

UNE VARIANTE DE L'INÉGALITÉ DE CHEEGER POUR LES CHAÎNES DE MARKOV FINIES

LAURENT MICLO

RÉSUMÉ. Soit P un noyau markovien réversible par rapport à une probabilité μ sur un ensemble fini de cardinal $n \in \mathbb{N}^*$, on montre que l'on a toujours $2I \geq \lambda \geq I/(n-1)$, où λ et I sont respectivement le trou spectral et la constante isopérimétrique associés à P et μ . La constante $n-1$ ci-dessus est optimale, on verra toutefois comment il est possible de l'améliorer dans certaines situations particulières (arbres pointés radiaux à nombre fini de générations). Une application de l'encadrement précédent est de retrouver immédiatement le comportement à basse température des trous spectraux associés aux algorithmes de recuit simulé généralisés finis. Une approche similaire permet également d'obtenir très simplement une caractérisation de l'énergie virtuelle due à Trouvé.

1. INTRODUCTION

Dans un cadre de géométrie riemannienne, Cheeger (1970) a introduit une inégalité permettant de minorer un trou spectral en termes d'une constante isopérimétrique. Ce résultat a ensuite été étendu à d'autres situations (voir par exemple Pignataro et Sullivan (1986), Mohar (1989)) et notamment à celle des chaînes de Markov réversibles (e.g. Lawler et Sokal (1988), Diaconis et Stroock (1991), Saloff-Coste (1997) et les références qui y sont données). Commençons donc par rappeler la forme que prend cette inégalité dans ce dernier contexte :

Soit (S, \mathcal{S}) un ensemble mesurable sur lequel on se donne un noyau markovien P admettant une probabilité réversible μ . La constante isopérimétrique $I(P, \mu)$ associée est alors définie par

$$I(P, \mu) = \inf_{A \in \mathcal{S}, 0 < \mu(A) \leq 1/2} \frac{\mu(\mathbf{1}_A P(\mathbf{1}_{A^c}))}{\mu(A)}$$

(si μ est une masse de Dirac, on convient plutôt que $I(P, \mu) = 1$).

D'autre part P se prolonge uniquement en un opérateur (alors autoadjoint, positif et de norme 1) sur $\mathbb{L}^2(\mu)$ et on note $\lambda(P, \mu)$ le trou spectral de $I - P$, où I est l'opérateur identité, c'est-à-dire que l'on pose

$$\lambda(P, \mu) = \inf_{f \in \mathbb{L}^2(\mu) \setminus \{0\}, \mu(f)=0} \frac{\mu(f(I - P)f)}{\mu(f^2)}$$

(avec la convention que $\lambda(P, \mu) = 2$, si μ est une masse de Dirac).

L'inégalité de Cheeger s'énonce alors

$$\lambda(P, \mu) \geq \frac{I^2(P, \mu)}{2}$$

En général l'exposant 2 ci-dessus ne peut pas être amélioré, comme le montre l'exemple classique de la marche symétrique aux plus proches voisins

sur $\mathbb{Z}/(n\mathbb{Z})$, quand $n \in \mathbb{N}^*$ devient grand. Mais qu'en est-il si on impose à S d'être un ensemble fini de cardinal donné?

Cette note a pour objectif de vérifier que dans cette situation où l'espace est figé, $I(P, \mu)$ et $\lambda(P, \mu)$ sont plutôt du même ordre, asymptotiquement quand l'une de ces quantités devient petite.

2. UNE VARIANTE LINÉAIRE DE L'INÉGALITÉ DE CHEEGER

Soit $n \in \mathbb{N}^*$ fixé, considérons $S = \{0, \dots, n\}$ muni de sa tribu usuelle \mathcal{S} , et notons \mathcal{R}_n l'ensemble des couples (P, μ) formés d'un noyau markovien P réversible par rapport une probabilité μ sur (S, \mathcal{S}) . Notre but est de montrer le résultat suivant, qui est plus intéressant que l'inégalité de Cheeger dès que $I(P, \mu) \leq 2/n$, et moins bon sinon :

Théorème 2.1. *Pour tout $(P, \mu) \in \mathcal{R}_n$, on a*

$$\lambda(P, \mu) \geq \frac{1}{n} I(P, \mu)$$

Rappelons d'autre part qu'il est bien connu que l'on a toujours trivialement $\lambda(P, \mu) \leq 2I(P, \mu)$ (pour le voir, il est commode d'introduire l'autre constante isopérimétrique $\tilde{I}(P, \mu) = \inf_{A \in \mathcal{S}, 0 < \mu(A) < 1} \mu(\mathbf{1}_A P(\mathbf{1}_{A^c})) / [\mu(A)\mu(A^c)]$, qui vérifie $2I(P, \mu) \geq \tilde{I}(P, \mu) \geq I(P, \mu)$ et clairement $\tilde{I}(P, \mu) \geq \lambda(P, \mu)$, car la définition de $\tilde{I}(P, \mu)$ calque celle de $\lambda(P, \mu)$, mais où dans l'infimum on ne retient que les fonctions de la forme $\mathbf{1}_A - \mu(A)$, pour $A \in \mathcal{S}$), ce qui permet d'obtenir l'encadrement

$$(2.1) \quad 2I(P, \mu) \geq \lambda(P, \mu) \geq \frac{1}{n} I(P, \mu)$$

Avant de donner la démonstration du théorème 2.1, revenons au cas général d'un couple réversible (P, μ) sur un espace mesurable (S, \mathcal{S}) quelconque, et introduisons la quantité

$$\bar{\lambda}(P, \mu) = \inf_{f \in H(\mu)} \frac{\mu(f(I - P)f)}{\mu(f^2)}$$

avec $H(\mu) = \{f \in \mathbb{L}^2(\mu) / f \geq 0, \mu(\{f = 0\}) \geq 1/2\}$.

Cette constante $\bar{\lambda}(P, \mu)$ a déjà été considérée par Mathieu (1997) et se compare très bien au trou spectral :

Lemme 2.1. *De manière générale, même si μ est seulement supposée invariante pour P , on a*

$$\bar{\lambda}(P, \mu) \leq \lambda(P, \mu) \leq 2\bar{\lambda}(P, \mu)$$

Ce résultat peut s'interpréter comme une version \mathbb{L}^2 des inégalités $2I(P, \mu) \geq \tilde{I}(P, \mu) \geq I(P, \mu)$.

Preuve. Quitte à remplacer P par $(P + P^*)/2$, où P^* est l'adjoint de P dans $\mathbb{L}^2(\mu)$, on peut se ramener au cas où μ est réversible pour P .

Pour prouver la première inégalité, soit une fonction $f \in \mathbb{L}^2(\mu)$, soit a une de ses médianes (i.e. a est un réel satisfaisant $\mu(f \leq a) \geq 1/2$ et $\mu(f \geq a) \geq 1/2$) et posons $g_+ = 0 \vee (f - a)$ et $g_- = 0 \vee (a - f)$, qui sont bien des éléments de $H(\mu)$.

En considérant les différents cas qui se présentent ($f(x) - a = g_+(x)$ ou $-g_-(x)$ et $f(y) - a = g_+(y)$ ou $-g_-(y)$), on obtient que pour tous $x, y \in S$,

$$(f(y) - f(x))^2 \geq (g_+(y) - g_+(x))^2 + (g_-(y) - g_-(x))^2$$

ainsi en utilisant que

$$\mu(f(I - P)f) = \frac{1}{2} \int (f(y) - f(x))^2 \mu(dx)P(x, dy)$$

on fait apparaître que

$$\begin{aligned} \mu(f(I - P)f) &\geq \mu(g_+(I - P)g_+) + \mu(g_-(I - P)g_-) \\ &\geq \bar{\lambda}(P, \mu)[\mu(g_+^2) + \mu(g_-^2)] \\ &= \bar{\lambda}(P, \mu)\mu((f - a)^2) \\ &\geq \bar{\lambda}(P, \mu)\mu((f - \mu(f))^2) \end{aligned}$$

d'où en fin de compte $\lambda(P, \mu) \geq \bar{\lambda}(P, \mu)$.

L'inégalité inverse $\bar{\lambda}(P, \mu) \geq \lambda(P, \mu)/2$ est due à Mathieu (1997), rappelons son argument dans notre contexte :

Soit $f \in H(\mu)$, on a

$$\begin{aligned} \mu(f(I - P)(f)) &= \mu((f - \mu(f))(I - P)(f - \mu(f))) \\ &\geq \lambda(P, \mu)[\mu(f^2) - \mu(f)^2] \end{aligned}$$

et il reste à remarquer que par l'inégalité de Cauchy-Schwarz,

$$\mu(f)^2 \leq \mu(f^2)\mu(\mathbf{1}_{f>0}) \leq \frac{1}{2}\mu(f^2)$$

□

Un des intérêts de la quantité $\bar{\lambda}(P, \mu)$ est qu'elle est aussi relativement proche de la constante isopérimétrique, car remarquons que $I(P, \mu)$ se définit comme $\bar{\lambda}(P, \mu)$, mais où on impose en plus dans l'infimum aux fonctions de $H(\mu)$ d'être des indicatrices, ce qui prouve que l'on a toujours $I(P, \mu) \geq \bar{\lambda}(P, \mu)$.

