

Inégalité de log-Sobolev pour les chaînes de Markov réversibles

Lucas Babet et Elena Dorbec

Mémoire 1A

Introduction

Dans ce mémoire, nous étudions la vitesse de convergence des chaîne de Markov et, plus particulièrement les majorations qu'on peut obtenir sur leur temps de mélange t_{ml} . L'objectif de notre étude est de montrer que l'inégalité de log-Sobolev modifiée (ILSM) implique celle de log-Sobolev (ILS) dans le cas réversible. Ce résultat très récent démontré par Justin Salez, Konstantin Tikhomirov et Pierre Youssef dans l'article [5] est intéressant car il permet de lier la contraction de l'entropie due à l'inégalité de log-Sobolev modifiée au phénomène d'hypercontractivité.

Aussi, nous commencerons ici par introduire le contexte d'étude et donc par faire une présentation générale des chaîne de Markov et de l'étude de leurs vitesse de convergence — et plus particulièrement, sur le contrôle de leur temps de mélange. Nous verrons ainsi qu'alors que l'inégalité de Poincaré (IP) permet un contrôle de la convergence en variation totale, celle de log-Sobolev modifiée fournit un contrôle L^1 et celle de log-Sobolev un contrôle L^2 voire L^∞ . Nous présenterons également une interprétation de ces trois inégalités en terme de contractions : contraction de la variance pour l'inégalité de Poincaré, contraction de l'entropie pour l'inégalité de log-Sobolev modifiée et hypercontractivité pour l'inégalité de log-Sobolev. Cette vision des inégalités permettra de comprendre précisément leurs effets et leurs dépendances classiques.

En effet, on a dans le cas général la chaîne d'implications suivante :

$$(ILS) \implies (ILSM) \implies (IP).$$

Nous présenterons donc la preuve de ce résultat au fil de notre étude. Enfin, nous pourrions conclure sur la preuve de l'implication réciproque dans le cas réversible :

$$(ILSM) \implies (ILS)$$

Nous verrons aussi qu'on peut étendre ce contrôle à des matrices non réversibles par l'ajout d'un coefficient de comparaison.

1 Cadre de l'étude

Tous les résultats présentés ci-dessous sont tirés du cours [1] de Justin Salez et des enseignements d'Anna Ben-Hamou.

1.1 Introduction aux chaînes de Markov

Définition 1.1. Tout au long de ce rapport, nous considérons (\mathcal{X}, P, μ_0) une **chaîne de Markov**, c'est-à-dire que

- \mathcal{X} est un ensemble *fini* non vide appelé **espace des états** ;
- $P : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ une **matrice de transition** vérifiant :

$$\forall x \in \mathcal{X}, \sum_{y \in \mathcal{X}} P(x, y) = 1;$$

(P est stochastique)

- μ_0 est une distribution de probabilités sur \mathcal{X} appelée **loi initiale**.

On peut se représenter toute chaîne de Markov par un graphe orienté avec des probabilités sur les branches correspondant à une probabilité de déplacement.

Nous nous intéressons alors à la suite : $(\mu_n)_{(n \in \mathbb{N})}$ avec μ_0 définie ci-dessus et :

$$\forall n \in \mathbb{N}, \mu_{n+1} = \mu_n P.$$

Cette suite correspond donc à des probabilités successives de se trouver en chacun des sommets en partant d'une configuration initiale.

On peut même définir une version continue de cette suite en notant pour tout réel $t : \bar{\mu}_t = \mu_0 H_t$ où

$$H_t = \exp(t(P - I)).$$

On verra la pertinence de cette définition dans les propriétés qui suivront.

Comme on étudie la convergence de la suite μ , on est naturellement amenés à considérer la quantité suivante.

Définition 1.2. Une **mesure stationnaire** de (\mathcal{X}, P, μ_0) , qu'on notera π est une mesure de probabilités telle que $\pi P = \pi$.

Cette mesure correspond donc à une situation stable par l'action de P , donc une limite potentielle de μ .

Pour étudier la vitesse de convergence, on aura tendance à comparer l'effet de la chaîne de Markov (\mathcal{X}, P, μ_0) à une chaîne de Markov rendant la convergence instantanée, ce qui motive la définition suivante.

Définition 1.3. Pour toute chaîne de Markov (\mathcal{X}, P, μ_0) admettant une mesure stationnaire π , on notera Π la matrice telle que pour tout $x, y \in \mathcal{X}$, $\Pi(x, y) = \pi(y)$.

