Corrigé de l'examen du cours de M2 : Méthodes numériques probabilistes

16 janvier 2018, 13h00-16h00.

Exercice : Métastabilité à basse température

1. La variable aléatoire τ^x est un temps d'arrêt.
2. On a $u(x_1) = 0$. Par ailleurs, on calcule

$$Tu'(x) = \int_{z=x}^{+\infty} \exp\left(\frac{V(x) - V(z)}{T}\right) dz$$

et donc

$$Tu''(x) = -1 + \frac{1}{T}\int_{z=x}^{+\infty} V'(x) \exp\left(\frac{V(x) - V(z)}{T}\right) dz.$$

Par conséquent

$$\begin{aligned} Lu(x) &= Tu''(x) - V'(x)u'(x) \\ &= -1 + \int_{z=x}^{+\infty} V'(x) \exp\left(\frac{V(x) - V(z)}{T}\right) dz - V'(x)\int_{z=x}^{+\infty} \exp\left(\frac{V(x) - V(z)}{T}\right) dz \\ &= -1. \end{aligned}$$

3. On a que $\tilde{u}(x) = \phi(x)u(x)$ est nulle pour $x < x_1 - 1$. Pour $x \in [x_1 - 1, \bar{x}]$, on écrit que

$$\begin{aligned} Tu(x) &= \int_{y=x_1}^{x}\int_{z=y}^{+\infty} \exp\left(\frac{V(y) - V(z)}{T}\right) dzdy \\ &\leq \int_{y=x_1}^{x} \exp(V(y)/T) \int_{z=y}^{\bar{x}} \exp(-V(z)/T) dzdy \\ &\quad + \int_{y=x_1}^{x} \exp(V(y)/T) \int_{z=\bar{x}}^{\infty} \exp(-(a + bz^p)/T) dzdy \end{aligned}$$

et ces fonctions sont clairement bornées. Pour $x > \bar{x}$, on a

$$\begin{aligned} Tu(x) &= Tu(\bar{x}) + \int_{y=\bar{x}}^{x}\int_{z=y}^{+\infty} \exp\left(\frac{V(y) - V(z)}{T}\right) dzdy \\ &\leq Tu(\bar{x}) + \int_{y=\bar{x}}^{x} \exp(ay^p/T) \int_{z=y}^{\infty} \exp(-az^p/T) dzdy. \end{aligned}$$

On estime maintenant

$$\begin{aligned} \int_{z=y}^{\infty} \exp(-az^p/T) dz &\leq \frac{T}{apy^{p-1}} \int_{z=y}^{\infty} \frac{apz^{p-1}}{T} \exp(-az^p/T) dz \\ &= \frac{T}{apy^{p-1}} \exp(-ay^p/T). \end{aligned}$$

Par conséquent,

$$\begin{aligned}
Tu(x) &\le Tu(\bar{x}) + \int_{y=\bar{x}}^{x} \exp(ay^p/T)\frac{T}{apy^{p-1}}\exp(-ay^p/T)\,dy \\
&= Tu(\bar{x}) + \frac{T}{ap}\int_{y=\bar{x}}^{x} \frac{1}{y^{p-1}}\,dy \\
&\le Tu(\bar{x}) + \frac{T}{ap}\int_{y=\bar{x}}^{\infty} \frac{1}{y^{p-1}}\,dy \\
&= Tu(\bar{x}) + \frac{T}{ap(2-p)}\bar{x}^{2-p} < \infty.
\end{aligned}$$

en utilisant le fait que $p > 2$. Ceci donne une borne L^∞ sur u. Les bornes sur u' et u'' s'obtiennent de la même façon.

4. Il suffit d'appliquer un calcul d'itô sur $\tilde{u}(X_t^x)$, en utilisant les résultats de régularité donnée dans la question précédente.
5. L'égalité est évidente. Il reste donc à montrer que l'espérance de $\int_{s=0}^t \mathbf{1}_{\{s\le\tau^x\}}\tilde{u}'(X_s^x)\mathrm{d}B_s$ est nulle. Pour cela, il suffit de montrer que $\int_0^t \mathbb{E}(|\mathbf{1}_{\{s\le\tau^x\}}\tilde{u}'(X_s^x)|^2)\,ds < \infty$ ce qui est clair en majorant simplement $|\mathbf{1}_{\{s\le\tau^x\}}\tilde{u}'(X_s^x)|$ par $\|\tilde{u}'\|_{L^\infty}$.
6. En prenant l'égalité de la Question 4 en $t = t\wedge\tau^x$, on a :

$$\tilde{u}(X_{t\wedge\tau^x}^x) = \tilde{u}(x) + \int_{s=0}^{t\wedge\tau^x} L\tilde{u}(X_s^x)\mathrm{d}s + \sqrt{2T}\int_{s=0}^{t\wedge\tau^x} \tilde{u}'(X_s^x)\mathrm{d}B_s.$$

d'où, en utilisant le fait que $L\tilde{u}(x) = -1$ pour $x \ge x_1 - 1$ et en prenant l'espérance : pour $x \ge x_1 - 1$,

$$\mathbb{E}(\tilde{u}(X_{t\wedge\tau^x}^x)) = u(x) - \mathbb{E}(t\wedge\tau^x).$$

