Agrégation à poids exponentiels: Algorithmes d'échantillonnage

Duy Tung LUU[1], Jalal FADILI[1], Christophe CHESNEAU[2]
[1]Normandie Univ, ENSICAEN, CNRS, GREYC, France [2] Normandie Univ, UNICAEN, CNRS, LMNO, France
duy-tung.luu@ensicaen.fr, Jalal.Fadili@ensicaen.fr, Christophe.Chesneau@unicaen.fr

Résumé – Un problème classique en traitement du signal et des images vise à estimer un signal/image à partir de ses mesures sous-déterminées et bruitées. Un estimateur populaire pour ce faire est l'agrégation à poids exponentiels. L'implémentation de ce dernier nécessite de résoudre un problème d'intégration très complexe en grande dimension, mais pour lequel les méthodes d'échantillonnage stochastique s'avèrent utiles. Dans cet article, nous proposons des algorithmes d'éclatement proximal pour échantillonner à partir de distributions dont les densités ne sont pas nécessairement lisses ni log-concaves. Notre approche allie des outils issus de l'analyse variationnelle et l'optimisation non-lisse d'une part, et d'autre part des equations de diffusion stochastiques, et plus précisément la diffusion de Langevin. Nous établissons en particulier les garanties de consistance de nos algorithmes vus comme des schémas de discrétisation dans ce contexte. Ces algorithmes sont appliqués pour calculer l'agrégateur à poids exponentiels pour divers problèmes inverses impliquant a priori non-lisses favorisant une certaine notion de simplicité/faible complexité.

Abstract – A classical problem in signal and image processing aims at recovering a signal/image from a set of underdetermined and noisy measurements. A popular estimator to do so is based exponential weighted aggregation. Implementing the latter necessitates to solve a challenging integration problem in high dimension, but for which stochastic sampling methods can prove useful. In this paper, we propose proximal splitting-type algorithms for sampling from distributions whose densities are not necessarily smooth nor log-concave. Our approach brings together tools from, on the one hand, variational analysis and non-smooth optimization, and on the other hand, stochastic diffusion equations, and in particular the Langevin diffusion. We establish in particular consistency guarantees of our algorithms seen as discretization schemes in this context. These algorithms are applied to compute the exponentially weighted aggregates for for some inverse problems involving non-smooth priors encouraging some notion of simplicity/complexity.

1 Introduction

1.1 Estimation en grande dimension

On considère le modèle linéaire direct suivant

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{\xi}, \tag{1}$$

où $\boldsymbol{y} \in \mathbb{R}^n$ est le vecteur des observations, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ est l'opérateur de mesure. $\boldsymbol{X}$ joue le rôle de l'opérateur de dégradation dans un problème de restauration en traitement du signal/image, ou encore la matrice des covariables pour un problème de régression en statistique. L'objectif est alors d'estimer le vecteur $\boldsymbol{\theta}_0 \in \mathbb{R}^p$ à partir des observations $\boldsymbol{y}$. En général, le problème (1) est soit sous-déterminé ($p < n$), ou justement déterminé ($p = n$) mais $\boldsymbol{X}$ est mal-conditionnée. Ainsi, on se trouve confronté à un problème mal-posé. Toutefois, $\boldsymbol{\theta}_0$ présente en général une structure simple se traduisant par le fait qu'il vit dans un sous-ensemble de $\mathbb{R}^p$ qui est à la fois structuré est de faible dimension intrinsèque. C'est ce qu'on appellera ici une notion de simplicité ou de faible complexité. On peut alors imposer ce type de structure simple par l'introduction d'a priori la favorisant.

Agrégation à poids exponentiels (EWA) L'EWA consiste à calculer l'espérance

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{EWA}} = \int_{\mathbb{R}^p} \boldsymbol{\theta}\widehat{\mu}(\boldsymbol{\theta})d\boldsymbol{\theta},\ \ \widehat{\mu}(\boldsymbol{\theta}) \propto \exp\left(-V(\boldsymbol{\theta})/\beta\right), \tag{2}$$

où $\beta > 0$ est appelé paramètre de température, et

$$V(\boldsymbol{\theta}) \stackrel{\text{def}}{=} F(\boldsymbol{X}\boldsymbol{\theta}, \boldsymbol{y}) + J_\lambda(\boldsymbol{\theta}).$$

$F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ est une fonction de perte générale supposée différentiable, J_λ est une pénalité (ou une régularisation) dépendant d'un paramètre $\lambda > 0$ favorisant une notion spécifique de simplicité/faible complexité. Il est utile de rappeler que (2) peut être vu comme une espérance conditionnelle a posteriori, mais uniquement dans certains cas.