Par ailleurs, le lemme 2.1 montre qu'il suffit de prouver la

PROPOSITION 2.2. *Pour tout $(P, \mu) \in \mathcal{R}_n$, on a*

$$\bar{\lambda}(P, \mu) \geq \frac{1}{n}I(P, \mu)$$

La clé technique de cette proposition est donnée par le résultat suivant :

Lemme 2.3. *Soit $(P, \mu) \in \mathcal{R}_n$ et f une fonction positive définie sur S telle que $\mu(\{f = 0\}) \geq 1/2$, alors si $1 \leq p \leq n + 1$ désigne le nombre de valeurs différentes prises par f (avec une probabilité strictement positive sous μ), on a*

$$\int_0^{+\infty} \mu((t - f)\mathbf{1}_{f \leq t}P(\mathbf{1}_{f > t})) dt \geq \frac{I(P, \mu)}{2(p - 1)}\mu(f^2)$$

Preuve. Notons $0 = f_0 < f_1 < \dots < f_{p-1}$ les différentes valeurs prises par f avec une probabilité strictement positive sous μ . On peut évidemment supposer que f ne prend que ces valeurs, quitte à modifier éventuellement cette fonction sur un ensemble μ -négligeable, ce qui ne change rien au problème. Soit \mathcal{A} la sous-tribu de \mathcal{S} engendrée par f . On pose $P_{\mathcal{A}} = E_{\mathcal{A}}PI_{\mathcal{A}}$, où $E_{\mathcal{A}}$ est l'opérateur de projection orthogonale de $\mathbb{L}^2(S, \mathcal{S}, \mu)$ sur

$\mathbb{L}^2(S, \mathcal{A}, \mu)$ (i.e. d'espérance conditionnelle sachant \mathcal{A}) et $I_{\mathcal{A}}$ est l'inclusion de $\mathbb{L}^2(S, \mathcal{A}, \mu)$ dans $\mathbb{L}^2(S, \mathcal{S}, \mu)$. Alors $P_{\mathcal{A}}$ est un noyau markovien sur (S, \mathcal{A}) réversible par rapport à $\mu_{\mathcal{A}}$, la restriction de μ à \mathcal{A} , et il apparaît immédiatement que $I(P_{\mathcal{A}}, \mu_{\mathcal{A}}) \geq I(P, \mu)$. Clairement on peut ainsi se ramener à ne considérer que le cas où $\mathcal{S} = \mathcal{A}$. On supposera donc désormais que $S = \{0, \dots, p-1\}$ muni de sa tribu usuelle \mathcal{S} , et on représentera P et μ respectivement par la matrice $(P_{i,j})_{0 \leq i, j \leq p-1}$ et le vecteur $(\mu_i)_{0 \leq i \leq p-1}$ dont toutes les coordonnées sont strictement positives. A une permutation près de S , on peut également faire l'hypothèse que pour tout $i \in S$, $f(i) = f_i$.

Exprimons avec ces notations les termes apparaissant dans le lemme :

$$\begin{aligned}
\int_0^\infty \mu(f \mathbf{1}_{f \leq t} P(\mathbf{1}_{f > t})) dt &= \sum_{0 \leq i \leq p-2} \int_{f_i}^{f_{i+1}} \mu(f \mathbf{1}_{f \leq t} P(\mathbf{1}_{f > t})) dt \\
&= \sum_{0 \leq i \leq p-2} \int_{f_i}^{f_{i+1}} \sum_{0 \leq j \leq i} \mu_j f_j \sum_{i < k \leq p-1} p_{j,k} dt \\
&= \sum_{0 \leq j < i < k \leq p-1} \mu_j p_{j,k} (f_{i+1} - f_i) f_j \\
&= \sum_{1 \leq l \leq j \leq i < k \leq p-1} \alpha_{j,k} h_{i+1} h_l
\end{aligned}$$

où on a posé pour tous $0 \leq j, k \leq p-1$, $\alpha_{j,k} = \mu_j p_{j,k}$ et $h_j = f_j - f_{j-1} \geq 0$ (en convenant que $f_{-1} = 0$).

De manière similaire on obtient

$$\begin{aligned}
&\int_0^\infty t \mu(\mathbf{1}_{f \leq t} P(\mathbf{1}_{f > t})) dt \\
&= \frac{1}{2} \sum_{0 \leq j < i < k \leq p-1} \mu_j p_{j,k} (f_{i+1}^2 - f_i^2) \\
&= \frac{1}{2} \sum_{0 \leq j < i < k \leq p-1} \alpha_{j,k} h_{i+1}^2 + \sum_{0 \leq j < i < k \leq p-1} \sum_{1 \leq l \leq i} \alpha_{j,k} h_{i+1} h_l
\end{aligned}$$

de sorte que finalement

$$\begin{aligned}
(2.2) \quad &\int_0^{+\infty} \mu((t-f) \mathbf{1}_{f \leq t} P(\mathbf{1}_{f > t})) dt \\
&= \frac{1}{2} \sum_{0 \leq j < i < k \leq p-1} \alpha_{j,k} h_{i+1}^2 + \sum_{0 \leq j < l \leq i < k \leq p-1} \alpha_{j,k} h_{i+1} h_l \\
&\geq \frac{1}{2} \sum_{0 \leq j < i < k \leq p-1} \alpha_{j,k} h_{i+1}^2
\end{aligned}$$

Or notons que par définition de $I(P, \mu)$, et en utilisant le fait que $\mu(f = 0) \geq 1/2$, on a pour $1 \leq i \leq p-2$ fixé,

$$\sum_{0 \leq j < i < k \leq p-1} \alpha_{j,k} \geq I(P, \mu) \sum_{i < k \leq p-1} \mu_k$$

d'où en fin de compte

$$\int_0^{+\infty} \mu((t-f)\mathbf{1}_{f \leq t} P(\mathbf{1}_{f > t})) dt \geq \frac{I(P, \mu)}{2} \sum_{1 \leq i \leq k \leq p-1} \mu_k h_i^2$$

Par ailleurs, on a

$$(2.3) \quad \begin{aligned} \mu(f^2) &= \sum_{0 \leq i \leq p-1} \mu_i \left(\sum_{1 \leq j \leq i} h_j \right)^2 \\ &\leq (p-1) \sum_{1 \leq j \leq i \leq p-1} \mu_i h_j^2 \end{aligned}$$

ce qui permet de conclure au résultat annoncé. \square

Nous sommes désormais en mesure de prouver la proposition 2.2, et tout d'abord identifions le terme du membre de gauche apparaissant dans le lemme avec $\mu(f(I-P)f)/2$.

En effet, en écrivant que

$$f = \int_0^{\infty} \mathbf{1}_{f \geq t} dt$$

et en utilisant la symétrie de

$$\mathbb{R}_+^2 \ni (t, s) \mapsto \mu(\mathbf{1}_{f \geq t}(I-P)(\mathbf{1}_{f \geq s}))$$

on fait apparaître que

$$\begin{aligned} \mu(f(I-P)f) &= \int_{[0, \infty]^2} dt ds \mu(\mathbf{1}_{f \geq t}(I-P)\mathbf{1}_{f \geq s}) \\ &= 2 \int_{[0, \infty]^2} dt ds \mathbf{1}_{t \geq s} \mu(\mathbf{1}_{f \geq t}(I-P)\mathbf{1}_{f \geq s}) \\ &= 2 \int_0^{\infty} dt \mu(\mathbf{1}_{f \geq t}(I-P)(t \wedge f)) \\ &= 2 \int_0^{\infty} dt \mu(\mathbf{1}_{f \geq t}(I-P)(t\mathbf{1}_{f \geq t} + f\mathbf{1}_{f < t})) \end{aligned}$$

Mais remarquons d'une part que

$$\begin{aligned} \mu(\mathbf{1}_{f \geq t}(I-P)(\mathbf{1}_{f \geq t})) &= \mu(\mathbf{1}_{f \geq t}(I-P)(\mathbf{1} - \mathbf{1}_{f < t})) \\ &= -\mu(\mathbf{1}_{f \geq t}(I-P)(\mathbf{1}_{f < t})) \\ &= \mu(\mathbf{1}_{f \geq t} P(\mathbf{1}_{f < t})) \end{aligned}$$

et d'autre part que

$$\begin{aligned} \mu(\mathbf{1}_{f \geq t}(I-P)(f\mathbf{1}_{f < t})) &= -\mu(\mathbf{1}_{f \geq t} P(f\mathbf{1}_{f < t})) \\ &= -\mu(f\mathbf{1}_{f < t} P(\mathbf{1}_{f \geq t})) \end{aligned}$$

ce qui prouve la formule

$$\mu(f(I-P)f) = 2 \int_0^{+\infty} \mu((t-f)\mathbf{1}_{f \leq t} P(\mathbf{1}_{f > t})) dt$$

La proposition 2.2 s'en déduit immédiatement, via le lemme 2.1.

Dans la démonstration du lemme 2.3, seul a importé qu'il existe des réels $0 = f_0 < f_1 < \dots < f_{p-1}$ tels que $\sum_{0 \leq i \leq p-1} \mu(\{f = f_i\}) = 1$ (et

$\mu(\{f = 0\}) \geq 1/2$), et le résultat présenté reste satisfait sous cette hypothèse, sans autre condition sur l'espace mesurable (S, \mathcal{S}) , la finitude de S ne nous ayant servi qu'à assurer la validité de cette hypothèse en permettant de majorer p par $\text{card}(S)$, mais on peut plus précisément le majorer par $n + 2 - \min\{\text{card}(A) / A \subset S, \mu(A) \geq 1/2\}$, ce qui montre que dans le théorème 2.1 et dans la proposition 2.2, on peut remplacer n par

$$n + 1 - \min\{\text{card}(A) / A \subset S, \mu(A) \geq 1/2\}$$

mais ce nombre, qui dépend alors de μ , est toujours minoré par $n/2$, ce qui au mieux permet de gagner un facteur 2.

3. OPTIMALITÉ DE L'INÉGALITÉ PRÉCÉDENTE

Nous allons donc vérifier que $1/n$ est la meilleure constante indépendante de $(P, \mu) \in \mathcal{R}_n$ que l'on peut espérer mettre dans l'inégalité du théorème 2.1, et ceci même asymptotiquement pour $I(P, \mu)$ petit, i.e.

$$(3.1) \quad \lim_{i \rightarrow 0^+} \sup_{(P, \mu) \in \mathcal{R}_n, 0 < I(P, \mu) \leq i} \frac{I(P, \mu)}{\lambda(P, \mu)} = n$$

D'ailleurs dès que $n \geq 2$, la constante n n'est optimale qu'asymptotiquement pour $I(P, \mu)$ petit, au sens où pour tout $0 < i \leq 1$,

$$(3.2) \quad \sup_{(P, \mu) \in \mathcal{R}_n, i \leq I(P, \mu)} \frac{I(P, \mu)}{\lambda(P, \mu)} < n$$

Pour prouver l'affirmation (3.1), on va utiliser un exemple suggéré par la preuve du lemme 2.3 (si l'on voulait seulement montrer que $1/n$ est la meilleure constante possible dans l'inégalité de la proposition 2.2, il suffirait de considérer ci-dessous $\eta = 1/2$ et faire tendre ϵ vers 0_+):

Pour $0 < \eta \leq 1/2$ et $0 < \epsilon < 2\eta/(2n-1)$, soit $P_{\eta, \epsilon}$ la matrice markovienne donnée par

$$\forall 0 \leq i, j \leq n,$$

$$P_{\eta, \epsilon}(i, j) = \begin{cases} 1/2 & , \text{ si } 1 \leq i \leq n-1 \text{ et } j = i+1 \\ 1/2 & , \text{ si } 1 \leq i \leq n-1 \text{ et } j = i-1 \\ \epsilon/(2-2\eta) & , \text{ si } i=0 \text{ et } j=1 \\ \epsilon/(2\eta-2(n-1)\epsilon) & , \text{ si } i=n \text{ et } j=n-1 \\ 1-\epsilon/(2-2\eta) & , \text{ si } i=0=j \\ 1-\epsilon/(2\eta-2(n-1)\epsilon) & , \text{ si } i=n=j \\ 0 & , \text{ sinon} \end{cases}$$

laquelle est clairement réversible par rapport à la probabilité $\mu_{\eta, \epsilon}$ décrite par

$$\forall 0 \leq i \leq n, \quad \mu_{\eta, \epsilon}(i) = \begin{cases} 1-\eta & , \text{ si } i=0 \\ \epsilon & , \text{ si } 1 \leq i \leq n-1 \\ \eta-(n-1)\epsilon & , \text{ si } i=n \end{cases}$$