7. On note que $\mathbb{E}(t\wedge\tau^x) = u(x) - \mathbb{E}(\tilde{u}(X_{t\wedge\tau^x}^x)) \le u(x)$ et donc, par convergence monotone, en faisant tendre $t\to\infty$, $\mathbb{E}(\tau^x) \le u(x) < \infty$. Par conséquent, par convergence dominée, on a $\mathbb{E}(\tilde{u}(X_{t\wedge\tau^x}^x))$ converge vers $\mathbb{E}(\tilde{u}(X_{\tau^x}^x)) = \tilde{u}(x_1) = u(x_1) = 0$ quand $t\to\infty$. On a donc finalement : pour $x \ge x_1$,

$$\mathbb{E}(\tau^x) = u(x).$$

8. Il s'agit de montrer un équivalent par la méthode de Laplace.

On commence par une remarque sur l'hypothèse d'intégrabilité. Par hypothèse, $M(\xi) \le M(\xi_0)$, et donc $\exp(M(\xi)/\eta) = \exp((M(\xi)-M(\xi_0))/\eta)\exp(M(\xi_0)/\eta)$ avec $M(\xi)-M(\xi_0) \le 0$. En particulier, pour tout $\eta\in(0,\eta_0]$, $\int_{\xi\in D}\exp\left(\frac{M(\xi)}{\eta}\right)\mathrm{d}\xi \le \int_{\xi\in D}\exp\left(\frac{M(\xi)-M(\xi_0)}{\eta_0}\right)\mathrm{d}\xi\exp(M(\xi_0)/\eta) < +\infty$.

Montrons maintenant le résultat asymptotique. Soit $\epsilon > 0$. Il existe $r > 0$ tel que pour tout $\xi\in D$, si $|\xi-\xi_0| \le r$, alors

$$\left|M(\xi) - M(\xi_0) - \frac{1}{2}(\xi-\xi_0)^T\nabla^2 M(\xi_0)(\xi-\xi_0)\right| \le \epsilon|\xi-\xi_0|^2.$$

On écrit alors

$$\begin{aligned}\int_{\xi\in D}\exp\left(\frac{M(\xi)}{\eta}\right)d\xi &= \int_{\xi\in D,|\xi-\xi_0|\leq r}\exp\left(\frac{M(\xi)}{\eta}\right)d\xi + \int_{\xi\in D,|\xi-\xi_0|>r}\exp\left(\frac{M(\xi)}{\eta}\right)d\xi \\ &= \eta\exp\left(\frac{M(\xi_0)}{\eta}\right)\left(\eta^{-1}\int_{\xi\in D,|\xi-\xi_0|\leq r}\exp\left(\frac{M(\xi)-M(\xi_0)}{\eta}\right)d\xi\right. \\ &\qquad \left. + \eta^{-1}\int_{\xi\in D,|\xi-\xi_0|>r}\exp\left(\frac{M(\xi)-M(\xi_0)}{\eta}\right)d\xi\right). \qquad (1)\end{aligned}$$

Pour le premier terme, par un changement de variable $\xi = \xi_0 + \sqrt{\eta}\zeta$, on a

$$\begin{aligned}&\eta^{-1}\int_{\xi\in D,|\xi-\xi_0|\leq r}\exp\left(\frac{M(\xi)-M(\xi_0)}{\eta}\right)d\xi \\ &= \int_{\mathbb{R}^2}1_{\xi_0+\sqrt{\eta}\zeta\in D,|\zeta|\leq r/\sqrt{\eta}}\exp\left(\frac{M(\xi_0+\sqrt{\eta}\zeta)-M(\xi_0)}{\eta}\right)d\zeta =: I(\eta)\end{aligned}$$

On note en utilisant le développement de Taylor ci-dessus que pour tout ζ tel que la fonction indicatrice est non nulle,

$$\frac{\eta}{2}\zeta^T\nabla^2M(\xi_0)\zeta - \eta\epsilon|\zeta|^2 \leq M(\xi_0+\sqrt{\eta}\zeta) - M(\xi_0) \leq \frac{\eta}{2}\zeta^T\nabla^2M(\xi_0)\zeta + \eta\epsilon|\zeta|^2.$$

On peut supposer depuis le début ϵ assez petit pour que $\nabla^2M(\xi_0)+\epsilon\mathrm{Id}$ soit symétrique définie négative, de sorte que $\int_{\mathbb{R}^2}\exp\left(\frac{\eta}{2}\zeta^T\nabla^2M(\xi_0)\zeta \pm \eta\epsilon|\zeta|^2\right)d\zeta < \infty$. Comme D est ouvert, $1_{\xi_0+\sqrt{\eta}\zeta\in D,|\zeta|\leq r/\sqrt{\eta}}$ converge vers 1 quand $\eta\to 0$, et donc, par convergence dominée,

$$\begin{aligned}&\lim_{\eta\to0}\int_{\mathbb{R}^2}\exp\left(\frac{\eta}{2}\zeta^T\nabla^2M(\xi_0)\zeta - \eta\epsilon|\zeta|^2\right)d\zeta \leq \liminf_{\eta\to0} I(\eta) \\ &\qquad \leq \limsup_{\eta\to0} I(\eta) \leq \lim_{\eta\to0}\int_{\mathbb{R}^2}\exp\left(\frac{\eta}{2}\zeta^T\nabla^2M(\xi_0)\zeta + \eta\epsilon|\zeta|^2\right)d\zeta.\end{aligned}$$

On a donc, en calculant les bornes sup et inf :

$$\frac{2\pi}{\sqrt{\det(-\nabla^2M(\xi_0)+2\epsilon\mathrm{Id})}} \leq \liminf_{\eta\to0} I(\eta) \leq \limsup_{\eta\to0} I(\eta) \leq \frac{2\pi}{\sqrt{\det(-\nabla^2M(\xi_0)-2\epsilon\mathrm{Id})}}.$$