1.2 La diffusion de Langevin

L'implémentation de $\widehat{\boldsymbol{\theta}}_n^{\mathrm{EWA}}$ correspond à un problème d'intégration qui devient complexe à résoudre analytiquement, ou même numériquement, en grande dimension. Une approche classique est de l'approcher via des méthodes MCMC consistant à échantillonner à partir de μ en construisant une chaîne de Markov appropriée dont la distribution stationnaire est μ, et de calculer la moyenne empirique des échantillons de la chaîne. Ici, nous considérons celles issues du processus de diffusion de Langevin. Une diffusion de Langevin $\boldsymbol{L} \in \mathbb{R}^p$, $p \geq 1$ est un processus de Markov homogène défini par l'équation différentielle stochastique (EDS)

$$d\boldsymbol{L}(t) = \frac{1}{2}\boldsymbol{\rho}(\boldsymbol{L}(t))dt + d\boldsymbol{W}(t),\ t > 0,\ \boldsymbol{L}(0) = \boldsymbol{l}_0, \tag{3}$$

où $\boldsymbol{\rho} = \nabla \log \mu$, μ est une densité sur $\mathbb{R}^p$ qui est non nulle partout et suffisamment lisse, $\boldsymbol{W}$ est un processus Brownien de dimension p et $\boldsymbol{l}_0 \in \mathbb{R}^p$ est la condition initiale. Sous des hypothèses faibles, l'EDS (3) a une unique solution dite forte et, $\boldsymbol{L}(t)$ suit une distribution stationnaire dont la densité est μ [10, Theorem 2.1]. Cela suggère une approximation naturelle de l'espérance $\int_{\mathbb{R}^p} f(\boldsymbol{\theta})\mu(\boldsymbol{\theta})d\boldsymbol{\theta}$, où $f : \mathbb{R}^p \to \mathbb{R}$, par la valeur moyenne d'une diffusion de Langevin, i.e., $\frac{1}{T}\int_0^T f(\boldsymbol{L}(t))dt$ pour $T > 0$ suffisamment grand.

Toutefois, pour calculer la trajectoire de diffusion en pratique, la dynamique définie par l'EDS (3) ne peut en général pas être suivie exactement. On doit alors la discrétiser. Une discrétisation classique est donnée par un schéma d'Euler explicite (Euler-Maruyama) qui s'écrit

$$\boldsymbol{L}_{k+1} = \boldsymbol{L}_k + \frac{\delta}{2}\boldsymbol{\rho}(\boldsymbol{L}_k) + \sqrt{\delta}\boldsymbol{Z}_k,\ t > 0,\ \boldsymbol{L}_0 = \boldsymbol{l}_0,$$

où $\delta > 0$ est le pas de discrétisation qui doit être suffisamment petit et $\{\boldsymbol{Z}_k\}_k$ sont iid $\sim \mathcal{N}(0, \mathbf{I}_p)$. Alors, la valeur moyenne $\frac{1}{T}\int_0^T \boldsymbol{L}(t)dt$ peut être approximée par la somme de Riemann $\frac{\delta}{T}\sum_{k=0}^{\lfloor T/\delta \rfloor - 1}\boldsymbol{L}_k$, où $\lfloor T/\delta \rfloor$ est la partie entière de T/δ qui est une approximation naturelle de $\widehat{\boldsymbol{\theta}}$. Cependant, cette approche exige que μ soit suffisamment lisse. Pour une revue complète sur l'échantillonnage par la diffusion de Langevin à partir de densités lisses et log-concaves, nous renvoyons à [1]. Pour faire face à des densités non lisses, quelques travaux récents ont proposé de remplacer $\log \mu$ par une version lissée en utilisant la régularisation/enveloppe de Moreau-Yosida [2, 8, 3, 4].

1.3 Contributions

Dans ce travail, nos principales contributions sont les suivantes :

- On vise à élargir la famille de μ couverte par [2, 8, 3, 4] en relâchant les conditions sous-jacentes. Plus précisément, dans notre contexte, μ est structurée comme $\widehat{\mu}$ dans (2) mais n'est pas nécessairement différentiable ni log-concave.
- Nous proposons deux algorithmes de type explicite-implicite proximal avec les garanties de consistance correspondantes.
- Ces algorithmes sont appliqués pour calculer l'estimateur EWA avec diverses penalités populaires dans la littérature.

Toutes les preuves des résultats énoncés peuvent être trouvées dans la version longue [6].

