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

6 janvier 2012, 09h00 - 12h00.

Les notes de cours sont autorisées.

Exercice 1 : Grandes déviations et échantillonnage d'importance

Première partie : grandes déviations

1 La loi forte des grands nombres dit que $\overline{X}_n/n$ converge vers $\mathbb{E}(X_1)$ presque sûrement, dans la limite $n \to \infty$. Par conséquent, par le théorème de convergence dominé, pour $x > \mathbb{E}(X_1)$ fixé, $\mathbb{P}(\overline{X}_n/n \geq x) = \mathbb{E}\left(1_{\overline{X}_n/n \geq x}\right)$ converge vers 0, dans la limite $n \to \infty$.

Le théorème central limite montre que $\sqrt{n}\left(\overline{X}_n/n - \mathbb{E}(X_1)\right)$ converge en loi vers une gaussienne centrée de variance $\sigma^2 = \mathrm{Var}(X_1)$. Par conséquent,

$$\begin{aligned} p_n(\mathbb{E}(X_1) + y/\sqrt{n}) &= \mathbb{P}\left(\overline{X}_n/n \geq \mathbb{E}(X_1) + y/\sqrt{n}\right) \\ &= \mathbb{P}\left(\sqrt{n}\left(\overline{X}_n/n - \mathbb{E}(X_1)\right) \geq y\right) \end{aligned}$$

et donc

$$\lim_{n\to\infty} p_n(\mathbb{E}(X_1) + y/\sqrt{n}) = \int_y^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-z^2/(2\sigma^2))\, dz.$$

2 La variable aléatoire $\overline{G}_n = \sum_{i=1}^n G_i$ est gaussienne centrée et de variance n. On a donc pour $x > 0$,

$$\begin{aligned} p_n(x) &= \mathbb{P}(\overline{G}_n/n \geq x) \\ &= \int_{xn}^\infty \frac{1}{\sqrt{2\pi n}} \exp(-y^2/(2n))\, dy \\ &= \int_{x\sqrt{n}}^\infty \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\, dz. \end{aligned}$$

Dans la limite $n \to \infty$, il faut donc estimer la queue d'une densité gaussienne. On a classiquement, pour $x > 0$,

$$\begin{aligned} \int_x^\infty \exp(-z^2/2)\, dz &\leq \int_x^\infty \frac{z}{x} \exp(-z^2/2)\, dz \\ &= \frac{1}{x} \int_x^\infty z \exp(-z^2/2)\, dz \\ &= \frac{\exp(-x^2/2)}{x}. \end{aligned}$$

Par ailleurs

$$\begin{aligned}
\int_x^{\infty} \exp(-z^2/2)\,dz &\geq \int_x^{2x} \exp(-z^2/2)\,dz \\
&\geq \int_x^{2x} \frac{z}{2x} \exp(-z^2/2)\,dz \\
&= \frac{1}{2x} \int_x^{2x} z \exp(-z^2/2)\,dz \\
&= \frac{1}{2x} \left(\exp(-x^2/2) - \exp(-2x^2) \right).
\end{aligned}$$

La borne inférieure ci-dessus est assez grossière mais suffisante pour la suite. Une estimation plus précise peut être obtenue par intégration par parties :

$$\begin{aligned}
\int_x^{\infty} \exp(-z^2/2)\,dz &= \int_x^{\infty} \frac{1}{z} z \exp(-z^2/2)\,dz \\
&= -\int_x^{\infty} \frac{1}{z^2} \exp(-z^2/2)\,dz + \frac{1}{x}\exp(-x^2/2) \\
&= -\int_x^{\infty} \frac{1}{z^3} z \exp(-z^2/2)\,dz + \frac{1}{x}\exp(-x^2/2) \\
&= \int_x^{\infty} \frac{3}{z^4} \exp(-z^2/2)\,dz - \frac{1}{x^3}\exp(-x^2/2) + \frac{1}{x}\exp(-x^2/2) \\
&\geq \exp(-x^2/2) \left(\frac{1}{x} - \frac{1}{x^3} \right).
\end{aligned}$$

On a donc

$$\begin{aligned}
\frac{1}{n} \ln p_n(x) &\leq \frac{1}{n} \ln \left(\frac{\exp(-x^2 n/2)}{x\sqrt{n}\sqrt{2\pi}} \right) \\
&= \frac{1}{n} \left(-x^2 n/2 - \ln(x\sqrt{n}\sqrt{2\pi}) \right)
\end{aligned}$$

et le membre de droite tend vers $-x^2/2$ dans la limite $n \to \infty$. De même

$$\begin{aligned}
\frac{1}{n} \ln p_n(x) &\geq \frac{1}{n} \ln \left(\frac{1}{2x\sqrt{n}\sqrt{2\pi}} \left(\exp(-x^2 n/2) - \exp(-2x^2 n) \right) \right) \\
&= \frac{1}{n} \ln \left(\exp(-x^2 n/2) \frac{1}{2x\sqrt{n}\sqrt{2\pi}} \left(1 - \exp(-3x^2 n/2) \right) \right) \\
&= \frac{1}{n} \left(-x^2 n/2 - \ln\left(2x\sqrt{n}\sqrt{2\pi}\right) + \ln\left(1 - \exp(-3x^2 n/2)\right) \right)
\end{aligned}$$

et le membre de droite tend vers $-x^2/2$ dans la limite $n \to \infty$. On a donc

$$\lim_{n\to\infty} \frac{1}{n} \ln p_n(x) = -x^2/2.$$

3 On suppose que X_1 est une gaussienne centrée réduite. On a donc

$$\begin{aligned}
\mathbb{E}(\exp(\theta X_1)) &= \int_{\mathbb{R}} \exp(\theta x) \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\,dx \\
&= \exp(\theta^2/2) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp(-(x-\theta)^2/2)\,dx \\
&= \exp(\theta^2/2).
\end{aligned}$$

On a donc $\Gamma(\theta) = \ln \mathbb{E}(\exp(\theta X_1)) = \theta^2/2$, et $\Gamma^*(x) = \sup_{\theta\in\mathbb{R}}(\theta x - \theta^2/2) = x^2/2$.