On considère maintenant le second terme dans (1). On sait qu'il existe $\delta>0$ tel que pour $\xi\in D$ et $|\xi-\xi_0|>r$, $M(\xi)\leq M(\xi_0)-\delta$. On écrit :

$$\begin{aligned}&\eta^{-1}\int_{\xi\in D,|\xi-\xi_0|>r}\exp\left(\frac{M(\xi)-M(\xi_0)}{\eta}\right)d\xi \\ &= \eta^{-1}\exp\left(-\frac{\delta/2}{\eta}\right)\int_{\xi\in D,|\xi-\xi_0|>r}\exp\left(\frac{M(\xi)-M(\xi_0)+\delta/2}{\eta}\right)d\xi.\end{aligned}$$

Par convergence dominée, puisque pour tout $\eta\leq\eta_0$, $\exp\left(\frac{M(\xi)-M(\xi_0)+\delta/2}{\eta}\right)\leq\exp\left(\frac{M(\xi)-M(\xi_0)+\delta/2}{\eta_0}\right)$ et que $\int_{\xi\in D,|\xi-\xi_0|>r}\exp\left(\frac{M(\xi)-M(\xi_0)+\delta/2}{\eta_0}\right)d\xi<\infty$, on a

$$\lim_{\eta\to0}\int_{\xi\in D,|\xi-\xi_0|>r}\exp\left(\frac{M(\xi)-M(\xi_0)+\delta/2}{\eta}\right)d\xi = 0.$$

On en déduit que $\lim_{\eta\to 0} \eta^{-1} \int_{\xi\in D, |\xi-\xi_0|>r} \exp\left(\frac{M(\xi)-M(\xi_0)}{\eta}\right) d\xi = 0$.
On en déduit que

$$\frac{2\pi}{\sqrt{\det(-\nabla^2 M(\xi_0) + 2\epsilon \mathrm{Id})}} \le \liminf_{\eta\to 0} J(\eta) \le \limsup_{\eta\to 0} J(\eta) \le \frac{2\pi}{\sqrt{\det(-\nabla^2 M(\xi_0) - 2\epsilon \mathrm{Id})}}$$

où $J(\eta) = \frac{1}{\eta}\exp\left(-\frac{M(\xi_0)}{\eta}\right) \int_{\xi\in D} \exp\left(\frac{M(\xi)}{\eta}\right) d\xi$. On conclut ensuite en prenant la limite $\epsilon \to 0$.

9. On considère la fonction $M(y,z) = V(y) - V(z)$ sur le domaine $D_x = \{(y,z), x_1 < y < x, z > y\}$, avec $x > x_*$. On a clairement que M atteint son maximum sur D_x en (x_*, x_2). Par ailleurs, on a $M < 0 < \max_{D_x} M$ sur $D_x \cap \{y > x_2\}$. De plus, pour $r > 0$, si $|y - x_*| \ge r$ et $x_1 < y \le x_2$, on a $V(y) \le V(x_*) - \delta$ pour un $\delta > 0$ et donc, pour tout $(y,z) \in D_x$ avec $|y - x_*| \ge r$ et $y \le x_2$, $M(y,z) = V(y) - V(z) \le V(x_*) - \delta - \min_{D_x} V(z) = V(x_*) - \delta - V(x_2)$. Enfin, on calcule la hessienne de M en (x_*, x_2) :

$$\nabla^2 M(x_*, x_2) = \begin{bmatrix} V''(x_*) & 0 \\ 0 & -V''(x_2) \end{bmatrix}.$$

C'est une matrice définie négative puisqu'on a supposé $V''(x_2) > 0$ et $V''(x_*) < 0$. En utilisant le résultat de la Question 8, et le fait que $\mathbb{E}(\tau^x) = u(x) = \frac{1}{T}\int_{(y,z)\in D_x} \exp\left(\frac{V(y)-V(z)}{T}\right) \mathrm{d}z\mathrm{d}y$, on a donc, pour $x > x_*$, quand $T \to 0$,

$$\mathbb{E}(\tau^x) \sim \exp\left(\frac{V(x_*) - V(x_2)}{T}\right) \frac{2\pi}{\sqrt{V''(x_2)|V''(x_*)|}}.$$

10. Cette formule s'appelle la formule d'Eyring-Kramers (ou loi d'Arrhenius).

Problème : Algorithme de Metropolis Hastings

Première partie : l'algorithme de Metropolis Hastings.

1. On a

$$P(x,y) = \begin{cases} Q(x,y)a(x,y) \text{ si } x \neq y \\ 1 - \sum_{y\neq x} Q(x,y)a(x,y) \text{ si } x = y. \end{cases}$$

2. On a (X_n) réversible par rapport à π si et seulement si, pour tout x, y $\pi(x)P(x,y) = \pi(y)P(y,x)$, i.e. si et seulement si pour tout $x \neq y$ tel que $Q(x,y) > 0$, $\pi(x)Q(x,y)a(x,y) = \pi(y)Q(y,x)a(y,x)$, ce qui est équivalent à : pour tout $x \neq y$ tel que $Q(x,y) > 0$, $a(x,y) = r(x,y)a(y,x)$.
Soit maintenant une fonction F telle que $F(z) = zF(1/z)$. On pose $\lambda(x,y) = a(x,y)/F(r(x,y))$. Pour tout $x \neq y$ tel que $Q(x,y) > 0$, on a

$$\begin{aligned} a(x,y) = r(x,y)a(y,x) &\iff \lambda(x,y)F(r(x,y)) = r(x,y)\lambda(y,x)F(r(y,x)) \\ &\iff \lambda(x,y)F(r(x,y)) = r(x,y)\lambda(y,x)F(1/r(x,y)) \\ &\iff \lambda(x,y)F(r(x,y)) = \lambda(y,x)F(r(x,y)) \\ &\iff \lambda(x,y) = \lambda(y,x). \end{aligned}$$

On vérifie facilement que les règles de Barker et de Metropolis Hastings donnent bien des chaînes réversibles par rapport à π.

