLM), puisque les patches P_i utilisés pour le calcul ne sont pas forcément les plus proches de \tilde{P} , mais peuvent être choisis dans l'ensemble de l'image, voire au-delà - dans d'autres images. Il reste à définir un ensemble de patches P_i convenable. Traditionnellement, on se donne une certaine zone de l'image, centrée au pixel d'intérêt, dans laquelle on prendra tous les patches possibles. Une autre idée, issue des méthodes de Monte-Carlo pour le calcul intégral, serait de considérer un immense ensemble de patches issus d'images au bruit négligeable. Cette procédure, dite *shotgun*, a été proposée et testée par Anat Levin et Boaz Nadler, qui, grâce à la considérable source de données que représente le net, ont pu avoir à leur disposition pas moins de 20000 images, soit l'équivalent de près de 10 milliards de patches, pour leur débruitage.

Si de tels algorithmes seront très bons dans une majorité des cas, ils pourraient s'avérer catastrophiques sur certaines images. C'est pour cette raison que nous allons introduire un nouveau critère, qui nous donnera l'assurance de ne pas avoir ces exceptions pour le moins fâcheuses.

2 Théorie minimax

Le risque bayésien offre un risque moyen, mais n'offre aucune garantie quant aux écarts possibles à la moyenne. La connaissance d'un risque bayésien n'écarte donc en rien la possibilité d'un résultat désastreux dans certains cas rares. L'idée du minimax est de combler cette lacune en étudiant systématiquement les pires cas possibles. La théorie minimax donne par conséquent des résultats extrêmement forts, qui consistent en des bornes inaliénables sur la performance des algorithmes. Cette théorie nous permettra en effet de donner un critère d'optimalité absolue pour nos algorithmes.

2.1 Premières définitions

Risque, biais et variance Soit u l'image sans bruit, et \hat{u} l'image débruitée. On définit comme suit le risque quadratique :

$$R(\hat{u}, u) = \mathbb{E} (\|\hat{u} - u\|_2^2),$$

que l'on pourra avec profit décomposer en la somme du biais et de la variance :

$$\begin{aligned} b^2 &= \|\mathbb{E}(\hat{u}) - u\|_2^2, \\ \sigma^2 &= \mathbb{E} (\|\hat{u} - \mathbb{E}(\hat{u})\|_2^2). \end{aligned}$$

Risque minimax et estimateur du minimax On définit comme suit le *risque minimax* de l'algorithme \mathcal{A} sur la classe \mathcal{C} :

$$R(\mathcal{A}) = \sup_{u \in \mathcal{C}} \mathbb{E} (\|\hat{u}_{\mathcal{A}} - u\|_2^2).$$

On cherche donc

$$\mathcal{A} = \arg \min_{\mathcal{A} \in \mathcal{A}} R(\mathcal{A}),$$

avec \mathcal{A} l'ensemble de **tous** les algorithmes possibles et $\hat{u}_{\mathcal{A}}$ l'estimateur hérité de l'algorithme \mathcal{A} . On cherche ainsi l'algorithme dont le pire cas est le moins catastrophique. Reste maintenant à choisir une classe \mathcal{C} opportune.

Classe horizon et images horizon Dans la mesure où les algorithmes de débruitage procèdent à des moyennes afin d'éliminer le bruit, il apparaît clairement que les bords seront les écueils entre lesquels un bon algorithme devra savoir louvoyer. D'où l'idée d'étudier leurs performances sur des images à la fois simples à concevoir théoriquement, mais posant de réelles difficultés pratiques : ce que nous appellerons les images horizon. Soit

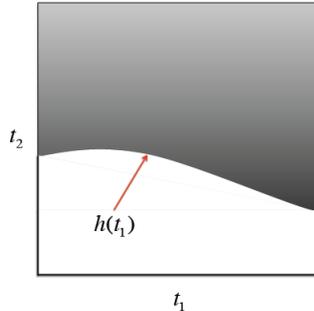
$$\Sigma(\beta, L) = \left\{ g \in C^k([0, 1]) \mid \forall x, x' \in [0, 1], |g^{(k)}(x) - g^{(k)}(x')| \leq L|x - x'|^{\beta-k} \right\},$$

avec k la partie entière de β .

La classe horizon β , notée $\mathcal{C}(\beta)$, est alors l'ensemble des fonctions g de $\Sigma(\beta, L)$ dont le graphe est compris dans un segment strictement inclus dans $[0, 1]$. On définit alors une image horizon comme étant de la forme

$$u : (x, y) \rightarrow \mathbf{1}_{h(x) \leq y}$$

avec h dans une certaine classe horizon (on omet en pratique le β puisqu'on le prend presque systématiquement égal à 2).



Généralisation du risque $R(\hat{\theta}_n, \psi_n, \theta) = \mathbb{E}_\theta \left(w(\psi_n^{-1}d(\hat{\theta}_n, \theta)) \right)$ où w est appelée fonction de perte et ψ_n une suite normalisatrice strictement positive dont le bon choix sera l'enjeu de ce qui suit.

Le risque maximal (qui ne dépend plus de θ) sera $r(\hat{\theta}_n, \psi_n) = \sup_{\theta \in \Sigma} R(\hat{\theta}_n, \theta, \psi_n)$

Vitesse de convergence On dit que ψ_n est un *minorant de la vitesse de convergence* pour l'ensemble Σ et la pseudo-métrique d s'il existe un réel positif C_0 tel que :

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} r(\hat{\theta}_n, \psi_n) \geq C_0.$$

Le principe de cette définition consiste à étudier les limites à la vitesse de convergence de notre estimateur vers la quantité estimée.

On dit que ψ_n est une *vitesse minimax de convergence* pour l'ensemble Σ et la pseudo-métrique d si ψ_n est un minorant de la vitesse de convergence et s'il existe C_1 un réel et θ_n^* un estimateur tel que :

$$\limsup_{n \rightarrow \infty} r(\theta_n^*, \psi_n) \leq C_1.$$

Dans ce cas, on dit que θ_n^* est un *estimateur optimal*, et on définira la *risque minimax* comme suit : $r_n^* = \inf_{\hat{\theta}_n} r(\hat{\theta}_n, \psi_n)$.

On remarque que ces notions sont relatives à la fonction de perte. Pour $w(x) = x^2$ on parlera de S-vitesse de convergence, pour $w(x) = |x|$ on parlera de A-vitesse de convergence et si w est une indicatrice, on parlera de I-vitesse de convergence. dans ce dernier cas, la définition précédente n'est pas très heureuse car elle est vérifiée par tous les estimateurs.

Correction pour les I-notions On dit que ψ_n est un *minorant de la I-vitesse de convergence* pour l'ensemble Σ et la pseudo-métrique d s'il existe un réel positif C tel que :

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} r(\hat{\theta}_n, \psi_n, C) \geq p_0,$$

avec p_0 positif et $r(\hat{\theta}_n, \psi_n, C) = \sup_{\theta \in \Sigma} \mathbb{P}_\theta \left(\psi_n^{-1}d(\hat{\theta}_n, \theta) \leq C \right)$

De même, ψ_n est appelée *I-vitesse minimax de convergence* pour l'ensemble Σ et la pseudo-métrique d si ψ_n est un minorant de la I-vitesse de convergence et s'il existe θ_n^* un estimateur (que l'on qualifiera d' *I-optimal* tel que :

$$\lim_{C \rightarrow \infty} \limsup_{n \rightarrow \infty} r(\theta_n^*, \psi_n, C) = 0.$$

Remarque importante : les I-notions sont plus fortes que les S-notions. Cela sera utile dans la démonstration du théorème qui vient.

Vraisemblance, rapport de vraisemblances Soit un modèle statistique dominé (i.e. pour tout θ , P_θ est dominé par une mesure μ , appelée mesure dominante). On définit alors une vraisemblance $p(\theta, x) = \frac{dP_\theta}{d\mu}(x)$ (qui n'est pas unique) et un rapport de vraisemblance

$$\Lambda(x; \theta_1, \theta_2) = \frac{dP_{\theta_1}}{dP_{\theta_2}}(x) = \frac{p(\theta_1, x)}{p(\theta_2, x)}$$

2.2 Le Saint Graal du débruitage

Avant de présenter une borne absolue à l'efficacité des algorithmes de débruitage, il faut mentionner un écueil d'ordre pratique : l'obtention du risque minimax pour un algorithme n'est pas chose aisée, et se fait traditionnellement que pour $n \rightarrow \infty$, n étant le nombre d'échantillons (pixels). Si n sera souvent appelé taille de l'image, il faut donc garder à l'esprit que faire croître revient à mieux échantillonner, et non pas à agrandir. On trouve aussi dans la littérature des études pour $\sigma \rightarrow 0$, ce qui revient au même. Cela peut paraître paradoxal, mais il s'agit avant tout d'un critère d'ordre théorique, que l'on espère en phase avec la réalité du terrain. Le théorème qui suit est **fondamental** car il nous donne une limite infranchissable.

Théorème (Saint Graal du débruitage) Soit une image de taille n (i.e à n "pixels"). On prend les X_i des points de cette image (éventuellement tirés aléatoirement indépendamment les uns des autres), et l'on observe $Y_i = X_i + \xi_i$ avec les ξ_i des variables gaussiennes indépendantes deux à deux et indépendantes des X_i .