Généralement, Γ^* est une fonction convexe comme suprémum de fonctions convexes. Par Jensen, on a

$$\mathbb{E}(\exp(\theta X_1)) \geq \exp\left(\mathbb{E}(\theta X_1)\right)$$

et donc $\Gamma(\theta) \geq \theta\mathbb{E}(X_1)$. Par conséquent, pour tout $\theta \in \mathbb{R}$, $\theta\mathbb{E}(X_1) - \Gamma(\theta) \leq 0$ et

$$\Gamma^*(\mathbb{E}(X_1)) = 0$$

(puisque 0 est atteint pour $\theta = 0$). Pour $x > \mathbb{E}(X_1)$ et pour $\theta \leq 0$, on a $\theta x - \Gamma(\theta) \leq \theta(x - \mathbb{E}(X_1)) \leq 0$, et par conséquent, pour $x > \mathbb{E}(X_1)$

$$\Gamma^*(x) = \sup_{\theta\in\mathbb{R}}(\theta x - \Gamma(\theta)) = \sup_{\theta>0}(\theta x - \Gamma(\theta))$$

puique $\Gamma^*(x) \geq 0$ (considérer $\theta = 0$). On en déduit que pour $x \geq y \geq \mathbb{E}(X_1)$,

$$\begin{aligned}\Gamma^*(x) &= \sup_{\theta>0}(\theta x - \Gamma(\theta))\\ &\geq \sup_{\theta>0}(\theta y - \Gamma(\theta))\\ &= \Gamma^*(y).\end{aligned}$$

On vérifie facilement toutes ces propriétés sur le cas de la gaussienne.

4 On a, pour $x > \mathbb{E}(X_1)$, et pour $\theta > 0$

$$\begin{aligned}\frac{1}{n}\ln p_n(x) &= \frac{1}{n}\ln\left(\mathbb{P}(\overline{X}_n/n \geq x)\right)\\ &= \frac{1}{n}\ln\left(\mathbb{E}\left(1_{\theta(\overline{X}_n/n - x)\geq 0}\right)\right)\\ &\leq \frac{1}{n}\ln\left(\mathbb{E}\left(\exp\left[\theta\left(\overline{X}_n/n - x\right)\right]\right)\right)\\ &= \frac{1}{n}\ln\left(\mathbb{E}\left(\exp\left[(\theta/n)\sum_{i=1}^n (X_i - x)\right]\right)\right)\\ &= \frac{1}{n}\ln\left(\mathbb{E}\left(\prod_{i=1}^n \exp\left[(\theta/n)\left(X_i - x\right)\right]\right)\right)\\ &= \frac{1}{n}\ln\left[\left(\mathbb{E}\left(\exp\left[(\theta/n)\left(X_1 - x\right)\right]\right)\right)^n\right]\\ &= \ln\left(\mathbb{E}\left(\exp\left[(\theta/n)\left(X_1 - x\right)\right]\right)\right)\\ &= \Gamma(\theta/n) - \theta x/n\\ &\leq \inf_{\theta>0}(\Gamma(\theta) - \theta x)\\ &= -\Gamma^*(x).\end{aligned}$$

Seconde partie : échantillonnage d'importance

5 On a

$$\begin{aligned}\mathbb{E}\left(1_{Z\geq x}\frac{p(Z)}{q(Z)}\right) &= \int_{\mathbb{R}} 1_{z\geq x}\frac{p(z)}{q(z)}q(z)\,dz \\ &= \int_{\mathbb{R}} 1_{z\geq x}p(z)\,dz \\ &= \mathbb{E}(1_{Y\geq x}) \\ &= \mathbb{P}(Y\geq x).\end{aligned}$$

La variance de l'estimateur avec échantillonnage d'importance est

$$\mathrm{Var}\left(1_{Z\geq x}\frac{p(Z)}{q(Z)}\right) = \mathbb{E}\left(1_{Z\geq x}\frac{p^2(Z)}{q^2(Z)}\right) - (\mathbb{P}(Y\geq x))^2.$$

On veut résoudre le problème :

$$\min_{q\text{ densité de probabilité}} \mathbb{E}\left(1_{Z\geq x}\frac{p^2(Z)}{q^2(Z)}\right).$$

Par Cauchy-Schwarz, on a, pour toute densité de probabilité q,

$$\left(\mathbb{E}\left(1_{Z\geq x}\frac{p(Z)}{q(Z)}\right)\right)^2 \leq \mathbb{E}\left(1_{Z\geq x}\frac{p^2(Z)}{q^2(Z)}\right),$$

et pour $q(z) = q_*(z) = \dfrac{1_{z\geq x}\,p(z)}{\int 1_{z\geq x}\,p(z)\,dz}$, on a ($Z_*$ étant distribué suivant $q_*(z)\,dz$),

$$\begin{aligned}\mathbb{E}\left(1_{Z_*\geq x}\frac{p^2(Z_*)}{q_*^2(Z_*)}\right) &= \int\left(1_{z\geq x}\frac{p^2(z)}{q_*(z)}\right)\,dz \\ &= \left(\int 1_{z\geq x}\,p(z)\,dz\right)\int 1_{z\geq x}p(z)\,dz \\ &= \left(\int 1_{z\geq x}\,p(z)\,dz\right)^2 = (\mathbb{P}(Y\geq x))^2.\end{aligned}$$

Donc, la variance minimum est nulle, et atteinte en $q_*(z) = \dfrac{1_{z\geq x}\,p(z)}{\int 1_{z\geq x}\,p(z)\,dz}$. Noter que c'est la loi de Y, conditionnellement au fait que $Y\geq x$. Bien sûr, ce résultat reste délicat à utiliser en pratique puisque la constante de normalisation $\int 1_{z\geq x}\,p(z)\,dz = \mathbb{P}(Y\geq x)$ est précisément la quantité qu'on veut calculer. Simuler la variable aléatoire Z n'est pas simple en pratique.