Puisque pour tout $0 \leq i \leq n-1$, on a $\mu_{\eta, \epsilon}(i)P_{\eta, \epsilon}(i, i+1) = \epsilon/2$, on se convainc facilement que $I(P_{\eta, \epsilon}, \mu_{\eta, \epsilon}) = \epsilon/(2\eta)$ (le minimum étant atteint en

$A = \{1, \dots, n\}$). D'autre part soit \tilde{f} la fonction définie par $\tilde{f}(i) = i$, pour tout $0 \leq i \leq n$. On calcule que

$$\mu_{\eta, \epsilon}(\tilde{f}(I - P_{\eta, \epsilon})(\tilde{f})) = n\epsilon/2$$

et

$$\begin{aligned} \mu_{\eta, \epsilon}(\tilde{f}^2) - \mu_{\eta, \epsilon}(\tilde{f})^2 &= \\ \frac{(n-1)n(2n-1)}{6}\epsilon + (\eta - \epsilon(n-1))n^2 - \left[\frac{(n-1)n}{2}\epsilon + (\eta - \epsilon(n-1))n \right]^2 \end{aligned}$$

Ainsi, η étant fixé comme ci-dessus, si on fait tendre ϵ vers 0_+ , on s'aperçoit que

$$\begin{aligned} \lim_{\epsilon \rightarrow 0_+} \frac{I(P_{\eta, \epsilon}, \mu_{\eta, \epsilon})}{\lambda(P_{\eta, \epsilon}, \mu_{\eta, \epsilon})} &\geq \lim_{\epsilon \rightarrow 0_+} \frac{\epsilon \mu_{\eta, \epsilon}[(\tilde{f} - \mu_{\eta, \epsilon}(\tilde{f}))^2]}{2\eta \mu_{\eta, \epsilon}(\tilde{f}(I - P_{\eta, \epsilon})(\tilde{f}))} \\ &= (1 - \eta)n \end{aligned}$$

et il reste à faire tendre η vers 0_+ pour obtenir (3.1) (plus directement, on pouvait aussi par exemple poser $\epsilon = \eta^2$ pour η assez petit et faire tendre η vers 0_+).

Montrons maintenant que la constante n n'est jamais atteinte dès que $n \geq 2$, en prouvant que pour tout $0 < i \leq 1$ fixé,

$$(3.3) \quad \sup_{(P, \mu) \in \mathcal{R}_n, i \leq I(P, \mu)} \frac{I(P, \mu)}{\lambda(P, \mu)} < n$$

(d'ailleurs remarquons que pour $i > 4/n$, l'inégalité de Cheeger et le lemme 2.1 affirment que ce supremum est majoré par $4/i < n$), ce qui entrainera (3.2).

Par l'absurde, supposons donc qu'il existe une suite $((P_m, \mu_m))_{m \in \mathbb{N}}$ d'éléments de \mathcal{R}_n telle que

$$(3.4) \quad \begin{aligned} \forall m \in \mathbb{N}, \quad I(P_m, \mu_m) &\geq i \\ \lim_{m \rightarrow \infty} \frac{I(P_m, \mu_m)}{\lambda(P_m, \mu_m)} &= n \end{aligned}$$

Par compacité de \mathcal{R}_n (que l'on aura muni de la topologie héritée de celle de $(\mathbb{R}^{n+1} \otimes \mathbb{R}^{n+1}) \times \mathbb{R}^{n+1}$) et quitte à extraire une sous-suite, on peut faire l'hypothèse que les (P_m, μ_m) admettent pour m grand une limite $(P, \mu) \in \mathcal{R}_n$. De la même manière et d'après (3.4) (car on convient que $\bar{\lambda}(P, \mu) = 1$ si μ est une masse de Dirac), on se ramène au cas où aucune des probabilités μ_m n'est une masse de Dirac, ce qui nous assure pour tout $m \in \mathbb{N}$ de l'existence d'une fonction $f_m \in H(\mu_m) \setminus \{0\}$ telle que

$$\bar{\lambda}(P_m, \mu_m) = \frac{\mu_m(f_m(I - P_m)(f_m))}{\mu_m(f_m^2)}$$

A une renormalisation près, on peut supposer que $\max_S f_m = n$, et à une réindexation près, que f_m est croissante sur $\{0, \dots, n\}$. On peut encore faire l'hypothèse que $f = \lim_{m \rightarrow \infty} f_m$ existe.

Nous allons maintenant utiliser les étapes de la preuve du lemme 2.3 pour obtenir une contradiction.

L'inégalité (2.3) implique nécessairement, si l'on veut que (3.4) soit satisfait, d'une part que f est la fonction définie par $f(j) = j$ pour tout $0 \leq j \leq n$

(et notamment pour m assez grand, f_m prend $p = n + 1$ valeurs distinctes), et d'autre part que

$$\forall 1 \leq j \leq n - 1, \quad \lim_{m \rightarrow \infty} \frac{\mu_m(j)}{\mu_m(n)} = 0$$

Ces dernières convergences ont pour conséquence que $P(n, 0) > 0$. En effet, si l'on devait avoir $P(n, 0) = 0$, écrivons que

$$\begin{aligned} I(P_m, \mu_m) &\leq \frac{\mu_m(\mathbf{1}_{\{n\}} P_m(\mathbf{1}_{\{0, \dots, n-1\}}))}{\mu_m(\{n\})} \\ &= \frac{\sum_{1 \leq j < n} \mu_m(j) P_m(j, n) + \mu_m(n) P_m(n, 0)}{\mu_m(n)} \end{aligned}$$

pour faire apparaître que cette expression converge vers 0 pour m grand, ce qui n'est pas autorisé. Ainsi $P(n, 0) > 0$. En tenant compte de ce qui précède, on voit alors dans le second membre de (2.2) (où des m doivent être mis en indice partout), puisque $n \geq 2$, que le dernier terme n'est pas asymptotiquement négligeable devant le premier, ce qui finalement nous fait aboutir à une contradiction.

Si $n = 1$, pour tout $(P, \mu) \in \mathcal{R}_1$, $I(P, \mu) = \bar{\lambda}(P, \mu)$, et (3.3) est donc faux. Pour voir que (3.2) n'est pas plus juste, on peut utiliser directement la paramétrisation suivante de $\tilde{\mathcal{R}}_1$, en convenant désormais de noter $\tilde{\mathcal{R}}_n$ l'ensemble des couples $(P, \mu) \in \mathcal{R}_n$ tels que μ ne soit pas une masse de Dirac : soit $\Theta = \{(\eta, a) \in]0, 1[\times]0, 1[/ a(1 - \eta) \leq \eta\}$ et pour $(\eta, a) \in \Theta$, notons

$$\begin{aligned} \mu_{\eta, a} &= (1 - \eta, \eta) \\ P_{\eta, a} &= \begin{pmatrix} 1 - a & a \\ (1 - \eta)a/\eta & 1 - (1 - \eta)a/\eta \end{pmatrix} \end{aligned}$$

de sorte que $\tilde{\mathcal{R}}_1 = \{(P_{\eta, a}, \mu_{\eta, a}); (\eta, a) \in \Theta\}$.

Cependant on calcule que pour $(\eta, a) \in \Theta$,

$$\begin{aligned} I(P_{\eta, a}, \mu_{\eta, a}) &= \frac{(1 - \eta)a}{\eta \wedge (1 - \eta)} \\ \lambda(P_{\eta, a}, \mu_{\eta, a}) &= \frac{a}{\eta} \end{aligned}$$

et il suffit donc de prendre $a = \eta/(1 - \eta)$ et de faire tendre η vers 0_+ pour réaliser que pour tout $0 < i \leq 1$,

$$\sup_{(P, \mu) \in \mathcal{R}_1, i \leq I(P, \mu)} \frac{I(P, \mu)}{\lambda(P, \mu)} = 1$$

Les considérations précédentes nous ont fait effleurer la question de la continuité des applications $I : \mathcal{R}_n \ni (P, \mu) \mapsto I(P, \mu)$ et $\lambda : \mathcal{R}_n \ni (P, \mu) \mapsto \lambda(P, \mu)$ (ainsi que celle de \tilde{I} et $\bar{\lambda}$). L'exemple précédent des $(P_{\eta, \epsilon}, \mu_{\eta, \epsilon})$, avec $0 < \eta \leq 1/2$ et $0 < \epsilon < 2\eta/(2n - 1)$, est suffisant pour nous convaincre qu'elles ne sont pas continues sur \mathcal{R}_n , toutefois notons qu'elles y sont semi-continues supérieurement (vu les conventions faites, qui assurent que le maximum de ces applications est notamment atteint en les couples $(P, \mu) \in \mathcal{R}_n \setminus \tilde{\mathcal{R}}_n$), et on peut également vérifier que l'ensemble de continuité de ces

applications est $\mathcal{R}_n^* \cup \mathcal{R}_n^\dagger$, où \mathcal{R}_n^* est l'ouvert formé des couples $(P, \mu) \in \mathcal{R}_n$ tels que μ charge tous les points de $\{0, \dots, n\}$ et où \mathcal{R}_n^\dagger est le fermé $\{(P, \mu) \in \mathcal{R}_n / I(P, \mu) = 0\}$ des couples réversibles réductibles.

4. EXEMPLES D'AMÉLIORATIONS POSSIBLES

Bien que de manière générale l'inégalité du théorème 2.1 soit optimale, on peut toutefois l'améliorer dans des situations particulières. Pour ceci, remarquons que dans la démonstration du théorème 2.1, il suffit dans la preuve de l'inégalité $\lambda(P, \mu) \geq \bar{\lambda}(P, \mu)$ du lemme 2.1 de prendre pour f l'un des vecteurs propres (non nul) associé à $\lambda(P, \mu)$, puis de considérer par la suite $(f - a)_+$ et $(f - a)_-$ où a est une médiane de f , ainsi si pour certaines raisons (de symétrie par exemple) on sait qu'il existe un tel vecteur ne prenant qu'un nombre donné de valeurs $\tilde{n} + 1$, on peut alors substituer \tilde{n} (respectivement $2\tilde{n}$) à n dans le théorème 2.1 (resp. dans la proposition 2.2).