Sur la classe horizon β définie précédemment, la S-vitesse minimax de convergence vaut $n^{-\frac{\beta}{\beta+1}}$.

Un algorithme atteignant ces performances sera donc optimal, au sens du minimax. Dans le cas étudié ($\beta = 2$), on cherche à atteindre $n^{-\frac{2}{3}}$ (ou bien $n^{-\frac{4}{3}}$ quand n désigne la largeur de l'image). Afin de démontrer ce résultat, on a besoin du théorème suivant. Plutôt que d'estimer l'image, on préfère considérer que l'on cherche à estimer G , l'ensemble situé sous la courbe (ie sur lequel l'image vaut 1). La métrique utilisée sera celle de la différence symétrique, qui correspond à la distance L^1 des fonctions associées à chaque ensemble.

Théorème principal sur les bornes inférieures (TPBI) : On considère $\theta_0, \dots, \theta_M$ des éléments de Σ avec $M = M_n \rightarrow \infty$, ce qui signifie que nous sommes davantage dans le test que dans l'estimation. On fait les hypothèses suivantes :

1. Pour tous j, k tels que $0 \leq j < k \leq M$, alors on a l'inégalité

$$d(\theta_j, \theta_k) \geq s_n > 0$$

2. Les rapports de vraisemblance ont la forme suivante :

$$\Lambda(\theta_0, \theta_j) = \exp(\Delta_j - \lambda_j \log M),$$

avec $0 < \lambda_j < \lambda < 1$ et les Δ_j des variables aléatoires telles que

$$\mathbb{P}_j(\Delta_j \geq 0) \geq p_* > 0.$$

3. Enfin, on demande que λ et p_* soient indépendants de j et de n .

Dans le cadre des hypothèses précédentes, **pour tout estimateur** $\hat{\theta}_n$ on a

$$\max_{0 \leq j \leq M} \mathbb{P}_j \left(d(\hat{\theta}_n, \theta_j) \geq s_n \right) \geq \frac{p_*}{2}.$$

Ce théorème aux hypothèses fantaisistes (démontré dans [2]) facilite grandement la démonstration du Graal. En effet, afin de le démontrer, nous voulons montrer que pour $\psi_n = n^{\frac{\beta}{\beta+1}}$, on a

$$\liminf_{n \rightarrow \infty} \inf_{\hat{G}_n} \sup_{G \in \mathcal{G}} \mathbb{P}_G(\psi_n^{-1} d(G, \hat{G}_n) \geq C) \geq p_0.$$

Ce qui est exactement le résultat annoncé par le TPBI.

Il suffit donc d'appliquer le TPBI aux bons éléments !

2.3 Construction des θ_i

Posons $\delta_n = n^{-\frac{1}{\beta+1}}$ et supposons que $m = m_n = \frac{1}{\delta_n}$ est entier.

Lemme de Varshamov-Gilbert Posons $M = \lfloor 2^{\frac{m}{8}} \rfloor$, il existe $M + 1$ vecteurs binaires de taille m notés $\bar{\omega}_j$ tels que la distance de Hamming ρ_H (nombre de composantes différentes) de deux vecteurs distincts quelconques parmi eux dépasse toujours $\frac{m}{16}$. On les numérote de 0 à M .

Fonction de base pour la classe $\Sigma(\beta, L)$ Il s'agit d'une fonction ϕ de \mathbb{R} dans lui-même telle que

- ϕ est \mathcal{C}^∞
- ϕ est nulle en dehors de $]-\frac{1}{2}; \frac{1}{2}[$ et strictement positive à l'intérieur.
- Le maximum de sa dérivée d'ordre $k + 1$ est plus petit que L , avec k la partie entière de β .

Soient $u_l = l\delta_n$ pour l de 0 à m , $U_i = [u_i, u_{i+1}]$ et b_i le centre de U_i pour i de 1 à m . Soient $f_{kn} = \delta_n^\beta \phi(\delta_n^{-1}(x - b_k))$, pour k de 1 à m . On définit

$$f(\cdot, \bar{\omega}_j) = \sum_{k=1}^m (\bar{\omega}_j)_k f_{kn}(\cdot)$$

et on pose enfin :

$$\theta_j = f(\cdot, \bar{\omega}_j) + \frac{1}{2}$$

On notera

$$G_j = \{(x_1, x_2) | 0 \leq x_2 \leq g(x_1, \bar{\omega}_j)\},$$

et $\mathbb{P}_{\theta_j} = \mathbb{P}_{G_j}$ pour j de 0 à m .

Vérification des hypothèses du TPBI

$$d(\theta_j, \theta_k) = \|f(\cdot, \bar{\omega}_j) - f(\cdot, \bar{\omega}_k)\| = c\delta^{\beta+1} \rho_H(\bar{\omega}_j, \bar{\omega}_k) \geq 2s_n,$$

avec

$$s_n = \frac{c}{32} n^{-\frac{\beta}{\beta+1}},$$

donc on a le premier point.

Les rapports de vraisemblance sont de la forme suivante :

$$\begin{aligned} \frac{d\mathbb{P}_{G_0}}{d\mathbb{P}_{G_j}} &= \exp\left(-\sum_{i=1}^n \mathbf{1}_{X_i \in G_l - G_0} Y_i + \frac{1}{2} \text{Card}(i | X_i \in G_l - G_0)\right) \\ &= \exp\left(\Delta_j - \frac{n}{2} \lambda(G_j - G_0)\right), \end{aligned}$$

avec

$$\Delta_l = - \sum_{i=1}^n \mathbf{1}_{X_i \in G_l - G_0} \xi_i + \frac{\mathbb{E}(\Lambda_l) - \Lambda_l}{2},$$

or

$$\lambda(G_j - G_0) = d(\theta_j, \theta_0) = c\delta^{\beta+1} \rho_H(\bar{\omega}_j, \bar{\omega}_0) \leq cm\delta^{\beta+1},$$

Et donc $\lambda(G_j - G_0) \leq \lambda \log M$ équivaut à $c\delta^{\beta+1} \leq \lambda \frac{\log 2}{8}$ qui est vraie pour n assez grand et c assez petit. D'où une première partie du deuxième point. On a par ailleurs

$$\Lambda_l = \sum \mathbf{1}_{X_i \in G_l - G_0},$$

qui est en fait une somme de variables aléatoires de Bernoulli deux à deux indépendantes de même paramètre $p = d(\theta_j, \theta_0)$, ce qui permet donc d'affirmer que les Λ_l sont tous égaux en loi.

$$\begin{aligned} & \mathbb{P}_{G_l}(\Delta_l \geq 0) \\ & \geq \mathbb{P} \left(- \sum_{i=1}^n \mathbf{1}_{X_i \in G_l - G_0} \xi_i \geq 0, \quad \mathbb{E}(\Lambda_l) - \Lambda_l \geq 0 \right) \\ & \geq \mathbb{E} \left(\mathbb{P} \left(- \sum_{i=1}^n \mathbf{1}_{X_i \in G_l - G_0} \xi_i \geq 0 \mid X_1, \dots, X_n \right) \mathbf{1}_{\mathbb{E}(\Lambda_l) - \Lambda_l \geq 0} \right). \end{aligned}$$

Connaissant X_1, \dots, X_n , la distribution de $-\sum_{i=1}^n \mathbf{1}_{X_i \in G_l - G_0} \xi_i$ est gaussienne (éventuellement dégénérée) donc on a :

$$\mathbb{P} \left(- \sum_{i=1}^n \mathbf{1}_{X_i \in G_l - G_0} \xi_i \geq 0 \mid X_1, \dots, X_n \right),$$

qui vaut $\frac{1}{2}$ dans le cas non-dégénéré et 1 dans le cas dégénéré.

Donc on a le second point, le TPBI s'applique alors aux θ_i et le résultat est démontré.

Retour sur le calcul des vraisemblances On a $Y_i = \mathbf{1}_{X_i \in G} + \xi_i$ où l'on suppose ξ un vecteur gaussien standard de taille n . On a donc

$$\mathbb{P}_G(Y_i = y \mid X_i) = \frac{1}{\sqrt{2\pi}} \exp \left(- \frac{(\mathbf{1}_{X_i \in G} - y)^2}{2} \right)$$

Soit p la densité des X_i . Une vraisemblance sera alors, par indépendance des X_i entre eux :

$$\mathbb{P}_G(dy_1, \dots, dy_n, dx_1, \dots, dx_n) = \prod_{i=1}^n \mathbb{P}_G(Y_i = y \mid X_i) p(x_1, \dots, x_n) dx_1 \cdots dx_n dy_1 \cdots dy_n.$$

Lorsque l'on fait le rapport des vraisemblances, la constante $\frac{1}{\sqrt{2\pi}}$ et la densité des X_i se simplifient pour donner :

$$\begin{aligned} \frac{d\mathbb{P}_{G_0}}{d\mathbb{P}_{G_l}} &= \prod_{i=1}^n \frac{\exp\left(-\frac{(\mathbf{1}_{X_i \in G_0} - Y_i)^2}{2}\right)}{\exp\left(-\frac{(\mathbf{1}_{X_i \in G_l} - Y_i)^2}{2}\right)} \\ &= \exp\left(-\frac{1}{2} \sum_{i=1}^n (-2Y_i(\mathbf{1}_{X_i \in G_0} - \mathbf{1}_{X_i \in G_l})) - \frac{1}{2} \sum_{i=1}^n (\mathbf{1}_{X_i \in G_0} - \mathbf{1}_{X_i \in G_l})\right). \end{aligned}$$

D'où l'expression annoncée, si l'on suppose que pour tout l , $G_0 \subset G_l$. Toutefois, même si ce n'est pas le cas, la démonstration reste vraie en remplaçant $\mathbf{1}_{X_i \in G_l - G_0}$ par $(\mathbf{1}_{X_i \in G_0} - \mathbf{1}_{X_i \in G_l})$

2.4 Applications de la théorie minimax

On va maintenant pouvoir quantifier l'efficacité des méthodes les plus standard de débruitage à la lumière des critères introduits précédemment.