6 On sait que

$$\mathrm{Var}(I_K) = \frac{1}{K}\mathrm{Var}\left(1_{\overline{X}_n/n\geq x}\right)$$

et le théorème central limite dit que l'erreur est d'ordre $\sqrt{\mathrm{Var}(I_K)}$. L'erreur relative est donc de l'ordre

$$\frac{1}{\sqrt{K}}\sqrt{\mathrm{Var}\left(1_{\overline{X}_n/n\geq x}\right)}/p_n(x).$$

Par ailleurs,

$$\begin{aligned}\mathrm{Var}\left(1_{\overline{X}_n/n\geq x}\right) &= \mathbb{E}\left(1_{\overline{X}_n/n\geq x}\right) - \left(\mathbb{E}\left(1_{\overline{X}_n/n\geq x}\right)\right)^2 \\ &= p_n(x) - (p_n(x))^2.\end{aligned}$$

L'erreur relative est donc de l'ordre

$$\frac{1}{\sqrt{K}}\sqrt{\mathrm{Var}\left(1_{\overline{X}_n/n\geq x}\right)}/p_n(x) = \frac{1}{\sqrt{K}}\sqrt{1/p_n(x)-1}.$$

Dans la limite $n\to\infty$, on sait, par (2), que $p_n(x)$ se comporte comme $\exp(-n\Gamma^*(x))$, et donc l'erreur relative se comporte comme

$$\frac{\exp(n\Gamma^*(x)/2)}{\sqrt{K}}.$$

A erreur relative fixée, il faut donc que K augmente exponentiellement vite avec n.

7 En utilisant comme fonction test $\varphi = 1/L_n$, on a $\mathbb{E}(1/L_n(\overline{X}_n)) = \mathbb{E}((1/L_n)(\overline{Y}_n)L_n(\overline{Y}_n)) = 1$.

On a à nouveau

$$\mathrm{Var}\left(\tilde{I}_K\right) = \frac{1}{K}\mathrm{Var}\left(1_{\overline{Y}_n/n\geq x}L_n(\overline{Y}_n)\right),$$

et

$$\mathrm{Var}\left(1_{\overline{Y}_n/n\geq x}L_n(\overline{Y}_n)\right) = M_n(x) - (p_n(x))^2.$$

Par l'inégalité de Cauchy-Schwarz (ou bien le fait que $\mathrm{Var}\left(1_{\overline{Y}_n/n\geq x}L_n(\overline{Y}_n)\right)\geq 0$), on a

$$M_n(x)\geq (p_n(x))^2.$$

On voit donc que l'erreur relative s'écrit

$$\frac{1}{\sqrt{K}}\sqrt{\mathrm{Var}\left(1_{\overline{Y}_n/n\geq x}L_n(\overline{Y}_n)\right)}/p_n(x) = \frac{1}{\sqrt{K}}\sqrt{M_n(x)/(p_n(x))^2-1}.$$

Pour qu'à nombre de réalisations K fixé, l'erreur relative reste bornée, il faut donc que $M_n(x)$ se comporte comme $(p_n(x))^2$ dans la limite $n\to\infty$, c'est-à-dire décroisse à la vitesse $\exp(-2n\Gamma^*(x))$.

Plus précisément, l'erreur relative reste bornée si et seulement si le rapport $M_n(x)/(p_n(x))^2$ est majoré par une constante, et donc, si et seulement si, il existe une constante C telle que dans la limite $n\to\infty$,

$$\frac{1}{n}\ln(M_n(x)) \leq \frac{C}{n} - 2\Gamma^*(x).$$

8 On vérifie que

$$\begin{aligned}
\mathbb{E}(1/L_n(\overline{X}_n)) &= \mathbb{E}(\exp(\theta\overline{X}_n))\exp(-n\Gamma(\theta)) \\
&= \mathbb{E}\left(\exp\left(\theta\sum_{i=1}^{n}X_i\right)\right)\exp(-n\Gamma(\theta)) \\
&= \mathbb{E}\left(\prod_{i=1}^{n}\exp(\theta X_i)\right)\exp(-n\Gamma(\theta)) \\
&= \left(\mathbb{E}\left(\exp(\theta X_1)\right)\right)^n\exp(-n\Gamma(\theta)) \\
&= \exp(n\Gamma(\theta))\exp(-n\Gamma(\theta)) \\
&= 1.
\end{aligned}$$

On introduit les $(Y_i)_{i\geq 1}$ une suite i.i.d. de loi donnée par : pour toute fonction φ mesurable bornée,

$$\mathbb{E}(\varphi(Y_1)) = \frac{\mathbb{E}(\varphi(X_1)\exp(\theta X_1)}{\mathbb{E}(\exp(\theta X_1))}.$$

La loi de Y_1 est donc celle de X_1 multipliée par la fonction d'importance $\exp(\theta x_1)$ normalisée. Par indépendance, la loi de $(Y_1,\ldots,Y_n)$ est donc celle de $(X_1,\ldots,X_n)$ multipliée par la fonction d'importance $\prod_{i=1}^{n}\exp\left(\theta x_i\right)$ normalisée Pour cette suite $(Y_i)_{i\geq 1}$, on a, pour toute fonction φ mesurable bornée,

$$\begin{aligned}
\mathbb{E}\left(\varphi\left(\sum_{i=1}^{n}Y_i\right)\right) &= \mathbb{E}\left(\varphi\left(\sum_{i=1}^{n}X_i\right)\prod_{i=1}^{n}\exp(\theta X_i-\Gamma(\theta))\right) \\
&= \mathbb{E}\left(\varphi\left(\sum_{i=1}^{n}X_i\right)\exp\left(\theta\sum_{i=1}^{n}X_i-n\Gamma(\theta)\right)\right) \\
&= \mathbb{E}\left(\varphi\left(\sum_{i=1}^{n}X_i\right)1/L_n\left(\sum_{i=1}^{n}X_i\right)\right)
\end{aligned}$$

et donc, pour toute fonction φ mesurable bornée,

$$\mathbb{E}\left(\varphi\left(\sum_{i=1}^{n}Y_i\right)L_n\left(\sum_{i=1}^{n}Y_i\right)\right) = \mathbb{E}\left(\varphi(\overline{X}_n)\right)$$

ce qui montre que $\overline{Y}_n$ a même loi que $\sum_{i=1}^{n}Y_i$.

