Algorithmes de Monte Carlo par chaînes de Markov et méthodes particulaires

Benjamin JOURDAIN

4 mars 2025

Table des matières

Chapitre 1

Algorithme de Metropolis Hastings et convergence de chaînes de Markov

1.1 noyau markovien et chaîne de Markov

On munit l'espace d'états E d'une tribu $\mathcal{E}$.

Définition 1.1.1. — *On appelle noyau markovien sur $(E, \mathcal{E})$, une application $P : E \times \mathcal{E} \mapsto [0, 1]$ telle que*
- — *$\forall x \in E$, $\mathcal{E} \ni A \mapsto P(x, A)$ est une probabilité,*
- — *$\forall A \in \mathcal{E}$, $E \ni x \mapsto P(x, A)$ est mesurable.*

— *On appelle chaîne de Markov de noyau de transition P, un processus $(X_n)_{n\in\mathbb{N}}$ à valeurs dans E adapté à la filtration $(\mathcal{F}_n)_{n\geq 0}$ tel que*

$$\forall n \in \mathbb{N},\ \forall A \in \mathcal{E},\ \mathbb{P}(X_{n+1} \in A|\mathcal{F}_n) = P(X_n, A).$$

On adoptera les notations suivantes :
- — $\mathcal{P}(E)$ désigne l'ensemble des probabilités sur $(E, \mathcal{E})$,
- — pour $\varphi : E \to \mathbb{R}$ mesurable positive ou telle que $\forall x \in E$, $\int_E |\varphi(y)|P(x, dy) < +\infty$, on note $P\varphi$ la fonction de E dans $\mathbb{R} \cup \{+\infty\}$ définie par $P\varphi(x) = \int_E \varphi(y)P(x, dy)$,
- — pour μ dans l'espace $\mathcal{P}(E)$ des probabilités sur $(E, \mathcal{E})$, on note μP la probabilité sur $(E, \mathcal{E})$ définie par $\forall A \in \mathcal{E}$, $\mu P(A) = \int_E P(x, A)\mu(dx)$,
- — la probabilité $\pi \in \mathcal{P}(E)$ est dite invariante par P si $\pi P = \pi$,
- — pour μ une mesure positive sur $(E, \mathcal{E})$ et $\varphi : E \to \mathbb{R}$ mesurable positive ou telle que $\int_E |\varphi(x)|\mu(dx) < \infty$, on note $\mu(\varphi) = \int_E \varphi(x)\mu(dx)$,
- — pour P et Q deux noyaux markoviens sur $(E, \mathcal{E})$, on note PQ le noyau markovien sur $(E, \mathcal{E})$ défini par $\forall x \in E$, $\forall A \in \mathcal{E}$, $PQ(x, A) = \int_E Q(y, A)P(x, dy)$ et par $(P^n)_{n\in N}$ la suite de noyaux définis par $P^0(x, A) = 1_A(x)$ (i.e. $P^0(x, dy) = \delta_x(dy)$) et la relation de récurrence $P^n := PP^{n-1} = P^{n-1}P$ pour $n \geq 1$. Notons que $P^1 = P$.

Si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de noyau de transition P avec X_0 de loi de probabilité μ, alors pour tout $n \in \mathbb{N}$, la loi de X_n est μP^n.

Exercice 1.1.2. *Montrer cette propriété (on pourra commencer par remarquer que pour $\varphi : E \to \mathbb{R}$ mesurable bornée et $k \in \mathbb{N}$, $\mathbb{E}[\varphi(X_{k+1})|\mathcal{F}_k] = P\varphi(X_k)$).*

1.2 Algorithme de Metropolis-Hastings

Cet algorithme est utilisé lorsque, sur l'espace d'états E, on veut simuler suivant la probabilité π de la forme $\pi(dx) = \frac{\eta(x)\lambda(dx)}{\int_E \eta(y)\lambda(dy)}$ où

— λ est une mesure positive de référence sur $(E, \mathcal{E})$,

— η est une fonction mesurable de E dans $\mathbb{R}_+$ telle que $\int_E \eta(x)\lambda(dx) \in]0, \infty[$.

Il va permettre de construire une chaîne de Markov de probabilité invariante π sans nécessairement connaître la constante de normalisation $\frac{1}{\int_E \eta(y)\lambda(dy)}$ de la densité η. Parmi ses très nombreuses applications, on peut notamment citer :

— dans le cas où λ est la mesure de Lebesgue sur $E = \mathbb{R}^d$, la simulation en physique statistique de la probabilité de Boltzmann-Gibbs de densité proportionnelle à $\eta(x) = e^{-\frac{1}{k_B T}V(x)}$ où k_B désigne la constante de Boltzmann, T la température et $V : \mathbb{R}^d \to \mathbb{R}$ est la fonction potentiel,

— dans le cas où E est l'espace des valeurs possibles d'un paramètre θ en statistique bayésienne, la simulation suivant la loi à posteriori de ce paramètre sachant que l'on a observé $Y = y$: si on note $p_{Y|\Theta}(y|\theta)$ la densité de l'observation Y lorsque le paramètre vaut θ et $p_\Theta(\theta)$ la densité à priori de Θ par rapport à λ, la formule de Bayes assure que sa densité à posteriori est

$$p_{\Theta|Y}(\theta|y) = \frac{p_{Y|\Theta}(y|\theta)p_\Theta(\theta)}{\int_E p_{Y|\Theta}(y|\vartheta)p_\Theta(\vartheta)\lambda(d\vartheta)}.$$

Notons que dans ces deux exemples, le calcul de la constante de normalisation est difficile.

Soit $q : E \times E \to \mathbb{R}_+$ une fonction mesurable telle que $\forall x \in E$, $\int_E q(x,y)\lambda(dy) = 1$ ($Q(x,dy) = q(x,y)\lambda(dy)$ est alors un noyau markovien) et on sait simuler suivant la probabilité $q(x,y)\lambda(dy)$. On pose

$$\alpha(x,y) = \begin{cases} \min\left(1, \frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) \text{ si } \eta(x)q(x,y) > 0 \\ 1 \text{ si } \eta(x)q(x,y) = 0 \end{cases}. \tag{1.1}$$

Partant d'une variable aléatoire X_0 à valeurs dans E, le principe de l'algorithme de Metropolis-Hastings est de construire $(X_n)_{n\in\mathbb{N}}$ par la récurrence suivante :

— sachant $(X_0, \dots, X_n)$, on génère une proposition Y_{n+1} suivant la probabilité $q(X_n, y)\lambda(dy)$ et indépendamment une variable aléatoire U_{n+1} uniforme sur $[0,1]$,

— on pose $X_{n+1} = Y_{n+1}1_{\{U_{n+1}\le\alpha(X_n,Y_{n+1})\}} + X_n 1_{\{U_{n+1}>\alpha(X_n,Y_{n+1})\}}$, c'est-à-dire que la proposition est acceptée avec probabilité $\alpha(X_n, Y_{n+1})$ et la position X_n est conservée sinon.

Ainsi pour $f : E \to \mathbb{R}$ mesurable bornée,

$$\begin{aligned}
&\mathbb{E}[f(X_{n+1})|X_0, X_1, \dots, X_n] \\
&= \mathbb{E}[\mathbb{E}[f(Y_{n+1})1_{\{U_{n+1}\le\alpha(X_n,Y_{n+1})\}} + f(X_n)1_{\{U_{n+1}>\alpha(X_n,Y_{n+1})\}}|X_0, X_1, \dots, X_n, Y_{n+1}|X_0, X_1, \dots, X_n] \\
&= \mathbb{E}[f(Y_{n+1})\alpha(X_n, Y_{n+1}) + f(X_n)(1-\alpha(X_n, Y_{n+1}))|X_0, X_1, \dots, X_n] \\
&= \int_E f(y)\alpha(X_n, y)q(X_n, y)\lambda(dy) + f(X_n)\int_E (1-\alpha(X_n, y))q(X_n, y)\lambda(dy) \\
&= \int_E f(y)P(X_n, dy)
\end{aligned}$$

où

$$P(x,dy) = \alpha(x,y)q(x,y)\lambda(dy) + R(x)\delta_x(dy) \text{ avec } R(x) = \int_E (1-\alpha(x,z))q(x,z)\lambda(dz). \tag{1.2}$$

Ainsi $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de noyau de transition P. Pour $y \neq x$,

$$\begin{aligned}\eta(x)q(x,y)\alpha(x,y) &= \begin{cases}\eta(x)q(x,y)\min\left(1, \frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) \text{ si } \eta(x)q(x,y) > 0\\ \eta(x)q(x,y)\times 1 \text{ si } \eta(x)q(x,y) = 0\end{cases}\\ &= \min(\eta(x)q(x,y), \eta(y)q(y,x)).\end{aligned}$$

est une fonction symétrique de (x,y) i.e. $\eta(x)q(x,y)\alpha(x,y) = \eta(y)q(y,x)\alpha(y,x)$. Cela entraîne que

$$\eta(x)\lambda(dx)\alpha(x,y)q(x,y)\lambda(dy) = \eta(y)\lambda(dy)\alpha(y,x)q(y,x)\lambda(dx).$$

Comme clairement, $\eta(x)\lambda(dx)R(x)\delta_x(dy) = \eta(y)\lambda(dy)R(y)\delta_y(dx)$, on en déduit que $\pi(dx)P(x,dy) = \pi(dy)P(y,dx)$. On dit que la probabilité π est réversible pour le noyau P. Comme $P(y,E) = 1$, par intégration en x sur E, cela implique que $\pi P = \pi$: la probabilité π est invariante par le noyau P. Notons que dans le cas particulier où $E = \mathbb{R}^d$ et $q(x,y) = \psi(y-x)$ pour une densité de probabilité paire ψ par rapport à la mesure de Lebesgue λ (exemple : $\psi(z) = e^{-\frac{|z|^2}{2\sigma^2}}/(2\pi\sigma^2)^{d/2}$), alors le rapport $\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}$ se simplifie en $\frac{\eta(y)}{\eta(x)}$. Comme les variables aléatoires $Y_{n+1} - X_n$ sont alors i.i.d. suivant la densité ψ, on parle d'algorithme de Metropolis-Hastings par marche aléatoire.

Remarque 1.2.1. — *Notons que la réversibilité de π pour le noyau P est préservée pour*

$$\alpha(x,y) = \begin{cases} a\left(\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) \text{ si } \eta(x)q(x,y) > 0\\ 1 \text{ si } \eta(x)q(x,y) = 0\end{cases},$$

où la fonction $a : \mathbb{R}_+ \to [0,1]$ vérifie $a(0) = 0$ et $\forall u > 0,\ a(u) = ua(1/u)$ si bien que $a(u) \leq \min(1,u)$. Le choix $a(u) = \min(1,u)$ fait dans l'algorithme de Metropolis-Hastings maximise la probabilité d'accepter les propositions. Il est privilégié en pratique pour ses meilleures propriétés asymptotiques (voir le paragraphe 1.6) mais d'autres sont possibles comme $a(u) = \frac{\sum_{k=1}^K u^k}{1+\sum_{k=1}^K u^k}$ pour $K \in \mathbb{N}^*$.

— *Notons enfin que le choix de la valeur de $\alpha(x,y)$ lorsque $q(x,y) = 0$ ne change rien au noyau P. En revanche, choisir $\alpha(x,y) = 1$ lorsque $\eta(x) = 0$ et $q(x,y) > 0$ permet de maximiser la probabilité de bouger de l'état x.*

Exercice 1.2.2. *On suppose que π est invariante pour le noyau P. Soient $f, g : E \to \mathbb{R}$ mesurables et telles que $\pi(f^2 + g^2) < \infty$.*

1. *Montrer que $(Pf)^2 \leq Pf^2$ et en déduire à l'aide de l'inégalité de Cauchy-Schwarz que $\pi(|gPf|) < \infty$.*
2. *Lorsque π est réversible pour P, montrer que $\pi(gPf) = \pi(fPg)$.*

Dans la suite de ce chapître, nous allons donner des conditions sur le noyau de transition P d'une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ à valeurs dans E assurant que

— la chaîne admet une unique probabilité invariante π et, lorsque $n \to \infty$, la loi de X_n converge à vitesse géométrique vers π dans une norme bien choisie,

— pour toute fonction $f : E \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$, les moyennes ergodiques $\frac{1}{n}\sum_{k=0}^{n} f(X_k)$ convergent p.s. vers $\pi(f)$,

— le comportement asymptotique de l'erreur dans la loi des grands nombres ergodique précédente est régi par un théorème de la limite centrale.

1.3 Ergodicité d'une chaîne de Markov à espace d'état général

Soit P un noyau markovien sur $(E, \mathcal{E})$. Nous allons nous intéresser à des conditions assurant que pour toute probabilité $\mu \in \mathcal{P}(E)$, μP^n converge lorsque $n \to \infty$ vers π où $\pi \in \mathcal{P}(E)$ est invariante c'est-à-dire vérifie $\pi P = \pi$. Pour cela, nous commençons par introduire une distance permettant de quantifier cette convergence.

1.3.1 distance en variation totale

Définition 1.3.1. *On appelle distance en variation totale sur $\mathcal{P}(E)$ la distance $d_{\mathrm{TV}}(\mu, \sigma) := \sup_{A \in \mathcal{E}} |\mu(A) - \sigma(A)|$.*

Pour obtenir une expression alternative de cette distance, nous introduisons, la décomposition de Jordan-Hahn de la mesure signée $\mu - \sigma$: $\mu - \sigma = \xi^+ - \xi^-$ où ξ^+ et ξ^- sont deux mesures positives sur $(E, \mathcal{E})$ telles qu'il existe $B \in \mathcal{E}$ vérifiant $\xi^-(B) = \xi^+(B^c) = 0$.

Exercice 1.3.2. *Montrer l'existence et l'unicité de la décomposition de Jordan-Hahn lorsque l'espace d'états E est fini ou dénombrable et $\mathcal{E}$ l'ensemble des parties de E (tribu discrète).*

Proposition 1.3.3. *La mesure $\mu \wedge \sigma$ sur $(E, \mathcal{E})$ définie par $\forall A \in \mathcal{E},\ \mu \wedge \sigma(A) = \sigma(A \cap B) + \mu(A \cap B^c)$ est la plus grande mesure positive majorée à la fois par μ et par σ. En outre,*

$$d_{\mathrm{TV}}(\mu, \sigma) = 1 - \mu \wedge \sigma(E) = \sup_{\varphi : |\varphi| \leq 1/2} |\mu(\varphi) - \sigma(\varphi)|,$$

où le supremum porte sur l'ensemble des fonctions mesurables $\varphi : E \to \mathbb{R}$ telles que $\sup_{x \in E} |\varphi(x)| \leq 1/2$.

Attention, généralement, en dehors des cas $A \subset B$ et $A \subset B^c$, $\mu \wedge \sigma(A) \neq \mu(A) \wedge \sigma(A)$.

Exercice 1.3.4. *Montrer que pour tout $\gamma > 0$, $d_{\mathrm{TV}}(\mu, \sigma) = \frac{1}{2\gamma} \sup_{\varphi : |\varphi| \leq \gamma} |\mu(\varphi) - \sigma(\varphi)|$.*

Exercice 1.3.5. *On suppose que $\mu(dx) = f(x)\lambda(dx)$, $\sigma(dx) = g(x)\lambda(dx)$ avec $f, g : E \to \mathbb{R}_+$ deux fonctions mesurables telles que $\int_E f(x)\lambda(dx) = \int_E g(x)\lambda(dx) = 1$.*

1. *Calculer $\mu(A) - \sigma(A)$ pour $A \in \mathcal{E}$ et en déduire que*

$$\begin{aligned} d_{\mathrm{TV}}(\mu, \sigma) &= \int_E \max(f(x) - g(x), 0)\lambda(dx) = \int_E \max(g(x) - f(x), 0)\lambda(dx) \\ &= \frac{1}{2}\int_E |g(x) - f(x)|\lambda(dx). \end{aligned}$$

2. *Vérifier que* $\xi^+(dx) = 1_{\{f(x)\geq g(x)\}}(f(x)-g(x))\lambda(dx)$ *et* $\xi^-(dx) = 1_{\{g(x)\geq f(x)\}}(g(x)-f(x))\lambda(dx)$ *est la décomposition de Jordan-Hahn de* $\mu-\sigma$ *et préciser les ensembles* B *et* B^c.
3. *Vérifier que* $\mu\wedge\sigma(dx) = (f(x)\wedge g(x))\lambda(dx)$.

Pour μ et σ quelconques, le théorème de Radon-Nikodym assure que μ et σ admettent des densités par rapport à $\lambda = \mu+\sigma$ et l'exercice précédent permet d'en déduire l'existence de la décomposition de Jordan-Hahn.

Démonstration : Pour $A \in \mathcal{E}$, comme, $\mu(A\cap B)-\sigma(A\cap B) = \xi^+(A\cap B) \geq 0$ et $\sigma(A\cap B^c)-\mu(A\cap B^c) = \xi^-(A\cap B^c)\geq 0$, $\mu\wedge\sigma(A) \leq \min(\mu(A),\sigma(A))$.

Par ailleurs, si η est une mesure positive telle que $\forall C \in \mathcal{E}$, $\eta(C) \leq \min(\mu(C),\sigma(C))$, pour $A\in\mathcal{E}$, $\eta(A) = \eta(A\cap B)+\eta(A\cap B^c) \leq \sigma(A\cap B)+\mu(A\cap B^c) = \mu\wedge\sigma(A)$.

Notons que comme $0 = \mu(E)-\sigma(E) = \xi^+(E\cap B)-\xi^-(E\cap B^c) = \xi^+(B)-\xi^-(B^c)$, $\xi^+(B) = \xi^-(B^c)$. Pour $A\in\mathcal{E}$, comme $\mu(A)-\sigma(A) = \xi^+(A\cap B)-\xi^-(A\cap B^c)$,

$$-\xi^-(B^c) \leq \mu(A)-\sigma(A) \leq \xi^+(B),$$

où la première inégalité est une égalité pour $A = B^c$ et la seconde pour $A = B$. On en déduit que

$$\begin{aligned} d_{\mathrm{TV}}(\mu,\sigma) &= \xi^+(B) = \mu(B)-\sigma(B) = (1-\mu(B^c))-\sigma(B) = 1-(\mu(E\cap B^c)+\sigma(E\cap B)) \\ &= 1-\mu\wedge\sigma(E). \end{aligned} \tag{1.3}$$

Pour $\varphi : E\to\mathbb{R}$ mesurable telle que $\sup_{x\in E}|\varphi(x)| \leq 1/2$, $\xi^+(\varphi)$ (resp. $\xi^-(\varphi)$) est maximum (resp. minimum) pour φ égale à $1/2$ sur B (resp. à $-1/2$ sur B^c) et ne dépend pas des valeurs prises par φ sur B^c (resp. B). Ainsi,

$$\begin{aligned} \mu(\varphi)-\sigma(\varphi) &= \xi^+(\varphi)-\xi^-(\varphi) \leq \xi^+(1_B-1/2)-\xi^-(1_B-1/2) \\ &= \mu(1_B-1/2)-\sigma(1_B-1/2) = \mu(B)-\sigma(B) = d_{\mathrm{TV}}(\mu,\sigma). \end{aligned}$$

Par symétrie, $\sigma(\varphi)-\mu(\varphi) \leq \sigma(B^c)-\mu(B^c) = d_{\mathrm{TV}}(\mu,\sigma)$, ce qui achève la démonstration. □

Remarque 1.3.6. *Pour* $A\in\mathcal{E}$,

$$\mu(A)-\mu\wedge\sigma(A) = \mu(A)-(\sigma(A\cap B)+\mu(A\cap B^c)) = \mu(A\cap B)-\sigma(A\cap B) = \xi_+(A)$$

si bien que $\xi_+ = \mu-\mu\wedge\sigma$, *ce qui entraîne* $\xi_- = \sigma-\mu+\xi_+ = \sigma-\mu\wedge\sigma$.

Lemme 1.3.7. *Pour tous* $\mu,\sigma\in\mathcal{P}(E)$, *on a* $d_{\mathrm{TV}}(\mu,\sigma) = \inf_{X\sim\mu,Y\sim\sigma}\mathbb{P}(X\neq Y)$ *où l'infimum porte sur tous les couples de variables aléatoires* (X,Y) *avec* X *de loi* μ *et* Y *de loi* σ.

Démonstration : Pour $X\sim\mu$ et $Y\sim\sigma$ définies sur le même espace de probabilité et $A\in\mathcal{E}$, on a

$$\begin{aligned} \mu(A)-\sigma(A) &= \mathbb{P}(X\in A, X=Y)+\mathbb{P}(X\in A, X\neq Y)-\mathbb{P}(Y\in A, X=Y)-\mathbb{P}(Y\in A, X\neq Y) \\ &= \mathbb{P}(X\in A, X\neq Y)-\mathbb{P}(Y\in A, X\neq Y) \end{aligned}$$

si bien que $-\mathbb{P}(X \neq Y) \leq -\mathbb{P}(Y \in A, X \neq Y) \leq \mu(A) - \sigma(A) \leq \mathbb{P}(X \in A, X \neq Y) \leq \mathbb{P}(X \neq Y)$. On en déduit que $\mathbb{P}(X \neq Y) \geq \sup_{A\in\mathcal{E}} |\mu(A) - \sigma(A)| = d_{\mathrm{TV}}(\mu, \sigma)$. Lorsque $d_{\mathrm{TV}}(\mu, \sigma) = 1$, l'inégalité est une égalité pour tout couple (X, Y) avec $X \sim \mu$ et $Y \sim \sigma$. Lorsque $d_{\mathrm{TV}}(\mu, \sigma) = 0$, c'est une égalité pour $Y = X \sim \mu$. Nous allons maintenant exhiber un couple pour lequel c'est une égalité lorsque $0 < d_{\mathrm{TV}}(\mu, \sigma) < 1$ i.e. $0 < \mu \wedge \sigma(E) < 1$. Soit $U \sim \mathcal{U}[0, 1]$ et $(Z, \tilde{X}, \tilde{Y})$ un triplet indépendant avec $Z \sim \frac{\mu\wedge\sigma}{\mu\wedge\sigma(E)}$, $\tilde{X} \sim \frac{\mu-\mu\wedge\sigma}{1-\mu\wedge\sigma(E)}$ et $\tilde{Y} \sim \frac{\sigma-\mu\wedge\sigma}{1-\mu\wedge\sigma(E)}$ (on peut par exemple prendre Z, $\tilde{X}$ et $\tilde{Y}$ indépendantes). On pose

$$(X, Y) = 1_{\{U\leq\mu\wedge\sigma(E)\}}(Z, Z) + 1_{\{U>\mu\wedge\sigma(E)\}}(\tilde{X}, \tilde{Y}).$$

Alors pour $f : E \to \mathbb{R}$ mesurable bornée, par indépendance de U et $(Z, \tilde{X})$ on a

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[1_{\{U\leq\mu\wedge\sigma(E)\}}f(Z)] + \mathbb{E}[1_{\{U>\mu\wedge\sigma(E)\}}f(\tilde{X})] \\
&= \mathbb{P}(U \leq \mu \wedge \sigma(E))\mathbb{E}[f(Z)] + \mathbb{P}(U > \mu \wedge \sigma(E))\mathbb{E}[f(\tilde{X})] \\
&= \mu \wedge \sigma(E) \int_E f(x)\frac{\mu \wedge \sigma(dx)}{\mu \wedge \sigma(E)} + (1 - \mu \wedge \sigma(E)) \int_E f(x)\frac{\mu(dx) - \mu \wedge \sigma(dx)}{1 - \mu \wedge \sigma(E)} \\
&= \int_E f(x)\mu(dx)
\end{aligned}$$

si bien que $X \sim \mu$. Un calcul analogue de $\mathbb{E}[f(Y)]$ assure que $Y \sim \sigma$. Par ailleurs, $\mathbb{P}(X \neq Y) \leq \mathbb{P}(U > \mu \wedge \sigma(E)) = 1 - \mu \wedge \sigma(E) = d_{\mathrm{TV}}(\mu, \sigma)$. □

Exercice 1.3.8. *Soit $A \in \mathcal{E}$ t.q. $\mu(A) = 1$.*

1. *Comment comparer $\mu(A^c \cap B)$ et $\sigma(A^c \cap B)$? En déduire que $\sigma(A \cap B) = \sigma(B)$.*
2. *Montrer que $\mu(A \cap B^c) = \mu(B^c)$ et conclure que $\mu \wedge \sigma(A) = 1 - d_{\mathrm{TV}}(\mu, \sigma)$.*

Proposition 1.3.9. *L'espace $\mathcal{P}(E)$ muni de la distance en variation totale est complet.*

Démonstration : Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de Cauchy. Pour tout $n \in \mathbb{N}$, μ_n est absolument continue par rapport à la probabilité $\mu = \sum_{n\in\mathbb{N}} \frac{\mu_n}{2^{n+1}}$ et admet donc une densité $p_n : \mathbb{E} \to \mathbb{R}_+$ par rapport à μ :

$$\forall A \in \mathcal{E},\ \mu_n(A) = \int_E 1_A(x)p_n(x)\mu(dx).$$

Comme le supremum à la seconde ligne est atteint pour $\varphi(x) = \frac{1}{2}\mathrm{signe}(p_n(x) - p_m(x))$,

$$\begin{aligned}
d_{\mathrm{TV}}(\mu_n, \mu_m) &= \sup_{\varphi:|\varphi|\leq 1/2} \left| \int_E \varphi(x)(p_n(x) - p_m(x))\mu(dx) \right| \\
&= \int_E \frac{1}{2}|p_n(x) - p_m(x)|\mu(dx) = \frac{1}{2}\|p_n - p_m\|_{L^1(\mu)},
\end{aligned}$$

la suite $(p_n)_{n\in\mathbb{N}}$ est de Cauchy dans $L^1(\mu)$ qui est complet. Elle converge donc vers p_∞. Comme $\mu(dx)$ p.p. $p_n(x) \geq 0$, $\mu(dx)$ p.p. $\max(-p_\infty(x), 0) \leq |p_n(x) - p_\infty(x)|$. Ainsi $\int_E \max(-p_\infty(x), 0)\mu(dx) \leq \int_E |p_n(x) - p_\infty(x)|\mu(dx)$, puis par passage à la limite $n \to \infty$, $\int_E \max(-p_\infty(x), 0)\mu(dx) = 0$ i.e. $\mu(dx)$ p.p. $p_\infty(x) \geq 0$. Par ailleurs, le passage à la limite $n \to \infty$ dans l'inégalité $|\int_E p_\infty(x)\mu(dx) - \int_E p_n(x)\mu(dx)| \leq \int_E |p_n(x) - p_\infty(x)|\mu(dx)$ assure que $\int_E p_\infty(x)\mu(dx) = 1$. Ainsi p_∞ est une densité de probabilité par rapport à μ. Notons μ_∞ la probabilité qui admet cette densité. Comme précédemment $d_{\mathrm{TV}}(\mu_n, \mu_\infty) = \frac{1}{2}\|p_n - p_\infty\|_{L^1(\mu)}$, où le second membre tend vers 0 lorsque $n \to \infty$. □

1.3.2 Conditions de Doeblin et ergodicité uniforme

Définition 1.3.10. *On dit que le noyau markovien P satisfait*

— *la condition de Doeblin si*

$$\exists \alpha \in]0,1],\ \forall x, y \in E,\ P(x,.) \wedge P(y,.)(E) \geq \alpha.$$

— *la condition de Doeblin uniforme si*

$$\exists \nu \in \mathcal{P}(E),\ \exists \alpha \in]0,1],\ \forall A \in \mathcal{E},\ \inf_{x \in E} P(x, A) \geq \alpha \nu(A).$$

La condition de Doeblin uniforme est plus forte que la condition de Doeblin en ce que pour tous $x, y \in E$, la mesure $P(x,.) \wedge P(y,.)$ domine la même mesure $\alpha\nu$ de masse totale α.

Théorème 1.3.11. *Sous la condition de Doeblin, le noyau markovien P admet une unique probabilité invariante π et*

$$\forall \mu \in \mathcal{P}(E),\ \forall n \in \mathbb{N},\ d_{\mathrm{TV}}(\mu P^n, \pi) \leq (1-\alpha)^n d_{\mathrm{TV}}(\mu, \pi).$$

Remarque 1.3.12. *Comme $\sup_{x \in E} |\varphi(x)| \leq 1/2$ implique que $\forall x \in E$, $|P\varphi(x)| \leq \int_E |\varphi(y)| P(x, dy) \leq 1/2$, on a toujours*

$$d_{\mathrm{TV}}(\mu P, \sigma P) = \sup_{\varphi : |\varphi| \leq 1/2} |\mu(P\varphi) - \sigma(P\varphi)| \leq \sup_{\psi : |\psi| \leq 1/2} |\mu(\psi) - \sigma(\psi)| = d_{\mathrm{TV}}(\mu, \sigma), \quad (1.4)$$

même sans la condition de Doeblin.

La preuve du théorème repose sur le résultat suivant

Lemme 1.3.13. *Sous la condition de Doeblin, $\forall x, y \in E$, $|P\varphi(x) - P\varphi(y)| \leq 2(1-\alpha) \sup_{z \in E} |\varphi(z)|$ et*

$$\forall \mu, \sigma \in \mathcal{P}(E),\ d_{\mathrm{TV}}(\mu P, \sigma P) \leq (1-\alpha) d_{\mathrm{TV}}(\mu, \sigma). \quad (1.5)$$

Nous allons en déduire le théorème avant de démontrer le lemme.
Démonstration : D'après (1.5), l'application $\mathcal{P}(E) \ni \mu \mapsto \mu P \in \mathcal{P}(E)$ est contractante pour d_{TV} distance qui rend l'espace $\mathcal{P}(E)$ complet d'après la proposition 1.3.9. D'après le théorème de point fixe de Picard, elle admet un unique point fixe π qui est l'unique probabilité invariante par P. Enfin, en itérant l'inégalité (1.5), on obtient

$$\forall \mu, \sigma \in \mathcal{P}(E),\ \forall n \in \mathbb{N},\ d_{\mathrm{TV}}(\mu P^n, \sigma P^n) \leq (1-\alpha)^n d_{\mathrm{TV}}(\mu, \sigma)$$

et l'inégalité énoncée dans le théorème en découle pour le choix $\sigma = \pi$. $\square$

Démonstration : Pour $\varphi : E \to \mathbb{R}$ une fonction mesurable bornée et $x, y \in E$, on a

$$\begin{aligned}
|P\varphi(x) - P\varphi(y)| = \Big| \int_E \varphi(z)P(x,dz) - \int_E \varphi(z)P(x,.) \wedge P(y,.)(dz) \\
+ \int_E \varphi(z)P(x,.) \wedge P(y,.)(dz) - \int_E \varphi(z)P(y,dz)\Big| \\
\leq \left|\int_E \varphi(z)(P(x,dz) - P(x,.) \wedge P(y,.)(dz))\right| + \left|\int_E \varphi(z)(P(y,dz) - P(x,.) \wedge P(y,.)(dz))\right| \\
\leq \int_E |\varphi(z)|(P(x,dz) - P(x,.) \wedge P(y,.)(dz)) + \int_E |\varphi(z)|(P(y,dz) - P(x,.) \wedge P(y,.)(dz)) \\
\leq \sup_{z\in E} |\varphi(z)| \left(\int_E (P(x,dz) - P(x,.) \wedge P(y,.)(dz)) + \int_E (P(y,dz) - P(x,.) \wedge P(y,.)(dz))\right) \\
\leq 2(1-\alpha) \sup_{z\in E} |\varphi(z)|,
\end{aligned}$$

où on a utilisé la positivité des mesures $P(x,.)-P(x,.)\wedge P(y,.)$ et $P(y,.)-P(x,.)\wedge P(y,.)$ pour la seconde inégalité et la condition de Doeblin pour la quatrième inégalité. En posant $c_\varphi = (1-\alpha)\sup_{z\in E}|\varphi(z)| - \sup_{y\in E} P\varphi(y)$, on en déduit que pour $x \in E$,

$$P\varphi(x) + c_\varphi = (1-\alpha)\sup_{z\in E}|\varphi(z)| + \inf_{y\in E}(P\varphi(x) - P\varphi(y)) \geq -(1-\alpha)\sup_{z\in E}|\varphi(z)|,$$

D'autre part,

$$P\varphi(x) + c_\varphi \leq P\varphi(x) + (1-\alpha)\sup_{z\in E}|\varphi(z)| - P\varphi(x) = (1-\alpha)\sup_{z\in E}|\varphi(z)|,$$

si bien que

$$\sup_{x\in E}|P\varphi(x) + c_\varphi| \leq (1-\alpha)\sup_{z\in E}|\varphi(z)|.$$

Ainsi

$$\begin{aligned}
d_{\mathrm{TV}}(\mu P, \sigma P) &= \sup_{\varphi:|\varphi|\leq 1/2} |\mu P(\varphi) - \sigma P(\varphi)| = \sup_{\varphi:|\varphi|\leq 1/2} |\mu(P\varphi) - \sigma(P\varphi)| \\
&= \sup_{\varphi:|\varphi|\leq 1/2} |\mu(P\varphi + c_\varphi) - \sigma(P\varphi + c_\varphi)| \\
&\leq \sup_{\psi:|\psi|\leq (1-\alpha)/2} |\mu(\psi) - \sigma(\psi)| = (1-\alpha)d_{\mathrm{TV}}(\mu,\sigma),
\end{aligned}$$

où on a utilisé $\mu(c_\varphi) = c_\varphi = \sigma(c_\varphi)$ pour la troisième égalité et le résultat de l'exercice 1.3.4 pour la dernière. □

Proposition 1.3.14. *La condition de Doeblin pour une itérée P^m de P est équivalente à l'ergodicité uniforme, à savoir l'existence de $\pi \in \mathcal{P}(E)$ telle que $\lim_{n\to\infty}\sup_{x\in E} d_{\mathrm{TV}}(P^n(x,.),\pi) = 0$.*

Démonstration : Si P est uniformément ergodique, choisir m assez grand pour que $\sup_{x\in E} d_{\mathrm{TV}}(P^m(x,.),\pi) \leq \frac{1-\alpha}{2}$ assure par inégalité triangulaire que

$$\forall x, y \in E,\ 1-\alpha \geq d_{\mathrm{TV}}(P^m(x,.), P^m(y,.)) = 1 - P^m(x,.) \wedge P^m(y,.)(E).$$

Réciproquement, s'il existe $m \geq 1$ t.q. le noyau P^m satisfait la condition de Doeblin, alors le théorème 1.3.11 entraîne que P^m admet une unique probabilité invariante π. Comme $\pi P = \pi P^m P = \pi P P^m$, la probabilité πP est invariante pour P^m. Par unicité, $\pi P = \pi$ et π est invariante pour P. C'est la seule probabilité invariante pour P puisque toute probabilité invariante pour P l'est pour P^m. En combinant l'égalité $\mu P^n = \mu P^{n-\lfloor \frac{n}{m} \rfloor m} P^m \cdots P^m$ où P^m apparaît $\lfloor \frac{n}{m} \rfloor$ fois, l'estimation du théorème appliqué au noyau P^m et (1.4), on obtient que

$$\forall n \in \mathbb{N},\ d_{\mathrm{TV}}(\mu P^n, \pi) = d_{\mathrm{TV}}(\mu P^n, \pi P^n) \leq (1-\alpha)^{\lfloor n/m \rfloor} d_{\mathrm{TV}}(\mu, \pi).$$

Puisque la distance en variation totale est majorée par 1 et que, si δ_x désigne la masse de Dirac en $x \in E$, $\delta_x P^n = P^n(x, .)$, on conclut que $\lim_{n\to\infty} \sup_{x\in E} d_{\mathrm{TV}}(P^n(x,.), \pi) = 0$. $\square$

1.3.3 Algorithme de recuit simulé

Cet algorithme a pour objectif de minimiser une fonction $V : E \to \mathbb{R}$ mesurable. Pour $\beta \geq 0$, on note $Z_\beta = \int_E e^{-\beta V(x)} \lambda(dx)$ où λ est une mesure positive de référence sur $(E, \mathcal{E})$. Dans le cas d'une probabilité de Boltzmann-Gibbs, $\beta = \frac{1}{k_B T}$, c'est-à-dire que β est l'inverse de la température au facteur multiplicatif près $1/k_B$ avec k_B la constante de Boltzmann. On note également $v_\star = \inf\{v \in \mathbb{R} : \lambda(\{x : V(x) \leq v\}) > 0\}$ l'essentiel infimum de V pour la mesure λ et on suppose que $v_\star > -\infty$. On suppose enfin que

$$B := \{\beta \geq 0 : Z_\beta < \infty\} \neq \emptyset.$$

Comme $Z_\beta = e^{-\beta v_\star} \int_E e^{-\beta(V(x)-v_\star)} \lambda(dx)$ où l'intégrale est une fonction décroissante de β, B est un intervalle de la forme $[\underline{\beta}, +\infty[$ ou $]\underline{\beta}, +\infty[$ avec $\underline{\beta} \in [0, +\infty[$. Pour tout $\beta \in B$, $\pi_\beta(dx) = \frac{1}{Z_\beta} e^{-\beta V(x)} \lambda(dx)$ est une probabilité sur $(E, \mathcal{E})$. Notons que pour $\beta \geq 0$, $Z_\beta \leq e^{-\beta v_\star} \lambda(E)$ si bien que lorsque $\lambda(E) < \infty$, alors $B = [0, +\infty[$. Le lemme suivant assure que lorsque $\beta \to \infty$ (i.e. la température tend vers 0), π_β se concentre sur les minimas de V.

Lemme 1.3.15. *Pour tout $\varepsilon > 0$, $\lim_{\beta\to\infty} \pi_\beta(\{x \in E : V(x) \geq v_\star + \varepsilon\}) = 0$.*

Démonstration : On remarque que

$$\pi_\beta(dx) = \frac{1}{\hat{Z}_\beta} e^{-\beta(V(x)-v_\star)} \lambda(dx) \text{ où } \hat{Z}_\beta = \int_E e^{-\beta(V(x)-v_\star)} \lambda(dx). \tag{1.6}$$

Donc

$$\begin{aligned}
\pi_\beta(\{x \in E : V(x) \geq v_\star + \varepsilon\}) &= \frac{\int_E 1_{\{V(x) \geq v_\star+\varepsilon\}} e^{-\beta(V(x)-v_\star)} \lambda(dx)}{\int_E e^{-\beta(V(x)-v_\star)} \lambda(dx)} \\
&\leq \frac{\int_E 1_{\{V(x) \geq v_\star+\varepsilon\}} e^{-\beta(V(x)-v_\star)} \lambda(dx)}{e^{-\frac{\beta\varepsilon}{2}} \lambda(\{x : V(x) \leq v_\star + \varepsilon/2\})} \\
&\leq \frac{\int_E 1_{\{V(x) \geq v_\star+\varepsilon\}} e^{-\frac{\beta}{2}(V(x)-v_\star)} \lambda(dx)}{\lambda(\{x : V(x) \leq v_\star + \varepsilon/2\})}.
\end{aligned}$$

où $\lambda(\{x : V(x) \leq v_\star + \varepsilon/2\}) > 0$ par définition de $v_\star$ et, pour la dernière inégalité, on a utilisé que $V(x) - v_\star \geq \varepsilon$ est équivalent à $V(x) - v_\star \geq \frac{\varepsilon}{2} + \frac{V(x)-v_\star}{2}$. Lorsque $\beta \geq 2(\underline{\beta}+1)$, $e^{-(\underline{\beta}+1)(V(x)-v_\star)} \geq 1_{\{V(x)\geq v_\star+\varepsilon\}} e^{-\frac{\beta}{2}(V(x)-v_\star)}$ avec le membre de droite qui tend vers 0 lorsque $\beta \to \infty$. Le théorème de convergence dominée assure que $\lim_{\beta\to\infty} \int_E 1_{\{V(x)\geq v_\star+\varepsilon\}} e^{-\frac{\beta}{2}(V(x)-v_\star)}\lambda(dx) = 0$, ce qui conclut la démonstration. □

Le lemme suivant contrôle la dépendance en le paramètre β de la mesure de Gibbs π_β lorsque l'essentiel supremum $\bar{v} = \sup\{v \in \mathbb{R} : \lambda(\{x \in E : V(x) \geq v\}) > 0\}$ de V pour la mesure λ est fini. Notons que comme $Z_\beta \geq e^{-\beta\bar{v}}\lambda(E)$, si $\bar{v} < \infty$, alors le fait que $Z_\beta < \infty$ pour $\beta > \underline{\beta}$ entraîne que $\lambda(E) < \infty$ si bien que $B = [0, +\infty[$.

Lemme 1.3.16. *On suppose que $\bar{v} < \infty$. Pour tous $\tilde{\beta}, \beta \geq 0$, on a*

$$d_{\mathrm{TV}}(\pi_{\tilde{\beta}}, \pi_\beta) \leq (\bar{v} - v_\star)|\tilde{\beta} - \beta|.$$

Démonstration : Pour fixer les idées, on suppose que $\tilde{\beta} > \beta$ ce qui entraîne que $\hat{Z}_{\tilde{\beta}} \leq \hat{Z}_\beta$. D'après l'exercice 1.3.5 et (1.6),

$$\begin{aligned}
2d_{\mathrm{TV}}(\pi_{\tilde{\beta}}, \pi_\beta) &= \int_E \left| \frac{e^{-\tilde{\beta}(V(x)-v_\star)}}{\hat{Z}_{\tilde{\beta}}} - \frac{e^{-\beta(V(x)-v_\star)}}{\hat{Z}_\beta} \right| \lambda(dx) \\
&= \int_E \left| e^{-\tilde{\beta}(V(x)-v_\star)} \left(\frac{1}{\hat{Z}_{\tilde{\beta}}} - \frac{1}{\hat{Z}_\beta} \right) + \frac{e^{-\tilde{\beta}(V(x)-v_\star)} - e^{-\beta(V(x)-v_\star)}}{\hat{Z}_\beta} \right| \lambda(dx) \\
&\leq \left(\frac{1}{\hat{Z}_{\tilde{\beta}}} - \frac{1}{\hat{Z}_\beta} \right) \int_E e^{-\tilde{\beta}(V(x)-v_\star)} \lambda(dx) + \frac{1}{\hat{Z}_\beta} \int_E \left(e^{-\beta(V(x)-v_\star)} - e^{-\tilde{\beta}(V(x)-v_\star)} \right) \lambda(dx) \\
&= 2\left(1 - \frac{\hat{Z}_{\tilde{\beta}}}{\hat{Z}_\beta} \right).
\end{aligned}$$

Comme $1 - e^{-y} \leq y$ pour $y \geq 0$, on obtient que

$$\begin{aligned}
\hat{Z}_\beta - \hat{Z}_{\tilde{\beta}} &= \int_E (1 - e^{-(\tilde{\beta}-\beta)(V(x)-v_\star)}) e^{-\beta(V(x)-v_\star)} \lambda(dx) \\
&\leq (\tilde{\beta} - \beta) \int_E (V(x) - v_\star) e^{-\beta(V(x)-v_\star)} \lambda(dx) \leq (\tilde{\beta} - \beta)(\bar{v} - v_\star)\hat{Z}_\beta.
\end{aligned}$$

Cette inégalité se récrit $1 - \frac{\hat{Z}_{\tilde{\beta}}}{\hat{Z}_\beta} \leq (\tilde{\beta} - \beta)(\bar{v} - v_\star)$, ce qui permet de conclure. □

On suppose que la densité de proposition q par rapport à la mesure de référence λ de l'algorithme de Metropolis-Hastings est symétrique : $\boxed{\forall x, y \in E,\ q(x, y) = q(y, x)}$. Pour la probabilité cible π_β, on a $\alpha(x, y) = \min(1, e^{-\beta(V(y)-V(x))}) = e^{-\beta(V(y)-V(x))^+}$ (où, par rapport à (1.1), le remplacement de 1 par $e^{-\beta(V(y)-V(x))}$ pour $q(x, y) = 0$ et $V(y) > V(x)$ est sans conséquence d'après le second point dans la remarque 1.2.1). Notons $Q(x, dy) = q(x, y)\lambda(dy)$ le noyau de proposition et pour tout $\beta \geq 0$

$$\begin{aligned}
&P_\beta(x, dy) = e^{-\beta(V(y)-V(x))^+} q(x, y)\lambda(dy) + R_\beta(x)\delta_x(dy) \\
&\text{où } R_\beta(x) = \int_E (1 - e^{-\beta(V(z)-V(x))^+}) q(x, z)\lambda(dz).
\end{aligned}$$

On a $P_0 = Q$ et pour tout $\beta \in B$, P_β est le noyau de Metropolis-Hastings associé à π_β.

Proposition 1.3.17. *On suppose que $\bar{v} < \infty$ (si bien que $B = [0, +\infty[$) et qu'il existe $m \in \mathbb{N}^*$ tel que Q^m satisfait la condition de Doeblin pour la constante $\alpha > 0$. Alors*

$$\kappa := \sup \left\{ v : \int_{E \times E} 1_{\{(V(y)-V(x))^+ \geq v\}} q(x,y)\lambda(dx)\lambda(dy) > 0 \right\} \leq \bar{v} - v_\star,$$

et pour toute constante $h \in]0, \frac{1}{m\kappa}[$ et toute probabilité μ_0 sur E la suite $\mu_n = \mu_0 P_{\beta_1} P_{\beta_2} \dots P_{\beta_n}$ où $(\beta_n = h \ln(n))_{n \geq 1}$ vérifie $\lim_{n\to\infty} d_{\mathrm{TV}}(\mu_n, \pi_{\beta_n}) = 0$ si bien que $\lim_{n\to\infty} \mu_n(\{x \in E : V(x) \geq v_\star + \varepsilon\}) = 0$ pour tout $\varepsilon > 0$.

Exercice 1.3.18. *En remarquant que pour $\varepsilon > 0$, $(V(y) - V(x))^+ \geq \bar{v} - v_\star + \varepsilon \Rightarrow V(y) \geq \bar{v} + \frac{\varepsilon}{2}$ ou $V(x) \leq v_\star - \frac{\varepsilon}{2}$, vérifier que $\kappa \leq \bar{v} - v_\star$.*

Remarque 1.3.19. *Si on note $(X_n)_n$ une chaîne de Markov inhomogène à valeurs dans E telle que $X_0 \sim \mu_0$ et*

$$\forall n \in \mathbb{N},\ \mathbb{P}(X_{n+1} \in A | (X_0, \dots, X_n)) = P_{\beta_{n+1}}(X_n, A)$$

alors μ_n est la loi de X_n. Sous les hypothèses de la proposition, pour $h \in]0, \frac{1}{m\kappa}[$, $\lim_{n\to\infty} \mathbb{P}(V(X_n) \geq v_\star + \varepsilon) = 0$ pour tout $\varepsilon > 0$, c'est-à-dire que la suite $(V(X_n))_n$ converge en probabilité vers $v_\star$.

Notons que $\mu_1(dy) = \mu_0 P_0(dy) = \mu_0 Q(dy) = \int_{x\in E} q(x,y)\mu_0(dx)\lambda(dy)$, si bien que μ_1 possède la densité $f_1(y) = \int_{x\in E} q(x,y)\mu_0(dx)$ par rapport à la mesure de référence λ. Supposons que pour $n \in \mathbb{N}^*$, μ_n possède la densité f_n par rapport à λ, alors $\mu_{n+1}(dy) = \int_{x\in E} e^{-\beta_{n+1}(V(y)-V(x))^+} q(x,y)\mu_n(dx)\lambda(dy) + R_{\beta_{n+1}}(y) f_n(y)\lambda(dy) = f_{n+1}(y)\lambda(dy)$ pour $f_{n+1}(y) = \int_{x\in E} e^{-\beta_{n+1}(V(y)-V(x))^+} q(x,y)\mu_n(dx) + R_{\beta_{n+1}}(y) f_n(y)$. Ainsi, nous avons montré par récurrence que pour tout $n \in \mathbb{N}^*$, μ_n possède une densité f_n par rapport à la mesure de référence λ. Soit $A = \left\{ x \in E : \int_E 1_{\{(V(y)-V(x))^+ > \kappa\}} q(x,y)\lambda(dy) = 0 \right\}$. Par définition de κ, $\lambda(A^c) = 0$. Modifions désormais la définition de P_β en

$$P_\beta(x, dy) = e^{-\beta((V(y)-V(x))^+ \wedge \kappa)} q(x,y)\lambda(dy) + \int_E (1 - e^{-\beta((V(z)-V(x))^+ \wedge \kappa)}) q(x,z)\lambda(dz)\delta_x(dy).$$

Pour tout $x \in A$, la probabilité $P_\beta(x, .)$ est inchangée. Comme $\lambda(A^c) = 0$, pour tous $\beta > 0$ et $\mu \in \mathcal{P}(E)$ qui possède une densité par rapport à λ, la valeur de μP_β n'est pas affectée par cette modification. Ainsi par récurrence sur $n \in \mathbb{N}^*$, l'égalité $\mu_n = \mu_1 P_{\beta_2} \dots P_{\beta_n}$ reste vraie malgré la modification. Comme $P_{\beta_1} = P_0$ est inchangée, on a $\mu_1 = \mu_0 P_{\beta_1}$ si bien que l'égalité $\mu_n = \mu_0 P_{\beta_1} P_{\beta_2} \dots P_{\beta_n}$ reste également vraie. Comme π_β possède une densité par rapport à λ, la modification ne change rien à la valeur de $\pi_\beta P_\beta$ si bien que π_β reste invariante par P_β. L'intérêt de la modification est d'assurer le résultat suivant.

Lemme 1.3.20. *Si pour $m \in \mathbb{N}^*$, Q^m satisfait la condition de Doeblin pour la constante $\alpha \in]0, 1]$, alors pour $\beta_1, \dots, \beta_m \geq 0$, $P_{\beta_1} P_{\beta_2} \dots P_{\beta_m}$ satisfait la condition de Doeblin pour la constante $\alpha e^{-\kappa \sum_{n=1}^m \beta_n}$.*

Démonstration du lemme 1.3.20 : La définition modifiée de P_β entraîne que pour $\beta \geq 0$, $P_\beta \geq e^{-\beta\kappa} Q$ au sens où pour tout $x \in E$, la mesure $e^{-\beta\kappa} Q(x, dy)$ est majorée par la probabilité $P_\beta(x, dy)$. En itérant, on en déduit que $P_{\beta_1} P_{\beta_2} \dots P_{\beta_m} \geq$

$e^{-\kappa\sum_{n=1}^m \beta_n}Q^m$ si bien que pour tous $x,y\in E$, $P_{\beta_1}P_{\beta_2}\dots P_{\beta_m}(x,.)\wedge P_{\beta_1}P_{\beta_2}\dots P_{\beta_m}(y,.)\geq e^{-\kappa\sum_{n=1}^m \beta_n}\left(Q^m(x,.)\wedge Q^m(y,.)\right)$. □

Démonstration de la proposition 1.3.17 : La majoration de κ découle de l'exercice 1.3.18. Comme β_n tend vers $+\infty$ avec n, pour $\varepsilon>0$ et $A_\varepsilon=\{x\in E : V(x)\geq v_\star+\varepsilon\}$, le lemme 1.3.16 entraîne que $\lim_{n\to\infty}\pi_{\beta_n}(A_\varepsilon)=0$. Comme $|\pi_{\beta_n}(A_\varepsilon)-\mu_n(A_\varepsilon)|\leq d_{\mathrm{TV}}(\pi_{\beta_n},\mu_n)$, il suffit de montrer que $d_{\mathrm{TV}}(\pi_{\beta_n},\mu_n)$ tend vers 0 lorsque $n\to\infty$ pour conclure que $\mu_n(A_\varepsilon)$ aussi.

Pour $n,k\in\mathbb{N}^*$, comme par invariance de π_β par P_β,

$$\forall \ell\in\{1,\dots,k\},\ (\pi_{\beta_{n+\ell}}-\pi_{\beta_{n+\ell-1}})P_{\beta_{n+\ell}}\dots P_{\beta_{n+k}}=\pi_{\beta_{n+\ell}}P_{\beta_{n+\ell+1}}\dots P_{\beta_{n+k}}-\pi_{\beta_{n+\ell-1}}P_{\beta_{n+\ell}}\dots P_{\beta_{n+k}},$$

on a $\pi_{\beta_{n+k}}=\sum_{\ell=1}^k(\pi_{\beta_{n+\ell}}-\pi_{\beta_{n+\ell-1}})P_{\beta_{n+\ell}}\dots P_{\beta_{n+k}}+\pi_{\beta_n}P_{\beta_{n+1}}\dots P_{\beta_{n+k}}$. Comme $\mu_{n+k}=\mu_nP_{\beta_{n+1}}\dots P_{\beta_{n+k}}$, avec la définition de d_{TV}, on en déduit que

$$\begin{aligned}
d_{\mathrm{TV}}(\pi_{\beta_{n+k}},\mu_{n+k})\leq&\sum_{\ell=1}^k d_{\mathrm{TV}}(\pi_{\beta_{n+\ell}}P_{\beta_{n+\ell}}\dots P_{\beta_{n+k}},\pi_{\beta_{n+\ell-1}}P_{\beta_{n+\ell}}\dots P_{\beta_{n+k}})\\
&+d_{\mathrm{TV}}(\pi_{\beta_n}P_{\beta_{n+1}}\dots P_{\beta_{n+k}},\mu_nP_{\beta_{n+1}}\dots P_{\beta_{n+k}})\\
\leq&\sum_{\ell=1}^k d_{\mathrm{TV}}(\pi_{\beta_{n+\ell}},\pi_{\beta_{n+\ell-1}})+d_{\mathrm{TV}}(\mu_n,\pi_{\beta_n}) \qquad (1.7)\\
\leq&(\bar v-v_\star)h\sum_{\ell=1}^k(\ln(n+\ell)-\ln(n+\ell-1))+d_{\mathrm{TV}}(\mu_n,\pi_{\beta_n})\\
=&(\bar v-v_\star)h\ln\left(\frac{n+k}{n}\right)+d_{\mathrm{TV}}(\mu_n,\pi_{\beta_n})
\end{aligned}$$

où on a utilisé (1.4) pour la seconde inégalité puis le lemme 1.3.16 pour la troisième. On pose $\beta_0=0$ si bien que l'inégalité reste vraie pour $n=0$ avec $\ln\left(\frac{n+k}{n}\right)$ remplacé par $\ln(k)$ au second membre. Pour $n=\ell m$ avec $\ell\in\mathbb{N}$ et $k\in\{1,\dots,m-1\}$, cela implique que $d_{\mathrm{TV}}(\pi_{\beta_{\ell m+k}},\mu_{\ell m+k})\leq(\bar v-v_\star)h\ln\left(\frac{\ell m+m-1}{(\ell m)\vee 1}\right)+d_{\mathrm{TV}}(\pi_{\beta_{\ell m}},\mu_{\ell m})$. Comme $\lim_{\ell\to\infty}\ln\left(\frac{\ell m+m-1}{\ell m}\right)=0$, il suffit donc de montrer que $d_{\mathrm{TV}}(\pi_{\beta_{\ell m}},\mu_{\ell m})$ tend vers 0 lorsque $\ell\to\infty$ pour conclure. Pour $k=m$ on peut améliorer l'inégalité (1.7) en remarquant que d'après les lemmes 1.3.20 et 1.3.13,

$$\begin{aligned}
d_{\mathrm{TV}}(\pi_{\beta_n}P_{\beta_{n+1}}\dots P_{\beta_{n+m}},\mu_nP_{\beta_{n+1}}\dots P_{\beta_{n+m}})&\leq(1-\alpha e^{-\kappa\sum_{\ell=1}^m\beta_{n+\ell}})d_{\mathrm{TV}}(\pi_{\beta_n},\mu_n)\\
&\leq(1-\alpha e^{-m\kappa\beta_{n+m}})d_{\mathrm{TV}}(\pi_{\beta_n},\mu_n)
\end{aligned}$$

si bien que

$$d_{\mathrm{TV}}(\pi_{\beta_{n+m}},\mu_{n+m})\leq(\bar v-v_\star)h\ln\left(\frac{n+m}{n\vee 1}\right)+(1-\alpha e^{-m\kappa\beta_{n+m}})d_{\mathrm{TV}}(\mu_n,\pi_{\beta_n}).$$

En posant pour $\ell\in\mathbb{N}$, $z_\ell=d_{\mathrm{TV}}(\pi_{\beta_{\ell m}},\mu_{\ell m})$, $\alpha_{\ell+1}=\alpha e^{-m\kappa\beta_{(\ell+1)m}}=\alpha((\ell+1)m)^{-m\kappa h}$ et $b_{\ell+1}=(\bar v-v_\star)h\ln\left(\frac{(\ell+1)m}{(\ell m)\vee 1}\right)$ on en déduit que

$$z_{\ell+1}\leq(1-\alpha_{\ell+1})z_\ell+b_{\ell+1}.$$

Pour $\ell \to \infty$, on a $\alpha_{\ell+1} \sim \alpha(m\ell)^{-m\kappa h}$ et comme pour $\ell \geq 1$, $b_{\ell+1} = (\bar{v} - v_\star)h \ln\left(1 + \frac{1}{\ell}\right)$, $b_{\ell+1} \sim (\bar{v} - v_\star)h/\ell$. Si $m\kappa h \leq 1$, alors $\sum_{\ell \geq 1} \alpha_\ell = \infty$ et si $m\kappa h < 1$, alors $\frac{b_{\ell+1}}{\alpha_{\ell+1}} \sim \frac{1}{\alpha}(\bar{v} - v_\star)hm^{m\kappa h}\ell^{m\kappa h-1}$ si bien que $\lim_{\ell\to\infty} \frac{b_\ell}{\alpha_\ell} = 0$. Avec le lemme suivant on conclut que si $h < \frac{1}{m\kappa}$ alors $z_\ell = d_{\mathrm{TV}}(\mu_{\ell m}, \pi_{\beta_{\ell m}})$ tend vers 0 lorsque $\ell \to \infty$ et donc $d_{\mathrm{TV}}(\mu_n, \pi_{\beta_n})$ tend également vers 0 lorsque $n \to \infty$.

$\square$

Lemma 1.3.21. *Soit $(z_\ell)_{\ell\in\mathbb{N}}$ une suite de $\mathbb{R}_+$ vérifiant*

$$\forall \ell \in \mathbb{N},\ z_{\ell+1} \leq (1 - \alpha_{\ell+1})z_\ell + b_{\ell+1}. \tag{1.8}$$

avec $(\alpha_\ell)_{\ell\geq 1}$ suite de $]0,1[$ telle que $\sum_{\ell\geq 1} \alpha_\ell = \infty$ et $(b_\ell)_{\ell \geq 1}$ suite de $\mathbb{R}_+$ telle que $\lim_{\ell\to\infty} \frac{b_\ell}{\alpha_\ell} = 0$. Alors $\lim_{\ell\to\infty} z_\ell = 0$.

Démonstration : Posons $A_0 = 1$ et pour $\ell \geq 1$, $A_\ell = \prod_{k=1}^{\ell} \frac{1}{1-\alpha_k}$. Notons que $(A_\ell)_{\ell\in\mathbb{N}}$ est une suite croissante. Comme $-\ln(A_\ell) = \sum_{k=1}^{\ell} \ln(1 - \alpha_k) \leq -\sum_{k=1}^{\ell} \alpha_k$ et $\sum_{k\geq 1} \alpha_k = \infty$, A_ℓ tend vers $+\infty$ avec ℓ. En multipliant (1.8) par $A_{\ell+1}$ et en utilisant que $A_{\ell+1}(1-\alpha_{\ell+1}) = A_\ell$, on obtient que $A_{\ell+1}z_{\ell+1} \leq A_\ell z_\ell + A_{\ell+1}b_{\ell+1}$. Par récurrence, on en déduit que

$$A_\ell z_\ell \leq A_0 z_0 + \sum_{k=1}^{\ell} A_k b_k.$$

Comme $A_0 = 1$ et $\alpha_k A_k = A_k - A_{k-1}$, on obtient que

$$\forall \ell \in \mathbb{N},\ z_\ell \leq \frac{z_0}{A_\ell} + \frac{1}{A_\ell}\sum_{k=1}^{\ell}(A_k - A_{k-1})\frac{b_k}{\alpha_k}.$$

Avec la croissance de la suite $(A_k)_{k\in\mathbb{N}}$ on en déduit que pour $\ell_0 \in \mathbb{N}^*$,

$$\forall \ell \geq \ell_0,\ z_\ell \leq \frac{1}{A_\ell}\left(z_0 + \sum_{k=1}^{\ell_0}(A_k - A_{k-1})\frac{b_k}{\alpha_k}\right) + \max_{k>\ell_0} \frac{b_k}{\alpha_k}.$$

Comme le second terme est arbitrairement petit pour ℓ_0 assez grand tandis qu'à ℓ_0 fixé le premier terme tend vers 0 lorsque $\ell \to \infty$, on conclut que $\lim_{\ell\to\infty} z_\ell = 0$. $\square$

En pratique, on travaille avec un nombre fini N d'itérations de l'algorithme et on utilise des suites $(\beta_n)_{1\leq n\leq N}$ qui croissent beaucoup plus vite que la suite $h\ln(n)$ (typiquement des puissances positives de n).

Notons que l'hypothèse que Q satisfait la condition de Doeblin est assez restrictive sur le couple (E, λ).