Convolution par un filtre linéaire Soit g un filtre réel, symétrique ($g_{i,j} = g_{-i,-j} = g_{i,-j}$) et soit G sa transformée de Fourier de dimension 2. On dit que ce filtre est isotrope s'il existe $F : \mathbb{R} \rightarrow \mathbb{C}$, tel que

$$G(\omega_1, \omega_2) = F\left(\sqrt{\omega_1^2 + \omega_2^2}\right).$$

Soit g un tel filtre. On suppose de plus que $\|\text{grad}(G(\omega_1, \omega_2))\|_2 \leq C$ avec C une constante. Alors, avec f une image de côté n , et \hat{f}_g l'image débruitée par convolution, on a :

$$\inf_g \sup_{f \in \mathcal{C}(2)} R_n(f, \hat{f}_g) = \Theta(n^{-\frac{2}{3}})$$

Filtre de Yaroslavsky Soient y l'image bruitée, et x l'image originelle (quand on la verra comme une fonction, on l'appellera f). On définit comme suit l'estimateur donné par le filtre de Yaroslavsky :

$$\hat{f}_{\Delta_n, \tau}^{SY}(i, j) = \frac{\sum_{m=i-\Delta_n}^{i+\Delta_n} \sum_{l=j-\Delta_n}^{j+\Delta_n} \omega_{i,j}^Y(m, l) y_{m,l}}{\sum_{m=i-\Delta_n}^{i+\Delta_n} \sum_{l=j-\Delta_n}^{j+\Delta_n} \omega_{i,j}^Y(m, l)},$$

avec

$$\omega_{i,j}^{SY}(m, l) = e^{-\frac{(y_{m,l} - x_{i,j})^2}{2\tau^2}}.$$

La présence du $x_{i,j}$ permet de faciliter les calculs, mais l'algorithme obtenu est alors un algorithme à semi-oracle.

On a alors, pour une image f de côté n :

$$\inf_{\tau, \Delta_n} \sup_{f \in \mathcal{C}(2)} R_n(f, \hat{f}^{SY}) = \Omega(n^{-\frac{2}{3}}).$$

Algorithme NLM On construit un estimateur de f comme suit :

$$\hat{f}_{i,j}^N = \frac{\sum_{(i,j)} \omega_{i,j}^N(m,l) y_{m,l}}{\sum_{(i,j)} \omega_{i,j}^N(m,l)},$$

avec

$$\omega_{i,j}^N(m,l) = \begin{cases} 1 & \text{si } d_{\delta_n}^2(y_{i,j}, y_{m,l}) \leq \sigma^2 + t_n, \\ 0 & \text{sinon} \end{cases},$$

où l'on définit

$$d_{\delta_n}^2(y_{i,j}, y_{n,p}) = \frac{1}{(2\delta_n + 1)^2 - 1} \sum_{|l| \leq \delta_n, |l| \leq \delta_n} |y_{i+l, j+m} - y_{n+l, p+m}|^2 - |y_{i,j} - y_{n,p}|^2.$$

On définit le SNLM (NLM avec semi-oracle) en remplaçant la moitié des valeurs bruitées par les valeurs réelles, c'est à dire que la distance s'écrit désormais :

$$d_{\delta_n}^2(y_{i,j}, y_{n,p}) = \frac{1}{(2\delta_n + 1)^2 - 1} \sum_{|l| \leq \delta_n, |l| \leq \delta_n} |x_{i+l, j+m} - y_{n+l, p+m}|^2 - |x_{i,j} - y_{n,p}|^2.$$

Soit $\varepsilon > 0$, on considère l'algorithme NLM avec $\delta_n = 2 \log^{\frac{1}{2} + \varepsilon} n$ et $t_n = \frac{2\sigma^2}{\log^{\frac{1}{2} + \varepsilon} n}$. Alors le risque de l'algorithme sur la classe horizon est en

$$O\left(\frac{\log^{\frac{1}{2} + \varepsilon} n}{n}\right).$$

Sous certaines hypothèses (voir [3] et [4]), on montre que le risque du SNLM est en $\frac{1}{n}$. Ce risque ne dépend pas de β , ce qui veut dire que l'algorithme NLM ne sait pas tirer parti de la régularité de l'image. Il devient donc sous-optimal pour β assez grand (supérieur à $\frac{1}{2}$)

Récapitulatif	Estimateur localement polynomial	$\Theta(n^{-\frac{2}{3}})$
	Semi-Yaroslavsky Filter	$\Theta(n^{-\frac{2}{3}})$
	Non-Local Means	$O\left(\frac{\log^{\frac{1}{2} + \varepsilon} n}{n}\right)$
	Optimalité	$\Theta(n^{-\frac{4}{3}})$

Les critères du minimax sont peu flatteurs pour les algorithmes présentés : il nous reste donc du travail avant de pouvoir prétendre à l'optimalité. La partie suivante propose des versions modifiées de l'algorithme NLM dont on établira l'optimalité selon les critères du minimax.

3 ANLM, ou optimisation de l'algorithme NLM

Il a été montré que l'algorithme de NLM standard n'atteignait pas la borne optimale (en terme de minimax) du débruitage. L'idée est d'améliorer cet algorithme, en ne se basant plus sur des voisinages isotropiques des pixels, comme c'était pour l'instant le cas ; mais sur des voisinages différents (anisotropiques). Il sera démontré que ce faisant, on s'approche du risque minimax optimal. Trois différents algorithmes ANLM seront présentés.

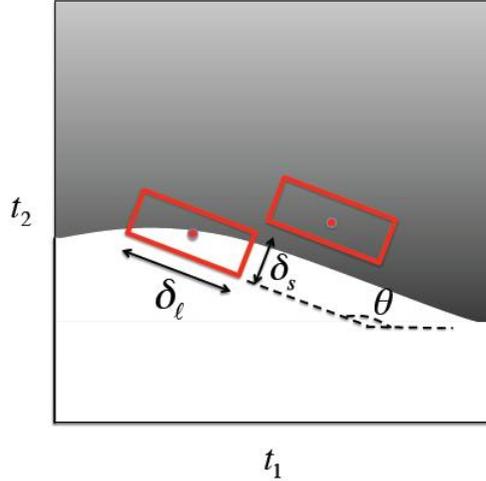
- Le premier, ANLM avec Oracle (OANLM), supposera une connaissance parfaite du contour, et il sera utilisé pour l'étude théorique de l'optimalité.
- Le second, ANLM avec angle Discret (DANLM), optimise le choix du voisinage anisotropique pour chaque pixel, sans aucune information du type oraculaire, pour atteindre une performance proche de l'optimum.
- Enfin, le troisième, ANLM basé sur le Gradient (GANLM), utilisera les informations issues du gradient pour estimer l'orientation du contour.

3.1 OANLM

Voisinages directionnels Soient $S=[0,1]^2$ et $R_{\nu,\mu}^\theta(\cdot)$ la rotation de centre (ν,μ) , d'angle θ . On définit le voisinage anisotropique $(\theta, \delta_s, \delta_l)$ d'un point (t_1, t_2) par

$$I_{\theta, \delta_s, \delta_l}(t_1, t_2) = R_{t_1, t_2}^\theta \left(\left[t_1 - \frac{\delta_l}{2}; t_1 + \frac{\delta_l}{2} \right] \times \left[t_2 - \frac{\delta_s}{2}; t_2 + \frac{\delta_s}{2} \right] \right) \cap S,$$

où θ représente l'orientation, δ_l la longueur, et δ_s la largeur du voisinage. (En pratique, on prendra $\delta_s \leq \delta_l$)



On découpe à présent $I_{\theta, \delta_s, \delta_l}(t_1, t_2)$ en $n_s \times n_l$ parties, données par

$$I_{\theta, \delta_s, \delta_l}^{j_1, j_2}(t_1, t_2) = R_{t_1, t_2}^\theta \left(\left[t_1 - \frac{j_1 \delta_l}{2n_l}; t_1 + \frac{j_1 \delta_l}{2n_l} \right] \times \left[t_2 - \frac{j_2 \delta_s}{2n_s}; t_2 + \frac{j_2 \delta_s}{2n_s} \right] \right) \cap S,$$

pour $1 \leq j_1 \leq n_l$ et $1 \leq j_2 \leq n_s$.