De plus, on a

$$\begin{aligned}
M_n(x) &= \mathbb{E}\left(1_{\overline{Y}_n/n\geq x}\,L_n^2(\overline{Y}_n)\right) \\
&= \mathbb{E}\left(1_{\overline{Y}_n/n\geq x}\,\exp\left(-2\theta\overline{Y}_n+2n\Gamma(\theta)\right)\right) \\
&\leq \mathbb{E}\left(1_{\overline{Y}_n/n\geq x}\,\exp\left(-2\theta nx+2n\Gamma(\theta)\right)\right) \\
&\leq \mathbb{E}\left(\exp\left(-2\theta nx+2n\Gamma(\theta)\right)\right) \\
&= \exp\left(-2n(\theta x-\Gamma(\theta))\right).
\end{aligned}$$

En choisissant $\theta = \theta^*$ qui réalise le suprémum dans

$$\Gamma^*(x) = \sup_{\theta \in \mathbb{R}} (\theta x - \Gamma(\theta))$$

on a

$$M_n(x) \leq \exp(-2n\Gamma^*(x)).$$

D'après la question précédente, en choisissant $\theta = \theta^*$, on obtient donc un estimateur qui réalise une erreur relative bornée dans la limite $n \to \infty$ (à nombre de réalisations K fixé).

9 Bien sûr, ces résultats restent théoriques, puisqu'il faut savoir simuler $\overline{Y}_n$ en pratique, et cela n'est pas simple en général. Comme vu dans la question précédente, $\overline{Y}_n$ a même loi que $\sum_{i=1}^n Y_i$ avec $(Y_i)_{i \geq 1}$ i.i.d. de loi définie par : pour toute fonction φ mesurable bornée,

$$\mathbb{E}(\varphi(Y_1)) = \frac{\mathbb{E}(\varphi(X_1) \exp(\theta^*(X_1)))}{\mathbb{E}(\exp(\theta^*(X_1)))}.$$

La question est donc de savoir comment simuler les $(Y_i)_{i \geq 1}$.

Dans le cas gaussien (X_1 est une normale centrée réduite), la formule ci-dessus s'écrit : pour toute fonction φ mesurable bornée,

$$\begin{aligned}
\mathbb{E}(\varphi(Y_1)) &= \frac{\mathbb{E}(\varphi(X_1) \exp(\theta^*(X_1)))}{\mathbb{E}(\exp(\theta^*(X_1)))} \\
&= \frac{\int_{\mathbb{R}} \varphi(z) \exp(\theta^* z) \exp(-z^2/2)}{\int_{\mathbb{R}} \exp(\theta^* z) \exp(-z^2/2)} \\
&= \frac{\int_{\mathbb{R}} \varphi(z) \exp(-(z-\theta^*)^2/2)}{\int_{\mathbb{R}} \exp(-(z-\theta^*)^2/2)}.
\end{aligned}$$

Autrement dit, les $(Y_i)_{i \geq 1}$ sont i.i.d. de loi gaussienne centrée en θ^* et de variance 1. Par conséquent $\overline{Y}_n$ est de loi gaussienne centrée en $n\theta^*$ et de variance n. De plus, dans le cas gaussien, le θ optimal est analytique : $\theta^* = x$. La variable aléatoire $\overline{Y}_n$ est donc facilement simulable dans le cas gaussien.

Pour en savoir plus sur le sujet, on pourra parcourir le livre : J.A. Bucklew, Introduction to rare event simulation*, Springer, 2004.*

Exercice 2 : Dynamique *overdamped Langevin* et distribution quasi-stationnaire

1 Puisque u est supposée régulière, on peut différencier $s \mapsto u(s, X_s^x)$, où $s \in [0,t]$, en utilisant le calcul d'Itô. On obtient

$$\begin{aligned} du(s, X_s^x) &= (\partial_s u + Lu)(s, X_s^x)\, ds + \sqrt{2\beta^{-1}} \nabla u(s, X_s^x) \cdot dB_s \\ &= \sqrt{2\beta^{-1}} \nabla u(s, X_s^x) \cdot dB_s. \end{aligned}$$

On a donc, en intégrant entre 0 et $t \wedge T_W^x = \inf(t, T_W^x)$, presque sûrement,

$$u(t \wedge T_W^x, X_{t\wedge T_W^x}^x) - u(0, X_0^x) = \sqrt{2\beta^{-1}} \int_0^{t\wedge T_W^x} \nabla u(s, X_s^x) \cdot dB_s.$$

La fonction ∇u est régulière donc bornée sur $[0,t] \times W$, et le temps d'arrêt $t \wedge T_W^x$ est borné : on en déduit que l'espérance de l'intégrale stochastique est nulle. On a donc

$$\mathbb{E}\left(u(0, X_0^x)\right) = \mathbb{E}\left(u(t \wedge T_W^x, X_{t\wedge T_W^x}^x)\right),$$

soit

$$\begin{aligned} u(0,x) &= \mathbb{E}\left(u(t \wedge T_W^x, X_{t\wedge T_W^x}^x)\left(1_{T_W^x < t} + 1_{T_W^x \geq t}\right)\right) \\ &= \mathbb{E}\left(u(T_W^x, X_{T_W^x}^x) 1_{T_W^x < t}\right) + \mathbb{E}\left(u(t, X_t^x) 1_{T_W^x \geq t}\right) \\ &= \mathbb{E}\left(\varphi(X_{T_W^x}^x) 1_{T_W^x < t}\right) + \mathbb{E}\left(v_0(X_t^x) 1_{T_W^x \geq t}\right), \end{aligned}$$

ce qui démontre (6).

Soit v solution régulière de l'EDP (7). Pour $t > 0$ fixé, on pose $u(s,x) = v(t-s,x)$. On vérifie facilement que u est solution de l'EDP (5). Et donc

$$v(t,x) = u(0,x) = \mathbb{E}\left(\varphi(X_{T_W^x}^x) 1_{T_W^x < t}\right) + \mathbb{E}\left(v_0(X_t^x) 1_{T_W^x \geq t}\right).$$

Noter que t est arbitraire, donc cette relation est en fait valable pour tout $t > 0$.