3. Dans ce cas, on a $r(x,y) = \exp(-\beta(V(y) - V(x)))$ et donc $a(x,y) = \min(1, \exp(-\beta(V(y) - V(x)))$ (noter qu'on n'a donc pas besoin de connaître Z pour implémenter l'algorithme). On voit que $a = 1$ si $V(y) \leq V(x)$ i.e. si on propose un déplacement vers une position plus basse en énergie. On voit également que si on propose un déplacement vers une position plus haute en énergie ($V(y) > V(x)$), on accepte le déplacement avec la probabilité $\exp(-\beta(V(y) - V(x)))$ qui tend vers 0 quand la température tend vers 0 ($\beta \to \infty$).

Deuxième partie : généralités sur les chaînes de Markov réversibles.

4. On calcule : pour $(x_0, \ldots, x_n) \in M^{n+1}$,

$$\begin{aligned}\mathbb{P}((X_0, \ldots, X_n) = (x_0, \ldots, x_n)) &= \pi(x_0)P(x_0, x_1)P(x_1, x_2) \ldots P(x_{n-1}, x_n)\\ &= P(x_1, x_0)\pi(x_1)P(x_1, x_2) \ldots P(x_{n-1}, x_n)\\ &= P(x_1, x_0)P(x_2, x_1) \ldots P(x_n, x_{n-1})\pi(x_n)\\ &= \pi(x_n)P(x_n, x_{n-1}) \ldots P(x_2, x_1)P(x_1, x_0)\\ &= \mathbb{P}((X_0, \ldots, X_n) = (x_n, \ldots, x_0))\\ &= \mathbb{P}((X_n, \ldots, X_0) = (x_0, \ldots, x_n)).\end{aligned}$$

5. Si P est réversible par rapport à π, pour tout f, g deux vecteurs de $\mathbb{R}^d$, on a

$$\begin{aligned}\langle Pf, g\rangle_\pi &= \sum_{x,y} P(x,y)f(y)g(x)\pi(x)\\ &= \sum_{x,y} P(y,x)f(y)g(x)\pi(y)\\ &= \langle f, Pg\rangle_\pi.\end{aligned}$$

   Pour la réciproque, il suffit de considérer $f(x) = 1_{x=x_0}$ et $g(y) = 1_{y=y_0}$ pour montrer que $\langle Pf, g\rangle_\pi = \langle f, Pg\rangle_\pi$ implique $\pi(x_0)P(x_0, y_0) = \pi(y_0)P(y_0, x_0)$. Comme P est autoadjoint pour ce produit scalaire, il est diagonalisable dans une base orthonormée pour ce produit scalaire.

6. On a que pour tout f, pour tout $x \in M$, $|Pf(x)| = |\sum_y P(x,y)f(y)| \leq \|f\|_{L^\infty} \sum_y P(x,y) = \|f\|_{L^\infty}$ et donc $\|Pf\|_{L^\infty} \leq \|f\|_{L^\infty}$. Ceci implique que pour toute valeur propre λ, $|\lambda| \leq 1$. De plus, 1 est clairement valeur propre pour le vecteur propre $e_1 = (1, \ldots, 1)^T$.

7. On peut utiliser le fait que les mesures invariantes sont les vecteurs propres associées à la valeur propre 1 pour P^T, et que P et P^T ont même spectre, avec des sous-espaces propres associées de même dimension. On peut aussi le démontrer directement de la façon suivante. Soit ν une mesure invariante : $\nu P = \nu$, et donc $\sum_{x \in M} \nu(x)P(x,y) = \nu(y)$. Ceci implique que $\sum_{x \in M} \nu(x)P(y,x)\pi(y)/\pi(x) = \nu(y)$, soit $\sum_{x \in M} P(y,x)f(x) = f(y)$ avec $f(x) = \nu(x)/\pi(x)$. On en déduit que $\nu(x)/\pi(x)$ est constant, et donc $\nu = \pi$.

8. En utilisant le fait que $\langle \bar{f}, e_1\rangle_\pi = 0$, on a

$$\|P^n \bar{f}\|^2_{L^2(\pi)} = \sum_{i=2}^{d} \lambda_i^{2n} |\langle \bar{f}, e_i\rangle_\pi|^2 \leq \rho^{2n} \|\bar{f}\|^2_{L^2(\pi)}.$$

On écrit ensuite

$$\begin{aligned}
|\mathbb{E}^x(f(X_n)) - \mathbb{E}_\pi(f)|^2 &= |(P^n\bar{f})(x)|^2 \\
&= |(P^n\bar{f})(x)|^2\pi(x)\frac{1}{\pi(x)} \\
&\leq \sum_{y\in M} |(P^n\bar{f})(y)|^2\pi(y)\frac{1}{\pi(x)} \\
&= \|P^n\bar{f}\|^2_{L^2(\pi)}\frac{1}{\pi(x)}.
\end{aligned}$$

On utilise finalement le fait que $\|P^n\bar{f}\|^2_{L^2(\pi)} \leq \rho^{2n}\|\bar{f}\|^2_{L^2(\pi)} = \rho^{2n}\mathrm{Var}_\pi(f)$ pour conclure.

9. (a) C'est exactement l'interprétation probabiliste de la condition $\forall(x,y)\in M^2, \exists n\geq 1,\ P^n(x,y) > 0$. En effet

$$\begin{aligned}
P^n(x,y) &= \sum_{(x_0,x_1,\ldots,x_n),x_0=x,x_n=y} P(x_0,x_1)\ldots P(x_{n-1},x_n) \\
&= \sum_{(x_0,x_1,\ldots,x_n),x_0=x,x_n=y} \mathbb{P}((X_0,\ldots,X_n)=(x_0,\ldots,x_n))
\end{aligned}$$

et cette somme est positive strictement si et seulement si l'un de ses termes est positif strictement.