Définissons le voisinage de (t_1, t_2) comme suit :

$$y_{t_1, t_2}^{\theta, \delta_s, \delta_l}(j_1, j_2) = \frac{\delta_s \delta_l}{n_s n_l} \int \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_l}^{j_1, j_2}(t_1, t_2)} dY(s_1, s_2),$$

et

$$X(t_1, t_2) = y_{t_1, t_2}^{\theta, \delta_s, \delta_l}(1, 1) = \frac{\delta_s \delta_l}{n_s n_l} \int \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_l}^{1, 1}(t_1, t_2)} dY(s_1, s_2).$$

La distance anisotrope entre deux voisinages $(\theta, \delta_s, \delta_l)$ est définie par

$$d_{\theta, \delta_s, \delta_l}^2(dY(t_1, t_2), dY(s_1, s_2)) = \frac{1}{n_s n_l - 1} (\|y_{t_1, t_2}^{\theta, \delta_s, \delta_l} - y_{s_1, s_2}^{\theta, \delta_s, \delta_l}\|_2^2 - |y_{t_1, t_2}^{\theta, \delta_s, \delta_l}(0, 0) - y_{s_1, s_2}^{\theta, \delta_s, \delta_l}(0, 0)|^2).$$

On remarque que l'on définit la distance entre deux patchs de même orientation et de même taille seulement.

Enfin, pour un paramètre de seuillage τ_σ qu'on se fixe, on donne les poids suivants :

$$w_{t_1, t_2}^{\theta, \delta_s, \delta_l}(s_1, s_2) = \begin{cases} 1 & \text{si } d_{\theta, \delta_s, \delta_l}^2(dY(t_1, t_2), dY(s_1, s_2)) \leq 2 \frac{n_s n_l}{\delta_s \delta_l} \sigma^2 + \tau_\sigma \\ 0 & \text{sinon} \end{cases}$$

Le seuil se résume à la distance entre le patch réel et le patch bruité augmentée d'une marge d'erreur τ_σ .

L'estimateur ALNM au point (t_1, t_2) est alors

$$\hat{f}^{\theta, \delta_s, \delta_l}(t_1, t_2) = \frac{\int_S w_{t_1, t_2}^{\theta, \delta_s, \delta_l}(s_1, s_2) X(s_1, s_2) ds_1 ds_2}{\int_S w_{t_1, t_2}^{\theta, \delta_s, \delta_l}(s_1, s_2) ds_1 ds_2}.$$

C'est à dire que l'on fait la moyenne sur les patchs que l'on garde.

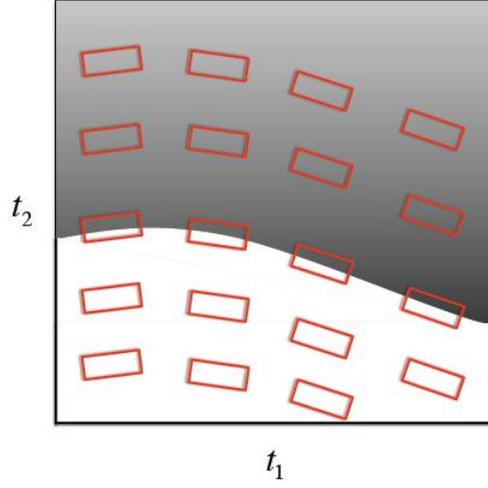
3.2 L'estimateur OANLM

Dans cette partie nous supposons donc que nous avons accès à l'orientation du contour en chaque point de l'image, grâce à un oracle.

On considère $f_h \in \mathcal{C}(\beta)$, et l'on note $h'(t_1)$ la dérivée du contour $h(t_1)$.

OANLM est l'application de ANLM avec les paramètres suivants :

- Une échelle quadratique pour les voisinages : nous prendrons $\delta_s = \delta_l^2$ et $\delta_s = 4\sigma^{2\alpha/(\alpha+1)} |\log \sigma|^{2\alpha/(\alpha+1)}$. Dans le cas qui nous intéresse $\alpha = 2$, cela donne $\delta_s = 4\sigma^{4/3} |\log \sigma|^{4/3}$
- Des voisinages alignés verticalement : le voisinage en un point (t_1, t_2) aura la même orientation que le voisinage au point $(t_1, h(t_1))$
- On suppose que $n_s n_l = \log^2 \sigma$
- Paramètre de seuillage : $\tau_\sigma = \frac{2}{\sqrt{|\log \sigma|}}$



Résultat Sous ces paramètres, et en désignant par \hat{f}^O l'estimateur OANLM, on a

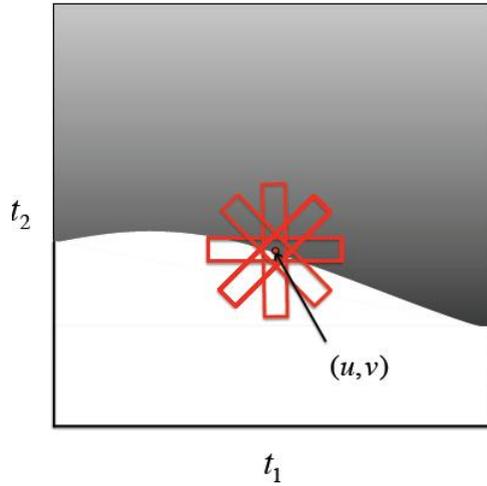
$$R(\mathcal{C}(2), \hat{f}^O) = O(\sigma^{4/3} |\log \sigma|^{4/3}) = O(\delta_s).$$

Ce résultat n'a évidemment qu'un intérêt purement théorique, puisque nous ne disposons que rarement dudit oracle.

3.3 DANLM

La principale idée du ANLM avec angle discret, c'est de calculer la distance entre plusieurs voisinages anisotropiques avec des angles différents.

Ainsi, on pose $q = \pi \sigma^{-2/3}$
 et $\theta_0 = 0, \theta_1 = \sigma^{2/3}, \theta_2 = 2\sigma^{2/3} \dots \theta_{q-1} = (q-1)\sigma^{2/3} = \pi - \sigma^{2/3}$
 et on définit $\Theta = \{\theta_0, \theta_1 \dots \theta_{q-1}\}$



Comme pour OANLM, on choisit $\delta_s = 4\sigma^{4/3} |\log \sigma|^{4/3}$ et $\delta_s = \delta_t^2$. Tout se

déroule de la même façon, à ceci près que l'on définit une nouvelle distance :

$$d_{\mathcal{A}}^2(dY(t_1, t_2), dY(s_1, s_2)) = \min_{\theta \in \Theta} d_{\theta, \delta_s, \delta_l}^2(dY(t_1, t_2), dY(s_1, s_2)),$$

où on rappelle que

$$d_{\theta, \delta_s, \delta_l}^2(dY(t_1, t_2), dY(s_1, s_2)) = \frac{1}{n_s n_l - 1} (\|y_{t_1, t_2}^{\theta, \delta_s, \delta_l} - y_{s_1, s_2}^{\theta, \delta_s, \delta_l}\|_2^2 - |y_{t_1, t_2}^{\theta, \delta_s, \delta_l}(0, 0) - y_{s_1, s_2}^{\theta, \delta_s, \delta_l}(0, 0)|^2).$$

Les poids sont définis comme suit

$$w_{t_1, t_2}^D(s_1, s_2) = \begin{cases} 1 & \text{si } d_{\mathcal{A}}^2(dY(t_1, t_2), dY(s_1, s_2)) \leq 2 \frac{n_s n_l}{\delta_s \delta_l} \sigma^2 + \tau_\sigma \\ 0 & \text{sinon} \end{cases}$$

et l'estimateur donné par

$$\hat{f}^D(t_1 t_2) = \frac{\int_S w_{t_1, t_2}^D(s_1, s_2) X(s_1, s_2) ds_1 ds_2}{\int_S w_{t_1, t_2}^D(s_1, s_2) ds_1 ds_2}.$$

Résultat Sous ses conditions, le risque minimax de DANLM vérifie

$$R(\mathcal{C}(2), \hat{f}^D) \leq O(\sigma^{4/3} |\log \sigma|^{4/3}).$$

3.4 Algorithmes ALNM pratiques

Extension aux images discrètes Cette section étudie la mise en pratique des voisinages ANLM sur des images discrètes. Soit une image $f(\frac{i}{n}, \frac{j}{n})$ de taille $n \times n$ dont nous ayons une observation bruitée

$$o_{i,j} = f\left(\frac{i}{n}, \frac{j}{n}\right) + \zeta_{i,j} \text{ où } \zeta_{i,j} \sim \mathcal{N}(0, \xi^2).$$

Soit $\bar{S} = \{\frac{1}{n}, \dots, 1\} \times \{\frac{1}{n}, \dots, 1\}$. Pour $B \in S$, on note $\bar{B} = B \cap \bar{S}$. En particulier, $\bar{I}_{\theta, \delta_s, \delta_l}(\frac{i}{n}, \frac{j}{n}) = I_{\theta, \delta_s, \delta_l}(\frac{i}{n}, \frac{j}{n}) \cap \bar{S}$.