2.1 On a

$$\begin{aligned} \int_W f L g \, d\mu &= \int_W f(-\nabla V \cdot \nabla g + \beta^{-1} \Delta g)\, Z^{-1} \exp(-\beta V)\, dx \\ &= -Z^{-1} \int_W f \nabla V \cdot \nabla g \exp(-\beta V)\, dx - \beta^{-1} Z^{-1} \int_W \nabla g \cdot \nabla\left(f \exp(-\beta V)\right) dx \\ &= -\beta^{-1} Z^{-1} \int_W \nabla g \cdot \nabla f \exp(-\beta V)\, dx \\ &= -\beta^{-1} \int_W \nabla f \cdot \nabla g \, d\mu. \end{aligned}$$

Cette expression est symétrique en f et g ce qui démontre par ailleurs que $\int_W f L g \, d\mu = \int_W g L f \, d\mu$.

2.2 En utilisant le fait que W est un ouvert borné, que V est minoré et majoré sur W, on a que l'application $f \in H_0^1(W) \mapsto f \in H_{\mu,0}^1(W)$ est continue, ainsi que son inverse. En utilisant l'inégalité de Poincaré sur $H_0^1(W)$, on a donc

$$\begin{aligned}\int_W |\nabla f|^2 d\mu &= Z^{-1} \int_W |\nabla f|^2 \exp(-\beta V) \\ &\geq C_1 \int_W |\nabla f|^2 \\ &\geq C_2 \int_W |f|^2 \\ &\geq C_3 \int_W |f|^2 d\mu\end{aligned}$$

et donc $\lambda_1 \geq \beta^{-1} C_2 > 0$.

Soit maintenant une suite minisante $f_n \in H_{\mu,0}^1(W)$, telle que

$$\lim_{n\to\infty} \frac{\beta^{-1}\int_W |\nabla f_n|^2 \, d\mu}{\int_W (f_n)^2 \, d\mu} = \lambda_1.$$

Quitte à changer f_n en $f_n / \left(\int_W f_n^2 \, d\mu\right)^{1/2}$, on peut supposer que $\int_W f_n^2 \, d\mu = 1$. La suite f_n est bornée dans $H_0^1(W)$, donc (quitte à extraire une sous-suite) elle converge faiblement dans $H_0^1(W)$, et fortement dans $L^2(W)$ vers une fonction $u_1 \in H_0^1(W)$. L'application $f \in H_0^1(W) \mapsto f \in H_{\mu,0}^1(W)$ étant (linéaire et) continue, f_n converge aussi faiblement dans $H_{\mu,0}^1(W)$, et fortement dans $L_\mu^2(W)$ vers la même fonction $u_1 \in H_{\mu,0}^1(W)$. On a donc $\int_W (u_1)^2 \, d\mu = 1$ et

$$\int_W |\nabla u_1|^2 \, d\mu \leq \liminf_{n\to\infty} \int_W |\nabla f_n|^2 \, d\mu.$$

On a donc

$$\frac{\beta^{-1}\int_W |\nabla u_1|^2 \, d\mu}{\int_W (u_1)^2 \, d\mu} \leq \lambda_1,$$

mais comme λ_1 est l'infimum sur toutes les fonctions $f \in H_{\mu,0}^1$ du rapport $\frac{\beta^{-1}\int_W |\nabla f|^2 \, d\mu}{\int_W f^2 \, d\mu}$, on a donc

$$\frac{\beta^{-1}\int_W |\nabla u_1|^2 \, d\mu}{\int_W (u_1)^2 \, d\mu} = \lambda_1.$$

2.3 La fonction u_1 est dans $H_0^1(W)$ et donc $|u_1|$ est dans $H_0^1(W)$ et telle que $\nabla |u_1| = \mathrm{sgn}(u_1) \nabla u_1$, où $\mathrm{sgn}(x) = 1_{x>0} - 1_{x\geq 0}$ est la fonction signe. En particulier, on a

$$\frac{\beta^{-1}\int_W |\nabla |u_1||^2 \, d\mu}{\int_W |u_1|^2 \, d\mu} = \frac{\beta^{-1}\int_W |\nabla u_1|^2 \, d\mu}{\int_W (u_1)^2 \, d\mu} = \lambda_1.$$

On peut donc supposer $u_1 \geq 0$ quitte à remplacer u_1 par $|u_1|$.

On écrit les équations d'Euler associée au problème de minisation. On sait que u_1 est tel que pour toute fonction $v \in H^1_{\mu,0}$, pour tout ϵ,

$$\frac{\int_W |\nabla(u_1+\epsilon v)|^2 \, d\mu}{\int_W (u_1+\epsilon v)^2 \, d\mu} \geq \int_W |\nabla u_1|^2 \, d\mu.$$

On a utilisé le fait que $\int_W (u_1)^2 \, d\mu = 1$. On a donc

$$\int_W |\nabla(u_1+\epsilon v)|^2 \, d\mu \geq \int_W |\nabla u_1|^2 \, d\mu \int_W (u_1+\epsilon v)^2 \, d\mu,$$

soit, en développant :

$$\begin{aligned}\int_W |\nabla u_1|^2 \, d\mu &+ 2\epsilon \int_W \nabla u_1 \cdot \nabla v \, d\mu + \epsilon^2 \int_W |\nabla v|^2 \, d\mu \\ &\geq \int_W |\nabla u_1|^2 \, d\mu \left(1 + 2\epsilon \int_W u_1 v \, d\mu + \epsilon^2 \int_W v^2 \, d\mu\right).\end{aligned}$$

Puisque $\int_W |\nabla u_1|^2 \, d\mu = \beta\lambda_1$, en considérant la limite $\epsilon \to 0$ (et en changeant v en $-v$), on a donc, pour toute fonction $v \in H^1_{\mu,0}$,

$$\int_W \nabla u_1 \cdot \nabla v \, d\mu = \beta\lambda_1 \int_W u_1 v \, d\mu.$$

En utilisant le résultat de la question 2.1, on a donc $-Lu_1 = \lambda_1 u_1$ (au sens faible), et $u_1 = 0$ au bord ∂W (puisque $u_1 \in H^1_{\mu,0}$). En d'autres termes, u_1 est une fonction propre de l'opérateur L avec conditions aux limites de Dirichlet, associée à la valeur propre $-\lambda_1$.