Lemme 1.3.22. *On suppose que $E = \mathbb{R}^d$ muni de la mesure de Lebesgue λ. Soit $Q(x, dz) = q(x, z)dz$ pour $q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ mesurable, symétrique et telle que $\int_{\mathbb{R}^d} q(x, z)dz = 1$ pour tout $x \in \mathbb{R}^d$. Alors $\sup_{m\in\mathbb{N}^*} \inf_{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d}(Q^m(x, .)\wedge Q^m(y, .))(\mathbb{R}^d) = 0$ i.e. aucune itérée Q^m de Q ne satisfait la condition de Doeblin.*

Remarque 1.3.23. — *Notons que sous les hypothèses du lemme, pour tout $\beta \geq 0$, P_β ne satisfait pas non plus la condition de Doeblin, même si, comme $B \neq \emptyset$ et $\bar{v} < \infty$ impliquent $\lambda(E) < \infty$, ce n'est pas restrictif en vue de la proposition 1.3.17. En effet, pour $x \neq y$, $P_\beta(x,.)$ et $P_\beta(y,.)$ admettent respectivement les densités $R_\beta(x)1_{\{z=x\}} + 1_{\mathbb{R}^d\setminus\{x,y\}}(z)e^{-\beta(V(z)-V(x))^+}q(x,z)$ et $R_\beta(y)1_{\{z=y\}} + 1_{\mathbb{R}^d\setminus\{x,y\}}(z)e^{-\beta(V(z)-V(y))^+}q(y,z)$ par rapport à la mesure de référence $\lambda(dz) + \delta_x(dz) + \delta_y(dz)$. Et, d'après l'exercice 1.3.5,*

$$\begin{aligned}(P_\beta(x,.) \wedge P_\beta(y,.))(\mathbb{R}^d) &= \int_{\mathbb{R}^d} (e^{-\beta(V(z)-V(x))^+}q(x,z)) \wedge (e^{-\beta(V(z)-V(y))^+}q(y,z))dz \\ &\leq \int_{\mathbb{R}^d} q(x,z) \wedge q(y,z)dz = (Q(x,.) \wedge Q(y,.))(\mathbb{R}^d).\end{aligned}$$

— *Toujours sous les hypothèses du lemme, pour le noyau P de l'algorithme de Metropolis-Hastings sur $\mathbb{R}^d$ défini par (1.2), un raisonnement analogue entraîne que pour $y \neq x$,*

$$\begin{aligned}(P(x,.) \wedge P(y,.))(\mathbb{R}^d) &= \int_{\mathbb{R}^d} (\alpha(x,z)q(x,z)) \wedge (\alpha(y,z)q(y,z))dz \\ &\leq \int_{\mathbb{R}^d} q(x,z) \wedge q(y,z)dz = (Q(x,.) \wedge Q(y,.))(\mathbb{R}^d)\end{aligned}$$

si bien que la condition de Doeblin n'est pas satisfaite par le noyau P.

— *Si, en revanche, E est un compact de $\mathbb{R}^d$ muni de la mesure de Lebesgue λ, pour la fonction symétrique $q(x,y) = \frac{1}{\lambda(E)}$, le noyau Q satisfait la condition de Doeblin.*

Démonstration : Notons que pour $m \geq 2$, $Q^m(x,dy) = q_m(x,y)dy$ où, en utilisant la symétrie de q pour la seconde égalité,

$$\begin{aligned}q_m(x,y) &= \int_{\mathbb{R}^{(m-1)d}} q(x,z_1)q(z_1,z_2)\ldots q(z_{m-2},z_{m-1})q(z_{m-1},y)dz_1\ldots dz_{m-1} \\ &= \int_{\mathbb{R}^{(m-1)d}} q(y,z_{m-1})q(z_{m-1},z_{m-2})\ldots q(z_2,z_1)q(z_1,x)dz_{m-1}\ldots dz_1 = q_m(y,x).\end{aligned}$$

Le noyau Q^m possèdant lui aussi une densité symétrique, il suffit de montrer que $\inf_{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d}(Q(x,.) \wedge Q(y,.))(\mathbb{R}^d) = 0$.

Supposons que $\forall x,y \in \mathbb{R}^d$, $(Q(x,.) \wedge Q(y,.))(\mathbb{R}^d) \geq \alpha > 0$ et obtenons une contradiction. D'après l'exercice 1.3.5, on a $(Q(x,.) \wedge Q(y,.))(\mathbb{R}^d) = \int_{\mathbb{R}^d} q(x,z) \wedge q(y,z)dz$. Fixons $x \in \mathbb{R}^d$. Comme, par convergence dominée, $\lim_{M\to\infty}\int 1_{\{|z-x|\geq M\}}q(x,z)dz = 0$, il existe $M_x \in]0,+\infty[$ tel que $\int 1_{\{|z-x|\geq M_x\}}q(x,z)dz \leq \frac{\alpha}{2}$. Alors

$$\begin{aligned}\forall y \in \mathbb{R}^d,\ \int_{\mathbb{R}^d} 1_{\{|z-x|<M_x\}}q(y,z)dz &\geq \int_{\mathbb{R}^d} 1_{\{|z-x|<M_x\}}q(x,z) \wedge q(y,z)dz \\ &\geq \int_{\mathbb{R}^d} q(x,z) \wedge q(y,z)dz - \int_{\mathbb{R}^d} 1_{\{|z-x|\geq M_x\}}q(x,z)dz \geq \frac{\alpha}{2}.\end{aligned}$$

Donc $\int_{\mathbb{R}^d\times\mathbb{R}^d} 1_{\{|z-x|<M_x\}}q(y,z)dzdy = \infty$ alors que, par la symétrie de q et le théorème de Fubini, cette intégrale est aussi égale à $\int_{\mathbb{R}^d} 1_{\{|z-x|<M_x\}}\int_{\mathbb{R}^d} q(z,y)dydz = \int_{\mathbb{R}^d} 1_{\{|z-x|<M_x\}}dz < \infty$. □

Cela a motivé l'étude de conditions plus locales qui font l'objet du paragraphe suivant.

1.3.4 Ergodicité géométrique sous condition de dérive

Ce chapître est inspiré de [35]. Nous allons travailler sous la condition de dérive suivante :

(D1) $\exists V : E \to \mathbb{R}_+$ mesurable, $\exists K \in \mathbb{R}_+,\ \exists \gamma \in]0,1[,\ \forall x \in E,\ PV(x) \leq \gamma V(x) + K$

(D2) $\exists R > \dfrac{2K}{1-\gamma},\ \exists \alpha \in]0,1],\ \forall x, y$ t.q. $V(x) + V(y) \leq R,\ P(x,.) \wedge P(y,.)(E) \geq \alpha.$

Notons que la condition de Doeblin implique la condition de dérive pour le choix $V \equiv 0$.

Notons également que (D2) est une conséquence de

$$\exists R > \frac{2K}{1-\gamma},\ \exists \nu \in \mathcal{P}(E),\ \exists \alpha \in]0,1],\ \forall A \in \mathcal{E},\ \inf_{x:V(x)\leq R} P(x,A) \geq \alpha\nu(A)$$

où l'ensemble $\{x \in E : V(x) \leq R\}$ est appelé “small set” dans la littérature (attention “petite set” désigne une notion reliée mais légèrement plus faible).

La fonction V qui porte le nom de fonction de Lyapunov tend en général vers l'infini à l'infini lorsque $E = \mathbb{R}^d$ (dans l'application à l'algorithme de Metropolis-Hastings par marche aléatoire du paragraphe 1.3.5, $V(x) = \eta^{-1/2}(x)$ où η est la densité cible). La condition (D1) signifie alors qu'en moyenne le noyau P rapproche de l'origine et donc du compact où une condition de Doeblin locale est satisfaite d'après (D2). L'inégalité $R > 2K/(1-\gamma)$ signifie que ce compact (d'autant plus grand que R l'est) doit être suffisamment grand par rapport à la force de rappel en moyenne vers l'origine (d'autant plus petite que K et γ sont grands). Plus généralement la condition de Doeblin est satisfaite localement sur l'ensemble de niveau $\{x : V(x) \leq \frac{R}{2}\}$ tandis que si $V(x) \geq \frac{R}{2}$ alors $K < (1-\gamma)V(x)$ si bien que $PV(x) \leq \gamma V(x) + K < V(x)$ et en moyenne V diminue par application du noyau P i.e. le noyau P rapproche de cet ensemble de niveau.

Définition 1.3.24. — *Pour $\beta \geq 0$, on note $\|\|_\beta$ la norme sur l'espace vectoriel des fonctions $\varphi : E \to \mathbb{R}$ mesurables et telles que $\sup_{x\in E} \frac{|\varphi(x)|}{1+\beta V(x)} < \infty$ définie par $\|\varphi\|_\beta = \sup_{x\in E} \frac{|\varphi(x)|}{1+\beta V(x)}$.*

- *Pour $\beta > 0$, on lui associe la distance d_β sur $\mathcal{P}_V(E) := \{\mu \in \mathcal{P}(E) : \mu(V) < \infty\}$ définie par $d_\beta(\mu,\sigma) = \sup_{\varphi:\|\varphi\|_\beta\leq 1} |\mu(\varphi) - \sigma(\varphi)|$. Pour $\mu, \sigma \in \mathcal{P}(E)$, on pose toujours $d_\beta(\mu,\sigma) = \sup_{\varphi:\|\varphi\|_\beta\leq 1} |\mu(\varphi) - \sigma(\varphi)|$ où rien ne garantit que le supremum soit fini lorsque V n'est pas bornée.*
- *Pour $\beta = 0$, on lui associe la distance d_0 sur $\mathcal{P}(E)$ définie par $d_0(\mu,\sigma) = \sup_{\varphi:|\varphi|\leq 1} |\mu(\varphi) - \sigma(\varphi)|$.*

Exercice 1.3.25. *Montrer que pour $\beta > 0$, d_β est une distance sur $\mathcal{P}_V(E)$.*

Notons que $\|\|_0$ est la norme du supremum sur l'espace des fonctions mesurables bornées sur E et que, d'après l'exercice 1.3.4, $d_0(\mu,\sigma) = 2d_{\mathrm{TV}}(\mu,\sigma)$. Pour $\beta > 0$, $\|\|_\beta$ est une norme sur l'espace plus grand lorsque $\sup_{x\in E} V(x) = \infty$

$$\mathcal{V} = \left\{\varphi : E \to \mathbb{R} \text{ mesurable t.q. } \sup_{x\in E} \frac{|\varphi(x)|}{1+V(x)} < \infty\right\}.$$

Pour $0 \leq \beta' \leq \beta$, comme $\|\varphi\|_\beta \leq \|\varphi\|_{\beta'}$, $d_{\beta'} \leq d_\beta$.

En reprenant la preuve de la proposition 1.3.9, et en remarquant que

$$\begin{aligned} d_\beta(\mu_n, \mu_m) &= \sup_{\varphi:|\varphi|\leq 1+\beta V} \left| \int_E \varphi(x)(p_n(x) - p_m(x))\mu(dx) \right| \\ &= \int_E (1+\beta V(x))|p_n(x) - p_m(x)|\mu(dx) = \|p_n - p_m\|_{L^1((1+\beta V)\mu)}, \end{aligned}$$

où $(1+\beta V)\mu$ désigne la mesure de densité $(1+\beta V(x))$ par rapport à $\mu(dx)$, on vérifie que

Proposition 1.3.26. *Pour $\beta > 0$, l'espace $\mathcal{P}_V(E)$ muni de d_β est complet.*

Théorème 1.3.27. *On suppose que le noyau markovien P satisfait (D1) et (D2). Alors il admet une unique probabilité invariante π. En outre, $\pi(V) \leq \frac{K}{1-\gamma}$ et*

$$\forall \beta > 0,\ \forall \mu \in \mathcal{P}(E),\ \forall n \in \mathbb{N},\ d_\beta(\mu P^n, \pi) \leq (\chi(\alpha, \beta, \gamma, K, R))^n d_\beta(\mu, \pi). \tag{1.9}$$

où $\chi(\alpha, \beta, \gamma, K, R) = (1 - \alpha + \beta K) \vee \frac{2+\beta\gamma R+2\beta K}{2+\beta R} \in]0, 1[$ si $\beta \in]0, \frac{\alpha}{K}[$.

Remarque 1.3.28. — *On a*

$$\begin{aligned} (2+\beta R)^2 \frac{d}{d\beta} \frac{2+\beta\gamma R+2\beta K}{2+\beta R} &= (\gamma R + 2K)(2+\beta R) - R(2+\beta\gamma R+2\beta K) \\ &= 2(2K - (1-\gamma)R) < 0. \end{aligned}$$

Ainsi $\mathbb{R}_+ \ni \beta \mapsto g(\beta) = (1 - \alpha + \beta K) - \frac{2+\beta\gamma R+2\beta K}{2+\beta R}$ est une fonction strictement croissante telle que $g(0) = -\alpha$ et $\lim_{\beta\to\infty} g(\beta) = +\infty$. Donc $\exists! \beta_\star > 0$ tel que $g(\beta_\star) = 0$. Pour $\beta \in [0, \beta_\star[$, $\chi(\alpha, \beta, \gamma, K, R) = \frac{2+\beta\gamma R+2\beta K}{2+\beta R} > \frac{2+\beta_\star\gamma R+2\beta_\star K}{2+\beta_\star R} = 1 - \alpha + \beta_\star K$. Pour $\beta \in]\beta_\star, +\infty[$, $\chi(\alpha, \beta, \gamma, K, R) = 1 - \alpha + \beta K > 1 - \alpha + \beta_\star K$. On conclut que $\inf_{\beta\geq 0} \chi(\alpha, \beta, \gamma, K, R) = \chi(\alpha, \beta_\star, \gamma, K, R)$.

— *Si pour $m \geq 2$, P^m satisfait (D1) et (D2), en raisonnant comme dans la Proposition 1.3.14, on vérifie que P admet une unique probabilité invariante π. Si, en outre, le noyau P lui-même vérifie la condition (D1), alors pour $\varphi \in \mathcal{V}$, la définition de $\|\|_\beta$ entraîne que*

$$\forall x \in E,\ |P\varphi(x)| \leq P|\varphi|(x) \leq \|\varphi\|_\beta P(1+\beta V)(x) \leq \|\varphi\|_\beta (1+\gamma\beta V(x)+\beta K), \tag{1.10}$$

si bien que $\|P\varphi\|_\beta \leq (1+\beta K)\|\varphi\|_\beta$. On en déduit que $d_\beta(\mu P, \sigma P) \leq (1+\beta K) d_\beta(\mu, \sigma)$ et avec l'estimation du théorème pour le noyau P^m que

$$\forall \mu \in \mathcal{P}(E),\ \forall n \in \mathbb{N},\ d_\beta(\mu P^n, \pi) \leq (1+\beta K)^{m-1} (\chi(\alpha, \beta, \gamma, K, R))^{\lfloor n/m \rfloor} d_\beta(\mu, \pi).$$

La preuve du théorème repose sur le résultat suivant

Lemme 1.3.29. *Sous (D1) et (D2) pour toute fonction $\varphi \in \mathcal{V}$,*

$$\forall \beta > 0, \forall x, y \in E,\ |P\varphi(x) - P\varphi(y)| \leq \chi(\alpha, \beta, \gamma, K, R)(2 + \beta(V(x) + V(y)))\|\varphi\|_\beta \tag{1.11}$$

$$\forall \mu, \sigma \in \mathcal{P}(E),\ d_\beta(\mu P, \sigma P) \leq \chi(\alpha, \beta, \gamma, K, R) d_\beta(\mu, \sigma). \tag{1.12}$$

Démontrons ce lemme avant de revenir à la preuve du théorème.

Démonstration : Soit $\varphi \in \mathcal{V}$ et $x, y \in E$. On distingue deux cas.

$\boxed{\text{Si } V(x) + V(y) \geq R}$, en utilisant (1.10) à la seconde inégalité, on obtient

$$|P\varphi(x) - P\varphi(y)| \leq |P\varphi(x)| + |P\varphi(y)| \leq (2 + \beta\gamma(V(x) + V(y)) + 2\beta K)\|\varphi\|_\beta.$$

La décroissance de $\mathbb{R}_+ \ni r \to f(r) := \frac{2+\beta\gamma r+2\beta K}{2+\beta r} = \gamma + \frac{2(1-\gamma+\beta K)}{2+\beta r}$ entraîne que

$$\begin{aligned}
2 + \beta\gamma(V(x) + V(y)) + 2\beta K &= \frac{2 + \beta\gamma(V(x) + V(y)) + 2\beta K}{2 + \beta(V(x) + V(y))}(2 + \beta(V(x) + V(y))) \\
&\leq \frac{2 + \beta\gamma R + 2\beta K}{2 + \beta R}(2 + \beta(V(x) + V(y))).
\end{aligned}$$

On conclut donc que

$$|P\varphi(x) - P\varphi(y)| \leq \frac{2 + \beta\gamma R + 2\beta K}{2 + \beta R}(2 + \beta(V(x) + V(y)))\|\varphi\|_\beta.$$

$\boxed{\text{Si } V(x) + V(y) \leq R}$, alors en utilisant (D2) à la quatrième inégalité puis (D1) à la cinquième, on obtient

$$\begin{aligned}
&|P\varphi(x) - P\varphi(y)| \\
&\leq \left|\int_E \varphi(z)(P(x,dz) - P(x,.) \wedge P(y,.)(dz))\right| + \left|\int_E \varphi(z)(P(y,dz) - P(x,.) \wedge P(y,.)(dz))\right| \\
&\leq \|\varphi\|_\beta\Big(\int_E (1 + \beta V(z))(P(x,dz) - P(x,.) \wedge P(y,.)(dz)) \\
&\qquad + \int_E (1 + \beta V(z))(P(y,dz) - P(x,.) \wedge P(y,.)(dz))\Big) \\
&\leq \|\varphi\|_\beta\Big(P(x,E) - P(x,.) \wedge P(y,.)(E) + \beta PV(x) + P(y,E) - P(x,.) \wedge P(y,.)(E) + \beta PV(y)\Big) \\
&\leq \|\varphi\|_\beta(2(1-\alpha) + \beta PV(x) + \beta PV(y)) \leq \|\varphi\|_\beta(2(1-\alpha) + \gamma\beta(V(x) + V(y)) + 2\beta K) \\
&\leq ((1 - \alpha + \beta K) \vee \gamma)(2 + \beta(V(x) + V(y)))\|\varphi\|_\beta.
\end{aligned}$$

Comme $\gamma = \lim_{r\to+\infty} f(r) < f(R) = \frac{2+\beta\gamma R+2\beta K}{2+\beta R}$, on a démontré (1.11).

Supposons maintenant $\|\varphi\|_\beta \leq 1$, notons χ à la place de $\chi(\alpha, \beta, \gamma, K, R)$ et posons $c_\varphi = \inf_{y\in E}(\chi(1 + \beta V(y)) - P\varphi(y))$. Pour $x \in E$, on a d'une part, d'après (1.11),

$$P\varphi(x)+c_\varphi = \inf_{y\in E}(P\varphi(x)-P\varphi(y)+\chi(2+\beta(V(x)+V(y)))-\chi(1+\beta V(x)) \geq -\chi(1+\beta V(x)),$$

ce qui assure que $c_\varphi > -\infty$. D'autre part, d'après la définition de c_φ,

$$P\varphi(x) + c_\varphi \leq P\varphi(x) + \chi(1 + \beta V(x)) - P\varphi(x) = \chi(1 + \beta V(x)).$$

Ainsi $\|P\varphi + c_\varphi\|_\beta \leq \chi$. On conclut que

$$\begin{aligned}
d_\beta(\mu P, \sigma P) &= \sup_{\varphi:\|\varphi\|_\beta\leq 1} |\mu P(\varphi) - \sigma P(\varphi)| = \sup_{\varphi:\|\varphi\|_\beta\leq 1} |\mu(P\varphi + c_\varphi) - \sigma(P\varphi + c_\varphi)| \\
&\leq \sup_{\psi:\|\psi\|_\beta\leq\chi} |\mu(\psi) - \sigma(\psi)| = \chi d_\beta(\mu, \sigma).
\end{aligned}$$

□

Démontrons maintenant le théorème 1.3.27.
Démonstration : Soit $\beta \in]0, \frac{\alpha}{K}[$. Pour alléger les notations, nous noterons χ à la place de $\chi(\alpha, \beta, \gamma, K, R)$. Montrons que $\chi \in]0, 1[$. La condition $\beta < \alpha/K$ assure que $1-\alpha+\beta K < 1$. Par ailleurs, les conditions $\gamma \in]0, 1[$, $\beta > 0$ et $R > \frac{2K}{1-\gamma}$ assurent que $\beta R > \beta\gamma R + 2\beta K$ si bien que $\frac{2+\beta\gamma R+2\beta K}{2+\beta R} < 1$. Ainsi, sous (D1) et (D2), $\chi \in]0, 1[$.

D'après (D1), pour $\mu \in \mathcal{P}_V(E)$,

$$\mu P(V) = \mu(PV) \leq \mu(\gamma V + K) = \gamma\mu(V) + K \tag{1.13}$$

si bien que $\mu P \in \mathcal{P}_V(E)$. D'après (1.12), l'application $\mathcal{P}_V(E) \ni \mu \mapsto \mu P \in \mathcal{P}_V(E)$ est contractante pour d_β distance qui rend l'espace $\mathcal{P}_V(E)$ complet d'après la proposition 1.3.26. D'après le théorème de point fixe de Picard, elle admet un unique point fixe π qui est l'unique élément de $\mathcal{P}_V(E)$ invariant par P. Pour le choix $\mu = \pi$, (1.13) s'écrit $\pi(V) \leq \gamma\pi(V) + K$ si bien que $\pi(V) \leq \frac{K}{1-\gamma}$.

L'inégalité $d_\beta(\mu P^n, \pi) \leq \chi^n d_\beta(\mu, \pi)$ s'obtient en itérant (1.12) pour le choix $\sigma = \pi$. Pour $x \in E$, comme $d_\beta(\delta_x, \pi) \leq \delta_x(1 + \beta V) + \pi(1 + \beta V) \leq 2 + \beta V(x) + \frac{\beta K}{1-\gamma} < \infty$, on en déduit que $\lim_{n\to\infty} d_\beta(P^n(x, .), \pi) = 0$.

Soit $\tilde{\pi}$ une autre probabilité invariante. On a $d_\beta(\pi, \tilde{\pi}) = d_\beta(\pi P, \tilde{\pi} P) \leq \chi d_\beta(\pi, \tilde{\pi})$ d'après (1.12). On en déduit que $d_\beta(\pi, \tilde{\pi}) = 0$ dès lors que $d_\beta(\pi, \tilde{\pi}) < \infty$, condition assurée par $\tilde{\pi} \in \mathcal{P}_V(E)$. Sans cette condition, il faut trouver un autre argument. Comme $d_\beta \geq d_0 = 2d_{\mathrm{TV}}$, pour tout $x \in E$, $P^n(x, .)$ converge en variation totale vers π et pour $A \in \mathcal{E}$, la suite $(P^n(x, A))_{n\geq 1}$ à valeurs dans $[0, 1]$ converge vers $\pi(A)$. Par convergence dominée, on en déduit que $\tilde{\pi} P^n(A) = \int_E P^n(x, A)\tilde{\pi}(dx)$ converge vers $\int_E \pi(A)\tilde{\pi}(dx) = \pi(A)$ lorsque $n \to \infty$. Comme par invariance de $\tilde{\pi}$, $\tilde{\pi}(A) = \tilde{\pi} P^n(A)$, on conclut que $\tilde{\pi} = \pi$.

□

1.3.5 Application à l'algorithme de Metropolis-Hastings

On se place dans le cas où

— $E = \mathbb{R}^d$ muni de sa tribu borélienne $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$,

— $\lambda(dx) = dx$ est la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$,

— $\pi(dx) = \frac{\eta(x)dx}{\int_{\mathbb{R}^d} \eta(y)dy}$ où η est une fonction mesurable de $\mathbb{R}^d$ dans $\mathbb{R}_+$ telle que $\int_{\mathbb{R}^d} \eta(y)dy \in]0, \infty[$.

On note $q(x, y)$ la densité du noyau de proposition par rapport à la mesure de Lebesgue. Notons que le noyau P s'écrit alors

$$P(x, dy) = \alpha(x, y)q(x, y)dy + \left(\int_{\mathbb{R}^d} (1 - \alpha(x, z))q(x, z)dz\right) \delta_x(dy)$$
$$\text{avec } \alpha(x, y) = 1_{\{\eta(x)q(x,y)>0\}} \min\left(\frac{\eta(y)q(y, x)}{\eta(x)q(x, y)}, 1\right) + 1_{\{\eta(x)q(x,y)=0\}}.$$

Le résultat principal de ce paragraphe, énoncé dans la proposition 1.3.35, est que si η est C^1 strictement positive et telle que $\lim_{|x|\to\infty} \frac{x}{|x|} \cdot \nabla \ln \eta(x) = -\infty$, $\limsup_{|x|\to\infty} \frac{x}{|x|} \cdot \frac{\nabla\eta(x)}{|\nabla\eta(x)|} < 0$

et $q(x,y)=\psi(y-x)$ pour une densité ψ paire et telle que $\forall M>0$, $\inf_{|x|\leq M}\psi(x)>0$, alors l'algorithme de Metropolis-Hastings par marche aléatoire, noté MHMA dans la suite, est géométriquement ergodique puisqu'il vérifie (D1) et (D2) pour $V(x)=\eta^{-1/2}(x)$. Notons que, d'après la remarque 1.3.23, pour cet algorithme, la condition de Doeblin n'est jamais satisfaite sur $\mathbb{R}^d$. Comme $q(x,y)=\psi(y-x)$ on peut même simplifier la preuve de ce résultat en remarquant que pour $x\neq y\in\mathbb{R}^d$, pour tout $z\in\mathbb{R}^d$, $|x-y|\leq|x-z|+|z-y|$ si bien que $|z-x|\vee|z-y|\geq\frac{|x-y|}{2}$ et $(\alpha(x,z)\psi(z-x))\wedge(\alpha(y,z)\psi(z-y))\leq 1_{\{|z-x|\geq\frac{|x-y|}{2}\}}\psi(z-x)+1_{\{|z-y|\geq\frac{|x-y|}{2}\}}\psi(z-y)$. Ainsi,

$$P(x,.)\wedge P(y,.)(\mathbb{R}^d)=\int_{\mathbb{R}^d}(\alpha(x,z)\psi(z-x))\wedge(\alpha(y,z)\psi(z-y))dz\leq 2\int_{|w|\geq\frac{|x-y|}{2}}\psi(w)dw$$

où, par convergence dominée, le membre de droite tend vers 0 lorsque $|x-y|\to\infty$.

Le lemme suivant donne une condition suffisante pour que la condition de Doeblin soit satisfaite lorsque le point de départ x est restreint à un compact.

Lemme 1.3.30. *Si $\forall M\in]0,+\infty[$, $\underline{q}_M:=\inf_{|x|\vee|y|\leq M}q(x,y)>0$ et $\bar{\eta}_M:=\sup_{|x|\leq M}\eta(x)<\infty$, alors pour tout M assez grand pour que $\int_{|z|\leq M}\eta(z)dz>0$,*

$$\forall|x|\leq M,\ \forall A\in\mathcal{B}(\mathbb{R}^d),\ P(x,A)\geq\alpha_M\nu_M(A),$$

où $\alpha_M:=\frac{\underline{q}_M}{\bar{\eta}_M}\int_{|z|\leq M}\eta(z)dz>0$ et $\nu_M(dy):=\frac{1_{\{|y|\leq M\}}\eta(y)dy}{\int_{|z|\leq M}\eta(z)dz}$ est une probabilité.

Remarque 1.3.31. *Les hypothèses sont satisfaites si η est continue et q est strictement positive et continue.*

Démonstration : En ne tenant pas compte de la contribution liée à la possibilité que l'algorithme partant de x reste en x si la proposition y n'est pas acceptée, on obtient que pour $|x|\leq M$ et $A\in\mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}P(x,A)&\geq\int_A 1_{\{\eta(y)>0\}}\left(\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\wedge 1\right)q(x,y)dy=\int_A 1_{\{\eta(y)>0\}}\left(\frac{q(y,x)}{\eta(x)}\wedge\frac{q(x,y)}{\eta(y)}\right)\eta(y)dy\\&\geq\int_A 1_{\{\eta(y)>0\}}\left(\frac{q(y,x)}{\eta(x)}\wedge\frac{q(x,y)}{\eta(y)}\right)1_{\{|y|\leq M\}}\eta(y)dy\geq\frac{\underline{q}_M}{\bar{\eta}_M}\int_A 1_{\{y\leq M\}}\eta(y)dy.\end{aligned}$$

Sous nos hypothèses, q ne s'annule pas et on utilise les conventions $\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}=\infty=\frac{q(y,x)}{\eta(x)}$ si $\eta(x)=0$. □

Il est plus difficile de vérifier (D1) et nous nous contenterons d'énoncer des conditions obtenues par Jarner et Hansen [39] pour que (D1) soit satisfaite par $V(x)=\eta^{-1/2}(x)$ dans le cas particulier de l'algorithme de Metropolis-Hastings par marche aléatoire où $q(x,y)=\psi(y-x)$ pour une densité de probabilité ψ paire sur $\mathbb{R}^d$.

Vérifions tout d'abord que pour le choix $V(x)=\eta^{-1/2}(x)$, $x\mapsto\frac{PV(x)}{V(x)}$ est borné. Ce résultat est vrai pour un espace d'état E quelconque, dès lors que la densité de proposition est symétrique.

Lemme 1.3.32. *Tout algorithme de Metropolis-Hastings avec une densité de proposition q symétrique ($\forall x,y\in E,\ q(x,y)=q(y,x)$) est tel que*

$$\forall x\in E\ t.q.\ \eta(x)>0,\ \frac{P\eta^{-1/2}(x)}{\eta^{-1/2}(x)}\leq\frac{5}{4}.$$

On en déduit que si la fonction $\eta^{-1/2}$ est localement bornée, il en va de même pour la fonction $P\eta^{-1/2}$.
Démonstration : Soit $x \in E$ t.q. $\eta(x) > 0$. Rappelons que partant de x, une proposition y est acceptée avec probabilité $\alpha(x,y) = \min(1, \frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}) = \min(1, \frac{\eta(y)}{\eta(x)})$. Notons $\mathcal{A}_x = \{y \in E : \eta(y) \geq \eta(x)\}$ l'ensemble des propositions qui sont acceptées avec probabilité 1 et $\mathcal{R}_x = \{y \in E : \eta(y) < \eta(x)\}$ son complémentaire. La proposition $y \in \mathcal{R}_x$ est acceptée avec probabilité $\frac{\eta(y)}{\eta(x)}$ et sinon, l'algorithme reste au point x. Donc

$$\begin{aligned}\frac{P\eta^{-1/2}(x)}{\eta^{-1/2}(x)} &= \int_{\mathcal{A}_x} \frac{\eta^{-1/2}(y)}{\eta^{-1/2}(x)} q(x,y)\lambda(dy) + \int_{\mathcal{R}_x} \frac{\eta^{-1/2}(y)}{\eta^{-1/2}(x)} \frac{\eta(y)}{\eta(x)} q(x,y)\lambda(dy) \\ &\quad + \int_{\mathcal{R}_x} \frac{\eta^{-1/2}(x)}{\eta^{-1/2}(x)} \left(1 - \frac{\eta(y)}{\eta(x)}\right) q(x,y)\lambda(dy) \\ &= \int_{\mathcal{A}_x} \frac{\eta^{1/2}(x)}{\eta^{1/2}(y)} q(x,y)\lambda(dy) + \int_{\mathcal{R}_x} \left(\frac{\eta^{1/2}(y)}{\eta^{1/2}(x)} + 1 - \frac{\eta(y)}{\eta(x)}\right) q(x,y)\lambda(dy) \qquad (1.14)\end{aligned}$$

Sur $\mathcal{A}_x$, $\frac{\eta^{1/2}(x)}{\eta^{1/2}(y)} \leq 1$ et sur $\mathcal{R}_x$, $\frac{\eta(y)}{\eta(x)} < 1$. Comme $\sup_{u\in[0,1]}(\sqrt{u} + 1 - u) = \frac{5}{4}$, on en déduit que

$$\frac{P\eta^{-1/2}(x)}{\eta^{-1/2}(x)} \leq \int_{\mathcal{A}_x} q(x,y)\lambda(dy) + \frac{5}{4}\int_{\mathcal{R}_x} q(x,y)\lambda(dy) \leq \frac{5}{4}\int_{\mathbb{R}^d} q(x,y)\lambda(dy) = \frac{5}{4}.$$

□

Le résultat de Jarner et Hansen porte sur la classe suivante de lois cibles.

Définition 1.3.33. *La loi π est dite sous-exponentielle si la fonction η est C^1, strictement positive et vérifie $\lim_{|x|\to\infty} \frac{x}{|x|} \cdot \nabla \ln \eta(x) = -\infty$.*

Si π est sous-exponentielle, pour $\beta > 0$, en choisissant M_β t.q. $\sup_{|x|\geq M_\beta} \frac{x}{|x|}.\nabla \ln \eta(x) \leq -\beta$, on obtient que pour $|y| \geq M_\beta$,

$$\ln \eta(y) - \ln \eta\left(\frac{M_\beta y}{|y|}\right) = \int_{M_\beta}^{|y|} \frac{y}{|y|}.\nabla \ln \eta\left(\frac{uy}{|y|}\right) du \leq -\beta(|y| - M_\beta).$$

Donc

$$\forall |y| \geq M_\beta,\ \eta(y) \leq \eta\left(\frac{M_\beta y}{|y|}\right) e^{-\beta(|y|-M_\beta)} \leq e^{\beta M_\beta} \sup_{|x|\leq M_\beta} \eta(x)e^{-\beta|y|}. \qquad (1.15)$$

Ainsi la densité $\frac{\eta(y)}{\int_{\mathbb{R}^d}\eta(x)dx}$ de π tend vers 0 plus vite que n'importe quelle exponentielle de $|y|$ lorsque $|y| \to \infty$, ce qui justifie la terminologie sous-exponentielle.

Exercice 1.3.34. *Vérifier que si les fonctions η_1 et η_2 sont proportionnelles aux densités de deux lois sous-exponentielles, alors les lois de densités proportionnelles aux fonctions $\eta_1\eta_2$ et $a_1\eta_1 + a_2\eta_2$ où $a_1, a_2 > 0$ sont sous-exponentielles.*

Dans la suite, nous noterons $\mathcal{S}_{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$ la sphère unité de $\mathbb{R}^d$ et ω_{d-1} la mesure de surface sur cette sphère.

Proposition 1.3.35. *Soit π sous-exponentielle et telle que $\limsup_{|x|\to\infty} \frac{x}{|x|} \cdot \frac{\nabla\eta(x)}{|\nabla\eta(x)|} < 0$.*

(i) *Tout algorithme MHMA avec la densité ψ paire et telle que* $\inf_{\zeta\in\mathcal{S}_{d-1}}\int_0^\infty \psi(r\zeta)r^{d-1}dr > 0$ *vérifie*

$$\exists K \in \mathbb{R}_+,\ \exists\gamma \in]0,1[,\ \forall x \in \mathbb{R}^d,\ P\eta^{-1/2}(x) \le \gamma\eta^{-1/2}(x) + K. \tag{1.16}$$

(ii) *Tout algorithme MHMA avec la densité ψ paire et telle que* $\forall M \in]0,+\infty[$, $\inf_{|x|\le M}\psi(x) > 0$ *est géométriquement ergodique.*

Exemple 1.3.36. *Les hypothèses de la proposition portant sur π sont satisfaites par la loi gaussienne $\mathcal{N}_d(\mu,\Gamma)$ dont la matrice de covariance $\Gamma \in \mathbb{R}^{d\times d}$ est supposée définie positive. En effet, $\nabla \ln \eta(x) = -\Gamma^{-1}(x-\mu)$ si bien que*

$$\frac{x}{|x|}\cdot\nabla\ln\eta(x) = -\frac{x}{|x|}\cdot\Gamma^{-1}x + \frac{x}{|x|}\cdot\Gamma^{-1}\mu \le |\Gamma^{-1}\mu| - |x|\inf_{\xi\in\mathcal{S}_{d-1}}\xi\cdot\Gamma^{-1}\xi$$

tend vers $-\infty$ lorsque $|x| \to \infty$. Ainsi π est sous-exponentielle. Par ailleurs, $\limsup_{|x|\to\infty}\frac{x}{|x|}\cdot\frac{\nabla\eta(x)}{|\nabla\eta(x)|} = -\inf_{\xi\in\mathcal{S}_{d-1}}\frac{\xi\cdot\Gamma^{-1}\xi}{|\xi||\Gamma^{-1}\xi|}$. *L'exercice qui suit permet de montrer que* $\inf_{\xi\in\mathcal{S}_{d-1}}\frac{\xi\cdot\Gamma^{-1}\xi}{|\Gamma^{-1}\xi|} = \frac{2\sqrt{\underline{\lambda}\bar{\lambda}}}{\underline{\lambda}+\bar{\lambda}}$ *où $\underline{\lambda}$ et $\bar{\lambda}$ désignent la plus petite et la plus grande valeurs propres de Γ^{-1}.*

Exercice 1.3.37. *On suppose que $M \in \mathbb{R}^{d\times d}$ est une matrice symétrique définie positive. On note $0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_d$ ses valeurs propres et on considère une base orthonormée de vecteurs propres de $\mathbb{R}^d$ associés à ces valeurs propres. Pour $\xi \in \mathcal{S}_{d-1}$, on note $a_1,\ldots,a_d$ les coordonnées de ξ dans cette base orthonormée.*

1. *Vérifier que* $\left(\frac{|M\xi|}{\xi\cdot M\xi}\right)^2 = \frac{\sum_{i=1}^d\lambda_i^2 p_i}{(\sum_{i=1}^d\lambda_i p_i)^2}$ *où* $(p_i = a_i^2)_{1\le i\le d}$ *est une probabilité. Que se passe-t-il si $\lambda_d = \lambda_1$?*
2. *Soit X est une variable aléatoire à valeurs dans $[\lambda_1,\lambda_d]$. Dans le cas où $\lambda_d > \lambda_1$, vérifier que si* $q = \mathbb{E}[\frac{X-\lambda_1}{\lambda_d-\lambda_1}]$, *on a* $q\in[0,1]$, $\lambda_1(1-q)+\lambda_d q = \mathbb{E}[X]$ *et* $\lambda_1^2(1-q)+\lambda_d^2 q = \mathbb{E}[X^2]+\mathbb{E}[(\lambda_d-X)(X-\lambda_1)]$. *Conclure que* $\frac{\mathbb{E}[X^2]}{\mathbb{E}^2[X]} \le \frac{\lambda_1^2(1-q)+\lambda_d^2 q}{(\lambda_1(1-q)+\lambda_d q)^2}$.
3. *Vérifier que* $\sup_{q\in[0,1]}\frac{\lambda_1^2+q(\lambda_d^2-\lambda_1^2)}{(\lambda_1+q(\lambda_d-\lambda_1))^2} = \frac{(\lambda_1+\lambda_d)^2}{4\lambda_1\lambda_d}$.
4. *Conclure que* $\sup_{\xi\in\mathcal{S}_{d-1}}\frac{|M\xi|}{\xi\cdot M\xi} = \frac{(\lambda_1+\lambda_d)}{2\sqrt{\lambda_1\lambda_d}}$.

Remarque 1.3.38. — *S'il existe $M \in]0,+\infty[$ t.q. $\inf_{|x|\le M}\psi(x) > 0$, alors* $\inf_{\zeta\in\mathcal{S}_{d-1}}\int_0^\infty\psi(r\zeta)r^{d-1}dr > 0$. *Ainsi l'hypothèse faite sur ψ dans* **(ii)** *implique celle faite dans* (**i**).

— *Les hypothèses que nous formulons dans la proposition et plus généralement dans tout ce paragraphe ne dépendent pas de la constante de proportionalité entre η et la densité de π : elles sont satisfaites par η si et seulement si elles sont satisfaites par $c\eta$ pour toute constante multiplicative $c \in]0,+\infty[$.*

— *Sous les hypothèses de la proposition, la condition* (1.16) *reste valable avec la fonction $\eta^{-1/2}$ remplacée par η^{-s} où $s \in]0,1[$ est arbitraire.*

La preuve repose sur les deux lemmes suivants.

Lemme 1.3.39. *Soit π telle que la fonction η est C^1, telle que $\nabla\eta$ ne s'annule pas en dehors d'un compact et que* $\ell := -\limsup_{|x|\to\infty}\frac{x}{|x|}\cdot\frac{\nabla\eta(x)}{|\nabla\eta(x)|} > 0$. *On suppose*

$$\exists\gamma \in]0,\ell[\ t.q.\ \inf_{\xi\in\mathcal{S}_{d-1}}\int_{r=0}^\infty\int_{\zeta\in\mathcal{S}_{d-1}}1_{\{|\zeta-\xi|\le\gamma\}}\psi(r\zeta)\omega_{d-1}(d\zeta)r^{d-1}dr > 0. \tag{1.17}$$

alors l'ensemble $\mathcal{A}_x = \{y\in\mathbb{R}^d : \eta(y)\ge\eta(x)\}$ est tel que $\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy > 0$.

Remarque 1.3.40. *Une condition suffisante pour que (1.17) soit satisfaite est que* $\inf_{\zeta\in\mathcal{S}_{d-1}}\int_0^\infty \psi(r\zeta)r^{d-1}dr > 0$. *Notons que la fonction* $\zeta \in \mathcal{S}_{d-1} \mapsto \int_0^\infty \psi(r\zeta)r^{d-1}dr$ *s'interprète comme la densité de* $\frac{Z}{|Z|}$ *par rapport à* ω_{d-1} *lorsque* Z *possède la densité* ψ.

Lemme 1.3.41. *Pour* $\varepsilon \in]0,1]$ *et* $x \in \mathbb{R}^d$, *on note*

$$\mathcal{C}_x^\varepsilon = \{y \in \mathbb{R}^d : \varepsilon \leq \frac{\eta(y)}{\eta(x)} \leq \frac{1}{\varepsilon}\}.$$

Si π *est sous-exponentielle, alors* $\limsup_{|x|\to\infty}\int_{\mathcal{C}_x^\varepsilon}\psi(y-x)dy = 0$.

Démontrons la proposition 1.3.35 avant de prouver les deux lemmes.
Démonstration : Montrons d'abord **(ii)** en supposant **(i)**. Comme $q(x,y) = \psi(y-x)$, pour $M \in]0,+\infty[$, $\inf_{|x|\vee|y|\leq M} q(x,y) \geq \inf_{|z|\leq 2M}\psi(z)$. La continuité et la stricte positivité de η permettent d'appliquer le lemme 1.3.30 et d'obtenir que pour tout $M > 0$, il existe une constante $\alpha_M > 0$ telle que

$$\forall x, y \in \mathbb{R}^d \text{ t.q. } |x| \vee |y| \leq M,\ P(x,.) \wedge P(y,.) \geq \alpha_M.$$

Comme (1.15) assure que $\lim_{|x|\to\infty}\eta(x) = 0$, on peut choisir M telle que $\sup_{|x|>M}\eta(x) < \frac{(1-\gamma)^2}{4K^2}$ où γ et K sont les constantes qui apparaissent dans (1.16). Alors $R := \inf_{|x|>M}\eta^{-1/2}(x) > \frac{2K}{1-\gamma}$ et $\eta^{-1/2}(x) \leq R \Rightarrow |x| \leq M$. Ainsi la condition de dérive (D1) et (D2) est satisfaite pour $V(x) = \eta^{-1/2}(x)$ et l'algorithme MHMA est géométriquement ergodique.

Il nous reste à démontrer **(i)**. Soit $\varepsilon > 0$. D'après (1.14) et comme $\int_{\mathcal{R}_x}\psi(y-x)dy = 1 - \int_{\mathcal{A}_x}\psi(y-x)dy$,

$$\frac{P\eta^{-1/2}(x)}{\eta^{-1/2}(x)} - 1 = \int_{\mathcal{C}_x^\varepsilon}\left(1_{\mathcal{R}_x}(y)\left(\frac{\eta^{1/2}(y)}{\eta^{1/2}(x)} - \frac{\eta(y)}{\eta(x)}\right) + 1_{\mathcal{A}_x}(y)\frac{\eta^{1/2}(x)}{\eta^{1/2}(y)}\right)\psi(y-x)dy$$
$$+ \int_{\mathcal{R}_x\cap(\mathcal{C}_x^\varepsilon)^c}\left(\frac{\eta^{1/2}(y)}{\eta^{1/2}(x)} - \frac{\eta(y)}{\eta(x)}\right)\psi(y-x)dy + \int_{\mathcal{A}_x\cap(\mathcal{C}_x^\varepsilon)^c}\frac{\eta^{1/2}(x)}{\eta^{1/2}(y)}\psi(y-x)dy - \int_{\mathcal{A}_x}\psi(y-x)dy$$

Pour $y \in \mathcal{R}_x$, $\eta(y) < \eta(x)$ et donc $0 \leq \frac{\eta^{1/2}(y)}{\eta^{1/2}(x)} - \frac{\eta(y)}{\eta(x)} \leq \frac{1}{4}1_{\mathcal{C}_x^\varepsilon}(y) + \sqrt{\varepsilon}1_{(\mathcal{C}_x^\varepsilon)^c}(y)$ (en utilisant $\max_{u\in[0,1]}\sqrt{u} - u = \frac{1}{4}$) tandis que pour $y \in \mathcal{A}_x$, $\eta(y) \geq \eta(x)$ si bien que $0 \leq \frac{\eta^{1/2}(x)}{\eta^{1/2}(y)} \leq 1_{\mathcal{C}_x^\varepsilon}(y) + \sqrt{\varepsilon}1_{(\mathcal{C}_x^\varepsilon)^c}(y)$. Ainsi

$$\begin{aligned}\frac{P\eta^{-1/2}(x)}{\eta^{-1/2}(x)} - 1 &\leq \int_{\mathcal{C}_x^\varepsilon}\psi(y-x)dx + \sqrt{\varepsilon}\int_{(\mathcal{R}_x\cup\mathcal{A}_x)\cap(\mathcal{C}_x^\varepsilon)^c}\psi(y-x)dy - \int_{\mathcal{A}_x}\psi(y-x)dy\\ &\leq \int_{\mathcal{C}_x^\varepsilon}\psi(y-x)dx + \sqrt{\varepsilon} - \int_{\mathcal{A}_x}\psi(y-x)dy.\end{aligned}$$

D'après le lemme 1.3.39 et la remarque 1.3.40, $\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy > 0$. Pour le choix $\varepsilon = \left(\frac{1}{3}\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy\right)^2$, le lemme 1.3.41 assure

$$\exists M < \infty,\ \forall |x| \geq M,\ P\eta^{-1/2}(x) \leq \left(1 - \frac{1}{3}\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy\right)\eta^{-1/2}(x).$$

Posons $\gamma = \left(1 - \frac{1}{3}\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy\right)$. Pour $|x| \le M$, d'après le lemme 1.3.32,

$$P\eta^{-1/2}(x) \le \frac{5}{4}\eta^{-1/2}(x) \le \gamma\eta^{-1/2}(x) + \left(\frac{5}{4} - \gamma\right)\sup_{|y|\le M}\eta^{-1/2}(y),$$

où le second membre est fini car η est continue et strictement positive. Donc (1.16) est satisfaite pour $K = \left(\frac{5}{4} - \gamma\right)\sup_{|y|\le M}\eta^{-1/2}(y)$. □

Remarque 1.3.42. *Soit π sous-exponentielle et ψ paire et t.q. $\forall M \in]0,+\infty[$, $\inf_{|x|\le M}\psi(x) > 0$. La preuve qui précède assure que l'algorithme MHMA est géométriquement ergodique dès lors que $\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy > 0$.*
Par ailleurs, comme $P(x,\{x\}) = \int_{\mathcal{R}_x}\left(1 - \frac{\eta(y)}{\eta(x)}\right)\psi(y-x)dy$, on a

$$\begin{aligned}
P(x,\{x\}) - 1 + \int_{\mathcal{A}_x}\psi(y-x)dy &= -\int_{\mathcal{R}_x\cap\mathcal{C}_x^\varepsilon}\frac{\eta(y)}{\eta(x)}\psi(y-x)dy \\
&\ge -\int_{\mathcal{R}_x\cap\mathcal{C}_x^\varepsilon}\psi(y-x)dy - \varepsilon\int_{\mathcal{R}_x\cap(\mathcal{C}_x^\varepsilon)^c}\psi(y-x)dy \\
&\ge -\int_{\mathcal{C}_x^\varepsilon}\psi(y-x)dy - \varepsilon.
\end{aligned}$$

Avec le lemme 1.3.41 qui s'applique comme π est sous-exponentielle, on en déduit que $\lim_{|x|\to\infty}\left(P(x,\{x\}) - 1 + \int_{\mathcal{A}_x}\psi(y-x)dy\right) = 0$. Le théorème 5.1 p103 [67], assure que si $\limsup_{|x|\to\infty}P(x,\{x\}) = 1$, l'algorithme MHMA n'est pas géométrique ergodique. Donc $\liminf_{|x|\to\infty}\int_{\mathcal{A}_x}\psi(y-x)dy > 0$ est une condition nécessaire et suffisante d'ergodicité géométrique dans ce cadre (voir le théorème 4.1 p349 [39]).

Démontrons maintenant le lemme 1.3.39.
Démonstration : Soit $\gamma \in]0,\ell[$ t.q. $\inf_{\xi\in\mathcal{S}_{d-1}}\int_{r=0}^{\infty}\int_{\zeta\in\mathcal{S}_{d-1}}1_{\{|\zeta-\xi|\le\gamma\}}\psi(r\zeta)\omega_{d-1}(d\zeta)r^{d-1}dr > 0$. Comme par convergence dominée, $\lim_{K\to\infty}\int_{r=K}^{\infty}\int_{\zeta\in\mathcal{S}_{d-1}}\psi(r\zeta)\omega_{d-1}(d\zeta)r^{d-1}dr = 0$ on peut choisir $K > 0$ tel que $\inf_{\xi\in\mathcal{S}_{d-1}}\int_{r=0}^{K}\int_{\zeta\in\mathcal{S}_{d-1}}1_{\{|\zeta-\xi|\le\gamma\}}\psi(r\zeta)\omega_{d-1}(d\zeta)r^{d-1}dr > 0$.

Pour $x \in \mathbb{R}^d$ t.q. $|x| > 0$, soit $\mathcal{W}_x = \{x + r\zeta :\ 0 < r \le K \text{ et } |\zeta - \frac{x}{|x|}| \le \gamma\}$. Comme $\int_{\mathcal{W}_x}\psi(y-x)dy = \int_{r=0}^{K}\int_{\zeta\in\mathcal{S}_{d-1}}1_{\{|\zeta-\frac{x}{|x|}|\le\gamma\}}\psi(r\zeta)\omega_{d-1}(d\zeta)r^{d-1}dr$, il suffit de montrer que $\mathcal{W}_x \in \mathcal{A}_x$ pour $|x|$ assez grand pour conclure.

D'après l'hypothèse sur π,

$$\exists M > 0,\ \forall |y| \ge M,\ \frac{y}{|y|}.\frac{\nabla\eta(y)}{|\nabla\eta(y)|} < -\frac{\ell+\gamma}{2}. \tag{1.18}$$

Par ailleurs, pour $x \in \mathbb{R}^d$ tel que $|x| > K$ et y dans la boule $B(x,K)$ fermée centrée en x et de rayon K ,

$$\left|\frac{y}{|y|} - \frac{x}{|x|}\right| \le \left|\frac{|x|-|y|}{|x||y|}y\right| + \left|\frac{y-x}{|x|}\right| \le 2\frac{|y-x|}{|x|} \le \frac{2K}{|x|}.$$

Soit $|x| \ge \frac{4K}{\ell-\gamma} \vee (M+K)$. Pour $y \in \mathcal{W}_x$, en utilisant la définition de $\mathcal{W}_x$ et les deux inégalités précédentes, on a donc

$$\frac{x-y}{|x-y|}.\frac{\nabla\eta(y)}{|\nabla\eta(y)|} \le \left|\frac{x-y}{|x-y|} - \frac{x}{|x|}\right| + \left|\frac{x}{|x|} - \frac{y}{|y|}\right| + \frac{y}{|y|}.\frac{\nabla\eta(y)}{|\nabla\eta(y)|} < \gamma + \frac{\ell-\gamma}{2} - \frac{\ell+\gamma}{2} < 0.$$

Pour $z \neq x$ t.q. $\eta(z) = \eta(x)$, comme la fonction $f(t) = \eta(x + t(z - x)) - \eta(x)$ s'annule en $t = 0$ et $t = 1$, le théorème de Rolle assure l'existence de $t \in]0, 1[$ tel que $f'(t) = 0$ i.e. $(z - x).\nabla\eta(x + t(z - x)) = 0$. Ainsi $y = x + t(z - x)$ vérifie $\frac{x-y}{|x-y|}.\frac{\nabla\eta(y)}{|\nabla\eta(y)|} = 0$ si bien que $y \notin \mathcal{W}_x$. Comme, pour $z \in \mathcal{W}_x$, $x + t(z - x) \in \mathcal{W}_x$ pour tout $t \in]0, 1[$, on conclut que $\forall z \in \mathcal{W}_x$, $\eta(z) \neq \eta(x)$. L'inégalité (1.18) écrite au point x assure l'existence de $\varepsilon \in]0, \frac{K}{|x|}]$ tel que $\eta((1 - \varepsilon)x) > \eta(x)$. Soit $y \in \mathcal{W}_x$. Comme $(1 - \varepsilon)x \in \mathcal{W}_x$ et que $\mathcal{W}_x$ est convexe, le segment $[(1 - \varepsilon)x, y]$ est inclus dans $\mathcal{W}_x$. La fonction continue $z \mapsto \eta(z) - \eta(x)$ ne s'annulant pas sur ce segment, $\eta(y) - \eta(x) > 0$, c'est-à-dire que $y \in \mathcal{A}_x$. $\square$

Nous terminerons ce paragraphe par la démonstration du lemme 1.3.41

Démonstration : Soit $\alpha > 0$. On veut montrer que pour $|x|$ grand, $\int_{\mathcal{C}_x^\varepsilon} \psi(y - x)dy \leq \alpha$. On commence par choisir $K \in]0, +\infty[$ t.q. $\int_{|z|>K} \psi(z) \leq \frac{\alpha}{3}$ et on note $B(x, K)$ la boule fermée de $\mathbb{R}^d$ centrée en x et de rayon K. Notons que $\int 1_{\{|y-x|>K\}}\psi(y-x)dy \leq \frac{\alpha}{3}$. Ensuite le théorème de convergence dominée assure que $\int \psi(z)1_{\{\psi(z)>m\}}dz$ converge vers 0 lorsque m tend vers l'infini si bien que l'on peut choisir $m_\alpha > 0$ tel que $\int \psi(z)1_{\{\psi(z)>m_\alpha\}}dz \leq \frac{\alpha}{3}$.

$$\begin{aligned}\int_{\mathcal{C}_x^\varepsilon} \psi(y - x)dy \leq& \int 1_{\{|y-x|>K\}}\psi(y - x)dy + \int 1_{\{\psi(y-x)>m_\alpha\}}\psi(y - x)dy \\ &+ \int_{\mathcal{C}_x^\varepsilon \cap B(x,K)} 1_{\{\psi(y-x)\leq m_\alpha\}}\psi(y - x)dy \\ \leq& \frac{\alpha}{3} + \frac{\alpha}{3} + m_\alpha\lambda(\mathcal{C}_x^\varepsilon \cap B(x, K)).\end{aligned} \tag{1.19}$$

Il suffit donc de majorer $\lambda(\mathcal{C}_x^\varepsilon \cap B(x, K))$ pour $|x|$ grand. Soit $x \in \mathbb{R}^d$ t.q. $|x| > K$. On pose $T_x = \{\xi \in \mathcal{S}_{d-1} : \exists r \in [0, +\infty),\ r\xi \in B(x, K)\}$. En passant en coordonnées polaires puis en utilisant que pour $y \in B(x, K)$, $|x| - K \leq |y| \leq |x| + K$, on obtient

$$\begin{aligned}\lambda(\mathcal{C}_x^\varepsilon \cap B(x, K)) =& \int_{\mathcal{S}_{d-1}} \int_0^\infty 1_{\mathcal{C}_x^\varepsilon \cap B(x,K)}(r\xi)r^{d-1}dr\omega_{d-1}(d\xi) \\ \leq& \int_{T_x} \int_{|x|-K}^{+\infty} 1_{\mathcal{C}_x^\varepsilon}(r\xi)(|x| + K)^{d-1}dr\omega_{d-1}(d\xi)\end{aligned}$$

Pour $\beta > 0$, en raisonnant comme pour obtenir (1.15), on vérifie que pour tout $\xi \in \mathcal{S}_{d-1}$, $\forall r \geq \rho \geq M_\beta$, $\eta(r\xi) \leq e^{-\beta(r-\rho)}\eta(\rho\xi)$. Donc si $|x| - K \geq M_\beta$ et $\xi \in T_x$, en notant $\underline{r}_\xi = \inf\{r \in [|x|-K, |x|+K] : r\xi \in \mathcal{C}_x^\varepsilon\}$, on a $\eta(\underline{r}_\xi\xi) \leq \frac{\eta(x)}{\varepsilon}$ par définition de $\mathcal{C}_x^\varepsilon$ et $\forall r > \underline{r}_\xi - \frac{2\ln\varepsilon}{\beta}$, $\eta(r\xi) \leq e^{-\beta(r-\underline{r}_\xi)}\eta(\underline{r}_\xi\xi) < \varepsilon\eta(x)$ si bien que $r\xi \notin \mathcal{C}_x^\varepsilon$. Donc $\int_{|x|-K}^{+\infty} 1_{\mathcal{C}_x^\varepsilon}(r\xi)dr \leq -\frac{2\ln\varepsilon}{\beta}$. Ainsi,

$$\forall |x| > K + M_\beta,\ \lambda(\mathcal{C}_x^\varepsilon \cap B(x, K)) \leq -\frac{2\ln\varepsilon}{\beta}(|x| + K)^{d-1}\omega_{d-1}(T_x). \tag{1.20}$$

Pour majorer $\omega_{d-1}(T_x)$, on remarque que pour $\xi \in T_x$ et $r_\xi \in [0, +\infty[$ tel que $r_\xi\xi \in B(x, K)$, $|r_\xi\xi| \in [|x| - K, |x| + K]$ si bien que $r_\xi \in [|x| - K, |x| + K]$. Donc pour $r \in [|x| - K, |x| + K]$, $|r - r_\xi| \leq 2K$ et $|x - r\xi| \leq |x - r_\xi\xi| + |(r_\xi - r)\xi| \leq K + 2K \leq 3K$. Ainsi $\{r\xi :\ (r, \xi) \in [|x| - K, |x| + K] \times T_x\} \subset B(x, 3K)$ et

$$\lambda(B(x, 3K)) \geq \omega_{d-1}(T_x) \int_{|x|-K}^{|x|+K} r^{d-1}dr \geq 2K(|x| - K)^{d-1}\omega_{d-1}(T_x).$$

Avec (1.20), on conclut que

$$\forall|x| > K + M_\beta,\ \lambda(\mathcal{C}_x^\varepsilon \cap B(x,K)) \leq -\frac{\ln\varepsilon}{K\beta}\left(\frac{|x|+K}{|x|-K}\right)^{d-1}\lambda(B(0,3K)).$$

Pour $|x| \geq K+1$, $\left(\frac{|x|+K}{|x|-K}\right)^{d-1} \leq (2K+1)^{d-1}$. D'après (1.19), le choix $\beta = -\frac{3m_\alpha \ln\varepsilon}{\alpha K}(2K+1)^{d-1}\lambda(B(0,3K))$ assure que

$$\forall|x| \geq K + (M_\beta \vee 1),\ \int_{\mathcal{C}_x^\varepsilon} \psi(y-x)dy \leq \alpha.$$

□

1.4 Loi forte des grands nombres pour les moyennes ergodiques

La loi des grands nombres que nous allons énoncer repose sur le théorème ergodique de Birkhoff (voir par exemple [16] corollary 6.23 p115) :

Théorème 1.4.1. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $T : \Omega \to \Omega$ une application mesurable qui préserve la probabilité $\mathbb{P}$ ($\forall A \in \mathcal{A}$, $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$) et qui est ergodique : $\forall A \in \mathcal{A}$, $T^{-1}(A) = A \Rightarrow \mathbb{P}(A) \in \{0,1\}$. Alors, pour toute variable aléatoire Y intégrable, $\mathbb{P}\left(\{\omega : \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} Y(T^k(\omega)) = \mathbb{E}[Y]\}\right) = 1$ où on pose $T^0(\omega) = \omega$ et pour $k \in \mathbb{N}$, $T^{k+1}(\omega) = T(T^k(\omega))$ (c'est-à-dire que T^k est l'itérée k fois de T).*

Soit $(X_n)_{n\in\mathbb{N}}$ chaîne de Markov de noyau de transition P sur $(E, \mathcal{E})$. Pour mettre en valeur le rôle joué par la loi μ de X_0, on note $\mathbb{P}_\mu$ la probabilité sur l'espace Ω sous-jacent. Dans le cas particulier $\mu = \delta_x$ où $x \in E$, on adopte la notation simplifiée $\mathbb{P}_x$ au lieu de $\mathbb{P}_{\delta_x}$. On note $\mathbb{E}_\mu$ (resp. $\mathbb{E}_x$) l'espérance sous $\mathbb{P}_\mu$ (resp. $\mathbb{P}_x$).