2 Notations et préliminaires

Vecteurs et matrices Pour $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ symétrique définie positive $\langle\cdot,\cdot\rangle_{\boldsymbol{M}} = \langle\cdot, \boldsymbol{M}\cdot\rangle$, où $\langle\cdot,\cdot\rangle$ est le produit scalaire euclidien sur $\mathbb{R}^d$. Soit $\left\|.\right\|_r$, pour $r \geq 1$, la norme ℓ_r d'un vecteur avec l'adaptation habituelle pour $r = +\infty$. $\mathbf{I}_d$ est la matrice identité sur $\mathbb{R}^d$. Pour $I \subset \{1, \ldots, d\}$, $\boldsymbol{x}_I$ est le sous-vecteur dont les entrées sont celles de $\boldsymbol{x} \in \mathbb{R}^d$ restreintes aux indices dans I.

Fonctions Pour une fonction C^1-lisse f, $\nabla f(\boldsymbol{x})$ désigne son gradient. Le domaine effectif de f est défini par $\mathrm{dom}(f) = \left\{\boldsymbol{x} \in \mathbb{R}^d \ :\ f(\boldsymbol{x}) < +\infty\right\}$. Une fonction f est propre si $f(\boldsymbol{x}) > -\infty$ pour tout $\boldsymbol{x}$ et $\mathrm{dom}(f) \neq \emptyset$. Une fonction f est semi-continue inférieurement (sci) en $\boldsymbol{x}_0$ si $\liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$.

L'opérateur proximal et l'enveloppe de Moreau Soit $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ symétrique définie positive. Pour une fonction f propre et sci et $\gamma > 0$, l'opérateur proximal et l'enveloppe Moreau sont définies respectivement par

$$\mathrm{prox}_{\gamma f}^{\boldsymbol{M}}(\boldsymbol{x}) \stackrel{\text{def}}{=} \underset{\boldsymbol{w}\in\mathbb{R}^d}{\mathrm{Argmin}} \left\{\frac{1}{2\gamma}\left\|\boldsymbol{w}-\boldsymbol{x}\right\|_{\boldsymbol{M}}^2 + f(\boldsymbol{w})\right\},$$
$${}^{\boldsymbol{M},\gamma}f(\boldsymbol{x}) \stackrel{\text{def}}{=} \inf_{\boldsymbol{w}\in\mathbb{R}^d} \left\{\frac{1}{2\gamma}\left\|\boldsymbol{w}-\boldsymbol{x}\right\|_{\boldsymbol{M}}^2 + f(\boldsymbol{w})\right\},$$

Ici, $\mathrm{prox}_{\gamma f}^{\boldsymbol{M}}(\boldsymbol{x})$ peut être un ensemble car le minimiseur, s'il existe, n'est pas nécessairement unique. Lorsque $\boldsymbol{M} = \mathbf{I}_d$, on utilise les notations allégées $\mathrm{prox}_{\gamma f}$ et ${}^{\gamma}f$.

3 Langvin Monte-Carlo proximal

Considérons la distribution μ dans l'EDS (3), et supposons qu'elle est structurée comme

$$\mu(\boldsymbol{\theta}) \propto \exp\left(-\big(L(\boldsymbol{\theta}) + H(\boldsymbol{D}^\top\boldsymbol{\theta})\big)\right), \qquad (4)$$

où $L \in C^1(\mathbb{R}^p)$, $\boldsymbol{D} \in \mathbb{R}^{p\times q}$ et $H : \mathbb{R}^q \to \mathbb{R} \cup \{+\infty\}$. On suppose qu'il existe $\boldsymbol{M} \in \mathbb{R}^{q\times q}$ une matrice symétrique définie positive, telle que les hypothèses suivantes soient vraies.

($\mathbf{H}.\mathbf{1}$) H est propre, sci et bornée inférieurement.

($\mathbf{H}.\mathbf{2}$) $\mathrm{prox}_{\gamma H}^{\boldsymbol{M}}$ est univoque.

($\mathbf{H}.\mathbf{3}$) $\mathrm{prox}_{\gamma H}^{\boldsymbol{M}}$ est localement Lipschitz continue.

($\mathbf{H}.\mathbf{4}$) $\exists C_1 > 0, \forall\boldsymbol{\theta} \in \mathbb{R}^p, \left\langle \boldsymbol{D}^\top\boldsymbol{\theta}, \mathrm{prox}_{\gamma H}^{\boldsymbol{M}}(\boldsymbol{D}^\top\boldsymbol{\theta})\right\rangle_{\boldsymbol{M}} \leq C_1(1 + \left\|\boldsymbol{\theta}\right\|_2^2)$.

($\mathbf{H}.\mathbf{5}$) $\exists C_2 > 0, \forall\boldsymbol{\theta} \in \mathbb{R}^p, \langle\boldsymbol{\theta}, \nabla L(\boldsymbol{\theta})\rangle \leq C_2(1 + \left\|\boldsymbol{\theta}\right\|_2^2)$.