Nous allons illustrer ceci pour les noyaux P réversibles définissant naturellement une structure « d'arbre pointé radial » :

Soit $(P, \mu) \in \mathcal{R}_n$, comme d'habitude, on associe à P un graphe (non orienté, sans arêtes multiples mais avec d'éventuelles boucles) sur S en imposant pour $x, y \in S$, à $\{x, y\}$ d'être une arête si et seulement si $P(x, y) > 0$. On note E l'ensemble des arêtes ainsi définies. Le graphe (S, E) est appelé un arbre pointé radial s'il est connexe, s'il existe un élément particulier (la racine) de S , disons 0, un entier $N \in \mathbb{N}$ (le nombre de générations plus 1) et une suite finie $(d_k)_{1 \leq k \leq N}$ d'entier(s) strictement positif(s), tels qu'en notant pour $0 \leq k \leq N + 1$, S_k l'ensemble des éléments de S à une distance k de 0 (rappelons que la distance entre deux points est la longueur du plus petit chemin les reliant dans (S, E)), on ait $S_{N+1} = \emptyset$ et pour tout $1 \leq k \leq N$ et tout $x \in S_k$, que x a $1 + d_k$ voisins différents de x , un appartenant à S_{k-1} (son père, appelé $p(x)$) et d_k appartenant à S_{k+1} (ses fils, dont l'ensemble est noté $F(x)$). On a ainsi $d_N = 0$ et on pose $d_0 = \text{card}(S_1)$.

On dit que P est compatible avec cette structure de graphe, si de plus pour tout $0 \leq k \leq N - 1$, tout $x \in S_k$ et tout $y \in F(x)$, $d_k P(x, y)$ ne dépend que de k , et vaut disons $p_k > 0$, et si pour tout $1 \leq k \leq N$, tout $x \in S_k$, $q_k \stackrel{\text{déf.}}{=} P(x, p(x)) > 0$ ne dépend aussi que de k . Nécessairement pour tout $0 \leq k \leq N$, $r_k \stackrel{\text{déf.}}{=} P(x, x)$ ne dépend pas non plus du choix de x dans la $(k + 1)^{\text{ième}}$ génération S_k .

Une fonction f définie sur S sera dite spéciale (sous-entendu relativement à E) si sur toute génération S_k , pour $1 \leq k \leq N$, f ne prend au plus que deux valeurs. Notamment le cardinal de l'image d'une telle application est borné par $2N + 1$. Rappelons qu'une fonction f est dite radiale pour (S, E) si elle est constante sur chacune des générations S_k , $0 \leq k \leq N$, elle est donc aussi spéciale.

Nous allons vérifier ci-dessous que sous les conditions précédentes, si λ est une valeur propre de P , alors il existe un vecteur propre associé qui est spécial.

En effet, remarquons que la formule explicite (cf. le lemme 3.1 p. 177 du livre de Freidlin et Wentzell (1984)) qui donne μ en termes de P (car ce dernier est irréductible d'après ce qui précède) montre que l'application $S \ni$

$x \mapsto \mu(\{x\})$ est radiale. Soit D la matrice diagonale indexée par $S \times S$, dont les éléments diagonaux sont donnés par $D_{x,x} = \sqrt{\mu(\{x\})}$ pour tout $x \in S$. Soit $\tilde{P} = DPD^{-1}$ (il s'agit de « l'opérateur de Schrödinger discret » associé à la matrice markovienne P) et notons que par l'hypothèse de réversibilité, \tilde{P} est symétrique. De plus P et \tilde{P} ont même spectre et leur vecteurs propres respectifs se déduisent réciproquement par application de D ou de D^{-1} . Pour obtenir l'affirmation précédente, il suffit donc d'appliquer à \tilde{P} le lemme suivant.

Introduisons tout d'abord une notation : à l'arbre pointé radial (S, E) on associe $\mathcal{M}(S, E)$ l'ensemble des matrices symétriques réelles $M = (M_{i,j})_{i,j \in S}$ telles que

- pour $i, j \in S$, si $\{i, j\} \notin E$, alors $M_{i,j} = 0$
- pour tout $0 \leq k \leq N-1$ et tout $i \in S_{k+1}$, $m_k \stackrel{\text{déf.}}{=} M_{p(i),i}$ ne dépend que de k et est non nul
- pour tout $0 \leq k \leq N$ et tout $i \in S_k$, $l_k \stackrel{\text{déf.}}{=} M_{i,i}$ ne dépend que de k .

Lemme 4.1. *Soit $M \in \mathcal{M}(S, E)$ et soit λ une valeur propre de M . Il existe alors un vecteur propre spécial qui lui est associé.*

Preuve. La démonstration repose sur une récurrence sur $N \in \mathbb{N}$: supposons donc le résultat établi pour tout arbre pointé radial (S', E') dont le nombre de générations est plus petit que N et tout $M' \in \mathcal{M}(S', E')$, et soit $M \in \mathcal{M}(S, E)$, où le nombre total de générations de (S, E) est $N+1$. Considérons une valeur propre λ de M et f un vecteur propre associé. Pour $x \in S_{N+1}$, on a

$$m_N f(p(x)) + l_{N+1} f(x) = \lambda f(x)$$

Distinguons deux cas :

- Si $\lambda = l_{N+1}$, nécessairement $f(x) = 0$ pour tout $x \in S_N$. Soit $S' = S_0 \sqcup \dots \sqcup S_{N-1}$ que l'on munit de la structure d'arbre pointé radiale héritée naturellement de celle de (S, E) . Posons aussi $M' = (M_{i,j})_{i,j \in S'}$ et $f' = (f(i))_{i \in S'}$ et remarquons que

$$M' f' = \lambda f'$$

Dans un premier temps, supposons que f' n'est pas la fonction identiquement nulle, ce qui assure que λ est aussi une valeur propre de M' . Par l'hypothèse de récurrence on peut alors trouver un vecteur propre spécial \tilde{f}' associé à la valeur propre λ sur S' .

Posons alors

$$\forall 0 \leq k \leq N+1, \forall x \in S_k,$$

$$\tilde{f}(x) = \begin{cases} \tilde{f}'(x) & , \text{ si } 0 \leq k \leq N-1 \\ 0 & , \text{ si } k = N \\ -m_{N-1} \tilde{f}(p(p(x))) / (d_N m_N) & , \text{ si } k = N+1 \end{cases}$$

En tenant compte des égalités

$$m_{N-1} \tilde{f}(p(x)) + l_N \tilde{f}(x) + m_N \sum_{y \in F(x)} \tilde{f}(y) = 0 = \lambda \tilde{f}(x)$$

valables pour $x \in S_N$, et de l'hypothèse $\lambda = l_{N+1}$, il est clair que \tilde{f} est une fonction spéciale non nulle satisfaisant $M\tilde{f} = \lambda\tilde{f}$.

Venons-en maintenant aux cas où $f' \equiv 0$ (ou $N = 1$), qui est la situation intéressante car c'est elle qui introduit deux valeurs différentes dans une génération. Nécessairement (sinon $f \equiv 0$) $d_N > 1$, et pour chaque $x \in S_N$, choisissons un de ses fils $r(x)$, puis notons $R = \{r(x); x \in S_N\}$. Il suffit alors de poser

$$\forall 0 \leq k \leq N + 1, \forall x \in S_k,$$

$$\tilde{f}(x) = \begin{cases} 0 & , \text{ si } k \leq N \\ -(d_N - 1) & , \text{ si } k = N + 1 \text{ et si } x \in R \\ 1 & , \text{ si } k = N + 1 \text{ et si } x \notin R \end{cases}$$

car on se convainc facilement que cette fonction est spéciale, non nulle et propre pour λ .

• Si $\lambda \neq l_{N+1}$, nécessairement pour $x \in S_{N+1}$, $f(x) = m_N f(p(x)) / (\lambda - l_{N+1})$. Ceci nous amène à considérer $S' = S_0 \sqcup \dots \sqcup S_N$ muni de la structure d'arbre pointé radiale héritée naturellement de celle de (S, E) et $M' = (M'_{i,j})_{i,j \in S'} \in \mathcal{M}(S', E')$ la matrice définie par

$$\forall i, j \in S', \quad M'_{i,j} = \begin{cases} l_N + d_N m_N (\lambda - l_{N+1})^{-1} & , \text{ si } i = j \in S_N \\ M_{i,j} & , \text{ sinon} \end{cases}$$

Soit également $f' = (f'(i))_{i \in S'}$, il apparaît que $f' \not\equiv 0$, sinon $f \equiv 0$, et que $M'f' = \lambda f'$. Par l'hypothèse de récurrence, il existe \tilde{f}' une fonction spéciale qui est propre pour la valeur propre λ de M' .

En conséquence, si on pose

$$\forall 0 \leq k \leq N + 1, \forall x \in S_k,$$

$$\tilde{f}(x) = \begin{cases} \tilde{f}'(x) & , \text{ si } 0 \leq k \leq N \\ m_N \tilde{f}'(p(x)) / (\lambda - l_{N+1}) & , \text{ si } k = N + 1 \end{cases}$$

on obtient un des vecteurs propres spéciaux que nous cherchions.

Reste à noter que l'initiation de la récurrence en $N = 0$ est triviale. \square

En fait la condition de symétrie des matrices de $\mathcal{M}(S, E)$ est inutile, elle ne nous a servi qu'à assurer qu'elles étaient diagonalisables. Sinon il faut considérer les valeurs propres et les fonctions spéciales complexes (évidemment il faut rajouter l'hypothèse que $n_k \stackrel{\text{déf.}}{=} M_{i,p(i)}$ ne dépend pas du choix de $i \in S_{k+1}$, $0 \leq k \leq N - 1$, et est non nul, en revanche, on peut aussi autoriser les matrices à être complexes).

La preuve précédente permet également de préciser un peu plus la possibilité du choix d'un « bon » vecteur propre : soit il en existe un qui est radial, soit il en existe un, disons f , qui a la structure suivante ; il existe $0 \leq k_0 < N$ tel que pour tout $x \in S_0 \sqcup \dots \sqcup S_{k_0}$, $f(x) = 0$, tel qu'à la génération $k_0 + 1$, f prenne deux valeurs, l'une strictement positive et l'autre strictement négative (avec pour tout $x \in S_{k_0}$, $\sum_{y \in F(x)} f(y) = 0$), et tel que si l'on ne considère que les descendants d'un élément x_0 de S_{k_0+1} , la restriction de f est radiale sur ce nouveau sous-arbre planté en $\{x_0\}$.