Définition 1.4.2. *Une probabilité π invariante par P (i.e. qui vérifie $\pi P = \pi$) est dite extrémale s'il n'existe pas deux probabilités invariantes distinctes π_1 et π_2 et $t \in]0,1[$ tels que $\pi = t\pi_1 + (1-t)\pi_2$.*

Remarque 1.4.3.
- *Si π et σ sont deux probabilités invariantes pour P, alors pour tout $t \in [0,1]$, $(t\pi + (1-t)\sigma)P = t\pi P + (1-t)\sigma P = t\pi + (1-t)\sigma$ si bien que la probabilité $t\pi + (1-t)\sigma$ est invariante. Ainsi l'ensemble des probabilités invariantes pour P est un sous-ensemble convexe de $\mathcal{P}(E)$.*
- *Si π et σ sont deux probabilités invariantes extrémales distinctes pour P, alors $\pi \wedge \sigma(E) = 1 - d_{\mathrm{TV}}(\pi,\sigma) < 1$. Comme π et σ majorent toutes deux $\pi \wedge \sigma$, on obtient que $\pi \wedge \sigma = \pi P \wedge \sigma P \geq (\pi \wedge \sigma)P$. Comme l'égalité $P(x,E) = 1$ valable pour tout $x \in E$ entraîne que $\pi \wedge \sigma(E) = (\pi \wedge \sigma)P(E)$, on en déduit que $\pi \wedge \sigma = (\pi \wedge \sigma)P$. Si $\pi \wedge \sigma(E) > 0$, la probabilité $\frac{\pi\wedge\sigma}{\pi\wedge\sigma(E)}$ est alors invariante par P et elle est distincte soit de π soit de σ. Dans le premier cas, comme, $\pi = \pi - \pi\wedge\sigma + \pi\wedge\sigma = (1 - \pi\wedge\sigma(E))\frac{\pi - \pi\wedge\sigma}{1-\pi\wedge\sigma(E)} + \pi\wedge\sigma(E)\frac{\pi\wedge\sigma}{\pi\wedge\sigma(E)}$, la probabilité $\frac{\pi-\pi\wedge\sigma}{1-\pi\wedge\sigma(E)}$ est également invariante par linéarité et distincte de $\frac{\pi\wedge\sigma}{\pi\wedge\sigma(E)}$ car $\pi \neq \frac{\pi\wedge\sigma}{\pi\wedge\sigma(E)}$, ce qui contredit l'extrémalité de π. Dans le second cas, par un raisonnement analogue, on contredit l'extrémalité de σ. Donc $\pi \wedge \sigma(E) = 0$ et il existe $B \in \mathcal{E}$ tel que $\pi(B) = 1 = \sigma(B^c)$, d'après (1.3).*

Proposition 1.4.4. *Soit π une probabilité invariante extrémale pour P. Alors $\mathbb{P}_\pi((X_n)_{n\in\mathbb{N}} \in A) \in \{0,1\}$ pour tout ensemble $A \in \mathcal{E}^{\otimes\mathbb{N}}$ invariant par l'opérateur de décalage $T((x_k)_{k\in\mathbb{N}}) = (x_k)_{k\geq 1}$ au sens où $T^{-1}(A) = A$. En outre, pour toute fonction $f : E \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$, $\mathbb{P}_\pi\left(\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \to_{n\to\infty} \pi(f)\right) = 1$.*

Démonstration : Soit A invariant par l'opérateur de décalage et pour $n \in \mathbb{N}$, $Y_n = \mathbb{E}_\pi[1_A((X_k)_{k\in\mathbb{N}})|\mathcal{F}_n]$ où $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$. La martingale $(Y_n)_{n\in\mathbb{N}}$ est fermée par la variable aléatoire $1_A((X_k)_{k\in\mathbb{N}})$ qui est mesurable par rapport à la plus petite tribu contenant $\bigcup_{k\in\mathbb{N}} \mathcal{F}_k$. Elle converge donc $\mathbb{P}_\pi$ p.s. vers $1_A((X_k)_{k\in\mathbb{N}})$.

Comme

$$T^{-1}(A) = \{(x_k)_{k\in\mathbb{N}} \in E^\mathbb{N} : T((x_k)_{k\in\mathbb{N}}) \in A\} = \{(x_k)_{k\in\mathbb{N}} \in E^\mathbb{N} : (x_k)_{k\geq 1} \in A\},$$

par invariance de A par T, $1_A((x_k)_{k\in\mathbb{N}}) = 1_A((x_k)_{k\geq 1})$ et par récurrence $1_A((x_k)_{k\in\mathbb{N}}) = 1_A((x_k)_{k\geq n})$ pour tout $n \in \mathbb{N}$. Ainsi $Y_n = \mathbb{E}_\pi[1_A((X_k)_{k\geq n})|\mathcal{F}_n]$ et la propriété de Markov assure que $Y_n = \varphi(X_n)$ où la fonction $\varphi : E \to \mathbb{R}$ mesurable est définie par $\varphi(x) = \mathbb{E}_x[1_A((X_k)_{k\geq 0})]$. La suite $(\varphi(X_n))_{n\in\mathbb{N}}$ converge donc $\mathbb{P}_\pi$ p.s. vers $1_A((X_k)_{k\in\mathbb{N}})$ de même que $(Z_n := \sup_{k\geq n} \varphi(X_k))_{n\in\mathbb{N}}$. Comme π est invariante par P, sous $\mathbb{P}_\pi$, $(X_k)_{k\geq n}$ a même loi que $(X_k)_{k\geq 0}$, ce qui assure que la loi de Z_n et celle de $\varphi(X_n)$ ne dépendent pas de n. Ainsi, pour tout $n \in \mathbb{N}$, $\varphi(X_n)$ et Z_n ont même loi que $1_A((X_k)_{k\in\mathbb{N}})$. Comme $\mathbb{P}_\pi(\varphi(X_n) \leq Z_n) = 1 = \mathbb{P}_\pi(Z_{n+1} \leq Z_n)$, cela implique que $\mathbb{P}_\pi(\varphi(X_n) = \sup_{k\geq n} \varphi(X_k)) = 1 = \mathbb{P}_\pi(\sup_{k\geq n} \varphi(X_k) = \sup_{k\geq n+1} \varphi(X_k))$. On en déduit que $\mathbb{P}_\pi$ p.s., la suite $(\varphi(X_n))_{n\in\mathbb{N}}$ est constante et égale à $1_A((X_k)_{k\in\mathbb{N}})$. Pour $B = \varphi^{-1}(\{1\})$, on en déduit que $\mathbb{P}_\pi$ p.s.,

$$1_B(X_0) = 1_B(X_1) = 1_A((X_k)_{k\in\mathbb{N}}),$$

et donc $\pi(B) = \mathbb{P}_\pi((X_n)_{n\in\mathbb{N}} \in A)$.

Supposons que $0 < \pi(B) < 1$. En notant $\pi(.|B)$ et $\pi(.|B^c)$ les probabilités définies sur $(E, \mathcal{E})$ par $\pi(C|B) = \frac{\pi(C\cap B)}{\pi(B)}$ et $\pi(C|B^c) = \frac{\pi(C\cap B^c)}{\pi(B^c)}$ pour tout $C \in \mathcal{E}$, on a

$$\begin{aligned}\pi(C|B) &= \frac{\mathbb{E}_\pi[1_{C\cap B}(X_1)]}{\pi(B)} = \frac{\mathbb{E}_\pi[1_C(X_1)1_B(X_1)]}{\pi(B)} = \frac{\mathbb{E}_\pi[1_C(X_1)1_B(X_0)]}{\pi(B)} \\ &= \int_{E\times E} 1_C(y)P(x,dy)\frac{1_B(x)\pi(dx)}{\pi(B)} = \pi(.|B)P(C).\end{aligned}$$

Ainsi $\pi(.|B)$ est invariante par P et, de manière analogue, $\pi(.|B^c)$ l'est également. Comme $\pi(.|B)$ et $\pi(.|B^c)$ sont distinctes ($\pi(B|B) = 1 = \pi(B^c|B^c)$) et comme $\pi = \pi(B)\pi(.|B) + \pi(B^c)\pi(.|B^c)$, on contredit l'extrémalité de π. Donc $\mathbb{P}_\pi((X_n)_{n\in\mathbb{N}} \in A) = \pi(B) \in \{0,1\}$.

Sur $(E^\mathbb{N}, \mathcal{E}^{\otimes\mathbb{N}})$ muni de la loi de $(X_n)_{n\geq 0}$ sous $\mathbb{P}_\pi$, l'opérateur de décalage $T : (x_n)_{n\geq 0} \to (x_n)_{n\geq 1}$ est donc ergodique. Il préserve cette loi par invariance de π pour P. Le théorème ergodique de Birkhoff entraîne alors que pour toute fonction $F : E^\mathbb{N} \to \mathbb{R}$ mesurable et telle que $\mathbb{E}_\pi[|F((X_k)_{k\in\mathbb{N}})|] < \infty$, $(\frac{1}{n}\sum_{j=0}^{n-1} F((X_k)_{k\geq j})_{n\geq 1}$ converge $\mathbb{P}_\pi$ p.s. vers $\mathbb{E}_\pi[F((X_k)_{k\in\mathbb{N}})]$ lorsque $n \to \infty$. Dans le cas particulier où $F((x_k)_{k\in\mathbb{N}}) = f(x_0)$ avec $f : E \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$, on en déduit que $\mathbb{P}_\pi$ p.s., $\frac{1}{n}\sum_{j=0}^{n-1} f(X_j)$ converge vers $\pi(f)$ lorsque $n \to \infty$. □

Exercice 1.4.5. *Soit $(\Omega, \mathcal{A})$ un espace mesurable et $T : \Omega \to \Omega$ une application mesurable.*

1. *Montrer que l'ensemble $\mathcal{P}$ des probabilités $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ qui sont préservées par T ($\forall A \in \mathcal{A},\ \mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$) est convexe.*
2. *Montrer que si $\mathbb{P} \in \mathcal{P}$ est extrêmale i.e. ne s'écrit pas $\mathbb{P} = t\mathbb{P}_1 + (1-t)\mathbb{P}_2$ pour deux éléments distincts $\mathbb{P}_1$ et $\mathbb{P}_2$ de $\mathcal{P}$ et $t \in]0,1[$, alors T est ergodique pour $\mathbb{P}$.*

Nous sommes maintenant prêts à énoncer la loi forte des grands nombres sous condition de dérive.

Théorème 1.4.6. *On suppose que le noyau markovien P satisfait (D1) et (D2) et on note π sa probabilité invariante dont l'existence et l'unicité sont assurées par le théorème 1.3.27. Alors pour toute fonction $f : E \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$,*

$$\forall \mu \in \mathcal{P}(E),\ \mathbb{P}_\mu \left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \longrightarrow_{n\to\infty} \pi(f) \right) = 1.$$

Remarque 1.4.7. — *Comme, d'après le théorème 1.3.27, $\pi(V) < \infty$, $\sup_{x\in E} \frac{|f(x)|}{1+V(x)} < \infty$ (i.e. $f \in \mathcal{V}$) est une condition suffisante pour que $\pi(|f|) < \infty$.*

— *Notons que si $f : E \to \mathbb{R}_+$ est mesurable et vérifie $\pi(f) = +\infty$, alors appliquer le théorème avec $f \wedge m$ où $m \in \mathbb{N}$ entraîne que pour tout $x \in E$, $\mathbb{P}_x \left(\liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \geq \pi(f \wedge m)\right) = 1$. Comme, par convergence monotone, $\lim_{m\to\infty} \pi(f \wedge m) = \pi(f) = \infty$, $\mathbb{P}_x \left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = +\infty\right) = 1$.*

Démonstration : Comme P admet π comme unique probabilité invariante, celle-ci est extrémale. D'après la proposition 1.4.4, pour $f : E \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$, $\mathbb{P}_\pi \left(\frac{1}{n} \sum_{j=0}^{n-1} f(X_j) \to_{n\to\infty} \pi(f)\right) = 1$.

Reste à démontrer que la convergence a aussi lieu $\mathbb{P}_\mu$ p.s. où $\mu \in \mathcal{P}(E)$ est quelconque. Comme

$$\mathbb{P}_\mu \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(X_j) = \pi(f) \right) = \int_E \mathbb{P}_x \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(X_j) = \pi(f) \right) \mu(dx),$$

il suffit de montrer qu'elle a lieu $\mathbb{P}_x$ p.s. pour tout $x \in E$. Soit $k \in \mathbb{N}$, comme $\lim_{n\to\infty} \frac{n+k}{n} = 1$ et $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{k-1} f(X_j) = 0$, $\frac{1}{n} \sum_{j=0}^{n-1} f(X_j)$ et $\frac{1}{n} \sum_{j=k}^{k+n-1} f(X_j) = \frac{n+k}{n} \times \frac{1}{n+k} \sum_{j=0}^{k+n-1} f(X_j) - \frac{1}{n} \sum_{j=0}^{k-1} f(X_j)$ ont mêmes valeurs d'adhérence lorsque $n \to \infty$. Avec la propriété de Markov et en posant $\varphi(y) = \mathbb{P}_y \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(X_j) = \pi(f)\right)$, on en déduit que

$$\begin{aligned}
\mathbb{P}_x \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(X_j) = \pi(f) \right) &= \mathbb{P}_x \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=k}^{k+n-1} f(X_j) = \pi(f) \right) \\
&= P^k\varphi(x) \geq \pi(\varphi) - |\pi(\varphi) - P^k\varphi(x)| \\
&= \mathbb{P}_\pi \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(X_j) = \pi(f) \right) - |\pi(\varphi - 1/2) - P^k(\varphi - 1/2)(x)| \\
&\geq 1 - \sup_{\psi : |\psi| \leq 1/2} |\pi(\psi) - P^k(x, \psi)| = 1 - d_{\mathrm{TV}}(\pi, P^k(x, .)).
\end{aligned}$$

D'après le théorème 1.3.27,

$$d_{\mathrm{TV}}(P^k(x,.),\pi) = d_{\mathrm{TV}}(\delta_x P^k,\pi) \leq \frac{\chi^k}{2} d_\beta(\delta_x,\pi) \leq \frac{\chi^k}{2}(2+\beta V(x)+\frac{\beta K}{1-\gamma})$$

tend vers 0 lorsque $n \to \infty$ pour $\beta \in]0, \frac{\alpha}{K}[$, ce qui achève la preuve. □

Dans le cas où P satisfait la condition de Doeblin uniforme, la probabilité invariante π est explicite et l'exercice suivant explique comment obtenir la conclusion du théorème 1.4.6 à l'aide de la loi forte des grands nombres usuelle.

Exercice 1.4.8. *On suppose que P satisfait la condition de Doeblin uniforme de la définition 1.3.10. Sous cette condition, $\alpha\nu(dy) = g(x,y)P(x,dy)$ pour une densité g définie sur $E \times E$, mesurable et à valeurs dans $[0,1]$. On se donne $(U_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. suivant la loi uniforme sur $[0,1]$ indépendante de la chaîne de Markov $(X_n)_{n\geq 0}$ de noyau P. On note $\tau_0 = 0$ et pour $k \geq 1$,*

$$\tau_k = \inf\{n > \tau_{k-1} : U_n \leq g(X_{n-1},X_n)\}.$$

On note Q le noyau markovien défini par $Q(x,dy) = \frac{P(x,dy)-\alpha\nu(dy)}{1-\alpha}$.

1. *Montrer que $\pi = \alpha \sum_{n\in\mathbb{N}}(1-\alpha)^n \nu Q^n$ est l'unique probabilité invariante de P.*

Soit $f : E \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$.

2. *Pour tout $x \in E$, montrer que pour $\varphi : E \to \mathbb{R}$ mesurable bornée, $\int_E \varphi(y)(1 - g(x,y))P(x,dy) = (1-\alpha)\int_E \varphi(y)Q(x,dy)$.*
3. *Soient $\psi_1 : \mathbb{N}^* \to \mathbb{R}$ et $\psi_2, \psi_3 : \bigcup_{n\in\mathbb{N}^*}\{n\} \times E^n \to \mathbb{R}$ mesurables bornée. En décomposant sur les valeurs de (τ_1,τ_2,τ_3), montrer que*

$$\begin{aligned}&\mathbb{E}\left[\psi_1(\tau_1)\psi_2(\tau_2-\tau_1, X_{\tau_1},\ldots,X_{\tau_2-1})\psi_3(\tau_3-\tau_2,X_{\tau_2},\ldots,X_{\tau_3-1})\right]\\ &\quad= \sum_{n\in\mathbb{N}^*}\alpha(1-\alpha)^{n-1}\psi_1(n)\prod_{i=2}^{3}\sum_{n\in\mathbb{N}^*}\alpha(1-\alpha)^{n-1}\int_{E^n}\psi_i(n,x_1,\ldots,x_n)\nu(dx_1)\prod_{k=1}^{n-1}Q(x_k,dx_{k+1}).\end{aligned}$$

En déduire que les variables aléatoires $\sum_{j=\tau_1}^{\tau_2-1} f(X_j)$ et $\sum_{j=\tau_2}^{\tau_3-1} f(X_j)$ sont i.i.d. intégrables d'espérance commune $\frac{\pi(f)}{\alpha}$. Comment ce résultat se généralise-t-il ?

4. *En déduire que p.s., $\frac{1}{k}\sum_{j=0}^{\tau_k-1} f(X_j) \longrightarrow_{k\to\infty} \frac{\pi(f)}{\alpha}$ et $\frac{\tau_k}{k} \longrightarrow_{k\to\infty} \frac{1}{\alpha}$.*
5. *Pour $n \in \mathbb{N}$, on note $\kappa(n) = \max\{k \in \mathbb{N} : \tau_k \leq n\}$. Montrer que p.s. $\lim_{n\to\infty}\kappa(n) = +\infty$. Lorsque $f \geq 0$, remarquer que*

$$\frac{1}{\tau_{\kappa(n)+1}}\sum_{j=0}^{\tau_{\kappa(n)}-1} f(X_j) \leq \frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \leq \frac{1}{\tau_{\kappa(n)}}\sum_{j=0}^{\tau_{\kappa(n)+1}-1} f(X_j)$$

et en déduire que p.s., $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) = \pi(f)$. Comment généraliser ce résultat sans supposer f de signe constant ?

1.5 Théorème de la limite centrale pour les moyennes ergodiques

Pour contrôler l'erreur commise dans la loi forte ergodique étudiée au paragraphe précédent, on s'intéresse maintenant au comportement asymptotique lorsque $n \to \infty$

de

$$\xi_n = \sqrt{n}\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \pi(f)\right).$$

S'il existe une fonction $F : E \to \mathbb{R}$ mesurable et telle que $\forall x \in E$, $P|F|(x) < \infty$ solution de l'équation de Poisson

$$\forall x \in E,\ F(x) - PF(x) = f(x) - \pi(f), \tag{1.21}$$

alors, pour $n \geq 1$, on peut récrire $\xi_n = \frac{F(X_0)-PF(X_{n-1})}{\sqrt{n}} + \frac{1}{\sqrt{n}}M^F_{n-1}$ où $M^F_{n-1} = \sum_{k=1}^{n-1}(F(X_k) - PF(X_{k-1}))$. Introduisons la filtration $\mathcal{F}_k = \sigma(X_0, \dots, X_k)$. Comme d'après la propriété de Markov, $\mathbb{E}[F(X_k)|\mathcal{F}_{k-1}] = PF(X_{k-1})$, $(M^F_n)_{n\in\mathbb{N}}$ est une martingale locale nulle en 0. On peut alors étudier le comportement asymptotique de $(\xi_n)_{n\geq 1}$ à l'aide du théorème de la limite centrale pour les martingales. C'est la motivation de l'étude qui suit de l'équation de Poisson (1.21). Notons que si F est solution, alors pour toute constante $c \in \mathbb{R}$, $F + c$ est également solution.

Proposition 1.5.1. *On suppose que le noyau Markovien P satisfait (D1) et (D2). Pour $f \in \mathcal{V}$, la série de terme général $(P^n f - \pi(f))_{n\in\mathbb{N}}$ converge normalement dans $\mathcal{V}$ muni de $\|.\|_1$, $F = \sum_{n\in\mathbb{N}}(P^n f - \pi(f))$ est solution de l'équation de Poisson* (1.21) *et si $\tilde{F}$ désigne une autre solution dans $\mathcal{V}$, $F - \tilde{F} = \pi(F - \tilde{F})$.*

Remarque 1.5.2. — *Notons que, pour tout $\beta > 0$, comme pour tout $x \in E$, $\frac{\min(\beta,1)}{1+\beta V(x)} \leq \frac{1}{1+V(x)} \leq \frac{\max(\beta,1)}{1+\beta V(x)}$,*

$$\min(\beta,1)\|\|_\beta \leq \|\|_1 \leq \max(\beta,1)\|\|_\beta.$$

La convergence normale a donc lieu simultanément pour toutes les normes $\|\|_\beta$.

— *Comme $F \in \mathcal{V}$, (D1) implique que $\forall x \in E$, $P|F|(x) < \infty$.*

Démonstration : Soit $f \in \mathcal{V}$, $\beta \in]0, \frac{\alpha}{K}[$ et $x \in E$. Avec la définition de d_β et le théorème 1.3.27, on obtient que

$$|P^n f(x) - \pi(f)| \leq d_\beta(P^n(x,.),\pi)\|f\|_\beta \leq \chi^n d_\beta(\delta_x,\pi)\|f\|_\beta, \tag{1.22}$$

avec $\chi \in]0,1[$ ne dépendant ni de n ni de x. Comme

$$d_\beta(\delta_x,\pi) \leq 1 + \beta V(x) + \pi(1+\beta V) \leq 2 + \frac{\beta K}{1-\gamma} + \beta V(x) \leq \max(2 + \frac{\beta K}{1-\gamma}, \beta)(1 + V(x)),$$

on obtient que

$$\|P^n f - \pi(f)\|_1 \leq \max(2 + \frac{\beta K}{1-\gamma}, \beta)\chi^n\|f\|_\beta, \tag{1.23}$$

d'où le caractère normalement convergent de la série. Soit $F = \sum_{n\in\mathbb{N}}(P^n f - \pi(f))$. Pour $\varphi \in \mathcal{V}$, $\|P\varphi\|_1 \leq \|\varphi\|_1\|P(1+V)\|_1 \leq \|\varphi\|_1\|1 + K + \gamma V\|_1 \leq (1+K)\|\varphi\|_1$ où on a utilisé (D1) pour la seconde inégalité. Ainsi l'application linéaire $\varphi \to P\varphi$ de $\mathcal{V}$ muni de $\|\|_1$ dans lui-même est continue. On en déduit que $PF = \sum_{n\in\mathbb{N}} P(P^n f - \pi(f)) = \sum_{n\in\mathbb{N}}(P^{n+1}f - \pi(f)) = \sum_{n\geq 1}(P^n f - \pi(f))$, d'où $F - PF = P^0 f - \pi(f) = f - \pi(f)$.

Soit $\tilde{F}$ une autre solution de l'équation de Poisson dans $\mathcal{V}$. Alors $G := F - \tilde{F} \in \mathcal{V}$ vérifie $G = PG$ si bien que $G = P^n G$ et $G - \pi(G) = P^n G - \pi(G)$. L'estimation (1.23) reste valable pour f remplacée par $G-\pi(G)$, si bien qu'avec le premier point de la remarque précédente, on obtient

$$\begin{aligned}\|G-\pi(G)\|_1 = \|P^nG-\pi(G)\|_1 &\leq \max(2+\frac{\beta K}{1-\gamma},\beta)\chi^n\|G-\pi(G)\|_\beta \\ &\leq \max(2+\frac{\beta K}{1-\gamma},\beta)\chi^n\max(\beta^{-1},1)\|G-\pi(G)\|_1.\end{aligned}$$

En choisissant n assez grand pour que $\max(2+\frac{\beta K}{1-\gamma},\beta)\chi^n\max(\beta^{-1},1)<1$ on conclut que $\|G-\pi(G)\|_1=0$, c'est-à-dire que $G=\pi(G)$.

□

L'inégalité de Jensen et (D1) entraînent que

$$P\sqrt{V}(x)\leq\sqrt{PV(x)}\leq\sqrt{\gamma V(x)+K}\leq\sqrt{\gamma}\sqrt{V(x)}+\sqrt{K}.$$

Comme $\sqrt{V(x)}+\sqrt{V(y)}\leq\sqrt{R}$ implique $V(x)+V(y)\leq R$ et $\sqrt{R}>\frac{2\sqrt{K}}{1-\sqrt{\gamma}}\Leftrightarrow R>\frac{4K}{(1-\sqrt{\gamma})^2}$, en remplaçant (V,K,γ) par $(\sqrt{V},\sqrt{K},\sqrt{\gamma})$ dans la proposition 1.5.1, on obtient le corollaire suivant :

Corollaire 1.5.3. *Supposons que le noyau markovien P satisfait (D1) et*

$$(\mathrm{D2}')\ \exists R>\frac{4K}{(1-\sqrt{\gamma})^2},\ \exists\alpha\in]0,1],\ \forall x,y\ t.q.\ V(x)+V(y)\leq R,\ P(x,.)\wedge P(y,.)(E)\geq\alpha.$$

Alors pour $f:E\to\mathbb{R}$ mesurable et telle que $f^2\in\mathcal{V}$, $F=\sum_{n\in\mathbb{N}}(P^nf-\pi(f))$ est solution de l'équation de Poisson (1.21) et vérifie $F^2\in\mathcal{V}$.

Remarque 1.5.4.
- *Comme $\gamma\in]0,1[$, $0<\gamma<\sqrt{\gamma}<1$ si bien que $(1-\sqrt{\gamma})^2<1-\sqrt{\gamma}<1-\gamma$, $\frac{2K}{1-\gamma}<\frac{2K}{(1-\sqrt{\gamma})^2}$ et (D2') implique (D2).*
- *La convergence de la série de terme général $(P^nf)_{n\in\mathbb{N}}$ est normale pour toute norme de la forme $\|g\|=\sup_{x\in E}\frac{|g(x)|}{1+\beta\sqrt{V(x)}}$ avec $\beta>0$.*
- *D'après la preuve de la proposition 1.5.1 (voir en particulier (1.22)), pour tout $\beta\in]0,\frac{\alpha}{\sqrt{K}}[$ et tout $n\in\mathbb{N}$,*

$$\sup_{x\in E}\frac{|P^nf(x)-\pi(f)|}{1+\sqrt{V(x)}}\leq\max(2+\frac{\beta\sqrt{K}}{1-\sqrt{\gamma}},\beta)\chi^n\sup_{x\in E}\frac{|f(x)|}{1+\beta\sqrt{V(x)}}\tag{1.24}$$

où $\chi=\chi(\alpha,\beta,\sqrt{\gamma},\sqrt{K},\sqrt{R})\in]0,1[$ d'après le théorème 1.3.27

Exercice 1.5.5. *On suppose que P satisfait la condition de Doeblin uniforme de la définition 1.3.10 et on note Q le noyau markovien défini par $Q(x,dy)=\frac{P(x,dy)-\alpha\nu(dy)}{1-\alpha}$. Montrer que pour $f:E\to\mathbb{R}$ mesurable bornée, $F=\sum_{n\in\mathbb{N}}(1-\alpha)^nQ^nf$ est solution de l'équation de Poisson $F-PF=f-\pi(f)$.*

Théorème 1.5.6. *Soit $(X_n)_{n\geq0}$ chaîne de Markov dont le noyau P satisfait (D1) et (D2'). Alors pour toute fonction $f:E\to\mathbb{R}$ mesurable et telle que $\sup_{x\in E}\frac{f^2(x)}{1+V(x)}<\infty$ et quelle que soit la loi $\mu\in\mathcal{P}(E)$ de X_0, la suite $\left(\xi_n=\sqrt{n}\left(\frac{1}{n}\sum_{k=0}^{n-1}f(X_k)-\pi(f)\right)\right)_{n\geq1}$ converge en loi vers $\mathcal{N}_1(0,\sigma^2(f))$ où $\sigma^2(f)=\pi(F^2)-\pi((PF)^2)$ avec F solution de l'équation de Poisson (1.21) dont l'existence est assurée par le corollaire 1.5.3.*

Remarque 1.5.7. — *Notons que comme $F^2 \in \mathcal{V}$ et $\pi(1+V) \leq 1 + \frac{K}{1-\gamma}$ (d'après le théorème 1.3.11), $\pi(F^2) < \infty$. Par ailleurs l'inégalité de Jensen et l'invariance de π par P assurent que $\pi((PF)^2) \leq \pi(PF^2) = \pi(F^2)$ si bien que $0 \leq \sigma^2(f) < \infty$.*

— *Notons que $F^2 - (PF)^2 = (F - PF)(F - PF + 2PF) = (f - \pi(f))^2 + 2(f - \pi(f))\sum_{k\geq 1}(P^k f - \pi(f))$ avec $\sum_{k\in\mathbb{N}} \sup_{x\in E} \frac{|P^k f(x) - \pi(f)|}{1+\sqrt{V(x)}} < \infty$ si bien que $\sum_{k\in\mathbb{N}} \sup_{x\in E} \frac{|(f(x)-\pi(f))(P^k f(x)-\pi(f))|}{1+V(x)} < \infty$. Ainsi la variance asymptotique admet l'expression alternative*

$$\sigma^2(f) = \pi((f - \pi(f))^2) + 2\sum_{k\geq 1} \mathbb{E}_\pi[(f(X_0) - \pi(f))(f(X_k) - \pi(f))],$$

où le premier terme du membre de droite est la variance de $f(X_0)$ et le terme d'indice $k \geq 1$ de la somme est la covariance entre $f(X_k)$ et $f(X_0)$, lorsque X_0 est distribuée suivant la probibilité invariante π.

— *Par ailleurs, par l'inégalité de Jensen,*

$$(F(X_k) - PF(X_{k-1}))^2 \leq 2(F^2(X_k) + (PF(X_{k-1}))^2) \leq 2(F^2(X_k) + PF^2(X_{k-1}))).$$

Donc, pour $x \in E$, $\mathbb{E}_x\left[(F(X_k) - PF(X_{k-1}))^2\right] \leq 4P^k F^2(x)$ où le membre de droite est fini d'après (D1) puisque $F^2 \in \mathcal{V}$. Ainsi, sous $\mathbb{P}_x$, $(M_n^F)_{n\geq 1}$ est une martingale de carré intégrable.
Notons que $< M^F >_{n-1} = \sum_{k=0}^{n-2} \mathbb{E}[(F(X_{k+1}) - PF(X_k))^2|\mathcal{F}_k] = \sum_{k=0}^{n-2}(PF^2(X_k) - (PF(X_k))^2)$ si bien que, par la loi forte des grands nombres ergodique, $\frac{<M^F>_{n-1}}{n}$ converge presque sûrement vers $\pi(PF^2 - (PF)^2) = \sigma^2(f)$ par invariance de π pour P.

Démonstration : Le fait que pour $\mu \in \mathcal{P}(E)$ et $\varphi : \mathbb{R} \to \mathbb{R}$ continue bornée, $\mathbb{E}_\mu[\varphi(\xi_n)] = \int_E \mathbb{E}_x[\varphi(\xi_n)]\mu(dx)$ et le théorème de convergence dominée assurent qu'il suffit de démontrer le résultat sous $\mathbb{P}_x$ avec $x \in E$ quelconque.

Avant de voir comment contrôler l'erreur introduite en approchant F par une fonction bornée, nous allons commencer par montrer le résultat en supposant F bornée par m

On a $\xi_n = \frac{F(X_0)-PF(X_{n-1})}{\sqrt{n}} + \frac{1}{\sqrt{n}} M_{n-1}^F$ où $M_{n-1}^F = \sum_{k=1}^{n-1}(F(X_k) - PF(X_{k-1}))$ et $\frac{F(X_0)-PF(X_{n-1})}{\sqrt{n}}$ converge p.s. vers 0 lorsque $n \to \infty$ car $|F| \leq m$ et $|PF| \leq P|F| \leq m$ entraîne que $\frac{|F(X_0)-PF(X_{n-1})|}{\sqrt{n}} \leq \frac{2m}{\sqrt{n}}$. Le théorème de Slutsky assure qu'il suffit de montrer que $\frac{M_{n-1}^F}{\sqrt{n}}$ converge en loi vers $\mathcal{N}_1(0, \sigma^2(f))$.

Pour cela, utilisons la fonction caractéristique et introduisons $a_k^n = \mathbb{E}_x\left[e^{i\frac{u}{\sqrt{n}}(F(X_k)-PF(X_{k-1}))}|\mathcal{F}_{k-1}\right]$. Comme $|F| \leq m$ et $|PF| \leq P|F| \leq m$,

$$\left|e^{i\frac{u}{\sqrt{n}}(F(X_k)-PF(X_{k-1}))} - 1 - \frac{iu}{\sqrt{n}}(F(X_k) - PF(X_{k-1})) + \frac{u^2}{2n}(F(X_k) - PF(X_{k-1}))^2\right| \leq \frac{4m^3|u|^3}{3n^{3/2}}.$$

En prenant l'espérance conditionnelle du terme entre valeurs absolues et en remarquant que $\mathbb{E}_x[F(X_k) - PF(X_{k-1})|\mathcal{F}_{k-1}] = 0$ et

$$\mathbb{E}_x[(F(X_k) - PF(X_{k-1}))^2|\mathcal{F}_{k-1}] = [PF^2 - (PF)^2](X_{k-1}),$$

on en déduit que

$$\left| a_k^n - 1 + \frac{u^2}{2n}[PF^2 - (PF)^2](X_{k-1}) \right| \leq \frac{4m^3|u|^3}{3n^{3/2}}. \tag{1.25}$$

En particulier, il existe $C \in]0, +\infty[$ telle que $\forall n \geq 1, \forall 1 \leq k \leq n-1,\ |a_k^n| \geq 1 - \frac{C}{n}$. On suppose maintenant $n > C$ si bien que $\left|\frac{1}{\prod_{k=1}^{n-1} a_k^n}\right| \leq \left(1 - \frac{C}{n}\right)^{-n}$.

Comme $\frac{e^{i\frac{u}{\sqrt{n}}M_{n-2}^F}}{\prod_{k=1}^{n-1} a_k^n}$ est $\mathcal{F}_{n-2}$-mesurable,

$$\begin{aligned}\mathbb{E}_x\left[\frac{e^{i\frac{u}{\sqrt{n}}M_{n-1}^F}}{\prod_{k=1}^{n-1} a_k^n}\right] &= \mathbb{E}_x\left[e^{i\frac{u}{\sqrt{n}}(F(X_{n-1})-PF(X_{n-2}))}\frac{e^{i\frac{u}{\sqrt{n}}M_{n-2}^F}}{\prod_{k=1}^{n-1} a_k^n}\right] \\ &= \mathbb{E}_x\left[a_{n-1}^n\frac{e^{i\frac{u}{\sqrt{n}}M_{n-2}^F}}{\prod_{k=1}^{n-1} a_k^n}\right] = \mathbb{E}_x\left[\frac{e^{i\frac{u}{\sqrt{n}}M_{n-2}^F}}{\prod_{k=1}^{n-2} a_k^n}\right],\end{aligned}$$

et, en itérant ce raisonnement, $\mathbb{E}_x\left[\frac{e^{i\frac{u}{\sqrt{n}}M_{n-1}^F}}{\prod_{k=1}^{n-1} a_k^n}\right] = 1$. Ainsi,

$$\begin{aligned}\left|\mathbb{E}_x\left[e^{i\frac{u}{\sqrt{n}}M_{n-1}^F}\right] - e^{-\frac{u^2\sigma_F^2}{2}}\right| &= \left|\mathbb{E}_x\left[e^{i\frac{u}{\sqrt{n}}M_{n-1}^F}\left(1 - \frac{e^{-\frac{u^2\sigma_F^2}{2}}}{\prod_{k=1}^{n-1} a_k^n}\right)\right]\right| \leq \mathbb{E}_x\left|\frac{\prod_{k=1}^{n-1} a_k^n - e^{-\frac{u^2\sigma_F^2}{2}}}{\prod_{k=1}^{n-1} a_k^n}\right| \\ &\leq \left(1 - \frac{C}{n}\right)^{-n}\mathbb{E}_x\left|\prod_{k=1}^{n-1} a_k^n - e^{-\frac{u^2\sigma_F^2}{2}}\right|\end{aligned}$$

où le premier facteur au second membre tend vers e^C lorsque $n \to \infty$. Pour montrer que le second facteur tend vers 0, on pose $b_k^n = e^{-\frac{u^2}{2n}[PF^2-(PF)^2](X_{k-1})}$ et on écrit

$$\mathbb{E}_x\left|\prod_{k=1}^{n-1} a_k^n - e^{-\frac{u^2\sigma_F^2}{2}}\right| \leq \left|\prod_{k=1}^{n-1} a_k^n - \prod_{k=1}^{n-1} b_k^n\right| + \mathbb{E}_x\left|e^{-\frac{u^2}{2n}\sum_{k=0}^{n-2}[PF^2-(PF)^2](X_k)} - e^{-\frac{u^2\sigma_F^2}{2}}\right|.$$

La positivité et la convergence presque sûre de $\frac{n-1}{n} \times \frac{1}{n-1}\sum_{k=0}^{n-2}[PF^2 - (PF)^2](X_k)$ vers $\sigma_F^2 := \pi(PF^2 - (PF)^2)$ déduite du théorème 1.4.6 assurent que le second terme au membre de droite de l'inégalité tend vers 0 lorsque $n \to \infty$ par convergence dominée. Montrons que

$$\prod_{k=1}^{n-1} a_k^n - \prod_{k=1}^{n-1} b_k^n = \sum_{k=1}^{n-1}\prod_{j=1}^{k-1} a_j^n(a_k^n - b_k^n)\prod_{j=k+1}^{n-1} b_j^n$$

tend également vers 0, ce qui terminera la preuve dans le cas où F est bornée.

Comme $0 \leq PF^2 - (PF)^2 \leq m^2$,

$$\left|1 - \frac{u^2}{2n}[PF^2 - (PF)^2](X_{k-1}) - b_k^n\right| \leq \frac{u^4m^4}{8n^2} \text{ et avec (1.25), } |a_k^n - b_k^n| \leq \frac{4m^3|u|^3}{3n^{3/2}} + \frac{u^4m^4}{8n^2}.$$

En outre, les définitions de a_k^n et b_k^n assurent que $\forall 1 \leq k \leq n-1,\ |a_k^n| \vee |b_k^n| \leq 1$. On en déduit que

$$\left|\prod_{k=1}^{n-1} a_k^n - \prod_{k=1}^{n-1} b_k^n\right| \leq \frac{4m^3|u|^3}{3\sqrt{n}} + \frac{u^4m^4}{8n},$$

où le second membre tend vers 0 lorsque $n \to \infty$.

Traitons maintenant le cas général où F n'est plus supposée bornée. Clairement $\frac{F(X_0)}{\sqrt{n}}$ tend p.s. vers 0 lorsque $n \to \infty$. Par ailleurs, en utilisant l'inégalité de Jensen puis (D1), on obtient

$$\mathbb{E}_x\left[\left(\frac{PF(X_{n-1})}{\sqrt{n}}\right)^2\right] \leq \frac{1}{n}\mathbb{E}_x[PF^2(X_{n-1})] = \frac{P^nF^2(x)}{n} \leq \frac{\|F^2\|_1}{n}\left(1+\gamma^nV(x)+K\sum_{k=0}^{n-1}\gamma^k\right),$$

ce qui assure que $\frac{PF(X_{n-1})}{\sqrt{n}}$ converge vers 0 en moyenne quadratique lorsque $n \to \infty$. Le théorème de Slutsky assure qu'il suffit de montrer que $\frac{M^F_{n-1}}{\sqrt{n}}$ converge en loi vers $\mathcal{N}_1(0,\sigma^2(f))$ pour conclure la démonstration. Pour cela, nous allons remplacer F par la fonction bornée $F_m = -m \vee F \wedge m$ où $m \in \mathbb{N}$. En introduisant la notation $G_m = F - F_m$, nous avons $M^F_{n-1} - M^{F_m}_{n-1} = M^{G_m}_{n-1}$ si bien qu'avec la propriété de martingale de $(M^{G_m}_k)_{k\geq 1}$,

$$\begin{aligned}\mathbb{E}_x\left[\left(\frac{M^F_{n-1}-M^{F_m}_{n-1}}{\sqrt{n}}\right)^2\right] &= \frac{1}{n}\sum_{k=1}^{n-1}\mathbb{E}_x\left[(G_m(X_k)-PG_m(X_{k-1}))^2\right]\\ &= \frac{1}{n}\sum_{k=1}^{n-1}\mathbb{E}_x\left[G_m^2(X_k)-(PG_m(X_{k-1}))^2\right] \leq \frac{1}{n}\sum_{k=1}^{n-1}P^kG_m^2(x).\end{aligned}$$

Comme $G_m^2 \leq F^2$, la fonction G_m^2 est dans $\mathcal{V}$ et, en remplaçant f par G_m^2 dans (1.22), on obtient que $\lim_{k\to\infty} P^kG_m^2(x) = \pi(G_m^2)$ puis, par passage à la moyenne de Cesaro, que $\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}P^kG_m^2(x) = \pi(G_m^2)$. Comme $G_m(y) = 0$ pour $m \geq |F(y)|$ et $G_m^2 \leq F^2$ avec $\pi(F^2) < \infty$, le théorème de convergence dominée assure que $\lim_{m\to\infty}\pi(G_m^2) = 0$ et donc que

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{E}_x\left[\left(\frac{M^F_{n-1}-M^{F_m}_{n-1}}{\sqrt{n}}\right)^2\right] = 0.$$

Montrons d'autre part que $\lim_{m\to\infty}\sigma^2_{F_m} = \sigma^2(f)$. Comme $F^2 \in \mathcal{V}$, $F \in \mathcal{V}$. La condition (D1), l'inégalité $|F_m| \leq |F|$ ainsi que la convergence de F_m vers F assurent que pour tout $y \in \mathbb{E}$, $PF_m(y)$ converge vers $PF(y)$ par convergence dominée lorsque $m \to \infty$. Comme $(PF_m)^2 \leq P(F_m)^2 \leq PF^2$ et que $\pi(PF^2) = \pi(F^2) < \infty$, le théorème de convergence dominée assure que $\lim_{m\to\infty}\pi((PF_m)^2) = \pi((PF)^2)$. De même $\lim_{m\to\infty}\pi(F_m^2) = \pi(F^2)$ et donc $\lim_{m\to\infty}\sigma^2_{F_m} = \sigma^2(f)$.

Pour $u \in \mathbb{R}^*$, comme $\mathbb{R} \ni x \to e^{iux}$ est lipschitzienne de constante $|u|$,

$$\begin{aligned}&\left|\mathbb{E}_x\left[e^{iu\frac{M^F_{n-1}}{\sqrt{n}}}\right] - e^{-\frac{\sigma_F^2u^2}{2}}\right|\\ &\leq \left|\mathbb{E}_x\left[e^{iu\frac{M^F_{n-1}}{\sqrt{n}}} - e^{iu\frac{M^{F_m}_{n-1}}{\sqrt{n}}}\right]\right| + \left|\mathbb{E}_x\left[e^{iu\frac{M^{F_m}_{n-1}}{\sqrt{n}}}\right] - e^{-\frac{\sigma_{F_m}^2u^2}{2}}\right| + \left|e^{-\frac{\sigma_{F_m}^2u^2}{2}} - e^{-\frac{\sigma_F^2u^2}{2}}\right|\\ &\leq |u|\mathbb{E}_x^{1/2}\left[\left(\frac{M^F_{n-1}-M^{F_m}_{n-1}}{\sqrt{n}}\right)^2\right] + \left|\mathbb{E}_x\left[e^{iu\frac{M^{F_m}_{n-1}}{\sqrt{n}}}\right] - e^{-\frac{\sigma_{F_m}^2u^2}{2}}\right| + \left|e^{-\frac{\sigma_{F_m}^2u^2}{2}} - e^{-\frac{\sigma_F^2u^2}{2}}\right|.\end{aligned} \tag{1.26}$$

Pour $\varepsilon > 0$ fixé, on peut choisir m_ε tel que $\left|e^{-\frac{\sigma^2_{F_{m_\varepsilon}}u^2}{2}} - e^{-\frac{\sigma^2_F u^2}{2}}\right| \leq \frac{\varepsilon}{3}$ et $\limsup_{n\to\infty} \mathbb{E}_x^{1/2}\left[\left(\frac{M^F_{n-1}-M^{F_{m_\varepsilon}}_{n-1}}{\sqrt{n}}\right)^2\right] \leq \frac{\varepsilon}{6|u|}$. Alors il existe n^1_ε tel que pour $n \geq n^1_\varepsilon$, $\mathbb{E}_x^{1/2}\left[\left(\frac{M^F_{n-1}-M^{F_{m_\varepsilon}}_{n-1}}{\sqrt{n}}\right)^2\right] \leq \frac{\varepsilon}{3|u|}$. Par ailleurs, la première étape de la démonstration, appliquée à la fonction bornée F_{m_ε}, assure l'existence de n^2_ε t.q. $\forall n \geq n^2_\varepsilon$, $\left|\mathbb{E}_x\left[e^{iu\frac{M^{F_{m_\varepsilon}}_{n-1}}{\sqrt{n}}}\right] - e^{-\frac{\sigma^2_{F_{m_\varepsilon}}u^2}{2}}\right| \leq \frac{\varepsilon}{3}$. En choisissant $m = m_\varepsilon$ dans (1.26), on conclut que

$$\forall n \geq n^1_\varepsilon \vee n^2_\varepsilon,\ \left|\mathbb{E}_x\left[e^{iu\frac{M^F_{n-1}}{\sqrt{n}}}\right] - e^{-\frac{\sigma^2_F u^2}{2}}\right| \leq \varepsilon,$$

ce qui achève la démonstration.

□

1.6 Comparaison des variances asymptotiques

Nous allons donner une condition suffisante due Tierney [73] permettant de comparer les variances asymptotiques $\sigma^2_0(f) = \pi(F_0^2) - \pi((P_0F_0)^2)$ et $\sigma^2_1(f) = \pi(F_1^2) - \pi((P_1F_1)^2)$ associées à deux noyaux markoviens P_0 et P_1 dans le théorème 1.5.6.

Proposition 1.6.1. *Supposons que la probabilité π est réversible pour les deux noyaux markoviens P_0 et P_1 et que $\pi(g(P_1 - P_0)g) \geq 0$ pour toute fonction $g : E \to \mathbb{R}$ mesurable vérifiant $\pi(g^2) < \infty$. Supposons également que*

$$\exists V : E \to \mathbb{R}_+ \text{ mesurable}, \exists K \in \mathbb{R}_+,\ \exists \gamma \in]0,1[,\ \forall x \in E,\ P_0V \vee P_1V(x) \leq \gamma V(x) + K,$$
$$\exists R > \frac{4K}{(1-\sqrt{\gamma})^2},\ \exists \alpha \in]0,1], \forall x, y \text{ t.q. } V(x) + V(y) \leq R,$$
$$(P_0(x,.) \wedge P_0(y,.)(E)) \wedge (P_1(x,.) \wedge P_1(y,.)(E)) \geq \alpha.$$

Alors $\sigma^2_0(f) \leq \sigma^2_1(f)$ pour toute fonction $f : E \to \mathbb{R}$ mesurable et t.q. $\sup_{x\in E} \frac{|f(x)|}{1+\sqrt{V(x)}} < \infty$.

Remarque 1.6.2. *Pour $g : E \to \mathbb{R}$ mesurable et telle que $\pi(g^2) < \infty$ et P un noyau markovien admettant π comme probabilité invariante, $\pi(P|g|) = \pi(|g|) \leq \sqrt{\pi(g^2)}$ ce qui assure que $Pg(x)$ est bien défini $\pi(dx)$ p.p.. Par l'inégalité de Cauchy-Schwarz, celle de Jensen qui assure que $(Pg)^2 \leq Pg^2$ puis l'invariance de π par P,*

$$(\pi(|gPg|))^2 \leq \pi(g^2)\pi((Pg)^2) \leq \pi(g^2)\pi(Pg^2) = \left(\pi(g^2)\right)^2 < \infty.$$

Comme $|g(P_1 - P_0)g| \leq |gP_1g| + |gP_0g|$, on en déduit que $\pi(|g(P_1 - P_0)g|) < \infty$.

Le lemme suivant donne une condition suffisante pour que l'hypothèse de la proposition soit satisfaite.

Lemme 1.6.3. *Soient P_0 et P_1 deux noyaux markoviens admettant π comme probabilité invariante. Si pour tout $A \in \mathcal{E}$ et tout $x \in E \setminus A$, $P_0(x, A) \geq P_1(x, A)$, alors $\pi(g(P_1 - P_0)g) \geq 0$ pour toute fonction $g : E \to \mathbb{R}$ mesurable vérifiant $\pi(g^2) < \infty$.*

Démonstration du lemme 1.6.3 : Soit $P(x, dy) = (\delta_x(dy)+P_0(x, dy)-P_1(x, dy))$. Pour $x \in E$ et $A \in \mathcal{E}$, $\delta_x(A) + P_0(x, A) - P_1(x, A)$ est égal à $P_0(x, A) - P_1(x, A) \geq 0$ si $x \notin A$ et à $1-P_1(x, A)+P_0(x, A) \geq P_0(x, A) \geq 0$ si $x \in A$. Comme $\delta_x(E)+P_0(x, E)-P_1(x, E) = 1$ on en déduit que P est un noyau markovien. Comme la probabilité π est invariante par les noyaux $\delta_x(dy)$, P_0 et P_1, elle est invariante par P, ce qui implique que $\int_E g^2(x)\pi(dx) = \frac{1}{2}\int_{E\times E}(g^2(x) + g^2(y))\pi(dx)P(x, dy)$. Ainsi

$$\begin{aligned}
\pi(g(P_1g - P_0g)) &= \int_{E\times E} g(x)g(y)\pi(dx)(P_1(x, dy) - P_0(x, dy)) \\
&= \int_{E\times E} g(x)g(y)\pi(dx)(\delta_x(dy) - P(x, dy)) \\
&= \int_E g^2(x)\pi(dx) - \int_{E\times E} g(x)g(y)\pi(dx)P(x, dy) \\
&= \frac{1}{2}\int_{E\times E} (g^2(x) + g^2(y) - 2g(x)g(y))\pi(dx)P(x, dy) \\
&= \frac{1}{2}\int_{E\times E} (g(x) - g(y))^2\pi(dx)P(x, dy) \geq 0.
\end{aligned}$$

□

Pour un espace d'état E fini, Peskun [65] a montré que $\sigma_0^2(f) \leq \sigma_1^2(f)$ lorsque π est réversible par rapport aux deux noyaux P_0 et P_1 qui vérifient la condition de comparaison en hypothèse de ce lemme. Cette condition permet de vérifier l'optimalité en terme de variance asymptotique du choix de Metropolis-Hastings pour la probabilité d'accepter la proposition dans l'algorithme du même nom.

Example 1.6.4. *Pour l'algorithme de Metropolis-Hastings décrit au paragraphe 1.2, soit*

$$P_1(x, dy) = \alpha(x, y)q(x, y)\lambda(dy) + \left(\int_E (1 - \alpha(x, z))q(x, z)\lambda(dz)\right)\delta_x(dy),$$

avec

$$\alpha(x, y) = \begin{cases} a\left(\frac{\eta(y)q(y,x)}{\eta(x)q(x,y)}\right) & si\ \eta(x)q(x, y) > 0 \\ 1\ si\ \eta(x)q(x, y) = 0 \end{cases},$$

pour une fonction $a : \mathbb{R}_+ \to [0, 1]$ vérifiant $a(0) = 0$ et $\forall u > 0,\ a(u) = ua(1/u)$. On note P_0 et α_0 le noyau et la probabilité d'acceptation obtenus pour le choix $a_0(u) = \min(u, 1)$ de Metropolis-Hastings. Pour $u > 0$, comme a est à valeurs dans $[0, 1]$, $a(u) \leq 1$ et $a(u) = ua(1/u) \leq u$ si bien que $a(u) \leq a_0(u)$ pour tout $u \geq 0$. Ainsi $\alpha(x, y) \leq \alpha_0(x, y)$ pour tous $x, y \in E$ et pour tout $A \in \mathcal{E}$ t.q. $x \notin A$,

$$P_1(x, A) = \int_A \alpha(x, y)q(x, y)\lambda(dy) \leq \int_A \alpha_0(x, y)q(x, y)\lambda(dy) = P_0(x, A).$$

Comme $\pi(dx) = \frac{\eta(x)\lambda(dx)}{\int_E \eta(y)\lambda(dy)}$ est réversible pour P_0 et P_1, la proposition et le lemme assurent, sous réserve que les conditions de dérive en hypothèse de la proposition soient *satisfaites, que la variance asymptotique correspondant au choix de Metropolis-Hastings est minimale parmi tous les choix de fonction $a : \mathbb{R}_+ \to [0,1]$ vérifiant $a(0) = 0$ et $\forall u > 0,\ a(u) = ua(1/u)$.*

Démontrons maintenant la proposition 1.6.1.

Démonstration : Soit $f : E \to \mathbb{R}$ mesurable et t.q. $\sup_{x\in E} \frac{|f(x)|}{1+\sqrt{V(x)}} < \infty$. Comme les variances asymptotiques pour les fonctions f et $f - \pi(f)$ sont égales, nous supposerons que $\pi(f) = 0$ sans que cela ne soit restrictif.

Pour $t \in [0,1]$, notons $P_t = tP_1 + (1-t)P_0$. Comme $P_tV \leq P_0V \vee P_1V$ et $P_t(x,.) \wedge P_t(y,.) \geq t(P_1(x,.) \wedge P_1(y,.)) + (1-t)(P_0(x,.) \wedge P_0(y,.))$, les hypothèses du théorème 1.3.27, du corollaire 1.5.3 et du théorème 1.5.6 sont satisfaites pour le noyau P_t. La probabilité π étant réversible pour P_0 et P_1, elle l'est également pour P_t et c'est l'unique probabilité invariante pour ce noyau d'après le théorème 1.3.27. D'après le théorème 1.5.6, la variance asymptotique de l'estimateur ergodique de $\pi(f)$ sous le noyau P_t est $\sigma_t^2(f) = \pi(F_t^2) - \pi((P_tF_t)^2)$ où $F_t = \sum_{n\in\mathbb{N}} P_t^n f$ est solution de l'équation de Poisson

$$F_t - P_tF_t = f. \tag{1.27}$$

Une relecture attentive des preuves du théorème 1.3.27 et de la proposition 1.5.1 montre que (1.24) se généralise en : $\forall\beta \in]0, \frac{\alpha}{\sqrt{K}}[,\ \exists\chi \in]0,1[,\ \forall n \in \mathbb{N},\ \forall t_1,\dots,t_n \in [0,1]$,

$$\sup_{x\in E} \frac{|P_{t_1}P_{t_2}\dots P_{t_n}f(x)|}{1+\sqrt{V(x)}} \leq \max\left(2 + \frac{\beta\sqrt{K}}{1-\sqrt{\gamma}}, \beta\right)\chi^n \sup_{x\in E} \frac{|f(x)|}{1+\beta\sqrt{V(x)}}. \tag{1.28}$$

Choisissons désormais $\beta = \frac{\alpha}{2\sqrt{K}}$ et posons $C = \max\left(2 + \frac{\beta\sqrt{K}}{1-\sqrt{\gamma}}, \beta\right) \sup_{x\in E} \frac{|f(x)|}{1+\beta\sqrt{V(x)}} < \infty$.

Nous allons vérifier que $t \mapsto \sigma_t^2(f)$ est dérivable sur $[0,1]$ de dérivée $\frac{d\sigma_t^2}{dt}(f) = 2\pi(F_t(P_1F_t - P_0F_t))$. Comme l'hypothèse appliquée à $g = F_t$ assure que cette dérivée est positive, cela entraîne que $\sigma_1^2(f) \geq \sigma_0^2(f)$.

Pour $g : E \to \mathbb{R}$ mesurable telle que $\sup_{y\in E} \frac{|g(y)|}{1+\sqrt{V(y)}} < \infty$ et $x \in E$, $[0,1] \ni t \mapsto P_t^n g(x)$ est un polynôme de degré n donc une fonction C^∞. Pour identifier $\frac{d}{dt}P_t^n f(x)$ remarquons que pour $t \in [0,1]$ et $h \neq 0$ tel que $t+h \in [0,1]$,

$$\frac{1}{h}(P_{t+h}^n f(x) - P_t^n f(x)) = \sum_{m=0}^{n-1} P_{t+h}^m \frac{P_{t+h} - P_t}{h} P_t^{n-1-m} f(x) = \sum_{m=0}^{n-1} P_{t+h}^m (P_1 - P_0) P_t^{n-1-m} f(x)$$

converge vers $\frac{d}{dt}P_t^n f(x) = \sum_{m=0}^{n-1} P_t^m (P_1 - P_0)P_t^{n-1-m} f(x)$ lorsque $h \to 0$ puisque $\sup_{y\in E} \frac{|(P_1-P_0)P_t^{n-1-m}f(y)|}{1+\sqrt{V(y)}} \leq \sup_{y\in E} \frac{|P_1P_t^{n-1-m}f(y)|}{1+\sqrt{V(y)}} + \sup_{y\in E} \frac{|P_0P_t^{n-1-m}f(y)|}{1+\sqrt{V(y)}} \leq 2C\chi^{n-m}$, d'après (1.28). En outre, de manière analogue, $\sup_{t\in[0,1],y\in E} \frac{|\frac{d}{dt}P_t^n f(y)|}{1+\sqrt{V(y)}} \leq 2nC\chi^n$ où le second membre est sommable sur $n \in \mathbb{N}$. Le théorème des accroissements finis et le théorème de convergence dominée entraînent alors que $t \mapsto F_t(x) = \sum_{n\in\mathbb{N}} P_t^n f(x)$ est dérivable sur $[0,1]$ de dérivée $\frac{dF_t(x)}{dt} = \sum_{n\in\mathbb{N}} \frac{d}{dt}P_t^n f(x) = \sum_{n\in\mathbb{N}^*}\sum_{m=0}^{n-1} P_t^m(P_1 - P_0)P_t^{n-1-m}f(x)$ avec $\sup_{t\in[0,1],y\in E} \frac{|\frac{dF_t(y)}{dt}|}{1+\sqrt{V(y)}} \leq 2C\chi\sum_{n\in\mathbb{N}^*} n\chi^{n-1} = \frac{2C\chi}{(1-\chi)^2}$. Avec l'hypothèse de croissance faite sur f, cela implique que $\sup_{t\in[0,1],y\in E} \frac{|f(y)\frac{dF_t(y)}{dt}|}{1+V(y)} < \infty$. Comme $\pi(1+V) < \infty$, on en déduit que $\pi(\sup_{t\in[0,1]}|f\frac{dF_t(y)}{dt}|) < \infty$. D'après l'équation de Poisson,

$$F_t^2 - (P_tF_t)^2 = (F_t - P_tF_t)(2F_t - (F_t - P_tF_t)) = f(2F_t - f),$$

si bien que $\sigma_t^2(f) = \pi(f(2F_t - f))$. Une nouvelle application du théorème des accroissements finis et du théorème de convergence dominée assure que $t \mapsto \sigma_t^2(f)$ est dérivable sur $[0,1]$ de dérivée

$$\frac{d\sigma_t^2(f)}{dt} = 2\pi\left(f\frac{dF_t}{dt}\right) = 2\pi\left((F_t - P_tF_t)\frac{dF_t}{dt}\right).$$

La réversibilité de π pour P_t s'écrit $\pi(dx)P_t(x,dy) = \pi(dy)P_t(y,dx)$ et implique que

$$\begin{aligned}\pi\left(\frac{dF_t}{dt}P_tF_t\right) &= \int_{E\times E} F_t(y)\frac{dF_t}{dt}(x)\pi(dx)P_t(x,dy) = \int_{E\times E} F_t(y)\frac{dF_t}{dt}(x)\pi(dy)P_t(y,dx)\\ &= \pi\left(F_tP_t\frac{dF_t}{dt}\right),\end{aligned}$$

si bien que $\frac{d\sigma_t^2(f)}{dt} = 2\pi\left(F_t\left(\frac{dF_t}{dt} - P_t\frac{dF_t}{dt}\right)\right)$. Par ailleurs, en dérivant formellement l'équation de Poisson (1.27), on obtient que $\frac{dF_t}{dt} - P_t\frac{dF_t}{dt}$ doit être égal à $\frac{dP_t}{dt}F_t = (P_1 - P_0)F_t$. Pour le montrer rigoureusement, on écrit

$$\begin{aligned}P_t\frac{dF_t}{dt} &= \sum_{n\in\mathbb{N}^*}\sum_{m=0}^{n-1} P_t^{m+1}(P_1 - P_0)P_t^{n-1-m}f = \sum_{n\in\mathbb{N}^*}\sum_{m=1}^{n} P_t^{m}(P_1 - P_0)P_t^{n-m}f\\ &= \sum_{n\geq 2}\sum_{m=1}^{n-1} P_t^{m}(P_1 - P_0)P_t^{n-1-m}f = \sum_{n\geq 2}\sum_{m=0}^{n-1} P_t^{m}(P_1 - P_0)P_t^{n-1-m}f - \sum_{n\geq 2}(P_1 - P_0)P_t^{n-1}f\\ \text{et } \frac{dF_t}{dt} - P_t\frac{dF_t}{dt} &= (P_1 - P_0)f + \sum_{n\geq 2}(P_1 - P_0)P_t^{n-1}f = (P_1 - P_0)\sum_{n\in\mathbb{N}^*} P_t^{n-1}f\\ &= (P_1 - P_0)F_t.\end{aligned}$$

On conclut que $\frac{d\sigma_t^2(f)}{dt} = 2\pi(F_t(P_1 - P_0)F_t)$. □

Chapitre 2

Systèmes de particules en interaction

Lorsque l'on veut calculer $\mathbb{E}[f_n(X_{0:n})]$ où $X_{0:n} = (X_0, \dots, X_n)$ avec $(X_k)_{k\in\mathbb{N}}$ chaîne de Markov à valeurs dans E issue de X_0 distribuée suivant la probabilité η_0 et $f_n : E^{n+1} \to \mathbb{R}$ telle que $\mathbb{P}(f_n(X_{0:n}) \neq 0)$ est très faible (événement rare), alors la précision de l'estimateur de Monte-Carlo standard va être mauvaise. L'échantillonnage préférentiel consiste à effectuer un changement de probabilité de manière à privilégier les trajectoires de la chaîne de Markov qui se dirigent vers les zones où f_n est non nulle. À cet effet, on peut introduire des fonctions $g_p : E^{p+1} \mapsto]0, +\infty[$ bornées et définir une nouvelle probabilité $\mathbb{Q}_n$ par $\frac{d\mathbb{Q}_n}{d\mathbb{P}} = \frac{\prod_{p=0}^{n-1} g_p(X_{0:p})}{\mathbb{E}\left[\prod_{p=0}^{n-1} g_p(X_{0:p})\right]}$.