Une grande famille de fonctions H vérifie les hypothèses ($\mathbf{H}.\mathbf{1}$)-($\mathbf{H}.\mathbf{3}$). En effet, on peut montrer que les fonctions dites prox-régulières (et a fortiori convexes) [9], vérifient ces hypothèses. La proposition suivante assure la différentiabilité de ${}^{\boldsymbol{M},\gamma}H$ et exprime $\nabla^{\boldsymbol{M},\gamma}H$ via $\mathrm{prox}_{\gamma H}^{\boldsymbol{M}}$.

Proposition 3.1. *Supposons que* ($\mathbf{H}.\mathbf{1}$) *et* ($\mathbf{H}.\mathbf{2}$) *sont satisfaites. Alors* ${}^{\boldsymbol{M},\gamma}H \in C^1(\mathbb{R}^q)$, *et* $\nabla^{\boldsymbol{M},\gamma}H = \gamma^{-1}\boldsymbol{M}\left(\mathbf{I}_q - \mathrm{prox}_{\gamma H}^{\boldsymbol{M}}\right)$.

On considère la diffusion de Langevin $\boldsymbol{L} \in \mathbb{R}^p$ définie par l'EDS suivante

$$d\boldsymbol{L}(t) = -\frac{1}{2}\nabla\left(L + ({}^{\boldsymbol{M},\gamma}H)\circ\boldsymbol{D}^\top\right)(\boldsymbol{L}(t))dt + d\boldsymbol{W}(t), \quad (5)$$

lorsque $t > 0$ et $\boldsymbol{L}(0) = \boldsymbol{l}_0$.

Proposition 3.2. *Supposons que* $(\mathbf{H.1})$-$(\mathbf{H.5})$ *sont satisfaites. Pour chaque point initial* $\boldsymbol{L}(0)$ *tel que* $\mathbb{E}\left[\left\|\boldsymbol{L}(0)\right\|_2^2\right] < \infty$, *l'EDS* (5) *admet une unique solution qui est fortement Markovienne, non-explosive et admet une unique mesure invariante de densité* $\mu_\gamma(\boldsymbol{\theta}) \propto \exp\left(-\left(L(\boldsymbol{\theta}) + ({}^{\boldsymbol{M},\gamma}H) \circ \boldsymbol{D}^\top(\boldsymbol{\theta})\right)\right)$.

La proposition suivante répond à la question naturelle du comportement de $\mu_\gamma - \mu$ par rapport à γ.

Proposition 3.3. *Supposons que* $(\mathbf{H.1})$ *est satisfaite. Alors* μ_γ *converge vers* μ *en variation totale lorsque* $\gamma \to 0$.

En appliquant Proposition 3.1 à (5), cette dernière devient

$$d\boldsymbol{L}(t) = \mathcal{A}(\boldsymbol{L}(t))dt + d\boldsymbol{W}(t),\ \boldsymbol{L}(0) = \boldsymbol{l}_0,\ t > 0, \quad (6)$$

où l'opérateur $\mathcal{A} = -\frac{1}{2}(\nabla L + \gamma^{-1}\boldsymbol{D}\boldsymbol{M}(\mathbf{I}_q - \operatorname{prox}_{\gamma H}^{\boldsymbol{M}}) \circ \boldsymbol{D}^\top)$.

Considérons maintenant la discrétisation Euler explicite de (6) avec $\delta > 0$, laquelle peut s'écrire comme

$$\boldsymbol{L}_{k+1} = \boldsymbol{L}_k + \delta\mathcal{A}(\boldsymbol{L}_k) + \sqrt{\delta}\boldsymbol{Z}_k,\ t > 0,\ \boldsymbol{L}_0 = \boldsymbol{l}_0. \quad (7)$$

Une solution approchée est définie par l'extension continue de (7)

$$\boldsymbol{L}^\delta(t) \stackrel{\text{def}}{=} \boldsymbol{L}_0 + \int_0^t \mathcal{A}(\overline{\boldsymbol{L}}(s))ds + \int_0^t d\boldsymbol{W}(s), \quad (8)$$

où $\overline{\boldsymbol{L}}(t) = \boldsymbol{L}_k$ pour $t \in [k\delta, (k+1)\delta[$. Notons que $\boldsymbol{L}^\delta(k\delta) = \overline{\boldsymbol{L}}(k\delta) = \boldsymbol{L}_k$. Ainsi, $\boldsymbol{L}^\delta(t)$ et $\overline{\boldsymbol{L}}(t)$ sont des extensions en temps continu de la chaîne discrète $\{\boldsymbol{L}_k\}_k$. On établit ainsi que l'espérance du schéma d'approximation (7) converge vers celle de $\boldsymbol{L}$ en norme ℓ_2.