Remplacer P par cette matrice dans la chaîne de Markov revient donc à atteindre une situation stable en une étape.

Se pose alors la question de l'existence d'une telle mesure stationnaire, d'où la proposition suivante.

Proposition 1.4. *Toute chaîne de Markov admet une mesure stationnaire*

Démonstration. Par compacité, la suite $(\frac{1}{n+1} \sum_{k=0}^n \mu_0 P^k)_{(n \in \mathbb{N})}$ est une sous-suite convergente dont la limite convient. \square

Cette mesure stationnaire est un élément particulièrement intéressant dans notre étude puisqu'on regarde le comportement asymptotique de la suite μ , c'est pourquoi on étudie des chaînes de Markov particulières permettant une meilleure connaissance de la mesure stationnaire et donc de cette limite.

Définition 1.5. Une chaîne de Markov est dite **irréductible** si

$$\forall x, y \in \mathcal{X}, \exists t \in \mathbb{N}, P^t(x, y) > 0.$$

Dans une chaîne de Markov irréductible, tout état est atteignable depuis tout autre état : le graphe correspondant est connexe. On peut alors facilement imaginer à partir d'une chaîne de Markov quelconque se restreindre à une chaîne de Markov réductible : il s'agit de considérer seulement les composantes connexes du graphe et la matrice extraite correspondante.

Remarque. Si une chaîne de Markov est irréductible, chaque état est asymptotiquement accessible donc chacune de ses mesures stationnaires n'admet aucun coefficient nul.

Cette propriété permet même d'obtenir le résultat suivant.

Proposition 1.6. *Si (\mathcal{X}, P, μ_0) est irréductible, alors elle admet une unique mesure stationnaire.*

Si elle converge, $(\mu_n)_{n \in \mathbb{N}}$ converge alors nécessairement vers π .

Démonstration. Soient π_1 et π_2 deux mesures stationnaires.

L'irréductibilité de (\mathcal{X}, P, μ_0) impose que ces mesures n'aient aucun coefficient nul. Alors, le théorème des valeurs intermédiaires appliqué à m définie sur $[0, 1]$ par $m(t) = \min_{x \in \mathcal{X}}(\pi_1(x) - t\pi_2(x))$ permet de construire une troisième mesure stationnaire telle qu'un de ses coefficients soit nul, d'où le résultat par l'absurde. \square

Pour un résultat encore plus fort, on regarde maintenant une classe de matrices imposant la convergence de la suite μ .

Définition 1.7. Une chaîne de Markov est dite **ergodique** si

$$\exists t \in \mathbb{N}, \forall x, y \in \mathcal{X}, \mathbb{N}, P^t(x, y) > 0.$$

Il existe alors une longueur de chemin permettant de relier n'importe quel couple d'états. Une chaîne de Markov ergodique est donc en particulier irréductible.

Proposition 1.8. *Si (\mathcal{X}, P, μ_0) est ergodique, alors $\mu_{n \in \mathbb{N}}$ converge vers π .*

Nous présenterons la preuve de ce résultat lorsque nous verrons la proposition 1.19.

Le fait d'être ergodique semble être une contrainte très forte, mais on peut en fait relaxer cette contrainte en considérant un autre type de chaîne de Markov.

Définition 1.9. Une chaîne de Markov est dite **paresseuse** si

$$\forall x \in \mathcal{X}, P(x, x) \geq \frac{1}{2}.$$

Il y a alors une grande probabilité que le système ne se modifie pas d'un instant à l'autre.

Remarque. Si (\mathcal{X}, P, μ_0) est une chaîne de Markov, remplacer P par $\frac{P+I}{2}$ permet de la rendre paresseuse. Cela revient à imaginer qu'on lance une pièce et qu'on ne bouge selon la loi de probabilité de la chaîne de Markov (\mathcal{X}, P, μ_0) seulement si on tombe sur pile et qu'on reste immobile sinon. Elles auront alors évidemment la même loi stationnaire.

En outre, on peut remarquer que rendre une matrice irréductible paresseuse par ce même procédé la rend également ergodique ! En effet, on permet alors à la suite de faire des pauses sur son trajet le long du graphe. L'existence d'un chemin entre chaque couple de sommets revient alors à l'existence d'un chemin de même longueur entre tous les couples de sommets : il suffit d'attendre une fois arrivé à destination.

On a donc remarqué que la suite μ convergeait sous des contraintes finalement assez faibles.