Exemple 2.0.1. *Si $E = \mathbb{R}$ et que $f_n(x_{0:n}) = \varphi(x_n)$ avec $\varphi(x)$ non nulle seulement pour de très grandes valeurs de x, on peut choisir $g_p(x_{0:p}) = \psi(x_p)$ avec ψ strictement positive et croissante pour privilégier les trajectoires qui sont déjà hautes à l'instant p ou bien $g_p(x_{0:p}) = \psi(x_p - x_{p-1})$ pour privilégier les trajectoires qui montent entre l'instant $p-1$ et l'instant p supposé supérieur ou égal à 1.*

Alors on a

$$\mathbb{E}[f_n(X_{0:n})] = \mathbb{E}^{\mathbb{Q}_n}\left[\tilde{f}_n(X_{0:n})\right]\mathbb{E}\left[\prod_{p=0}^{n-1} g_p(X_{0:p})\right] \quad \text{où} \quad \tilde{f}_n(x_{0:n}) := f_n(x_{0:n})\prod_{p=0}^{n-1} g_p^{-1}(x_{0:p}). \tag{2.1}$$

Pour implémenter l'échantillonnage préférentiel correspondant au changement de $\mathbb{P}$ à $\mathbb{Q}_n$, il faut à la fois être capable de simuler $(X_k)_{0\leq k\leq n}$ sous la probabilité $\mathbb{Q}_n$ et de calculer la constante de normalisation $\mathbb{E}[\prod_{p=0}^{n-1} g_p(X_{0:p})]$. Notons que cette constante de normalisation est une espérance du même type que celle de départ mais portant sur la trajectoire $X_{0:n-1}$ raccourcie d'une unité de temps, ce qui suggère d'itérer la transformation effectuée sur l'espérance de départ. Posons $\boxed{\gamma_0 = \eta_0}$ et pour $p \in \{1, \dots, n\}$, introduisons la mesure positive γ_p sur E^{p+1} définie par

$$\forall h_p : E^{p+1} \to \mathbb{R} \text{ mesurable bornée }, \boxed{\gamma_p(h_p) = \mathbb{E}\left[h_p(X_{0:p})\prod_{k=0}^{p-1} g_k(X_{0:k})\right]}$$

ainsi que la probabilité η_p sur E^{p+1} obtenue par normalisation de γ_p :

$$\boxed{\eta_p(h_p) = \frac{\gamma_p(h_p)}{\gamma_p(1)} = \frac{\mathbb{E}\left[h_p(X_{0:p})\prod_{k=0}^{p-1} g_k(X_{0:k})\right]}{\mathbb{E}\left[\prod_{k=0}^{p-1} g_k(X_{0:k})\right]}.} \tag{2.2}$$

En particulier $\mathbb{E}[f_n(X_{0:n})] = \mathbb{E}\left[\tilde{f}_n(X_{0:n})\prod_{k=0}^{n-1} g_k(X_{0:k})\right] = \gamma_n(\tilde{f}_n)$ et $\mathbb{E}^{\mathbb{Q}_n}[h_n(X_{0:n})] = \eta_n(h_n)$.

Comme $\gamma_p(1) = \gamma_{p-1}(g_{p-1}) = \eta_{p-1}(g_{p-1})\gamma_{p-1}(1)$ et $\gamma_0(1) = 1$, par récurrence sur p on obtient que

$$\text{pour } p \geq 1, \ \boxed{\gamma_p(1) = \prod_{k=0}^{p-1} \eta_k(g_k).}$$

On en déduit que

$$\mathbb{E}[f_n(X_{0:n})] = \gamma_n(\tilde{f}_n) = \eta_n(\tilde{f}_n)\gamma_n(1) = \eta_n(\tilde{f}_n)\prod_{k=0}^{n-1} \eta_k(g_k). \tag{2.3}$$

Plus généralement, pour tout $p \in \{1, \ldots, n-1\}$,

$$\boxed{\gamma_p(h_p) = \eta_p(h_p)\gamma_p(1) = \eta_p(h_p)\prod_{k=0}^{p-1} \eta_k(g_k).} \tag{2.4}$$

L'exercice suivant fournit un example où remplacer un calcul d'espérance par un calcul de produit d'espérances comme dans la formule (2.3) conduit à une réduction de variance significative.

Exercice 2.0.2. *On souhaite estimer une probabilité de la forme $p = \mathbb{P}(X \in A)$ avec p très petite (X désigne par exemple le vecteur des paramètres de fonctionnement d'une centrale nucléaire et A l'ensemble des valeurs de ces paramètres conduisant à la fusion du cœur de la centrale). On se donne pour cela une suite $(X_i)_{i\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi de X et on note $\hat{p}_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \in A\}}$.*

1. *Quel est le comportement asymptotique de $\hat{p}_n$ pour $n \to +\infty$? Donner la variance de $\hat{p}_n$.*

On suppose maintenant que l'on sait simuler indépendamment des X_i une suite $(Y_i)_{i\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi conditionnelle de X sachant $X \in B$ où B contient A. En particulier $\mathbb{P}(Y_i \in A) = \mathbb{P}(X \in A | X \in B)$. On note $\hat{q}_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \in B\}}$ et $\hat{r}_n = \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i \in A\}}$.

2. *Quel est le comportement asymptotique de $\hat{q}_n \times \hat{r}_n$ pour $n \to +\infty$? Calculer l'espérance et la variance de $\hat{q}_n \times \hat{r}_n$. Expliquer pourquoi, lorsque $\mathbb{P}(X \in B)$ et $\mathbb{P}(X \in A|X \in B)$ sont du même ordre, $\hat{q}_n \times \hat{r}_n$ approche p beaucoup plus précisément que $\hat{p}_n$.*

3. *On suppose plus généralement que $p = \prod_{j=1}^k p_j$ et que les estimateurs indépendants $(\hat{p}_n^j)_{1\leq j\leq k}$ sont respectivement les moyennes empiriques de n variables aléatoires de Bernoulli de paramètre p_j indépendantes.*
   (a) *Montrer que $\prod_{j=1}^k \hat{p}_n^j$ converge presque sûrement vers p lorsque $n \to \infty$.*
   (b) *Vérifier que la variance asymptotique $\mathcal{V}_k(p_1, \ldots, p_k) = \lim_{n\to\infty} n\mathrm{Var}\left(\frac{\prod_{j=1}^k \hat{p}_n^j}{\prod_{j=1}^k p_j}\right)$ qui mesure la précision de cet estimateur est donnée par $\mathcal{V}_k(p_1, \ldots, p_k) = \sum_{j=1}^k \frac{1-p_j}{p_j} = \sum_{j=1}^k \frac{1}{p_j} - k$.*

(c) *Montrer que* $\frac{1}{k}\sum_{j=1}^{k}\frac{1}{p_j} \geq e^{-\frac{1}{k}\sum_{j=1}^{k}\ln(p_j)}$ *et en déduire que* $\mathcal{V}_k(p^{1/k},\dots,p^{1/k}) = \inf_{(p_1,\dots,p_k)\in(\mathbb{R}_+^*)^k:\prod_{j=1}^k p_j=p}\mathcal{V}_k(p_1,\dots,p_k)$.

(d) *Comparer* $\lim_{k\to\infty}\mathcal{V}_k(p^{1/k},\dots,p^{1/k})$ *à* $\mathcal{V}_1(p)$.

Le choix sous forme produit du changement de probabilité $\frac{d\mathbb{Q}_n}{d\mathbb{P}} = \frac{\prod_{p=0}^{n-1} g_p(X_{0:p})}{\mathbb{E}\left[\prod_{p=0}^{n-1} g_p(X_{0:p})\right]} = \frac{\prod_{p=0}^{n-1} g_p(X_{0:p})}{\prod_{p=0}^{n-1} \eta_p(g_p)}$ qui peut sembler étrange de prime abord, va permettre de construire dynamiquement une suite de mesures empiriques $(\eta_p^M = \frac{1}{M}\sum_{m=1}^M \delta_{X_{0:p}^{p,m}})_{0\leq p\leq n}$ qui approchent les probabilités $(\eta_p)_{0\leq p\leq n}$ dans la limite $M\to\infty$ d'un grand nombre de particules. Il n'y a pas besoin de savoir simuler $(X_k)_{0\leq k\leq n}$ sous $\mathbb{Q}_n$ pour bénéficier en partie de la réduction de variance par échantillonnage préférentiel. En suivant (2.3), on approchera $\mathbb{E}\left[f_n(X_{0:n})\right]$ par $\eta_n^M(\tilde{f}_n)\prod_{p=0}^{n-1}\eta_p^M(g_p)$.

Avant d'expliquer la construction des mesures empiriques η_p^M, montrons que les mesures η_p apparaissent également naturellement dans des problèmes de filtrage. On considère un signal (par exemple la position d'un aéronef dans le ciel) modélisé par une chaîne de Markov $(X_k)_{k\in\mathbb{N}}$ avec X_0 distribuée suivant la probabilité η_0. On n'observe pas directement le signal mais des informations $Y_k = \varphi_k(X_k) + V_k$ (écho radar par exemple) où $\varphi_k : E\to\mathbb{R}^d$ et les bruits d'observation $(V_k)_{k\in\mathbb{N}}$ sont supposés indépendants à valeurs $\mathbb{R}^d$ de densités respectives $(q_k)_{k\in\mathbb{N}}$ par rapport à la mesure de Lebesgue et indépendants de la chaîne de Markov $(X_k)_{k\in\mathbb{N}}$. La densité conditionnelle de $Y_{0:p}$ sachant $X_{0:p+1} = x_{0:p+1}$ est $\prod_{k=0}^p q_k(y_k-\varphi_k(x_k))$. La formule de Bayes assure alors que

$$\mathbb{P}(X_{0:p+1}\in dx_{0:p+1}|Y_{0:p}=y_{0:p}) \propto \prod_{k=0}^p q_k(y_k-\varphi_k(x_k))\mathbb{P}(X_{0:p+1}\in dx_{0:p+1}).$$

Ainsi, pour le choix $g_k(x_{0:k}) = q_k(y_k-\varphi_k(x_k))$ pour tout $k\in\{0,\dots,p\}$, la loi conditionnelle de $X_{0:p+1}$ sachant $Y_{0:p}=y_{0:p}$ est η_{p+1}. La loi conditionnelle $\hat{\eta}_p$ de $X_{0:p}$ sachant $Y_{0:p}=y_{0:p}$ s'obtient par la formule de la loi marginale :

$$\hat{\eta}_p(h_p) = \int_{E^{p+2}} h_p(x_{0:p})\eta_{p+1}(dx_{0:p+1}) = \frac{\mathbb{E}[h_p(X_{0:p})\prod_{k=0}^p g_k(X_{0:k})]}{\gamma_{p+1}(1)} = \frac{\gamma_p(h_pg_p)}{\gamma_p(g_p)} = \frac{\eta_p(h_pg_p)}{\eta_p(g_p)}.$$

La construction des probabilités $(\eta_p^M)_{0\leq p\leq n}$ se fait par récurrence sur p. La chaîne de Markov $(X_k)_{k\geq 0}$ est supposée inhomogène au sens où, pour $h: E\to\mathbb{R}$ mesurable bornée, $\mathbb{E}[h(X_k)|X_{0:k-1}] = P_kh(X_{k-1})$ pour un noyau Markovien P_k pouvant dépendre de l'instant k. Notons que le cas où l'espace E_k où X_k prend ses valeurs dépend de k peut rentrer dans le cadre introduit ici en posant $E=\bigcup_{k\in\mathbb{N}}E_k$.

Initialisation : On génère un échantillon $(X_0^{0,m})_{1\leq m\leq M}$ i.i.d. suivant la loi η_0 de X_0.

Passage de $\eta_p^M = \frac{1}{M}\sum_{m=1}^M\delta_{X_{0:p}^{p,m}}$ **à** $\eta_{p+1}^M = \frac{1}{M}\sum_{m=1}^M\delta_{X_{0:p+1}^{p+1,m}}$ **:** pour $p\in\{0,\dots,n-1\}$ ce passage s'effectue en deux étapes :

Sélection : on génère des vecteurs aléatoires $(X_{0:p}^{p+1,m})_{1\leq m\leq M}$ conditionnellement indépendants sachant $\mathcal{F}_p = \sigma((X_{0:k}^{k,m})_{1\leq m\leq M, 0\leq k\leq p})$ et vérifiant

$$\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^M\delta_{X_{0:p}^{p+1,m}}\middle|\mathcal{F}_p\right] = \frac{1}{M\eta_p^M(g_p)}\sum_{m=1}^M g_p(X_{0:p}^{p,m})\delta_{X_{0:p}^{p,m}}. \tag{2.5}$$

Après avoir décrit la seconde étape, nous allons revenir sur plusieurs approches permettant d'effectuer cette étape de sélection. Notons que $\hat{\eta}_p^M = \frac{1}{M}\sum_{m=1}^M \delta_{X_{0:p}^{p+1,m}}$ (égale à la loi marginale des $p+1$ premières coordonnées sous η_{p+1}^M) constitue une approximation de la mesure $\hat{\eta}_p$ introduite dans le problème de filtrage présenté plus haut.

Mutation : sachant $\mathcal{G}_{p+1} = \sigma((X_{0:k}^{k,m})_{1\leq m\leq M, 0\leq k\leq p}, (X_{0:p}^{p+1,m})_{1\leq m\leq M})$, on génère des variables aléatoires $(X_{p+1}^{p+1,m})_{1\leq m\leq M}$ conditionnellement indépendantes et respectivement distribuées suivant $P_{p+1}(X_p^{p+1,m}, .)$. Pour tout $m \in \{1,\ldots,M\}$, on pose $X_{0:p+1}^{p+1,m} = (X_{0:p}^{p+1,m}, X_{p+1}^{p+1,m})$ et $\eta_{p+1}^M = \frac{1}{M}\sum_{m=1}^M \delta_{X_{0:p+1}^{p+1,m}}$.

Intuitivement, l'étape de sélection permet de se débarrasser du poids multiplicatif individuel $g_p(X_{0:p}^{p,m})$ affecté à chaque trajectoire entre l'instant p et l'instant $p+1$ en répliquant les trajectoires pour lesquelles ce poids est élevé et en supprimant celles pour lesquelles ce poids est faible. L'étape de mutation consiste juste à ajouter une nouvelle position tirée suivant le noyau markovien P_{p+1} à chaque trajectoire issue de l'étape de sélection.

Pour l'étape de sélection, partant de M valeurs $(y^m)_{1\leq m\leq M}$ (les trajectoires $(X_{0:p}^{p,m})_{1\leq m\leq M}$) dans un espace d'états $\mathcal{Y}$ quelconque (l'espace produit E^{p+1}) et d'une probabilité $(p^m)_{1\leq m\leq M}$ $((g_p(X_{0:p}^{p,m})/\sum_{\ell=1}^M g_p(X_{0:p}^{p,l}))_{1\leq m\leq M})$, il faut être capable de générer des variables aléatoires indépendantes $(Y^m)_{1\leq m\leq M}$ à valeurs dans $\{y^1,\ldots,y^M\}$,

$$\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^M \delta_{Y^m}\right] = \sum_{m=1}^M p^m \delta_{y^m}, \tag{2.6}$$

i.e pour toute fonction $f : \mathcal{Y} \to \mathbb{R}$ mesurable bornée, $\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^M f(Y^m)\right] = \sum_{m=1}^M p^m f(y^m)$. Notons que, lorsque les y^ℓ sont distincts, $\delta_{Y^m} = \sum_{\ell=1}^M 1_{\{Y^m=y^\ell\}}\delta_{y^\ell}$ et (2.6) est équivalent à $\mathbb{E}\left[\sum_{m=1}^M 1_{\{Y^m=y^\ell\}}\right] = Mp^\ell$ pour tout $\ell \in \{1,\ldots,M\}$. Sans supposer les y^ℓ distincts, (2.6) est équivalent à $\mathbb{E}\left[\sum_{m=1}^M 1_{\{Y^m=y^\ell\}}\right] = M\sum_{j=1}^M 1_{\{y^j=y^\ell\}}p^j$ pour tout $\ell \in \{1,\ldots,M\}$. Il y a bien sûr de multiples façons de générer de telles variables aléatoires :

Échantillonnage multinomial : on génére les Y^m i.i.d. suivant la probabilité $\sum_{\ell=1}^M p^\ell \delta_{y^\ell}$ $(\forall \ell, m \in \{1,\ldots,M\},\ \mathbb{P}(Y^m = y^\ell) = \sum_{j=1}^M 1_{\{y^j=y^\ell\}}p^j)$. Notons que $\mathbb{E}[1_{\{Y^m=y^\ell\}}] = \sum_{j=1}^M 1_{\{y^j=y^\ell\}}p^j$ pour tous $m, \ell \in \{1,\ldots,M\}$.

Échantillonnage résiduel : Pour $x \in \mathbb{R}$, on note $\lfloor x \rfloor$ et $\lceil x \rceil$ les parties entières inférieure et supérieure de x, c'est-à-dire les entiers relatifs vérifiant $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$ et $\lceil x \rceil - 1 < x \leq \lceil x \rceil$ et $\{x\} = x - \lfloor x \rfloor \in [0,1[$ sa partie fractionnaire. L'échantillonnage résiduel consiste à choisir $Y^m = y^\ell$ pour $\sum_{j=1}^{\ell-1}\lfloor Mp^j \rfloor < m \leq \sum_{j=1}^{\ell}\lfloor Mp^j \rfloor$ (on réplique donc $\lfloor Mp^\ell \rfloor$ fois la valeur y^ℓ) et, lorsque $\sum_{j=1}^M \lfloor Mp^j \rfloor < M$ à générer les $r = M - \sum_{j=1}^M \lfloor Mp^j \rfloor = \sum_{j=1}^M \{Mp^j\}$ variables aléatoires $Y_{\sum_{j=1}^M \lfloor Mp^j \rfloor + 1}, \ldots, Y_M$ i.i.d. suivant $\frac{1}{M-\sum_{j=1}^M \lfloor Mp^j \rfloor}\sum_{\ell=1}^M \{Mp^\ell\}\delta_{y^\ell}$. Alors, si les y^ℓ sont distincts, pour tout $\ell \in \{1,\ldots,M\}$, $\sum_{m=1}^{M-r} 1_{\{Y^m=y^\ell\}} = \lfloor Mp^\ell \rfloor$ et, lorsque $r > 0$, pour tout $m \in \{M-r+1,\ldots,M\}$, $\mathbb{E}[1_{\{Y^m=y^\ell\}}] = \frac{\{Mp^\ell\}}{r}$.

Échantillonnage stratifié : à partir de M variables aléatoires $(U^m)_{1\leq m\leq M}$ i.i.d. suivant la loi uniforme sur $[0,1]$ on construit

$$Y^m = \sum_{\ell=1}^{M} 1_{\{\sum_{j=1}^{\ell-1} p^j < (m-U^m)/M \leq \sum_{j=1}^{\ell} p^j\}} y^\ell \text{ pour } m \in \{1,\ldots,M\}.$$

Échantillonnage systématique : on prend les variables U^m toutes égales à une seule variable U uniforme sur $[0,1]$ dans la formule précédente. Notons que si y^ℓ est distinct de y^j pour $j \neq \ell$, alors cette valeur est tirée $\lfloor M\sum_{j=1}^{\ell} p^j + U\rfloor - \lfloor M\sum_{j=1}^{\ell-1} p^j + U\rfloor$ fois.

Avec l'échantillonnage multinomial, l'échantillonnage résiduel et l'échantillonnage stratifié, les variables aléatoires Y^m sont indépendantes (pour l'échantillonnage résiduel, on pourra remarquer qu'une variable aléatoire constante est indépendant de toute variable aléatoire), alors qu'avec l'échantillonnage systématique, on perd cette propriété d'indépendance (sauf si pour tout $m \in \{1,\ldots,M\}$, Mp^m est entier, auquel cas les échantillonnages résiduel, stratifié et systématique coincident). C'est pourquoi seules les trois premières techniques d'échantillonnage sont éligibles pour l'étape de sélection de la méthode particulaire décrite plus haut.

Notons que lorsque les y^ℓ sont distincts, alors la mesure empirique $\frac{1}{M}\sum_{m=1}^{M}\delta_{Y^m}$ s'écrit $\frac{1}{M}\sum_{\ell=1}^{M}\sum_{m=1}^{M} 1_{\{Y^m=y^\ell\}}\delta_{y^\ell}$. L'objectif du problème suivant est de comparer les variances des variables aléatoires $\sum_{m=1}^{M} 1_{\{Y^m=y^\ell\}}$ pour les différentes technique d'échantillonnage.

Problème 2.0.3. *On suppose les y^ℓ distincts. Pour $\ell \in \{1,\ldots,M\}$, la variable aléatoire $\sum_{m=1}^{M} 1_{\{Y^m=y^\ell\}}$ est respectivement notée*

— *N_ℓ dans l'échantillonnage multinomial,*
— *$\tilde{N}_\ell$ dans l'échantillonnage résiduel,*
— *$\bar{N}_\ell$ dans l'échantillonnage stratifié,*
— *$\hat{N}_\ell$ dans l'échantillonnage systématique.*

Condition (2.6) **:** *Remarquer que $\frac{1}{M}\sum_{m=1}^{M} f(Y^m) = \sum_{\ell=1}^{M}\left(\frac{1}{M}\sum_{m=1}^{M} 1_{\{Y^m=y^\ell\}} f(y^\ell)\right)$ et en déduire que* (2.6) *est équivalente à $\forall \ell \in \{1,\ldots,M\}$, $\mathbb{E}\left[\sum_{m=1}^{M} 1_{\{Y^m=y^\ell\}}\right] = Mp^\ell$.*

Échantillonnage multinomial *:*

1. *Pour $\ell \in \{1,\ldots,M\}$ vérifier que $\mathbb{E}(N_\ell) = Mp^\ell$ et calculer $\mathrm{Var}(N_\ell)$.*
2. *Vérifier que le vecteur $(N_1,\ldots,N_M)$ suit la loi multinomiale de paramètre $(M,p^1,\ldots,p^M)$: $\forall (n_1,\ldots,n_M) \in \mathbb{N}^M$*

$$\mathbb{P}(N_1 = n_1,\ldots,N_M = n_M) = 1_{\{\sum_{m=1}^{M} n_m = M\}} M! \prod_{m=1}^{M} \frac{(p^m)^{n_m}}{n_m!}.$$

Les variables aléatoires N_m sont-elles indépendantes ?

Variance optimale : *Soit N une variable aléatoire à valeurs dans $\mathbb{Z}$ intégrable d'espérance ν. On note $x_+ = \max(x,0)$ et $x_- = \max(-x,0)$ pour $x \in \mathbb{R}$.*

1. *Vérifier que $\mathbb{E}[(N-\nu)_+] = \mathbb{E}[(N-\nu)_-]$. Montrer que $(N-\nu)_+ \geq (1-\{\nu\})1_{\{N>\nu\}}$ et $(N-\nu)_- \geq \{\nu\}1_{\{N\leq\nu\}}$.*
2. *En déduire que $\mathrm{Var}(N) = \mathbb{E}[(N-\nu)_+^2 + (N-\nu)_-^2] \geq \mathbb{E}[(N-\nu)_+] \geq \{\nu\}(1-\{\nu\})$ et exhiber une loi sous laquelle les deux dernières inégalités sont des égalités.*

Échantillonnage résiduel :

3. *Que se passe-t-il si* $r = 0$ *?*

On suppose désormais $r > 0$.

4. *Pour* $\ell \in \{1, \ldots, M\}$ *vérifier que* $\mathbb{E}(\tilde{N}_\ell) = Mp^\ell$ *et calculer* $\mathrm{Var}(\tilde{N}_\ell)$. *En déduire que la répartition résiduelle est optimale en termes de variance si* $r = 1$ *mais pas si* $r \geq 2$.
5. *Vérifier que* $r \leq M(1-p^\ell)+\{Mp^\ell\}$ *et en déduire que* $\mathrm{Var}(\tilde{N}_\ell) \leq \frac{M\{Mp^\ell\}(1-p^\ell)}{M(1-p^\ell)+\{Mp^\ell\}}$. *Conclure que* $\mathrm{Var}(\tilde{N}_\ell) \leq \mathrm{Var}(N_\ell)$. *Quel intérêt voyez-vous à l'échantillonnage résiduel par rapport à l'échantillonnage multinomial ?*

Échantillonnage stratifié :

6. *Vérifier que pour* $\ell, m \in \{1, \ldots, M\}$,
$$\mathbb{E}\left[1_{\{p^1+\ldots+p^{\ell-1} < \frac{m-U_m}{M} \leq p^1+\ldots+p^\ell\}}\right] = M \int_{\frac{m-1}{M}}^{\frac{m}{M}} 1_{\{p^1+\ldots+p^{\ell-1} < u \leq p^1+\ldots+p^\ell\}} du$$
et en déduire que $\mathbb{E}[\bar{N}_\ell] = Mp^\ell = \mathbb{E}[\hat{N}_\ell]$.
7. *Pour* x *et* y *deux réels tels que* $x < y$ *et* $j \in \mathbb{Z}$, *vérifier*
$$]x, y] \cap]j-1, j] = \begin{cases}]j-1, j] \text{ si et seulement si } \lceil x \rceil + 1 \leq j \leq \lfloor y \rfloor \\ \emptyset \text{ si } j \leq \lfloor x \rfloor \text{ ou } j \geq \lceil y \rceil + 1 \end{cases}.$$
En déduire que $\bar{N}_\ell$ *prend ses valeurs dans* $\{\lfloor n(p_1 + \ldots + p_\ell)\rfloor - \lceil n(p_1 + \ldots + p_{\ell-1})\rceil, \lfloor n(p_1+\ldots+p_\ell)\rfloor - \lfloor n(p_1+\ldots+p_{\ell-1})\rfloor, \lceil n(p_1+\ldots+p_\ell)\rceil - \lfloor n(p_1+\ldots+p_{\ell-1})\rfloor\}$. *Vérifier que pour* X *variable aléatoire à valeurs dans* $[a, b]$, $\mathrm{Var}(X) \leq \frac{(b-a)^2}{4}$ *et en déduire que* $\mathrm{Var}(\bar{N}_\ell) \leq 1$.
8. *Soit* x *et* y *deux réels. Vérifier que* $y \geq \lfloor y - x\rfloor + \lfloor x \rfloor$ *et* $x \geq \lfloor y \rfloor - \lceil y - x \rceil$. *En déduire que* $\lfloor y - x \rfloor \leq \lfloor y \rfloor - \lfloor x \rfloor \leq \lceil y - x \rceil \leq \lfloor y - x \rfloor + 1$. *Lorsque* $x < y$, *vérifier que* $]x, y] \cap \mathbb{N}$ *est égal à* $\emptyset$ *si* $\lfloor y \rfloor = \lfloor x \rfloor$ *et comprend les* $\lfloor y \rfloor - \lfloor x \rfloor$ *éléments* $\{\lfloor x \rfloor + 1, \ldots, \lfloor y \rfloor\}$ *sinon. En déduire que l'intervalle*
$$]M(p_1 + \ldots + p_{\ell-1}) + U, M(p_1 + \ldots + p_\ell) + U]$$
contient soit $\lfloor Mp_\ell \rfloor$ *soit* $\lfloor Mp_\ell \rfloor + 1$ *éléments de* $\{1, \ldots, M\}$. *En déduire que* $\hat{N}_\ell - \lfloor Mp_i \rfloor$ *suit une loi de Bernoulli de paramètre à préciser et conclure que* $\mathrm{Var}(\hat{N}_\ell) = \{Mp_\ell\}(1 - \{Mp_\ell\})$, *c'est-à-dire que la répartition systématique est optimale en termes de variance. Vérifier que ce résultat reste vrai pour* $\bar{N}_\ell$ *lorsque* $M(p^1 + \ldots + p^{\ell-1})$ *ou* $M(p^1 + \ldots + p^\ell)$ *sont entiers, propriété toujours satisfaite pour* $\ell \in \{1, M\}$.
9. *Vérifier que*
$$\mathrm{Var}(\bar{N}_\ell) = M\left(p^\ell - M\sum_{m=1}^{M}\left(\int_{\frac{m-1}{M}}^{\frac{m}{M}} 1_{\{p^1+\ldots+p^{\ell-1}<u\leq p^1+\ldots+p^\ell\}} du\right)^2\right)$$
et en déduire par l'inégalité de Cauchy-Schwarz que $\mathrm{Var}(\bar{N}_\ell) \leq \mathrm{Var}(N_\ell)$.

10. *À l'aide de la question 7, montrer que le cardinal de*

$$\{m \in \{1,\dots,M\} : \int_{\frac{m-1}{M}}^{\frac{m}{M}} 1_{\{p^1+\dots+p^{\ell-1}<u\leq p^1+\dots+p^\ell\}}du = 1/M\}$$

est égal soit à $\lfloor Mp^\ell \rfloor$ *soit à* $\lfloor Mp^\ell \rfloor - 1$. *En remarquant que dans le premier cas, si* $a,b \in [0,1/M]$ *avec* $a+b = \{Mp^\ell\}/M$, $a^2+b^2 \geq \frac{\{Mp^\ell\}^2}{2M^2}$ *et dans le second, si* $a,b \in [0,1/M]$ *avec* $a+b = (1+\{Mp^\ell\})/M$, $a^2+b^2 \geq \frac{(1+\{Mp^\ell\})^2}{2M^2}$, *en déduire que*

$$\sum_{m=1}^{M}\left(\int_{\frac{m-1}{M}}^{\frac{m}{M}} 1_{\{p^1+\dots+p^{\ell-1}<u\leq p^1+\dots+p^\ell\}}du\right)^2 \\ \geq \min\left(\frac{\lfloor Mp^\ell \rfloor + \{Mp^\ell\}^2/2}{M^2}, \frac{Mp^\ell + (\{Mp^\ell\}^2-1)/2}{M^2}\right).$$

Conclure que $\mathrm{Var}(\bar{N}_\ell) \leq \max\left(\{Mp^\ell\}(1-\frac{\{Mp^\ell\}}{2}), \frac{1-\{Mp^\ell\}^2}{2}\right)$.
Pour $M=3$ *et* $p^1 = \frac{1}{12}$, $p^2 = \frac{1}{2}$ *et* $p^3 = \frac{5}{12}$, *vérifier que* $\mathrm{Var}(\bar{N}_2) = \frac{3}{8} > \frac{1}{4} = \mathrm{Var}(\tilde{N}_2)$.

Exercice 2.0.4. *On suppose que* $\underline{p} := \min_{1\leq m\leq M} p^m > 0$ *(on peut restreindre l'ensemble des valeurs* $(y^m)_{1\leq m\leq M}$ *pour s'y ramener).*

1. *Montrer que* $M\underline{p} \leq 1$. *Que vaut* $(p^1,\dots,p^M)$ *en cas d'égalité ? Que donnent alors l'échantillonnage résiduel et l'échantillonnage stratifié ?*
2. *On suppose* $M\underline{p} < 1$, *on se donne des variables aléatoires* $(Z^m)_{1\leq m\leq M}$ *indépendantes et telles que* $\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^M \delta_{Z^m}\right] = \frac{1}{1-M\underline{p}}\sum_{m=1}^M (p^m - \underline{p})\delta_{y^m}$ *et indépendamment une suite* $(U^m)_{1\leq m\leq M}$ *de variables aléatoires indépendantes et uniformes sur* $[0,1]$.
   (a) *Vérifier que si pour* $m \in \{1,\dots,M\}$, $Y^m = 1_{\{U_m\leq M\underline{p}\}}y^m + 1_{\{U_m>M\underline{p}\}}Z^m$ *alors*

$$\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^M \delta_{Y^m}\right] = \sum_{m=1}^M p^m\delta_{y^m} \tag{2.7}$$

   et les variables aléatoires $(Y^m)_{1\leq m\leq M}$ *sont indépendantes.*
   (b) *Si* $(Y^1,\dots,Y^M) = 1_{\{U_1\leq M\underline{p}\}}(y^1,\dots,y^M) + 1_{\{U_1>M\underline{p}\}}(Z^1,\dots,Z^M)$, *vérifier que* (2.7) *reste vraie mais que les variables aléatoires* $(Y^m)_{1\leq m\leq M}$ *ne sont pas indépendantes.*

En suivant la formule (2.4), on introduit l'approximation suivante de la mesure non normalisée γ_p : $\gamma_0^M = \eta_0^M$ et

$$\text{pour } p \geq 1, \ \boxed{\gamma_p^M(h_p) = \eta_p^M(h_p)\prod_{k=0}^{p-1}\eta_k^M(g_k).} \tag{2.8}$$

Notons que $\eta_p^M(h_p) = \frac{\gamma_p^M(h_p)}{\gamma_p^M(1)}$ si bien que la probabilité empirique η_p^M s'obtient par normalisation de la mesure ponctuelle γ_p^M.

Proposition 2.0.5. *On suppose que pour tout $p \in \{0,\dots,n-1\}$, la fonction $g_p : E^{p+1} \to]0,+\infty[$ est mesurable et bornée. Pour tout $p \in \{0,\dots,n\}$ et toute fonction $h_p : E^{p+1} \to \mathbb{R}$ mesurable bornée,*

$$\mathbb{E}\left[\gamma_p^M(h_p)\right] = \gamma_p(h_p).$$

En particulier, si $\tilde{f}_n$ est bornée, l'estimateur $\eta_n^M(\tilde{f}_n)\prod_{p=0}^{n-1}\eta_p^M(g_p) = \gamma_n^M(\tilde{f}_n)$ de $\mathbb{E}\left[f_n(X_{0:n})\right]$ est sans biais.

Démonstration : On raisonne par récurrence sur p. On a $\gamma_0^M = \eta_0^M = \frac{1}{M}\sum_{m=1}^M \delta_{X_0^{0,m}}$ où les variables $(X_0^{0,m})_{1\leq m\leq M}$ sont toutes distribuées suivant $\eta_0 = \gamma_0$, ce qui assure que $\mathbb{E}[\gamma_0^M(h_0)] = \gamma_0(h_0)$. Ainsi, l'hypothèse de récurrence est satisfaite au rang $p = 0$. Supposons-la satisfaite au rang p. Rappelons que $\mathcal{F}_p = \sigma((X_{0:k}^{k,m})_{1\leq m\leq M, 0\leq k\leq p})$ et $\mathcal{G}_{p+1} = \sigma((X_{0:k}^{k,m})_{1\leq m\leq M, 0\leq k\leq p}, (X_{0:p}^{p+1,m})_{1\leq m\leq M})$. En notant $P_{p+1}h_{p+1}(x_{0:p}) = \int_E h_{p+1}(x_{0:p+1})P_{p+1}(x_p, dx_{p+1})$, on a d'après l'étape de mutation

$$\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{G}_{p+1}] = \frac{1}{M}\sum_{m=1}^M P_{p+1}h_{p+1}(X_{0:p}^{p+1,m}).$$

Ensuite comme $\mathcal{F}_p \subset \mathcal{G}_{p+1}$,

$$\begin{aligned}
\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p] = \mathbb{E}[\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{G}_{p+1}]|\mathcal{F}_p] &= \mathbb{E}\left[\frac{1}{M}\sum_{m=1}^M P_{p+1}h_{p+1}(X_{0:p}^{p+1,m})\middle|\mathcal{F}_p\right] \\
&= \frac{\eta_p^M(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)},
\end{aligned} \tag{2.9}$$

où on a utilisé (2.5) pour la dernière égalité. Comme $\prod_{k=0}^p \eta_k^M(g_k)$ est $\mathcal{F}_p$ mesurable, on en déduit que

$$\begin{aligned}
\mathbb{E}\left[\eta_{p+1}^M(h_{p+1})\prod_{k=0}^p \eta_k^M(g_k)\right] &= \mathbb{E}\left[\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p]\prod_{k=0}^p \eta_k^M(g_k)\right] \\
&= \mathbb{E}\left[\eta_p^M(g_pP_{p+1}h_{p+1})\prod_{k=0}^{p-1}\eta_k^M(g_k)\right] = \mathbb{E}\left[\gamma_p^M(g_pP_{p+1}h_{p+1})\right].
\end{aligned}$$

L'hypothèse de récurrence au rang p et la définition de γ_p assurent alors que

$$\mathbb{E}\left[\eta_{p+1}^M(h_{p+1})\prod_{k=0}^p \eta_k^M(g_k)\right] = \gamma_p(g_pP_{p+1}h_{p+1}) = \mathbb{E}\left[P_{p+1}h_{p+1}(X_{0:p})\prod_{k=0}^p g_k(X_{0:k})\right].$$

Comme la propriété de Markov assure que $P_{p+1}h_{p+1}(X_{0:p}) = \mathbb{E}\left[h_{p+1}(X_{0:p+1})|X_{0:p}\right]$ et que $\prod_{k=0}^p g_k(X_{0:k})$ est une fonction de $X_{0:p}$,

$$\begin{aligned}
\mathbb{E}\left[P_{p+1}h_{p+1}(X_{0:p})\prod_{k=0}^p g_k(X_{0:k})\right] &= \mathbb{E}\left[\mathbb{E}\left[h_{p+1}(X_{0:p+1})|X_{0:p}\right]\prod_{k=0}^p g_k(X_{0:k})\right] \\
&= \mathbb{E}\left[h_{p+1}(X_{0:p+1})\prod_{k=0}^p g_k(X_{0:k})\right] = \gamma_{p+1}(h_{p+1}),
\end{aligned}$$

ce qui achève la démonstration. □

Remarque 2.0.6. — *Notons que nous venons d'établir que* $\gamma_{p+1}(h_{p+1}) = \gamma_p(g_p P_{p+1} h_{p+1})$. *Comme* $\gamma_{p+1}(1) = \gamma_p(g_p) = \eta_p(g_p)\gamma_p(1)$, *avec* (2.2), *on en déduit que*

$$\eta_{p+1}(h_{p+1}) = \frac{\gamma_{p+1}(h_{p+1})}{\gamma_{p+1}(1)} = \frac{\gamma_p(g_p P_{p+1} h_{p+1})}{\eta_p(g_p)\gamma_p(1)} = \frac{\eta_p(g_p P_{p+1} h_{p+1})}{\eta_p(g_p)}. \tag{2.10}$$

— *En regardant attentivement la preuve précédente on constate que la conclusion de la proposition 2.0.5 reste vraie sans indépendance dans l'étape d'initialisation et sans indépendance conditionnelle dans les étapes de sélection et de mutation. En particulier, elle reste vraie si l'étape de sélection repose sur l'échantillonnage systématique.*

Théorème 2.0.7. *On suppose que pour tout* $p \in \{0, \dots, n-1\}$, $0 < \underline{g}_p := \inf_{x_{0:p} \in E^{p+1}} g_p(x_{0:p}) \leq \bar{g}_p := \sup_{x_{0:p} \in E^{p+1}} g_p(x_{0:p}) < \infty$ *et que les propriétés d'indépendance et d'indépendance conditionnelle sont satisfaites à l'initialisation et lors des sélections et des mutations. Alors pour tout* $p \in \{0, \dots, n\}$ *et toute fonction* $h_p : E^{p+1} \to \mathbb{R}$ *mesurable bornée (par* $\overline{|h_p|} < \infty$*),*

$$\sup_{M \geq 1} \sqrt{M} \mathbb{E}^{1/4} \left[(\eta_p^M(h_p) - \eta_p(h_p))^4 \right] < \infty \tag{2.11}$$

$$\sup_{M \geq 1} \sqrt{M} \mathbb{E}^{1/4} \left[(\gamma_p^M(h_p) - \gamma_p(h_p))^4 \right] < \infty. \tag{2.12}$$

En outre $\mathbb{P}\left(\lim_{M\to\infty} \eta_p^M(h_p) = \eta_p(h_p)\right) = 1 = \mathbb{P}\left(\lim_{M\to\infty} \gamma_p^M(h_p) = \gamma_p(h_p)\right)$.

Pour des résultats plus fins (convergence presque sûre sous des hypothèses plus faibles, théorème de la limite centrale associé, inégalités de concentration), nous renvoyons au livre de Del Moral [22] et au problème 2.0.9 pour un théorème de la limite centrale dans le cas particulier de l'échantillonnage multinomial.

Comme $\mathbb{E}[f_n(X_{0:n})] = \gamma_n(\tilde{f}_n)$ avec $\tilde{f}_n(x_{0:n}) = f_n(x_{0:n}) \prod_{p=0}^{n-1} g_p^{-1}(x_{0:p})$ bornée si f_n est bornée et les fonctions g_p minorées par une constante positive, le théorème implique le corollaire suivant.

Corollaire 2.0.8. *Sous les hypothèses d'indépendance et de minoration/majoration des fonctions* g_p *du théorème 2.0.7, si* $f_n : E^{n+1} \to \mathbb{R}$ *est bornée,*

$$\sup_{M \geq 1} \sqrt{M} \mathbb{E}^{1/4} \left[(\gamma_n^M(\tilde{f}_n) - \mathbb{E}[f_n(X_{0:n})])^4 \right] < \infty$$

et $\mathbb{P}\left(\lim_{M\to\infty} \gamma_n^M(\tilde{f}_n) = \mathbb{E}[f_n(X_{0:n})]\right) = 1$.

Cette fois, la propriété d'indépendance conditionnelle à l'étape de sélection va jouer un rôle clé dans la preuve du théorème, si bien que celle-ci ne s'applique pas à l'échantillonnage systématique.

Démonstration : La condition (2.11) implique que $\mathbb{E}\left[\sum_{M\geq 1} (\eta_p^M(h_p) - \eta_p(h_p))^4\right] < \infty$ d'où $\mathbb{P}\left(\sum_{M\geq 1} (\eta_p^M(h_p) - \eta_p(h_p))^4 < \infty\right) = 1 = \mathbb{P}(\lim_{M\to\infty} (\eta_p^M(h_p) - \eta_p(h_p))^4 = 0)$. Par le même raisonnement, (2.12) implique que $\mathbb{P}(\lim_{M\to\infty} (\gamma_p^M(h_p) - \gamma_p(h_p))^4 = 0) = 1$.

Montrons l'estimation (2.11) par récurrence sur p. Pour $p=0$, on a

$$\mathbb{E}[(\eta_0^M(h_0)-\eta_0(h_0))^4]=\frac{1}{M^4}\sum_{m_1,m_2,m_3,m_4=1}^{M}\mathbb{E}\left[\prod_{i=1}^{4}(h_0(X_0^{0,m_i})-\eta_0(h_0))\right].$$

Parmi les M^4 espérances qui figurent dans la somme au second membre beaucoup sont nulles : en effet dès que l'un des quatre indices m_i est différent des trois autres, par indépendance des $(X_0^{0,m})_{1\le m\le M}$, l'espérance est égale à l'espérance du produit des termes correspondant aux trois autres indices par $\mathbb{E}[(h_0(X_0^{0,m_i})-\eta_0(h_0)]=0$.
Il reste donc seulement les contributions des cas où

- les 4 indices sont égaux soit M termes en $\mathbb{E}[(h_0(X_0^{0,1})-\eta_0(h_0))^4]$
- 2 des indices prennent une valeur et les 2 autres une autre valeur soit $3M(M-1)$ termes en $\mathbb{E}^2[(h_0(X_0^{0,1})-\eta_0(h_0))^2]$ (3 pour le choix de l'indice $m_k,\ k=2,3$ ou 4, égal au premier indice m_1, $M(M-1)$ pour le choix de deux valeurs différentes parmi M).

Ainsi

$$\mathbb{E}[(\eta_0^M(h_0)-\eta_0(h_0))^4]=\frac{1}{M^3}\mathbb{E}[(h_0(X_0^{0,1})-\eta_0(h_0))^4]+\frac{3(M-1)}{M^3}(\mathrm{Var}\ (h_0(X_0^{0,1})))^2.$$

L'hypothèse de récurrence est donc satisfaite au rang $p=0$. Supposons la satisfaite au rang p et déduisons la au rang $p+1$. Reprenons les notations de la preuve de la proposition 2.0.5. Comme, d'après (2.9), $\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p]=\frac{\eta_p^M(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)}$ et d'après (2.10), $\eta_{p+1}(h_{p+1})=\frac{\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p(g_p)}$,

$$\eta_{p+1}^M(h_{p+1})-\eta_{p+1}(h_{p+1})=\eta_{p+1}^M(h_{p+1})-\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p]+\frac{\eta_p^M(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)}-\frac{\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p(g_p)}.$$

Donc

$$\begin{aligned}\mathbb{E}[(\eta_{p+1}^M(h_{p+1})-\eta_{p+1}(h_{p+1}))^4]&\le 8\mathbb{E}\left[\left(\eta_{p+1}^M(h_{p+1})-\mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p]\right)^4\right]\\&+8\mathbb{E}\left[\left(\frac{\eta_p^M(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)}-\frac{\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p(g_p)}\right)^4\right].\end{aligned}\tag{2.13}$$

Pour le second terme au second membre, on utilise la décomposition suivante

$$\begin{aligned}\frac{\eta_p^M(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)}-\frac{\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p(g_p)}=&\frac{\eta_p^M(g_pP_{p+1}h_{p+1})-\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)}\\&+\frac{\eta_p(g_pP_{p+1}h_{p+1})(\eta_p(g_p)-\eta_p^M(g_p))}{\eta_p^M(g_p)\eta_p(g_p)}.\end{aligned}$$

Comme $0<\frac{1}{\eta_p^M(g_p)}\le\frac{1}{\underline{g}_p}$ et $\left|\frac{\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p(g_p)}\right|\le\overline{|h_{p+1}|}$, on en déduit que

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\eta_p^M(g_pP_{p+1}h_{p+1})}{\eta_p^M(g_p)}-\frac{\eta_p(g_pP_{p+1}h_{p+1})}{\eta_p(g_p)}\right)^4\right]\le&\frac{8}{\underline{g}_p^4}\mathbb{E}[(\eta_p^M(g_pP_{p+1}h_{p+1})-\eta_p(g_pP_{p+1}h_{p+1}))^4]\\&+\frac{8\overline{|h_{p+1}|}^4}{\underline{g}_p^4}\mathbb{E}[(\eta_p(g_p)-\eta_p^M(g_p))^4].\end{aligned}$$

L'hypothèse de récurrence au rang p appliquée avec $h_p = g_p P_{p+1} h_{p+1}$ et avec $h_p = g_p$ assure donc que

$$\sup_{M\geq 1} M^2 \mathbb{E}\left[\left(\frac{\eta_p^M(g_p P_{p+1} h_{p+1})}{\eta_p^M(g_p)} - \frac{\eta_p(g_p P_{p+1} h_{p+1})}{\eta_p(g_p)}\right)^4\right] < \infty. \tag{2.14}$$

Pour traiter le premier terme au second membre de (2.13), on utilise les propriétés de l'étape de mutation puis de l'étape de sélection qui assurent que les variables aléatoires $(X_{0:p+1}^{p+1,m})_{1\leq m\leq M}$ sont conditionnellement indépendantes sachant $\mathcal{F}_p$. Comme

$$\begin{aligned}
&\mathbb{E}\left[\left(\eta_{p+1}^M(h_{p+1}) - \mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p]\right)^4\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{1}{M}\sum_{m=1}^M (h_{p+1}(X_{0:p+1}^{p+1,m}) - \mathbb{E}[h_{p+1}(X_{0:p+1}^{p+1,m})|\mathcal{F}_p])\right)^4 \middle| \mathcal{F}_p\right]\right] \\
&= \frac{1}{M^4}\sum_{m_1,m_2,m_3,m_4=1}^M \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^4 (h_{p+1}(X_{0:p+1}^{p+1,m_i}) - \mathbb{E}[h_{p+1}(X_{0:p+1}^{p+1,m_i})|\mathcal{F}_p]) \middle| \mathcal{F}_p\right]\right].
\end{aligned}$$

L'indépendance conditionnelle assure que dès que l'un des quatre indices m_i est distinct des trois autres $\mathbb{E}\left[\prod_{i=1}^4 (h_{p+1}(X_{0:p+1}^{p+1,m_i}) - \mathbb{E}[h_{p+1}(X_{0:p+1}^{p+1,m_i})|\mathcal{F}_p]) \middle| \mathcal{F}_p\right] = 0$. Comme dans le cas $p = 0$, on en déduit avec la bornitude de h_{p+1} que

$$\sup_{M\geq 1} M^2 \mathbb{E}\left[\left(\eta_{p+1}^M(h_{p+1}) - \mathbb{E}[\eta_{p+1}^M(h_{p+1})|\mathcal{F}_p]\right)^4\right] < \infty.$$

En combinant cette inégalité, (2.14) et (2.13), on conclut que l'hypothèse de récurrence est vraie au rang p, ce qui achève la démonstration de (2.11). Pour en déduire (2.12), on note $h_k = g_k$ pour $k \in \{0, \ldots, p-1\}$ de manière à ce que

$$\gamma_p^M(h_p) - \gamma_p(h_p) = \prod_{k=0}^p \eta_k^M(h_k) - \prod_{k=0}^p \eta_k(h_k) = \sum_{k=0}^p \left(\prod_{j=0}^{k-1} \eta_j^M(h_j)\right)(\eta_k^M(h_k) - \eta_k(h_k))\left(\prod_{j=k+1}^p \eta_j(h_j)\right).$$

Par conséquent,

$$\begin{aligned}
(\gamma_p^M(h_p) - \gamma_p(h_p))^4 &\leq (p+1)^3 \sum_{k=0}^p \left(\prod_{j=0}^{k-1} \eta_j^M(h_j) \prod_{j=k+1}^p \eta_j(h_j)\right)^4 (\eta_k^M(h_k) - \eta_k(h_k))^4 \\
&\leq (p+1)^3 \sum_{k=0}^p \left(\prod_{\substack{j=0\\ j\neq k}}^p \overline{|h_j|}\right)^4 (\eta_k^M(h_k) - \eta_k(h_k))^4.
\end{aligned}$$

En prenant l'espérance dans cette inégalité, on déduit (2.12) de (2.11). □

Problème 2.0.9. *On se place dans le cadre et les notations de ce chapitre. On suppose que pour tout $p \in \mathbb{N}$ la fonction $g_p : E^{p+1} \to \mathbb{R}$ est mesurable et telle que*

$$0 < \underline{g}_p := \inf_{x_{0:p}\in E^{p+1}} g_p(x_{0:p}) \leq \bar{g}_p := \sup_{x_{0:p}\in E^{p+1}} g_p(x_{0:p}) < \infty.$$

On suppose que les positions initiales des particules $(X_0^{0,m})_{m\geq 1}$ sont i.i.d. suivant la probabilité η_0 et que la sélection est effectuée par échantillonnage multinomial, résiduel ou stratifié. Pour $h_p : E^{p+1} \to \mathbb{R}$ mesurable bornée, on pose $\hat{\eta}_p^M(h_p) = \frac{\eta_p^M(g_p h_p)}{\eta_p^M(g_p)}$ et $\hat{\eta}_p(h_p) = \frac{\eta_p(g_p h_p)}{\eta_p(g_p)}$. Pour $h_{p+1} : E^{p+2} \to \mathbb{R}$ mesurable bornée, on note $P_{p+1}h_{p+1} : E^{p+1} \to \mathbb{R}$ la fonction définie par $P_{p+1}h_{p+1}(x_{0:p}) = \int_{x_{p+1}\in E} h_{p+1}(x_{0:p+1})P_{p+1}(x_p, dx_{p+1})$.

1. *Que peut-on dire de $\mathbb{P}\left(\lim_{M\to\infty}(\eta_p^M(g_p h_p), \eta_p^M(g_p)) = (\eta_p(g_p h_p), \eta_p(g_p))\right)$? En déduire que $\mathbb{P}\left(\lim_{M\to\infty}\hat{\eta}_p^M(h_p) = \hat{\eta}_p(h_p)\right) = 1$.*
2. *Vérifier que $\eta_{p+1}(h_{p+1}) = \hat{\eta}_p(P_{p+1}h_{p+1})$.*
3. *Pour quelle fonction $h_{p+1} : E^{p+2} \to \mathbb{R}$, a-t-on $\frac{1}{M}\sum_{m=1}^M h_p(X_{0:p}^{p+1,m}) = \eta_{p+1}^M(h_{p+1})$? Vérifier qu'alors $\eta_{p+1}(h_{p+1}) = \hat{\eta}_p(h_p)$ et en déduire que $\mathbb{P}\left(\lim_{M\to\infty}\frac{1}{M}\sum_{m=1}^M h_p(X_{0:p}^{p+1,m}) = \hat{\eta}_p(h_p)\right) = 1$.*
4. *Montrer que lorsque $M \to \infty$, pour $h_0 : E \to \mathbb{R}$ mesurable bornée, $\sqrt{M}(\eta_0^M(h_0) - \eta_0(h_0))$ converge en loi vers $\mathcal{N}_1(0, \mathbb{V}_0(h_0))$ en explicitant $\mathbb{V}_0(h_0)$.*

On suppose désormais que l'étape de sélection est effectuée suivant l'échantillonnage multinomial. L'objectif de la suite du problème est de montrer par récurrence sur p que pour $h_p : E^{p+1} \to \mathbb{R}$ mesurable bornée, $\sqrt{M}(\eta_p^M(h_p) - \eta_p(h_p))$ converge en loi lorsque $M \to \infty$ vers $\mathcal{N}_1(0, \mathbb{V}_p(h_p))$. On suppose l'hypothèse de récurrence satisfaite au rang p et on se donne $h_{p+1} : E^{p+2} \to \mathbb{R}$ mesurable bornée.

5. *Pour $\hat{h}_p : E^{p+1} \to \mathbb{R}$ mesurable bornée, on pose $\tilde{h}_p = \frac{g_p}{\eta_p(g_p)}(\hat{h}_p - \hat{\eta}_p(\hat{h}_p))$. Remarquer que $\eta_p(\tilde{h}_p) = 0$ et $\hat{\eta}_p^M(\hat{h}_p) - \hat{\eta}_p(\hat{h}_p) = \frac{\eta_p(g_p)}{\eta_p^M(g_p)}\eta_p^M(\tilde{h}_p)$. Vérifier que $\sqrt{M}\left(\hat{\eta}_p^M(\hat{h}_p) - \hat{\eta}_p(\hat{h}_p) - \eta_p^M(\tilde{h}_p)\right)$ converge en probabilité vers 0 lorsque $M \to \infty$ et en déduire que $\sqrt{M}(\hat{\eta}_p^M(\hat{h}_p) - \hat{\eta}_p(\hat{h}_p))$ converge en loi vers $\mathcal{N}_1\left(0, \mathbb{V}_p(\tilde{h}_p)\right)$.*
6. *Vérifier que pour $v \in \mathbb{R}$ et $\mathcal{F}_p = \sigma((X_{0:k}^{k,m})_{1\leq m\leq M, 0\leq k\leq p})$,*
$$\mathbb{E}\left[e^{iv\sqrt{M}(\eta_{p+1}^M(h_{p+1})-\eta_{p+1}(h_{p+1}))}\right] = \mathbb{E}\Bigg[e^{iv\sqrt{M}(\hat{\eta}_p^M(P_{p+1}h_{p+1})-\hat{\eta}_p(P_{p+1}h_{p+1}))} \times \mathbb{E}\left[e^{iv\sqrt{M}\left(\eta_{p+1}^M(h_{p+1})-\hat{\eta}_p^M(P_{p+1}h_{p+1})\right)}\middle|\mathcal{F}_p\right]\Bigg].$$
7. *On pose $\hat{\eta}_p^M P_{p+1}(dx_{0:p+1}) := \hat{\eta}_p^M(dx_{0:p})P_{p+1}(x_p, dx_{p+1})$. Vérifier que $\hat{\eta}_p^M P_{p+1}(h_{p+1}) = \hat{\eta}_p^M(P_{p+1}h_{p+1})$ et montrer que conditionnellement à $\mathcal{F}_p$, les trajectoires $(X_{0:p+1}^{p+1,m})_{1\leq m\leq M}$ sont i.i.d. suivant $\hat{\eta}_p^M P_{p+1}(dx_{0:p+1})$.*
8. *Vérifier que pour $x \in \mathbb{R}$, $\left|e^{ix} - 1 - ix + \frac{x^2}{2}\right| \leq \frac{|x|^3}{6}$ puis, en notant $\overline{|h_{p+1}|} = \sup_{x_{0:p+1}\in E^{p+2}}|h_{p+1}(x_{0:p+1})|$, montrer que pour $v \in \mathbb{R}$,*
$$\Bigg|\mathbb{E}\left[e^{i\frac{v}{\sqrt{M}}(h_{p+1}(X_{0:p+1}^{p+1,1})-\hat{\eta}_p^M(P_{p+1}h_{p+1}))}\middle|\mathcal{F}_p\right] - 1 + \frac{v^2}{2M}\left(\hat{\eta}_p^M(P_{p+1}h_{p+1}^2) - (\hat{\eta}_p^M(P_{p+1}h_{p+1}))^2\right)\Bigg| \leq \frac{4|v|^3\overline{|h_{p+1}|}^3}{3M^{3/2}}.$$
9. *On pose $\mathbb{W}_{p+1}(h_{p+1}) := \hat{\eta}_p(P_{p+1}h_{p+1}^2) - (\hat{\eta}_p(P_{p+1}h_{p+1}))^2$ et*
$$\varepsilon_M = M\left(\mathbb{E}\left[e^{i\frac{v}{\sqrt{M}}(h_{p+1}(X_{0:p+1}^{p+1,1})-\hat{\eta}_p^M(P_{p+1}h_{p+1}))}\middle|\mathcal{F}_p\right] - 1 + \frac{v^2}{2M}\mathbb{W}_{p+1}(h_{p+1})\right)$$

Vérifier que

$$|\varepsilon_M| \leq \frac{4|v|^3\overline{|h_{p+1}|}^3}{3\sqrt{M}} + \frac{v^2}{2}|\hat{\eta}_p^M(P_{p+1}h_{p+1}^2) - \hat{\eta}_p(P_{p+1}h_{p+1}^2)|$$
$$+ \frac{v^2}{2}|(\hat{\eta}_p^M(P_{p+1}h_{p+1}))^2 - (\hat{\eta}_p(P_{p+1}h_{p+1}))^2|$$

et en déduire le comportement de la suite $(\varepsilon_M)_{M\geq 1}$ *lorsque* $M \to \infty$.

10. *Comment comparer* $P_{p+1}h_{p+1}^2$ *et* $(P_{p+1}h_{p+1})^2$ *? En déduire que* $\mathbb{W}_{p+1}(h_{p+1}) \geq 0$.
11. *À l'aide de la formule du binôme, vérifier que pour* $M \geq \frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{4}$,

$$\left|\left(1 - \frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{2M} + \frac{\varepsilon_M}{M}\right)^M - \left(1 - \frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{2M}\right)^M\right| \leq \sum_{k=1}^M \frac{|\varepsilon_M|^k}{k!} \leq e^{|\varepsilon_M|} - 1$$

et en déduire que $\mathbb{E}\left[e^{iv\sqrt{M}\left(\eta_{p+1}^M(h_{p+1}) - \hat{\eta}_p^M(P_{p+1}h_{p+1})\right)}|\mathcal{F}_p\right]$ *converge presque sûrement vers* $e^{-\frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{2}}$ *lorsque* $M \to \infty$*. puis que*

$$\lim_{M\to\infty} \mathbb{E}\left[\left|\mathbb{E}\left[e^{iv\sqrt{M}\left(\eta_{p+1}^M(h_{p+1}) - \hat{\eta}_p^M(P_{p+1}h_{p+1})\right)}|\mathcal{F}_p\right] - e^{-\frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{2}}\right|\right] = 0.$$

12. *Vérifier que*

$$\left|\mathbb{E}\left[e^{iv\sqrt{M}(\eta_{p+1}^M(h_{p+1}) - \eta_{p+1}(h_{p+1}))}\right] - e^{-\frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{2}}\mathbb{E}\left[e^{iv\sqrt{M}(\hat{\eta}_p^M(P_{p+1}h_{p+1}) - \hat{\eta}_p(P_{p+1}h_{p+1}))}\right]\right|$$
$$\leq \mathbb{E}\left[\left|\mathbb{E}\left[e^{iv\sqrt{M}\left(\eta_{p+1}^M(h_{p+1}) - \hat{\eta}_p^M(P_{p+1}h_{p+1})\right)}|\mathcal{F}_p\right] - e^{-\frac{v^2\mathbb{W}_{p+1}(h_{p+1})}{2}}\right|\right]$$

et conclure que, lorsque $M \to \infty$, $\sqrt{M}(\eta_{p+1}^M(h_{p+1}) - \eta_{p+1}(h_{p+1}))$ *converge en loi vers* $\mathcal{N}_1(0, \mathbb{V}_{p+1}(h_{p+1}))$ *avec*

$$\mathbb{V}_{p+1}(h_{p+1}) = \mathbb{W}_{p+1}(h_{p+1}) + \mathbb{V}_p(\mathbb{L}_p h_{p+1}) \text{ où } \mathbb{L}_p h_{p+1} = \frac{g_p}{\eta_p(g_p)}\left(P_{p+1}h_{p+1} - \hat{\eta}_p(P_{p+1}h_{p+1})\right).$$

Chapitre 3

Stability and convergence of stochastic algorithms

In this chapter, we will present the sufficient conditions given by Andrieu, Moulines and Priouret [4] for the stability and the convergence of deterministic recursions. We will check that these conditions ensure the convergence of the Robins-Monro stochastic algorithm without the need of the Robbins-Sigmund lemma. These conditions will also be used in the next chapter to study the convergence of some adaptive Monte Carlo Markov Chain algorithms which mix a Metropolis Hasting sampling procedure with a stochastic algorithm.