Théorème 3.1. *Supposons que* $(\mathbf{H.1})$-$(\mathbf{H.5})$ *sont satisfaites, et* $\mathbb{E}\left[\left\|\boldsymbol{L}(0)\right\|_2^m\right] < \infty$ *pour tout* $m \geq 2$. *Alors*

$$\left\|\mathbb{E}\left[\boldsymbol{L}^\delta(T)\right] - \mathbb{E}\left[\boldsymbol{L}(T)\right]\right\|_2 \leq \mathbb{E}\sup_{0\leq t\leq T}\left\|\boldsymbol{L}^\delta(t) - \boldsymbol{L}(t)\right\|_2 \xrightarrow[\delta\to 0]{} 0.$$

4 Algorithmes explicites-implicites

Nous en venons maintenant à notre objectif original : le calcul numérique de $\widehat{\boldsymbol{\theta}}_n^{\text{EWA}}$ dans (2) par l'échantillonnage de $\widehat{\mu}$. Dans la suite, on pose $F_\beta = F(\boldsymbol{X}\cdot, \boldsymbol{y})/\beta$. On suppose que J_λ satisfait $(\mathbf{H.1})$ et $F(\boldsymbol{X}\cdot, \boldsymbol{y}) \in C^1(\mathbb{R}^p)$ satisfait $(\mathbf{H.5})$. Nous allons maintenant décrire deux algorithmes type explicite-implicite pour implémenter (2) en tirant profit de la structuration de $\widehat{\mu}$. Ces algorithmes reposent sur une spécialisation adéquate des résultats de la section précédente.

4.1 LMC Explicite-Implicite

Dans (4), prenons $H = F_\beta + J_\lambda/\beta$, $L \equiv 0$ et $\boldsymbol{D} = \mathbf{I}_p$, avec F une fonction de perte quadratique, i.e., $F_\beta(\boldsymbol{\theta}) = \left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\right\|_2^2/\beta$. Observons que H satisfait $(\mathbf{H.1})$ par hypothèse sur J_λ. Afin de relier $(\mathbf{H.2})$ et $(\mathbf{H.3})$ à $\operatorname{prox}_{\gamma J_\lambda/\beta}$, nous allons proposer une métrique $\boldsymbol{M}$ dans laquelle $\operatorname{prox}_{\gamma H}^{\boldsymbol{M}}$ s'exprime en fonction de $\operatorname{prox}_{\gamma J_\lambda/\beta}$. Cette idée est mise en oeuvre dans le lemme suivant issu de [5, Lemme 7.1].

Lemme 4.1. *Supposons que* $\gamma \in \left]0, \beta/(2\left|\!\left|\!\left|\boldsymbol{X}\right|\!\right|\!\right|^2)\right[$ *et que* J_λ *vérifie* $(\mathbf{H.1})$. *On définit* $\boldsymbol{M}_\gamma \stackrel{\text{def}}{=} \mathbf{I}_p - \frac{2\gamma}{\beta}\boldsymbol{X}^\top\boldsymbol{X}$, *qui est symétrique définie positive. Alors*

$$\operatorname{prox}_{\gamma H}^{\boldsymbol{M}_\gamma} = \operatorname{prox}_{\gamma J_\lambda/\beta} \circ (\mathbf{I}_p - \gamma\nabla F_\beta). \quad (9)$$

Ainsi, on obtient que $\operatorname{prox}_{\gamma H}^{\boldsymbol{M}_\gamma}$ satisfait $(\mathbf{H.2})$ et $(\mathbf{H.3})$. Pour appliquer les résultats de la Section 3, il faut vérifier $(\mathbf{H.4})$, ce qui est vrai dès lors qu'on impose que J_λ vérifie

$(\mathbf{H.4'})$ $\exists C_1' > 0, \forall\boldsymbol{\theta} \in \mathbb{R}^p, \left\langle \operatorname{prox}_{\gamma J_\lambda/\beta}(\boldsymbol{\theta} - \gamma\nabla F_\beta(\boldsymbol{\theta})), \boldsymbol{\theta}\right\rangle_{\boldsymbol{M}_\gamma} \leq C_1'(1 + \left\|\boldsymbol{\theta}\right\|_2^2)$.