Comme on veut étudier le comportement asymptotique de la suite, on s'intéresse maintenant à la probabilité asymptotique de passer de l'état x à l'état y . On est alors amenés à considérer la quantité $\pi(x)P(x, y)$ qui rend qualitativement compte du flux asymptotique porté par chaque arrête du graphe. Cette étude motive la définition suivante.

Définition 1.10. Si (\mathcal{X}, P, μ_0) est irréductible, on définit P^* l'**adjoint** de P par :

$$P^*(x, y) = \frac{\pi(y)}{\pi(x)} P(y, x).$$

Cet adjoint a la même mesure stationnaire que P et correspond à une inversion des sens des flux asymptotiques.

Sur l'espace $L^2_{\mathbb{C}}(\mathcal{X}, \mathbb{C})$ on a en fait défini le produit scalaire suivant :

$$\langle f, g \rangle = \sum_{x \in \mathcal{X}} f(x) \overline{g(x)} \pi(x).$$

Naturellement, comme on parle d'adjoint avec un produit scalaire fixé, on est amené à considérer les matrices auto-adjointes.

Définition 1.11. Une chaîne de Markov est dite **réversible** si

$$\forall (x, y) \in \mathcal{X}^2, \pi(x)P(x, y) = \pi(y)P(y, x).$$

La matrice est alors auto-adjointe pour le produit scalaire défini ci-dessus. Cette propriété signifie que le flux à travers chaque arrête est asymptotiquement symétrique.

Remarque. Si (\mathcal{X}, P, μ_0) est une chaîne de Markov, remplacer P par $\frac{P+P^*}{2}$ permet de la rendre réversible. La distribution stationnaire reste alors la même.

Proposition 1.12. *Si une chaîne de Markov (\mathcal{X}, P, μ_0) est irréductible et réversible, P est diagonalisable.*

On note alors (λ_i) les valeurs propres ordonnées dans l'ordre décroissant et (φ_i) une base orthonormée pour le produit scalaire sus-défini de vecteurs propres associés.

Démonstration. On utilise ici le théorème spectral. □

Remarque. Comme P est stochastique, on a toujours nécessairement $\lambda_1 = 1$.

1.2 Notations et objets utiles

On liste ici les différentes distances définies sur les mesures de probabilités qu'on utilisera pour étudier la vitesse de convergence de la suite μ .

On notera donc $\langle \cdot, \cdot \rangle$ le produit scalaire défini ci-dessus :

$$\forall (f, g) \in L_{\mathbb{C}}^2(\mathcal{X}, \mathbb{C})^2, \langle f, g \rangle = \sum_{x \in \mathcal{X}} f(x) \overline{g(x)} \pi(x)$$

Définition 1.13. Aussi, sur l'espace des mesures de probabilités sur \mathcal{X} , on définit la **norme p relativement à π** (et donc la distance associée) comme suit :

$$\|\mu\|_p = \left(\sum_{x \in \mathcal{X}} \mu(x)^p \pi(x) \right)^{\frac{1}{p}}.$$

On aura alors bien pour toute distribution μ : $\langle \mu, \mu \rangle = \|\mu\|_1^2 := \|\mu\|^2$

Définition 1.14. Pour deux mesures de probabilité μ et ν sur \mathcal{X} , on définit la **distance en variation totale** comme suit :

$$d_{\text{TV}}(\mu, \nu) = \sup \{ \mu(A) - \nu(A), A \subset \mathcal{X} \}.$$

Cette distance correspond donc à la plus grande distance dans \mathbb{R} observable entre deux mesures de probabilités.

Proposition 1.15.

d_{TV} est convexe et, pour toutes mesures μ, ν , on a :

$$d_{\text{TV}}(\mu P, \nu P) \leq d_{\text{TV}}(\mu, \nu).$$

On a en outre :

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sup_{f: \mathcal{X} \rightarrow [-1, 1]} |\mu(f) - \nu(f)|.$$

Ces résultats seront utiles pour des calculs que nous effectuerons plus tard.

Pour manipuler efficacement les probabilités, on définit les quantités suivantes.

Définition 1.16. Pour toute fonction $f : \mathcal{X} \rightarrow \mathbb{R}$ notons :

- $\mathbb{E}(f) = \sum_{x \in \mathcal{X}} f(x)\pi(x)$ l'espérance de f ;
- $\mathbb{V}(f) = \mathbb{E}(f^2) - \mathbb{E}(f)^2$ sa variance ;
- $\text{Ent}(f) = \sum_{x \in \mathcal{X}} f(x) \log(f(x))\pi(x)$ son entropie.