Soit $\mathcal{P} = \{(r, q) \in \mathbb{Z}^2 / (\frac{i+r}{n}, \frac{j+q}{n}) \in \bar{I}_{\theta, \delta_s, \delta_l}(\frac{i}{n}, \frac{j}{n})\}$. On définit alors la distance anisotropique $(\theta, \delta_s, \delta_l)$ entre deux pixels $o_{i,j}, o_{m,l}$ par

$$\bar{d}_{\theta, \delta_s, \delta_l}^2(o_{i,j}, o_{m,l}) = \frac{1}{|\mathcal{P}|} \sum_{(r,q) \in \mathcal{P}} (o_{i+r, j+q} - o_{m+r, l+q})^2.$$

On donne les poids

$$\bar{w}_{i,j}^{\theta, \delta_s, \delta_l}(m, l) = \begin{cases} 1 & \text{si } \bar{d}_{\theta, \delta_s, \delta_l}^2(o_{i,j}, o_{m,l}) \leq 2\xi^2 + \tau_n \\ 0 & \text{sinon} \end{cases}$$

et l'estimateur en version discrète :

$$\hat{f}_{i,j}^{\theta, \delta_s, \delta_l} = \frac{\sum_m \sum_l \bar{w}_{i,j}^{\theta, \delta_s, \delta_l}(m, l) o_{m,l}}{\sum_m \sum_l \bar{w}_{i,j}^{\theta, \delta_s, \delta_l}(m, l)}.$$

GANLM Si DANLM peut être implémenté en pratique, et s'il est théoriquement optimal, il n'en reste pas moins extrêmement gourmand : en effet, il faut calculer pour chaque pixel un grand nombre de distances, pour choisir ensuite la minimale, avant même de se lancer dans le calcul de l'estimateur. Sa complexité croît linéairement avec le nombre de voisinages q que l'on considère, qui est de l'ordre de $n^{2/3}$. C'est pour cela que nous proposons l'algorithme GANLM, qui choisit l'orientation des voisinages d'après une connaissance du contour fournie par le gradient de l'image.

Pour en améliorer la rapidité, on se fixe un seuil sur la valeur du gradient, qui indiquera à l'algorithme d'utiliser NLM dans les régions avec peu de contours, un voisinage anisotropique se révélant presque inutile dans de telles régions. L'idée étant d'estimer l'orientation du contour via le gradient, de nombreuses erreurs peuvent apparaître pour des images très bruitées. Aussi sera-t-il préférable de débruiter dans un premier temps avec un autre algorithme, comme NLM par exemple, afin d'avoir une estimation plus fiable de l'image. Le pseudo-code correspondant à cette démarche est donné ici :

Algorithme GANLM

Entrée \hat{f} : estimation de l'image
 δ_l, δ_s : taille des voisinages
 λ : seuil
 Estimer le gradient de l'image $(g_h(i, j), g_v(i, j))$ en chaque pixel (i, j)
 Pour chaque pixel faire :

$$g(i, j) = \sqrt{g_h(i, j)^2 + g_v(i, j)^2}$$

$$\theta_{i, j} = \arctan\left(\frac{g_v(i, j)}{g_h(i, j)}\right)$$

Si $g(i, j) \geq \lambda$ appliquer ANLM avec les paramètres $\theta_{i, j}, \delta_l, \delta_s$

Si non appliquer NLM avec $\delta = \sqrt{\delta_l \delta_s}$

Tests pratiques

On compare les performances du NLM et du GANLM sur deux images : l'une synthétique, correspond à la courbe d'un sinus, l'autre est une image naturelle (tirée de Playboy, 1972). On remarque que si GANLM est souvent meilleur en terme de PSNR (ce n'est même pas le cas ici...), le gain sur l'image naturelle est minime (de l'ordre de 0.1 dB). Pour l'image synthétique, les choses se passent un peu mieux, mais il faut tout de même restreindre le calcul du PSNR à une bande voisine de la discontinuité pour obtenir un gain en performance de l'ordre du décibel.



Avec bruit de variance 100



Avec NLM, PSNR de 33.4141



Avec GANLM, PSNR de 34.0365



Lena bruitée (Variance de 100)



Avec NLM, PSNR de 32.72



Avec GANLM, PSNR de 32.47

Conclusion

La recherche de l'optimum contribue à la diversification de la faune des algorithmes de débruitage. Si toutefois l'algorithme ANLM atteint bien l'optimum au sens de la théorie minimax, nos expériences n'ont pas permis de mettre en évidence une supériorité écrasante face au NLM standard. On note aussi la présence de paramètres qu'il n'est pas évident de bien choisir. Par conséquent, le succès que représente le développement d'un algorithme optimal ne doit pas mettre fin à notre quête d'un meilleur débruitage. En outre, il convient de réfléchir aux critères qui nous ont conduit à annoncer l'optimalité car ceux-ci ne sont pas toujours en accord avec le PSNR, et *a fortiori* avec des critères purement esthétiques. L'approche anisotrope reste néanmoins intéressante, et parfaitement justifiée du point de vue conceptuel : on peut espérer des développements prometteurs, soit dans l'amélioration des algorithmes anisotropes déjà existants, soit dans l'exportation du concept à d'autres algorithmes.

Bibliographie

- [1] A.Levin, B.Nadler, Natural Image Denoising : Optimality and Inherent Bounds, 2011
- [2] A.P.Korostelev, A.B.Tsybakov, Minimax theory of image reconstruction, Springer-Verlag, 1993.
- [3] R.G.Baraniuk, A.Maleki, M.Narayan, Suboptimality of Non-Local Means for image with sharp edges, preprint, 2011
- [4] R.G.Baraniuk, A.Maleki, M.Narayan, Anisotropic Non-Local Means denoising, preprint, 2011
- [5] M.Lebrun, M.Colom, A.Buades, J-M.Morel, Secrets of image denoising cuisine, published online, Apr 2012

Remerciements

Un grand merci à nos tuteurs, Jean-Michel Morel et Mauricio Delbracio, pour nous avoir guidés tout au long de ce stage. Nous remercions également Alexandre Tsybakov, pour avoir éclairé un point crucial de notre travail, ainsi que Richard Baraniuk qui a eu la gentillesse de nous envoyer le code de son algorithme, qui a nourri notre réflexion.

Annexes

Filtre de Wiener

Soit $(\varphi_n)_{n \geq 0}$ une base orthogonale. L'image \tilde{u} s'écrit sur cette base $\tilde{u} = \sum \langle \tilde{u} | \varphi_i \rangle$. On va rajouter un coefficient, sur lequel on va pouvoir jouer : ainsi, on pose

$$\hat{u}(a_i) = \sum a_i \cdot \langle \tilde{u} | \varphi_i \rangle \varphi_i.$$

L'idée à présent est de trouver les a_i tels que l'erreur soit minimale, autrement dit, on cherche a_i tels que $a_i = \arg \min_{a_i} \mathbb{E} \|u - \hat{u}(a_i)\|^2$

Or, en supposant le bruit n de loi $\mathcal{N}(0, \sigma^2)$, on a $\langle \tilde{u} | \varphi_i \rangle = \langle u | \varphi_i \rangle + n_i$.

D'où $\|u - \hat{u}(a_i)\|^2 = \sum [\langle u | \varphi_i \rangle - a_i (\langle u | \varphi_i \rangle + n_i)]^2$
et $\mathbb{E} \|u - \hat{u}(a_i)\|^2 = \sum (\sigma^2 a_i^2 + \langle u | \varphi_i \rangle^2 (1 - a_i)^2)$ On cherche le minimum de cette fonction en a_i :

$$\frac{\partial}{\partial a_i} \mathbb{E} \|u - \hat{u}(a_i)\|^2 = 2 \sum (\sigma^2 a_i - (1 - a_i) \langle u | \varphi_i \rangle^2) = 0,$$