Let $d_\theta \in \mathbb{N}^*$, Θ be an open subset of $\mathbb{R}^{d_\theta}$ and $h : \Theta \to \mathbb{R}^{d_\theta}$. We are interested in the asymptotic behaviour as $n \to \infty$ of a Θ-valued sequence $(\theta_n)_{n\geq 0}$ satisfying the induction equation

$$\forall n \in \mathbb{N},\ \theta_{n+1} = \theta_n + \gamma_{n+1}(h(\theta_n) + \xi_{n+1}) \tag{3.1}$$

where $(\gamma_n, \xi_n)_{n\geq 1}$ is a $\mathbb{R}_+ \times \mathbb{R}^{d_\theta}$-valued sequence. Notice that (3.1) alone does not imply that $(\theta_n)_{n\in\mathbb{N}}$ is Θ-valued.
We will use a Lyapunov function $w : \Theta \to \mathbb{R}_+$. For any $M' > M \geq 0$, let $\mathcal{W}^M = \{\theta \in \Theta : w(\theta) \leq M\}$ and $\mathcal{W}_M^{M'} = \{\theta \in \Theta : M \leq w(\theta) \leq M'\}$.

Example 3.0.1. *In the Robbins-Monro algorithm,* $\Theta = \mathbb{R}^{d_\theta}$*,*

$$\theta_{n+1} = \theta_n + \gamma_{n+1} Y_{n+1}$$

where $(Y_n)_n$ *is a sequence of square integrable random variables such that for* $\mathcal{F}_n = \sigma(\theta_0, Y_1, \ldots, Y_n)$*,* $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = h(\theta_n)$ *and* $(\gamma_n)_{n\geq 1}$ *a deterministic sequence of positive numbers. This algorithm is aimed at computing* $\theta_\star$ *such that* $h(\theta_\star) = 0$ *when* $\forall \theta \neq \theta_\star$*,* $h(\theta).(\theta_\star - \theta) > 0$*. Typically, this condition is satisfied when* h *is equal to minus the gradient of a strictly convex function minimal at* $\theta_\star$*. The introduction of the martingale increment* $\xi_{n+1} = Y_{n+1} - \mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_{n+1} - h(\theta_n)$ *permits to rewrite the Robbins-Monro dynamics as a special case of* (3.1)*. In Example 3.0.8 below, the Lyapunov function* $w(\theta) = |\theta - \theta_\star|^2$ *is used to prove the convergence of* $(\theta_n)_n$ *to* $\theta_\star$ *under additional assumptions on* h *and the sequences* $(\gamma_n)_{n\geq 1}$ *and* $(\xi_n)_{n\geq 1}$*.*

Proposition 3.0.2. *Assume that* h *is continuous,* w *is continuously differentiable and there exist* $0 < M_0 < M_1 < \infty$ *such that* $\mathcal{W}^{M_1}$ *is a compact subset of* $\mathbb{R}^{d_\theta}$ *and* $\nabla w.h$ *is strictly negative on* $\mathcal{W}_{M_0}^{M_1}$*.*

For all M, M' such that $M_0 < M < M' < M_1$, there exists $\delta, \eta > 0$ such that if $\theta_0 \in \mathcal{W}^M$, the sequence $(\gamma_n)_{n\geq 1}$ is $[0,\delta]$-valued and the sequence $(\xi_n)_{n\geq 1}$ admits a decomposition $\xi_n = \xi_n^1 + \xi_n^2$ such that $\sup_{m\geq 1}(|\sum_{k=1}^m \gamma_k \xi_k^1| + |\sum_{k=1}^m \gamma_k \xi_k^2 1_{\mathcal{W}^{M_1}}(\theta_{k-1})|) \leq \eta$, then $\forall n \in \mathbb{N}, \theta_n \in \mathcal{W}^{M'} \subset \mathcal{W}^{M_1}$.

The next corollary states conditions under which the recurrence of the sequence $(\theta_n)_n$ i.e. return of this sequence infinitely often to a compact subset of $\mathbb{R}^{d_\theta}$ implies that the sequence stays in a larger compact subset after a finite time, a property called stability.

Corollary 3.0.3. *Under the assumptions of Proposition 3.0.2, if $\lim_{n\to\infty}\gamma_n = 0$, the sequence $(\xi_n)_{n\geq 1}$ admits a decomposition $\xi_n = \xi_n^1 + \xi_n^2$ such that $\lim_{n\to\infty}\sup_{m\geq n}(|\sum_{k=n}^m \gamma_k \xi_k^1| + |\sum_{k=n}^m \gamma_k \xi_k^2 1_{\mathcal{W}^{M_1}}(\theta_{k-1})|) = 0$ and there exists $M \in (M_0, M_1)$ such that $(\theta_n)_{n\geq 1}$ is infinitely often in $\mathcal{W}^M$, then for any $M' \in (M, M_1)$, $\exists N < \infty$, $\forall n \geq N$, $\theta_n \in \mathcal{W}^{M'}$.*

Remark 3.0.4. *In view of applications to stochastic algorithms where a martingale increment contributes to each of the ξ_j (see the above example of the Robins-Monro algorithm), it is really important that the assumption is formulated on $\sup_{k\geq n}\left(\left|\sum_{j=n}^k \gamma_j \xi_j^1\right| + \left|\sum_{j=n}^k \gamma_j \xi_j^2 1_{\mathcal{W}^{M_1}}(\theta_{j-1})\right|\right)$ and not $\sup_{k\geq n}\left(\sum_{j=n}^k \gamma_j \left|\xi_j^1\right| + \sum_{j=n}^k \gamma_j 1_{\mathcal{W}^{M_1}}(\theta_{j-1})\left|\xi_j^2\right|\right)$, even if this complicates the proof. The introduction of $1_{\mathcal{W}^{M_1}}(\theta_{j-1})$ may help to check the hypothesis. Nevertheless, the corollary states that for n large enough $\theta_n \in \mathcal{W}^{M'} \subset \mathcal{W}^{M_1}$ and therefore $\forall k \geq n$, $\sum_{j=n}^k \gamma_j \xi_j^2 1_{\mathcal{W}^{M'}}(\theta_{j-1}) = \sum_{j=n}^k \gamma_j \xi_j^2$ and $\lim_{n\to\infty}\sup_{k\geq n}\left|\sum_{j=n}^k \gamma_j \xi_j\right| = 0$.*

Proof of Corollary 3.0.3: Let $M' \in (M, M_1)$ and $\delta, \eta > 0$ denote the constants associated with (M, M') in Proposition 3.0.2. The vanishing limits assumptions ensure the existence of $n_{\delta,\eta}$ such that $\forall n \geq n_{\delta,\eta}$, $\gamma_n \in [0,\delta]$ and $\sup_{m\geq n}(|\sum_{k=n}^m \gamma_k \xi_k^1| + |\sum_{k=n}^m \gamma_k \xi_k^2 1_{\mathcal{W}^{M_1}}(\theta_{k-1})|) \leq \eta$. The set $\{n \geq n_{\delta,\eta} : \theta_n \in \mathcal{W}^M\}$ is infinite and therefore non empty. Let N denote its minimum. One concludes that $\theta_n \in \mathcal{W}^{M'}$ for all $n \geq N$, by applying the statement in Proposition 3.0.2 to the shifted sequence $(\theta_{N+n})_{n\in\mathbb{N}}$. ∎

The proof of Proposition 3.0.2 relies on the following lemma and is postponed after its proof.

Lemma 3.0.5. *Assume that h is continuous, w is continuously differentiable and there exists $M_1 \in (0,\infty)$ such that $\mathcal{W}^{M_1}$ is a compact subset of $\mathbb{R}^{d_\theta}$. Then*

$$\begin{aligned}\forall \varepsilon > 0,\ \exists \hat{\eta}_\varepsilon > 0,\ \forall \hat{\theta} \in \mathcal{W}^{M_1}, \forall \tilde{\theta} \in \mathbb{R}^{d_\theta} \text{ such that } |\tilde{\theta} - \hat{\theta}| \leq \hat{\eta}_\varepsilon,&\\ \tilde{\theta} \in \Theta \text{ and } |w(\tilde{\theta}) - w(\hat{\theta})| \vee |h(\tilde{\theta}) - h(\hat{\theta})| \leq \varepsilon.&\end{aligned} \tag{3.2}$$

Moreover, for any non empty compact subset $\mathcal{K}$ of $\mathcal{W}^{M_1}$ such that $\nabla w.h$ is strictly negative on $\mathcal{K}$,

$$\begin{aligned}\exists \delta, \lambda, z > 0, \forall \bar{\theta} \in \mathcal{K},\ \forall \gamma \in [0,\delta],\ \forall \xi \in \mathbb{R}^{d_\theta} \text{ with } |\xi| \leq \lambda,&\\ \bar{\theta} + \gamma(h(\bar{\theta}) + \xi) \in \Theta \text{ and } w(\bar{\theta} + \gamma(h(\bar{\theta}) + \xi)) \leq w(\bar{\theta}) - z\gamma.&\end{aligned} \tag{3.3}$$

Last, if moreover there exists $M_0 \in (0, M_1)$ such that $\nabla w.h$ is strictly negative on $\mathcal{W}^{M_1}_{M_0}$, then

$$\forall M \in (M_0, M_1),\ \exists \delta, \lambda > 0, \forall \bar\theta \in \mathcal{W}^M,\ \forall \gamma \in [0, \delta], \\ \forall \xi \in \mathbb{R}^{d_\theta} \textit{ with } |\xi| \le \lambda,\ \bar\theta + \gamma(h(\bar\theta) + \xi) \in \mathcal{W}^M. \tag{3.4}$$

Proof: Let us prove the second assertion and first check that the compactness of $\mathcal{W}^{M_1}$ combined with the continuity of ∇w on Θ imply that for fixed $\varepsilon > 0$,

$$\exists \eta_\varepsilon > 0,\ \forall \hat\theta \in \mathcal{W}^{M_1},\ \forall \tilde\theta \in \mathbb{R}^{d_\theta} \text{ such that } |\tilde\theta - \hat\theta| \le \eta_\varepsilon,\ \tilde\theta \in \Theta \text{ and } |\nabla w(\tilde\theta) - \nabla w(\hat\theta)| \le \varepsilon.$$

The same reasoning applied to the continuous functions w and h in place of ∇w leads to (3.2).

Since Θ is open and ∇w is continuous, for each $\theta \in \Theta$ there exists $\eta_\theta > 0$ such that the open ball $B(\theta, \eta_\theta)$ is included in Θ and $\forall \hat\theta \in B(\theta, \eta_\theta)$, $|\nabla w(\hat\theta) - \nabla w(\theta)| \le \frac{\varepsilon}{2}$. The compact set $\mathcal{W}^{M_1}$ included in Θ is covered by finitely many balls $B(\theta, \eta_\theta/2)$. Let $\eta_\varepsilon > 0$ denote the minimal radius of these balls. For $\hat\theta \in \mathcal{W}^{M_1}$ there exists one of these finitely many balls such that $\hat\theta \in B(\theta, \eta_\theta/2)$ and $\eta_\varepsilon \le \eta_\theta/2$. If $\tilde\theta \in \mathbb{R}^{d_\theta}$ is such that $|\tilde\theta - \hat\theta| \le \eta_\varepsilon$, then $\hat\theta, \tilde\theta \in B(\theta, \eta_\theta)$ so that $\tilde\theta \in \Theta$ and

$$|\nabla w(\hat\theta) - \nabla w(\tilde\theta)| \le |\nabla w(\hat\theta) - \nabla w(\theta)| + |\nabla w(\theta) - \nabla w(\tilde\theta)| \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Let $\mathcal{K}$ be a non-empty compact subset of $\mathcal{W}^{M_1}$ such that $\nabla w.h$ is strictly negative on $\mathcal{K}$. One has $\sup_{\theta\in\mathcal{K}} \nabla w.h(\theta) < 0$. Let $\varepsilon = \frac{-\sup_{\theta\in\mathcal{K}} \nabla w.h(\theta)}{2\sup_{\theta\in\mathcal{K}} |h(\theta)|}$ and $\bar\theta \in \mathcal{K}$. One has $\varepsilon \le \frac{\sup_{\theta\in\mathcal{K}} |\nabla w(\theta)|}{2}$ and

$$\begin{aligned}
&\forall \theta \in \mathbb{R}^{d_\theta} \text{ such that } |\theta - \bar\theta| \le \eta_\varepsilon,\ \theta \in \Theta,\\
&|\nabla w(\theta)| \le |\nabla w(\bar\theta)| + |\nabla w(\theta) - \nabla w(\bar\theta)| \le \sup_{\tilde\theta\in\mathcal{K}} |\nabla w(\tilde\theta)| + \varepsilon \le \frac{3}{2} \sup_{\tilde\theta\in\mathcal{K}} |\nabla w(\tilde\theta)|, \text{ and}\\
&\nabla w(\theta).h(\bar\theta) \le \nabla w(\bar\theta).h(\bar\theta) + |\nabla w(\theta) - \nabla w(\bar\theta)| \sup_{\tilde\theta\in\mathcal{K}} |h(\tilde\theta)| \le \frac{1}{2} \sup_{\tilde\theta\in\mathcal{K}} \nabla w.h(\tilde\theta).
\end{aligned}$$

Let $\gamma \in [0, \frac{\eta_\varepsilon}{2\sup_{\theta\in\mathcal{K}} |h(\theta)|}]$ and $|\xi| \le \sup_{\theta\in\mathcal{K}} |h(\theta)|$. Then $|\bar\theta + \gamma(h(\bar\theta) + \xi) - \bar\theta| \le \eta_\varepsilon$ so that, for all $t \in [0, 1]$, $\bar\theta + \gamma t(h(\bar\theta) + \xi) \in \Theta$, $|\nabla w(\bar\theta + \gamma t(h(\bar\theta) + \xi))| \le \frac{3}{2} \sup_{\tilde\theta\in\mathcal{K}} |\nabla w(\tilde\theta)|$ and $\nabla w(\bar\theta + \gamma t(h(\bar\theta) + \xi)).h(\bar\theta) \le \frac{1}{2} \sup_{\theta\in\mathcal{K}} \nabla w.h(\theta)$. Hence

$$\begin{aligned}
w(\bar\theta + \gamma(h(\bar\theta) + \xi)) - w(\bar\theta) &= \gamma \int_0^1 \nabla w(\bar\theta + \gamma t(h(\bar\theta) + \xi)).(h(\bar\theta) + \xi)dt\\
&\le \gamma \left(\frac{1}{2} \sup_{\theta\in\mathcal{K}} \nabla w.h(\theta) + \frac{3|\xi|}{2} \sup_{\theta\in\mathcal{K}} |\nabla w(\theta)| \right),
\end{aligned}$$

where the right-hand side is non positive when $|\xi| \le -\sup_{\theta\in\mathcal{K}} \nabla w.h(\theta)/(3\sup_{\theta\in\mathcal{K}} |\nabla w(\theta)|)$ which implies $|\xi| \le \frac{\sup_{\theta\in\mathcal{K}} |h(\theta)|}{3}$. If moreover $|\xi| \le -\sup_{\theta\in\mathcal{K}} \nabla w.h(\theta)/(6\sup_{\theta\in\mathcal{K}} |\nabla w(\theta)|)$, then

$$w(\bar\theta + \gamma(h(\bar\theta) + \xi)) \le w(\bar\theta) + \frac{\gamma}{4} \sup_{\theta\in\mathcal{K}} \nabla w.h(\theta)$$

and the second statement holds with $\delta = \eta_\varepsilon/(2\sup_{\theta\in\mathcal{K}}|h(\theta)|)$ for $\varepsilon = -\sup_{\theta\in\mathcal{K}}\nabla w.h(\theta)/(2\sup_{\theta\in\mathcal{K}}|h(\theta)|)$, $\lambda = -\sup_{\theta\in\mathcal{K}}\nabla w.h(\theta)/(6\sup_{\theta\in\mathcal{K}}|\nabla w(\theta)|)$ and $z = -\sup_{\theta\in\mathcal{K}}\nabla w.h(\theta)/4$.

Let us now moreover assume that there exists $M_0 \in (0, M_1)$ such that $\nabla w.h$ is strictly negative on $\mathcal{W}^{M_1}_{M_0}$. Let $M \in (M_0, M_1)$. If $\mathcal{W}^M$ is empty then the conclusion is immediate. So we now suppose that $\mathcal{W}^M$ is non empty and even that both $\mathcal{W}^{M_0}$ and $\mathcal{W}^M_{M_0}$ are non empty. Let us deal with the case $\bar\theta \in \mathcal{W}^{M_0}$ using the first assertion before using the case $w(\bar\theta+\gamma(h(\bar\theta)+\xi)) - w(\bar\theta) \leq 0$ in the previous reasoning applied with $\mathcal{K} = \mathcal{W}^{M_1}_{M_0}$ to deal with $\bar\theta \in \mathcal{W}^M_{M_0}$.

If $\bar\theta \in \mathcal{W}^{M_0}$, $\gamma \in [0, \frac{\hat\eta_{M-M_0}}{1+\sup_{\theta\in\mathcal{W}^{M_0}}|h(\theta)|}]$ and $|\xi| \leq 1$ then $|\bar\theta+\gamma(h(\bar\theta)+\xi)-\bar\theta| \leq \hat\eta_{M-M_0}$ so that $\bar\theta+\gamma(h(\bar\theta)+\xi) \in \Theta$ and

$$w(\bar\theta+\gamma(h(\bar\theta)+\xi)) \leq w(\bar\theta) + M - M_0 \leq M_0 + M - M_0 = M,$$

which imply that $\bar\theta+\gamma(h(\bar\theta)+\xi) \in \mathcal{W}^M$.
The set $\mathcal{W}^{M_1}_{M_0} = \{\theta \in \mathcal{W}^{M_1} : w(\theta) \geq M_0\}$ is compact since from any sequence in this set, by compacity of $\mathcal{W}^{M_1}$, one can extract a subsequence converging to some limit in $\mathcal{W}^{M_1}$ and this limit belongs to $\mathcal{W}^{M_1}_{M_0}$ by continuity of w. The case $w(\bar\theta+\gamma(h(\bar\theta)+\xi)) - w(\bar\theta) \leq 0$ in the previous reasoning applied with $\mathcal{K} = \mathcal{W}^{M_1}_{M_0}$ ensures that the last statement holds with $\delta = \frac{\hat\eta_{M-M_0}}{1+\sup_{\theta\in\mathcal{W}^{M_0}}|h(\theta)|} \wedge \frac{\eta_\varepsilon}{2\sup_{\theta\in\mathcal{W}^{M_1}_{M_0}}|h(\theta)|}$ for $\varepsilon = \frac{-\sup_{\theta\in\mathcal{W}^{M_1}_{M_0}}\nabla w.h(\theta)}{2\sup_{\theta\in\mathcal{W}^{M_1}_{M_0}}|h(\theta)|)}$, $\lambda = 1 \wedge \frac{-\sup_{\theta\in\mathcal{W}^{M_1}_{M_0}}\nabla w.h(\theta)}{3\sup_{\theta\in\mathcal{W}^{M_1}_{M_0}}|\nabla w(\theta)|}$.
When $\mathcal{W}^{M_0}$ (resp. $\mathcal{W}^M_{M_0}$) is empty, one may choose δ and λ equal to the respective constants after (resp. before) $\wedge$ in their previous definitions.

∎

Proof of Proposition 3.0.2: Let M, M' be such that $M_0 < M < M' < M_1$. We define inductively another sequence $(\bar\theta_n)_{n\in\mathbb{N}}$ by $\bar\theta_0 = \theta_0$ and for all $n \in \mathbb{N}$, $\bar\theta_{n+1} = \bar\theta_n + \gamma_{n+1}h(\theta_n)$. One easily checks by induction on n that

$$\forall n \in \mathbb{N},\ \theta_n - \bar\theta_n = \sum_{k=1}^{n}\gamma_k\xi_k. \tag{3.5}$$

Let $\hat\eta_\varepsilon$ (resp. δ, λ) be the positive constants given by the first (resp. third) assertion in Lemma 3.0.5 and $\eta = \hat\eta_{M'-M} \wedge \hat\eta_\lambda$. Let us suppose that the sequence $(\gamma_n)_{n\geq 1}$ takes values in $[0, \delta]$ and that $\max_{n\geq 1}(|\sum_{k=1}^n \gamma_k\xi^1_k| + |\sum_{k=1}^n \gamma_k\xi^2_k 1_{\mathcal{W}^{M_1}}(\theta_{k-1})|) \leq \eta$. Let us check by induction on n that for all $n \in \mathbb{N}$, for all $k \in \{0,\ldots,n\}$, $\bar\theta_k \in \mathcal{W}^M$ and $\theta_k \in \mathcal{W}^{M'}$. The induction hypothesis holds for $n = 0$ since $\bar\theta_0 = \theta_0 \in \mathcal{W}^M \subset \mathcal{W}^{M'}$. Assuming that that it holds for n, it is enough to check that $\bar\theta_{n+1} \in \mathcal{W}^M$ and $\theta_{n+1} \in \mathcal{W}^{M'}$ to conclude the proof. Since

$$\sum_{k=1}^{n}\gamma_k\xi_k = \sum_{k=1}^{n}\gamma_k\xi_k 1_{\mathcal{W}^{M_1}}(\theta_{k-1}) = \sum_{k=1}^{n}\gamma_k\xi^1_k + \sum_{k=1}^{n}\gamma_k\xi^2_k 1_{\mathcal{W}^{M_1}}(\theta_{k-1}),$$

by (3.5), $|\theta_n - \bar\theta_n| \leq |\sum_{k=1}^n\gamma_k\xi^1_k| + |\sum_{k=1}^n\gamma_k\xi^2_k 1_{\mathcal{W}^{M_1}}(\theta_{k-1})| \leq \eta \leq \hat\eta_\lambda$ and by the first statement in Lemma 3.0.5, $|h(\theta_n) - h(\bar\theta_n)| \leq \lambda$. Since $\bar\theta_{n+1} = \bar\theta_n + \gamma_{n+1}h(\bar\theta_n) + \gamma_{n+1}(h(\theta_n) - h(\bar\theta_n))$, the third statement in this lemma ensures that $\bar\theta_{n+1} \in \mathcal{W}^M$. Moreover $\sum_{k=1}^{n+1}\gamma_k\xi_k =$

$\sum_{k=1}^{n+1}\gamma_k\xi_k^1+\sum_{k=1}^{n+1}\gamma_k\xi_k^2 1_{\mathcal{W}^{M_1}}(\theta_{k-1})$ so that, by (3.5), $|\theta_{n+1}-\bar{\theta}_{n+1}|\leq\hat{\eta}_{M'-M}$. By the first statement in Lemma 3.0.5, $w(\theta_{n+1})\leq w(\bar{\theta}_{n+1})+M'-M\leq M'$ so that $\theta_{n+1}\in\mathcal{W}^{M'}$. ∎

Let us now introduce

$$\mathcal{L}=\{\theta\in\Theta:\nabla w.h(\theta)=0\}.$$

Theorem 3.0.6. *Assume that h is continuous, w is continuously differentiable and there exist $0<M_0<M_1<\infty$ such that $\mathcal{L}\cap\mathcal{W}_{M_0}^{M_1}=\emptyset$, $\mathcal{W}^{M_1}$ is a compact subset of $\mathbb{R}^{d_\theta}$ and $\nabla w.h$ is non-positive on $\mathcal{W}^{M_1}$. Assume moreover that $\sum_{n\geq1}\gamma_n=\infty$, $\lim_{n\to\infty}\gamma_n=0$ and that the sequence $(\xi_n)_{n\geq1}$ admits a decomposition $\xi_n=\xi_n^1+\xi_n^2$ such that $\lim_{n\to\infty}\sup_{k\geq n}\left(\left|\sum_{j=n}^k\gamma_j\xi_j^1\right|+\left|\sum_{j=n}^k\gamma_j\xi_j^2 1_{\mathcal{W}^{M_1}}(\theta_{j-1})\right|\right)=0$. If the image $\mathcal{D}$ of $\mathcal{L}\cap\mathcal{W}^{M_1}$ by w has an empty interior and there exists $M\in(M_0,M_1)$ such that for n large enough $\theta_n\in\mathcal{W}^M$, then $w(\theta_n)$ converges to some limit in $\mathcal{D}$ as $n\to\infty$ and $\lim_{n\to\infty}\inf_{\theta\in\mathcal{L}\cap\mathcal{W}^{M_1}}|\theta_n-\theta|=0$.*

Remark 3.0.7. — *The assumptions $\mathcal{L}\cap\mathcal{W}_{M_0}^{M_1}=\emptyset$ and that $\nabla w.h$ is non-positive on $\mathcal{W}^{M_1}$ imply that this function is negative on $\mathcal{W}_{M_0}^{M_1}$. Hence the assumptions are stronger than those of Corollary 3.0.3 so that if for some $M'\in(M_0,M_1)$, $(\theta_n)_n$ is infinitely often in $\mathcal{W}^{M'}$, then for any $M\in(M',M_1)$, $\theta_n\in\mathcal{W}^M$ for n large enough.*

— *If $\mathcal{L}\cap\mathcal{W}^{M_1}$ is finite then $\mathcal{D}$ is also finite and has an empty interior.*

— *When $\mathcal{L}\cap\mathcal{W}^{M_1}$ is a singleton $\{\theta_\star\}$, then the conclusion is that $\lim_{n\to\infty}\theta_n=\theta_\star$ and the proof is much simpler. Let $\underline{\theta}\in\mathcal{W}^{M_1}$ be such that $w(\underline{\theta})=\inf_{\theta\in\mathcal{W}^{M_1}}w(\theta)$. The set $\mathcal{W}^M$ is not empty since it contains θ_n for n large. Hence $w(\underline{\theta})\leq M<M_1$. Since Θ is open and w continuous, for $\varepsilon>0$ small enough, $\{\theta\in\mathbb{R}^{d_\theta}:|\theta-\underline{\theta}|<\varepsilon\}\subset\mathcal{W}^{M_1}$ and therefore $\nabla w(\underline{\theta})=0$, by the first order optimality condition. Since $\nabla w.h(\theta)<0$ for $\theta\in\mathcal{W}^{M_1}\setminus\{\theta_\star\}$, $\underline{\theta}=\theta_\star$. Hence for $\theta\in\mathcal{W}^{M_1}\setminus\{\theta_\star\}$, $w(\theta)>w(\theta_\star)$ and for $\varepsilon>0$, by compactness of $\{\theta\in\mathcal{W}^{M_1}:|\theta-\theta_\star|\geq\varepsilon\}$, $\inf_{\theta\in\mathcal{W}^{M_1}:|\theta-\theta_\star|\geq\varepsilon}w(\theta)>w(\theta_\star)$ so that it is enough to check that $w(\theta_n)$ tends to $w(\theta_\star)$ as $n\to\infty$ to conclude that θ_n tends to $\theta_\star$.*

For $\varepsilon\in(0,M_1-w(\theta_\star))$, $\mathcal{W}_{w(\theta_\star)+\frac{\varepsilon}{3}}^{M_1}$ is a compact subset of $\mathcal{W}^{M_1}$ on which $\nabla w.h$ is strictly negative. By the second assertion in Lemma 3.0.5,

$$\begin{aligned}\exists\delta,\lambda,z>0,\ \forall\bar{\theta}\in\mathcal{W}_{w(\theta_\star)+\frac{\varepsilon}{3}}^{M_1},\ \forall\gamma\in[0,\delta],\ \forall\xi\in\mathbb{R}^{d_\theta}\text{ with }|\xi|\leq\lambda,&\\ \bar{\theta}+\gamma(h(\bar{\theta})+\xi)\in\Theta\text{ and }w(\bar{\theta}+\gamma(h(\bar{\theta})+\xi))\leq w(\bar{\theta})-z\gamma.&\quad(3.6)\end{aligned}$$

The assumptions ensure the existence of $n_\varepsilon\in\mathbb{N}$ such that for $n\geq n_\varepsilon$,

— *$\theta_n\in\mathcal{W}^M$,*

— *$\left|\sum_{j=n_\varepsilon}^n\gamma_{j+1}\xi_{j+1}\right|=\left|\sum_{j=n_\varepsilon}^n\gamma_{j+1}\xi_{j+1}^1+\sum_{j=n_\varepsilon}^n\gamma_{j+1}\xi_{j+1}^2 1_{\mathcal{W}^{M_1}}(\theta_j)\right|\leq\hat{\eta}_\lambda\wedge\hat{\eta}_{M_1-M}\wedge\hat{\eta}_{\varepsilon/3}$ where $\hat{\eta}_\varepsilon$ is introduced in the first statement of Lemma 3.0.5,*

— *and $\gamma_{n+1}\in[0,\delta\wedge\frac{\hat{\eta}_{\varepsilon/3}}{\sup_{\theta\in\mathcal{W}^{M_1}}|h(\theta)|}]$.*

Let $\bar{\theta}_{n_\varepsilon}=\theta_{n_\varepsilon}$ and

$$\text{for } n>n_\varepsilon,\bar{\theta}_n=\theta_n-\sum_{j=n_\varepsilon}^{n-1}\gamma_{j+1}\xi_{j+1}=\theta_{n_\varepsilon}+\sum_{j=n_\varepsilon}^{n-1}\gamma_{j+1}h(\theta_j).$$

By the first assertion in Lemma 3.0.5,

$$\forall n\geq n_\varepsilon,\ \bar{\theta}_n\in\Theta,\ |h(\bar{\theta}_n)-h(\theta_n)|\leq\lambda\text{ and }|w(\bar{\theta}_n)-w(\theta_n)|\leq(M_1-M)\wedge\varepsilon/3,\quad(3.7)$$

which implies $w(\bar\theta_n) \leq w(\theta_n) + M_1 - M$ so that $\bar\theta_n \in \mathcal{W}^{M_1}$. Moreover,

$$\forall n \geq n_\varepsilon,\ \bar\theta_{n+1} = \bar\theta_n + \gamma_{n+1}h(\theta_n)$$

so that, by the first assertion in Lemma 3.0.5 and since $|\gamma_{n+1}h(\theta_n)| \leq \hat\eta_{\varepsilon/3}$, $w(\bar\theta_{n+1}) \leq w(\bar\theta_n) + \frac{\varepsilon}{3}$. If $\bar\theta_n \in \mathcal{W}^{M_1}_{w(\theta_\star)+\frac{\varepsilon}{3}}$, writing $\bar\theta_{n+1} = \bar\theta_n + \gamma_{n+1}(h(\bar\theta_n) + \xi)$ where $\xi = h(\theta_n) - h(\bar\theta_n)$ is such that $|\xi| \leq \lambda$ and using (3.6), one obtains that $w(\bar\theta_{n+1}) \leq w(\bar\theta_n) - z\gamma_{n+1}$. Hence

$$\forall n \geq n_\varepsilon,\ w(\bar\theta_{n+1}) \leq w(\bar\theta_n) + 1_{\{w(\bar\theta_n)<w(\theta_\star)+\frac{\varepsilon}{3}\}}\frac{\varepsilon}{3} - 1_{\{w(\bar\theta_n)\geq w(\theta_\star)+\frac{\varepsilon}{3}\}}z\gamma_{n+1}. \tag{3.8}$$

Hence if $w(\bar\theta_k) \geq w(\theta_\star) + \frac{\varepsilon}{3}$ for $n_\varepsilon \leq k \leq n$, then $w(\bar\theta_{n+1}) \leq w(\bar\theta_{n_\varepsilon}) - z\sum_{k=n_\varepsilon}^n \gamma_{k+1}$ so that with $\sum_{k\geq n_\varepsilon}\gamma_{k+1} = \infty$, we deduce that $\tilde n_\varepsilon = \inf\{n \geq n_\varepsilon : w(\bar\theta_n) < w(\theta_\star) + \frac{\varepsilon}{3}\} < \infty$. By an easy inductive reasoning using (3.8), we obtain that for $n \geq \tilde n_\varepsilon$, $w(\bar\theta_n) < w(\theta_\star) + \frac{2\varepsilon}{3}$. With (3.7), we conclude that for $n \geq \tilde n_\varepsilon$, $w(\theta_n) < w(\theta_\star) + \varepsilon$.

Proof: Since $\nabla w.h$ is continuous, $\mathcal{L}\cap\mathcal{W}^{M_1}$ is compact and its image $\mathcal{D}$ by the continuous function w is also compact. For $\eta > 0$, let $\mathcal{D}_\eta = \{u \in \mathbb{R}, \exists v \in \mathcal{D} : |u - v| < \eta\}$. The set $\mathcal{D}_\eta$ is an open subset of $\mathbb{R}$ and therefore an at most countable union of disjoint open intervals. If $u \in \mathcal{D}_\eta$ then there exists $v \in \mathcal{D}$ such that $u \in (v-\eta, v+\eta)$ and $(v-\eta, v+\eta) \subset \mathcal{D}_\eta$ so that the length of each open interval in the union is at least 2η. Last, since $\sup_{u\in\mathcal{D}_\eta}|u| \leq \sup_{v\in\mathcal{D}}|w| + \eta$ with $\sup_{v\in\mathcal{D}}|v| < \infty$ by compacity of $\mathcal{D}$, $\mathcal{D}_\eta$ is bounded and $\mathcal{D}_\eta$ is a finite union of intervals.

Let $\eta > 0$. The set $w^{-1}(\mathcal{D}_{\eta/2})$ is open by continuity of w and $\mathcal{K} = \mathcal{W}^{M_1} \cap (\mathbb{R}^{d_\theta} \setminus w^{-1}(\mathcal{D}_{\eta/2}))$ is compact and such that $\nabla w.h$ is strictly negative on $\mathcal{K}$. By the second assertion in Lemma 3.0.5,

$$\begin{aligned}\exists\delta,\lambda,z > 0, \forall\bar\theta \in \mathcal{K},\ \forall\gamma \in [0,\delta],\ \forall\xi \in \mathbb{R}^{d_\theta} \text{ with } |\xi| \leq \lambda,\\ \bar\theta + \gamma(h(\bar\theta)+\xi) \in \Theta \text{ and } w(\bar\theta + \gamma(h(\bar\theta)+\xi)) \leq w(\bar\theta) - z\gamma.\end{aligned} \tag{3.9}$$

The assumptions ensure the existence of $n_\mathcal{K} \in \mathbb{N}$ such that for $n \geq n_\mathcal{K}$, $\theta_n \in \mathcal{W}^M$, $\gamma_{n+1} \in [0,\delta]$, $\left|\sum_{j=n_\mathcal{K}}^n \gamma_{j+1}\xi_{j+1}\right| = \left|\sum_{j=n_\mathcal{K}}^n \gamma_{j+1}\xi^1_{j+1} + \sum_{j=n_\mathcal{K}}^n \gamma_{j+1}\xi^2_{j+1}1_{\mathcal{W}^{M_1}}(\theta_j)\right| \leq \hat\eta_{M_1-M}\wedge\hat\eta_{\eta/4}\wedge\hat\eta_\lambda$ where $\hat\eta_\varepsilon$ is introduced in the first statement of Lemma 3.0.5. Let $\bar\theta_{n_\mathcal{K}} = \theta_{n_\mathcal{K}}$ and for $n > n_\mathcal{K}$, $\bar\theta_n = \theta_n - \sum_{j=n_\mathcal{K}}^{n-1}\gamma_{j+1}\xi_{j+1}$. By the first assertion in Lemma 3.0.5, for all

$$\forall n \geq n_\mathcal{K},\ \bar\theta_n \in \Theta \text{ and } |w(\bar\theta_n) - w(\theta_n)| \leq (M_1 - M) \wedge \eta/4, \tag{3.10}$$

which implies $w(\bar\theta_n) \leq w(\theta_n) + M_1 - M$ so that $\bar\theta_n \in \mathcal{W}^{M_1}$ and $|h(\bar\theta_n) - h(\theta_n)| \leq \lambda$. Moreover,

$$\forall n \geq n_\mathcal{K},\ \bar\theta_{n+1} = \bar\theta_n + \gamma_{n+1}h(\theta_n) = \bar\theta_n + \gamma_{n+1}(h(\bar\theta_n) + (h(\theta_n) - h(\bar\theta_n))). \tag{3.11}$$

If there exists $n_1 \geq n_\mathcal{K}$ such that for all $n \geq n_1$, $w(\bar\theta_n) \notin \mathcal{D}_{\eta/2}$, then for $n \geq n_1$, $\bar\theta_n \in \mathcal{K}$ so that, combining an inductive reasoning, (3.9) and (3.11), we get that for $n > n_1$, $w(\bar\theta_n) \leq w(\bar\theta_{n_1}) - z\sum_{j=n_1+1}^n \gamma_j$. Since $\sum_{j\geq n_1+1}\gamma_j = \infty$, this contradicts the positivity of the Lyapunov function w. Hence $(w(\bar\theta_n))_n$ is infinitely often in $\mathcal{D}_{\eta/2}$.

One has $\mathcal{D}_{\eta/2} = \bigcup_{k=1}^{k_{\eta/2}}(a_k, b_k)$ for $k_{\eta/2} \in \mathbb{N}^*$ and $-\infty = b_0 < a_1 < b_1 < a_2 < b_2 < \ldots < a_{k_{\eta/2}} < b_{k_{\eta/2}} < a_{k_{\eta/2}+1} = +\infty$. Let $k^\star_{\eta/2} \in \{1,\ldots,k_{\eta/2}\}$ be such that $(w(\bar\theta_n))_n$ is infinitely often in $(a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}})$. We now check that $w(\bar\theta_n)$ is in $(a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}} + \frac{\eta}{4})$ for n large enough. Let $0 < \varepsilon < (a_{k^\star_{\eta/2}+1} - b_{k^\star_{\eta/2}}) \wedge (a_{k^\star_{\eta/2}} - b_{k^\star_{\eta/2}-1}) \wedge \frac{\eta}{4}$ and $n_1 \geq n_{\mathcal{K}}$ be such that for all $n \geq n_1$, $\gamma_{n+1} \sup_{\theta \in \mathcal{W}^{M_1}} |h(\theta)| \leq \hat\eta_\varepsilon$. Combining the first assertion in Lemma 3.0.5, (3.9) and (3.11), we get that

$$\forall n \geq n_1,\ w(\bar\theta_{n+1}) \leq w(\bar\theta_n) - z\gamma_{n+1} 1_{\mathcal{K}}(\bar\theta_n) + \varepsilon 1_{\mathcal{D}_{\eta/2}}(w(\bar\theta_n)).$$

Let $n_2 \geq n_1$ be such that $w(\bar\theta_{n_2}) \in (a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}})$. Since $\varepsilon < a_{k^\star_{\eta/2}+1} - b_{k^\star_{\eta/2}}$, one easily checks that for all $n \geq n_2$, $w(\bar\theta_n) < b_{k^\star_{\eta/2}} + \varepsilon$ by induction and considering the two cases $w(\bar\theta_n) < b_{k^\star_{\eta/2}}$ and $w(\bar\theta_n) \in [b_{k^\star_{\eta/2}}, a_{k^\star_{\eta/2}+1}) \subset \mathcal{K}$. In the same way, since $\varepsilon < a_{k^\star_{\eta/2}} - b_{k^\star_{\eta/2}-1}$, if $w(\bar\theta_{n_3}) \leq a_{k^\star_{\eta/2}}$ for some $n_3 \geq n_2$, then for all $n \geq n_3$, $w(\bar\theta_n) \leq a_{k^\star_{\eta/2}}$ which contradicts the fact that $(w(\bar\theta_n))_n$ is infinitely often in $(a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}})$. Hence for all $n \geq n_2$, $w(\bar\theta_n) \in (a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}} + \frac{\eta}{4})$.

Since $\mathcal{D}_{\eta/2} \subset \mathcal{D}_\eta$ where $\mathcal{D}_\eta$ is a finite union of disjoint intervals, $(a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}})$ is included in one of these intervals that we denote $(a_{k^\star_\eta}, b_{k^\star_\eta})$. For $u \in (a_{k^\star_{\eta/2}}, b_{k^\star_{\eta/2}})$ there exists $v \in \mathcal{D}$ such that $|v-u| < \eta/2$ and if $u' \in \mathbb{R}$ is such that $|u'-u| \leq \eta/2$, then $|v-u'| \leq |v-u| + |u-u'| < \eta$ so that $u' \in \mathcal{D}_\eta$. Hence $(a_{k^\star_{\eta/2}} - \eta/2, b_{k^\star_{\eta/2}} + \eta/2) \subset (a_{k^\star_\eta}, b_{k^\star_\eta})$. With (3.10), we deduce that for $n \geq n_2$, $w(\theta_n) \in (a_{k^\star_\eta}, b_{k^\star_\eta})$.

Therefore any limit point of the sequence $(w(\theta_n))_n$ belongs to the interval $\bigcap_{\eta>0}(a_{k^\star_\eta}, b_{k^\star_\eta})$ which is included in $\bigcap_{\eta>0}\mathcal{D}_\eta$. For $u \in \bigcap_{\eta>0}\mathcal{D}_\eta$, for each $n \in \mathbb{N}^*$, there exists $v_n \in \mathcal{D}$ such that $|u - v_n| \leq 1/n$. By compacity of $\mathcal{D}$, a subsequence of $(v_n)_n$ converges to some $v_\infty \in \mathcal{D}$ so that $u = v_\infty$. Hence $\bigcap_{\eta>0}\mathcal{D}_\eta \subset \mathcal{D}$. Since $\mathcal{D}$ has an empty interior, the interval $\bigcap_{\eta>0}(a_{k^\star_\eta}, b_{k^\star_\eta})$ is a singleton $\{v^\star\}$ and the sequence $w(\theta_n)_n$ converges to $v^\star$ as $n \to \infty$.

Let for $\eta > 0$, $[\mathcal{L} \cap \mathcal{W}^{M_1}]_{\eta/2} = \{\theta \in \mathcal{W}^{M_1} : \exists \hat\theta \in \mathcal{L} \cap \mathcal{W}^{M_1} : |\hat\theta - \theta| < \eta/2\}$. We now suppose that η is small enough enough so that, by the first statement in Lemma 3.0.5, $\forall \theta \in \mathbb{R}^{d_\theta}$ such that $\exists \hat\theta \in \mathcal{W}^{M_1}$ with $|\hat\theta - \theta| \leq \eta/2$, $\theta \in \Theta$. Then $[\mathcal{L} \cap \mathcal{W}^{M_1}]_{\eta/2} = \{\theta \in \mathbb{R}^{d_\theta} : \exists \hat\theta \in \mathcal{L} \cap \mathcal{W}^{M_1} : |\hat\theta - \theta| < \eta/2\}$ is an open subset of $\mathbb{R}^{d_\theta}$ and $\mathcal{K}_\eta = \mathcal{W}^{M_1} \setminus [\mathcal{L} \cap \mathcal{W}^{M_1}]_{\eta/2}$ is compact as a closed subset of the compact $\mathcal{W}^{M_1}$. Reasoning like in the beginning of the proof and using the convergence of $w(\theta_n)$ to $v^\star$, one obtains the existence of $z > 0$, $n_1 \geq n_{\mathcal{K}_\eta}$ such that if for $n \geq n_{\mathcal{K}_\eta}$, $\bar\theta_n = \theta_n - \sum_{j=n_{\mathcal{K}_\eta}}^{n-1} \gamma_{j+1}\xi_{j+1}$, then $\bar\theta_n \in \mathcal{W}^{M_1}$ for all $n \geq n_{\mathcal{K}_\eta}$ and

$$\forall n \geq n_1,\ |w(\theta_n) - w(\bar\theta_n)| \vee |w(\theta_n) - v^\star| \leq \frac{\eta z}{16 \sup_{\theta \in \mathcal{W}^{M_1}} |h(\theta)|},\ \sup_{k \geq n} \left| \sum_{j=n}^{k} \gamma_j \xi_j \right| \leq \frac{\eta}{4}$$
$$\text{and } w(\bar\theta_{n+1}) \leq w(\bar\theta_n) - z\gamma_{n+1} 1_{\mathcal{K}_\eta}(\bar\theta_n) + \varepsilon 1_{[\mathcal{L} \cap \mathcal{W}^{M_1}]_\eta}(\bar\theta_n).$$

For $n \geq n_1$, let $\tau(n) = \inf\{k \geq n : \bar\theta_k \in [\mathcal{L} \cap \mathcal{W}^{M_1}]_{\eta/2}\}$. Since for $n \leq k \leq \tau(n)$, $w(\bar\theta_k) \leq w(\bar\theta_n) - z\sum_{j=n}^{k-1}\gamma_{j+1}$ and $\sum_{j \geq n}\gamma_{j+1} = \infty$, the positivity of w implies that $\tau(n) < \infty$. Moreover,

$$\begin{aligned}
\sum_{j=n}^{\tau(n)-1} \gamma_{j+1} &\leq \frac{1}{z}\left(w(\bar\theta_n) - w(\bar\theta_{\tau(n)})\right) \\
&\leq \frac{1}{z}\left(|w(\bar\theta_n) - w(\theta_n)| + |w(\theta_n) - v^\star| + |v^\star - w(\theta_{\tau(n)})| + |w(\theta_{\tau(n)}) - w(\bar\theta_{\tau(n)})|\right) \\
&\leq \frac{\eta}{4 \sup_{\theta \in \mathcal{W}^{M_1}} |h(\theta)|}.
\end{aligned}$$

Therefore, for $n \geq n_1$,

$$|\theta_n - \theta_{\tau_n}| = \left| \sum_{j=n}^{\tau(n)-1} \gamma_{j+1}(h(\theta_j) + \xi_{j+1}) \right| \leq \sup_{\theta \in \mathcal{W}^{M_1}} |h(\theta)| \sum_{j=n}^{\tau(n)-1} \gamma_{j+1} + \left| \sum_{j=n+1}^{\tau(n)} \gamma_j \xi_j \right| \leq \frac{\eta}{4} + \frac{\eta}{4} = \frac{\eta}{2}.$$

Since $\theta_{\tau(n)} \in [\mathcal{L} \cap \mathcal{W}^{M_1}]_{\eta/2}$, we conclude that for $n \geq n_1$, $\inf_{\theta \in \mathcal{L} \cap \mathcal{W}^{M_1}} |\theta - \theta_n| \leq \eta$.

∎

Example 3.0.8. *Let us turn back to the Robbins-Monro example :*

$$\theta_{n+1} = \theta_n + \gamma_{n+1} Y_{n+1}$$

where $(Y_n)_n$ is a sequence of square integrable random variables such that for $\mathcal{F}_n = \sigma(\theta_0, Y_1, \ldots, Y_n)$, $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = h(\theta_n)$ and $(\gamma_n)_{n\geq 1}$ a deterministic sequence of positive numbers. Let $\xi_{n+1} = Y_{n+1} - h(\theta_n) = Y_{n+1} - \mathbb{E}[Y_{n+1}|\mathcal{F}_n]$ so that $\theta_{n+1} = \theta_n + \gamma_{n+1}(h(\theta_n) + \xi_{n+1})$. We assume that

- *$h : \mathbb{R}^{d_\theta} \to \mathbb{R}$ is continuous and such that $h(\theta_\star) = 0$, $h(\theta).(\theta_\star - \theta) > 0$ for all $\theta \neq \theta_\star$ and $\kappa_h := \sup_{\theta \in \mathbb{R}^{d_\theta}} \frac{|h(\theta)|^2}{1+|\theta-\theta_\star|^2} < \infty$,*
- *$\sum_{n\geq 1} \gamma_n = \infty$ and $\sum_{n \geq 1} \gamma_n^2 < \infty$,*
- *$\forall n \in \mathbb{N}$, $\mathbb{E}\left[|\xi_{n+1}|^2|\mathcal{F}_n\right] \leq s^2(\theta_n)$ for some function $s : \mathbb{R}^{d_\theta} \to \mathbb{R}$ such that $\kappa_s := \sup_{\theta \in \mathbb{R}^{d_\theta}} \frac{s^2(\theta)}{1+|\theta-\theta_\star|^2} < \infty$,*
- *$\mathbb{E}[|\theta_0 - \theta_\star|^2] < \infty$, hypothesis satisfied when θ_0 is deterministic.*

The square integrability of $(\theta_0 - \theta_\star)$ and the one of Y_{k+1} for all $k \in \mathbb{N}$ ensures that for all $n \in \mathbb{N}$, $\theta_n - \theta_\star = \theta_0 - \theta_\star + \sum_{k=0}^{n-1} \gamma_{k+1} Y_{k+1}$ is square integrable. Let us check that $\sup_{n \in \mathbb{N}} \mathbb{E}[|\theta_n - \theta_\star|^2] < \infty$.

$$\begin{aligned}
|\theta_{n+1} - \theta_\star|^2 =& |\theta_n - \theta_\star + \gamma_{n+1}(h(\theta_n) + \xi_{n+1})|^2 \\
=& |\theta_n - \theta_\star|^2 + 2\gamma_{n+1}(\theta_n - \theta_\star).h(\theta_n) + \gamma_{n+1}^2 |h(\theta_n)|^2 \\
&+ 2\gamma_{n+1}(\theta_n - \theta_\star + \gamma_{n+1} h(\theta_n)).\xi_{n+1} + \gamma_{n+1}^2 |\xi_{n+1}|^2 \\
\leq& |\theta_n - \theta_\star|^2 + \gamma_{n+1}^2 |h(\theta_n)|^2 + 2\gamma_{n+1}(\theta_n - \theta_\star + \gamma_{n+1} h(\theta_n)).\xi_{n+1} + \gamma_{n+1}^2 |\xi_{n+1}|^2,
\end{aligned}$$

using that $h(\theta_n).(\theta_\star - \theta_n) \geq 0$ for the inequality. Since $\theta_n - \theta_\star + \gamma_{n+1} h(\theta_n)$ is $\mathcal{F}_n$-measurable and $\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = 0$, the expectation of the second term in the right-hand side vanishes so that setting $\kappa = \kappa_h + \kappa_s$, we obtain

$$\begin{aligned}
\mathbb{E}[|\theta_{n+1} - \theta_\star|^2] &\leq \mathbb{E}[|\theta_n - \theta_\star|^2] + \gamma_{n+1}^2 \mathbb{E}[|h(\theta_n)|^2 + s^2(\theta_n)] \\
&\leq \mathbb{E}[|\theta_n - \theta_\star|^2] + \kappa \gamma_{n+1}^2 (1 + \mathbb{E}[|\theta_n - \theta_\star|^2]) \\
&= (1 + \kappa \gamma_{n+1}^2) \mathbb{E}[|\theta_n - \theta_\star|^2] + \kappa \gamma_{n+1}^2.
\end{aligned}$$

Iterating this inequality, we deduce that for $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}[|\theta_n - \theta_\star|^2] &\leq \mathbb{E}[|\theta_0 - \theta_\star|^2] \prod_{l=1}^{n} (1 + \kappa \gamma_l^2) + \kappa \sum_{k=1}^{n} \gamma_k^2 \prod_{l=k+1}^{n} (1 + \kappa \gamma_l^2) \\
&\leq \left(\mathbb{E}[|\theta_0 - \theta_\star|^2] + \kappa \sum_{k \geq 1} \gamma_k^2 \right) e^{\sum_{l \geq 1} \ln(1 + \kappa \gamma_l^2)} \leq \left(\mathbb{E}[|\theta_0 - \theta_\star|^2] + \kappa \sum_{k \geq 1} \gamma_k^2 \right) e^{\kappa \sum_{l \geq 1} \gamma_l^2},
\end{aligned}$$

with the right-hand side finite and not depending on n*. The process* $(m_k = \sum_{j=1}^k \gamma_j \xi_j)_{k\in\mathbb{N}}$ *(convention* $m_0 = 0$*) is a* $\mathcal{F}_k$*-martingale. Since for* $k \geq 1$*,*

$$\mathbb{E}\left[|m_k|^2\right] = \sum_{j=1}^k \gamma_j^2 \mathbb{E}\left[|\xi_j|^2\right] \leq \sum_{j=1}^k \gamma_j^2 \mathbb{E}\left[s^2(\theta_{j-1})\right]$$

$$\leq \kappa_s \sum_{j=1}^k \gamma_j^2 \left(1 + \mathbb{E}[|\theta_{j-1} - \theta_\star|^2]\right) \leq \kappa_s \left(1 + \sup_{j\geq 1} \mathbb{E}[|\theta_{j-1} - \theta_\star|^2]\right) \sum_{j=1}^k \gamma_j^2,$$

this martingale is bounded in L^2 *and therefore a.s. convergent. Since for* $k \geq n \geq 1$*,* $\left|\sum_{j=n}^k \gamma_j \xi_j\right| = |m_k - m_{n-1}|$*, a.s.* $\lim_{n\to\infty} \max_{k\geq n} \left|\sum_{j=n}^k \gamma_j \xi_j\right| = 0$*.*

We set $\Theta = \mathbb{R}^{d_\theta}$*,* $w(\theta) = |\theta - \theta_\star|^2$ *so that* $\nabla w(\theta) = 2(\theta - \theta_\star)$ *and* $\mathcal{L} := \{\theta \in \mathbb{R}^{d_\theta} : \nabla w.h(\theta) = 0\} = \{\theta_\star\}$*. For all* $M_1 \in \mathbb{N}^* \setminus \{1\}$*,* $\mathcal{W}^{M_1}$ *is compact and* $\nabla w.h$ *negative on* $\mathcal{W}_1^{M_1}$*. Since, by the Fatou Lemma,* $\mathbb{E}[\liminf_{n\to\infty} |\theta_n - \theta_\star|^2] \leq \liminf_{n\to\infty} \mathbb{E}[|\theta_n - \theta_\star|^2] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[|\theta_n - \theta_\star|^2] < \infty$*, one has* $\mathbb{P}(\liminf_{n\to\infty} |\theta_n - \theta_\star|^2 < \infty) = 1$ *and therefore a.s. there exists an integer* $M_1 \geq 2$ *such that* θ_n *is infinitely often in* $\mathcal{W}^{\frac{1+M_1}{2}}$ *and, by Corollary 3.0.3 (*γ_n *tends to* 0 *as* $n \to \infty$ *since* $\sum_{n\geq 1} \gamma_n^2 < \infty$*),* θ_n *remains in* $\mathcal{W}^{\frac{1+3M_1}{4}}$ *for* n *large enough. Since for all integers* $M_1 \geq 2$*,* $\mathcal{L} \cap \mathcal{W}^{M_1} = \{\theta_\star\}$ *and the image of this set by* w *is equal to* $\{0\}$*, Theorem 3.0.6 implies that a.s.,* θ_n *converges to* $\theta_\star$ *as* $n \to \infty$*.*

Chapitre 4

Adaptive Metropolis Hastings algorithms

4.1 the Self-Healing Umbrella Sampling algorithm

4.1.1 Framework

Let $\pi(dx) = \eta(x)\lambda(dx)$ where $\eta : E \to \mathbb{R}_+$ is a probability density with respect to the reference measure $\lambda : \int_E \eta(x)\lambda(dx) = 1$. We consider

— a partition $(E_i)_{1\leq i\leq \mathcal{I}}$ of the state space E into measurable subsets,

— $I : E \to \{1, \ldots, \mathcal{I}\}$ defined by $I(x) = \sum_{i=1}^{\mathcal{I}} i 1_{E_i}(x)$,

— $\theta_\star = (\theta_\star(i))_{1\leq i\leq \mathcal{I}}$ with $\theta_\star(i) = \pi(E_i) = \int_{E_i} \eta(x)\lambda(dx)$. We suppose that for each $i \in \{1, \ldots, \mathcal{I}\}$, $\theta_\star(i) > 0$,

— $\Theta = \{\theta = (\theta(1), \ldots, \theta(\mathcal{I})) \in]0, 1[^{\mathcal{I}} \text{ s.t. } \sum_{i=1}^{\mathcal{I}} \theta(i) = 1\}$.

The metastability is the fact that regions with high probability under π are separated by regions with very low probability so that moving from one region with high probability to another will be very difficult for a Metropolis-Hastings algorithm targeting π. In the above framework, the metastability can be characterized by the fact that the probability measure $\theta_\star$ on $\{1, \ldots, \mathcal{I}\}$ is very far from uniformity.

For $\theta \in \Theta$, let $\pi_\theta(dx) = \eta_\theta(x)\lambda(dx)$ where

$$\eta_\theta(x) = \frac{1}{Z_\theta} \sum_{i=1}^{\mathcal{I}} \frac{\eta(x)}{\theta(i)} 1_{E_i}(x) = \frac{1}{Z_\theta} \times \frac{\eta(x)}{\theta(I(x))}$$

and the normalization constant Z_θ is such that π_θ is a probability measure i.e. $1 = \frac{1}{Z_\theta} \sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta(i)}$ so that $Z_\theta = \sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta(i)}$.

Exercice 4.1.1. *Find $\theta \in \Theta$ such that $\pi_\theta = \pi$.*

We have $Z_{\theta_\star} = \mathcal{I}$, so that $\pi_{\theta_\star}(E_i) = \frac{1}{Z_{\theta_\star}} \times \frac{\theta_\star(i)}{\theta_\star(i)} = \frac{1}{\mathcal{I}}$. The subsets E_i in the partition are equiprobable under $\pi_{\theta_\star}$. Targeting the measure $\pi_{\theta_\star}$ will drastically reduce the metastability. To overcome the fact that $\theta_\star$ is unknown, adaptive algorithms construct a couple (X_n, θ_n) at time n with θ_n an estimator of $\theta_\star$ computed as a function of the past values $(X_0, \theta_0, X_1, \theta_1, \ldots, X_n)$ and sample X_{n+1} given $(X_0, \theta_0, X_1, \theta_1, \ldots, X_n, \theta_n)$ according to a Metropolis-Hastings kernel targeting π_{θ_n}.

Nevertheless, when sampling $X \sim \pi_{\theta_\star}$, to compute $\int_E f(x)\pi(dx)$ for some function $f : E \to \mathbb{R}$ measurable and nonnegative or integrable with respect to π, one needs to perform importance sampling according to the formula

$$\int_E f(x)\pi(dx) = \mathbb{E}\left[f(X)\mathcal{I}\sum_{i=1}^{\mathcal{I}} 1_{E_i}(X)\theta_\star(i)\right].$$

Since $\mathbb{P}(X \in E_i) = \frac{1}{\mathcal{I}}$, the variance of the importance sampling factor

$$\mathrm{Var}\left[\mathcal{I}\sum_{i=1}^{\mathcal{I}} 1_{E_i}(X)\theta_\star(i)\right] = \mathcal{I}\sum_{i=1}^{\mathcal{I}}\theta_\star^2(i) - \left(\sum_{i=1}^{\mathcal{I}}\theta_\star(i)\right)^2 = \mathcal{I}\sum_{i=1}^{\mathcal{I}}\theta_\star^2(i) - 1$$

may be large if $\theta_\star$ is far from uniformity (which is the case in presence of metastability). In order to reduce this variance, it is sensible to mitigate the strength of the biasing by introducing some parameter $a \in [0,1]$ and setting $\pi_\theta^a(dx) = \eta_\theta^a(x)\lambda(dx)$ where

$$\eta_\theta^a(x) = \frac{1}{Z_\theta^a}\sum_{i=1}^{\mathcal{I}}\frac{\eta(x)}{\theta^a(i)}1_{E_i}(x) = \frac{1}{Z_\theta^a}\times\frac{\eta(x)}{\theta^a(I(x))} \tag{4.1}$$

and $Z_\theta^a = \sum_{i=1}^{\mathcal{I}}\frac{\theta_\star(i)}{\theta^a(i)}$ so that $Z_{\theta_\star}^a = \sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)$. We have $\pi_\theta^0 = \pi$ and $\pi_\theta^1 = \pi_\theta$. For $X \sim \pi_{\theta_\star}^a$,

$$\int_E f(x)\pi(dx) = \mathbb{E}\left[f(X)Z_{\theta_\star}^a\sum_{i=1}^{\mathcal{I}} 1_{E_i}(X)\theta_\star^a(i)\right].$$

Since $\mathbb{P}(X \in E_i) = \frac{\theta_\star^1(i)}{Z_{\theta_\star}^a\theta_\star^a(i)} = \frac{\theta_\star^{1-a}(i)}{Z_{\theta_\star}^a}$, we have

$$\begin{aligned}\mathrm{Var}\left[Z_{\theta_\star}^a\sum_{i=1}^{\mathcal{I}} 1_{E_i}(X)\theta_\star^a(i)\right] &= (Z_{\theta_\star}^a)^2\sum_{i=1}^{\mathcal{I}}\theta_\star^{2a}(i)\times\frac{\theta_\star^{1-a}(i)}{Z_{\theta_\star}^a} - \left(\sum_{i=1}^{\mathcal{I}}\theta_\star(i)\right)^2 = Z_{\theta_\star}^a\sum_{i=1}^{\mathcal{I}}\theta_\star^{1+a}(i) - 1\\ &= \sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)\sum_{i=1}^{\mathcal{I}}\theta_\star^{1+a}(i) - 1.\end{aligned}$$

The next lemma shows that this variance is a non-decreasing function of the strength parameter a of the biasing.