Définition 1.17. On appelle **forme de Dirichlet** de deux fonctions $f, g : \mathcal{X} \rightarrow \mathbb{R}$ la quantité :

$$\mathcal{E}_P(f, g) = \sum_{x, y \in \mathcal{X}} (f(x) - f(y))(g(x) - g(y))\pi(x)P(x, y).$$

On interprète cette quantité comme une covariance locale entre deux fonctions relativement au flux asymptotique.

1.3 Étude de la vitesse de convergence des chaînes de Markov

Définition 1.18. Si la distribution stationnaire est unique, on appelle **distance à l'équilibre** d'une chaîne de Markov la fonction définie par :

$$\mathcal{D}_P(t) = \max_{\nu \in \mathcal{P}(\mathcal{X})} d_{\text{TV}}(\nu P^t, \pi).$$

Cette fonction quantifie en effet l'écart maximal qu'on peut avoir avec la distribution stationnaire après t itérations.

Proposition 1.19. On a pour toute chaîne de Markov ergodique (\mathcal{X}, P, μ_0) :

$$\mathcal{D}_P(t)^{\frac{1}{t}} \xrightarrow{t \rightarrow +\infty} \max \{|\lambda|, \lambda \in Sp(P) \setminus \{1\}\}$$

Cette convergence implique donc directement la convergence de μ vers π dans le cas ergodique, et même une décroissance géométrique de la suite !

Démonstration.

Par un calcul que nous n'écrivons pas ici, on montre que $(2\mathcal{D}_P(t))_{t \in \mathbb{N}}$ est sous-multiplicative. Le lemme de Fekete permet alors d'obtenir la convergence suivante :

$$(\mathcal{D}_P(t))^{\frac{1}{t}} \rightarrow \inf \left\{ \mathcal{D}_P(t)^{\frac{1}{t}} \mid t \in \mathbb{N} \right\}.$$

□

Le principe du min-max de Courant-Fisher appliqué à $P - \Pi$ nous permet ainsi de caractériser cette limite par un critère variationnel.

Définition 1.20.

On appelle **temps de mélange** d'une chaîne de Markov la quantité :

$$t_{\text{ml}}^\varepsilon = \inf \{ t > 0, \mathcal{D}_p(t) < \varepsilon \};$$

Dans le cas continu, on définit de même :

$$t_{\text{ml}}^\varepsilon = \inf \{ t > 0, \max_{\mu}(\bar{\mu}(t)) < \varepsilon \}.$$

Sans spécification, on prendra $\varepsilon = \frac{1}{4}$.

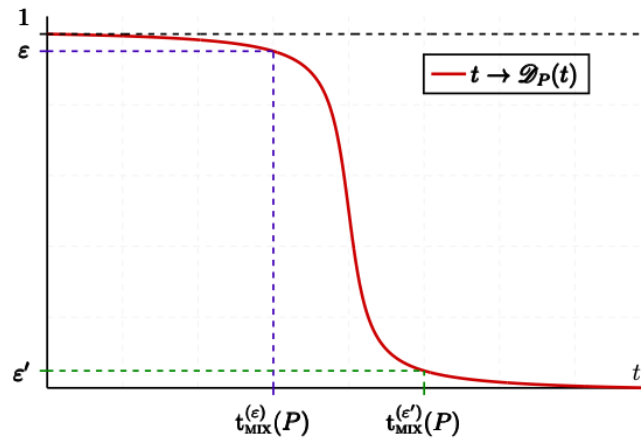


FIGURE 1 – Définitions du temps de mélange

1.4 Les inégalités fonctionnelles

Les trois équations que nous étudions ici sont les suivantes.