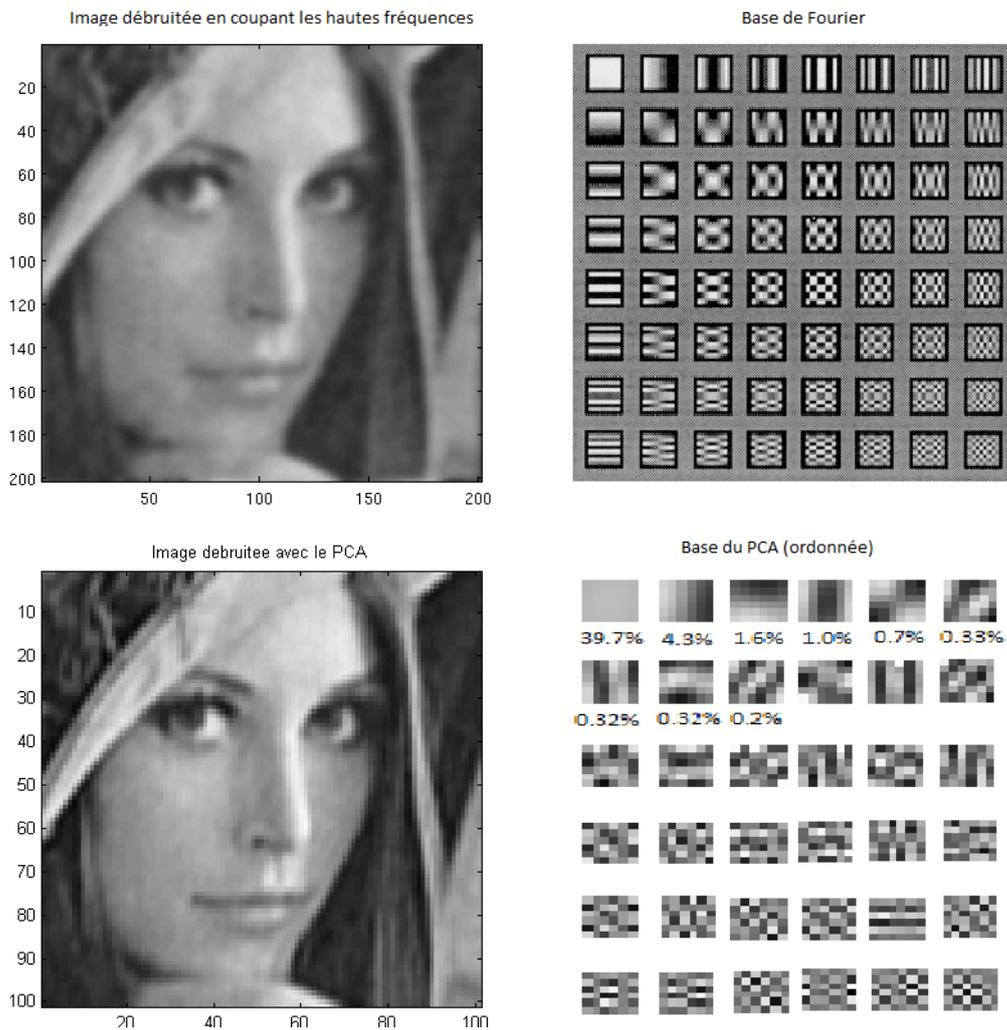
ce qui donne :

$$a_i = \frac{\langle u | \varphi_i \rangle^2}{\sigma^2 + \langle u | \varphi_i \rangle^2},$$

soit encore, puisqu'on ne connaît pas $\langle u | \varphi_i \rangle$,

$$a_i = \frac{\langle \tilde{u} | \varphi_i \rangle^2 - \sigma^2}{\langle \tilde{u} | \varphi_i \rangle^2}.$$

On prend bien conscience, en regardant la première égalité, que si le signal est prépondérant sur le bruit, c'est-à-dire si $\langle u | \varphi_i \rangle^2$ est grand devant σ^2 , alors a_i sera proche de 1 ; et qu'*a contrario*, si c'est le bruit qui domine, a_i sera proche de 0.



Risque minimax d'un filtre linéaire

On considère $f_h(t_1, t_2)$ pour $h(t) = \frac{1}{2}$ (c'est à dire une image carrée constituée de la juxtaposition de deux rectangles de tailles identiques, l'un blanc, l'autre noir). On suppose que n , la taille de l'image, est un nombre pair. On représente la TFD bidimensionnelle d'un signal x par :

$$X(k_1, k_2) = \frac{1}{n} \sum_{l_1} \sum_{l_2} x_{l_1, l_2} e^{-j \frac{2\pi k_1 l_1}{n}} e^{-j \frac{2\pi k_2 l_2}{n}}.$$

Comme $y = x + z$ avec z gaussien centré, la TFD de \hat{f}_g^{LF} vaut :

$$\hat{F}_g^{LF}(k_1, k_2) = Y(k_1, k_2)G(k_1, k_2) = X(k_1, k_2)G(k_1, k_2) + Z(k_1, k_2)G(k_1, k_2).$$

L'invariance de la loi gaussienne centrée par TFD donne alors $Z \sim \mathcal{N}(0, \sigma^2)$.

Pour l'image étudiée, on établit que :

$$X(k_1, k_2) = \begin{cases} 0 & \text{si } k_1 \neq 0, \\ \frac{1-e^{-j\pi k_2}}{1-e^{-j\frac{2\pi k_2}{n}}} & \text{si } k_1 = 0, \end{cases}$$

Notre définition du risque amène :

$$R_n(f, \hat{f}_g^{LF}) = \frac{1}{n^2} E(\|X - \hat{F}_g^{LF}\|_F^2),$$

avec

$$\|Y\|_F^2 = \sum_{k_1, k_2} |Y(k_1, k_2)|^2.$$

Soit $B(\hat{f})$ le biais de l'estimateur \hat{f} , on a :

$$\begin{aligned} B^2(\hat{f}_g^{LF}) &= \frac{1}{n^2} \|X - E(\hat{F}_g^{LF})\|^2 = \frac{1}{n^2} \sum_{1 \leq k_2 \leq n, k_2 \text{ impair}} |1 - G(0, k_2)|^2 \frac{1}{\sin^2 \frac{\pi k_2}{n}} \\ &\geq \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{\frac{2}{3}}, k_2 \text{ impair}} |1 - G(0, k_2)|^2 \frac{1}{\sin^2 \frac{\pi k_2}{n}}. \end{aligned}$$

Considérons désormais sa variance définie par :

$$V(\hat{f}_g^{LF}) = \frac{1}{n^2} E\left(\|\hat{F}_g^{LF} - E(\hat{F}_g^{LF})\|^2\right) = \frac{1}{n^2} \sum_{k_1, k_2} |G(k_1, k_2)|^2 \sigma^2.$$

Osons une généralisation bidimensionnelle des sommes de Riemann pour une fonction continue à gradient borné :

$$\frac{1}{n^2} \sum_{k_1, k_2} |G(k_1, k_2)|^2 = \iint |\hat{G}(\omega_1, \omega_2)|^2 + O(n^{-1}).$$

Par passage en polaire, en exploitant la radialité de g :

$$\iint |\hat{G}(\omega_1, \omega_2)|^2 \geq 2\pi \int_0^{2\pi} r |F(r)| dr = 2\pi \int_0^{2\pi} \omega_2 |\hat{G}(0, \omega_2)|^2 d\omega_2.$$

Donc, en combinant ce qui précède :

$$V(\hat{f}_g^{LF}) = \frac{4\pi^2}{n^2} \sum_{k_2} k_2 |G(0, k_2)|^2 \sigma^2 - O(n^{-1}).$$

En sommant le biais et la variance, on obtient le risque. Mais avant cela, on adapte $G(0, k_2)$ pour leur donner le même ordre de grandeur. Pour cela, on prend :

$$G(0, k_2) = \frac{1}{1 + \frac{4\pi^4 \sigma^2 k_2^3}{n^2}}$$

pour k_2 impair, et nul sinon. Dans ce cas, le risque est minoré par le double du biais, donc étudions directement le biais en y injectant la valeur obtenue de $G(0, k_2)$.

$$\begin{aligned} B^2(\hat{f}_g^{LF}) &\geq \frac{1}{n^4} \left(\frac{4\pi^4 \sigma^2}{1 + 4\pi^4 \sigma^2} \right)^2 \sum_{1 \leq k_2 \leq n^{2/3}, k_2 \text{ impair}} k_2^4 \\ &\geq K^2 \left(n^{-\frac{2}{3}} + o(n^{-2/3}) \right). \end{aligned}$$

On obtient ce résultat en deux temps :

1. Pour n assez grand, la somme sur les termes impairs est un Θ de la somme sur tous les termes.
2. Pour tout $K \in \mathbb{N}$,

$$\sum_{k=1}^K k^4 \geq \sum_{k=1}^K k^3 = \left(\frac{K(K+1)}{2} \right)^2.$$

Risque minimax du filtre de Yaroslavsky

Lemme 1 On suppose que $x_{i,j}$ vaut 0. Si $x_{m,l} = x_{i,j}$, alors $E(\omega_{m,l} y_{m,l}) = 0$. De plus, si $|x_{i,j} - x_{m,l}| = 1$ (c'est à dire $x_{m,l} = 1$), alors $E(\omega_{m,l} y_{m,l}) > \frac{\tau}{\sqrt{\sigma^2 + \tau^2}} e^{-\frac{1}{2(\sigma^2 + \tau^2)}}$.

Démonstration La première assertion est évidente par symétrie.

Les calculs donnent :

$$\begin{aligned} E(\omega_{m,l}) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-1)^2}{2\tau^2} - \frac{z^2}{2\sigma^2}} dz \\ &= \frac{\tau e^{-\frac{1}{2(\sigma^2 + \tau^2)}}}{\sqrt{\sigma^2 + \tau^2}}, \end{aligned}$$

et

$$E(z\omega_{m,l}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{(z-1)^2}{2\tau^2} - \frac{z^2}{2\sigma^2}} dz \geq 0.$$

Donc, comme

$$E(\omega_{m,l} y_{m,l}) = E(\omega_{m,l} x_{m,l}) + E(\omega_{m,l} z_{m,l}),$$

les résultats qui précèdent démontrent la seconde assertion.