À partir des Lemmes 3.1 et 4.1, on obtient

$$\nabla^{\boldsymbol{M}_\gamma,\gamma}H = \gamma^{-1}\boldsymbol{M}_\gamma\left(\mathbf{I}_p - \operatorname{prox}_{\gamma J_\lambda/\beta} \circ (\mathbf{I}_p - \gamma\nabla F_\beta)\right).$$

Au vu de cette expression, la discrétisation d'Euler explicite de l'EDS (5) se lit dans ce cas

$$\begin{aligned}\boldsymbol{L}_{k+1} = {} & (\mathbf{I} - \tfrac{\delta}{2\gamma}\boldsymbol{M}_\gamma)\boldsymbol{L}_k + \tfrac{\delta}{2\gamma}\boldsymbol{M}_\gamma\operatorname{prox}_{\gamma J_\lambda/\beta}(\boldsymbol{L}_k - \gamma\nabla F_\beta(\boldsymbol{L}_k)) \\ & + \sqrt{\delta}\boldsymbol{Z}_k,\ t > 0,\ \boldsymbol{L}_0 = \boldsymbol{l}_0, \end{aligned} \quad (10)$$

et $\widehat{\boldsymbol{\theta}}_n^{\text{EWA}}$ est alors approximé par $(\delta/T\sum_{k=1}^{\lfloor T/\delta\rfloor - 1}\boldsymbol{L}_k)$.

Remarque 4.1. *Observons qu'hormis le terme de diffusion stochastique,* (10) *est une forme relaxée de l'algorithme explicite-implicite dans la métrique* $\boldsymbol{M}_\gamma$. *De plus, supposons qu'au lieu de* (5), *on considère l'EDS*

$$d\boldsymbol{L}(t) = -\frac{1}{2}\nabla\left((L + ({}^{\boldsymbol{M},\gamma}H) \circ \boldsymbol{D}^\top) \circ \boldsymbol{M}^{-1/2}\right)(\boldsymbol{L}(t))dt + d\boldsymbol{W}(t),$$

dont la distribution stationnaire converge aussi vers μ *sous les mêmes hypothèses que Proposition 3.3. En discrétisant cette dernière EDS par Euler explicite et en appliquant le changement de variable* $\boldsymbol{U}_k = {\boldsymbol{M}_\gamma}^{-1/2}\boldsymbol{L}_k$, *on obtient*

$$\begin{aligned}\boldsymbol{U}_{k+1} = {} & (\mathbf{I} - \tfrac{\delta}{2\gamma})\boldsymbol{U}_k + \tfrac{\delta}{2\gamma}\operatorname{prox}_{\gamma J_\lambda/\beta}(\boldsymbol{U}_k - \gamma\nabla F_\beta(\boldsymbol{U}_k)) \\ & + \sqrt{\delta}{\boldsymbol{M}_\gamma}^{-1/2}\boldsymbol{Z}_k,\ t > 0,\ \boldsymbol{L}_0 = \boldsymbol{l}_0,\end{aligned}$$

ce qui n'est pas sans rappeler la forme relaxée de l'algorithme explicite-implicite. Ceci donne d'ailleurs une justification théorique à l'intuition numérique de [8].

4.2 LMC Semi Explicite-Implicite

Une limitation essentielle de (10) est que l'opérateur proximal de J_λ soit être facilement calculable, i.e. une fonction simple. Ceci n'est typiquement pas vrai si par exemple J_λ est un a priori analyse, i.e., $J_\lambda = W_\lambda \circ \boldsymbol{D}^\top$ où $\boldsymbol{D} \in \mathbb{R}^{p\times q}$ est l'opérateur analyse et $W_\lambda : \mathbb{R}^q \to \mathbb{R} \cup \{+\infty\}$. On note $W_{\beta,\lambda} = W_\lambda/\beta$. On suppose que $W_{\beta,\lambda}$ satisfait $(\mathbf{H.1})$-$(\mathbf{H.4})$ avec $\boldsymbol{M} = \mathbf{I}_q$, et $F_\beta \in C^1(\mathbb{R}^p)$ satisfait $(\mathbf{H.5})$. Il est alors légitime d'appliquer les résultats de la section précédente en prenant $L = F_\beta$, $H = W_{\beta,\lambda}$, et $\boldsymbol{M} = \mathbf{I}_q$. Ainsi, Proposition 3.1 donne

$$\nabla({}^\gamma H \circ \boldsymbol{D}^\top)(\boldsymbol{\theta}) = \gamma^{-1}\boldsymbol{D}(\boldsymbol{D}^\top\boldsymbol{\theta} - \operatorname{prox}_{\gamma W_{\beta,\lambda}}(\boldsymbol{D}^\top\boldsymbol{\theta})),$$

et la a discrétisation d'Euler explicite de (5) devient

$$\begin{aligned}\boldsymbol{L}_{k+1} = {} & \boldsymbol{L}_k - \frac{\delta}{2}\nabla F_\beta(\boldsymbol{L}_k) - \frac{\delta}{2\gamma}\boldsymbol{D}\big(\boldsymbol{D}^\top\boldsymbol{L}_k \\ & - \operatorname{prox}_{\gamma W_{\beta,\lambda}}(\boldsymbol{D}^\top\boldsymbol{L}_k)\big) + \sqrt{\delta}\boldsymbol{Z}_k,\ t > 0,\ \boldsymbol{L}_0 = \boldsymbol{l}_0.\end{aligned}$$