On dit que la chaîne de Markov vérifie :

- L'**inégalité de Poincaré** (IP) si la constante γ suivante est bien définie. Auquel cas, on l'appelle **constante de Poincaré**.

$$\gamma(P) = \inf \left\{ \frac{\mathcal{E}_P(f, f)}{\mathbb{V}(f)}, f \text{ non constante} \right\}$$

- L'**inégalité de log-Sobolev** (ILS) si la constante α suivante est bien définie. Auquel cas, on l'appelle **constante de log-Sobolev**.

$$\alpha(P) = \inf \left\{ \frac{\mathcal{E}_P(\sqrt{f}, \sqrt{f})}{\text{Ent}(f)}, f \text{ positive, } \|f\|_2 = 1 \right\}$$

- L'**inégalité de log-Sobolev modifiée** (ILSM) si la constante β suivante est bien définie. Auquel cas, on l'appelle **constante de log-Sobolev modifiée**.

$$\beta(P) = \inf \left\{ \frac{\mathcal{E}_P(f, \log(f))}{\text{Ent}(f)}, f \text{ positive, } \|f\|_2 = 1 \right\}$$

2 Effet des inégalités sur le temps de mélange.

On notera dans toute la suite $\lambda = \lambda_2$ la plus grande valeur propre de P différente de 1.

2.1 Première majoration classique

Proposition 2.1 (Première majoration classique du temps de mélange).

On a, si P irréductible et réversible :

$$t_{\text{ml}}^\varepsilon \leq \frac{1}{2\varepsilon \log(\lambda) \sqrt{\min(\pi)}}.$$

Démonstration. On a :

$$\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{2 \leq i \leq n} \lambda_i^t \varphi_i(x) \overline{\varphi_i(y)}.$$

Pour la norme associée au produit scalaire on obtient donc :

$$\left\| \frac{P^t(x, \cdot)}{\pi(\cdot)} - 1 \right\| \leq \frac{\lambda^{2t}}{\pi(x)},$$

d'où, avec Cauchy-Schwartz :

$$d_{\text{TV}}(\mu, \pi) \leq \frac{1}{2} \left\| \frac{\mu}{\pi} - 1 \right\|$$

puis

$$d_{\text{TV}}(\mu P, \pi) \leq \frac{\lambda^t}{2\pi(x)}.$$

Ceci étant vrai pour toute mesure initiale μ , on obtient notre borne par passage au logarithme. \square

Ce résultat peut toutefois être facilement amélioré en considérant les inégalités fonctionnelles introduites ci-dessus.

2.2 L'inégalité de Poincaré

Proposition 2.2 (Rappel de l'inégalité). *On dit qu'une chaîne de Markov vérifie l'inégalité de Poincaré si il existe une constante Γ tel que pour toutes fonctionnelles f*

$$\Gamma \mathbb{V}(f) \leq \mathcal{E}(f, f).$$

La constante de Poincaré notée $\gamma(P)$ est alors la constante optimale.

Remarquons que pour toute fonctionnelle f , $\mathcal{E}_{\Pi}(f, f) = \mathbb{V}(f)$. Intuitivement, on compare donc ici la convergence classique et la convergence en un pas ; d'où le lien avec l'étude de la vitesse de convergence de la suite.

Cette comparaison est particulièrement utile car les chaînes de Markov vérifient presque toutes cette inégalité. On a même une caractérisation simple de la constante optimale.

Proposition 2.3. *Toute chaîne de Markov irréductible (\mathcal{X}, P, μ_0) vérifie l'inégalité de Poincaré et on a*

$$\gamma(P) = 1 - \lambda.$$

On rappelle que λ est la plus grande valeur propre de P différente de 1.

Démonstration. Remarquons tout d'abord que la forme de Dirichlet et la variance sont inchangées par passage à l'adjoint et même à $\frac{P+P^*}{2}$. Aussi, en cas d'existence :

$$\gamma(P) = \gamma(P^*) = \gamma\left(\frac{P+P^*}{2}\right).$$

On pourra donc ici supposer donc P réversible sans perte de généralité. On a donc ici facilement gagné une hypothèse par rapport à la première majoration présentée du temps de mélange !

Soit maintenant (\mathcal{X}, P, μ_0) une chaîne de Markov irréductible. On a alors :

$$\mathcal{E}_P(f, f) = \langle (I - P)f, f \rangle = \sum_{i=2}^n (1 - \lambda_i) \langle f, \varphi_i \rangle^2.$$

De plus

$$\mathbb{V}(f) = \sum_{i=2}^n \langle f, \varphi_i \rangle^2.$$

D'où on a directement

$$\mathbb{V}(f) \leq (1 - \lambda) \mathcal{E}_P(f, f)$$

avec égalité en $f = \varphi_2$.

On a donc bien $\gamma(P) = 1 - \lambda$. □

Théorème 2.4 (Majoration de Poincaré).