3 On introduit $v(t,x) = \mathbb{E}\left(f(X_t^x) 1_{t \leq T_W^x}\right)$ et $\bar{v}(t,x) = \mathbb{P}(t \leq T_W^x) = \mathbb{E}\left(1_{t \leq T_W^x}\right)$. En utilisant le résultat de la question 1, on a que v et $\bar{v}$ satisfont :

$$\begin{cases} \partial_t v(t,x) = Lv(t,x) \text{ pour } t \geq 0 \text{ et } x \in W, \\ v(t,x) = 0 \text{ pour } t \geq 0 \text{ et } x \in \partial W, \\ v(0,x) = f(x) \text{ pour } x \in W, \end{cases}$$

et

$$\begin{cases} \partial_t \bar{v}(t,x) = L\bar{v}(t,x) \text{ pour } t \geq 0 \text{ et } x \in W, \\ \bar{v}(t,x) = 0 \text{ pour } t \geq 0 \text{ et } x \in \partial W, \\ \bar{v}(0,x) = 1 \text{ pour } x \in W. \end{cases}$$

On a donc (en utilisant la question 2.1)

$$\begin{aligned}\frac{d}{dt} \int_W v(t,x) \, \nu(dx) &= \int_W \partial_t v(t,x) \, \nu(dx) \\ &= \frac{\int_W (Lv) u_1 \, d\mu}{\int_W u_1 \, d\mu} \\ &= \frac{\int_W v L u_1 \, d\mu}{\int_W u_1 \, d\mu} \\ &= -\lambda_1 \frac{\int_W v u_1 \, d\mu}{\int_W u_1 \, d\mu} \\ &= -\lambda_1 \int_W v \, d\nu.\end{aligned}$$

On en déduit que (en utilisant le fait que $v(0,x) = f(x)$),

$$\int_W v\, d\nu = \int_W f\, d\nu \exp(-\lambda_1 t).$$

De même,

$$\int_W \bar{v}\, d\nu = \exp(-\lambda_1 t).$$

On en déduit que

$$\int_W v\, d\nu = \int_W f\, d\nu \int_W \bar{v}\, d\nu,$$

soit

$$\int_W \mathbb{E}\left(f(X_t^x) 1_{t \le T_W^x}\right) \nu(dx) = \int_W f\, d\nu \int_W \mathbb{P}\left(t \le T_W^x\right) \nu(dx).$$

Si on suppose que X_0 est distribué suivant ν, on a donc, pour toute fonction f,

$$\frac{\mathbb{E}\left(f(X_t) 1_{t \le T_W}\right)}{\mathbb{P}\left(t \le T_W\right)} = \int_W f\, d\nu,$$

soit

$$\mathbb{E}\left(f(X_t) \,|\, t \le T_W\right) = \int_W f\, d\nu.$$

Autrement dit, ν est une mesure de probabilité telle que, si X_0 est distribué suivant ν, après un temps $t > 0$ et conditionnellement au fait qu'on n'a pas quitté le domaine W durant l'intervalle de temps $[0,t]$, X_t est aussi distribué suivant ν. C'est une première manière de voir la distribution quasi-stationnaire.

4 On vérifie facilement que ν a pour densité q par rapport à la mesure de Lebesgue ($d\nu = q(x)\, dx$) avec

$$q(x) = \frac{u_1(x) \exp(-\beta V(x))}{\int_W u_1(x) \exp(-\beta V(x))\, dx}.$$

On note dans la suite $C_q = \int_W u_1(x) \exp(-\beta V(x))\, dx$.

On introduit

$$\begin{aligned} g(t) &= \int_W \mathbb{E}\left(1_{T_W^x < t}\, \varphi(X_{T_W}^x)\right) \nu(dx) \\ &= \int_W v(t,x) q(x)\, dx, \end{aligned}$$

où $v(t,x) = \mathbb{E}\left(1_{T_W^x < t}\, \varphi(X_{T_W}^x)\right)$. En utilisant la question 1, on sait que v satisfait :

$$\begin{cases} \partial_t v(t,x) = L v(t,x) \text{ pour } t \ge 0 \text{ et } x \in W, \\ \quad v(t,x) = \varphi(x) \text{ pour } t \ge 0 \text{ et } x \in \partial W, \\ \quad v(0,x) = 0 \text{ pour } x \in W. \end{cases}$$

On a donc, par intégrations par parties, en utilisant le fait que $q = 0$ sur ∂W,

$$\begin{aligned}
g'(t) &= \int_W \partial_t v(t,x) q(x)\, dx \\
&= \int_W Lv(t,x) q(x)\, dx \\
&= \int_W (-\nabla V \cdot \nabla v + \beta^{-1} \Delta v) q \\
&= \int_W (-\nabla V \cdot \nabla v) q - \beta^{-1} \int_W \nabla v \cdot \nabla q \\
&= C_q^{-1} \int_W (-\nabla V \cdot \nabla v) u_1 \exp(-\beta V) - \beta^{-1} C_q^{-1} \int_W \nabla v \cdot \nabla (u_1 \exp(-\beta V)) \\
&= -\beta^{-1} C_q^{-1} \int_W (\nabla v \cdot \nabla u_1) \exp(-\beta V) \\
&= \beta^{-1} C_q^{-1} \int_W v \operatorname{div}(\nabla u_1 \exp(-\beta V)) - \beta^{-1} C_q^{-1} \int_{\partial W} v (\partial_n u_1) \exp(-\beta V) \\
&= C_q^{-1} \int_W v (\beta^{-1} \Delta u_1 - \nabla V \cdot \nabla u_1) \exp(-\beta V) - \beta^{-1} C_q^{-1} \int_{\partial W} \varphi (\partial_n u_1) \exp(-\beta V) \\
&= -\lambda_1 C_q^{-1} \int_W v u_1 \exp(-\beta V) - \beta^{-1} C_q^{-1} \int_{\partial W} \varphi (\partial_n u_1) \exp(-\beta V) \\
&= -\lambda_1 g(t) - \beta^{-1} C_q^{-1} \int_{\partial W} \varphi (\partial_n u_1) \exp(-\beta V).
\end{aligned}$$