5 Applications et expériences numériques

On applique nos algorithmes afin de calculer l'estimateurs EWA pour divers problèmes inverses avec des a priori de type analyse par groupes. Soient $\{\mathcal{G}_1, \ldots, \mathcal{G}_L\}_{\mathcal{G}_l \subset \{1,\ldots,q\}}$ où $\bigcup_{l=1}^{L} \mathcal{G}_l = \{1, \ldots, q\}$ et $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset$ lorsque $i \neq j$. La pénalité J_λ est définie par $J_\lambda(\boldsymbol{\theta}) = \sum_{l=1}^{L} w_\lambda(\left\|[\boldsymbol{D}^\top \boldsymbol{\theta}]_{\mathcal{G}_l}\right\|_2)$, où $w_\lambda : \mathbb{R}^+ \to \mathbb{R}$ satisfait les hypothèse suivantes

$(\mathbf{W.1})$ w_λ est une fonction croissante et bornée inférieurement sur $(0, +\infty)$.

$(\mathbf{W.2})$ w_λ est continument différentiable sur $(0, +\infty)$ et le problème $\min_{t \in [0,+\infty)} \{t + \frac{\gamma}{\beta} w'_\lambda(t)\}$ admet une solution unique en 0 pour quelque $\gamma > 0$.

Le lemme suivant détaille le calcul numérique du terme $\operatorname{prox}_{\gamma W_\lambda/\beta}$ dans les algorithmes.

Lemme 5.1. *Supposons que* $(\mathbf{W.1})$ *et* $(\mathbf{W.2})$ *sont satisfaites pour quelque* $\gamma > 0$. *Pour tout* $\boldsymbol{u} \in \mathbb{R}^q$, *on a* $\operatorname{prox}_{\gamma W_\lambda/\beta}(\boldsymbol{u}) = \left(\operatorname{prox}_{\gamma w_\lambda/\beta}(\|\boldsymbol{u}_{\mathcal{G}_1}\|_2) \frac{\boldsymbol{u}_{\mathcal{G}_1}^T}{\|\boldsymbol{u}_{\mathcal{G}_1}\|_2}, \ldots, \operatorname{prox}_{\gamma w_\lambda/\beta}(\|\boldsymbol{u}_{\mathcal{G}_L}\|_2) \frac{\boldsymbol{u}_{\mathcal{G}_L}^T}{\|\boldsymbol{u}_{\mathcal{G}_L}\|_2}\right)^T$ *où* $\forall t \geq 0$,

$$\operatorname{prox}_{\gamma w_\lambda/\beta}(t) = \begin{cases} 0 & \textit{if } t \leq \frac{\gamma}{\beta} w'_\lambda(0^+), \\ t - \frac{\gamma}{\beta} w'_\lambda\Big(\operatorname{prox}_{\gamma w_\lambda/\beta}(t)\Big) & \textit{if } t > \frac{\gamma}{\beta} w'_\lambda(0^+). \end{cases} \tag{11}$$

L'objetif est de former une famille de pénalités qui permet d'établir $(\mathbf{H.1'})$ - $(\mathbf{H.3'})$, $(\mathbf{H.4'} - \mathbf{FB})$ et $(\mathbf{H.4'} - \mathbf{SFB})$. En effet, on peut vérifier ces hypothèses en utilisant Lemme 5.1 et [6, Lemme 7.3].

Nos expériences numériques illustrent la performance de l'estimateur EWA avec trois pénalités $\ell_{1,2}$, i.e. $w_\lambda(t) = \lambda t$, FIRM et SCAD (voir [6, Section 7]) (désignés respectivement par EWA-$\ell_{1,2}$, EWA-FIRM et EWA-SCAD) pour résoudre des problèmes d'inpainting, de déconvolution et d'échantillonnage compressée (CS) d'image avec un bruit additif centré Gaussien blanc. Ces pénalités vérifient toutes les hypothèses $(\mathbf{W.1})$ et $(\mathbf{W.2})$. Le niveau de bruit σ est en fonction du rapport signal sur $\sigma = \frac{\left\|\boldsymbol{X}\boldsymbol{\theta}_0\right\|_2}{\sqrt{n}10^{\mathrm{RSB}/10}}$. Dans nos expériences, $\mathrm{RSB} = 5\mathrm{dB}$. L'opérateur d'analyse $\boldsymbol{D}^\top$ correspond au gradient discret d'une image avec des conditions aux bords de Neumann (ceci revient à un a priori de type variation totale "isotrope" avec donc des blocs de taille 2). Du fait de la présence d'un opérateur analyse qui n'est pas unitaire, nous avons appliqué l'algorithme semi explicite-implicite pour calculer EWA avec $\beta = 1/(pn)$, $\gamma = \beta$, et $\delta = \{5\beta/10^3, 5\beta/10^2, \beta/10^6\}$ respectivement associés aux problèmes d'inpainting, de déconvolution et d'échantillonnage compressée. Les valeurs de β et γ sont choisies à partir de [7, Théorème 4.1] et [6, Section 7], et le pas de discrétisation δ est suffisamment petit pour que l'algorithme converge.