Cependant vérifions sur l'exemple suivant (avec $N = 1$) que l'on n'a pas toujours de vecteur propre radial, même dans le seul espace propre associé au trou spectral. On considère pour $0 < a \leq 1/2$ et $0 < b \leq 1$ la matrice

$$\begin{pmatrix} 1 - 2a & a & a \\ b & 1 - b & 0 \\ b & 0 & 1 - b \end{pmatrix}$$

On calcule que ses valeurs propres sont 1 , $1 - b$ et $1 - 2a - b$ et que l'espace propre associé à $1 - b$ (le trou spectral vallant b) est la droite vectorielle engendrée par $(0, 1, -1)$.

Revenons au cas général d'un noyau P compatible avec une structure d'arbre pointé radial, il apparaît donc que dans cette situation, on peut remplacer dans le théorème 2.1, $n = \text{card}(S) - 1 = d_0(1 + d_1(\cdots(1 + d_{N-1})\cdots))$ par $2N$.

Ceci fournit d'ailleurs une famille d'exemples où l'on peut appliquer une variante du théorème 2.1 à des ensembles infinis, car ce qui importe ci-dessus c'est que le nombre de générations soit borné et non pas que le graphe soit localement fini :

Soit $N \in \mathbb{N}^*$ et donnons-nous une famille $((E_i, \mathcal{E}_i, m_i))_{1 \leq i \leq N}$ d'espaces mesurables séparables probabilisés. Supposons que pour tout $1 \leq i \leq N$ fixé, il existe une suite $(B_n^{(i)})_{n \geq 0} = ((B_n^{(i,j)})_{1 \leq j \leq d_{i,n}})_{n \geq 0}$ de partitions de plus en plus fines de E_i , telle que $\mathcal{E}_i = \sigma(B_n^{(i,j)}; n \geq 0, 1 \leq j \leq d_{i,n})$ et satisfaisant $m_i(B_n^{(i,j)}) = m_i(B_n^{(i,k)})$, pour tout $n \geq 0$ et tous $1 \leq j, k \leq d_{i,n}$. Comme exemple typique, on peut prendre pour tout $1 \leq i \leq N$, $(E_i, \mathcal{E}_i, m_i) = ([0, 1], \mathcal{B}([0, 1]), m_{|[0,1]})$, où m est la mesure de Lebesgue.

Pour $1 \leq i \leq N$, notons $S_i = E_1 \times \cdots \times E_i$, muni de la tribu $\mathcal{S}_i = \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_i$ et posons $S = \{0\} \sqcup S_1 \sqcup \cdots \sqcup S_N$, $\mathcal{S} = \sigma(\mathcal{S}_1, \cdots, \mathcal{S}_N)$. Pour $x = (e_1, \cdots, e_i) \in S_i$, $1 \leq i \leq N$, on note $p(x) = (e_1, \cdots, e_{i-1}) \in S_{i-1}$ (avec la convention que $p(x) = 0$, si $i = 1$) et si $i < N$, $m^{(x)}$ désignera la probabilité $\delta_{e_1} \otimes \cdots \otimes \delta_{e_i} \otimes m_{i+1}$ sur S_{i+1} .

Enfin donnons-nous trois familles $(p_i)_{0 \leq i \leq N}$, $(r_i)_{0 \leq i \leq N}$ et $(q_i)_{0 \leq i \leq N}$ de nombres positifs tels que $q_0 = p_N = 0$, $p_i + r_i + q_i = 1$ pour tout $0 \leq i \leq N$, et $p_i > 0$ et $q_{i+1} > 0$ pour tout $0 \leq i < N$. On suppose que la matrice markovienne \tilde{P} définie par

$$\forall 0 \leq i, j \leq N, \quad \tilde{P}(i, j) = \begin{cases} p_i & , \text{ si } j = i + 1 \\ r_i & , \text{ si } i = j \\ q_i & , \text{ si } j = i - 1 \\ 0 & , \text{ sinon} \end{cases}$$

admet une probabilité réversible $\nu = (\nu(i))_{0 \leq i \leq N}$.

Considérons maintenant le noyau markovien P défini par

$$P(x, \cdot) = \begin{cases} r_0 \delta_x(\cdot) + p_0 m_1(\cdot) & , \text{ si } x = 0 \\ q_i \delta_{p(x)}(\cdot) + r_i \delta_x(\cdot) + p_i m^{(x)}(\cdot) & , \text{ si } x \in S_i, \text{ avec } 1 \leq i \leq N - 1 \\ q_N \delta_{p(x)}(\cdot) + r_N \delta_x(\cdot) & , \text{ si } x \in S_N \end{cases}$$

on vérifie qu'il est réversible par rapport à la probabilité $\mu = \sum_{0 \leq i \leq N} \nu^{(i)} \mu_i$, où μ_i est la probabilité $m_1 \otimes \cdots \otimes m_i$ sur S_i , pour $1 \leq i \leq N$, et $m_0 = \delta_0$.

On a alors

$$(4.1) \quad \frac{I(P, \mu)}{2N} \leq \lambda(P, \mu) \leq 2I(P, \mu)$$

En effet, ceci s'obtient par une procédure d'approximation : pour $1 \leq i \leq N$ fixé, soit $(\mathcal{A}_n^{(i)})_{n \in \mathbb{N}}$ la suite croissante de sous-tribus finies de \mathcal{E}_i définie par $\mathcal{A}_n^{(i)} = \sigma(B_n^{(i,j)}; 1 \leq j \leq d_{i,n})$ pour tout $n \in \mathbb{N}$. Tout naturellement on pose pour $n \in \mathbb{N}$, $\mathcal{S}_n^{(i)} = \mathcal{A}_n^{(1)} \otimes \cdots \otimes \mathcal{A}_n^{(i)}$, puis $\mathcal{S}_n = \sigma(\mathcal{S}_n^{(i)}; 1 \leq i \leq N)$. On a clairement $\mathcal{S} = \sigma(\mathcal{S}_n; n \geq 0)$, de sorte que l'on soit assuré de

$$\begin{aligned} I(P, \mu) &= \lim_{n \rightarrow \infty} I(P_{\mathcal{S}_n}, \mu_{\mathcal{S}_n}) \\ \lambda(P, \mu) &= \lim_{n \rightarrow \infty} \lambda(P_{\mathcal{S}_n}, \mu_{\mathcal{S}_n}) \end{aligned}$$

or d'après ce qui précède, pour tout $n \in \mathbb{N}$,

$$\frac{I(P_{\mathcal{S}_n}, \mu_{\mathcal{S}_n})}{2N} \leq \lambda(P_{\mathcal{S}_n}, \mu_{\mathcal{S}_n})$$

ce qui à la limite prouve (4.1).

Remarquons également que pour ce type d'exemples, la constante isopérimétrique se calcule relativement facilement : ainsi on a, si $\nu(0) \geq 1/2$,

$$I(P, \mu) = \min_{1 \leq k \leq N} \frac{q_k}{1 + s_k + s_k s_{k+1} + \cdots + s_k \cdots s_{N-1}}$$

avec pour $2 \leq i \leq N$, $s_i = p_i/q_{i+1}$ (en effet, considérer d'abord le cas où les E_i , $1 \leq i \leq N$, sont finis, sous l'hypothèse que $\mu(\{0\}) \geq 1/2$, il est clair que le minimum définissant $I(P, \mu)$ est atteint en un sous-arbre formé d'un individu et de tous ses descendants, et noter que par réversibilité, la probabilité μ s'exprime très simplement à partir de $\mu(\{0\})$ en fonction des données).

Néanmoins l'hypothèse de structure est ici très forte : si l'on veut que (4.1) reste satisfait « avec des β en indice », au sens donné dans l'introduction, il faut exiger du potentiel perturbateur U qu'il soit radial.

La remarque du début de cette section peut également s'appliquer à certains exemples de marches aléatoires sur un groupe fini : on suppose que l'ensemble fini S est muni d'une structure de groupe et d'une probabilité m . La chaîne de Markov associée est celle dont les probabilités de transitions sont données par la matrice P définie par

$$\forall x, y \in S, \quad P(x, y) = m(yx^{-1})$$

Soit μ la probabilité équilibrée sur S , elle est clairement invariante pour P et sera réversible si m est symétrique (i.e. $m(x) = m(x^{-1})$ pour tout $x \in S$). Outre cette symétrie, supposons que m est une fonction de classe, c'est-à-dire que $m(x)$ ne dépend que de la classe de conjugaison à laquelle appartient $x \in S$. A partir du théorème 6 p. 49 du livre de Diaconis (1988), on constate alors sans difficulté que chaque valeur propre de P admet un caractère du groupe comme vecteur propre (et donc aussi, soit sa partie réelle, soit sa partie imaginaire), et il est bien connu que ceux-ci sont des fonctions de classe. Sous ces hypothèses, il apparaît donc que l'on peut

remplacer dans le théorème 2.1, la constante n par le nombre de classes de conjugaison du groupe moins 1. Ceci est évidemment sans intérêt si S est abélien, mais par exemple si $S = S_N$, le groupe symétrique d'ordre N , on est ainsi amené à substituer à $N! - 1$ ($\sim \sqrt{2\pi N} \exp(N \ln(N) - N)$ pour N grand, par la formule de Stirling) le nombre de partitions de N moins 1, qui asymptotiquement pour N grand est équivalent à $\exp(\sqrt{2/3}\pi\sqrt{N})/(4\sqrt{3}N)$ (cf. le théorème 2.1 p. 146 du livre de Ayoub (1963)).

APPENDICE : SUR LE RECUIT SIMULÉ

Le théorème 2.1 montre que si l'on dispose d'une famille $((P_\beta, \mu_\beta))_{\beta \geq 0}$ d'éléments de \mathcal{R}_n telle que la limite de $\beta^{-1} \ln(I(P_\beta, \mu_\beta))$ existe pour β grand, alors les quantités $\beta^{-1} \ln(\lambda(P_\beta, \mu_\beta))$ convergent également et admettent la même limite.