Définition On définit le Δ -voisinage d'un pixel (m, l) par :

$$\mathcal{C}_{m,l}^\Delta = \{(i, j) : |i - m| \leq \Delta, |j - l| \leq \Delta\} \cap \mathcal{S}.$$

Lemme 2 Soit $\Omega_n = (2\Delta_n + 1)^2$. On a

$$\mathbb{P} \left(\frac{1}{\Omega_n} \left(\sum_{(m,l) \in \mathcal{C}_{i,j}^{\Delta_n}} \omega_{m,l}^{Sy} - \sum_{(m,l) \in \mathcal{C}_{i,j}^{\Delta_n}} E(\omega_{m,l}^{Sy}) \right) \geq t \right) \leq 2e^{-2\Omega_n t^2}.$$

Démonstration C'est une simple application de l'inégalité de Hoeffding aux variables aléatoires centrées, indépendantes et bornées :

$$\frac{1}{\Omega_n} \left(\omega_{m,l}^{Sy} - E(\omega_{m,l}^{Sy}) \right).$$

Démonstration du théorème Dans un premier temps, on montre par l'absurde que le choix optimal de Δ_n est en $\Omega(\log n)$. On constate en effet que si $\Delta_n = O(\log n)$ alors pour l'image nulle, l'estimateur obtenu est non biaisé mais sa variance est en $\Omega\left(\frac{1}{\log^2 n}\right)$. Nous sommes donc bien loin de l'optimum trouvé dans le cas des filtres linéaires.

On étudie désormais la même image que précédemment avec n pair, et l'on définit deux régions de cette image :

1. $P_1 = \left\{ (i, j) : \frac{n}{2} \leq j \leq \frac{n}{2} + \frac{\Delta_n}{2} \right\}$,
2. $P_2 = \left\{ (i, j) : j > \frac{n}{2} + \Delta_n \right\}$.

On remarque qu'au moins un quart des pixels voisins (au sens de Δ_n) d'un pixel de P_1 ont une valeur débruitée égale à 1, et qu'il en va de même pour tous les pixels voisins d'un pixel de P_2 .

Cas 1 $(i, j) \in P_1$

Par l'inégalité de Jensen, on a

$$E \left(x_{i,j} - \frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l} y_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l}} \right)^2 \geq \left(E \frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l} y_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l}} \right)^2.$$

On définit les deux constantes suivantes :

$$\begin{aligned} m_0 &= E(\omega_{i,j}^{SY}(m,l) | x_{i,j} = 0, x_{m,l} = 0), \\ m_1 &= E(\omega_{i,j}^{SY}(m,l) | x_{i,j} = 0, x_{m,l} = 1). \end{aligned}$$

Il est clair que $m_0 > m_1$. Pour ε quelconque, on définit l'événement :

$$A = \left\{ \sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l} \sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} E(\omega_{m,l}) \leq \Delta_n^{2-\varepsilon} \right\}.$$

On a :

$$\begin{aligned}
E \left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l} y_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l}} \right) &\geq E \left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l} y_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l}} \middle| A \right) \mathbb{P}(A) \\
&\geq E \left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l} y_{m,l}}{4\Delta_n^2 m_0 + \Delta_n^{2-\varepsilon}} \middle| A \right) \mathbb{P}(A) \\
&\geq E \left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l} y_{m,l}}{4\Delta_n^2 m_0 + \Delta_n^{2-\varepsilon}} \right) - \mathbb{P}(A^C) \\
&\geq \left(\frac{c_0 \Delta_n^2}{4\Delta_n^2 m_0 + \Delta_n^{2-\varepsilon}} \right) - \mathbb{P}(A^C).
\end{aligned}$$

La première inégalité vient du lemme 2 et du fait que $m_0 \geq m_1$, la dernière étape utilise le lemme 1 avec $c_0 = \frac{\tau}{\sqrt{\sigma^2 + \tau^2} e^{\frac{1}{2(\sigma^2 + \tau^2)}}}$ et le fait que, comme $\mathcal{C}_{m,l}^{\Delta_n}$ a $(2\Delta_n + 1)^2$ pixels, au moins Δ_n^2 ont la valeur sans bruit 1. Comme de plus, $\Delta_n = \Omega(\log n)$, le lemme 2 montre que $\mathbb{P}(A^C) = o(1)$, et donc que le biais est, pour les pixels dans P_1 , de l'ordre d'une constante.

Cas 2 $(i, j) \in P_2$ Comme tous les pixels autour des pixels de P_2 ont une valeur sans bruit égale à 1,

$$E \left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l} y_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l}} \right)^2 = \left(E \frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l} z_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l}} \right)^2.$$

$$\begin{aligned}
& E \left(\left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l} z_{m,l}}{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l}} \right)^2 \middle| A \right) \mathbb{P}(A) \\
& \geq E \left(\left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^\Delta} \omega_{m,l} z_{m,l}}{4\Delta_n^2 m_0 + \Delta_n^{2-\varepsilon}} \right)^2 \middle| A \right) \mathbb{P}(A) \\
& \geq E \left(\frac{\sum_{(m,l) \in \mathcal{C}_{m,l}^{\Delta_n}} \omega_{m,l} y_{m,l}}{4\Delta_n^2 m_0 + \Delta_n^{2-\varepsilon}} \right)^2 - \mathbb{P}(A^C) \\
& = \frac{4\Delta_n^2 E(\omega_{i,j} z_{i,j})^2}{(4\Delta_n^2 m_0 + \Delta_n^{2-\varepsilon})^2} - \mathbb{P}(A^C).
\end{aligned}$$

Etant donné $\Delta_n = O(\log n)$, par le lemme 2, on a $\mathbb{P}(A^C) = o(\frac{1}{n^2})$. Ainsi, l'expression du dessus est en $\frac{\gamma}{\Delta_n^2}$. Donc en combinant les minoration sur P_1 et P_2 on obtient un minorant du type $\frac{\beta\Delta_n}{n} + \frac{\gamma}{\Delta_n^2}$, ce qui donne un Δ_n optimal de l'ordre de $n^{1/3}$ pour lequel :

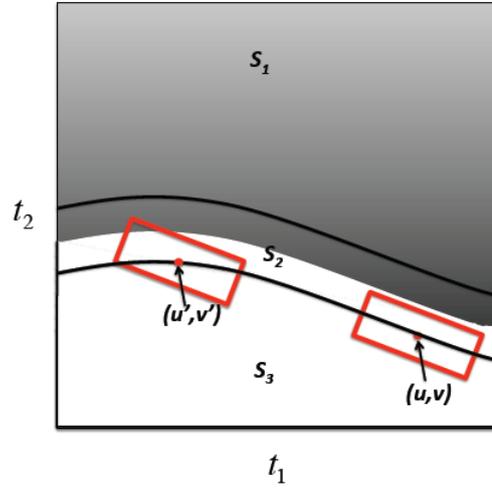
$$\inf_{\Delta_n, \tau} R_n(f, \hat{f}^{SY}) = \Omega(n^{-\frac{2}{3}}).$$

On retrouve l'idée de conditionner sur un événement favorable dans la démonstration suivante.

Risque du OANLM

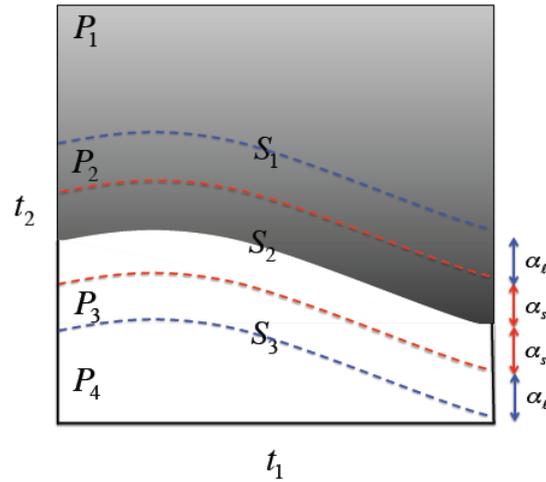
Soit S l'image, on définit la partition suivante :

$$\begin{aligned}
S_1 &= \left\{ (v, u) \mid u > h(v) + \left(1 + \frac{C}{2}\right) \delta_s \right\}, \\
S_2 &= \left\{ (v, u) \mid h(v) - \left(1 + \frac{C}{2}\right) \delta_s < u < h(v) + \left(1 + \frac{C}{2}\right) \delta_s \right\}, \\
S_3 &= \left\{ (v, u) \mid u < h(v) - \left(1 + \frac{C}{2}\right) \delta_s \right\}.
\end{aligned}$$



Comme on le voit sur les figures, les voisinages anisotropes bien orientés de points de S_1 (resp de S_3) seront toujours strictement au dessus (resp en dessous) de la courbe de h . On partitionne S_1 en P_1 et P_2 , et S_3 en P_3 et P_4 définis ci-dessous (l'idée est que les voisinages anisotropes de points de P_1 (resp P_4) soient toujours au dessus (resp en dessous) de la courbe de h quelle que soit leur orientation.

$$\begin{aligned}
 P_1 &= \left\{ (v, u) \mid h(v) + 2\delta_l + \frac{C}{2}\delta_s \leq u \right\}, \\
 P_2 &= \left\{ (v, u) \mid h(v) + \left(1 + \frac{C}{2}\right)\delta_s \leq u \leq h(v) + 2\delta_l + \frac{C}{2}\delta_s \right\}, \\
 P_3 &= \left\{ (v, u) \mid h(v) - \left(1 + \frac{C}{2}\right)\delta_s \geq u \geq h(v) - 2\delta_l - \frac{C}{2}\delta_s \right\}, \\
 P_4 &= \left\{ (v, u) \mid u \leq h(v) - 2\delta_l - \frac{C}{2}\delta_s \right\}.
 \end{aligned}$$