Si P est irréductible et paresseuse (donc ergodique), on a :

$$t_{\text{ml}}^\varepsilon \leq \frac{2}{\gamma(P)} \log \left(\frac{1}{2\varepsilon \sqrt{\min(\pi)}} \right)$$

On lit ici que l'hypothèse de l'irréductibilité seule ne permet pas d'obtenir une borne sur le temps de mélange, mais qu'il faut ajouter l'hypothèse paresseuse pour obtenir le contrôle.

Toutefois, l'ajout de l'hypothèse paresseuse nous coûte en réalité assez peu, comme on a vu dans l'introduction qu'il était facile de se ramener à une matrice paresseuse.

En outre, la constante optimale de Poincaré permet pour un faible coût de transférer le résultat entre deux matrices proches, comme le montre le résultat suivant. Cette inégalité — dont on présente la démonstration à la fin de la sous-section — est donc en fait vérifiée avec une grande généralité.

Proposition 2.5 (Constante de comparaison).

Pour P et Q ayant la même loi stationnaire, on a l'inégalité suivante :

$$\gamma(P) \geq \gamma(Q) \cdot \gamma(P : Q),$$

où l'on a posé la constante de comparaison de P par rapport à Q :

$$\gamma(P : Q) = \inf \left\{ \frac{\mathcal{E}_P(f, f)}{\mathcal{E}_Q(f, f)}, f \text{ non constante} \right\}.$$

Démonstration. On a pour tout f non constante :

$$\frac{\mathcal{E}_P(f, f)}{\mathcal{E}_Q(f, f)} = \frac{\mathcal{E}_P(f, f)\mathbb{V}(f)}{\mathbb{V}(f)\mathcal{E}_Q(f, f)} \geq \gamma(P : Q)$$

d'où

$$\frac{\mathcal{E}_P(f, f)}{\mathbb{V}(f)} \geq \frac{\mathcal{E}_Q(f, f)}{\mathbb{V}(f)} \gamma(P : Q) \geq \gamma(Q) \gamma(P : Q).$$

On en déduit le résultat par passage à l'infimum. □

Ainsi, un contrôle sur $\gamma(Q)$ induit un contrôle sur $\gamma(P)$ avec un coût $\gamma(P : Q)$. Ce résultat est donc particulièrement intéressant pour relaxer les contraintes sur la chaîne de Markov.

Remarque. Lorsque P est paresseuse, on obtient même : $\gamma(P) \leq \gamma(P^*P)$ car un calcul direct montre alors que $\mathcal{E}_P(f, f) \leq \mathcal{E}_{P^*P}(f, f)$. Ce résultat est particulièrement important quand on se penche sur une autre interprétation de l'inégalité de Poincaré : son aspect contractant. Il nous servira également à démontrer l'inégalité sur le temps de mélange.

Proposition 2.6 (Contraction de la variance). *Pour $f : \mathcal{X} \rightarrow \mathbb{R}$, on a*

$$\mathbb{V}(Pf) \leq (1 - \gamma(P^*P))\mathbb{V}(f).$$

Démonstration. On procède comme lors de la démonstration que toute matrice irréductible vérifie l'inégalité de Poincaré :

$$\begin{aligned}\mathbb{V}(f) - \mathbb{V}(Pf) &= \langle f, f \rangle - \langle Pf, Pf \rangle \\ &= \langle f, f \rangle - \langle P^*Pf, f \rangle \\ &= \langle (I - P^*P)f, f \rangle.\end{aligned}$$

De plus,

$$\begin{aligned}\langle (I - P^*P)f, f \rangle &= \mathcal{E}_P(f, f) \\ &\geq \gamma(P^*P)\mathbb{V}(f)\end{aligned}$$

d'où le résultat. □

Remarque. Aussi, avec la remarque qui précède, si P est paresseuse, on a directement :

$$\mathbb{V}(Pf) \leq (1 - \gamma(P))\mathbb{V}(f).$$

Armés de toutes ces inégalités, on peut donc se placer dans le cas paresseux et obtenir la majoration suivante sur le temps de mélange.