(b) Si $X_0\sim\pi$, la loi du couple (X_0,X_1) est donnée par

$$\mathbb{P}((X_0,X_1)=(x,y)) = \pi(x)P(x,y).$$

On a donc

$$\begin{aligned}
\mathbb{E}^\pi((f(X_1)-f(X_0))^2) &= \sum_{x,y}(f(y)-f(x))^2\pi(x)P(x,y) \\
&= \mathcal{E}(f,P).
\end{aligned}$$

On peut maintenant développer le terme sous la somme :

$$\begin{aligned}
\sum_{x,y}(f(y)-f(x))^2\pi(x)P(x,y) &= \sum_{x,y}(f^2(y)+f^2(x)-2f(x)f(y))\pi(x)P(x,y) \\
&= \sum_{x,y} f^2(y)\pi(x)P(x,y) + \sum_x f^2(x)\pi(x) - \sum_{x,y} 2f(x)f(y)\pi(x)P(x,y) \\
&= 2\sum_x f^2(x)\pi(x) - 2\sum_x f(x)Pf(x)\pi(x)
\end{aligned}$$

où on a utilisé le fait que $\pi P = \pi$. On en déduit le résultat :

$$\sum_{x,y}(f(y)-f(x))^2\pi(x)P(x,y) = 2\mathbb{E}^\pi(f\cdot(f-Pf)(X_0)).$$

(c) On suppose P irréductible. Si $Pf=f$, on a $0=\mathcal{E}(f,P)=\sum_{x,y}(f(y)-f(x))^2\pi(x)P(x,y)$ et donc, pour tout x,y, si on considère un chemin $(x_0,x_1,\ldots,x_n)$ menant de x à y de probabilité positive strictement, on a pour tout i, $(f(x_{i+1}-f(x_i))^2\pi(x_i)P(x_i,x_{i+1})=0$ ce qui implique $f(x_i)=f(x_{i+1})$ (puisque $P(x_i,x_{i+1})>0$) et donc finalement $f(x)=f(y)$.
Par conséquent, si P est irréductible, alors $Pf=f$ implique f est constante, et donc $f\in\mathrm{Vect}(e_1)$. Ceci montre que 1 est valeur propre simple.

10. (a) On a :

$$\begin{aligned}\|\mathcal{T}^{n_0}(\nu)-\mathcal{T}^{n_0}(\mu)\|_{L^1} &= \sum_y \Big|\sum_x (\nu(x)-\mu(x))P^{n_0}(x,y)\Big| \\ &= \sum_y \Big|\sum_x (\nu(x)-\mu(x))(P^{n_0}(x,y)-\epsilon\pi(y))\Big| \\ &\le \sum_y\sum_x |\nu(x)-\mu(x)|(P^{n_0}(x,y)-\epsilon\pi(y)) \\ &= (1-\epsilon)\sum_x |\nu(x)-\mu(x)| = (1-\epsilon)\|\nu-\mu\|_{L^1}.\end{aligned}$$

On en déduit par le point fixe de Banach que $\mathcal{T}^{n_0}$ a un unique point fixe, et donc $\mathcal{T}$ a un unique point fixe, qui est nécessairement π.
Par une récurrence évidente, on a pour tout $q\in\mathbb{N}$,

$$\|\mathcal{T}^{qn_0}(\nu)-\mathcal{T}^{qn_0}(\mu)\|_{L^1}\le(1-\epsilon)^q\|\nu-\mu\|_{L^1}.$$

Ensuite, pour $n\ge 0$, on écrit $n=qn_0+r$ avec $q=\left\lfloor\frac{n}{n_0}\right\rfloor$ et $r\in\{0,\dots,n_0-1\}$. On a alors :

$$\begin{aligned}\|\mu P^n-\pi\|_{L^1} &= \|\mathcal{T}^n(\nu)-\mathcal{T}^n(\pi)\|_{L^1} \\ &= \|\mathcal{T}^{qn_0}(\mathcal{T}^r\nu)-\mathcal{T}^{qn_0}(\mathcal{T}^r\pi)\|_{L^1} \\ &\le (1-\epsilon)^q\|\mathcal{T}^r\nu-\mathcal{T}^r\pi\|_{L^1}\end{aligned}$$

et on conclut en notant que pour deux mesures de probabilités μ et ν, on a toujours $\|\mu-\nu\|_{L^1}\le\|\mu\|_{L^1}+\|\nu\|_{L^1}=2$.

(b) Si P est irréductible et apériodique, d'après la question précédente, on a que pour tout $x\in M$, $P^n(x,y)$ converge vers $\pi(y)$ quand $n\to\infty$. Supposons que -1 soit valeur propre : il existe $f\in\mathbb{R}^d$, $f\neq 0$ tel que $Pf=-f$. Par conséquent $(P^{2k+1}f)(x)=-f(x)$ et en faisant tendre k vers l'infini, on obtient $\mathbb{E}_\pi(f)=-f(x)$. f est donc un vecteur constant, nécessairement nul, ce qui est en contradiction avec $f\neq 0$.

Troisième partie : variances asymptotiques pour les chaînes de Markov réversibles.