Lemma 4.1.2. *The function $[0,1] \ni a \mapsto f(a) = \sum_{i=1}^{\mathcal{I}}\theta_\star^{1+a}(i)\sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)$ is non-decreasing.*

Proof:

$$\begin{aligned}\partial_a\ln(f(a)) &= \partial_a\ln\left(\sum_{i=1}^{\mathcal{I}}\theta_\star^{1+a}(i)\right) + \partial_a\ln\left(\sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)\right)\\ &= \frac{\sum_{i=1}^{\mathcal{I}}\ln(\theta_\star(i))\theta_\star^{1+a}(i)}{\sum_{i=1}^{\mathcal{I}}\theta_\star^{1+a}(i)} - \frac{\sum_{i=1}^{\mathcal{I}}\ln(\theta_\star(i))\theta_\star^{1-a}(i)}{\sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)} = g(1+a) - g(1-a)\end{aligned}$$

where for $b \in \mathbb{R}$, $g(b) = \frac{\sum_{i=1}^{\mathcal{I}} \ln(\theta_\star(i))\theta_\star^b(i)}{\sum_{i=1}^{\mathcal{I}} \theta_\star^b(i)}$ satisfies

$$g'(b) = \frac{\sum_{i=1}^{\mathcal{I}} \ln^2(\theta_\star(i))\theta_\star^b(i)}{\sum_{i=1}^{\mathcal{I}} \theta_\star^b(i)} - \left(\frac{\sum_{i=1}^{\mathcal{I}} \ln(\theta_\star(i))\theta_\star^b(i)}{\sum_{i=1}^{\mathcal{I}} \theta_\star^b(i)}\right)^2.$$

The right-hand side is non-negative as it can be interpreted as a variance (or by Cauchy-Schwarz inequality, $\left(\sum_{i=1}^{\mathcal{I}} \ln(\theta_\star(i))\theta_\star^b(i)\right)^2 \leq \sum_{i=1}^{\mathcal{I}} \ln^2(\theta_\star(i))\theta_\star^b(i) \sum_{i=1}^{\mathcal{I}} \theta_\star^b(i)$). Hence $\partial_a \ln(f(a)) \geq 0$. ∎

Hence the smaller a, the smaller is the variance due to the importance sampling factor whereas the larger a the weaker is the metastability of $\pi^a_{\theta_\star}$. Therefore some tradeoff between these opposite effects should be found.

4.1.2 the SHUS$^\alpha_a$ algorithm with parameters $a \in [0,1]$ and $\alpha \in (1/2, 1]$

Algorithm 1 (SHUS$^\alpha_a$)**.** *For a (possibly random) initial condition $(\tilde{\theta}_0, X_0)$ in $(\mathbb{R}_+^*)^{\mathcal{I}} \times E$, the SHUS$^\alpha_a$ algorithm consists in iterating the following three steps over $n \geq 0$: given $(\tilde{\theta}_n, X_n) \in (\mathbb{R}_+^*)^{\mathcal{I}} \times E$,*

— *Normalize $\tilde{\theta}_n$:*

$$\theta_n := \frac{\tilde{\theta}_n}{S_n} \in \Theta \text{ where } S_n := \sum_{i=1}^{\mathcal{I}} \tilde{\theta}_n(i), \tag{4.2}$$

— *Sample X_{n+1} according to the distribution $P^a_{\theta_n}(X_n, \cdot)$,*

— *Update, for all $i \in \{1, \dots, \mathcal{I}\}$,*

$$\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i) + \frac{\gamma}{g_\alpha(S_n)} S_n\, \theta_n^a(i)\, 1_{E_i}(X_{n+1}). \tag{4.3}$$

Here $\gamma > 0$ is some constant and

$$\text{for } s > 0,\ g_\alpha(s) = \begin{cases} s \text{ if } \alpha = 1 \\ (\ln(1+s))^{\frac{\alpha}{1-\alpha}} \text{ if } \alpha \in (1/2, 1) \end{cases}.$$

In the updating step $\tilde{\theta}_{n+1}(I(X_{n+1})) > \tilde{\theta}_n(I(X_{n+1}))$ while $\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i) = 0$ for $i \neq I(X_{n+1})$. Hence $S_{n+1} = \sum_{j=1}^{\mathcal{I}} \tilde{\theta}_{n+1}(j) > \sum_{j=1}^{\mathcal{I}} \tilde{\theta}_n(j) = S_n$ and, after normalization, $\theta_{n+1}(i) < \theta_n(i)$ for $i \neq I(X_{n+1})$ and $\theta_{n+1}(I(X_{n+1})) > \theta_n(I(X_{n+1}))$. Moreover $\sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\tilde{\theta}^a_{n+1}(j)} < \sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\tilde{\theta}^a_n(j)}$ and since $\pi^a_{\theta_{n+1}}(E_i) = \frac{\theta_\star(i)}{\tilde{\theta}^a_{n+1}(i) \sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\tilde{\theta}^a_{n+1}(j)}}$,

$$\text{for } i \neq I(X_{n+1}),\ \pi^a_{\theta_{n+1}}(E_i) = \frac{\theta_\star(i)}{\tilde{\theta}^a_n(i) \sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\tilde{\theta}^a_{n+1}(j)}} > \frac{\theta_\star(i)}{\tilde{\theta}^a_n(i) \sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\tilde{\theta}^a_n(j)}} = \pi^a_{\theta_n}(E_i)$$

$$\pi^a_{\theta_{n+1}}(E_{I(X_{n+1})}) = 1 - \sum_{i \neq I(X_{n+1})} \pi^a_{\theta_{n+1}}(E_i) < 1 - \sum_{i \neq I(X_{n+1})} \pi^a_{\theta_n}(E_i) = \pi^a_{\theta_n}(E_{I(X_{n+1})}).$$

The weight of the just visited stratum in the target probability measure $\pi^a_{\theta_{n+1}}$ at the next Metropolis-Hastings step gets smaller. In contrast, the weights of the non visited strata increase so that they get more easily visited in the future. This way, the algorithm struggles with the metastability.

The function g_α will determine the asymptotic behaviour of the step of the algorithm. Since $S_n = \frac{\tilde{\theta}_n(i)}{\theta_n(i)}$, setting $\gamma_{n+1} = \frac{\gamma}{g_\alpha(S_n)}$, the udating step also writes

$$\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i)\left(1 + \frac{\gamma}{g_\alpha(S_n)}\theta_n^{a-1}(i)1_{E_i}(X_{n+1})\right) = \tilde{\theta}_n(i)\left(1 + \gamma_{n+1}\theta_n^{a-1}(i)1_{E_i}(X_{n+1})\right).$$

When updating is performed according to this formula with $(\gamma_n)_{n\geq 1}$ a sequence of positive steps given in advance and the two other steps unchanged, the algorithm is called Wang-Landau.

Algorithm 2 (WL$_a$)**.** *For a (possibly random) sequence $(\gamma_n)_{n\geq 1}$ and initial conditions $(\tilde{\theta}_0, X_0)$ in $(\mathbb{R}_+^*)^{\mathcal{I}} \times E$, the WL_a algorithm consists in iterating the following three steps over $n \geq 0$: given $(\tilde{\theta}_n, X_n) \in (\mathbb{R}_+^*)^{\mathcal{I}} \times E$,*

— *Compute the normalizing constant S_n and the probability measure θ_n on $\{1, \ldots, \mathcal{I}\}$:*

$$S_n = \sum_{i=1}^{\mathcal{I}} \tilde{\theta}_n(i), \qquad \theta_n = \frac{\tilde{\theta}_n}{S_n} \in \Theta. \tag{4.4}$$

— *Sample X_{n+1} according to the distribution $P_{\theta_n}^a(X_n, \cdot)$.*

— *Compute, for all $i \in \{1, \ldots \mathcal{I}\}$,*

$$\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i)\left(1 + \gamma_{n+1}\,\theta_n^{a-1}(i)1_{E_i}(X_{n+1})\right). \tag{4.5}$$

Since $S_{n+1} = S_n(1+\gamma_{n+1}\theta_n^a(I(X_{n+1})))$, the stochastic approximation procedure corresponding to the updating and normalization steps writes

$$\theta_{n+1}(i) = \frac{\tilde{\theta}_{n+1}(i)}{S_{n+1}} = \frac{\tilde{\theta}_n(i)}{S_n} \times \frac{1+\gamma_{n+1}\theta_n^{a-1}(i)1_{E_i}(X_{n+1})}{1+\gamma_{n+1}\theta_n^a(I(X_{n+1}))} = \theta_n(i)\frac{1+\gamma_{n+1}\theta_n^{a-1}(i)1_{E_i}(X_{n+1})}{1+\gamma_{n+1}\theta_n^a(I(X_{n+1}))}. \tag{4.6}$$

Typically, to ensure convergence to $\theta_\star$ of the sequence $(\theta_n)_n$ generated by this procedure, one needs that $\sum_{n\geq 1}\gamma_n = +\infty$ and $\sum_{n\geq 1}\gamma_n^2 < \infty$ (see Chapter 3). The Wang-Landau algorithm is adaptive in the sense that $\theta_\star$ is learned during the computation trough Metropolis-Hastings steps targeting the probability measure corresponding to the current value θ_n. The SHUS$_a^\alpha$ algorithm is doubly adaptive in the sense that the sequence of steps $\frac{\gamma}{g_\alpha(S_n)}$ is computed automatically so that $0 < \liminf_{n\to+\infty} n^\alpha \frac{\gamma}{g_\alpha(S_n)}$ and $\limsup_{n\to+\infty} n^\alpha \frac{\gamma}{g_\alpha(S_n)} < +\infty$ (which, since $\alpha \in (1/2, 1]$, ensures $\sum_{n\geq 1}\frac{\gamma}{g_\alpha(S_n)} = +\infty$ and $\sum_{n\geq 1}\left(\frac{\gamma}{g_\alpha(S_n)}\right)^2 < +\infty$).

4.1.3 Intuition about the choice of g_α

We have

$$S_{n+1} = S_n + \frac{\gamma}{g_\alpha(S_n)}S_n\theta_n^a(I(X_{n+1})), \quad \text{for } n \in \mathbb{N}. \tag{4.7}$$

In the limit $n \to \infty$, we expect that θ_n converges to $\theta_\star$ and the law of X_{n+1} to $\pi_{\theta_\star}^a$ so that $\mathbb{P}(X_{n+1} \in E_i)$ converges to

$$\pi_{\theta_\star}^a(E_i) = \frac{\pi(E_i)}{Z_{\theta_\star}^a\theta_\star^a(i)} = \frac{\theta_\star^{1-a}(i)}{Z_{\theta_\star}^a} \text{ where } Z_{\theta_\star}^a = \sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i).$$

One expects in turn that $\mathbb{E}[\theta_n^a(I(X_{n+1}))]$ goes to

$$\sum_{i=1}^{\mathcal{I}} \theta_\star^a(i) \frac{\theta_\star^{1-a}(i)}{Z_{\theta_\star}^a} = \frac{1}{Z_{\theta_\star}^a} \sum_{i=1}^{\mathcal{I}} \theta_\star(i) = \frac{1}{Z_{\theta_\star}^a}.$$

So one may expect that $(S_n)_{n\in\mathbb{N}}$ asymptotically behaves like $(s(n))_{n\in\mathbb{N}}$ where $s(t)$ solves the Ordinary Differential Equation

$$\frac{ds(t)}{dt} = \frac{\gamma s(t)}{Z_{\theta_\star}^a g_\alpha(s(t))}.$$

if $\alpha = 1$, then $g_\alpha(s) = s$ so that $\frac{ds(t)}{dt} = \frac{\gamma}{Z_{\theta_\star}^a}$, $s(t) = s(0) + \frac{\gamma t}{Z_{\theta_\star}^a}$ and $\gamma_n = \frac{\gamma}{S_n}$ should be equivalent to $\frac{Z_{\theta_\star}^a}{n}$ as $n \to \infty$.

if $\alpha \in (1/2, 1)$, we introduce a new approximation by replacing $g_\alpha(s)$ by $(\ln(s))^{\frac{\alpha}{1-\alpha}}$ in the ODE :

$$\frac{d\ln(s(t))}{dt} = \frac{\gamma}{Z_{\theta_\star}^a}(\ln(s))^{-\frac{\alpha}{1-\alpha}} \text{ so that } \frac{d}{dt}\left((\ln(s(t)))^{\frac{1}{1-\alpha}}\right) = \frac{\gamma}{(1-\alpha)Z_{\theta_\star}^a}.$$

We deduce that as $t \to \infty$, $\ln(s(t)) \sim \left(\frac{\gamma t}{(1-\alpha)Z_{\theta_\star}^a}\right)^{1-\alpha}$ and $\gamma_n = \frac{\gamma}{g_\alpha(S_n)} \sim \frac{\gamma^{1-\alpha}((1-\alpha)Z_{\theta_\star}^a)^\alpha}{n^\alpha}$ as $n \to \infty$.

4.1.4 Intuition on the convergence when $\alpha = 1$

Let us suppose that θ_n converges to some limit θ_∞ and check that θ_∞ should then be equal to $\theta_\star$. We then expect that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n 1_{E_i}(X_k) = \pi_{\theta_\infty}^a(E_i) = \frac{\theta_\star(i)}{Z_{\theta_\infty}^a \theta_\infty^a(i)}$. Since

$$\tilde{\theta}_n(i) = \tilde{\theta}_0(i) + \gamma \sum_{k=1}^n \theta_{k-1}^a(i) 1_{E_i}(X_k)$$

we deduce that $\frac{\tilde{\theta}_n(i)}{n}$ converges to $\gamma\theta_\infty^a(i) \times \frac{\theta_\star(i)}{Z_{\theta_\infty}^a \theta_\infty^a(i)} = \frac{\gamma\theta_\star(i)}{Z_{\theta_\infty}^a}$ as $n \to \infty$. As a consequence, $\frac{S_n}{n} = \sum_{i=1}^{\mathcal{I}} \frac{\tilde{\theta}_n(i)}{n}$ converges to $\frac{\gamma}{Z_{\theta_\infty}^a}\sum_{i=1}^{\mathcal{I}} \theta_\star(i) = \frac{\gamma}{Z_{\theta_\infty}^a}$ and $\theta_n(i) = \frac{\tilde{\theta}_n(i)/n}{S_n/n}$ to $\frac{\gamma\theta_\star(i)}{Z_{\theta_\infty}^a} \times \frac{Z_{\theta_\infty}^a}{\gamma} = \theta_\star(i)$. This intuitive reasonning is not easy to generalize when $\alpha \in]1/2, 1[$.

4.2 Convergence of the SHUS$_a^\alpha$ algorithm

4.2.1 General assumptions

We study the convergence of the algorithm under the following assumption on the target probability measure $\eta(x)\lambda(dx)$ and on the strata $(E_i)_{i\in\{1,\ldots,\mathcal{I}\}}$:

A1 The density η of the target distribution is such that $\sup_E \eta < \infty$ and the strata $(E_i)_{i\in\{1,\ldots,\mathcal{I}\}}$ satisfy $\min_{1\le i\le\mathcal{I}} \pi(E_i) > 0$.

It is assumed that the Markov transition kernels $\{P_\theta^a, \theta \in \Theta\}$ satisfy :

A2 For any $\theta \in \Theta$, P^a_θ is a Metropolis-Hastings transition kernel with proposal kernel $q(x,y)\, d\lambda(y)$ where $q(x,y)$ is symmetric and satisfies $\inf_{E^2} q > 0$, and with invariant distribution $\eta^a_\theta\, d\lambda = \pi^a_\theta$, where η^a_θ is given by (4.1).

The assumption $\inf_{E^2} q > 0$ is particularly useful for the proof of recurrence property : we recall that the algorithm is recurrent if with probability one, the sequence $(\theta_n)_{n\geq 0}$ enters infinitely often a compact subset of Θ, see (4.15) below.

4.2.2 Convergence results

Our main result is the following convergence result.

Theorem 4.2.1. *Let $\gamma > 0$, $a \in [0,1)$ and $\alpha \in (\frac{1}{2}, 1]$. Assume A1, A2. Then the SHUS$^\alpha_a$ algorithm starting from any $(\mathbb{R}^*_+)^{\mathcal{I}} \times E$-valued random initial condition $(\tilde\theta_0, X_0)$ converges in the following sense :*

(i) $\lim_{n\to+\infty} \theta_n = \theta_\star$ $\mathbb{P}$*-a.s.*

(ii) For any bounded measurable function $f : E \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}\left[f(X_n)\right] = \int_E f(x)\pi^a_{\theta_\star}(dx) \text{ and } \mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f(X_k) = \int_E f(x)\pi^a_{\theta_\star}(dx)\right) = 1$$

(iii) For any bounded measurable function $f : E \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{\mathcal{I}} \theta_{n-1}^{1-a}(i)\right) \theta^a_{n-1}(I(X_n)) f(X_n)\right] = \int_E f(x)\pi(dx)\ ,$$

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \left(\sum_{i=1}^{\mathcal{I}} \theta_{k-1}^{1-a}(i)\right) \theta^a_{k-1}(I(X_k))\, f(X_k) = \int_E f(x)\pi(dx)\right) = 1$$

Note that $\frac{\eta(x)}{\eta^a_\theta(x)} = Z^a_\theta \theta^a(I(x))$ with $Z^a_\theta = \sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta^a(i)}$. In practice, since $\theta_\star$ is unknown, Z^a_θ is unkwown but may be approximated by $\sum_{i=1}^{\mathcal{I}} \theta^{1-a}(i)$ when θ is close to $\theta_\star$. This approximation is used in statement (iii).

According to [25], for the choice $a = 1$, the same convergence result holds under A1, A2 and if the target density η is bounded from below : $\inf_E \eta > 0$. This additional assumption on η is needed in to prove the recurrence property : $\limsup_{n\to\infty} \min_{1\leq i\leq \mathcal{I}} \theta_n(i) > 0$.

The key property for the proof of convergence of the sequence $(\theta_n)_{n\geq 0}$ to $\theta_\star$ is to rewrite the updating rule of the weight sequence $(\theta_n)_{n\geq 0}$ as in (4.9) below. This allows us to consider it as a stochastic approximation algorithm with *(i)* a random stepsize sequence $(\gamma_{n+1} = \frac{\gamma}{g_\alpha(S_n)})_{n\in\mathbb{N}}$ and *(ii)* Markovian inputs X_{n+1} with a distribution conditional to the past, depending on the current estimate θ_n of $\theta_\star$. According to (4.6), we have

$$\theta_{n+1}(i) = \theta_n(i) \frac{1 + \gamma_{n+1}\theta_n^{a-1}(i) 1_{E_i}(X_{n+1})}{1 + \gamma_{n+1}\theta^a_n(I(X_{n+1}))}.$$

With the identify $\frac{1+b}{1+c} = 1 + b - c + \frac{c(c-b)}{1+c}$ valid for $(b,c) \in \mathbb{R}\times\mathbb{R}^*$, we deduce that

$$\begin{aligned}\frac{\theta_{n+1}(i)}{\theta_n(i)} =& 1 + \gamma_{n+1}\left(\theta_n^{a-1}(i)1_{E_i}(X_{n+1}) - \theta^a_n(I(X_{n+1}))\right) \\ &+ \gamma^2_{n+1} \frac{\theta^a_n(I(X_{n+1}))\left(\theta^a_n(I(X_{n+1})) - \theta_n^{a-1}(i)1_{E_i}(X_{n+1})\right)}{1 + \gamma_{n+1}\theta^a_n(I(X_{n+1}))}.\end{aligned}$$

Hence

$$\begin{aligned}\theta_{n+1}(i) =&\theta_n(i) + \gamma_{n+1}\left(\theta_n^a(i)1_{E_i}(X_{n+1}) - \theta_n(i)\theta_n^a(I(X_{n+1}))\right)\\ &+ \gamma_{n+1}^2 \frac{\theta_n^a(I(X_{n+1}))\left(\theta_n(i)\theta_n^a(I(X_{n+1})) - \theta_n^a(i)1_{E_i}(X_{n+1})\right)}{1+\gamma_{n+1}\theta_n^a(I(X_{n+1}))}.\end{aligned}$$

Let us introduce the function $H : E \times \Theta \to \mathbb{R}^{\mathcal{I}}$ with i-th component

$$H_i(x,\theta) := \theta^a(i)\,1_{E_i}(x) - \theta(i)\,\theta^a(I(x)) = \theta^a(I(x))\,(1_{E_i}(x) - \theta(i)). \tag{4.8}$$

We do not indicate explicitly the dependence of H on a for the ease of notation. Setting

$$\Lambda_{n+1}(i) = \gamma_{n+1}\theta_n^{2a}(I(X_{n+1}))\frac{\theta_n(i) - 1_{E_i}(X_{n+1})}{1+\gamma_{n+1}\theta_n^a(I(X_{n+1}))},$$

we rewrite the evolution as

$$\theta_{n+1} = \theta_n + \gamma_{n+1}H(X_{n+1},\theta_n) + \gamma_{n+1}\Lambda_{n+1}. \tag{4.9}$$

Since $0 \leq \frac{\theta_n^{2a}(I(X_{n+1}))}{1+\gamma_{n+1}\theta_n^a(I(X_{n+1}))} \leq 1$ and $-1_{E_i}(X_{n+1}) \leq \theta_n(i) - 1_{E_i}(X_{n+1}) \leq \theta_n(i)$, the Euclidean norm of $\Lambda_{n+1} \in \mathbb{R}^{\mathcal{I}}$ with coordinates $\Lambda_{n+1}(i),\ 1 \leq i \leq \mathcal{I}$ satisfies

$$|\Lambda_{n+1}| \leq \gamma_{n+1}\sqrt{\sum_{i=1}^{\mathcal{I}}(\theta_n(i) - 1_{E_i}(X_{n+1}))^2} \leq \gamma_{n+1}\sqrt{\sum_{i=1}^{\mathcal{I}}(\theta_n^2(i) + 1_{E_i}(X_{n+1}))} \leq \gamma_{n+1}\sqrt{2}. \tag{4.10}$$

Let us compute the expectation of the function H_i under the probability measure π_θ^a :

$$\begin{aligned}h_i(\theta) &= \int_E H_i(x,\theta)\pi_\theta^a(dx) = \theta^a(i)\pi_\theta^a(E_i) - \theta(i)\sum_{j=1}^{\mathcal{I}}\theta^a(j)\pi_\theta^a(E_j)\\ &= \frac{1}{Z_\theta^a}\left(\theta^a(i) \times \frac{\theta_\star(i)}{\theta^a(i)} - \theta(i)\sum_{j=1}^{\mathcal{I}}\theta^a(j) \times \frac{\theta_\star(j)}{\theta^a(j)}\right) = \frac{\theta_\star(i) - \theta(i)}{Z_\theta^a}.\end{aligned}$$

Hence the function $h : \Theta \to \mathbb{R}^{\mathcal{I}}$ with coordinates $h_i,\ 1 \leq i \leq \mathcal{I}$ writes

$$h(\theta) = \frac{\theta_\star - \theta}{Z_\theta^a}. \tag{4.11}$$

The function h is called the mean-field function of the algorithm. The recursion (4.9) is a noisy version of the dynamics driven by this mean field function :

$$\theta_{n+1} = \theta_n + \gamma_{n+1}(h(\theta_n) + \xi_{n+1}) \text{ where } \xi_{n+1} := H(X_{n+1},\theta_n) - h(\theta_n) + \Lambda_{n+1}. \tag{4.12}$$

The proof of convergence of $(\theta_n)_{n\geq 0}$ relies on the results in Chapter 3 with the Lyapunov function $w(\theta) = \sum_{i=1}^{\mathcal{I}}\theta_\star(i)\ln(\theta_\star(i)/\theta(i))$. Notice that as soon as $\sum_{n\in\mathbb{N}}\gamma_{n+1}^2 < \infty$, then (4.10) ensures that $\lim_{n\to\infty}\sup_{k\geq n}|\sum_{j=n}^k \gamma_j\Lambda_j| = 0$. To deal with the other contribution $H(X_{n+1},\theta_n) - h(\theta_n)$ in ξ_{n+1}, we introduce the solution $\hat{H}_\theta$ (for notational simplicity we do not explicit the dependence of $\hat{H}_\theta$ on the parameter a which is fixed) of the Poisson equation

$$\hat{H}_\theta(x) - P_\theta^a\hat{H}_\theta(x) = H(x,\theta) - h(\theta),\ x \in E$$

permits to introduce a martingale increment by rewriting

$$\begin{aligned}H(X_{n+1},\theta_n)-h(\theta_n) &= \hat{H}_{\theta_n}(X_{n+1})-P^a_{\theta_n}\hat{H}_{\theta_n}(X_{n+1})\\ &= \hat{H}_{\theta_n}(X_{n+1})-P^a_{\theta_n}\hat{H}_{\theta_n}(X_n),\ \text{martingale increment since } X_{n+1}\sim P^a_{\theta_n}(X_n,.)\\ &+ P^a_{\theta_n}\hat{H}_{\theta_n}(X_n)-P^a_{\theta_{n+1}}\hat{H}_{\theta_{n+1}}(X_{n+1}),\ \text{telescopic}\\ &+ P^a_{\theta_{n+1}}\hat{H}_{\theta_{n+1}}(X_{n+1})-P^a_{\theta_n}\hat{H}_{\theta_n}(X_{n+1}),\ \text{small if } \theta\mapsto P^a_\theta\hat{H}_\theta(x) \text{ smooth.}\end{aligned}$$

After multiplication by the stepsize γ_{n+1}, the telescopic property is lost :

$$\begin{aligned}\sum_{k=0}^{n-1}\gamma_{k+1}\left(P^a_{\theta_k}\hat{H}_{\theta_k}(X_k)-P^a_{\theta_{k+1}}\hat{H}_{\theta_{k+1}}(X_{k+1})\right) =&\gamma_1 P^a_{\theta_0}\hat{H}_{\theta_0}(X_0)-\gamma_n P^a_{\theta_n}\hat{H}_{\theta_n}(X_n)\\ &+\sum_{k=1}^{n-1}(\gamma_{k+1}-\gamma_k)P^a_{\theta_k}\hat{H}_{\theta_k}(X_k)\end{aligned}$$

but the sum will be small if γ_k does not move too quickly with k. Notice that

$$\frac{1}{k^\alpha}-\frac{1}{(k+1)^\alpha}=\frac{1}{k^\alpha}\left(1-(1+1/k)^{-\alpha}\right)\sim\frac{\alpha}{k^{1+\alpha}}\ \text{as } k\to\infty$$

Theorem 4.2.1 is a consequence of Proposition 4.2.2 and Proposition 4.2.4 stated below. Proposition 4.2.2 shows that convergence holds as soon as the algorithm is recurrent and the stepsize sequence $(\gamma_n)_{n\geq1}$ a.s. satisfies the usual conditions $\sum_{n\geq1}\gamma_n^2<\infty$ and $\sum_{n\geq1}\gamma_n=\infty$ for the convergence of stochastic approximation algorithms. Proposition 4.2.4 ensures that those two conditions are actually satisfied in our setting. The proof of Proposition 4.2.4 is based in particular on some sufficient conditions on the sequence $(\gamma_n)_{n\geq1}$ for the recurrence of the algorithm, stated in Proposition 4.2.3.

Proposition 4.2.2. *Let $a\in[0,1)$. Assume we are given a density η and a family of kernels P^a_θ satisfying A1 and A2. Assume that we are given sequences $(\tilde{\theta}_n,X_n)_{n\geq0}$ and $(\gamma_n)_{n\geq1}$ generated by the WL_a algorithm (see Algorithm 2). In particular, for all $n\geq0$, conditionally on*

$$\mathcal{F}_n:=\sigma\left(\tilde{\theta}_0,X_0,X_1,\cdots,X_n\right), \tag{4.13}$$

X_{n+1} is generated according to the distribution $P^a_{\theta_n}(X_n,\cdot)$.

If, moreover,

- *the stepsize sequence $(\gamma_n)_{n\geq1}$ is predictable with respect to the filtration $(\mathcal{F}_n)_{n\geq0}$ (i.e. for all $n\geq1$, γ_n is $\mathcal{F}_{n-1}$-measurable) and such that*

$$\mathbb{P}\left((\gamma_n)_{n\geq1}\ \textit{is non-increasing,}\ \sum_{n\geq1}\gamma_n=\infty\ \textit{and}\ \sum_{n\geq1}\gamma_n^2<\infty\right)=1, \tag{4.14}$$

- *the algorithm is recurrent, in the following sense :*

$$\mathbb{P}\left(\limsup_{n\to\infty}\min_{1\leq i\leq\mathcal{I}}\theta_n(i)>0\right)=1, \tag{4.15}$$

then the conclusions (i)-(ii)-(iii) of Theorem 4.2.1 hold.

Some sufficient conditions on $(\gamma_n)_{n\geq 1}$ to ensure the recurrence of the WL_a algorithm are given in the following proposition.

Proposition 4.2.3. *Let $a \in [0,1)$. Assume we are given a density η and a family of kernels P_θ^a satisfying A1 and A2. Assume that we are given sequences $(\tilde{\theta}_n, X_n)_{n\geq 0}$ and $(\gamma_n)_{n\geq 1}$ satisfying the recurrence relation (4.5) of the WL_a algorithm.*

If the sequence $(\gamma_n)_{n\geq 1}$ is non-increasing, bounded from above by a deterministic sequence converging to 0 as $n \to \infty$ and such that

$$\bar{r}_{\mathcal{I},\gamma} := \sup_{n\geq 1} \frac{\gamma_n}{\gamma_{n+\mathcal{I}-1}} < +\infty, \tag{4.16}$$

then the algorithm is recurrent, in the sense of (4.15).

Notice that Propositions 4.2.2 and 4.2.3 give some sufficient conditions on $(\gamma_n)_{n\geq 1}$ for the convergence of the WL_a algorithm.

Proposition 4.2.4. *Let $\gamma > 0$, $a \in [0,1)$ and $\alpha \in (\frac{1}{2}, 1]$. Assume A1, A2. Then the sequences $\left(\tilde{\theta}_n, X_n\right)_{n\geq 0}$ and $\left(\gamma_{n+1} = \frac{\gamma}{g_\alpha(S_n)}\right)_{n\geq 0}$ generated by the $SHUS_a^\alpha$ algorithm starting from any $(\mathbb{R}_+^*)^{\mathcal{I}} \times E$-valued random initial condition $(\tilde{\theta}_0, X_0)$ are such that (4.14) and (4.15) hold.*

A useful corollary of the previous results gives the effective behavior of the stepsize sequence $(\gamma_{n+1})_{n\geq 0}$ in the longtime limit $n \to +\infty$, confirming the intuition given in Section 4.1.3.

Corollary 4.2.5. *Let $\gamma > 0$, $a \in [0,1)$ and $\alpha \in (\frac{1}{2}, 1]$. Assume A1, A2.*

Then, the stepsize sequence $\left(\gamma_{n+1} = \frac{\gamma}{g_\alpha(S_n)}\right)_{n\geq 0}$ generated by the $SHUS_a^\alpha$ algorithm starting from any $(\mathbb{R}_+^*)^{\mathcal{I}} \times E$-valued random initial condition $(\tilde{\theta}_0, X_0)$ has the following asymptotic behavior :

$$\mathbb{P}\left(\lim_{n\to\infty} (n^\alpha \gamma_n)^{1/\alpha} = \mathcal{C}_\alpha\, Z_{\theta_\star}^a\right) = 1 \quad \text{with} \quad \mathcal{C}_\alpha := \begin{cases} 1 & \text{if } \alpha = 1, \\ (1-\alpha)\,\gamma^{(1/\alpha)-1} & \text{if } \alpha \in (1/2, 1). \end{cases}$$

where, we recall $Z_{\theta_\star}^a = \sum_{i=1}^{\mathcal{I}} \theta_\star^{1-a}(i)$.

4.3 Proof of Proposition 4.2.2

To prove Proposition 4.2.2, we are first going to establish several preliminary results.

Proposition 4.3.1. *Under A1 and A2, there exists $\tilde{\alpha} \in (0,1)$ such that for all $\theta \in \Theta$, for all $x \in E$ and for all measurable set $A \subset E$, it holds :*

$$P_\theta^a(x, A) \geq \tilde{\alpha}\pi_\theta^a(A). \tag{4.17}$$

Hence the uniform Doeblin condition holds. By theorem 1.3.11, we deduce that

Corollary 4.3.2. *Under A1 and A2,*

$$\exists \tilde{\alpha} \in (0,1),\ \forall n \geq 0,\ \sup_{\theta\in\Theta}\sup_{x\in E} d_{\mathrm{TV}}((P_\theta^a)^n(x,\cdot), \pi_\theta^a) \leq (1-\tilde{\alpha})^n. \tag{4.18}$$

Proof of Proposition 4.3.1. By definition of the Metropolis-Hastings kernel and since q is symmetric by A2,

$$P_\theta^a(x,A) \geq \int_A q(x,y)\left(1 \wedge \frac{\eta_\theta^a(y)}{\eta_\theta^a(x)}\right) d\lambda(y) \geq \frac{\inf_{E\times E} q}{\sup_E \eta_\theta^a} \int_A \eta_\theta^a(y)\lambda(dy) = \frac{\inf_{E\times E} q}{\sup_E \eta_\theta^a} \pi_\theta^a(A).$$

Again by A2, $\inf_{E\times E} q > 0$. On the other hand, for $x \in E$,

$$\eta_\theta^a(x) = \sum_{i=1}^{\mathcal{I}} 1_{E_i}(x) \frac{\eta(x)/\theta^a(i)}{\sum_{j=1}^{\mathcal{I}} \theta_\star(j)/\theta^a(j)} \leq \sum_{i=1}^{\mathcal{I}} 1_{E_i}(x) \frac{\eta(x)/\theta^a(i)}{\theta_\star(i)/\theta^a(i)} \leq \frac{\sup_E \eta}{\min_{1\leq i\leq \mathcal{I}} \theta_\star(i)}$$

where the right-hand side is finite by A1. Hence the conclusion holds with

$$\tilde{\alpha} = \frac{\min_{1\leq i\leq \mathcal{I}} \theta_\star(i) \inf_{E\times E} q}{\sup_E \eta}.$$

□

The proof is now organized as follows. We first state three lemmas which quantify the dependence on θ of the invariant measure π_θ^a, the transition kernel P_θ^a, and the solution to a Poisson equation associated with P_θ^a. We then give the proof of Proposition 4.2.2.

Lemma 4.3.3. *For all $\theta_1, \theta_2 \in \Theta$,*

$$d_{\mathrm{TV}}(\pi_{\theta_1}^a, \pi_{\theta_2}^a) \leq (\mathcal{I}-1) \sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta_2^a(i)}{\theta_1^a(i)}\right|.$$

Notice that the result of course also holds with the symmetrized right-hand side

$$(\mathcal{I}-1) \min\left(\sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta_2^a(i)}{\theta_1^a(i)}\right|, \sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta_1^a(i)}{\theta_2^a(i)}\right|\right).$$

Proof: The proof is adapted from [24, Lemma 4.6]. By definition of η_θ^a, for $\ell \in \{1,2\}$,

$$\eta_{\theta_\ell}^a(x) = \sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)/\theta_\ell^a(i)}{\sum_{j=1}^{\mathcal{I}} [\theta_\star(j)/\theta_\ell^a(j)]} \times \frac{\eta(x)}{\theta_\star(i)}\ 1_{E_i}(x)\ ,$$

where $\frac{\eta(x)}{\theta_\star(i)}\ 1_{E_i}(x)$ is probability density with respect to λ. Hence,

$$\begin{aligned} 2d_{\mathrm{TV}}(\pi_{\theta_1}^a, \pi_{\theta_2}^a) &\leq \sum_{i=1}^{\mathcal{I}} \left| \frac{\theta_\star(i)/\theta_1^a(i)}{\sum_{k=1}^{\mathcal{I}} [\theta_\star(k)/\theta_1^a(k)]} - \frac{\theta_\star(i)/\theta_2^a(i)}{\sum_{k=1}^{\mathcal{I}} [\theta_\star(k)/\theta_2^a(k)]} \right| \\ &\leq \frac{\sum_{j=1}^{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} \theta_\star(i)\theta_\star(j) \left|1/[\theta_1^a(i)\theta_2^a(j)] - 1/[\theta_2^a(i)\theta_1^a(j)]\right|}{\sum_{k=1}^{\mathcal{I}} [\theta_\star(k)/\theta_1^a(k)]\ \sum_{l=1}^{\mathcal{I}} [\theta_\star(l)/\theta_2^a(l)]}. \end{aligned}$$

We denote by $N(\theta_1, \theta_2)$ the numerator in the right-hand side of the previous inequality. Then,

$$\begin{aligned} N(\theta_1,\theta_2) &= \sum_{j=1}^{\mathcal{I}} \sum_{i\neq j} \theta_\star(i)\theta_\star(j)\ \frac{|\theta_2^a(i)\,\theta_1^a(j) - \theta_1^a(i)\,\theta_2^a(j)|}{\theta_1^a(i)\,\theta_2^a(i)\,\theta_1^a(j)\,\theta_2^a(j)} \\ &\leq \sum_{j=1}^{\mathcal{I}} \sum_{i\neq j} \theta_\star(i)\theta_\star(j)\ \left|\frac{\theta_1^a(j) - \theta_2^a(j)}{\theta_1^a(i)\,\theta_1^a(j)\,\theta_2^a(j)}\right| + \sum_{j=1}^{\mathcal{I}} \sum_{i\neq j} \theta_\star(i)\theta_\star(j)\ \left|\frac{\theta_1^a(i) - \theta_2^a(i)}{\theta_1^a(i)\,\theta_2^a(i)\,\theta_1^a(j)}\right|. \end{aligned}$$

For the denominator, we use the lower bound

$$\forall i,j \in \{1,\ldots,\mathcal{I}\}, \qquad \sum_{k=1}^{\mathcal{I}}[\theta_\star(k)/\theta_1^a(k)] \sum_{l=1}^{\mathcal{I}}[\theta_\star(l)/\theta_2^a(l)] \geq \frac{\theta_\star(i)\,\theta_\star(j)}{\theta_1^a(i)\,\theta_2^a(j)} \vee \frac{\theta_\star(i)\,\theta_\star(j)}{\theta_1^a(j)\,\theta_2^a(i)}.$$

Therefore,

$$d_{\mathrm{TV}}(\pi_{\theta_1}^a, \pi_{\theta_2}^a) \leq \sum_{j=1}^{\mathcal{I}} \sum_{i \neq j} \frac{|\theta_1^a(j) - \theta_2^a(j)|}{\theta_1^a(j)} \leq (\mathcal{I}-1) \sum_{j=1}^{\mathcal{I}} \frac{|\theta_1^a(j) - \theta_2^a(j)|}{\theta_1^a(j)} \;,$$

which gives the claimed result. ∎

Lemma 4.3.4. *Assume that q is symmetric. Then for all $\theta_1, \theta_2 \in \Theta$ and $x \in E$*

$$d_{\mathrm{TV}}(P_{\theta_1}^a(x,\cdot), P_{\theta_2}^a(x,\cdot)) \leq 2 \max_{i \in \{1,\ldots,\mathcal{I}\}} \left|1 - \frac{\theta_2^a(i)}{\theta_1^a(i)}\right|. \tag{4.19}$$

Proof: The proof is adapted from [24, Lemma 4.7]. As P_θ^a is a Metropolis kernel, for any bounded measurable function f,

$$\begin{aligned}\left|P_{\theta_1}^a f(x) - P_{\theta_2}^a f(x)\right| &= \left|\int_E q(x,y)\,(\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y))\,(f(y) - f(x))\,d\lambda(y)\right| \\ &\leq 2 \sup_E |f| \sup_{E^2} |\alpha_{\theta_1} - \alpha_{\theta_2}|\end{aligned}$$

where $\alpha_\theta(x,y) := 1 \wedge (\eta_\theta^a(y)/\eta_\theta^a(x))$ since q is symmetric. For notational simplicity, we do not indicate explicitly the dependence of α_{θ_1} on a. Since, by Proposition 1.3.3, $d_{\mathrm{TV}}(P_{\theta_1}^a(x,\cdot), P_{\theta_2}^a(x,\cdot)) = \sup_{f:|f|\leq\frac{1}{2}} \left|P_{\theta_1}^a f(x) - P_{\theta_2}^a f(x)\right|$, one deduces that $d_{\mathrm{TV}}(P_{\theta_1}^a(x,\cdot), P_{\theta_2}^a(x,\cdot)) \leq \sup_{E^2} |\alpha_{\theta_1} - \alpha_{\theta_2}|$. Let us introduce the unnormalized density

$$\forall x \in E,\ \tilde{\eta}_\theta^a(x) := \sum_{i=1}^{\mathcal{I}} \frac{\eta(x)}{\theta^a(i)} 1_{E_i}(x),$$

which is such that $\forall x \in E, \eta_\theta^a(x) = (Z_\theta^a)^{-1} \tilde{\eta}_\theta^a(x)$ where $Z_\theta^a = \sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\theta^a(j)}$. Notice that

$$\forall x,y \in E,\ \alpha_\theta(x,y) = 1 \wedge \frac{\tilde{\eta}_\theta^a(y)}{\tilde{\eta}_\theta^a(x)}.$$

We now show that

$$\forall \theta_1, \theta_2 \in \Theta,\ \forall x,y \in E,\ |\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y)| \leq 2 \sup_E \left|1 - \frac{\tilde{\eta}_{\theta_1}^a}{\tilde{\eta}_{\theta_2}^a}\right| \tag{4.20}$$

which yields the result (4.19) since

$$\sup_E \left|1 - \frac{\tilde{\eta}_{\theta_1}^a}{\tilde{\eta}_{\theta_2}^a}\right| = \max_{i \in \{1,\ldots,\mathcal{I}\}} \left|1 - \frac{\theta_2^a(i)}{\theta_1^a(i)}\right|.$$

The proof (4.20) is performed by distingishing between four cases.

- If $\tilde{\eta}^a_{\theta_1}(x) \leq \tilde{\eta}^a_{\theta_1}(y)$ and $\tilde{\eta}^a_{\theta_2}(x) \leq \tilde{\eta}^a_{\theta_2}(y)$, then $\alpha_{\theta_1}(x,y) = 1 = \alpha_{\theta_2}(x,y)$ and $|\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y)| = 0$.
- If $\tilde{\eta}^a_{\theta_1}(x) \leq \tilde{\eta}^a_{\theta_1}(y)$ and $\tilde{\eta}^a_{\theta_2}(y) \leq \tilde{\eta}^a_{\theta_2}(x)$, then

$$\begin{aligned}
|\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y)| &= 1 - \frac{\tilde{\eta}^a_{\theta_2}(y)}{\tilde{\eta}^a_{\theta_2}(x)} \leq 1 - \frac{\tilde{\eta}^a_{\theta_2}(y)}{\tilde{\eta}^a_{\theta_2}(x)} \frac{\tilde{\eta}^a_{\theta_1}(x)}{\tilde{\eta}^a_{\theta_1}(y)} \leq \frac{|\tilde{\eta}^a_{\theta_2}(x) - \tilde{\eta}^a_{\theta_1}(x)|}{\tilde{\eta}^a_{\theta_2}(x)} + \frac{\tilde{\eta}^a_{\theta_1}(x)}{\tilde{\eta}^a_{\theta_2}(x)} \frac{|\tilde{\eta}^a_{\theta_1}(y) - \tilde{\eta}^a_{\theta_2}(y)|}{\tilde{\eta}^a_{\theta_1}(y)} \\
&\leq \frac{|\tilde{\eta}^a_{\theta_2}(x) - \tilde{\eta}^a_{\theta_1}(x)|}{\tilde{\eta}^a_{\theta_2}(x)} + \frac{\tilde{\eta}^a_{\theta_1}(x)}{\tilde{\eta}^a_{\theta_1}(y)} \frac{\tilde{\eta}^a_{\theta_2}(y)}{\tilde{\eta}^a_{\theta_2}(x)} \frac{|\tilde{\eta}^a_{\theta_1}(y) - \tilde{\eta}^a_{\theta_2}(y)|}{\tilde{\eta}^a_{\theta_2}(y)} \\
&\leq 2 \sup_E \left| 1 - \frac{\tilde{\eta}^a_{\theta_1}}{\tilde{\eta}^a_{\theta_2}} \right| .
\end{aligned}$$

- If $\tilde{\eta}^a_{\theta_1}(y) \leq \tilde{\eta}^a_{\theta_1}(x)$ and $\tilde{\eta}^a_{\theta_2}(y) \leq \tilde{\eta}^a_{\theta_2}(x)$, then

$$\begin{aligned}
|\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y)| &= \left| \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_1}(x)} - \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_2}(x)} + \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_2}(x)} - \frac{\tilde{\eta}^a_{\theta_2}(y)}{\tilde{\eta}^a_{\theta_2}(x)} \right| \\
&\leq \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_1}(x)} \frac{\left|\tilde{\eta}^a_{\theta_2}(x) - \tilde{\eta}^a_{\theta_1}(x)\right|}{\tilde{\eta}^a_{\theta_2}(x)} + \frac{\left|\tilde{\eta}^a_{\theta_1}(y) - \tilde{\eta}^a_{\theta_2}(y)\right|}{\tilde{\eta}^a_{\theta_2}(x)} \\
&\leq \frac{\left|\tilde{\eta}^a_{\theta_1}(x) - \tilde{\eta}^a_{\theta_2}(x)\right|}{\tilde{\eta}^a_{\theta_2}(x)} + \frac{\left|\tilde{\eta}^a_{\theta_1}(y) - \tilde{\eta}^a_{\theta_2}(y)\right|}{\tilde{\eta}^a_{\theta_2}(y)} \\
&\leq 2 \sup_E \left| 1 - \frac{\tilde{\eta}^a_{\theta_1}}{\tilde{\eta}^a_{\theta_2}} \right| .
\end{aligned}$$

- Assume $\tilde{\eta}^a_{\theta_1}(y) \leq \tilde{\eta}^a_{\theta_1}(x)$ and $\tilde{\eta}^a_{\theta_2}(x) \leq \tilde{\eta}^a_{\theta_2}(y)$. If $\tilde{\eta}^a_{\theta_1}(x) \leq \tilde{\eta}^a_{\theta_2}(y)$, it holds that

$$|\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y)| = 1 - \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_1}(x)} \leq 1 - \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_2}(y)} \leq \sup_E \left| 1 - \frac{\tilde{\eta}^a_{\theta_1}}{\tilde{\eta}^a_{\theta_2}} \right| .$$

Otherwise, we have $\tilde{\eta}^a_{\theta_2}(x) \leq \tilde{\eta}^a_{\theta_2}(y) \leq \tilde{\eta}^a_{\theta_1}(x)$ and we write

$$\begin{aligned}
|\alpha_{\theta_1}(x,y) - \alpha_{\theta_2}(x,y)| = 1 - \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_1}(x)} &= \frac{\tilde{\eta}^a_{\theta_2}(x)}{\tilde{\eta}^a_{\theta_1}(x)} \left(\frac{\tilde{\eta}^a_{\theta_1}(x)}{\tilde{\eta}^a_{\theta_2}(x)} - \frac{\tilde{\eta}^a_{\theta_2}(y)}{\tilde{\eta}^a_{\theta_2}(x)} \right) + \frac{\tilde{\eta}^a_{\theta_2}(y)}{\tilde{\eta}^a_{\theta_1}(x)} \left(1 - \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_2}(y)} \right) \\
&\leq \left(\frac{\tilde{\eta}^a_{\theta_1}(x)}{\tilde{\eta}^a_{\theta_2}(x)} - 1 \right) + \left| 1 - \frac{\tilde{\eta}^a_{\theta_1}(y)}{\tilde{\eta}^a_{\theta_2}(y)} \right| \\
&\leq 2 \sup_E \left| 1 - \frac{\tilde{\eta}^a_{\theta_1}}{\tilde{\eta}^a_{\theta_2}} \right| .
\end{aligned}$$

This shows that (4.20) holds, and thus concludes the proof. ∎

Lemma 4.3.5. *Assume A1 and A2. Then,* $\sup_{\theta \in \Theta} \sup_{x \in E} |H(x,\theta)| \leq \sqrt{2}$ *where* H *is defined by* (4.8). *In addition, for any* $\theta \in \Theta$, *there exists a unique function* $\widehat{H}_\theta$ *solving the Poisson equation*

$$\widehat{H}_\theta - P^a_\theta \widehat{H}_\theta = H(\cdot, \theta) - h(\theta), \qquad \pi^a_\theta \left(\widehat{H}_\theta \right) = 0, \tag{4.21}$$

where, we recall, $h(\theta) = \pi^a_\theta(H(\cdot,\theta))$ *is defined by* (4.11). *Moreover,* $\widehat{H}_\theta$ *is uniformly bounded :*

$$\sup_{\theta \in \Theta, x \in E} \left| \widehat{H}_\theta(x) \right| < \frac{2\sqrt{\mathcal{I}}}{\tilde{\alpha}}.$$

Notice that for notational simplicity, we do not indicate explicitly the dependence of $\widehat{H}_\theta$ on a.

Proof: For all $(x,\theta) \in E \times \Theta$, since $H_i(x,\theta) = \theta^a(I(x)))(1_{E_i}(x) - \theta(i))$,

$$\begin{aligned}|H(x,\theta)|^2 &= \sum_{i=1}^{\mathcal{I}} (H_i(x,\theta))^2 \leq \theta^{2a}(I(x)) \sum_{i=1}^{\mathcal{I}} \left(1_{E_i}(x) + \theta(i)^2\right) \\ &\leq 1 + \sum_{i=1}^{\mathcal{I}} \theta(i) = 2, \end{aligned} \tag{4.22}$$

so that $\sup_{\theta\in\Theta} \sup_{x\in E} |H(x,\theta)| \leq \sqrt{2}$. Set

$$\widehat{H}_\theta(x) := \sum_{n\geq 0} \left(\int_E (P^a_\theta)^n(x,dy) H(y,\theta) - h(\theta) \right).$$

By Proposition 1.3.3 and Corollary 4.3.2 and since $|H_i(x,\theta)| \leq 1$,

$$\begin{aligned}\sup_{\theta\in\Theta, x\in E} \left| \int_E (P^a_\theta)^n(x,dy) H_i(y,\theta) - h_i(\theta) \right| &\leq 2 \sup_{\theta\in\Theta, x\in E} d_{\mathrm{TV}}((P^a_\theta)^n(x,.), \pi^a_\theta) \\ &\leq 2(1-\tilde{\alpha})^n.\end{aligned}$$

Therefore

$$\widehat{H}_\theta(x) := \sum_{n\geq 0} \left(\int_E (P^a_\theta)^n(x,dy) H(y,\theta) - h(\theta) \right).$$

is well defined and

$$\sup_{\theta\in\Theta} \sup_{x\in E} \left| \widehat{H}_\theta(x) \right| \leq \sup_{\theta\in\Theta} \sup_{x\in E} \sum_{n\geq 0} |(P^a_\theta)^n H(\cdot,\theta)(x) - \pi^a_\theta(H(\cdot,\theta))| \leq \frac{2\sqrt{\mathcal{I}}}{\tilde{\alpha}}. \tag{4.23}$$

It is easily seen that this function satisfies (4.21) (see the proof of Proposition 1.5.1).

∎

Lemma 4.3.6. *Assume A1 and A2. There exists a finite constant C such that, for any $\theta_1, \theta_2 \in \Theta$,*

$$\sup_E \left\{ \left| \widehat{H}_{\theta_1} - \widehat{H}_{\theta_2} \right| + \left| P^a_{\theta_1} \widehat{H}_{\theta_1} - P^a_{\theta_2} \widehat{H}_{\theta_2} \right| \right\} \leq C \left(|\theta_1 - \theta_2| + \sum_{i=1}^{\mathcal{I}} \left| 1 - \frac{\theta^a_2(i))}{\theta^a_1(i))} \right| \right).$$

Proof: By the definition (4.8) of H,

$$\sup_{x\in E} |H_i(x,\theta_1) - H_i(x,\theta_2)| \leq \max_{j\in\{1,\ldots,\mathcal{I}\}} |\theta^a_1(j) - \theta^a_2(j)| + |\theta_1(i) - \theta_2(i)|.$$

Therefore, there exists a finite constant C' such that for any $\theta_1, \theta_2 \in \Theta$,

$$\sup_{x\in E} |H(x,\theta) - H(x,\theta_2)| \leq C' \left\{ \sum_{j=1}^{\mathcal{I}} \left| 1 - \frac{\theta^a_2(j)}{\theta^a_1(j)} \right| + |\theta_1 - \theta_2| \right\}. \tag{4.24}$$

Since $h(\theta) = \frac{\theta_\star - \theta}{\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta^a(i))}}$ where, $\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta^a(i))} \geq \sum_{i=1}^{\mathcal{I}} \theta_\star(i) = 1$, one has

$$
\begin{aligned}
|h(\theta_1) - h(\theta_2)| &\leq \frac{|\theta_1 - \theta_2|}{\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta_1^a(i)}} + |\theta_2| \left| \frac{1}{\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta_1^a(i)}} - \frac{1}{\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta_2^a(i)}} \right| \\
&\leq |\theta_1 - \theta_2| + \frac{|\theta_2|}{\sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\theta_1^a(j)}} \sum_{i=1}^{\mathcal{I}} \frac{\frac{\theta_\star(i)}{\theta_2^a(i)}}{\sum_{j=1}^{\mathcal{I}} \frac{\theta_\star(j)}{\theta_2^a(j)}} \left| \frac{\theta_2^a(i)}{\theta_1^a(i)} - 1 \right| \\
&\leq |\theta_1 - \theta_2| + \sum_{i=1}^{\mathcal{I}} \left| \frac{\theta_2^a(i)}{\theta_1^a(i)} - 1 \right|.
\end{aligned}
$$

Hence, with the Poisson equation $\widehat{H}_\theta - P_\theta^a \widehat{H}_\theta = H(\cdot, \theta) - h(\theta)$, it is enough to prove the estimation of $P_{\theta_1}^a \widehat{H}_{\theta_1} - P_{\theta_2}^a \widehat{H}_{\theta_2}$ to conclude.

Denoting (by a slight abuse of notation) $H_\theta(x)$ and P_θ in place of $H(x,\theta)$ and P_θ^a to lighten notations and following [26, Lemma 4.2], one writes

$$
P_{\theta_1} \widehat{H}_{\theta_1} - P_{\theta_2} \widehat{H}_{\theta_2} = (P_{\theta_1}^n - P_{\theta_2}^n) H_{\theta_1} + P_{\theta_2}^n (H_{\theta_1} - H_{\theta_2}) - \pi_{\theta_2}^a (H_{\theta_1} - H_{\theta_2}) + \pi_{\theta_2}^a (H_{\theta_1}) - \pi_{\theta_1}^a (H_{\theta_1}).
$$

Using the invariance of $\pi_{\theta_1}^a$ by P_{θ_1} for the second equality and denoting by $(P_{\theta_1}^k - \pi_{\theta_1}^a)$ the kernel $P_{\theta_1}^k(x, dy) - \pi_{\theta_1}^a(dy)$, one gets

$$
\begin{aligned}
(P_{\theta_1}^n - P_{\theta_2}^n) H_{\theta_1} &= \sum_{k=0}^{n-1} P_{\theta_1}^k (P_{\theta_1} - P_{\theta_2}) P_{\theta_2}^{n-k-1} H_{\theta_1} \\
&= \sum_{k=0}^{n-1} (P_{\theta_1}^k - \pi_{\theta_1}^a)(P_{\theta_1} - P_{\theta_2}) P_{\theta_2}^{n-k-1} H_{\theta_1} + \sum_{k=0}^{n-1} \left(\pi_{\theta_1}^a P_{\theta_2}^{n-k-1}(H_{\theta_1}) - \pi_{\theta_1}^a P_{\theta_2}^{n-k}(H_{\theta_1}) \right) \\
&= \sum_{k=0}^{n-1} (P_{\theta_1}^k - \pi_{\theta_1}^a)(P_{\theta_1} - P_{\theta_2}) P_{\theta_2}^{n-k-1} H_{\theta_1} + \pi_{\theta_1}^a (H_{\theta_1}) - \pi_{\theta_1}^a (P_{\theta_2}^n H_{\theta_1}).
\end{aligned}
$$

Also remarking that $(P_{\theta_1} - P_{\theta_2}) \pi_{\theta_2}^a (H_{\theta_1}) = \pi_{\theta_2}^a (H_{\theta_1}) - \pi_{\theta_2}^a (H_{\theta_1}) = 0$, one deduces that

$$
\begin{aligned}
P_{\theta_1}^n (H_{\theta_1} - \pi_{\theta_1}^a (H_{\theta_1})) - P_{\theta_2}^n (H_{\theta_2} - \pi_{\theta_2}^a (H_{\theta_2})) &= \sum_{k=0}^{n-1} (P_{\theta_1}^k - \pi_{\theta_1}^a)(P_{\theta_1} - P_{\theta_2}) \left(P_{\theta_2}^{n-k-1} H_{\theta_1} - \pi_{\theta_2}^a (H_{\theta_1}) \right) \\
&\quad + P_{\theta_2}^n (H_{\theta_1} - H_{\theta_2}) - \pi_{\theta_2}^a (H_{\theta_1} - H_{\theta_2}) + \pi_{\theta_2}^a (H_{\theta_1}) - \pi_{\theta_1}^a (P_{\theta_2}^n H_{\theta_1}).
\end{aligned}
\tag{4.25}
$$

By Proposition 1.3.3, (4.19), (4.18) and Lemma 4.3.5,

$$
\begin{aligned}
&\sup_E \left| (P_{\theta_1}^k - \pi_{\theta_1}^a)(P_{\theta_1} - P_{\theta_2}) \left(P_{\theta_2}^{n-k-1} H_{\theta_1} - \pi_{\theta_2}^a (H_{\theta_1}) \right) \right| \\
&\quad \leq 8 \sup_{x \in E} d_{\mathrm{TV}}(P_{\theta_1}^k(x, .), \pi_{\theta_1}^a) \sup_{x \in E} d_{\mathrm{TV}}(P_{\theta_1}^a(x, .), P_{\theta_2}^a(x, .)) \sup_{x \in E} d_{\mathrm{TV}}(P_{\theta_2}^{n-k-1}(x, .), \pi_{\theta_2}^a) \sup_{x \in E} |H_{\theta_1}(x)| \\
&\quad \leq 16\sqrt{2} \sup_{i \in \{1, \dots, \mathcal{I}\}} \left| 1 - \frac{\theta_2^a(i)}{\theta_1^a(i)} \right| (1 - \tilde{\alpha})^{n-1}.
\end{aligned}
$$

In the same way, by Proposition 1.3.3, (4.18) and (4.24),

$$
\begin{aligned}
\sup_E \left| P_{\theta_2}^n (H_{\theta_1} - H_{\theta_2}) - \pi_{\theta_2}^a (H_{\theta_1} - H_{\theta_2}) \right| &\leq 2 \sup_{x \in E} d_{\mathrm{TV}}(P_{\theta_2}^n(x, .), \pi_{\theta_2}^a) \sup_{x \in E} |H(\theta_1, x) - H(\theta_2, x)| \\
&\leq 2(1 - \tilde{\alpha})^n C' \left\{ |\theta_1 - \theta_2| + \sum_{i=1}^{\mathcal{I}} \left| 1 - \frac{\theta_2^a(i))}{\theta_1^a(i))} \right| \right\}.
\end{aligned}
$$

Last, remarking that, by invariance of $\pi^a_{\theta_2}$ by $P^n_{\theta_2}$, $\pi^a_{\theta_2}(H_{\theta_1}) - \pi^a_{\theta_1}(P^n_{\theta_2}H_{\theta_1}) = (\pi^a_{\theta_2} - \pi^a_{\theta_1})(P^n_{\theta_2}H_{\theta_1} - \pi^a_{\theta_2}(H_{\theta_1}))$ and using (4.18) and Lemmas 4.3.3 and 4.3.5, one gets

$$\begin{aligned}\left|\pi^a_{\theta_2}(H_{\theta_1}) - \pi^a_{\theta_1}(P^n_{\theta_2}H_{\theta_1})\right| &\leq 4d_{\mathrm{TV}}(\pi^a_{\theta_1}, \pi^a_{\theta_2}) \sup_{x\in E} d_{\mathrm{TV}}(P^n_{\theta_2}(x,.), \pi^a_{\theta_2}) \sup_{x\in E} |H_{\theta_1}(x)| \\ &\leq 4\sqrt{2}(\mathcal{I}-1)(1-\tilde{\alpha})^n \sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta^a_2(i))}{\theta^a_1(i))}\right|.\end{aligned}$$

Plugging these three estimations into (4.25), we obtain the existence of a finite constant C not depending on $\theta_1, \theta_2 \in \Theta$ such that

$$\sup_E \left|P^n_{\theta_1}(H_{\theta_1} - \pi_{\theta_1}(H_{\theta_1})) - P^n_{\theta_2}(H_{\theta_2} - \pi^a_{\theta_2}(H_{\theta_2}))\right| \leq Cn(1-\tilde{\alpha})^{n-1} \left\{|\theta_1 - \theta_2| + \sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta^a_2(i))}{\theta^a_1(i))}\right|\right\}.$$

One concludes by summation over $n \geq 1$ since $\sum_{n\geq 1} n(1-\tilde{\alpha})^{n-1} = \tilde{\alpha}^{-2}$.