Rappel : On cherche à montrer, si P est paresseuse, l'inégalité suivante :

$$t_{\text{ml}}^\varepsilon \leq \frac{2}{\gamma(P)} \log \left(\frac{1}{2\varepsilon\sqrt{\min(\pi)}} \right)$$

Démonstration. Grâce à Cauchy-Schwartz, on a :

$$d_{\text{TV}}(P^t(x, \cdot), \pi) \leq \frac{1}{2} \left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|.$$

Or, par définition, pour tout entier t ,

$$\frac{P^t(x, y)}{\pi(y)} = \frac{P^{*t}(y, x)}{\pi(x)} = P^{*t}f_x(y)$$

en notant

$$f_x = \frac{\delta_x(\cdot)}{\pi(x)}.$$

En outre, on a :

$$\pi P^{*t} f_x = \pi f_x = 1.$$

D'où, d'après la définition de la variance et le lemme de contraction de la variance :

$$\begin{aligned} \left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|^2 &= \mathbb{V}(P^{*t} f_x) \\ &\leq \frac{e^{\gamma(P)t}}{\min(\pi)}. \end{aligned}$$

Finalement

$$\forall t, D_p(t) \leq \frac{e^{\gamma(P)t}}{2 \min(\pi)}.$$

□

Par rapport à la vitesse de convergence, on a ici obtenu en norme 2 le résultat suivant.

Proposition 2.7. *Pour P irréductible on a dans l'espace $\mathcal{L}_2(\mathcal{X})$ le résultat suivant :*

$$\|p_n^x - 1\|_2^2 \leq \frac{1 - \pi(x)}{\pi(x)} (1 - \lambda)^{2n} \quad \text{avec } p_n^x = \frac{P^n(x, y)}{\pi(y)}.$$

Démonstration. Par la décomposition en base orthonormée de $\mathcal{L}_2(\mathcal{X})$, on obtient :

$$\|p_n^x - 1\|_2^2 = \langle p_n^x - 1, p_n^x - 1 \rangle = \sum_{i=2}^n (1 - \lambda_i)^{2n} |\varphi_i(x)|^2 \leq \frac{1 - \pi(x)}{\pi(x)} (1 - \lambda)^{2n}$$

par propriétés de la base (φ_i) . □

On a donc encore ici un contrôle par la quantité $\gamma(P) = 1 - \lambda$ sur la convergence discrète qui nous permet d'observer l'inégalité de Poincaré sous un autre angle.

On peut maintenant regarder l'effet qu'a cette inégalité sur la version continue de la suite. On obtient en fait ici un **contrôle en variation totale**.

Proposition 2.8. *Lorsque P est irréductible, on a l'inégalité suivante :*

$$4 d_{\text{TV}}(\bar{\mu}_t, \pi)^2 \leq \frac{e^{-2t\gamma(P)}}{\min(\pi)}.$$

Démonstration. On utilise la proposition précédente et le fait que $H_t = e^{-t(I-P)}$ et on obtient alors que, pour tout t , on a :

$$\|h_t^x - 1\|_2^2 \leq \frac{1 - \pi(x)}{\pi(x)} e^{2t(1-\lambda)}.$$

□

C'est d'ailleurs cette inégalité qui motive notre définition de H_t . On voit ainsi directement ici le lien entre H_t et notre base orthonormée.

2.3 L'inégalité de log-Sobolev modifiée

Cette inégalité est introduite par Serguey Bobkov et Prasad Tetali dans l'article [4] pour la contraction qu'elles induisent sur la forme de Dirichlet.

Proposition 2.9 (Rappel de l'inégalité). *On dit qu'une chaîne de Markov vérifie l'inégalité de log-Sobolev si il existe une constante B tel que pour toutes fonctionnelles f :*

$$B\text{Ent}(f) \leq \mathcal{E}(f, \log f).$$

La constante de log-Sobolev notée $\beta(P)$ est alors la constante optimale.

Proposition 2.10. *L'inégalité de log-Sobolev modifiée implique l'inégalité de Poincaré.*

Démonstration. Supposons que l'inégalité de log-Sobolev est vérifiée par la chaîne de Markov (\mathcal{X}, P, μ_0) et soit f une fonctionnelle positive.

Par le calcul, on remarque qu'il suffit de montrer que

$$\begin{aligned} \mathcal{E}_P(e^f, f) &\leq \frac{1}{2} \sum_{x,y \in \mathcal{X}^2} (f(x) - f(y))^2 e^{f(x)} P(x,y) \pi(x) \\ &\leq \sum_{x,y \in \mathcal{X}^2} (e^{f(x)} - e^{f(y)})^2 e^{-f(x)} P(x,y) \pi(x) \end{aligned}$$