11. On note $\bar f=f-\mathbb{E}_\pi(f)$. On a $\sum_{n\ge1}\mathrm{Cov}^\pi(f(X_0),f(X_n))=\sum_{n\ge1}\mathbb{E}_\pi\left(\bar fP^n\bar f\right)$. Or $\|P^n\bar f\|_{L^2(\pi)}\le\rho^n\|\bar f\|_{L^2(\pi)}$. On a donc $|\mathbb{E}_\pi\left(\bar fP^n\bar f\right)|\le\rho^n\|\bar f\|^2_{L^2(\pi)}$ et la série est absolument convergente, car $\rho<1$.
On suppose que la chaîne démarre à l'état stationnaire : $X_0\sim\pi$. On note $S(n)=\sum_{k=1}^n\mathrm{Cov}^\pi(f(X_0),f(X_k))$.
On a :

$$\begin{aligned}\mathbb{E}_\pi\left(\frac{1}{\sqrt n}\sum_{k=0}^{n-1}\left(\bar f(X_k)\right)\right)^2 &= \frac1n\sum_{k=0}^{n-1}\mathbb{E}_\pi(\bar f(X_k))^2+\frac2n\sum_{k=0}^{n-2}\sum_{l=k+1}^{n-1}\mathbb{E}_\pi(\bar f(X_k)\bar f(X_l)) \\ &= \mathrm{Var}_\pi(f)+\frac2n\sum_{k=0}^{n-2}\sum_{m=1}^{n-1-k}\mathbb{E}_\pi(\bar f(X_0)\bar f(X_m)) \\ &= \mathrm{Var}_\pi(f)+\frac2n\sum_{k=0}^{n-2}S(n-1-k) \\ &= \mathrm{Var}_\pi(f)+\frac2n\sum_{k=1}^{n-1}S(k)\end{aligned}$$

et on conclut par le Lemme de Césaro, puisque $\lim_{k\to\infty} S(k) = \sum_{n\geq 1} \mathrm{Cov}^\pi(f(X_0), f(X_n)) < \infty$. Si on ne démarre pas sous π, on a

$$\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(\bar{f}(X_k)\right)\right)^2 = \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}(\bar{f}(X_k))^2 + \frac{2}{n}\sum_{k=0}^{n-2}\sum_{l=k+1}^{n-1}\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l)).$$

La loi forte des grands nombres montre que $\frac{1}{n}\sum_{k=0}^{n-1}\bar{f}(X_k)^2$ converge vers $\mathrm{Var}_\pi(f)$ p.s. et dans L^1 et donc $\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}(\bar{f}(X_k))^2 \to \mathrm{Var}_\pi(f)$. Soit maintenant un n assez grand et L tel que $k+1 < k+L < n-1$. On découpe la deuxième somme :

$$\frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=k+1}^{n-1}\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l)) = \frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=k+1}^{k+L}\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l)) + \frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=k+L+1}^{n-1}\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l)).$$

Le premier terme converge vers $\sum_{l=1}^{L}\mathbb{E}_\pi(\bar{f}(X_0)\bar{f}(X_l))$ en appliquant la loi forte des grands nombres à la chaîne de Markov $(X_k, X_{k+1}, \ldots, X_{k+L})_{k\geq 0}$. Pour le deuxième terme, on utilise le fait que, pour $l \geq k$,

$$\begin{aligned}
|\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l))| &= |\mathbb{E}(\bar{f}(X_k)P^{l-k}\bar{f}(X_k))| \\
&\leq \|\bar{f}\|_{L^\infty}\|P^{l-k}\bar{f}\|_{L^\infty} \\
&\leq \|\bar{f}\|_{L^\infty}\frac{\rho^{(l-k)}}{\sqrt{\min\pi}}\sqrt{\mathrm{Var}_\pi(f)}
\end{aligned}$$

en utilisant le résultat de la Question 8, et le fait que $\pi > 0$. On a donc

$$\begin{aligned}
\left|\frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=k+L+1}^{n-1}\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l))\right| &\leq \frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=k+L+1}^{\infty}\mathbb{E}(|\bar{f}(X_k)\bar{f}(X_l)|) \\
&\leq C(f)\frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=k+L+1}^{\infty}\rho^{(l-k)} \\
&\leq C(f)\frac{1}{n}\sum_{k=0}^{n-2}\sum_{l=L+1}^{\infty}\rho^{l} \\
&\leq C(f)\rho^L,
\end{aligned}$$

pour une constante $C(f)$ qui dépend de f. Pour tout $\epsilon > 0$, en choisissant L assez grand pour que $C(f)\rho^L < \frac{\epsilon}{2}$, et que $\sum_{l=1}^{L}\mathbb{E}_\pi(\bar{f}(X_0)\bar{f}(X_l))$ soit proche à $\epsilon/2$ près de $\sum_{l=1}^{\infty}\mathbb{E}_\pi(\bar{f}(X_0)\bar{f}(X_l))$, puis en faisant tendre n vers l'infini, on obtient alors

$$\lim_{n\to\infty}\frac{1}{n}\left|\sum_{k=0}^{n-2}\sum_{l=k+1}^{n-1}\mathbb{E}(\bar{f}(X_k)\bar{f}(X_l)) - \sum_{l=1}^{\infty}\mathbb{E}_\pi(\bar{f}(X_0)\bar{f}(X_l))\right| \leq \epsilon.$$

Ceci permet de conclure en faisant tendre ϵ vers 0.