∎

Let us introduce the Lyapunov function $w : (0, +\infty)^{\mathcal{I}} \to \mathbb{R}$ defined by :

$$w(\theta) := \sum_{i=1}^{\mathcal{I}} \theta_\star(i) \ \ln\left(\frac{\theta_\star(i)}{\theta(i)}\right).$$

When $\theta \in \Theta$, $w(\theta)$ is the relative entropy of the probability measure $\theta_\star$ with respect to the probability measure θ on $\{1, \ldots, \mathcal{I}\}$.

Lemma 4.3.7. *We have*

$$\begin{aligned}&\forall\theta \in \Theta,\ w(\theta) \geq 0 \text{ and } w(\theta) = 0 \Leftrightarrow \theta = \theta_\star \\ &\forall\theta \in \Theta,\ h.\nabla w(\theta) \leq 0,\end{aligned}$$

where $h(\theta) = \frac{\theta_\star - \theta}{\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta^a(i)}}$ *is defined in* (4.11).

Proof: Applying Jensen's inequality with the strictly convex function $x\ln(x)$ (with derivative $\ln(x) + 1$ and second order derivative $1/x$), we get

$$w(\theta) = \sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta(i)} \ln\left(\frac{\theta_\star(i)}{\theta(i)}\right)\theta(i) \geq \left(\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta(i)}\theta(i)\right) \ln\left(\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\theta(i)}\theta(i)\right) = 1\ln(1) = 0,$$

with equality iff $\frac{\theta_\star(i)}{\theta(i)}$ does not depend on $i \in \{1, \ldots, \mathcal{I}\}$ i.e., in view of the normalization condition $\sum_{i=1}^{\mathcal{I}} \theta_\star(i) = 1 = \sum_{i=1}^{\mathcal{I}} \theta(i)$, $\theta = \theta_\star$. On the other hand, $\partial_{\theta(i)} w(\theta) = -\frac{\theta_\star(i)}{\theta(i)}$ so that

$$h.\nabla w(\theta) = -\frac{1}{Z^a_\theta} \sum_{i=1}^{\mathcal{I}} (\theta_\star(i) - \theta(i))\frac{\theta_\star(i)}{\theta(i)} = \frac{1}{Z^a_\theta}\left(1 - \sum_{i=1}^{\mathcal{I}} \frac{\theta^2_\star(i)}{\theta(i)}\right).$$

We conclude by remarking that, by Cauchy-Schwarz inequality,

$$1 = \left(\sum_{i=1}^{\mathcal{I}} \sqrt{\theta(i)}\frac{\theta_\star(i)}{\sqrt{\theta(i)}}\right)^2 \leq \sum_{i=1}^{\mathcal{I}} \theta(i) \sum_{i=1}^{\mathcal{I}} \frac{\theta^2_\star(i)}{\theta(i)} = 1 \times \sum_{i=1}^{\mathcal{I}} \frac{\theta^2_\star(i)}{\theta(i)}.$$

■

We are now in position to prove Proposition 4.2.2, by considering successively the three items in Theorem 4.2.1.

(i) The proof of the first item consists in verifying the sufficient conditions given in Theorem 3.0.6 in the chapter 3 dedicated to the stability and convergence of stochastic recursions. Let us rewrite (4.12)

$$\theta_{n+1} = \theta_n + \gamma_{n+1}(h(\theta_n) + \xi_{n+1}) \text{ with } \xi_{n+1} = H(X_{n+1}, \theta_n) - h(\theta_n) + \Lambda_{n+1}.$$

By Lemma 4.3.7, $\nabla w.h$ is non-positive on Θ and $\mathcal{L} := \{\theta \in \Theta : \nabla w.h(\theta) = 0\} = \{\theta_\star\}$.

In order to apply the results in Chapter 3, we modify the parametrization by introducing the open subset $\hat{\Theta} = \{\hat{\theta} \in (0,1)^{\mathcal{I}-1} : \sum_{i=1}^{\mathcal{I}-1} \hat{\theta}(i) < 1\}$ (notice that since for $\theta \in \Theta$, $\sum_{i=1}^{\mathcal{I}} \theta(i) = 1$, θ is contained in an hyperplane of $\mathbb{R}^{\mathcal{I}}$ and therefore not an open subset of $\mathbb{R}^{\mathcal{I}}$) and setting for $\theta \in \Theta$, $\theta_- = (\theta(1), \ldots, \theta(\mathcal{I}-1)) \in \hat{\Theta}$ and for $\hat{\theta} \in \mathbb{R}^{\mathcal{I}-1}$, $\hat{\theta}_+ = (\hat{\theta}(1), \ldots, \hat{\theta}(\mathcal{I}-1), 1 - \sum_{i=1}^{\mathcal{I}-1} \hat{\theta}(i))$. Notice that for $\hat{\theta} \in \hat{\Theta}$, $\hat{\theta}_+ \in \Theta$. We also set $\hat{h}(\hat{\theta}) = h(\hat{\theta}_+)$ and $\hat{w}(\hat{\theta}) = w(\hat{\theta}_+)$. The function $\hat{h}$ is continuous on $\hat{\Theta}$ and $\hat{w}$ is $\mathcal{C}^1$ on $\hat{\Theta}$.

Since for $\hat{\theta} \in \hat{\Theta}$ and $i \in \{1, \ldots, \mathcal{I}-1\}$, $\partial_{\hat{\theta}(i)} \hat{w}(\hat{\theta}) = \partial_{\theta(i)} w(\hat{\theta}_+) - \partial_{\theta(\mathcal{I})} w(\hat{\theta}_+)$, we have

$$\begin{aligned}
\sum_{i=1}^{\mathcal{I}} \frac{\theta_\star(i)}{\hat{\theta}_+^a(i)} \nabla \hat{w}.\hat{h}(\hat{\theta}) &= \sum_{i=1}^{\mathcal{I}-1} (\theta_\star(i) - \hat{\theta}(i)) \partial_{\theta(i)} w(\hat{\theta}_+) - \partial_{\theta(\mathcal{I})} w(\hat{\theta}_+) \sum_{i=1}^{\mathcal{I}-1} (\theta_\star(i) - \hat{\theta}(i)) \\
&= \sum_{i=1}^{\mathcal{I}} (\theta_\star(i) - \hat{\theta}_+(i)) \partial_{\theta(i)} w(\hat{\theta}_+) = \nabla w.h(\hat{\theta}_+),
\end{aligned}$$

where we used that $\sum_{i=1}^{\mathcal{I}-1} (\theta_\star(i) - \hat{\theta}(i)) = 1 - \sum_{i=1}^{\mathcal{I}-1} \hat{\theta}(i) - (1 - \sum_{i=1}^{\mathcal{I}-1} \theta_\star(i)) = \hat{\theta}_+(\mathcal{I}) - \theta_\star(\mathcal{I})$ for the second equality. Hence $\nabla \hat{w}.\hat{h}$ is non-positive on $\hat{\Theta}$ and $\hat{\mathcal{L}} := \{\hat{\theta} \in \hat{\Theta} : \nabla \hat{w}.\hat{h}(\hat{\theta}) = 0\} = \{\theta_{\star -}\}$.

One has

$$\begin{aligned}
\forall \theta \in \Theta,\ w(\theta) &\geq \sum_{i=1}^{\mathcal{I}} \theta_\star(i) \ln(\theta_\star(i)) - \min_{1 \leq i \leq \mathcal{I}} \theta_\star(i) \ln(\min_{1 \leq i \leq \mathcal{I}} \theta(i)) \\
\text{and } \min_{1 \leq i \leq \mathcal{I}} \theta(i) &\geq e^{\frac{1}{\min_{1 \leq i \leq \mathcal{I}} \theta_\star(i)} (\sum_{i=1}^{\mathcal{I}} \theta_\star(i) \ln(\theta_\star(i)) - w(\theta))}.
\end{aligned}$$

Hence $\min_{1 \leq i \leq \mathcal{I}} \hat{\theta}_+(i) \geq e^{\frac{1}{\min_{1 \leq i \leq \mathcal{I}} \theta_\star(i)} \sum_{i=1}^{\mathcal{I}} \theta_\star(i) \ln(\theta_\star(i)) - M_1}$ if $\hat{\theta} \in \mathcal{W}^{M_1} := \{\hat{\theta} \in \hat{\Theta} : \hat{w}(\hat{\theta}) \leq M_1\}$ for $M_1 > 0$. Since $\hat{\theta} \subset [0,1]^{\mathcal{I}-1}$ with the latter a compact subset of $\mathbb{R}^{\mathcal{I}-1}$, from any sequence in $\mathcal{W}^{M_1}$, one may extract a subsequence which converges to some limit $\hat{\theta}_\infty \in [0,1]^{\mathcal{I}-1}$. Moreover by continuity of $[0,1]^{\mathcal{I}-1} \ni \hat{\theta} \mapsto \min_{1 \leq i \leq \mathcal{I}} \hat{\theta}_+(i)$, $\min_{1 \leq i \leq \mathcal{I}} \hat{\theta}_{\infty +}(i) \geq e^{\sum_{i=1}^{\mathcal{I}} \theta_\star(i) \ln(\theta_\star(i)) - M_1}$ so that $\hat{\theta}_\infty \in \hat{\Theta}$. By continuity of w on $\hat{\Theta}$, we deduce that $\hat{w}(\hat{\theta}_\infty) \leq M_1$ so that $\hat{\theta}_\infty \in \mathcal{W}^{M_1}$. Therefore $\mathcal{W}^{M_1}$ is compact.

Let us check that we can find a decomposition $\xi_n = \xi_n^1 + \xi_n^2$ such that for any compact subset $\mathcal{K}$ of $\hat{\Theta}$,

$$\mathbb{P} - a.s. \qquad \lim_k \sup_{\ell \geq k} \left(\left| \sum_{n=k}^{\ell} \gamma_{n+1} \xi_{n+1}^1 \right| + \left| \sum_{n=k}^{\ell} \gamma_{n+1} \xi_{n+1}^2 1_{\theta_{n-} \in \mathcal{K}} \right| \right) = 0. \tag{4.26}$$

A.s., by the recurrence property (4.15) and since $\forall \hat\theta \in \hat\Theta,\ \hat w(\hat\theta) \le \sum_{i=1}^{\mathcal{I}} \theta_\star(i)\ln(\theta_\star(i)) - \ln(\min_{1\le i\le \mathcal{I}} \hat\theta_+(i))$, there exists an integer $M_1 \ge 2$ such that (θ_{n-}) is infinitely often in $\mathcal{W}^{\frac{1+M_1}{2}}$ and, by Corollary 3.0.3 applied with $M_0 = 1$, (γ_n tends to 0 as $n \to \infty$ since $\sum_{n\ge1}\gamma_n^2 < \infty$), θ_{n-} remains in $\mathcal{W}^{\frac{1+3M_1}{2}}$ for n large enough. Since for all integers $M_1 \ge 2$, $\hat{\mathcal{L}} \cap \mathcal{W}^{M_1} = \{\theta_{\star-}\}$ and the image of this set by $\hat w$ is equal to $\{0\}$, Theorem 3.0.6 implies that a.s., as $n \to \infty$, θ_{n-} converges to $\theta_{\star-}$ and therefore θ_n converges to $\theta_\star$.

Let us now check (4.26). By (4.10),

$$\mathbb{P}\left(\forall k, \sup_{\ell\ge k}\left|\sum_{n=k}^{\ell}\gamma_{n+1}\Lambda_{n+1}\right| \le \sqrt{2}\ \sum_{n\ge k}\gamma_{n+1}^2\right) = 1$$

so that (4.14) imply that $\sup_{\ell\ge k}\left|\sum_{n=k}^{\ell}\gamma_{n+1}\Lambda_{n+1}\right|$ converges to 0 a.s. as $k\to\infty$.

Using the Poisson equation (4.21), we write

$$H(X_{n+1},\theta_n) - h(\theta_n) = \widehat{H}_{\theta_n}(X_{n+1}) - P^a_{\theta_n}\widehat{H}_{\theta_n}(X_{n+1}) = \mathcal{E}_{n+1} + R^{(1)}_{n+1} + R^{(2)}_{n+1},$$

with

$$\begin{aligned}\mathcal{E}_{n+1} &= \widehat{H}_{\theta_n}(X_{n+1}) - P^a_{\theta_n}\widehat{H}_{\theta_n}(X_n)\ ,\\ R^{(1)}_{n+1} &= P^a_{\theta_n}\widehat{H}_{\theta_n}(X_n) - P^a_{\theta_{n+1}}\widehat{H}_{\theta_{n+1}}(X_{n+1})\ ,\\ R^{(2)}_{n+1} &= P^a_{\theta_{n+1}}\widehat{H}_{\theta_{n+1}}(X_{n+1}) - P^a_{\theta_n}\widehat{H}_{\theta_n}(X_{n+1})\ .\end{aligned}$$

Recall that γ_{n+1} is $\mathcal{F}_n$-measurable. Let us first check that the martingale $(M_k)_{k\ge1}$ defined by $M_k := \sum_{n=1}^k \gamma_n\mathcal{E}_n$ converges a.s. as $k\to\infty$, which will imply that a.s.

$$\lim_k \sup_{\ell\ge k}\left|\sum_{n=k}^{\ell}\gamma_{n+1}\mathcal{E}_{n+1}\right| = 0\ . \tag{4.27}$$

Let us first prove the result assuming that γ_1 is square integrable, which implies that M_k is also square integrable. Indeed, for all $k\ge1$, $|M_k| \le 2k\gamma_1\sup_{\theta\in\Theta}\sup_{x\in E}\left|\widehat{H}_\theta(x)\right|$. The latter inequality holds since $(\mathcal{E}_n)_{n\ge0}$ is bounded by $2\sup_{\theta\in\Theta}\sup_{x\in E}|\widehat{H}_\theta(x)|$ and, by (4.14), $(\gamma_n)_{n\ge1}$ is bounded by γ_1. Moreover, γ_{n+1} is $\mathcal{F}_n$-measurable and the conditional distribution of X_{n+1} given $\mathcal{F}_n$ is $P^a_{\theta_n}(X_n,\cdot)$, so that

$$\mathbb{E}(\gamma_{n+1}\mathcal{E}_{n+1}\,|\,\mathcal{F}_n) = \gamma_{n+1}\left[\mathbb{E}\left(\widehat{H}_{\theta_n}(X_{n+1})\middle|\,\mathcal{F}_n\right) - P^a_{\theta_n}\widehat{H}_{\theta_n}(X_n)\right] = 0\ .$$

In conclusion, $(M_k)_{k\ge1}$ is a square integrable $\mathcal{F}_k$-martingale. Since

$$\sum_n \mathbb{E}\left[(M_{n+1}-M_n)^2\middle|\,\mathcal{F}_n\right] = \sum_n \gamma_{n+1}^2\mathbb{E}(\mathcal{E}_{n+1}^2|\mathcal{F}_n)$$

is smaller than $C\sum_{n\ge1}\gamma_n^2$ which is a.s. finite by (4.14), $(M_k)_{k\ge1}$ converges a.s. by the martingale convergence theorem. Now, if γ_1 is not square integrable, one can apply the above argument upon replacing γ_1 by $\gamma_1\wedge\Gamma$ where $\Gamma\in\mathbb{N}$ is a constant. This shows that (4.27) holds almost surely on the event $\{\gamma_1<\Gamma\}$, and thus on the event $\cup_{\Gamma=1}^\infty\{\gamma_1<\Gamma\}$. Since the set $\cup_{\Gamma=1}^\infty\{\gamma_1<\Gamma\} = \{\gamma_1<\infty\}$ is of probability one, (4.27) holds almost surely.

We now consider the term $R^{(1)}_{n+1}$.

$$
\begin{aligned}
-\sum_{n=k}^{\ell} R^{(1)}_{n+1} &= \sum_{n=k}^{\ell} \gamma_{n+1} \left(P^a_{\theta_{n+1}} \hat{H}_{\theta_{n+1}}(X_{n+1}) - P^a_{\theta_n} \hat{H}_{\theta_n}(X_n) \right) \\
&= \sum_{n=k+1}^{\ell+1} \gamma_n P^a_{\theta_n} \hat{H}_{\theta_n}(X_n) - \sum_{n=k}^{\ell} \gamma_{n+1} P^a_{\theta_n} \hat{H}_{\theta_n}(X_n) \\
&= \gamma_{\ell+1} P^a_{\theta_{\ell+1}} \hat{H}_{\theta_{\ell+1}}(X_{\ell+1}) + \sum_{n=k+1}^{\ell} (\gamma_n - \gamma_{n+1}) P^a_{\theta_n} \hat{H}_{\theta_n}(X_n) - \gamma_{k+1} P^a_{\theta_k} \hat{H}_{\theta_k}(X_k)
\end{aligned}
$$

By the monotonicity of the sequence $(\gamma_n)_n$ and since, by (4.23), $\widehat{H}_\theta$ is uniformly bounded in (θ, x), we deduce the existence of a finite constant C such that

$$
\left| \sum_{n=k}^{\ell} R^{(1)}_{n+1} \right| \leq C \left(\gamma_{\ell+1} + \sum_{n=k+1}^{\ell} (\gamma_n - \gamma_{n+1}) + \gamma_{k+1} \right) = 2C\gamma_{k+1}
$$

Since, by (4.14), a.s. $\gamma_{k+1} \to 0$ as $k \to \infty$, we deduce that the same is true for $\sup_{\ell \geq k} \left| \sum_{n=k}^{\ell} \gamma_{n+1} R^{(1)}_{n+1} \right|$.

We now consider the term $R^{(2)}_{n+1}$. By Lemma 4.3.5, there exists a constant C such that for any $\theta_1, \theta_2 \in \Theta$

$$
\sup_E \left| P^a_{\theta_1} \widehat{H}_{\theta_1} - P^a_{\theta_2} \widehat{H}_{\theta_2} \right| \leq C \left(|\theta_1 - \theta_2| + \sum_{i=1}^{\mathcal{I}} \left| 1 - \frac{\theta^a_2(i)}{\theta^a_1(i)} \right| \right) . \tag{4.28}
$$

For any compact subset $\mathcal{K}$ of $\hat{\Theta}$, we have $\inf_{\hat{\theta} \in \mathcal{K}} \min_{1 \leq i \leq \mathcal{I}} \hat{\theta}_+(i) > 0$. Moreover, for $a \in (0,1)$ and $t, t' \in (0,1)$ either $t' \leq \frac{t}{2}$ and $|1 - \left(\frac{t'}{t}\right)^a| \leq 1 \leq \frac{2}{t}|t'-t|$ or $t' \geq \frac{t}{2}$ and

$$
\left| 1 - \left(\frac{t'}{t} \right)^a \right| = \frac{|(t')^a - t^a|}{t^a} \leq \frac{a}{t^a} \left(\frac{t}{2} \right)^{a-1} |t' - t| = \frac{a}{2^{a-1} t} |t' - t|.
$$

We deduce that there exists a constant C such that for any $n \geq 0$,

$$
\sum_{i=1}^{\mathcal{I}} \left| 1 - \frac{\theta^a_{n+1}(i)}{\theta^a_n(i)} \right| 1_{\theta_n - \in \mathcal{K}} \leq C \, |\theta_{n+1} - \theta_n| \, . \tag{4.29}
$$

Moreover, by (4.9), (4.10) and the boundedness of H there exists a finite constant C such that with probability one, for any $n \geq 0$,

$$
|\theta_{n+1} - \theta_n| \leq C \, \gamma_{n+1} \, . \tag{4.30}
$$

Therefore, combining (4.28)–(4.29)–(4.30), there exists a finite constant C such that

$$
\mathbb{P} \left(\forall k, \sup_{\ell \geq k} \left| \sum_{n=k}^{\ell} \gamma_{n+1} R^{(2)}_{n+1} 1_{\theta_n - \in \mathcal{K}} \right| \leq C \sum_{n \geq k} \gamma^2_{n+1} \right) = 1 \, .
$$

By (4.14), $\sup_{\ell \geq k} \left| \sum_{n=k}^{\ell} \gamma_{n+1} R^{(2)}_{n+1} 1_{\theta_n - \in \mathcal{K}} \right|$ tends to zero a.s. as $k \to \infty$. Therefore (4.26) holds for $\xi^1_n = \Lambda_n + \mathcal{E}_n + R^{(1)}_n$ and $\xi^2_n = R^{(2)}_n$. This concludes the proof of the a.s. convergence : $\lim_{n \to \infty} \theta_n = \theta_\star$.

(ii) Let $f : E \to \mathbb{R}$ be measurable and bounded and $n \geq k$. We have

$$\begin{aligned}\mathbb{E}[f(X_n)] - \pi^a_{\theta_\star}(f) =& \mathbb{E}[f(X_n) - (P^a_{\theta_{n-k}})^k f(X_{n-k})] + \mathbb{E}[(P^a_{\theta_{n-k}})^k f(X_{n-k}) - \pi^a_{\theta_{n-k}}(f)] \\ &+ \mathbb{E}[\pi^a_{\theta_{n-k}}(f)] - \pi^a_{\theta_\star}(f).\end{aligned}$$

The first difference in the right-hand side rewrites

$$\begin{aligned}\mathbb{E}[f(X_n)] - \mathbb{E}[(P^a_{\theta_{n-k}})^k f(X_{n-k})] &= \sum_{\ell=0}^{k-1} \mathbb{E}[(P^a_{\theta_{n-k}})^\ell f(X_{n-\ell}) - (P^a_{\theta_{n-k}})^{\ell+1} f(X_{n-\ell-1})] \\ &= \sum_{\ell=0}^{k-1} \mathbb{E}[P^a_{\theta_{n-\ell-1}}(P^a_{\theta_{n-k}})^\ell f(X_{n-\ell-1}) - (P^a_{\theta_{n-k}})^{\ell+1} f(X_{n-\ell-1})]\end{aligned}$$

using that for $\ell \leq k-1$, since θ_{n-k} is $\mathcal{F}_{n-\ell-1}$-measurable, $\mathbb{E}[(P^a_{\theta_{n-k}})^\ell f(X_{n-\ell}) | \mathcal{F}_{n-\ell-1}] = P^a_{\theta_{n-\ell-1}}(P^a_{\theta_{n-k}})^\ell f(X_{n-\ell-1})$ for the second equality. Since $\sup_E |(P^a_{\theta_{n-k}})^\ell f| \leq \sup_E |f|$, we have $|P^a_{\theta_{n-\ell-1}}(P^a_{\theta_{n-k}})^\ell f(X_{n-\ell-1}) - (P^a_{\theta_{n-k}})^{\ell+1} f(X_{n-\ell-1})| \leq 2 \sup_E |f| d_{\mathrm{TV}}(P^a_{\theta_{n-\ell-1}}, P^a_{\theta_{n-k}})$ by Proposition 1.3.3 and using Lemma 4.3.4, we deduce that

$$\begin{aligned}|\mathbb{E}[f(X_n)] - \mathbb{E}[(P^a_{\theta_{n-k}})^k f(X_k)]| &\leq 2 \sup_E |f| \sum_{\ell=0}^{k-1} \mathbb{E}\left[d_{\mathrm{TV}}(P^a_{\theta_{n-\ell-1}}, P^a_{\theta_{n-k}})\right] \\ &\leq 2 \sup_E |f| \sum_{\ell=0}^{k-1} \mathbb{E}\left[1 \wedge 2 \max_{1 \leq i \leq \mathcal{I}} \left|1 - \frac{\theta^a_{n-\ell-1}(i)}{\theta^a_{n-k}(i)}\right|\right].\end{aligned}$$

Next, by (4.18), $|(P^a_{\theta_{n-k}})^k f(X_k) - \pi^a_{\theta_{n-k}}(f)| \leq 2 \sup_E |f| \sup_{\theta \in \Theta} \sup_{x \in E} d_{\mathrm{TV}}((P^a_\theta)^n(x, \cdot), \pi^a_\theta) \leq 2 \sup_E |f| (1 - \tilde{\alpha})^k$. Last, by Lemma 4.3.3,

$$|\pi^a_{\theta_{n-k}}(f) - \pi^a_{\theta_\star}(f)| \leq 2 \sup_E |f| d_{\mathrm{TV}}(\pi^a_{\theta_{n-k}}, \pi^a_{\theta_\star}) \leq 2 \sup_E |f| \left(1 \wedge (\mathcal{I} - 1) \sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta^a_\star(i)}{\theta^a_{n-k}(i)}\right|\right). \tag{4.31}$$

Hence

$$\begin{aligned}\frac{1}{2 \sup_E |f|} |\mathbb{E}[f(X_n)] - \pi^a_{\theta_\star}(f)| &\leq (1 - \tilde{\alpha})^k \\ &+ \mathbb{E}\left[\sum_{\ell=0}^{k-1} 1 \wedge 2 \max_{1 \leq i \leq \mathcal{I}} \left|1 - \frac{\theta^a_{n-\ell-1}(i)}{\theta^a_{n-k}(i)}\right| + 1 \wedge (\mathcal{I} - 1) \sum_{i=1}^{\mathcal{I}} \left|1 - \frac{\theta^a_\star(i)}{\theta^a_{n-k}(i)}\right|\right].\end{aligned}$$

The a.s. convergence of $(\theta_n)_n$ to $\theta_\star$ ensures with Lebesgue theorem that the expectation in the right-hand side tends to 0 as $n \to \infty$. Hence $\limsup_{n \to \infty} |\mathbb{E}[f(X_n)] - \pi^a_{\theta_\star}(f)| \leq 2 \sup_E |f| (1 - \tilde{\alpha})^k$. By letting $k \to \infty$, we conclude that $\mathbb{E}[f(X_n)]$ tends to $\pi_{\theta_\star}(f)$ as $n \to \infty$.

Since, by (4.31), the a.s. convergence of θ_n to $\theta_\star$ as $n \to \infty$, the Cesaro mean $\frac{1}{n} \sum_{k=1}^n \pi^a_{\theta_{k-1}}(f)$ a.s. converges to $\pi^a_{\theta_\star}(f)$, to prove the a.s. convergence of $\frac{1}{n} \sum_{k=1}^n f(X_k)$ to $\pi_{\theta_\star}(f)$ it is enough to check the a.s. convergence of $\frac{1}{n} \sum_{k=1}^n (f(X_k) - \pi^a_{\theta_{k-1}}(f))$ to 0. By a reasoning similar (but easier since f does not depend on θ) to the proof of Lemma 4.3.5, we obtain that $F_\theta = \sum_{n \in \mathbb{N}} ((P^a_\theta)^n f - \pi^a_\theta(f))$ solves the Poisson equation $F_\theta - P^a_\theta F_\theta = f - \pi^a_\theta(f)$, $\pi^a_\theta(F_\theta) = 0$ and satisfies $\sup_{\theta \in \Theta, x \in E} |F_\theta(x)| \leq \frac{2 \sup_E |f|}{\tilde{\alpha}}$. Moreover, reasoning like in the proof of Lemma 4.3.6 (using Lemma 4.3.3 to estimate $\pi^a_{\theta_1}(f)-\pi^a_{\theta_2}(f)$), we obtain

$$\exists C<\infty,\ \forall \theta_1,\theta_2\in\Theta,\ \sup_{x\in E}|F_{\theta_1}(x)-F_{\theta_2}(x)|\leq C\sum_{i=1}^{\mathcal{I}}\left|1-\frac{\theta_2^a(i)}{\theta_1^a(i)}\right|. \tag{4.32}$$

We have

$$\frac{1}{n}\sum_{k=1}^{n}(f(X_k)-\pi^a_{\theta_{k-1}}(f))=\frac{1}{n}\sum_{k=1}^{n}(F_{\theta_{k-1}}(X_k)-P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_k))$$
$$=\frac{1}{n}\sum_{k=1}^{n}(F_{\theta_{k-1}}(X_k)-P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_{k-1}))+\frac{1}{n}\sum_{k=1}^{n}(P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_{k-1})-P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_k))$$

Since $\mathbb{E}[F_{\theta_{k-1}}(X_k)|\mathcal{F}_{k-1}]=P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_{k-1})$, the difference $m_k=F_{\theta_{k-1}}(X_k)-P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_{k-1})$ is a martingale increment. The finiteness of $\sup_{\theta\in\Theta,x\in E}|F_\theta(x)|$ ensures that $\sum_{k\geq 1}\frac{1}{k^2}\mathbb{E}\left[m_k^2\right]<\infty$ so that $\sum_{k=1}^n\frac{m_k}{k}$ a.s. converges as $n\to\infty$. Since

$$\frac{1}{n}\sum_{k=1}^{n}m_k=\frac{1}{n}\sum_{k=1}^{n}k\left(\sum_{j=1}^{k}\frac{m_j}{j}-\sum_{j=1}^{k-1}\frac{m_j}{j}\right)\overset{\ell=k-1}{=}\frac{1}{n}\sum_{k=1}^{n}k\sum_{j=1}^{k}\frac{m_j}{j}-\frac{1}{n}\sum_{\ell=1}^{n-1}(\ell+1)\sum_{j=1}^{\ell}\frac{m_j}{j}$$
$$=\sum_{j=1}^{n}\frac{m_j}{j}-\frac{1}{n}\sum_{\ell=1}^{n-1}\sum_{j=1}^{\ell}\frac{m_j}{j}$$

where, by Cesaro's lemma, the two terms in the right-hand side have the same limit as $n\to\infty$ $\frac{1}{n}\sum_{k=1}^n m_k$ a.s. converges to 0. On the other hand,

$$\frac{1}{n}\sum_{k=1}^{n}(P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_{k-1})-P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_k))=\frac{P^a_{\theta_0}F_{\theta_0}(X_0)-P^a_{\theta_{n-1}}F_{\theta_{n-1}}(X_n)}{n}$$
$$+\frac{1}{n}\sum_{k=1}^{n-1}(P^a_{\theta_k}F_{\theta_k}(X_k)-P^a_{\theta_{k-1}}F_{\theta_{k-1}}(X_k)).$$

By the finiteness of $\sup_{\theta\in\Theta,x\in E}|F_\theta(x)|$, the first term in the right-hand side a.s. converges to 0 as $n\to\infty$. Moreover, by (4.32) and (4.19)

$$|P^a_{\theta_1}F_{\theta_1}(x)-P^a_{\theta_2}F_{\theta_2}(x)|\leq|P^a_{\theta_1}F_{\theta_1}(x)-P^a_{\theta_1}F_{\theta_2}(x)|+|P^a_{\theta_1}F_{\theta_2}(x)-P^a_{\theta_2}F_{\theta_2}(x)|\leq C\sum_{i=1}^{\mathcal{I}}\left|1-\frac{\theta_2^a(i)}{\theta_1^a(i)}\right|.$$

With the a.s. convergence of θ_n to $\theta_\star$, we deduce that the second term in the right-hand side also a.s. converges to 0 and conclude that $\frac{1}{n}\sum_{k=1}^n f(X_k)$ a.s. converges to $\pi^a_{\theta_\star}(f)$.

(iii) Let $f:E\to\mathbb{R}$ be measurable and bounded. We have

$$\mathbb{E}\left[\left(\sum_{i=1}^{\mathcal{I}}\theta_{n-1}^{1-a}(i)\right)\left(\sum_{j=1}^{\mathcal{I}}\theta_{n-1}^{a}(j)\,f(X_n)1_{E_j}(X_n)\right)\right]-\int_E f(x)\pi(dx)$$
$$=\mathbb{E}\left[\sum_{j=1}^{\mathcal{I}}\left(\theta_{n-1}^a(j)\sum_{i=1}^{\mathcal{I}}\theta_{n-1}^{1-a}(i)-\theta_\star^a(j)\sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)\right)f(X_n)1_{E_j}(X_n)\right]$$
$$+\mathbb{E}\left[\sum_{j=1}^{\mathcal{I}}\theta_\star^a(j)\sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)f(X_n)1_{E_j}(X_n)\right]-\int_E f(x)\pi(dx) \tag{4.33}$$

The a.s. convergence of θ_n to $\theta_\star$ as $n \to \infty$ and the boundedness of f ensure that

$$\sum_{j=1}^{\mathcal{I}} \left(\theta_{n-1}^a(j) \sum_{i=1}^{\mathcal{I}} \theta_{n-1}^{1-a}(i) - \theta_\star^a(j) \sum_{i=1}^{\mathcal{I}} \theta_\star^{1-a}(i) \right) f(X_n) 1_{E_j}(X_n)$$

converges a.s. to 0. With the inequality

$$\sup_{\theta \in \Theta} \max_{1 \le j \le \mathcal{I}} \theta^a(j) \sum_{i=1}^{\mathcal{I}} \theta^{1-a}(i) \le \mathcal{I}, \tag{4.34}$$

and the boundedness of f, we deduce by Lebesgue theorem that the first term in the right-hand side of (4.33) converges to 0 as $n \to \infty$. Moreover, the function $E \ni x \mapsto \left(\sum_{i=1}^{\mathcal{I}} \theta_\star^{1-a}(i)\right) \sum_{j=1}^{\mathcal{I}} \theta_\star^a(j) \, f(x) 1_{E_j}(x)$ is measurable and bounded and such that $\int_E \sum_{i=1}^{\mathcal{I}} \theta_\star^{1-a}(i) \sum_{j=1}^{\mathcal{I}} \theta_\star^a(j) \, f(x) 1_{E_j}(x) \pi_{\theta_\star}^a(dx) = \int_E f(x) \pi(dx)$ so that the difference between the second and the third terms also converges to 0 by (ii). Hence

$$\lim_{n \to \infty} \mathbb{E} \left[\left(\sum_{i=1}^{\mathcal{I}} \theta_{n-1}^{1-a}(i) \right) \left(\sum_{j=1}^{\mathcal{I}} \theta_{n-1}^a(j)) \, f(X_n) 1_{E_j}(X_n) \right) \right] = \int_E f(x) \, \pi(dx).$$

Again by (ii), a.s.

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{\mathcal{I}} \theta_\star^{1-a}(i) \sum_{j=1}^{\mathcal{I}} \theta_\star^a(j) \, f(X_k) 1_{E_j}(X_k) = \int_E f(x) \, \pi(dx).$$

Since the Cesaro mean

$$\frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{\mathcal{I}} \left(\theta_{k-1}^a(j) \sum_{i=1}^{\mathcal{I}} \theta_{k-1}^{1-a}(i) - \theta_\star^a(j) \sum_{i=1}^{\mathcal{I}} \theta_\star^{1-a}(i) \right) f(X_k) 1_{E_j}(X_k)$$

converges a.s. to 0 as $n \to \infty$, we deduce that

$$a.s., \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{i=1}^{\mathcal{I}} \theta_{k-1}^{1-a}(i) \right) \left(\sum_{j=1}^{\mathcal{I}} \theta_{k-1}^a(j) \, f(X_k) 1_{E_j}(X_k) \right) = \int_E f(x) \, \pi(dx).$$

4.4 Proof of Proposition 4.2.3 : recurrence of the algorithm

In all this section, we consider that the sequence is generated by the WL_a algorithm 2.

The aim of this section is to give some sufficient conditions on $(\gamma_n)_{n \ge 1}$ such that $\mathbb{P}$-a.s., the sequence $(\theta_{n-})_{n \ge 0}$ visits a.s. infinitely often a compact subset of $\hat{\Theta}$. For $n \ge 0$, we set

$$\underline{\theta}_n := \min_{1 \le i \le \mathcal{I}} \theta_n(i).$$

The objective is thus to verify that a.s. the sequence $(\underline{\theta}_n)_{n \ge 0}$ takes infinitely often values in a compact subset of $(0, 1)$. We will show this property along a sequence of well chosen stopping times $(T_k)_{k \ge 0}$ defined inductively as follows.

Let $R > 1$. We set $T_0 = 0$ and for $k \in \mathbb{N}$, $T_{k+1} = \infty$ if $T_k = \infty$. Otherwise when $T_k < +\infty$, let for $m \in \mathbb{N}$, $\theta_{T_k+m\mathcal{I}}((1)_m) \leq \theta_{T_k+m\mathcal{I}}((2)_m) \leq \ldots \leq \theta_{T_k+m\mathcal{I}}((\mathcal{I})_m)$ denote the increasing reordering of $(\theta_{T_k+m\mathcal{I}}(i))_{1\leq i\leq \mathcal{I}}$ and

$$i_m := \max\{i \leq \mathcal{I} : \theta_{T_k+m\mathcal{I}}((i)_m) < R\underline{\theta}_{T_k+m\mathcal{I}}\}.$$

We then introduce an event corresponding to visiting successively the strata of small weights with indices $(i)_m$ for $i \in \{1, \ldots, i_m\}$, in decreasing order :

$$A_m := \Big\{X_{T_k+m\mathcal{I}+1} \in E_{(i_m)_m}, X_{T_k+m\mathcal{I}+2} \in E_{(i_m-1)_m}, \ldots, X_{T_k+m\mathcal{I}+i_m} \in E_{(1)_m}\Big\}. \tag{4.35}$$

The next stopping T_{k+1} is then defined by

$$T_{k+1}(\omega) = T_k(\omega) + \mathcal{I} \times \inf\{m \geq 1 : \omega \in A_{m-1}\} \text{ with convention } \inf \emptyset = +\infty.$$

Note that $T_k \geq k\mathcal{I}$ by definition. Let us first show some additional properties on this sequence of stopping times.

Lemma 4.4.1. *Assume A1 and A2. Then,* $\mathbb{P}(\forall k \in \mathbb{N},\ T_k < +\infty) = 1$ *and*

$$\exists p \in (0,1),\ \forall k, m \in \mathbb{N},\ \mathbb{P}\left(T_{k+1} - T_k > m\mathcal{I} | \mathcal{F}_{T_k}\right) \leq (1-p)^m. \tag{4.36}$$

In addition,

$$\mathbb{P}\left(\exists C_\star < +\infty,\ \forall k \in \mathbb{N},\ T_k \leq C_\star k\right) = 1. \tag{4.37}$$

Proof: The first two statements are a consequence of

$$\exists p \in (0,1),\ \forall k, m \in \mathbb{N} \text{ with } T_k < \infty,\ \mathbb{P}(A_m|\mathcal{F}_{T_k+m\mathcal{I}}) \geq p. \tag{4.38}$$

The main ingredient in the proof of this inequality is the following one :

$$\begin{aligned}
\forall x \in E,\ \forall i \in \{1,\ldots,\mathcal{I}\},\ P^a_\theta(x, E_i) &= \int_{E_i} q(x,y)\left(1 \wedge \frac{\theta^a(I(x))\,\eta(y)}{\theta^a(i)\,\eta(x)}\right)\lambda(dy) \\
&\geq \left(\inf_{E^2} q\right)\int_{E_i}\left(1 \wedge \frac{\theta^a(I(x))\,\eta(y)}{\theta^a(i)\,\sup_E \eta}\right)\lambda(dy) \\
&\geq \left(\inf_{E^2} q\right)\int_{E_i}\left(\frac{\eta(y)}{\sup_E \eta} \wedge \frac{\theta^a(I(x))\,\eta(y)}{\theta^a(i)\,\sup_E \eta}\right)\lambda(dy) \\
&= c\,\theta_\star(i)\left(\frac{\theta^a(I(x))}{\theta^a(i)} \wedge 1\right),
\end{aligned} \tag{4.39}$$

where $c = \frac{\inf_{E^2} q}{\sup_E \eta} > 0$ by A1 and A2. Now, by (4.9), for $j \in \{1, \ldots, i_m - 1\}$, it holds on the event $\{X_{T_k+m\mathcal{I}+1} \in E_{(i_m)_m}, \ldots, X_{T_k+m\mathcal{I}+j} \in E_{(i_m+1-j)_m}\}$,

$$\begin{aligned}
&\frac{\theta_{T_k+m\mathcal{I}+j}((i_m+1-j)_m)}{\theta_{T_k+m\mathcal{I}+j}((i_m-j)_m)} \\
&= \frac{\theta_{T_k+m\mathcal{I}}((i_m+1-j)_m)}{\theta_{T_k+m\mathcal{I}}((i_m-j)_m)} \times \left(1 + \gamma_{T_k+m\mathcal{I}+j}\theta^{a-1}_{T_k+m\mathcal{I}+j-1}((i_m+1-j)_m)\right)\ .
\end{aligned}$$

Both factors on the right-hand side are larger than 1 (the first one by definition of the ordered indices $(i)_m$), so that, by (4.39) and the monotonicity of ρ,

$$P^a_{\theta_{T_k+m\mathcal{I}+j}}\left(X_{T_k+m\mathcal{I}+j}, E_{(i_m-j)_m}\right) \geq c\,\theta_\star((i_m-j)_m) \geq c\,\underline{\theta}_\star\,,$$

where $\underline{\theta}_\star = \min_{1\leq i\leq \mathcal{I}} \theta_\star(i)$. Note that this implies in particular that $c\,\underline{\theta}_\star \leq 1$. Using successively the strong Markov property of the chain $(X_n, \theta_n)_{n\geq 0}$, a backward induction on n, the definition of i_m, together with (4.39), we have

$$\begin{aligned}
&\mathbb{P}(A_m|\mathcal{F}_{T_k+m\mathcal{I}})\\
&= \mathbb{E}\left(1_{\{X_{T_k+m\mathcal{I}+1}\in E_{(i_m)_m},\ldots,X_{T_k+m\mathcal{I}+i_m-1}\in E_{(2)_m}\}} P^a_{\theta_{T_k+m\mathcal{I}+i_m-1}}(X_{T_k+m\mathcal{I}+i_m-1}, E_{(1)_m})|\mathcal{F}_{T_k+m\mathcal{I}}\right)\\
&\geq c\,\underline{\theta}_\star \mathbb{P}\left(1_{\{X_{T_k+m\mathcal{I}+1}\in E_{(i_m)_m},\ldots,X_{T_k+m\mathcal{I}+i_m-2}\in E_{(3)_m}\}} P^a_{\theta_{T_k+m\mathcal{I}+i_m-2}}(X_{T_k+m\mathcal{I}+i_m-2}, E_{(2)_m})\Big|\,\mathcal{F}_{T_k+m\mathcal{I}}\right)\\
&\geq \left(c\,\underline{\theta}_\star\right)^{i_m-1} P^a_{\theta_{T_k+m\mathcal{I}}}(X_{T_k+m\mathcal{I}}, E_{(i_m)_m})\\
&\geq \left(c\,\underline{\theta}_\star\right)^{i_m-1} c\,\underline{\theta}_\star \frac{\theta^a_{T_k+m\mathcal{I}}(I(X_{T_k+m\mathcal{I}}))}{\theta^a_{T_k+m\mathcal{I}}((i_m)_m)}\\
&\geq \left(c\,\underline{\theta}_\star\right)^{\mathcal{I}} \frac{\underline{\theta}^a_{T_k+m\mathcal{I}}}{\theta^a_{T_k+m\mathcal{I}}((i_m)_m)} \geq \left(c\,\underline{\theta}_\star\right)^{\mathcal{I}} \frac{(\theta_{T_k+m\mathcal{I}}((i_m)_m)/R)^a}{\theta^a_{T_k+m\mathcal{I}}((i_m)_m)} \geq \left(c\,\underline{\theta}_\star\right)^{\mathcal{I}} R^{-a}.
\end{aligned}$$

The proof is therefore concluded by setting

$$p = \left(c\,\underline{\theta}_\star\right)^{\mathcal{I}} R^{-a}.$$

This concludes the proof of (4.38) and thus of the first two statements of Lemma 4.4.1. The third statement can be deduced from the second one by a coupling argument. Let us deduce from (4.36) that it is possible to couple a sequence $(\tilde{T}_k)_{k\geq 0}$ with the same law as $(T_k)_{k\geq 0}$ with a sequence $(\tau_k)_{k\geq 1}$ of independent geometric random variables with parameter p in such a way that

$$\text{a.s. }, \forall k \in \mathbb{N},\, \tilde{T}_{k+1} - \tilde{T}_k \leq \mathcal{I}\tau_{k+1}. \tag{4.40}$$

Let $(U_k)_{k\geq 1}$ be a sequence of random variables i.i.d. according to the uniform law on $(0,1)$. The càg pseudo-inverse $(0,1) \ni u \mapsto F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ of the cumulative distribution function $F(x) = 1_{\{x\geq 0\}}\left(1-(1-p)^{\lfloor x\rfloor}\right)$ of the geometric law with parameter p is given by $F^{-1}(u) = \lceil \frac{\ln(1-u)}{\ln(1-p)}\rceil$ and by the inverse sampling technique, the random variables $\tau_k = \lceil \frac{\ln(1-U_k)}{\ln(1-p)}\rceil$ are i.i.d. according to this geometric law. Now, define

$$F_{(T_0,\ldots,T_k)}(x) = \mathbb{P}\left(\frac{T_{k+1}-T_k}{\mathcal{I}} \leq x\,\middle|\,(T_0,\ldots,T_k)\right)$$

for $k \geq 0$. Since the random vector $(T_0,\ldots,T_k)$ is $\mathcal{F}_{T_k}$-measurable, for any $x \geq 0$,

$$\begin{aligned}
F_{(T_0,\ldots,T_k)}(x) \geq F_{(T_0,\ldots,T_k)}(\lfloor x\rfloor) &= 1 - \mathbb{E}\Big[\mathbb{P}\big(T_{k+1}-T_k > \lfloor x\rfloor\mathcal{I}\,\big|\,\mathcal{F}_{T_k}\big)\Big|(T_0,\ldots,T_k)\Big]\\
&\geq 1-(1-p)^{\lfloor x\rfloor} = F(x).
\end{aligned}$$

where we used (4.36) for the second inequality. Hence for all $k \geq 1$ and $u \in (0,1)$, $F_{(T_0,\ldots,T_k)}(u) := \inf\{x \in \mathbb{R} : F_{(T_0,\ldots,T_k)}(x) \geq u\} \leq F^{-1}(u)$.

The sequence $(\tilde{T}_k)_{k\geq 0}$ defined inductively by $\tilde{T}_0 = 0$ and

$$\forall k \in \mathbb{N}, \quad \tilde{T}_{k+1} = \tilde{T}_k + \mathcal{I}\, F^{-1}_{(\tilde{T}_0,\dots,\tilde{T}_k)}(U_{k+1}),$$

satisfies the required properties : it has the same law as $(T_k)_{k\geq 0}$ by the inverse sampling technique and

$$\tilde{T}_{k+1} - \tilde{T}_k = \mathcal{I}F^{-1}_{(\tilde{T}_0,\dots,\tilde{T}_k)}(U_{k+1}) \leq \mathcal{I}F^{-1}(U_k) = \mathcal{I}\tau_k.$$

As a consequence,

$$\mathbb{P}\left(\limsup_{k\to\infty} \frac{T_k}{k} \leq \frac{\mathcal{I}}{p}\right) = \mathbb{P}\left(\limsup_{k\to\infty} \frac{\tilde{T}_k}{k} \leq \frac{\mathcal{I}}{p}\right) \geq \mathbb{P}\left(\limsup_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} \tau_j \leq \frac{1}{p}\right) = 1\ ,$$

the last equality being a consequence of the strong law of large numbers. This concludes the proof of (4.37). ∎

We are now in position to state the main result of this section.

Lemma 4.4.2. *Let $a \in [0,1)$. Assume A1, A2 and that the sequence $(\gamma_n)_{n\geq 1}$ is non-increasing, bounded from above by a deterministic sequence converging to 0 as $n \to \infty$ and such that $\bar{r}_{\mathcal{I},\gamma} < \infty$, where, we recall (see (4.16))*

$$\bar{r}_{\mathcal{I},\gamma} = \sup_{n\geq 1} \frac{\gamma_n}{\gamma_{n+\mathcal{I}-1}}.$$

Then

$$\mathbb{P}\left(\limsup_{k\to\infty} \underline{\theta}_{T_k-\mathcal{I}} > 0\right) = 1.$$

Since $\lim_{k\to\infty} T_k = \infty$ and by Lemma 4.4.1, $\forall k \geq 0$, $T_k < \infty$ almost surely, Proposition 4.2.3 is an immediate consequence of Lemma 4.4.2.

Proof: The argument follows the proof of the second statement in [24, Proposition 4.1]. For $k \geq 1$, we set $Y_k := \underline{\theta}_{T_k-\mathcal{I}}$. As a preliminary result, let us first prove that there exists $\underline{k} \in \mathbb{N}\setminus\{0\}$ and $\bar{y} \in (0,1)$ such that

$$\forall k \geq \underline{k}, \quad Y_k \leq \bar{y} \Longrightarrow \mathbb{E}(\ln(Y_{k+1})|\mathcal{F}_{T_k}) \geq \ln(Y_k). \tag{4.41}$$

Let us recall (4.9) :

$$\forall i \in \{1,\dots,\mathcal{I}\},\ \forall n \in \mathbb{N},\ \theta_{n+1}(i) = \theta_n(i)\frac{1+\gamma_{n+1}1_{E_i}(X_{n+1})\theta_n^{a-1}(i)}{1+\gamma_{n+1}\theta_n^a(I(X_{n+1}))}.$$

One deduces that, on the one hand, for any index $i \in \{1,\dots,\mathcal{I}\}$ such that $\theta_{T_k-\mathcal{I}}(i) \geq R\underline{\theta}_{T_k-\mathcal{I}}$, one has

$$\begin{aligned}\theta_{T_{k+1}-\mathcal{I}}(i) &\geq \theta_{T_k-\mathcal{I}}(i)\prod_{j=T_k-\mathcal{I}+1}^{T_{k+1}-\mathcal{I}} \frac{1}{1+\gamma_j\,\theta^a_{j-1}(I(X_j))} \geq \theta_{T_k-\mathcal{I}}(i)\left(\frac{1}{1+\gamma_{T_k-\mathcal{I}+1}}\right)^{T_{k+1}-T_k} \\ &\geq R\left(\frac{1}{1+\gamma_{T_k-\mathcal{I}+1}}\right)^{T_{k+1}-T_k}\underline{\theta}_{T_k-\mathcal{I}},\end{aligned} \tag{4.42}$$

where we used the monotonicity of the sequence $(\gamma_n)_{n\geq 1}$ for the second inequality. On the other hand, by definition of T_k, any stratum with index $i \in \{1,\dots,\mathcal{I}\}$ such that $\theta_{T_k-\mathcal{I}}(i) < R\underline{\theta}_{T_k-\mathcal{I}}$ is visited at least once between the times $T_k - \mathcal{I} + 1$ and T_k so that, using that $\theta_n(i)$ decreases for n between $T_k - \mathcal{I}$ and this visit, as well as the monotonicities of $t \mapsto t^{a-1}$ and $n \mapsto \gamma_n$,

$$\begin{aligned}\theta_{T_{k+1}-\mathcal{I}}(i) &\geq \theta_{T_k-\mathcal{I}}(i)\left(1+\gamma_{T_k}(R\underline{\theta}_{T_k-\mathcal{I}})^{a-1}\right)\left(\frac{1}{1+\gamma_{T_k-\mathcal{I}+1}}\right)^{T_{k+1}-T_k}\\ &\geq \left(1+\gamma_{T_k}(R\underline{\theta}_{T_k-\mathcal{I}})^{a-1}\right)\left(\frac{1}{1+\gamma_{T_k-\mathcal{I}+1}}\right)^{T_{k+1}-T_k}\underline{\theta}_{T_k-\mathcal{I}}.\end{aligned} \tag{4.43}$$

Combining (4.42) and (4.43), one deduces that

$$Y_{k+1} \geq \left(R \wedge \left(1+\gamma_{T_k}(RY_k)^{a-1}\right)\right)\left(\frac{1}{1+\bar{r}_{\mathcal{I},\gamma}\gamma_{T_k}}\right)^{T_{k+1}-T_k} Y_k.$$

Taking the logarithm and remarking that the second statement in Lemma 4.4.1 implies that $\mathbb{E}(T_{k+1}-T_k|\mathcal{F}_{T_k}) \leq \frac{\mathcal{I}}{p}$, one obtains that

$$\mathbb{E}(\ln(Y_{k+1})|\mathcal{F}_{T_k}) - \ln(Y_k) \geq \left(\ln(R) \wedge \ln\left(1+\gamma_{T_k}(RY_k)^{a-1}\right)\right) - \frac{\mathcal{I}}{p}\ln\left(1+\bar{r}_{\mathcal{I},\gamma}\gamma_{T_k}\right).$$

Since $T_k \geq k\mathcal{I}$ and the sequence $(\gamma_n)_{n\geq 1}$ is bounded from above by a deterministic sequence converging to 0 one can choose $\underline{k} \in \mathbb{N}\setminus\{0\}$ such that $\forall k \geq \underline{k}$, $\gamma_{T_k} \leq \bar{\gamma} = \frac{R^{\frac{p}{\mathcal{I}}}-1}{\bar{r}_{\mathcal{I},\gamma}}$ so that $\ln(R) - \frac{\mathcal{I}}{p}\ln(1+\bar{r}_{\mathcal{I},\gamma}\gamma_{T_k}) \geq 0$. Last, since $\lim_{t\to 0^+} t^{a-1} = +\infty$, one may choose $\bar{y} \in (0, 1/R)$ such that

$$\inf_{t\in(0,R\bar{y})} t^{a-1} \geq \frac{1}{\bar{\gamma}}\left[\exp\left(\frac{\mathcal{I}\bar{\gamma}\bar{r}_{\mathcal{I},\gamma}}{p}\right) - 1\right],$$

so that for all $(y,\gamma) \in (0,\bar{y}) \times (0,\bar{\gamma})$,

$$\ln\left(1+\gamma(Ry)^{a-1}\right) \geq \frac{\gamma}{\bar{\gamma}}\ln\left(1+\bar{\gamma}(Ry)^{a-1}\right) \geq \frac{\mathcal{I}\bar{r}_{\mathcal{I},\gamma}\gamma}{p} \geq \frac{\mathcal{I}}{p}\ln(1+\bar{r}_{\mathcal{I},\gamma}\gamma),$$

where, for the first inequality, we used the fact that for $\alpha \in (0,1)$, for all $x \geq 0$, $\ln(1+\alpha x) \geq \alpha\ln(1+x)$ (by concavity of the logarithm) and for the third that $x \geq \ln(1+x)$. This concludes the proof of (4.41).

Now, to prove that a.s. $\limsup_{k\to\infty}\underline{\theta}_{T_k-\mathcal{I}} > 0$, let us introduce the stopping times $(\sigma_m)_{m\geq 0}$ and $(\tau_m)_{m\geq 1}$ such that $\sigma_0 = 0$, and for $m \geq 1$ (with the convention $\inf\emptyset = \infty$),

$$\tau_m = \inf\{k > \sigma_{m-1} : Y_k \leq \bar{y}\}, \qquad \sigma_m = \inf\{k > \tau_m : Y_k > \bar{y}\},$$

where $\bar{y}$ has been introduced in (4.41). On the event $\{Y_k > \bar{y}$ infinitely often$\}$, one has $\limsup_{k\to\infty} Y_k \geq \bar{y} > 0$. Notice that the complementary of the previous event writes $\{Y_k > \bar{y}$ infinitely often$\}^c = \{\exists m \geq 1, \tau_m < \infty = \sigma_m\}$. To prove the result on this event, let us consider, for any fixed $m \geq 1$ and $\ell \geq 1$, the process $(Z_k)_{k\geq \underline{k}\vee\ell}$ defined by

$$\forall k \geq \underline{k} \vee \ell,\ Z_k = -\ln(Y_{k\wedge\sigma_m})\,1_{\tau_m\leq\ell}$$

where $\underline{k}$ has been introduced in (4.41). The process $(Z_k)_{k\geq\underline{k}\vee\ell}$ is a non-negative $\mathcal{F}_{T_k}$-supermartingale by (4.41) and thus converges a.s. to a finite limit as $k \to \infty$. Hence, for any fixed $m \geq 1$, on $\{\tau_m < \infty\} = \cup_{\ell\geq 1}\{\tau_m \leq \ell\}$, the process $(-\ln(Y_{k\wedge\sigma_m}))_{k\geq 1}$ converges a.s. to a finite limit V_m. As a consequence, on $\{\exists m \geq 1 : \tau_m < \infty = \sigma_m\}$, $(Y_k)_{k\geq 1}$ converges a.s. to $\sum_{m\geq 1} 1_{\{\tau_m<\infty=\sigma_m\}}\mathrm{e}^{-V_m}$ which is positive on the event $\{\exists m \geq 1 : \tau_m < \infty = \sigma_m\}$. In conclusion, almost surely, $\limsup_{k\to\infty} Y_k > 0$. This concludes the proof. ∎

4.5 Proof of Proposition 4.2.4

We have checked in the previous section that the WL$_a$ algorithm (which encompasses the SHUS$_a^\alpha$ algorithm, see Section 4.1.2) is recurrent under mild conditions on the stepsize sequence $(\gamma_n)_{n\geq 1}$. In this section, we verify that for any $\alpha \in (\frac{1}{2}, 1]$, these conditions are satisfied for the stepsize sequence generated by the SHUS$_a^\alpha$ algorithm, as well as the usual conditions (4.14) of summability on the sequence $(\gamma_n)_{n\geq 1}$. This is the content of Proposition 4.5.1 below. Proposition 4.2.4 is then deduced from Propositions 4.2.3 and 4.5.1 by conditioning w.r.t. $\mathcal{F}_0$.

Proposition 4.5.1. *Assume A1 and A2. Let $a \in [0, 1)$ and $\alpha \in (\frac{1}{2}, 1]$. The random stepsize sequence $(\gamma_{n+1} = \frac{\gamma}{g_\alpha(S_n)})_{n\geq 0}$ generated by the SHUS$_a^\alpha$ algorithm started from a deterministic initial condition $(\tilde{\theta}_0, X_0) \in (\mathbb{R}_+^*)^{\mathcal{I}} \times E$ is decreasing, bounded from above by some deterministic sequence converging to 0 as $n \to \infty$ and such that $\mathbb{P}(\inf_{n\geq 1} n^\alpha \gamma_n > 0) = 1$. Moreover, $\bar{r}_{\mathcal{I},\gamma} = \sup_{n\geq 1} \frac{\gamma_n}{\gamma_{n+\mathcal{I}-1}} < \infty$ with the explicit upper bounds :*

— *if $\alpha \in (\frac{1}{2}, 1)$,*

$$\bar{r}_{\mathcal{I},\gamma} \leq \left(1 + \frac{(\mathcal{I}-1)\gamma}{\ln(1+S_0)^{\frac{1}{1-\alpha}}}\right)^{\frac{\alpha}{1-\alpha}},$$

— *if $\alpha = 1$,*

$$\bar{r}_{\mathcal{I},\gamma} \leq 1 + \frac{\gamma(\mathcal{I}-1)}{S_0}.$$

Finally, $\mathbb{P}\left(\sup_{n\geq 1} n^\alpha \gamma_n < +\infty\right) = 1$.

The property $\inf_{n\geq 1} n^\alpha \gamma_n > 0$ implies that $\sum_{n\geq 1} \gamma_n = \infty$, while $\sup_{n\geq 1} n^\alpha \gamma_n < +\infty$ implies that $\sum_{n\geq 1} \gamma_n^2 < \infty$ (see the assumptions (4.14) required on the stepsize sequence to prove convergence).

Let us also mention that, when $\alpha = 1$, the proof we give below implies that $\mathbb{P}(\inf_{n\geq 0} \min_{1\leq i\leq \mathcal{I}} \theta_n(i) > 0) = 1$ (using Equations (4.46) and (4.52) below), *i.e.* that the SHUS$_a^1$ algorithm is stable. This gives another way to prove the stability of the method in this specific setting, without following the two-step argument that we used for a general α, namely first proving the recurrence of the algorithm (see Propositions 4.2.3 and 4.2.4), and then using Theorem 3.0.6 (see the proof of Proposition 4.2.2).

Proof: We decompose the proof in several steps.

Deterministic upper bound on $(\gamma_n)_{n\geq 1}$. By (4.7), $(S_n)_{n\geq 0}$ is increasing so that $(\gamma_n)_{n\geq 1}$ is decreasing since g_α is increasing. Using (4.7) again, we have

$$S_{n+1} = S_n + \frac{\gamma}{g_\alpha(S_n)} \theta_n^{a-1}(I(X_{n+1}))\tilde{\theta}_n(I(X_{n+1})) = S_n \left(1 + \frac{\gamma \theta_n^a(I(X_{n+1}))}{g_\alpha(S_n)}\right)$$

and since by (4.3), for all $i \in \{1, \ldots, \mathcal{I}\}$, $(\tilde{\theta}_n(i))_{n\geq 0}$ is non-decreasing and $\inf_{t\in(0,1)} t^{1-a} = 1 = \sup_{t\in(0,1)} t^a$, it holds

$$S_n + \frac{\gamma \min_{1\leq i\leq \mathcal{I}} \tilde{\theta}_0(i)}{g_\alpha(S_n)} \leq S_{n+1} \leq S_n \left(1 + \frac{\gamma}{g_\alpha(S_n)}\right). \tag{4.44}$$

Hence $S_{n+1} \geq \psi(S_n)$, where

$$\forall x > 0,\ \psi(x) = x + \frac{\gamma \min_{1\leq i\leq \mathcal{I}} \tilde{\theta}_0(i)}{g_\alpha(x)}.$$

Let $(\underline{s}_n)_{n\geq 0}$ be defined inductively by $\underline{s}_0 = S_0$ and $\underline{s}_{n+1} = \psi(\underline{s}_n)$ for all $n \geq 0$. This sequence is increasing and goes to ∞ when $n \to \infty$ as $x \mapsto \gamma \min_{1\leq i\leq \mathcal{I}} \tilde{\theta}_0(i)/g_\alpha(x)$ is locally bounded away from 0 on $(0, +\infty)$. Moreover ψ is a convex function which is decreasing on $(0, x_\psi)$ and increasing on $(x_\psi, +\infty)$ for some $x_\psi \in (0, +\infty)$. The minimum of the function g is therefore attained at x_ψ. Since

$$S_1 \geq \underline{s}_1 = \psi(S_0) \geq \psi(x_\psi) = x_\psi + \frac{\gamma \min_{1\leq i\leq \mathcal{I}} \tilde{\theta}_0(i)}{g_\alpha(x)} > x_\psi,$$

it follows that $\min(S_n, \underline{s}_n) > x_\psi$ for any $n \geq 1$, and one easily checks by induction on n that $S_n \geq \underline{s}_n$ for all $n \geq 1$. Hence the sequence $(\gamma_n = \frac{\gamma}{g_\alpha(S_n)})_{n\geq 1}$ is bounded from above by a deterministic sequence converging to 0 as $n \to \infty$.