Ce résultat a été suggéré par le comportement à basse température de la constante isopérimétrique et du trou spectral associés aux algorithmes de recuit simulé généralisés. Pour ceux-ci, on dispose d'un noyau markovien Q sur (S, S) , irréductible, mais pas nécessairement réversible par rapport à sa probabilité invariante. On se donne également une fonction de coût $V : \{(x, y), x \neq y \in S\} \rightarrow \mathbb{R}_+ \sqcup \{+\infty\}$, satisfaisant $\forall x \neq y \in S, V(x, y) = +\infty \Leftrightarrow Q(x, y) = 0$. On pose ensuite pour tout $\beta \geq 0$,

$$\forall x, y \in S, \quad Q_\beta(x, y) = \begin{cases} \exp(-\beta V(x, y))Q(x, y) & , \text{ si } x \neq y \\ 1 - \sum_{z \in S \setminus \{x\}} Q_\beta(x, z) & , \text{ sinon} \end{cases}$$

ce qui définit un noyau Q_β irréductible sur (S, S) , admettant donc une unique probabilité invariante μ_β .

Les algorithmes de recuit simulé généralisés associés à la famille $(Q_\beta)_{\beta \geq 0}$ et à une fonction de classe C^1 , $\beta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ (qui représente l'évolution temporelle de l'inverse de la température) satisfaisant $\lim_{t \rightarrow +\infty} \beta_t = +\infty$, sont alors les processus de Markov inhomogènes en temps continu, qui à tout instant $t \geq 0$, admettent $Q_{\beta_t} - I$ comme générateur. Ce type d'algorithme a été beaucoup étudié, car initialement il servait à trouver les minima globaux de problèmes d'optimisation combinatoire, et la forme généralisée présentée ci-dessus apparaît par exemple quand on cherche à paralléliser ces processus (pour plus de précisions sur ce sujet, on renvoie à Trouvé (1996a), Trouvé (1996b) et aux références que contiennent ces articles).

L'une des étape cruciale de l'étude de ces algorithmes revient à évaluer pour β grand le comportement de $\beta^{-1} \ln(\lambda(P_\beta, \mu_\beta))$, où P_β est le noyau markovien réversible par rapport à μ_β donné par $(Q_\beta + Q_\beta^*)/2$ (symétrisé additif de Q_β), où Q_β^* est l'opérateur adjoint de Q_β dans $\mathbb{L}^2(\mu_\beta)$. Or par ailleurs, il est bien connu (cf. Freidlin et Wentzell (1984)) qu'il existe deux fonctions $\rho : S \rightarrow \mathbb{R}_+$ et $U : S \rightarrow \mathbb{R}_+$ (cette dernière pouvant se décrire uniquement en termes de V) telles que pour β grand

$$\forall x \in S, \quad \mu_\beta(x) \sim \rho(x) \exp(-\beta U(x))$$

ce qui permet d'obtenir immédiatement (écrire $I(P_\beta, \mu_\beta)$ sous sa forme symétrique : $\min_{A \in S, 0 < \mu_\beta(A) < 1} \mu_\beta(\mathbf{1}_A P_\beta(\mathbf{1}_{A^c})) / \min(\mu_\beta(A); \mu_\beta(A^c))$) que

$$\lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(I(P_\beta, \mu_\beta)) = -c$$

avec $c = \max_{A \subset S} (D(A) - \max\{U(A); U(A^c)\})$, où

$$\begin{aligned} D(A) &= \min_{x \in A, y \notin A} \min\{U(x) + V(x, y); U(y) + V(y, x)\} \\ U(A) &= \min_{x \in A} U(x) \end{aligned}$$

D'après l'encadrement (2.1), il apparaît donc que

$$(4.2) \quad \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(P_\beta, \mu_\beta)) = -c$$

ce qui en fournit une preuve alternative, peut-être plus intuitive que l'argument usuel par choix de chemins pour aboutir à des inégalités de Poincaré (voir par exemple Holley et Stroock (1988), Götze (1991), Miclo (1992) et Deuschel et Mazza (1994)). En remplaçant des considérations de chemins par une approche plus ensembliste, on peut également retrouver très simplement une caractérisation de Trounev de l'énergie virtuelle (cf. Trounev (1996a)), ce que nous allons maintenant faire, bien que cela ne soit pas lié au théorème 2.1, car cela aidera à traiter le cas du temps discret à la fin de cet appendice.

L'expression donnée pour $D(A)$ fait clairement apparaître que pour tout $A \in \mathcal{S}$, $D(A) = D(A^c)$, mais remarquons que l'on a aussi

$$(4.3) \quad D(A) = \min_{x \in A, y \notin A} U(x) + V(x, y)$$

En effet, on a même plus précisément, pour tout $\beta \geq 0$ et tout $A \in \mathcal{S}$,

$$(4.4) \quad \mu_\beta \left(\mathbf{1}_A \left(\frac{Q_\beta + Q_\beta^*}{2} \right) (\mathbf{1}_{A^c}) \right) = \mu_\beta(\mathbf{1}_A Q_\beta(\mathbf{1}_{A^c})) = \mu_\beta(\mathbf{1}_{A^c} Q_\beta(\mathbf{1}_A))$$

et (4.3) en découle en prenant le logarithme, en divisant par β , puis en passant à la limite pour β grand.

L'équation (4.4) montre également que l'on aurait aussi pu prendre pour $I(P_\beta, \mu_\beta)$ l'une des deux définitions équivalentes suivantes,

$$\begin{aligned} I(P_\beta, \mu_\beta) &= \inf_{A \in \mathcal{S}, 0 < \mu_\beta(A) \leq 1/2} \frac{\mu_\beta(\mathbf{1}_A Q_\beta(\mathbf{1}_{A^c}))}{\mu_\beta(A)} \\ I(P_\beta, \mu_\beta) &= \inf_{A \in \mathcal{S}, 0 < \mu_\beta(A) \leq 1/2} \frac{\mu_\beta(\mathbf{1}_{A^c} Q_\beta(\mathbf{1}_A))}{\mu_\beta(A)} \end{aligned}$$

ce qui peut encore s'énoncer en disant que la constante isopérimétrique associée à un couple (Q, μ) seulement supposé invariant n'est autre que la constante isopérimétrique associée au couple réversible symétrisé additivement $((Q + Q^*)/2, \mu)$.

Pour prouver (4.4), il suffit évidemment de montrer sa dernière égalité, et pour cela on écrit (en laissant tomber les indices β , qui ne jouent aucun rôle ici),

$$\begin{aligned} \mu(\mathbf{1}_A Q(\mathbf{1}_{A^c})) &= \mu(\mathbf{1}_A Q(\mathbf{1} - \mathbf{1}_A)) \\ &= \mu(\mathbf{1}_A) - \mu(\mathbf{1}_A Q(\mathbf{1}_A)) \\ &= \mu(Q(\mathbf{1}_A)) - \mu(\mathbf{1}_A Q(\mathbf{1}_A)) \\ &= \mu((\mathbf{1} - \mathbf{1}_A) Q(\mathbf{1}_A)) \\ &= \mu(\mathbf{1}_{A^c} Q(\mathbf{1}_A)) \end{aligned}$$

Pour $A \in \mathcal{S}$, notons

$$H(A) = D(A) - U(A)$$

qui est parfois appelé la hauteur de sortie de A (relativement à la fonction de coût V), on retrouve donc que

$$c = \min_{A \in \mathcal{S}} \min\{H(A); H(A^c)\}$$

(hauteur maximum des « puits secondaires ») et cette constante peut aussi se décrire de manière plus traditionnelle en termes de l'élévation associée à V : pour $x, y \in S$, notons $\mathcal{C}_{x,y}$ l'ensemble des chemins allant de x à y (i.e. des suites finies $(r_i)_{0 \leq i \leq m}$ d'éléments de S telles que $r_0 = x$, $r_m = y$ et satisfaisant $V(r_i, r_{i+1}) < +\infty$ pour tout $0 \leq i \leq m-1$), et si $r = (r_i)_{1 \leq i \leq m} \in \mathcal{C}_{x,y}$, son élévation est définie par $e(r) = \max_{0 \leq i \leq m-1} U(r_i) + V(r_i, r_{i+1})$. On pose ensuite

$$H(x, y) = \min_{r \in \mathcal{C}_{x,y}} e(r)$$

(en convenant que $H(x, x) = U(x)$ pour $x \in S$) l'élévation minimum pour aller de x à y , et on vérifie alors sans difficulté que

$$c = \max_{x, y \in S} H(x, y) - U(x) - U(y)$$

Il est intéressant de noter que les considérations précédentes fournissent en fait des caractérisations de la probabilité invariante et de l'énergie virtuelle, ainsi par exemple on a :

PROPOSITION 4.2. *Soit Q un noyau markovien sur un espace mesurable quelconque (S, \mathcal{S}) et μ une probabilité telle que l'on ait l'équilibre des flux :*

$$\forall A \in \mathcal{S}, \quad \mu(\mathbf{1}_A Q(\mathbf{1}_{A^c})) = \mu(\mathbf{1}_{A^c} Q(\mathbf{1}_A))$$

alors μ est invariante pour P .

Preuve. Il suffit de remarquer comme précédemment que pour tout $A \in \mathcal{S}$, on a d'une part

$$\mu(\mathbf{1}_A) = \mu(\mathbf{1}_A Q(\mathbf{1}_{A^c})) + \mu(\mathbf{1}_A Q(\mathbf{1}_A))$$

et d'autre part

$$\mu(Q(\mathbf{1}_A)) = \mu(\mathbf{1}_{A^c} Q(\mathbf{1}_A)) + \mu(\mathbf{1}_A Q(\mathbf{1}_A))$$

□

Ce résultat est sans intérêt en lui-même, néanmoins en revenant à S fini, si on rajoute des β en indice et si on ne considère que ce qui se passe dans une échelle logarithmique pour β grand, on obtient une heuristique pour la caractérisation suivante de l'énergie virtuelle :

Soit $\tilde{U} : S \rightarrow \mathbb{R}_+$ une fonction arbitraire satisfaisant seulement $\min_S \tilde{U} = 0$. Par analogie avec ce qui précède, posons pour $A \in \mathcal{S}$,

$$D_{\tilde{U}}(A) = \min_{x \in A, y \notin A} \tilde{U}(x) + V(x, y)$$

PROPOSITION 4.3. *Supposons que pour tout $A \in \mathcal{S} \setminus \{\emptyset, S\}$, on ait*

$$D_{\tilde{U}}(A) = D_{\tilde{U}}(A^c) < +\infty$$

alors $\tilde{U} = U$.

Remarquons que la finitude de $D_{\tilde{U}}(A)$ pour tout $A \in \mathcal{S} \setminus \{\emptyset, S\}$ est équivalente à l'irréductibilité du noyau de communication a priori Q .