12. On a

$$\mathrm{Var}_\pi(f) = \|\bar{f}\|_{L^2(\pi)}^2$$

et

$$\begin{aligned}
\sum_{n\geq 1} \mathrm{Cov}^\pi(f(X_0), f(X_n)) &= \sum_{n\geq 1} \langle \bar{f}, P^n \bar{f}\rangle_\pi \\
&= \langle \bar{f}, \sum_{n\geq 1} P^n \bar{f}\rangle_\pi \\
&= \langle \bar{f}, \sum_{n\geq 0} P^n \bar{f}\rangle_\pi - \|\bar{f}\|^2_{L^2(\pi)} \\
&= \langle \bar{f}, (\mathrm{Id} - P)^{-1} \bar{f}\rangle_\pi - \|\bar{f}\|^2_{L^2(\pi)}
\end{aligned}$$

d'où le résultat. On a ici utilisé le fait que $(\mathrm{Id} - P) : \mathrm{Vect}(e_1)^\perp \to \mathrm{Vect}(e_1)^\perp$ est inversible et que P est de norme matricielle (pour la norme triple associée à la norme $L^2(\pi)$, qui est telle que $\|P\|_{L^2(\pi)} = \max |\lambda_i|$) inférieure strictement à 1, ce qui donne la convergence normale de la série $\sum_{n\geq 0} P^n = (\mathrm{Id} - P)^{-1}$. Tout ceci est clair quand on écrit la série dans la base orthonormale $(e_i)_{i=1,\dots,d}$. Noter que de manière consistante, l'équation

$$(\mathrm{Id} - P)g_0 = \bar{f}$$

admet bien une unique solution dans $L^2_0(\pi) = (\mathrm{Ker}(\mathrm{Id} - P))^\perp = (\mathrm{Im}(\mathrm{Id} - P))$ ($L^2_0(\pi)$ est l'espace des vecteurs de moyenne nulle, muni de la norme L^2, introduit à la question suivante).

13. La matrice $\mathrm{Id} - P$ est autoadjointe pour le produit scalaire $\langle \cdot, \cdot\rangle_\pi$ et négative. De plus elle est autoadjointe *définie* négative sur $L^2_0(\pi) = \mathrm{Vect}(e_1)^\perp$. Par conséquent, la fonctionnelle

$$J : \begin{cases} L^2_0(\pi) \to \mathbb{R} \\ \qquad g \mapsto 2\langle \bar{f}, g\rangle_\pi - \langle g, (\mathrm{Id} - P)g\rangle_\pi \end{cases}$$

est continue, dérivable, tend vers l'infini à l'infini et est strictement concave. Elle admet donc un unique maximum en un vecteur g_0 solution de l'équation d'Euler : pour tout $h \in L^2_0(\pi)$,

$$2\langle \bar{f}, h\rangle_\pi - 2\langle h, (\mathrm{Id} - P)g_0\rangle_\pi = 0$$

ce qui donne :

$$(\mathrm{Id} - P)g_0 = \bar{f}.$$

Noter que cette équation admet bien une unique solution dans $L^2_0(\pi) = (\mathrm{Ker}(\mathrm{Id} - P))^\perp = (\mathrm{Im}(\mathrm{Id} - P))$. On a donc

$$\begin{aligned}
\sup_{g\in L^2_0(\pi)} J(g) &= J(g_0) \\
&= 2\langle \bar{f}, g_0\rangle_\pi - \langle g_0, (\mathrm{Id} - P)g_0\rangle_\pi \\
&= 2\langle \bar{f}, (\mathrm{Id} - P)^{-1}\bar{f}\rangle_\pi - \langle (\mathrm{Id} - P)^{-1}\bar{f}, \bar{f}\rangle_\pi \\
&= \langle \bar{f}, (\mathrm{Id} - P)^{-1}\bar{f}\rangle_\pi.
\end{aligned}$$

14. En utilisant l'expression $\mathcal{E}(f, P) = \sum_{x,y}(f(y) - f(x))^2\pi(x)P(x, y)$, il est clair que si $P \succ Q$, alors $\mathcal{E}(f, P) \geq \mathcal{E}(f, Q)$.

Remarque : Noter que la réciproque est fausse. On peut par exemple le vérifier sur le cas concret suivant. On prend $M = \{1, 2, 3\}$, π la mesure uniforme sur M, et on considère les matrices P et Q suivantes :

$$P = \begin{bmatrix} 0.55 & 0.13 & 0.32 \\ 0.13 & 0.65 & 0.22 \\ 0.32 & 0.22 & 0.46 \end{bmatrix} \qquad Q = \begin{bmatrix} 0.63 & 0.18 & 0.19 \\ 0.18 & 0.78 & 0.04 \\ 0.19 & 0.04 & 0.77 \end{bmatrix}.$$

Les matrices de transitions P et Q sont réversibles par rapport à π. On peut vérifier numériquement que le spectre de $Q-P$ est positif ce qui implique que pour tout $f : M \to \mathbb{R}$

$$\begin{aligned} 0 &\leq \langle f, (Q-P)f\rangle_\pi \\ &= \langle f, (I-P)f\rangle - \langle f, (I-Q)f\rangle_\pi \\ &= \frac{1}{2}\left(\mathcal{E}(f,P) - \mathcal{E}(f,Q)\right). \end{aligned}$$

Et pourtant $P(1,2) < Q(1,2)$, ce qui signifie que la relation $P \succ Q$ n'est pas vérifiée.

15. On utilise la formule donnée à la Question 13 pour la variance asymptotique. Si $P \succ Q$, on a que pour tout f et tout g, $4\langle \bar{f}, g\rangle_\pi - 2\mathcal{E}(g,P) - \|\bar{f}\|^2_{L^2(\pi)} \leq 4\langle \bar{f}, g\rangle_\pi - 2\mathcal{E}(g,Q) - \|\bar{f}\|^2_{L^2(\pi)}$. Par conséquent, pour tout $g \in L^2_0(\pi)$,

$$4\langle \bar{f}, g\rangle_\pi - 2\mathcal{E}(g,P) - \|\bar{f}\|^2_{L^2(\pi)} \leq \sup_{g\in L^2_0(\pi)} 4\langle \bar{f}, g\rangle_\pi - 2\mathcal{E}(g,Q) - \|\bar{f}\|^2_{L^2(\pi)}$$

et donc, en prenant le supremum du membre de gauche :

$$\sup_{g\in L^2_0(\pi)} 4\langle \bar{f}, g\rangle_\pi - 2\mathcal{E}(g,P) - \|\bar{f}\|^2_{L^2(\pi)} \leq \sup_{g\in L^2_0(\pi)} 4\langle \bar{f}, g\rangle_\pi - 2\mathcal{E}(g,Q) - \|\bar{f}\|^2_{L^2(\pi)}$$

ce qui donne

$$v(f,P) \leq v(f,Q).$$

On peut comprendre intuitivement ce résultat en se disant que la chaîne associée à P a plus de chance de bouger de x à $y \neq x$ que Q : elle "bouge plus" et il est donc naturel que la variance asymptotique associée à P soit plus petite que la variance asymptotique associée à Q.