Lower bound on $(\gamma_n)_{n\geq 0}$**.** When $\alpha \in (\frac{1}{2}, 1)$, to prove the lower bound on $(n^\alpha \gamma_n)_{n\geq 1}$, we remark that for all $n \in \mathbb{N}$, using that $S_n \leq S_{n+1}$ implies that $\frac{1+S_{n+1}}{1+S_n} \leq \frac{S_{n+1}}{S_n}$ for the first inequality, and the inequality $\ln(1+x) \leq x$ and the upper bound in (4.44) for the second one,

$$\begin{aligned} \ln(1+S_{n+1}) &= \ln(1+S_n) + \ln\left(\frac{1+S_{n+1}}{1+S_n}\right) \leq \ln(1+S_n) + \ln\left(\frac{S_{n+1}}{S_n}\right) \\ &\leq \ln(1+S_n) + \frac{\gamma}{(\ln(1+S_n))^{\frac{\alpha}{1-\alpha}}}. \end{aligned} \tag{4.45}$$

Therefore, denoting for simplicity $l_n = (\ln(1+S_n))^{\frac{1}{1-\alpha}}$ and using the monotonicity of the sequence $(l_n)_{n\geq 0}$ and the convexity of $x \mapsto (1+x)^{\frac{1}{1-\alpha}}$ on $\mathbb{R}_+$ for the second inequality, we get

$$\begin{aligned} l_{n+1} &\leq l_n \left(1 + \frac{\gamma}{l_n}\right)^{\frac{1}{1-\alpha}} \\ &\leq l_n \left(1 + \frac{l_0}{l_n}\left(\left(1 + \frac{\gamma}{l_0}\right)^{\frac{1}{1-\alpha}} - 1\right)\right). \end{aligned}$$

This implies by induction on n that

$$l_n \leq l_0 + n l_0 \left(\left(1 + \frac{\gamma}{l_0}\right)^{\frac{1}{1-\alpha}} - 1\right)$$

Raising this inequality to the power $-\alpha$ ensures that $\mathbb{P}\left(\inf_{n\geq 1} n^\alpha \gamma_n > 0\right) = 1$.

When $\alpha = 1$, since $g_1(s) = s$, the upper bound in (4.44) writes $S_{n+1} \leq S_n + \gamma$ so that, by induction on n,

$$S_n \leq S_0 + n\gamma. \tag{4.46}$$

Therefore

$$\gamma_{n+1} = \frac{\gamma}{S_n} \geq \frac{\gamma}{S_0 + n\gamma} = \frac{\gamma_1}{1 + n\gamma_1}.$$

Upper bounds on $\bar{r}_{\mathcal{I},\gamma}$**.** Let us now derive the upper bounds on $\bar{r}_{\mathcal{I},\gamma}$. When $\alpha = 1$, by (4.46), $S_{n+\mathcal{I}-1} \leq S_n + \gamma(\mathcal{I} - 1)$, which yields

$$\frac{\gamma_{n+1}}{\gamma_{n+\mathcal{I}}} = \frac{S_{n+\mathcal{I}-1}}{S_n} \leq 1 + \frac{\gamma(\mathcal{I}-1)}{S_n} \leq 1 + \frac{\gamma(\mathcal{I}-1)}{S_0}.$$

When $\alpha \in (\frac{1}{2}, 1)$, by (4.44), the inequality $\ln(1+x) \leq x$ on $\mathbb{R}_+$ and the monotonicity of $(S_n)_{n\geq 0}$, we have for any $0 \leq q \leq n$,

$$\ln(1+S_{n+1}) \leq \ln(1+S_n) + \ln\left(1 + \frac{\gamma S_n}{(1+S_n)\ln(1+S_n)^{\frac{\alpha}{1-\alpha}}}\right) \leq \ln(1+S_n) + \frac{\gamma}{\ln(1+S_0)^{\frac{\alpha}{1-\alpha}}}$$
$$\leq \ln(1+S_n) + \ln(1+S_q)\frac{\gamma}{\ln(1+S_0)^{\frac{1}{1-\alpha}}}.$$

Therefore,

$$\frac{\gamma_{n+1}}{\gamma_{n+\mathcal{I}}} = \frac{\ln(1+S_{n+\mathcal{I}-1})^{\frac{\alpha}{1-\alpha}}}{\ln(1+S_n)^{\frac{\alpha}{1-\alpha}}} \leq \left(1 + \frac{\gamma(\mathcal{I}-1)}{\ln(1+S_0)^{\frac{1}{1-\alpha}}}\right)^{\frac{\alpha}{1-\alpha}}.$$

Upper bounds on $(\gamma_n)_{n\geq 1}$: the case $\alpha \in (\frac{1}{2}, 1)$. To deal with the last assertion, we are going to derive lower bounds on

$$\underline{\tilde{\theta}}_n = \min_{1\leq i\leq \mathcal{I}} \tilde{\theta}_n(i).$$

By (4.3), for all $n \geq 0$ and for all $i \in \{1, \ldots, \mathcal{I}\}$,

$$\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i)\left(1 + \gamma_{n+1}\,\theta_n^{a-1}(i)\,1_{E_i}(X_{n+1})\right). \tag{4.47}$$

We use the sequence $(T_k)_{k\in\mathbb{N}}$ introduced in Section 4.4 for $R > 1$. Since $\inf_{t\in(0,1)} t^{a-1} = 1$, we deduce that for all $k \geq 0$ and for all $i \in \{1, \ldots, \mathcal{I}\}$,

$$\tilde{\theta}_{T_{k+1}}(i) \geq \tilde{\theta}_{T_{k+1}-\mathcal{I}}(i) \prod_{n=T_{k+1}-\mathcal{I}+1}^{T_{k+1}} (1 + \gamma_{n+1}1_{E_i}(X_n)).$$

Now, if i is the index of a stratum with large weight, namely $\tilde{\theta}_{T_{k+1}-\mathcal{I}}(i) \geq R\underline{\tilde{\theta}}_{T_{k+1}-\mathcal{I}}$, one simply uses the lower bound :

$$\tilde{\theta}_{T_{k+1}}(i) \geq R\underline{\tilde{\theta}}_{T_{k+1}-\mathcal{I}}. \tag{4.48}$$

If i is the index of a stratum with small weight, namely $\tilde{\theta}_{T_{k+1}-\mathcal{I}}(i) < R\underline{\tilde{\theta}}_{T_{k+1}-\mathcal{I}}$, by definition of the sequence $(T_k)_{k\in\mathbb{N}}$, this stratum is visited at least once between $T_{k+1} - \mathcal{I}$ and T_{k+1}, and thus, using the monotonicity of the sequence $(\gamma_n)_{n\geq 1}$, we get

$$\tilde{\theta}_{T_{k+1}}(i) \geq \left(1 + \gamma_{T_{k+1}}\right)\underline{\tilde{\theta}}_{T_{k+1}-\mathcal{I}}. \tag{4.49}$$

By combining (4.48) and (4.49), one thus obtains

$$\underline{\tilde{\theta}}_{T_{k+1}} \geq \left(R \wedge \left(1 + \gamma_{T_{k+1}}\right)\right)\underline{\tilde{\theta}}_{T_{k+1}-\mathcal{I}}.$$

Since $T_k \leq T_{k+1} - \mathcal{I}$, by the monotonicity of the sequence $(\underline{\tilde{\theta}}_n)_{n\geq 0}$, one concludes that

$$\forall k \in \mathbb{N},\ \underline{\tilde{\theta}}_{T_{k+1}} \geq \left(R \wedge \left(1 + \gamma_{T_{k+1}}\right)\right)\underline{\tilde{\theta}}_{T_k} \geq (1 + c\gamma_{T_{k+1}})\underline{\tilde{\theta}}_{T_k} \tag{4.50}$$

for $c = 1 \wedge ((R-1)/\gamma_1))$. Hence, using the concavity of the logarithm for the second inequality then (4.37) and the fact that, by the first step of the proof, $\mathbb{P}(\inf_n n^\alpha \gamma_n > 0) = 1$

for the third one, one gets that for the positive random variable $\tilde{c} = c\frac{\ln(1+\gamma_1)}{\gamma_1 C_\star^\alpha} \inf_{n\geq 1} n^\alpha \gamma_n$ and $k \geq 1$,

$$\begin{aligned}\ln \underline{\tilde{\theta}}_{T_k} &\geq \ln \underline{\tilde{\theta}}_0 + c\sum_{j=1}^k \ln(1+\gamma_{T_j}) \geq \ln \underline{\tilde{\theta}}_0 + c\frac{\ln(1+\gamma_1)}{\gamma_1}\sum_{j=1}^k \gamma_{T_j} \\ &\geq \ln \underline{\tilde{\theta}}_0 + \tilde{c}\sum_{j=1}^k j^{-\alpha} \geq \ln \underline{\tilde{\theta}}_0 + \frac{\tilde{c}}{1-\alpha}\left((k+1)^{1-\alpha}-1\right).\end{aligned}$$

With the monotonicity of $(\underline{\tilde{\theta}}_n)_n$ and (4.37), one deduces that a.s., for all $n \in \mathbb{N}$,

$$\underline{\tilde{\theta}}_n \geq \underline{\tilde{\theta}}_{T_{\lfloor n/C_\star \rfloor}} \geq \underline{\tilde{\theta}}_0 \exp\left(\frac{\tilde{c}}{1-\alpha}\left(\frac{n^{1-\alpha}}{C_\star^{1-\alpha}}-1\right)\right) \geq \underline{\tilde{\theta}}_0 \exp\left(\frac{\tilde{c}}{1-\alpha}\left(\frac{(n+1)^{1-\alpha}-1}{C_\star^{1-\alpha}}-1\right)\right).$$

Since

$$S_{n+1} \geq S_n + \gamma_{n+1}\underline{\tilde{\theta}}_n \tag{4.51}$$

setting $c_0 = \underline{\tilde{\theta}}_0 \left(\inf_{n\geq 1} n^\alpha \gamma_n\right) \exp\left(-\frac{\tilde{c}(1+C_\star^{1-\alpha})}{(1-\alpha)C_\star^{1-\alpha}}\right)$ and $c_1 = \frac{\tilde{c}}{(1-\alpha)C_\star^{1-\alpha}}$ one gets

$$\forall n \in \mathbb{N},\ S_{n+1} \geq S_n + \frac{c_0}{(n+1)^\alpha} \mathrm{e}^{c_1(n+1)^{1-\alpha}}.$$

Since $x \mapsto x^{-\alpha}\mathrm{e}^{c_1 x^{1-\alpha}}$ is increasing for $x \geq \left(\frac{\alpha}{c_1(1-\alpha)}\right)^{\frac{1}{1-\alpha}}$, one deduces that for all $n \geq n_1 := \left\lceil \left(\frac{\alpha}{c_1(1-\alpha)}\right)^{\frac{1}{1-\alpha}} \right\rceil$,

$$S_n \geq S_{n_1} + \int_{n_1}^n c_0 x^{-\alpha}\mathrm{e}^{c_1 x^{1-\alpha}} dx = S_{n_1} + \frac{c_0}{c_1(1-\alpha)}\left(\mathrm{e}^{c_1 n^{1-\alpha}} - \mathrm{e}^{c_1 n_1^{1-\alpha}}\right)$$

so that $\limsup_{n\to\infty}(n+1)^{\alpha-1}\ln(1+S_n) \geq c_1 > 0$. This concludes the proof since $(n+1)^{\alpha-1}\ln(1+S_n) > 0$ for all $n \geq 0$.

Upper bounds on $(\gamma_n)_{n\geq 1}$: the case $\alpha = 1$. The objective is to show that a.s. $\sup_{n\geq 1} n\gamma_n < \infty$. Since $S_n = \sum_{i=1}^{\mathcal{I}} \tilde{\theta}_n(i) \geq \underline{\tilde{\theta}}_n$, it is sufficient to prove that

$$\mathbb{P}\left(\inf_{n\geq 1} n^{-1}\underline{\tilde{\theta}}_n > 0\right) = 1. \tag{4.52}$$

Writing $\tilde{\theta}_n(i)\gamma_{n+1}\theta_n^{a-1}(i) = \gamma\tilde{\theta}_n(i)^a S_n^{-a}$ in (4.47) and using the monotonicity of the sequences $(\tilde{\theta}_n(i))_{n\geq 0}$ and $(S_n)_{n\geq 0}$, one deduces that (4.49) and (4.50) may respectively be replaced by :

— For all $k \in \mathbb{N}$, if i is the index of a stratum with small weight, namely $\tilde{\theta}_{T_{k+1}-d}(i) < R\underline{\tilde{\theta}}_{T_{k+1}-\mathcal{I}}$,

$$\tilde{\theta}_{T_{k+1}}(i) \geq \tilde{\theta}_{T_{k+1}-\mathcal{I}}(i) + \gamma\tilde{\theta}^a_{T_{k+1}-\mathcal{I}}(i) S^{-a}_{T_{k+1}-1}. \tag{4.53}$$

— By combining (4.48) and (4.53), one thus obtains

$$\underline{\tilde{\theta}}_{T_{k+1}} \geq \left(R\underline{\tilde{\theta}}_{T_k}\right) \wedge \left(\underline{\tilde{\theta}}_{T_k} + \gamma\underline{\tilde{\theta}}^a_{T_k} S^{-a}_{T_{k+1}-1}\right) \tag{4.54}$$

Therefore $k \mapsto \underline{\tilde{\theta}}_{T_k}$ grows at least geometrically with ratio R as long as the decreasing sequence $\left(1+\gamma\left(\underline{\tilde{\theta}}_{T_k}\right)^{a-1} S_{T_{k+1}-1}^{-a}\right)_{k\geq 1}$ is larger than R and there exists a random variable $K \in \mathbb{N}\setminus\{0\}$ such that a.s.

$$\forall k \geq K,\ \underline{\tilde{\theta}}_{T_{k+1}} \geq \underline{\tilde{\theta}}_{T_k} + \gamma \underline{\tilde{\theta}}_{T_k}^a S_{T_{k+1}-1}^{-a}.$$

Since by (4.46) and (4.37), there exists a positive random variable C such that a.s. $\forall k,\ \gamma S_{T_{k+1}-1}^{-a} \geq Ck^{-a}$ we deduce that a.s.,

$$\forall k \geq K,\ \underline{\tilde{\theta}}_{T_{k+1}}^{1-a} \geq \underline{\tilde{\theta}}_{T_k}^{1-a}\left(1+C\underline{\tilde{\theta}}_{T_k}^{a-1}k^{-a}\right)^{1-a}.$$

With the monotonicity of $(\underline{\tilde{\theta}}_n)_{n\geq 0}$ and the inequality $(1+x)^{1-a} \geq 1+(1-a)(1+x_0)^{-a}x$ valid for $0\leq x\leq x_0$, we deduce that a.s.

$$\forall k \geq K,\ \underline{\tilde{\theta}}_{T_{k+1}}^{1-a} \geq \underline{\tilde{\theta}}_{T_k}^{1-a} + (1-a)\left(1+C\underline{\tilde{\theta}}_0^{a-1}\right)^{-a} Ck^{-a}.$$

Summing this inequality and remarking that for $k > K$, $\sum_{\ell=K}^{k-1}\ell^{-a} \geq \int_K^k \frac{dx}{x^a} = \frac{k^{1-a}-K^{1-a}}{1-a}$, we deduce that $\mathbb{P}(\inf_{k\geq 1} k^{a-1}\underline{\tilde{\theta}}_{T_k}^{1-a} > 0) = 1$. Since the inequality $\forall n,\ T_{\lfloor n/C_\star\rfloor} \leq n$ deduced from (4.37) and the monotonicity of $(\underline{\tilde{\theta}}_n)_{n\geq 0}$ imply that $\underline{\tilde{\theta}}_n \geq \underline{\tilde{\theta}}_{T_{\lfloor n/C_\star\rfloor}}$, we conclude that $\mathbb{P}\left(\inf_{n\geq 1} n^{-1}\underline{\tilde{\theta}}_n > 0\right) = 1$.

∎

4.6 Proof of Corollary 4.2.5

Case $\alpha = 1$. By (4.7), for $n\in\mathbb{N}$,

$$\frac{S_n}{n} = \frac{S_0}{n} + \frac{\gamma}{n}\sum_{k=1}^n \theta_{k-1}^a(I(X_k)).$$

Therefore, to conclude the proof it is enough to check that a.s.,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \theta_{k-1}^a(I(X_k)) = \left(\sum_{j=1}^{\mathcal{I}}\theta_\star^{1-a}(j)\right)^{-1}. \tag{4.55}$$

This follows from the decomposition

$$\begin{aligned}\left(\sum_{j=1}^{\mathcal{I}}\theta_\star^{1-a}(j)\right)\frac{1}{n}\sum_{k=1}^n \theta_{k-1}^a(I(X_k)) &= \frac{1}{n}\sum_{k=1}^n\left(\sum_{j=1}^{\mathcal{I}}\theta_{k-1}^{1-a}(j)\right)\theta_{k-1}^a(I(X_k)) \\ &+ \frac{1}{n}\sum_{k=1}^n\left(\sum_{j=1}^{\mathcal{I}}\theta_\star^{1-a}(j) - \sum_{j=1}^{\mathcal{I}}\theta_{k-1}^{1-a}(j)\right)\theta_{k-1}^a(I(X_k)).\end{aligned}$$

where the first term in the right-hand side almost surely converges to 1 by item (iii) in Theorem 4.2.1 (choosing $f \equiv 1$). The absolute value of the second term in the right-hand side is bounded from above by

$$\frac{1}{n}\sum_{k=1}^{n}\left|\sum_{j=1}^{\mathcal{I}}\{\theta_\star^{1-a}(j)-\theta_{k-1}^{1-a}(j)\}\right|$$

which almost surely goes to zero thanks to item (i) in Theorem 4.2.1 and Cesàro Lemma.

Case $\alpha \in (\frac{1}{2}, 1)$. We remark that (4.7) implies that

$$\ln(1+S_{n+1}) = \ln(1+S_n) + \frac{\gamma\theta_n^a(I(X_{n+1}))}{(\ln(1+S_n))^{\frac{\alpha}{1-\alpha}}} + R_{n+1}^{(1)},$$

where, setting $h(x) = \ln(1+x) - x$,

$$R_{n+1}^{(1)} = h\left(\frac{\gamma S_n \theta_n^a(I(X_{n+1}))}{(1+S_n)(\ln(1+S_n))^{\frac{\alpha}{1-\alpha}}}\right) - \frac{\gamma\theta_n^a(I(X_{n+1}))}{(1+S_n)(\ln(1+S_n))^{\frac{\alpha}{1-\alpha}}}.$$

From now on, C denotes a positive random variable which may change from line to line. Since $0 \ge h(x) \ge -x^2/2$ for all $x > 0$ and

$$\sup_{x>0}\frac{(\ln(1+x))^{\frac{\alpha}{1-\alpha}}}{1+x} < +\infty,$$

it follows that

$$\forall n \ge 0, \ \left|R_{n+1}^{(1)}\right| \le c(\ln(1+S_n))^{-\frac{2\alpha}{1-\alpha}}$$

for some deterministic constant $c \in (0, +\infty)$. Writing

$$(\ln(1+S_{n+1}))^{\frac{1}{1-\alpha}} = (\ln(1+S_n))^{\frac{1}{1-\alpha}} \times \left(1 + \frac{\gamma\theta_n^a(I(X_{n+1}))}{(\ln(1+S_n))^{\frac{1}{1-\alpha}}} + \frac{R_{n+1}^{(1)}}{\ln(1+S_n)}\right)^{\frac{1}{1-\alpha}}$$

and remarking that $x \mapsto \frac{1}{x^2}\left|(1+x)^{\frac{1}{1-\alpha}} - 1 - \frac{x}{1-\alpha}\right|$ is locally bounded on $\mathbb{R}_+$, one deduces that

$$(\ln(1+S_{n+1}))^{\frac{1}{1-\alpha}} = (\ln(1+S_n))^{\frac{1}{1-\alpha}} + \frac{\gamma\theta_n^a(I(X_{n+1}))}{1-\alpha} + R_{n+1}^{(2)},$$

where, for all $n \ge 0$, $\left|R_{n+1}^{(2)}\right| \le c'(\ln(1+S_n))^{-\frac{\alpha}{1-\alpha}}$ for some constant $c' \in (0, +\infty)$ depending on S_0. One has

$$\frac{1}{n}\Big(\ln(1+S_n)\Big)^{\frac{1}{1-\alpha}} = \frac{1}{n}\Big(\ln(1+S_0)\Big)^{\frac{1}{1-\alpha}} + \frac{1}{n}\sum_{k=1}^{n}R_k^{(2)} + \frac{\gamma}{(1-\alpha)n}\sum_{k=1}^{n}\theta_{k-1}^a(I(X_k)).$$

Since $\gamma_{n+1} = \frac{\gamma}{g_\alpha(S_n)} = \frac{\gamma}{\ln(1+S_n)^{\frac{\alpha}{1-\alpha}}}$, the last statement in Proposition 4.5.1 ensures the existence of a positive random variable C such that $\forall n \ge 0$, $|R_{n+1}^{(2)}| \le Cn^{-\alpha}$, so that by Cesaro lemma, the second term in the right-hand side almost surely converge to zero as $n \to \infty$. So does the first term. Using (4.55), one thus obtains, almost surely,

$$\lim_{n\to\infty}\frac{1}{n}(\ln(1+S_n))^{\frac{1}{1-\alpha}} = \frac{\gamma}{1-\alpha}\left(\sum_{i=1}^{\mathcal{I}}\theta_\star^{1-a}(i)\right)^{-1}$$

which concludes the proof.

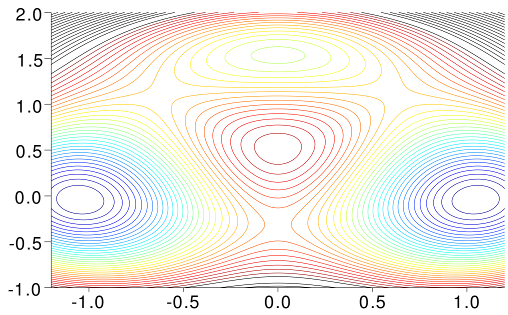

FIGURE 4.1 – Level sets of the potential V considered for the simulations. The minima are located at the positions $x_\pm \simeq (\pm 1.05, -0.04)$, which defines two metastable states, one on the left in the vicinity of x_-, and one on the right in the vicinity of x_+.

4.7 Numerical illustration

The results stated in Section 4.2 precisely describe the asymptotic behavior of the SHUS_a^α algorithm but do not give much information about the efficiency of the adaptive algorithm compared to the original non adaptive one. The aim of this section is to explore on a specific numerical example the interest of using the SHUS_a^α algorithm in terms of computational efficiency.

The state space is $E = [-R, R] \times \mathbb{R}$. The density of the target measure reads

$$\forall x = (x_1, x_2) \in E,\ \eta(x) = Z^{-1} 1_{[-R,R]}(x_1)\, \mathrm{e}^{-\beta V(x_1,x_2)},$$

for some positive inverse temperature β and $Z = \int_E \mathrm{e}^{-\beta V(x_1,x_2)}\, dx_1 dx_2$, with

$$V(x_1,x_2) = 3\exp\left(-x_1^2 - \left(x_2 - \frac{1}{3}\right)^2\right) - 3\exp\left(-x_1^2 - \left(x_2 - \frac{5}{3}\right)^2\right) \tag{4.56}$$
$$- 5\exp\left(-(x_1-1)^2 - x_2^2\right) - 5\exp\left(-(x_1+1)^2 - x_2^2\right) + 0.2x_1^4 + 0.2\left(x_2 - \frac{1}{3}\right)^4.$$

A plot of the level sets of the potential V is presented in Figure 4.1. The global minima of the potential are located at the points $x_- \simeq (-1.05, -0.04)$ and $x_+ \simeq (+1.05, -0.04)$. This induces two metastable states, located in the vicinities of each of the global minima.

We introduce $\mathcal{I}$ strata $(E_\ell = (a_\ell, a_{\ell+1}) \times \mathbb{R})_{\ell=1,\dots,\mathcal{I}}$, where $a_\ell = -R + 2(\ell-1)R/\mathcal{I}$ for $\ell = 1, \dots, \mathcal{I}+1$. In the following, we set $R = 1.2$ and $\mathcal{I} = 24$. The initial weight vector is $\tilde{\theta}_0 = (1/\mathcal{I}, \dots, 1/\mathcal{I})$ and the initial configuration is $X_0 = (-1, 0)$. For all θ, the kernel P_θ^a is defined by a Metropolis-Hastings step with target density η_θ^a and a two-dimensional Gaussian proposal distribution $q(x,y) = \frac{1}{2\eta\sigma^2}\exp(-|x-y|^2/\sigma^2)$. We choose in the following $\sigma^2 = 0.01$ (so that $\sigma = 2R/\mathcal{I}$) and we use the Mersenne-Twister random number generator as implemented in the GSL library.

The Metropolis-Hastings dynamics without adaptation (namely with the Gaussian proposal distribution and target η at each iteration) is metastable : it takes a long time (which becomes exponentially large in the limit $\beta \to \infty$) to go from the stratum containing x_- to the stratum containing x_+.

The SHUS_a^α algorithm is applied with $a \in [0,1]$. Moreover, we choose in all numerical simulations the parameter γ (which appears in the recurrence relation (4.3)) as a function of α as follows :

$$\gamma(\alpha) = \begin{cases} 1 \text{ if } \alpha = 1, \\ (1-\alpha)^{-\frac{\alpha}{1-\alpha}} \text{ if } \alpha \in (1/2, 1). \end{cases} \tag{4.57}$$

This is to avoid the degeneracy of the constant $\mathcal{C}_\alpha(\gamma)$ which appears in the asymptotic behavior of the stepsize sequence (see Corollary 4.2.5) when $\alpha \to 1$. Indeed, when γ does not depend on α, $\lim_{\alpha\to 1} \mathcal{C}_\alpha(\gamma) = 0$. With the choice (4.57), $\mathcal{C}_\alpha(\gamma(\alpha))$ no longer depends on α :

$$\forall \alpha \in (1/2, 1),\ \mathcal{C}_\alpha(\gamma(\alpha)) = \mathcal{C}_1(\gamma(1)) = 1.$$

4.7.1 Asymptotic behavior of the stepsize sequence

With the choice (4.57) of the parameter γ, it is expected from Corollary 4.2.5 that, in the large n limit and for $\alpha \in (1/2, 1)$,

$$\gamma_n \sim \left(\frac{\mathfrak{g}(a)}{n}\right)^\alpha \quad \text{with } \mathfrak{g}(a) = Z_{\theta_\star}^a = \sum_{i=1}^{\mathcal{I}} \theta_\star(i)^{1-a},$$

while, for $\alpha = 1$,

$$\mu\gamma_n \sim \frac{\mathfrak{g}(a)}{n}.$$

This is checked numerically on Figures 4.2 for various values of α. The reference values $\theta_\star(i)$ are obtained by a two-dimensional quadrature. The numerical results are in excellent agreement with the theoretical findings, and allow to go even beyond the theoretical predictions since the scaling of the steps also hold for values of $\alpha < 1/2$. Note however that the convergence is slower for smaller values of a and α. In particular, although not apparent on the scale of the plots, the values of $n\gamma_n^{1/\alpha}$ are very slowly increasing for $a = 0.2$ and $\alpha = 0.4$.

4.7.2 Exit times

In order to show the interest of using the SHUS_a^α algorithm to get out of metastable states, we now study the exit time from the left metastable state in the small temperature regime $\beta \to \infty$. More precisely, starting from the initial condition $X_0 = (-1, 0)$ close to the global minimum x_-, we consider the time it takes to the system to go to the vicinity of the global minimum x_+.

This section is organized as follows. In Section 4.7.2, exit times from metastable states for the SHUS_a^α algorithm are studied using a toy model with only three states. Following results in [23], this gives heuristic scalings for the exit times in the small temperature regime on the two-dimensional model presented above. We then compare in Section 4.7.2 these expected asymptotic behaviors with those numerically observed on the two-dimensional potential.

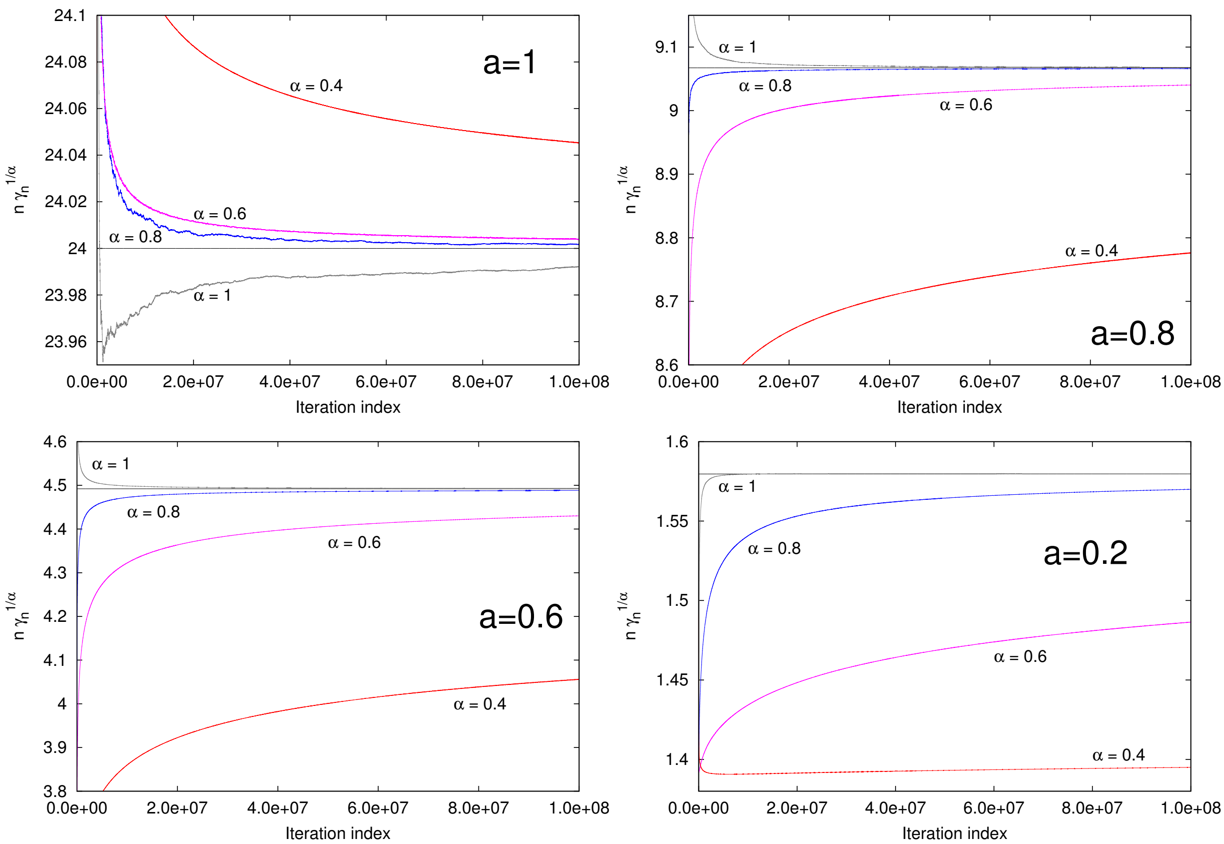

FIGURE 4.2 – Behavior of the effective time step as a function of the iteration index, at $\beta = 4$ and averaged over 5000 realizations. Each plot represents $n\gamma_n^{1/\alpha}$ as a function of the iteration index n for a different value of a, the various curves representing the results obtained for $\alpha \in \{0.4, 0.6, 0.8, 1\}$. Note that, in all cases, the convergence is faster for larger values of α. The horizontal line represents the limiting value $\mathfrak{g}(a)$ (namely $\mathfrak{g}(1) = 24$, $\mathfrak{g}(0.8) = 9.07$, $\mathfrak{g}(0.6) = 4.49$ and $\mathfrak{g}(0.2) = 1.58$).

Heuristic on a toy model

Following [23], one can derive a scaling of the time needed to leave a metastable state on a toy model with only three states.

Let us recall the model we consider in [23]. The toy model consists of three states and three strata, so that $E = \{1, 2, 3\}$ and $E_i = \{i\}$ for $i \in \{1, 2, 3\}$. The target probability is defined, for a small positive parameter $\varepsilon \in (0, 1)$, by

$$\eta(\{1\}) = \eta(\{3\}) = \frac{1}{2+\varepsilon} \text{ and } \eta(\{2\}) = \frac{\varepsilon}{2+\varepsilon}. \tag{4.58}$$

Notice that in this setting, for all $i \in \{1, 2, 3\}$, $\eta(\{i\}) = \theta_\star(i)$. The unbiased dynamics is a Metropolis-Hastings algorithm with target η and proposal matrix $Q \in \mathbb{R}^{3\times 3}$ defined by

$$Q = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix},$$

so that jumps are only allowed between neighboring states. Since the state 2 has small probability, there are two metastable states for this dynamics : state 1 and state 3. One can check that the number of iterations $n^{\mathrm{MH}}(\varepsilon)$ needed by the unbiased dynamics to go from state 1 to state 3 is of order

$$n^{\mathrm{MH}}_{1\to 3}(\varepsilon) \simeq \frac{6}{\varepsilon} \tag{4.59}$$

in the limit $\varepsilon \to 0$ (see [23, Proposition 3.1] for precise statements). One important feature of adaptive dynamics is that this exit time is drastically reduced (compare (4.59) and (4.61)-(4.62) below).

Let us now consider the SHUS$^\alpha_\rho$ algorithm applied to this toy model. Set $\tilde\theta_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $X_0 = 1$. The main output of the paper [23] is that the number of iterations needed to go from 1 to 3 using the adaptive algorithm can be estimated in the limit $\varepsilon \to 0$ by considering the minimum number of iterations required to reach the metastable state 2. This is what we estimate in the following, using some formal derivations, which could be made rigorous using the same techniques as in [23]. For the sake of conciseness, we do not give here rigorous proofs of these results. We will however check the consistency of these results with what is observed numerically in the next section.

As long as $(X_n)_{n\geq 0}$ stays in state 1, $\tilde\theta_n(2) = \tilde\theta_0(2) = \frac{1}{3}$, $\tilde\theta_n(3) = \tilde\theta_0(3) = \frac{1}{3}$ and $\tilde\theta_n(1) = u_n$ where $(u_n)_{n\geq 0}$ is a sequence satisfying, in view of (4.3) :

$$u_{n+1} = u_n + \gamma \frac{u_n + \frac{2}{3}}{g_\alpha\left(u_n + \frac{2}{3}\right)} \left(\frac{3u_n}{3u_n + 2}\right)^a, \tag{4.60}$$

with initial condition $u_0 = \frac{1}{3}$. Let us denote by

$$\zeta_n := \frac{(u_n, \frac{1}{3}, \frac{1}{3})}{u_n + \frac{2}{3}}$$

the associated normalized vector in Θ. Since $P^a_\theta(1,2) = \frac{1}{3}\left(\frac{\varepsilon\theta^a(1)}{\theta^a(2)} \wedge 1\right) = \frac{1}{3}\left(\frac{\varepsilon\tilde\theta^a(1)}{\tilde\theta^a(2)} \wedge 1\right)$, it follows that $P^a_{\zeta_n}(1,2) = \frac{1}{3}\left(\varepsilon(3u_n)^a \wedge 1\right)$. Thus, the probability of staying in state 1 from iteration 0 up to iteration $n+1$ is

$$\prod_{k=0}^{n}\left(1 - P^a_{\zeta_k}(1,2)\right) = \exp\left(\sum_{k=0}^{n} \ln\left(1 - \frac{1}{3}(\varepsilon(3u_k)^a \wedge 1)\right)\right).$$

For ε small, expanding the logarithm, this probability is of order $1/2$ when $\sum_{k=0}^{n} u_k^a$ is of order $1/\varepsilon$.

The case $\alpha = 1$. When $\alpha = 1$, in view of (4.60),

$$u_{n+1} = u_n + \gamma \left(\frac{3u_n}{3u_n + 2} \right)^a .$$

When n is large, u_n becomes large and $u_{n+1} \simeq u_n + \gamma$ so that $u_n \simeq \gamma n$ and then $\sum_{k=0}^{n} u_k^a \simeq \frac{\gamma^a}{1+a} n^{1+a}$. The probability to reach state 2 for the first time after at least $n+1$ iterations attains $1/2$ when this sum becomes of order $1/\varepsilon$ i.e. when n is of order $\gamma^{-\frac{a}{1+a}}(1+a)^{\frac{1}{1+a}}\varepsilon^{-\frac{1}{1+a}}$. We thus obtain that the number of iterations $n^1_{1\to3}$ needed to go from 1 to 3 in the case $\alpha = 1$ satisfies, in the limit $\varepsilon \to 0$,

$$n^1_{1\to3}(\varepsilon) \simeq C^1_{1\to3}\, \varepsilon^{-\frac{1}{1+a}}, \tag{4.61}$$

where $C^1_{1\to3} := \gamma^{-\frac{a}{1+a}}(1+a)^{\frac{1}{1+a}}$ is a constant.

The case $\alpha \in (\frac{1}{2}, 1)$. When $\alpha \in (\frac{1}{2}, 1)$, in view of (4.60),

$$u_{n+1} = u_n + \gamma \frac{u_n + 2/3}{\ln^{\frac{\alpha}{1-\alpha}}(1 + u_n + 2/3)} \left(\frac{3u_n}{3u_n + 2} \right)^a .$$

When n is large, u_n becomes large and

$$u_{n+1} \simeq u_n \left(1 + \frac{\gamma}{\ln^{\frac{\alpha}{1-\alpha}}(u_n)} \right),$$

so that

$$\ln(u_{n+1}) \simeq \ln(u_n) + \frac{\gamma}{\ln^{\frac{\alpha}{1-\alpha}}(u_n)}.$$

By comparison with the ordinary differential equation $\frac{dy}{dt}(t) = \gamma(y(t))^{-\frac{\alpha}{1-\alpha}}$ with solution $y(t) = \left(y(0)^{1/(1-\alpha)} + \frac{\gamma t}{1-\alpha}\right)^{1-\alpha}$, one obtains $u_n \simeq \exp\left[\left(\frac{\gamma n}{1-\alpha}\right)^{1-\alpha}\right]$. We now want to estimate $\sum_{k=0}^{n} u_k^a$. Since for $c = a\left(\frac{\gamma}{1-\alpha}\right)^{1-\alpha}$,

$$\int_0^x \mathrm{e}^{cy^{1-\alpha}} dy = \frac{x^\alpha \mathrm{e}^{cx^{1-\alpha}}}{c(1-\alpha)} - \frac{\alpha}{c(1-\alpha)} \int_0^x y^{\alpha-1} \mathrm{e}^{cy^{1-\alpha}} dy,$$

$\int_0^x \mathrm{e}^{cy^{1-\alpha}} dy \sim \frac{x^\alpha}{c(1-\alpha)} \mathrm{e}^{cx^{1-\alpha}}$ as $x \to +\infty$. Hence $\sum_{k=0}^{n} u_k^a \sim \frac{n^\alpha \mathrm{e}^{cn^{1-\alpha}}}{c(1-\alpha)}$. The probability to reach state 2 for the first time after at least $n+1$ iterations attains $1/2$ when this sum becomes of order $1/\varepsilon$ i.e. when n is of order $\frac{1-\alpha}{\gamma}\left(-\frac{\ln \varepsilon}{a}\right)^{1/(1-\alpha)}$. We thus obtain that the number of iterations $n^\alpha_{1\to3}$ needed to go from 1 to 3 in the case $\alpha \in (1/2, 1)$ satisfies, in the limit $\varepsilon \to 0$,

$$n^\alpha_{1\to3}(\varepsilon) \simeq C^\alpha_{1\to3}\, |\ln \varepsilon|^{1/(1-\alpha)} . \tag{4.62}$$

where $C^\alpha_{1\to3} := \frac{1-\alpha}{\gamma} a^{1/(\alpha-1)}$.

By comparing the transition time from 1 to 3 on the unbiased Metropolis-Hastings dynamics (see (4.59)), and for the SHUS$^\alpha_a$ algorithm (see (4.61) and (4.62)), the interest of using the adaptively biased dynamics is obvious : the exit times are much smaller for SHUS. Moreover, one can see that it is interesting to consider $\alpha < 1$ in order to reduce the exit time compared to $\alpha = 1$.

Numerical results on exit times

Average exit times for the two-dimensional model described at the beginning of Section 4.7 are obtained by performing independent realizations of the following procedure, for given values of a, α, β : initialize the system in the state $X_0 = (-1, 0)$, and run the dynamics until the first time index $\mathcal{N}$ such that the first component of $X_{\mathcal{N}}$ is larger than 1. We perform K independent realizations of this process. The corresponding empirical average first exit time is denoted by t_β^α. Since we work with a fixed maximal computational time (of about a week or two on our computing machines with our implementation of the code), K turns out to be of the order of a few hundreds for the largest exit times, while $K = 10^5$ in the easiest cases corresponding to the shortest exit times. In our numerical results, we checked that K is always sufficiently large so that the relative error on t_β^α is less than a few percents in the worst cases.

The first task is to identify the equivalent of the parameter ε in the toy model from Section 4.7.2. In the large β regime, using Laplace's method, the ratio between the probability of the stratum in the transition region (around the vertical axis $(0, y)$ for $y \in \mathbb{R}$) and the metastable states is of order $\bar{C} \exp(-\beta\delta_0)$ for some positive constants $\bar{C}$ and δ_0. In view of (4.58), this suggests the following formal equivalence between ε and β :

$$\varepsilon(\beta) = \bar{C}\mathrm{e}^{-\beta\delta_0}. \tag{4.63}$$

We next replace ε by the right-hand side of the above equality in the scalings found at the end of Section 4.7.2.

When $\alpha = 1$, it is expected from (4.63) and (4.61) that, in the regime $\beta \to \infty$,

$$t_\beta^1 \simeq \widetilde{C}\mathrm{e}^{\beta\delta_0 \frac{\mu}{\mu+a}} \tag{4.64}$$

for some constant $\widetilde{C}$. When $\alpha \in (1/2, 1)$, it is expected from (4.63) and (4.62) that, in the regime $\beta \to \infty$,

$$\ln(t_\beta^\alpha) \simeq \frac{1}{1-\alpha} \ln\beta. \tag{4.65}$$

We check in the sequel the scalings (4.64)-(4.65) by varying the parameters of the dynamics in several ways : (i) fix α (as well as μ for $\alpha = 1$) and vary a ; (ii) fix $\alpha = 1$ and a, and vary μ ; (iii) fix $\alpha = 1$ and vary $a = 1 - \mu$, which corresponds to the well-tempered metadynamics.

Dependence on a. The dependence of the exit times on a is studied in Figure 4.3 for $\alpha = 1$, and in Figure 4.4 for $\alpha \in (\frac{1}{2}, 1)$.

For $\alpha = 1$, we perform, for each value of a a least-square fit of $\ln t_\beta^1$ in terms of β to obtain the scaling $t_\beta^1 \sim \mathrm{e}^{\beta r(a)}$. In view of (4.64), we next compare $r(a)$ with $\delta_0/(1 + a)$. Numerically, we estimate $\delta_0 \sim 2.4$, in accordance with the value found in [23]. The numerical results on Figure 4.3 are therefore in excellent agreement with the scaling expected from the toy model.

For $\alpha \in (\frac{1}{2}, 1)$, we perform a fit of $\ln(t_\beta^\alpha)$ in terms of $\ln\beta$ to obtain the scaling $\ln(t_\beta^\alpha) \sim s(\alpha, a)\ln\beta$. In view of (4.65), we expect $s(\alpha, a)$ to be independent of a and close to $1/(1-\alpha)$. The slopes $s(\alpha, a)$ obtained in the numerical experiments displayed on Figure 4.4 for two values of α (namely $\alpha = 0.6$ and $\alpha = 0.8$) are reported in Table 4.1. They are

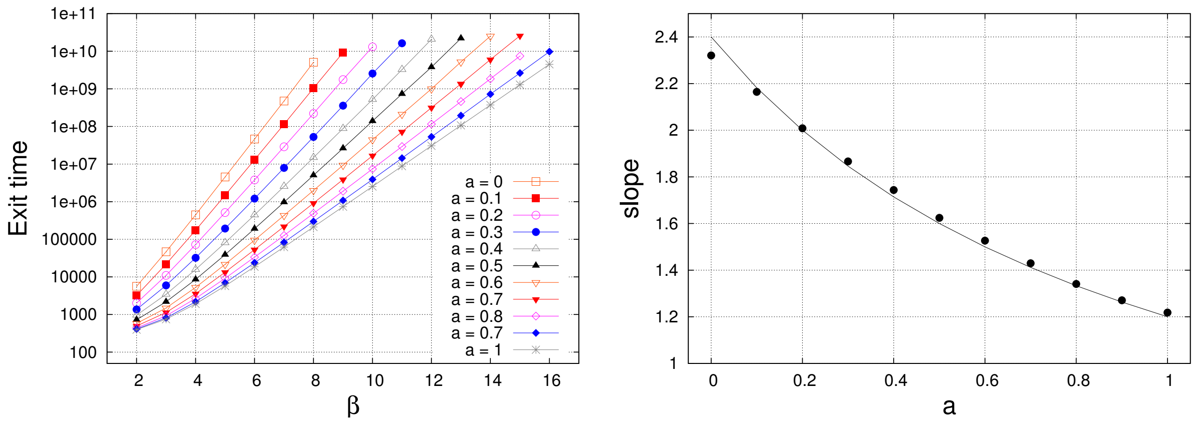

FIGURE 4.3 – Case $\alpha = 1$. Left : Exit times for various values of a. Right : Associated slopes $r(a)$, fitted by $2.4/(1+a)$.

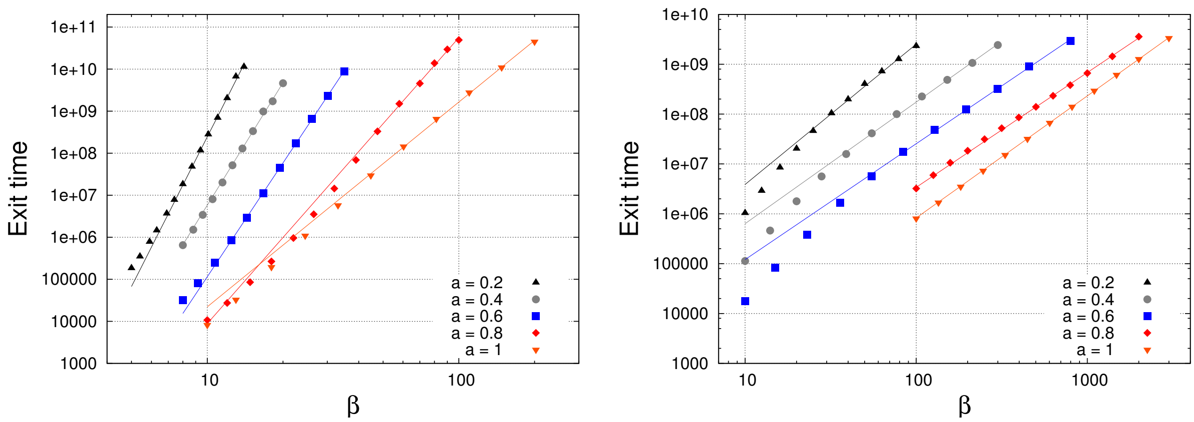

FIGURE 4.4 – Case $\alpha \in (1/2, 1)$. Left : Exit times for $\alpha = 0.8$. Right : Exit times for $\alpha = 0.6$. In both cases, a linear fit is superimposed in solid lines to the data.

	$a = 0.2$	$a = 0.4$	$a = 0.6$	$a = 0.8$	$a = 1$	expected
$\alpha = 0.6$	2.72	2.38	2.27	2.33	2.44	2.5
$\alpha = 0.8$	11.8	9.69	9.00	6.80	4.87	5

TABLE 4.1 – Numerically estimated slopes $s(\alpha, a)$ of the exit times $t_\beta^\alpha \sim \beta^{s(\alpha,a)}$, for various values of a (see Figure 4.4). The expected values are $s(\alpha) = 1/(1-\alpha)$, namely $s(0.6) = 2.5$ and $s(0.8) = 5$.

close to the expected values for $\alpha = 0.6$, as well as for $\alpha = 0.8$ when a is close to 1. The discrepancies observed for $\alpha = 0.8$ and small values of a may be due to the fact it is difficult to reach the asymptotic regime $\beta \to \infty$ in this setting. It might be that the slopes of the curves would decrease for much larger values of β.

Bibliographie

[1] Alfonsi, A. : On the discretization schemes for the CIR (and Bessel squared) processes, Monte Carlo Methods and Applications 11(4) :355-384, 2005.

[2] Alfonsi, A. : High order discretization schemes for the CIR process : Application to affine term structure and Heston models, Mathematics of Computation 79 :209-237, 2010.

[3] Alfonsi, A., Jourdain, B. and Kohatsu-Higa, A. : Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme, Ann. Appl. Probab. 24(3) :1049-1080, 2014

[4] Andrieu, C., Moulines, E. and Priouret, P. : Stability of stochastic approximation under verifiable conditions. SIAM J. Control Optim., 44 :283–312, 2005.

[5] Applebaum, D. : Lévy processes and stochastic calculus. Cambridge Studies in Advanced Mathematics, 93. Cambridge University Press, Cambridge, 2004.

[6] Bally, V. and Talay, D. : The law of the Euler scheme for stochastic differential equations (I) : convergence rate of the distribution function. Probability Theory and Related Fields, 104 :43-60, 1995.

[7] Bally, V. and Talay, D. : The law of the Euler scheme for stochastic differential equations (II) : convergence rate of the density. Monte Carlo Methods and Applications, 2 :93-128, 1996.

[8] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics : A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100 :020603, 2008.

[9] Barndorff-Nielsen, O. : Processes of normal inverse Gaussian type. Finance Stoch. 2(1) :41-68, 1998.

[10] Bally, V., Kohatsu-Higa, A. : A probabilistic interpretation of the parametrix method. Ann. Appl. Probab. 25(6) :3095-3138, 2015.

[11] Bertoin, J. : Lévy processes. Cambridge Tracts in Mathematics, 121. Cambridge University Press, Cambridge, 1996.

[12] Bencheikh, O. and Jourdain, B. : Convergence in total variation of the Euler-Maruyama scheme applied to diffusion processes with measurable drift coefficient and additive noise, Preprint arXiv :2005.09354.

[13] Beskos, A., Papaspiliopoulos, O. and Roberts, G. : Retrospective exact simulation of diffusion sample paths with applications. Bernoulli 12(6) :1077-1098, 2006.

[14] Beskos, A., Papaspiliopoulos, O., Roberts, G.and Fearnhead, P. : Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 68(3)333-382, 2006.

[15] Beskos, A. and Roberts, G. : Exact simulation of diffusions. Ann. Appl. Probab. 15(4) :2422-2444, 2005.

[16] Breiman, L. : Probability. Corrected reprint of the 1968 original. Classics in Applied Mathematics, 7. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

[17] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.*, 96 :090601, 2006.

[18] Carr, P., Geman, H., Madan, D.B. and Yor, M. : From local volatility to local Lévy models. Quant. Finance 4(5) :581–588, 2004.

[19] Y. Crespo, F. Marinelli, F. Pietrucci, and A. Laio. Metadynamics convergence law in a multidimensional system. *Phys. Rev. E*, 81(5) :055701, 2010.

[20] J.F. Dama, G.M. Hocky, R. Sun and G.A. Voth. Exploring valleys without climbing every peak : More efficient and forgiving metabasin metadynamics via robust on-the-fly bias domain restriction. *J. Chem. Theory Comput.*, 11(12) :5638-5650, 2015

[21] J.F. Dama, M. Parrinello, and G.A. Voth. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.*, 112 :240602(1–6), 2014.

[22] Del Moral, P. : Feynman-Kac formulae. Genealogical and interacting particle systems with applications. Probability and its Applications (New York). Springer-Verlag, 2004.

[23] G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre, and G. Stoltz. Efficiency of the Wang-Landau algorithm : A simple test case. *Appl. Math. Res. Express*, 2014(2) :275–311, 2014.

[24] G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre, and G. Stoltz. Convergence of the Wang-Landau algorithm. *Math. Comput.*, 84(295) :2297–2327, 2015.

[25] G. Fort, B. Jourdain, T. Lelièvre, and G. Stoltz. Self-Healing Umbrella Sampling : Convergence and efficiency. *Stat Comput.*, 27(1), 147-168, 2017.

[26] G. Fort, E. Moulines, and P. Priouret. Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.*, 39(6) :3262–3289, 2012.

[27] Gaines, J. and Lyons, T. : Random generation of stochastic area integrals. SIAM J. Appl. Math. 54(4) :1132-1146, August 1994.

[28] Giles, M. : Multi-level Monte Carlo path simulation, Operations Research, 56(3) :607-617, 2008.

[29] Giles, M. and Szpruch, L. : Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation, Ann. Appl. Probab. 24(4) :1585-1620, 2014.

[30] Gobet, E. : Revisiting the Greeks for European and American options. Proceedings of the "International Symposium on Stochastic Processes and Mathematical Finance" at Ritsumeikan University, Kusatsu, Japan, March 2003. 18 pages. 2004.

[31] Gobet, E. Weak approximation of killed diffusion using Euler schemes. Stochastic Processes and their Applications, 87 :167-197, 2000.

[32] Gobet, E. Euler Schemes and Half-Space Approximation for the Simulation of Diffusion in a Domain, ESAIM P&S, 5 :261-297, 2001.

[33] Gobet, E. Menozzi, S. Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme, Stochastic Processes and their Applications, 112(2) :201-223, 2004.

[34] Guyon, J. : Euler scheme and tempered distributions. Stochastic Processes and their Applications, 116(6) :877-904, 2006.

[35] Hairer, M., Mattingly, J. C., Yet another look at Harris' ergodic theorem for Markov chains. Seminar on Stochastic Analysis, Random Fields and Applications VI, 109-117, Progr. Probab., 63, Birkhäuser/Springer Basel AG, Basel, 2011.

[36] Hutzenthaler, M., Jentzen, A. and Kloeden, P. E., Strong convergence of an explicit numerical method for SDEs with non-globally Lipschitz continuous coefficients, Annals of Applied Probability 22(4) :1611-1641, 2012.

[37] Jacob, P.E. and Ryder, R.J. . The Wang-Landau algorithm reaches the flat histogram criterion in finite time. *Ann. Appl. Probab.*, 24(1) :34–53, 2014.

[38] Jacod, J., Kurtz, T.G., Méléard, S., Protter, P. : The approximate Euler method for Lévy driven stochastic differential, Ann. Inst. H. Poincaré Probab. Statist. 41(3) :523-558, 2005.

[39] Jarner, S.F., Hansen, E. : Geometric ergodicity of Metropolis algorithms, Stochastic Processes and their Applications, 85 :341-361, 2000.

[40] Jourdain, B. and Sbai, M. : Exact retrospective Monte Carlo computation of arithmetic average Asian options, Monte Carlo methods and Applications 13(2) :135-171, 2007.

[41] Jourdain, B. and Sbai, M. : High order discretization schemes for stochastic volatility models, Journal of Computational Finance 17(2), 2013 .

[42] Kloeden, P. E. and Platen, E. : Numerical Solution of Stochastic Differential Equations. Springer Berlin Heidelberg New York, 1992.

[43] Kemna, A. G. Z. and Vorst, A. C. F. : A Pricing Method for Options Based on Average Asset Values. Journal of Banking and Finance, 14 :113-129, March 1990.

[44] Kohatsu-Higa, A. and Tankov, P. : Jump adapted discretization schemes for Lévy driven SDEs, Stochastic Process. Appl., 120(11) :2258-2285, 2010.

[45] Kunita, H. : Stochastic flows and stochastic differential equations. Cambridge University Press, 1990.

[46] Kurtz, T. and Protter, P. : Wong-Zakai corrections, random evolutions, and simulation schemes for SDEs. Stochastic analysis, 331–346, Academic Press, Boston, MA, 1991.

[47] Kusuoka, S. : Approximation of expectation of diffusion process and mathematical finance. Taniguchi Conference on Mathematics Nara '98, 147-165, Adv. Stud. Pure Math., 31, Math. Soc. Japan, Tokyo, 2001.

[48] Kusuoka, S. : Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. Advances in mathematical economics. Vol. 6, 69-83, 2004.

[49] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A*, 99 :12562–12566, 2002.

[50] Lapeyre, B. and Temam, E. : Competitive Monte Carlo methods for the pricing of Asian Options. Journal of Computational Finance, 5(1) :39-59, Fall 2001

[51] Lyons, T. and Victoir, N. : Cubature on Wiener space. Stochastic analysis with applications to mathematical finance. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 460 :169-198, 2004.

[52] Mao, X and Yuan, C. : A note on the rate of convergence of the Euler-Maruyama method for stochastic differential equations. Stochastic Analysis and Applications, 26(2) :325-333, 2008.

[53] Milstein, G. N. : Numerical integration of stochastic differential equations. Translated and revised from the 1988 Russian original. Mathematics and its Applications, 313. Kluwer Academic Publishers Group, Dordrecht, 1995

[54] Madan, D., Carr, P. and Chang, E. : The Variance Gamma Process and Option Pricing, European Finance Review 2 :79-105, 1998.

[55] S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino. Self-healing Umbrella Sampling : A non-equilibrium approach for quantitative free energy calculations. *J. Phys. Chem. B*, 110(29) :14011–14013, 2006.

[56] M. McGovern and J. de Pablo. A boundary correction algorithm for metadynamics in multiple dimensions. *J. Chem. Phys.*, 139(8) :084102, 2013.

[57] Newton, N. : Variance reduction for simulated diffusions. SIAM J. Appl. Math. 54(6) :1780-1805, 1994.

[58] Newton, N. : Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations. SIAM J. Appl. Math. 51(2) :542-567, 1991.

[59] Ninomiya, S. : A new simulation scheme of diffusion processes : application of the Kusuoka approximation to finance problems. Math. Comput. Simulation 62 :479-486, 2003.

[60] Ninomiya, S. : A partial sampling method applied to the Kusuoka approximation. Monte Carlo Methods Appl. 9(1) :27-38, 2003.

[61] Ninomiya, S. and Victoir, N. : Weak approximation of stochastic differential equations and application to derivative pricing, Applied Mathematical Finance, 15(2) :107-121, 2008.

[62] Ninomiya, M. and Ninomiya, S. : A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. Finance and Stochastics 13 :415-443, 2009.

[63] Pagès, G. Convex order for path-dependent derivatives : a dynamic programming approach. Séminaire de Probabilités XLVIII, 33–96, Lecture Notes in Math., 2168, Springer, Cham, 2016.

[64] Pagès, G. Multi-step Richardson-Romberg extrapolation : remarks on variance control and complexity. Monte Carlo Methods Appl. 13(1) :37-70, 2007.

[65] Peskun, P.H. : Optimal Monte-Carlo sampling using Markov chains. Biometrika 60(3) :607-612, 1973.

[66] Protter, P. and Talay, D. : The Euler scheme for Lévy driven stochastic differential equations. Ann. Probab. 25(1) :393-423, 1997.

[67] Roberts, G.O., Tweedie, R.L. : Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika 83(1) :95-110, 1996.

[68] Rubenthaler, S. : Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process. Stochastic Process. Appl. 103(2) :311-349, 2003.

[69] Sato, K. : Lévy processes and infinitely divisible distributions. Cambridge Studies in Advanced Mathematics, 68. Cambridge University Press, Cambridge, 1999.

[70] Talay, D. and Tubaro, L. : Expansion of the global error for numerical schemes solving stochastic differential equations. Stochastic Anal. Appl. 8(4) :483-509, 1990.

[71] Tanaka, H. and Kohatsu-Higa, A. : An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs. Ann. Appl. Probab 19(3) :1026-1062, 2009.

[72] Temam, E. : Couverture approchée d'options exotiques. Pricing des options asiatiques. PhD thesis, Université Paris 6, 2001.

[73] Tierney, L., A note on Metropolis-Hastings kernels for general state spaces, Ann. Appl. Probab. 8(1) :1-9, 1998

[74] F. Wang and D.P. Landau. Determining the density of states for classical statistical models : A random walk algorithm to produce a flat histogram. *Phys. Rev. E*, 64 :056101, 2001.

[75] F. Wang and D.P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86(10) :2050–2053, 2001.