Lemme 1 : Soient Z_1, \dots, Z_r des variables gaussiennes standard iid. On a les inégalités de concentration suivantes pour la variable aléatoire $\sum_{i=1}^r Z_i^2$ de loi χ_r^2 :

$$\begin{aligned}\mathbb{P}\left(\frac{1}{r}\sum_i Z_i^2 - 1 > t\right) &\leq e^{-\frac{r}{2}(t - \log(1+t))}, \\ \mathbb{P}\left(\frac{1}{r}\sum_i Z_i^2 - 1 < -t\right) &\leq e^{-\frac{r}{2}(t + \log(1+t))}.\end{aligned}$$

Définitions :

$$\begin{aligned}F_{t_1, t_2}^{\theta, \delta_s, \delta_l}(j_1, j_2) &= \frac{n_s n_l}{\delta_s \delta_l} \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_l}^{j_1, j_2}} dW(s_1, s_2), \\ Z(t_1, t_2) &= \frac{n_s n_l}{\delta_s \delta_l} \int \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_l}^{j_1, j_2}(t_1, t_2)} dW(s_1, s_2), \\ F(t_1, t_2) &= \frac{n_s n_l}{\delta_s \delta_l} \int \int_{(s_1, s_2) \in I_{\theta, \delta_s, \delta_l}^{j_1, j_2}(t_1, t_2)} f(s_1, s_2) ds_1 ds_2,\end{aligned}$$

pour plus de simplicité, on omettra les exposants et les indices lorsqu'il n'y aura pas d'ambigüité.

$$\begin{aligned}A_1 &= \{(u, v) \in P_1 | w(u, v) = 1\}, \\ A_2 &= \{(u, v) \in P_4 | w(u, v) = 0\}.\end{aligned}$$

Lemme 3 : Prenons $\delta_l = 2\sigma^{\frac{2}{3}} |\log \sigma|^{\frac{2}{3}}$ et $\delta_s = 4\delta_l^2$; $n_s = 2|\log \sigma|^{\frac{2}{3}}$ et $n_l = |\log \sigma|^{\frac{4}{3}}$. On a alors :

$$\begin{aligned}\mathbb{P}(\lambda(P_1) - \lambda(A_1) > \varepsilon) &= O\left(\frac{\sigma^8}{\varepsilon}\right), \\ \mathbb{P}(\lambda(P_4) - \lambda(A_2) > \varepsilon) &= O\left(\frac{\sigma^8}{\varepsilon}\right).\end{aligned}$$

Ce résultat s'obtient en appliquant le lemme 2 au vecteur gaussien F dont la norme suit une loi χ^2 , ce qui donne un ordre de grandeur de $\mathbb{E}(\lambda(A_1))$ qui nous permet, via l'inégalité de Markov, de majorer $\mathbb{P}(\lambda(P_1) - \lambda(A_1) > \varepsilon)$. On procède de même pour la seconde inégalité.

Fort de ce résultat, on définit comme suit l'événement \mathcal{E} :

$$\mathcal{E} = \{\lambda(P_1) - \lambda(A_1) < \sigma^2\} \cap \{\lambda(P_4) - \lambda(A_2) < \sigma^2\}.$$

Le risque du OANLM au point $(t_1, t_2) \in S_1$ vaut alors $\mathbb{E}(U)$ avec

$$U = \left(\frac{\int_S w X du dv}{\int_S w du dv} \right)^2.$$

On a

$$\mathbb{E}(U) = \mathbb{E}(U|\mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{E}(U|\mathcal{E}^C)\mathbb{P}(\mathcal{E}^C) \leq \mathbb{E}(U|\mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}^C).$$

Par le lemme 3, ce dernier terme est en $O(\sigma^6)$ (on a pris $\varepsilon = \sigma^2$).

Par ailleurs, en intercalant les bons termes, on obtient :

$$\begin{aligned} \left| \int_{P_1} wXdudv - \int_{P_1} Xdudv \right| &= O(\lambda(P_1 - A_1)), \\ \left| \int_{P_4} wXdudv - \int_{P_4} Xdudv \right| &= O(\lambda(P_4 - A_2)), \\ \left| \int_{P_1} wdudv - \int_{P_1} dudv \right| &= O(\lambda(P_1 - A_1)), \\ \left| \int_{P_4} wdudv - \int_{P_4} dudv \right| &= O(\lambda(P_4 - A_2)). \end{aligned}$$

Donc, en posant $P_{14} = P_1 \cup P_4$ et $B = S - P_{14}$, on a

$$\mathbb{E}(U|\mathcal{E})\mathbb{P}(\mathcal{E}) = \mathbb{E} \left(\left(\frac{\int_{P_{14}} wXdudv + \int_B wXdudv}{\int_{P_{14}} wdudv + \int_B wdudv} \right)^2 \middle| \mathcal{E} \right) \mathbb{P}(\mathcal{E}).$$

On montre, en calculant la différence et en faisant apparaître en facteur des différences en $O(\lambda(P_i - A_j))$ (qui est contrôlable puisqu'on a conditionné par \mathcal{E} qu'on a effectivement :

$$\mathbb{E}(U|\mathcal{E})\mathbb{P}(\mathcal{E}) = \mathbb{E} \left(\left(\frac{\int_{P_1} Xdudv + \int_B wXdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 \middle| \mathcal{E} \right) \mathbb{P}(\mathcal{E}) + O(\sigma^2).$$

Partant de là, on a trivialement

$$\mathbb{E}(U|\mathcal{E})\mathbb{P}(\mathcal{E}) \leq \mathbb{E} \left(\frac{\int_{P_1} Xdudv + \int_B wXdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 + O(\sigma^2).$$

Puis, en décomposant X en $F + Z$, en développant le carré et en utilisant l'inégalité de Cauchy-Schwarz on obtient :

$$\begin{aligned} \mathbb{E}(U|\mathcal{E})\mathbb{P}(\mathcal{E}) &\leq \mathbb{E} \left(\frac{\int_B wFdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 + \mathbb{E} \left(\frac{\int_{P_1} Zdudv + \int_B wZdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 \\ &\quad + 2 \sqrt{\mathbb{E} \left(\frac{\int_{P_1} Fdudv + \int_B wFdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2} \sqrt{\mathbb{E} \left(\frac{\int_{P_1} Zdudv + \int_B wZdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2} + O(\sigma^2). \end{aligned}$$

Lemme 4 Dans le cadre étudié, on a :

$$\mathbb{E} \left(\frac{\int_B wFdudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 \leq O(\sigma^{\frac{4}{3}} |\log \sigma|^{\frac{4}{3}}).$$

On majore le premier terme par $\mathbb{E} \left(\frac{\int_B dudv}{\int_{P_1} dudv} \right)^2 = \left(\frac{\lambda(B)}{\lambda(P_1)} \right)^2$ et on n'a plus qu'à calculer l'aire de ces deux ensembles.

Lemme 5

$$\mathbb{E} \left(\frac{\int_{P_1} Z dudv + \int_B w Z dudv}{\int_{P_1} dudv + \int_B wdudv} \right)^2 = O(\sigma^{\frac{4}{3}} |\log \sigma|^{\frac{4}{3}}).$$

On majore d'abord cette expression par

$$\mathbb{E} \left(\frac{\int_{P_1} Z dudv + \int_B w Z dudv}{\int_{P_1} dudv} \right)^2$$

que l'on majore ensuite par $V_1 + V_2 + V_3$ (développement + Cauchy-Schwarz) avec

$$\begin{aligned} V_1 &= \mathbb{E} \left(\frac{\int_{P_1} Z dudv}{\int_{P_1} dudv} \right)^2, \\ V_2 &= \mathbb{E} \left(\frac{\int_B w Z dudv}{\int_{P_1} dudv} \right)^2, \\ V_3 &= 2\sqrt{V_1} \sqrt{V_2}. \end{aligned}$$

Pour majorer V_1 et V_2 , on remplace le carré au numérateur par une intégrale double (Fubini) avant d'intervertir espérance et intégration (re-Fubini). On se ramène alors à des calculs d'aires qui donnent :

$$\begin{aligned} V_1 &= O \left(\frac{\delta_s \delta_l}{n_s n_l} \right) = O(\sigma^2), \\ V_2 &= O(\delta_l^2) = O(\sigma^{\frac{4}{3}} |\log \sigma|^{\frac{4}{3}}). \end{aligned}$$

on obtient alors que V_1 et V_3 sont négligeables devant V_2 .

Bilan

- Pour tout point de S_1 , on vient de montrer que le risque minimax du OANLM pour la métrique L^2 et la fonction de perte $x \rightarrow x^2$ est en $O(\sigma^{\frac{4}{3}} |\log \sigma|^{\frac{4}{3}})$.
- Par symétrie, il en sera de même pour tout point de S_3 .
- Pour les points de S_2 , on majore simplement le risque en chaque point par 1 puisque l'aire de ce domaine décroît suffisamment vite avec $\sigma \rightarrow 0$ pour que le risque total soit négligeable devant les risques sur S_1 et S_3 .

Le théorème est ainsi démontré.