Preuve. Pour $\beta \geq 0$, soit ν_β la probabilité sur S définie par

$$\forall x \in S, \quad \nu_\beta(x) = \tilde{Z}_\beta^{-1} \exp(-\beta \tilde{U}(x))$$

avec $\tilde{Z}_\beta = \sum_{y \in S} \exp(-\beta \tilde{U}(y))$, et notons $h_\beta = d\mu_\beta/d\nu_\beta$. Par compacité de $[0, +\infty]^S$, soit $(\beta_n)_{n \in \mathbb{N}}$ une suite de réels positifs telle que $\lim_{n \rightarrow \infty} \beta_n = +\infty$ et telle que h_{β_n} converge vers une certaine fonction $h : S \rightarrow \overline{\mathbb{R}}_+$.

Soit $A = \{x \in S / h(x) = 0\}$, par l'absurde supposons que $A \neq \emptyset$ et écrivons que

$$\begin{aligned} & \exp(\beta_n D_{\tilde{U}}(A)) \mu_{\beta_n}(\mathbf{1}_A P_{\beta_n}(\mathbf{1}_{A^c})) \\ &= \tilde{Z}_{\beta_n}^{-1} \sum_{x \in A, y \notin A} h_{\beta_n}(x) q(x, y) \exp(-\beta_n [\tilde{U}(x) + V(x, y) - D_{\tilde{U}}(A)]) \end{aligned}$$

expression qui converge vers 0 pour n grand, par définition de A , de $D_{\tilde{U}}(A)$ et du fait que $\lim_{\beta \rightarrow +\infty} \tilde{Z}_\beta = \text{card}(\{x \in S / \tilde{U}(x) = 0\}) > 0$.

Par ailleurs,

$$\begin{aligned} & \exp(\beta_n D_{\tilde{U}}(A)) \mu_{\beta_n}(\mathbf{1}_{A^c} P_{\beta_n}(\mathbf{1}_A)) \\ &= \tilde{Z}_{\beta_n}^{-1} \sum_{x \notin A, y \in A} h_{\beta_n}(x) q(x, y) \exp(-\beta_n [\tilde{U}(x) + V(x, y) - D_{\tilde{U}}(A)]) \end{aligned}$$

or en considérant un couple $(x, y) \in A^c \times A$ où le minimum définissant $D_{\tilde{U}}(A^c) = D_{\tilde{U}}(A)$ est atteint, on constate que la liminf pour n grand de la quantité précédente est strictement positive (éventuellement même infinie). On obtient ainsi une contradiction qui montre qu'en fait $A = \emptyset$, i.e. $h > 0$. En considérant ensuite $A = \{x \in S / h(x) < +\infty\}$, on prouve d'une manière similaire que $h < +\infty$. On en déduit ensuite que pour tout $x \in S$,

$$0 < \liminf_{\beta \rightarrow +\infty} h_\beta(x) \leq \limsup_{\beta \rightarrow +\infty} h_\beta(x) < +\infty$$

(en fait $\lim_{\beta \rightarrow +\infty} h_\beta(x)$ existe) puis le résultat annoncé. \square

Rappelons maintenant la caractérisation donnée par Trounev (cf. la proposition 2.17 de l'article de Trounev (1996a)) de l'énergie virtuelle U par symétrie des élévations. Soit toujours $\tilde{U} : S \rightarrow \mathbb{R}_+$ une fonction telle que $\min_S \tilde{U} = 0$, on pose pour $r = (r_i)_{0 \leq i \leq m} \in \mathcal{C}_{x,y}$ et $x, y \in S$,

$$e_{\tilde{U}}(r) = \max_{0 \leq i \leq m-1} \tilde{U}(r_i) + V(r_i, r_{i+1})$$

puis

$$H_{\tilde{U}}(x, y) = \min_{r \in \mathcal{C}_{x,y}} e_{\tilde{U}}(r)$$

(toujours en convenant que $H_{\tilde{U}}(x, x) = \tilde{U}(x)$, pour tout $x \in S$).

Rappel 4.4. *Supposons Q irréductible, on a équivalence entre*

$$(i) \quad \forall x, y \in S, \quad H_{\tilde{U}}(x, y) = H_{\tilde{U}}(y, x)$$

$$(ii) \quad \tilde{U} = U.$$

La preuve originale est relativement complexe et est basée sur une récurrence sur le cardinal de S et des décompositions en cycles. La démonstration que nous proposons est plus « globale », ce qui nous laisse l'espoir de pouvoir la généraliser à d'autres situations. Elle repose sur le lemme suivant, qui peut être vu comme une simple réécriture de la condition (i).

Lemme 4.5. *Soit \tilde{U} comme ci-dessus, on a équivalence entre*

$$\begin{aligned} (i) \quad & \forall x, y \in S, \quad H_{\tilde{U}}(x, y) = H_{\tilde{U}}(y, x) \\ (ii) \quad & \forall A \in \mathcal{S} \setminus \{\emptyset, S\}, \quad D_{\tilde{U}}(A) = D_{\tilde{U}}(A^c). \end{aligned}$$

Preuve. Supposons (i) satisfait, soit $A \in \mathcal{S} \setminus \{\emptyset, S\}$, $x \in A$ et $y \notin A$ tels que $D_{\tilde{U}}(A) = \tilde{U}(x) + V(x, y)$. Considérons $r = (r_i)_{0 \leq i \leq m} \in \mathcal{C}_{y,x}$ tel que $e_{\tilde{U}}(r) = H(y, x) = H(x, y) \leq \tilde{U}(x) + V(x, y)$. Il existe $0 \leq i \leq m-1$ tel que $r_i \in A^c$ et $r_{i+1} \in A$, et on a alors $D_{\tilde{U}}(A^c) \leq \tilde{U}(r_i) + V(r_i, r_{i+1}) \leq e_{\tilde{U}}(r)$. Ainsi on a montré que pour tout $A \in \mathcal{S} \setminus \{\emptyset, S\}$ $D_{\tilde{U}}(A^c) \leq D_{\tilde{U}}(A)$, ce qui implique en fait (ii).

Réciproquement supposons (ii) satisfait, soit $x \neq y \in S$, puis posons $A = \{z \in S / H_{\tilde{U}}(y, z) \leq H_{\tilde{U}}(x, y)\} \cup \{y\}$. Par l'absurde supposons que $x \notin A$, ce qui nous assure que si $r \in \mathcal{C}_{x,y}$ alors $e_{\tilde{U}}(r) \geq D_{\tilde{U}}(A^c)$ et ainsi il apparaît que $D_{\tilde{U}}(A) = D_{\tilde{U}}(A^c) \leq H_{\tilde{U}}(x, y)$. Mais soit $z \in A$ et $z' \in A^c$ tels que $U(z) + V(z, z') = D_{\tilde{U}}(A)$, on aboutit à la contradiction que $H_{\tilde{U}}(z, z') \leq H_{\tilde{U}}(x, y)$, d'où finalement $x \in A$, et le résultat annoncé s'en déduit aisément. \square

Pour terminer, considérons la situation du temps discret ; de manière similaire au début de cet appendice, on dispose d'une évolution $\beta : \mathbb{N} \rightarrow \mathbb{R}_+$ et on s'intéresse aux chaînes de Markov inhomogènes dont le noyau de probabilités de transition à tout instant $n \in \mathbb{N}$, est Q_{β_n} . Pour étudier ces algorithmes, on est amené à évaluer plutôt le comportement pour β grand de $\beta^{-1} \ln(\lambda(Q_{\beta}^{r*} Q_{\beta}^r, \mu_{\beta}))$, où $r \geq 1$ est un entier, on aura remarqué que le noyau markovien $Q_{\beta}^{r*} Q_{\beta}^r$ est bien réversible par rapport à μ_{β} , il est parfois appelé le symétrisé multiplicatif de Q_{β}^r . Plus précisément, on a montré dans Miclo (1997) en considérant des chemins, que la limite suivante existe pour tout $r \geq 1$,

$$(4.5) \quad c_r = - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(Q_{\beta}^r Q_{\beta}^{*r}, \mu_{\beta})) \in [0, +\infty]$$

et surtout, du moins si V ne prend pas que les valeurs 0 ou $+\infty$, ce que l'on supposera désormais (sinon on est dans le cas homogène où Q_{β} ne dépend pas de $\beta \geq 0$, et il peut alors se poser des problèmes de périodicité), qu'il existe $r_0 \leq \text{card}(S) - 1$ tel que pour tout $r \geq r_0$, $c_r = c$.

En s'inspirant de ce qui précède, on peut retrouver plus simplement ces résultats : tout d'abord prolongeons V sur $S \times S$ en posant

$$\forall x \in S, \quad V(x, x) = - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(Q_{\beta}(x, x))$$

(ainsi $V(x, x) = 0$ dès que $Q(x, x) > 0$ ou s'il existe $y \neq x$ tel que $0 < V(x, y) < +\infty$, et $V(x, x) = +\infty$ sinon). On pose ensuite pour $r \geq 0$,

$$\forall x, y \in S, \quad V_r(x, y) = \min_{x_1, \dots, x_{r-1} \in S} V(x, x_1) + V(x_1, x_2) + \dots + V(x_{r-1}, y)$$

de sorte que pour tous $x, y \in S$ et tout $r \geq 1$,

$$\lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(Q_\beta^r(x, y)) = -V_r(x, y)$$

En utilisant alors que pour tout $A \subset S$, $\mu_\beta(\mathbf{1}_A Q_\beta^{r*} Q_\beta^r(\mathbf{1}_{A^c})) = \mu_\beta(Q_\beta^r(\mathbf{1}_A) Q_\beta^r(\mathbf{1}_{A^c}))$, on obtient que

$$\begin{aligned} & - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(I(Q_\beta^{r*} Q_\beta^r, \mu_\beta)) \\ & = \max_{\emptyset \neq A \subset S, A \neq S} \min_{x \in S, y \in A, z \in A^c} U(x) + V_r(x, y) + V_r(x, z) - U(A) \vee U(A^c) \end{aligned}$$

ce qui par le biais du théorème 2.1 montre que (4.5) existe bien et que c_r vaut l'expression ci-dessus.