L'image originale, masquée et convoluée, et les résultats numériques sont montrés en Figure 1.

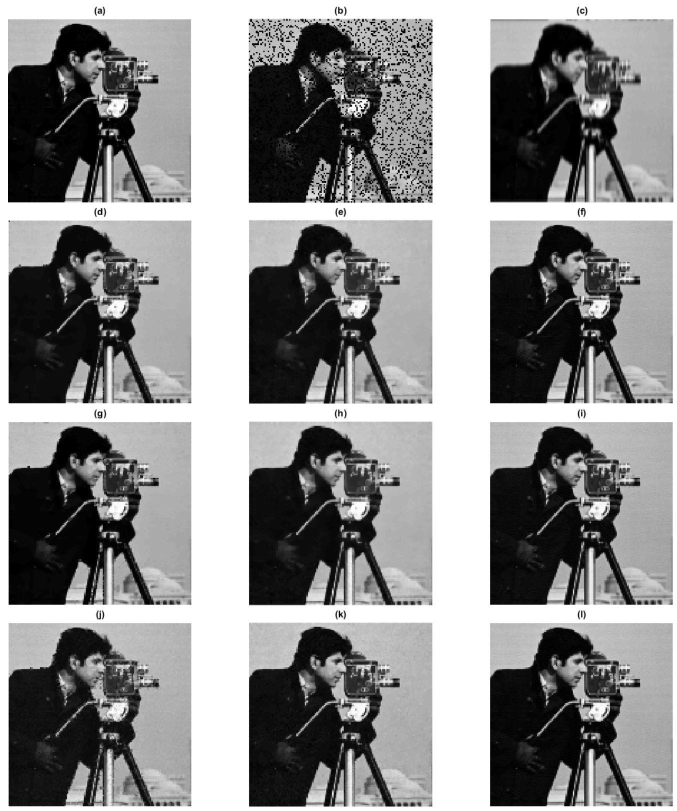

FIGURE 1 – (a, b, c) : Image originale, masquée et convoluée. (d, e, f) : Inpainting, CS et Déconvolution avec EWA - $\ell_{1,2}$. (g, h, i) : Inpainting, CS et Déconvolution avec EWA-FIRM. (j, k, l) : Inpainting, CS et Déconvolution avec EWA-SCAD.

Références

[1] A. S. Dalalyan. Theoretical guarantees for approximate sampling from a smooth and log-concave density. to appear in JRSS B 1412.7392, arXiv, December 2014.

[2] A. S. Dalalyan and A. B. Tsybakov. Sparse regression learning by aggregation and langevin monte-carlo. *J. Comput. Syst. Sci.*, 78(5) :1423–1443, Sept. 2012.

[3] A. Durmus and E. Moulines. Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm. Preprint hal-01176132, July 2015.

[4] A. Durmus, E. Moulines, and M. Pereyra. Sampling from convex non continuously differentiable functions, when Moreau meets Langevin. Preprint hal-01267115, 2016.

[5] T. Duy Luu, J. M. Fadili, and C. Chesneau. PAC-Bayesian risk bounds for group-analysis sparse regression by exponential weighting. Technical report, hal-01367742, 2016.

[6] T. D. Luu, J. Fadili, and C. Chesneau. Sampling from non-smooth distribution through Langevin diffusion. Technical report, hal-01492056, Mar. 2017.

[7] T. D. Luu, J. Fadili, and C. Chesneau. Sharp oracle inequalities for low-complexity priors. Technical Report arXiv : 1702.03166, 2017.

[8] M. Pereyra. Proximal markov chain monte carlo algorithms. *Statistics and Computing*, 26(4) :745–760, 2016.

[9] R. A. Poliquin and R. T. Rockafellar. Prox-regular functions in variational analysis. *Transaction of the American Mathematical Society*, 348(5), 1996.

[10] G. O. Roberts and R. L. Tweedie. Exponential Convergence of Langevin Distributions and Their Discrete Approximations. *Bernoulli*, 2(4) :341–363, 1996.