16. On calcule

$$\mathrm{Var}_\pi(f) = \|\bar{f}\|^2_{L^2(\pi)} = \sum_{i=2}^d |\langle \bar{f}, e_i\rangle_\pi|^2$$

et

$$\begin{aligned} \mathrm{Cov}^\pi(f(X_0), f(X_n)) &= \langle \bar{f}, P^n \bar{f}\rangle_\pi \\ &= \left\langle \sum_{i=2}^d \langle \bar{f}, e_i\rangle_\pi e_i, \sum_{j=2}^d \lambda_j^n \langle \bar{f}, e_j\rangle_\pi e_j \right\rangle_\pi \\ &= \sum_{i=2}^d \lambda_i^n |\langle \bar{f}, e_i\rangle_\pi|^2. \end{aligned}$$

On a donc

$$\begin{aligned} v(f,P) &= \sum_{i=2}^d |\langle \bar{f}, e_i\rangle_\pi|^2 \left(1 + 2\sum_{n\geq 1} \lambda_i^n\right) \\ &= \sum_{i=2}^d |\langle \bar{f}, e_i\rangle_\pi|^2 \left(1 + 2\left(\frac{1}{1-\lambda_i} - 1\right)\right) \\ &= \sum_{i=2}^d |\langle \bar{f}, e_i\rangle_\pi|^2 \frac{1+\lambda_i}{1-\lambda_i}. \end{aligned}$$

En utilisant la question précédente, on sait donc que pour tout fonction $f : M \to \mathbb{R}$,

$$\sum_{i=2}^{d} |\langle \bar{f}, e_i(P)\rangle_\pi|^2 \frac{1+\lambda_i(P)}{1-\lambda_i(P)} \leq \sum_{i=2}^{d} |\langle \bar{f}, e_i(Q)\rangle_\pi|^2 \frac{1+\lambda_i(Q)}{1-\lambda_i(Q)}.$$

On choisit alors $\bar{f} = e_2(P)$ et on obtient

$$\frac{1+\lambda_2(P)}{1-\lambda_2(P)} \leq \sum_{i=2}^{d} |\langle e_2(P), e_i(Q)\rangle_\pi|^2 \frac{1+\lambda_i(Q)}{1-\lambda_i(Q)} \leq \frac{1+\lambda_2(Q)}{1-\lambda_2(Q)}$$

en utilisant le fait que la fonction $\lambda \in (-1,1) \mapsto \frac{1+\lambda}{1-\lambda}$ est croissante et que $\sum_{i=2}^{d} |\langle e_2(P), e_i(Q)\rangle_\pi|^2 \leq |e_2(P)|^2 = 1$. Et donc, en utilisant à nouveau la croissance de cette fonction, on a $\lambda_2(P) \leq \lambda_2(Q)$.

Quatrième partie : retour sur l'algoritme de Metropolis Hastings.

17. Par la question 10, Q est irréductible et apériodique si et seulement si il existe $n_0 > 0$ tel que pour tout x et y, il existe un chemin $(x_0 = x, x_1, \ldots, x_{n_0} = y)$ de longueur n_0 reliant x à y et de probabilité strictement positive pour Q : pour tout $i \in \{0, \ldots, n_0 - 1\}$, $Q(x_i, x_{i+1}) > 0$. Comme la fonction a est strictement positive, on a aussi que pour tout $i \in \{0, \ldots, n_0 - 1\}$, $P(x_i, x_{i+1}) > 0$ et donc $\mathbb{P}(X_0 = x_0, \ldots, X_{n_0} = x_{n_0}) > 0$ ce qui implique que P est irréductible et apériodique. Les conséquences sur le spectre de P sont directement obtenues par la question 10.
18. On note P_{MH} le noyau de transition associé au choix a_{MH} pour la fonction d'acceptation. On considère un autre choix pour la fonction d'acceptation a, qui rende la chaîne de Markov réversible par rapport à π, et on note P le noyau de transition associé. On sait qu'il faut que pour tout $x \neq y$,
$$a(x,y) = a(y,x)r(x,y).$$
Donc, comme pour tout $(x,y) \in M^2$, $a(x,y) \leq 1$, on a $a(x,y) \leq r(x,y)$ et $a(x,y) \leq 1$ donc
$$a(x,y) \leq \min(1, r(x,y)) = a_{\mathrm{MH}}(x,y).$$
On en déduit que pour tout $x \neq y$, $P(x,y) = Q(x,y)a(x,y) \leq Q(x,y)a_{\mathrm{MH}}(x,y) = P_{\mathrm{MH}}(x,y)$ et donc $P_{\mathrm{MH}} \succ P$. Par la question 15, ceci implique que pour toute fonction $f : M \to \mathbb{R}$, $v(f, P_{\mathrm{MH}}) \leq v(f,P)$: la variance asymptotique pour le choix a_{MH} est donc plus petite que la variance asymptotique pour toute autre fonction d'acceptation a.