On remarque que, sur ∂W,

$$\begin{aligned}
\partial_n q &= C_q^{-1} \partial_n (u_1 \exp(-\beta V)) \\
&= C_q^{-1} \left(\partial_n u_1 - u_1 \partial_n V\right) \exp(-\beta V)) \\
&= C_q^{-1} \partial_n u_1 \exp(-\beta V),
\end{aligned}$$

puisque $u_1 = 0$ sur ∂W. On a donc finalement

$$g'(t) = -\lambda_1 g(t) - \beta^{-1} \int_{\partial W} \varphi\, \partial_n q.$$

On en déduit que

$$\frac{d}{dt} (\exp(\lambda_1 t) g(t)) = -\beta^{-1} \exp(\lambda_1 t) \int_{\partial W} \varphi\, \partial_n q,$$

et donc, puisque $g(0) = 0$,

$$g(t) = -\frac{\beta^{-1}}{\lambda_1} (1 - \exp(-\lambda_1 t)) \int_{\partial W} \varphi\, \partial_n q.$$

5 On considère tout d'abord $\varphi = 1$, de telle sorte que $g(t) = \mathbb{P}(t > T_W^x)$, et on a

$$g(t) = -\frac{\beta^{-1}}{\lambda_1} (1 - \exp(-\lambda_1 t)) \int_{\partial W} \partial_n q.$$

En faisant tendre $t \to \infty$, et puisqu'on sait que $g(t)$ tend vers 1, on a donc

$$-\frac{\beta^{-1}}{\lambda_1}\int_{\partial W} \partial_n q = 1.$$

On en déduit que $\mathbb{P}(t > T_W^x) = 1 - \exp(-\lambda_1 t)$ et que T_W est une variable aléatoire exponentielle de paramètre λ_1.

Ensuite, pour une fonction φ quelconque, on a :

$$\begin{aligned}\mathbb{E}(1_{T_W < t}\, \varphi(X_{T_W})) = g(t) &= (1 - \exp(-\lambda_1 t))\frac{\int_{\partial W} \varphi\, \partial_n q}{\int_{\partial W} \partial_n q} \\ &= \mathbb{P}(T_W < t)\frac{\int_{\partial W} \varphi\, \partial_n q}{\int_{\partial W} \partial_n q}.\end{aligned}$$

On en déduit que T_W et X_{T_W} sont indépendants, avec T_W de loi exponentielle de paramètre λ_1 et X_{T_W} de densité $\frac{\partial_n q}{\int_{\partial W} \partial_n q}$ par rapport à la mesure de Lebesgue sur ∂W. C'est une deuxième propriété fondamentale de la distribution quasi-stationnaire : en démarrant sous cette distribution, le temps de sortie est exponentiellement distribué et indépendant du point de sortie.

6 On considère $v(t,x) = \mathbb{E}\left(v_0(X_t^x) 1_{t \le T_W^x}\right)$. On sait que v est solution de

$$\begin{cases} \partial_t v(t,x) = Lv(t,x) \text{ pour } t \ge 0 \text{ et } x \in W, \\ v(t,x) = 0 \text{ pour } t \ge 0 \text{ et } x \in \partial W, \\ v(0,x) = v_0(x) \text{ pour } x \in W. \end{cases}$$

En utilisant la décomposition spectrale de l'opérateur L, on a donc

$$v(t,x) = \sum_{k \ge 1} (v_0, u_k)_\mu \exp(-\lambda_k t) u_k(x),$$

où $(f,g)_\mu = \int_W fg\, d\mu$ désigne le produit scalaire dans L^2_μ. Cette égalité a lieu, par exemple, dans L^2_μ.

7 Comme précédemment, on a

$$h(t) = \frac{\mathbb{E}\left(f(X_t) 1_{t \le T_W}\right)}{\mathbb{P}(t \le T_W)} = \frac{\int_W \mathbb{E}\left(f(X_t^x) 1_{t \le T_W^x}\right)\, \mu_0(dx)}{\int_W \mathbb{P}\left(t \le T_W^x\right)\, \mu_0(dx)} = \frac{\int_W v(t,x)\, \mu_0(dx)}{\int_W \bar{v}(t,x)\, \mu_0(dx)}.$$

où $v(t,x) = \mathbb{E}\left(f(X_t^x) 1_{t \le T_W^x}\right)$ et $\bar{v}(t,x) = \mathbb{P}\left(t \le T_W^x\right)$. D'après la question précédente, on a donc

$$\int_W v(t,x)\, \mu_0(dx) = \sum_{k \ge 1} (f, u_k)_\mu \exp(-\lambda_k t) \int_W u_k \mu_0(dx)$$

et

$$\int_W \bar{v}(t,x)\, \mu_0(dx) = \sum_{k \ge 1} (1, u_k)_\mu \exp(-\lambda_k t) \int_W u_k \mu_0(dx).$$

Noter que $\int_W u_k \mu_0(dx) = \int_W u_k q_0 \, d\mu < \infty$ puisque q_0 et u_k sont dans L^2_μ. Les séries sont convergentes puisque $(u_k)_{k\geq 1}$ est une base orthonormée de L^2_μ et donc

$$\begin{aligned}\left|\sum_{k\geq 1}(f,u_k)_\mu \exp(-\lambda_k t)\int_W u_k \mu_0(dx)\right| &\leq \sqrt{\sum_{k\geq 1}|(f,u_k)_\mu|^2}\sqrt{\sum_{k\geq 1}|(q_0,u_k)_\mu|^2} \\ &= \|f\|_\mu \|q_0\|_\mu.\end{aligned}$$

On a donc

$$\begin{aligned}h(t) &= \frac{\sum_{k\geq 1}(f,u_k)_\mu (q_0,u_k)_\mu \exp(-\lambda_k t)}{\sum_{k\geq 1}(1,u_k)_\mu (q_0,u_k)_\mu \exp(-\lambda_k t)} \\ &= \frac{\frac{(f,u_1)_\mu}{(1,u_1)_\mu} + \sum_{k\geq 2}\frac{(f,u_k)_\mu (q_0,u_k)_\mu}{(1,u_1)_\mu (q_0,u_1)_\mu}\exp(-(\lambda_k-\lambda_1)t)}{1+\sum_{k\geq 2}\frac{(1,u_k)_\mu (q_0,u_k)_\mu}{(1,u_1)_\mu (q_0,u_1)_\mu}\exp(-(\lambda_k-\lambda_1)t)}.\end{aligned}$$