Par ailleurs, remarquons, d'une part que chacun des termes $\mu_\beta(Q_\beta^r(\mathbf{1}_A) Q_\beta^r(\mathbf{1}_{A^c}))$ est croissant en $r \in \mathbb{N}^*$ (car $\mu_\beta((Q_\beta^r(\mathbf{1}_A))^2) \leq \mu_\beta(Q_\beta((Q_\beta^{r-1}(\mathbf{1}_A))^2)) = \mu_\beta((Q_\beta^{r-1}(\mathbf{1}_A))^2)$ et $\mu_\beta(Q_\beta^r(\mathbf{1}_A) Q_\beta^r(\mathbf{1}_{A^c})) = \mu_\beta(A) - \mu_\beta((Q_\beta^r(\mathbf{1}_A))^2)$), et d'autre part que pour tout $r \in \mathbb{N}^*$, $c_r \geq c$. En effet, en utilisant la définition de la probabilité invariante, il apparaît que pour tous $x, y \in S$, $U(x) + V(x, y) \geq U(y)$. Pour $x, y \in A$ et $z \in A^c$, soient $x_1, \dots, x_{r-1} \in S$ tels que $V_r(x, z) = V(x, x_1) + V(x_1, x_2) + \dots + V(x_{r-1}, z)$, et considérons un indice $0 \leq i \leq r-1$ tel que $x_i \in A$ et $x_{i+1} \in A^c$ (en convenant que $x_0 = x$), on a alors $U(x) + V_r(x, z) \geq U(x_i) + V(x_i, x_{i+1})$ et $V_r(x, y) \geq 0$, ce qui prouve que

$$\min_{x \in A, y \in A, z \in A^c} U(x) + V_r(x, y) + V_r(x, z) \geq D(A)$$

et d'une manière similaire, on obtient

$$\min_{x \in A^c, y \in A, z \in A^c} U(x) + V_r(x, y) + V_r(x, z) \geq D(A^c) = D(A)$$

dont découle l'inégalité $c_r \geq c$ annoncée.

Pour montrer l'existence d'un $r_0 \leq \text{card}(S) - 1$ tel que $c_r = c$ pour $r \geq r_0$, il suffit donc de voir que pour tout $\emptyset \neq A \subset S$, $A \neq S$ fixé, il existe $r_0(A) \leq \text{card}(S) - 1$ tel que

$$(4.6) \quad \min_{x \in S, y \in A, z \in A^c} U(x) + V_{r_0(A)}(x, y) + V_{r_0(A)}(x, z) \leq D(A)$$

Pour ceci, soit $(x_0, y_0) \in A \times A^c$ tel que $U(x_0) + V(x_0, y_0) = D(A)$ (de manière analogue, on peut considérer le cas où $U(y_0) + V(y_0, x_0) = D(A^c) = D(A)$). Si $V(x_0, y_0) > 0$, on a $V(x_0, x_0) = 0$, ainsi en prenant $r_0(A) = 1$, puis $x = x_0 = y$ et $z = y_0$ dans (4.6), on obtient le résultat voulu.

Sinon, soit $p = (p_i)_{0 \leq i \leq N} \in \mathcal{C}_{y_0, x_0}$ tel que $e(p) = H(y_0, x_0) = H(x_0, y_0) = U(x_0) + V(x_0, y_0) = D(A)$. On supposera de plus que p est injectif: pour tous $0 \leq i \neq j \leq N$, on a $p_i \neq p_j$.

S'il existe $0 \leq i \leq N-1$ tel que $V(p_i, p_{i+1}) > 0$, soit $j = \min\{i \leq k \leq N : \{p_k, p_{k+1}\} \not\subset A \text{ ou } \{p_k, p_{k+1}\} \not\subset A^c\}$, où on a convenu que $p_{N+1} = y_0$. Il suffit alors de prendre $r_0(A) = j - i + 1$ et, suivant les cas, $x = p_i$, $y = p_i$, $z = p_{j+1}$ ou $x = p_i$, $y = p_{j+1}$, $z = p_i$ dans (4.6), pour conclure au résultat escompté.

Si p ne satisfait pas la propriété précédente, notons qu'alors pour tout $0 \leq i \leq N$, $U(p_i) = U(x_0) = U(y_0)$. Ceci nous amène à définir C l'ensemble des $z \in S$ pour lesquels il existe un chemin plat allant de y_0 à z et un

chemin plat allant de z à y_0 , un chemin $q = (q_i)_{0 \leq i \leq N'}$ étant dit plat si pour tout $0 < i \leq N'$, $U(q_i) = U(q_0)$ et $V(q_{i-1}, q_i) = 0$. Notamment tout le chemin p est inclus dans le plateau C . Puisque V ne prend pas que les valeurs 0 ou $+\infty$, il existe nécessairement $x_1 \in C$ et $y_1 \in S$ tels que, soit $0 < V(x_1, y_1) < +\infty$, soit $V(x_1, y_1) = 0$ et $U(x_1) > U(y_1)$. Quitte à modifier p , on peut supposer qu'il passe par x_1 , et soit $0 \leq i \leq N$ tel que $x_i = x_1$ (p n'est plus nécessairement injectif, mais on suppose par contre que le chemin $(p_k)_{i \leq k \leq N}$ l'est). On considère alors l'indice $i \leq j \leq N$ défini comme précédemment. Si $V(x_1, y_1) > 0$, on procède de la même manière que ci-dessus pour se rendre compte que l'on peut prendre $r_0(A) = j - i + 1$. Si par contre on a $V(x_1, y_1) = 0$ et $U(y_1) < U(x_1)$, distinguons encore deux possibilités :

- Si x_1 et y_1 n'appartiennent pas tous les deux soit à A soit à A^c , on peut alors reprendre les considérations précédentes avec $x_0 = x_1$ et $y_0 = y_1$ pour voir qu'il existe un $r_0(A) \leq \text{card}(S) - 1$ qui convient.

- Si x_1 et y_1 appartiennent tous les deux soit à A soit à A^c , construisons une suite finie y_2, \dots, y_{j-i+1} telle que pour tout $1 \leq k \leq j - i$, $V(y_k, y_{k+1}) = 0$ (c'est possible car pour tout $z' \in S$, il existe au moins un $z'' \in S$ tel que $V(z', z'') = 0$). Suivant que $x_1 \in A$ ou $x_1 \in A^c$, tous les y_1, \dots, y_{j-i+1} restent dans A ou dans A^c . Il apparaît donc en considérant $x = x_1, y = y_{j-i+1}, z = p_{j+1}$ ou $x = x_1, y = p_{j+1}, z = y_{j-i+1}$, dans (4.6), que $r_0(A) = j - i + 1$ convient.

En étant un peu plus soigneux, les arguments précédents permettent d'obtenir un r_0 qui convient de la manière suivante : on appelle point plat un élément x de S tel que $V(x, x) = +\infty$ et tel que pour tout $y \in S$, $V(x, y) = 0$ implique que $U(y) = U(x)$. Un chemin injectif $p = (p_i)_{0 \leq i \leq N}$, avec $N \geq 2$, sera dit un pont, si p_0 n'est pas un point plat, si $V(p_0, p_1) = U(p_1) - U(p_0)$, si pour tout $1 \leq i \leq N - 2$, p_i est un point plat, et si $V(p_{N-1}, p_N) = +\infty$. Un chemin $p = (p_0, p_1)$ de longueur 1 sera quant à lui dit un pont, si $V(p_0, p_0) = 0$. On note \mathcal{C} l'ensemble des couples (x, y) de points $x \neq y \in S$ pour lesquels il existe un pont p les reliant et satisfaisant $e(p) = H(x, y)$. On définit ensuite $\bar{d}(x, y)$ comme étant la plus petite longueur d'un tel chemin. On peut alors prendre, toujours si V ne prend pas que les valeurs 0 ou $+\infty$,

$$r_0 = \max_{(x,y) \in \mathcal{C}} \bar{d}(x, y)$$

ce qui est une estimée de $\min\{r \geq 1 : c_r = c\}$ différente, et souvent un peu meilleure, de celle présentée dans Miclo (1997).

REMERCIEMENTS

Je tiens tout particulièrement à exprimer ma gratitude envers Laurent Saloff-Coste, qui m'a expliqué les inégalités de Cheeger et m'a fourni de nombreuses références à leur sujet.

RÉFÉRENCES

- R. Ayoub. *An introduction to the analytic theory of numbers*. Mathematical Surveys 10. American Mathematical Society, 1963.

- J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacien. In R. C. Gunning, editor, *Problems in Analysis: A Symposium in Honor of S. Bochner*, pages 195–199. Princeton University Press, 1970.
- J.-D. Deuschel et C. Mazza. L^2 convergence of time nonhomogeneous Markov processes: I. spectral estimates. *The Annals of Applied Probability*, 4(4):1012–1056, 1994.
- P. Diaconis. *Group Representations in Probability and Statistics*. Lecture Notes-Monograph Series 11. Institute of Mathematical Statistics, 1988.
- P. Diaconis et D. Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.
- M.I. Freidlin et A.D. Wentzell. *Random Perturbations of Dynamical Systems*. A Series of Comprehensive Studies in Mathematics 260. Springer-Verlag, 1984.
- F. Götze. Rate of convergence of simulated annealing processes. Préprint de l'Université Bielefeld, Décembre 1991.
- R. Holley et D. Stroock. Simulated annealing via Sobolev inequalities. *Communications in Mathematical Physics*, 115:553–569, 1988.
- G. Lawler and A. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Transactions of the American mathematical society*, 309(2):557–580, October 1988.
- P. Mathieu. Hitting times and spectral gap inequalities. *Annales de l'Institut Henri Poincaré*, 33(4):437–465, 1997.
- L. Miclo. Recuit simulé sans potentiel sur un ensemble fini. In J. Azéma, P.A. Meyer, and M. Yor, editors, *Séminaire de Probabilités XXVI*, Lecture Notes in Mathematics 1526, pages 47–60. Springer-Verlag, 1992.
- L. Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In J. Azéma, M. Emery, and M. Yor, editors, *Séminaire de Probabilités XXXI*, Lecture Notes in Mathematics 1655, pages 136–167. Springer-Verlag, Berlin, 1997.
- B. Mohar. Isoperimetric numbers of graphs. *Journal of combinatorial theory, Series B*, 47:274–291, 1989.
- T. Pignataro and D. Sullivan. Ground state and lowest eigenvalue of the Laplacian for non-compact hyperbolic surfaces. *Communication in Mathematical Physics*, 104:529–535, 1986.
- L. Saloff-Coste. Lectures on finite Markov chains. In P. Bernard, editor, *Lectures on Probability Theory and Statistics. Ecole d'Été de Probabilités de Saint-Flour XXVI-1996*, Lecture Notes in Mathematics 1665. Springer-Verlag, Berlin, 1997.
- A. Trouvé. Cycle decompositions and simulated annealing. *Society for Industrial and Applied Mathematics. Journal on Control and Optimization*, 34(3):966–986, May 1996.
- A. Trouvé. Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 32(3):299–348, 1996.

CNRS, UMR C5583, LABORATOIRE DE STATISTIQUE ET PROBABILITÉS, UNIVERSITÉ
PAUL SABATIER, 118, ROUTE DE NARBONNE, 31062 TOULOUSE CEDEX, FRANCE
E-mail address: `miclo@cict.fr`