Noter que

$$\begin{aligned}\left|\sum_{k\geq 2}\frac{(f,u_k)_\mu (q_0,u_k)_\mu}{(1,u_1)_\mu (q_0,u_1)_\mu}\exp(-(\lambda_k-\lambda_1)t)\right| &\leq C\exp(-(\lambda_2-\lambda_1)t)\sum_{k\geq 2}|(f,u_k)_\mu||(q_0,u_k)_\mu| \\ &\leq C\|f\|_\mu \|q_0\|_\mu \exp(-(\lambda_2-\lambda_1)t)\end{aligned}$$

et donc cette quantité tend vers 0 quand $t \to \infty$. Le même raisonnement permet de montrer que $\lim_{t\to\infty}\sum_{k\geq 2}\frac{(1,u_k)_\mu (q_0,u_k)_\mu}{(1,u_1)_\mu (q_0,u_1)_\mu}\exp(-(\lambda_k-\lambda_1)t) = 0$. On en déduit que

$$\lim_{t\to\infty} h(t) = \frac{(f,u_1)_\mu}{(1,u_1)_\mu} = \frac{\int_W f u_1 \, d\mu}{\int_W u_1 \, d\mu} = \int_W f \, d\nu.$$

8 On a donc montré que la loi de X_t, conditionnellement au fait que le processus reste dans W, converge en temps long vers la distribution quasi-stationnaire ν. C'est une troisième propriété fondamentale de la distribution quasi-stationnaire.

On a vu à la question 5 que si on démarre sous ν, le temps de sortie est exponentiel et indépendant du point de sortie. Ceci montre donc de manière informelle, que si le processus X_t reste "très longtemps" dans un état W, on peut supposer que le temps qu'il reste à y passer est exponentiel, et que le prochain état visité (qui dépend du point de sortie) est indépendant de ce temps. Ceci peut permettre de formaliser le lien entre une dynamique *overdamped Langevin* et des modèles de type *kinetic Monte Carlo* (chaîne de Markov continue en temps, et à espace d'états discret). Pour une formalisation rigoureuse, on renvoie notamment à l'article en référence à la fin de la correction.

9 On suppose que le processus X_t démarre sous la probabilité μ_0 à support dans W, et on considère le temps de sortie T_W. Sa loi est caractérisée par

$$k(t) = \mathbb{P}(T_W \geq t) = \int_W \mathbb{P}(T_W^x \geq t)\, \mu_0(dx) = \int_W \bar{v}(t,x)\, \mu_0(dx),$$

avec $\bar{v}$ solution de

$$\begin{cases} \partial_t \bar{v}(t,x) = L\bar{v}(t,x) \text{ pour } t \geq 0 \text{ et } x \in W, \\ \bar{v}(t,x) = 0 \text{ pour } t \geq 0 \text{ et } x \in \partial W, \\ \bar{v}(0,x) = 1 \text{ pour } x \in W. \end{cases}$$

Comme dans la question 7, on a :

$$k(t) = \sum_{k\geq 1} (1, u_k)_\mu (q_0, u_k)_\mu \exp(-\lambda_k t).$$

On suppose que T_W est exponentiellement distribué ce qui est équivalent au fait qu'il existe $\lambda > 0$ tel que $k(t) = \exp(-\lambda t)$. On a donc l'égalité :

$$\exp(-\lambda t) = \sum_{k\geq 1} (1, u_k)_\mu (q_0, u_k)_\mu \exp(-\lambda_k t).$$

On voit donc que si les valeurs propres sont non dégénérées (hypothèse (i) : $\forall i \neq j, \lambda_i \neq \lambda_j$) et si les modes propres sont de moyenne non nulle (hypothèse (ii) : $\int_W u_k d\mu \neq 0$), alors, nécessairement, $\mu_0 = \nu$. En effet, on a, dans la limite $t \to \infty$,

$$\sum_{k\geq 1} (1, u_k)_\mu (q_0, u_k)_\mu \exp(-\lambda_k t) \sim (1, u_1)_\mu (q_0, u_1)_\mu \exp(-\lambda_1 t).$$

Donc, nécessairement, $\lambda_1 = \lambda$ et $(1, u_1)_\mu (q_0, u_1)_\mu = 1$. On en déduit que

$$\sum_{k\geq 2} (1, u_k)_\mu (q_0, u_k)_\mu \exp(-\lambda_k t) = 0.$$

En utilisant l'hypothèses (i), on en déduit $(1, u_k)_\mu (q_0, u_k)_\mu = 0$ et, grâce à l'hypothèse (ii), $(q_0, u_k)_\mu = 0$. Ceci montre bien que $\mu_0 = \nu$, puisque

$$q_0(x) = \sum_{k\geq 1} (q_0, u_k)_\mu u_k(x) = (q_0, u_1)_\mu u_1(x) = \frac{u_1(x)}{(1, u_1)_\mu}.$$

Réciproquement, si l'hypothèse (i) n'est pas vérifiée, *i.e.* il existe i tel que $\lambda_i = \lambda_{i+1}$ (noter que nécessairement $i \geq 2$), alors il suffit de considérer

$$d\mu_0 = \left(\frac{u_1}{\int_W u_1 \, d\mu} + \varepsilon \left(\left(\int_W u_{i+1} \, d\mu \right) u_i - \left(\int_W u_i \, d\mu \right) u_{i+1} \right) \right) d\mu$$

pour ε assez petit pour obtenir une mesure de probabilité telle que le temps de sortie est exponentiel. De même si l'hypothèse (ii) n'est pas vérifiée, *i.e.* il existe i tel que $\int_W u_i \, d\mu = 0$ (noter que nécessairement $i \geq 2$), alors il suffit de considérer

$$d\mu_0 = \left(\frac{u_1}{\int_W u_1 \, d\mu} + \varepsilon u_i \right) d\mu$$

pour ε assez petit pour obtenir une mesure de probabilité telle que le temps de sortie est exponentiel. Dans les deux cas, μ_0 est différent de ν.

Pour en savoir plus sur ce sujet, on pourra consulter l'article : C. Le Bris, T. Lelièvre, M. Luskin et D. Perez, A mathematical formalization of the parallel replica dynamics, `http://arxiv.org/abs/1105.4636`.