Or, on peut toujours supposer P réversible dans perte de généralité, d'où on a :

$$\sum_{x,y \in \mathcal{X}} e^{f(x)} (f(x) - f(y))^2 P(x,y) \pi(x) = \sum_{x,y \in \mathcal{X}} \frac{e^{f(x)} + e^{f(y)}}{2} (f(x) - f(y))^2 P(x,y) \pi(x)$$

et

$$\sum_{x,y \in \mathcal{X}} e^{-f(x)} (e^{f(x)} - e^{f(y)})^2 P(x,y)\pi(x) = \sum_{x,y \in \mathcal{X}} \frac{e^{-f(x)} + e^{-f(y)}}{2(e^{f(x)} - e^{f(y)})^2} P(x,y)\pi(x).$$

De plus, par convexité, on obtient, pour h réel :

$$h(e^h - 1) \leq \frac{h^2(e^h + 1)}{2} \leq \frac{e^{-h} + 1}{2}(e^h - 1)^2.$$

On obtient alors la comparaison entre $\mathcal{E}_P(e^f, f)$ et $\mathbb{V}(e^f)$ lorsque f est positive.

Or, $\gamma(P)$ est invariant par restriction aux fonctions positives et on a l'inégalité suivante pour toute fonctionnelle f positive :

$$2\mathcal{E}_P(f, \log(f)) \leq \mathbb{V}(f).$$

On obtient alors le résultat souhaité en combinant ces inégalités. □

L'inégalité de log-Sobolev modifiée est donc plus forte que celle de Poincaré.

Sur les temps de mélange, on a cette fois l'inégalité suivante.

Théorème 2.11 (Majoration de log-Sobolev modifiée). Si P est irréductible, on a :

$$d_{\text{TV}}(\bar{\mu}_t, \pi)^2 \leq 2 \log \left(\frac{1}{\min(\pi)} \right) e^{-2\beta(P)t}.$$

Pour le démontrer et comprendre un peu mieux cette signification, on a le résultat suivant.

Proposition 2.12 (Contraction de l'entropie). *Pour tout f :*

$$\text{Ent}(\bar{\mu}_t) \leq \text{Ent}(\mu_0) e^{-2\beta(P)t}.$$

On lit ici une décroissance exponentielle de l'entropie, quand on avait une décroissance exponentielle de la variance avec l'inégalité de Poincaré. Cela permet donc de comparer les niveaux sur lesquels agissent les différents inégalités. En outre, comme on a vu que l'inégalité de log-Sobolev modifiée implique l'inégalité de Poincaré et qu'on a montré $2\mathcal{E}_P(f, \log(f)) \leq \mathbb{V}(f)$ dans la preuve, on remarque que cette double contraction induite par l'inégalité de log-Sobolev modifiée induit directement un contrôle sur la forme de Dirichlet. En effet, elle est encadrée par l'entropie et la variance, qui se contractent alors toutes deux.

Démonstration. Comme pour la propriété similaire que l'on verra ensuite pour log-Sobolev, on remarque que :

$$\partial_t \text{Ent}(fH_t) = \mathcal{E}_P(f, \log(f))$$

On obtient alors le résultat en appliquant le lemme de Gronwald et à l'inégalité de log-Sobolev. \square

Le résultat suivant permet en outre de relier le carré de la variation totale à l'entropie ce qui, outre nous permettre de voir l'entropie sous un autre jour, permettra de montrer l'inégalité sur le temps de mélange continu énoncée ci-dessus.

Proposition 2.13. *On a pour toute mesure ν de densité f par rapport à π :*

$$2d_{\text{TV}}(\nu, \pi)^2 \leq \text{Ent}(f)$$

Démonstration. On remarque que pour tout $x \geq 0$:

$$3(x-1)^2 \leq (4+2x)(x \log(x) - x - 1)$$

Il reste ensuite à appliquer cette inégalité dans la somme définissant $\text{Ent}(f)$. On obtient ensuite le résultat grâce à l'inégalité de Cauchy-Schwartz. \square

Rappel : On cherche à montrer, si P est irréductible, l'inégalité suivante :

$$d_{\text{TV}}(\bar{\mu}_t, \pi)^2 \leq 2 \log \left(\frac{1}{\min(\pi)} \right) e^{-2\beta(P)t}.$$

Démonstration. D'après la proposition précédente, on a pour tout t , en notant f_t la densité de $\bar{\mu}_t$ par rapport à π :

$$2d_{\text{TV}}(\bar{\mu}_t, \pi)^2 \leq \text{Ent}(f_t).$$

Or, par définition, $f_t = H_t f_0$ d'où le résultat d'après ce qui précède. \square

2.4 L'inégalité de log-Sobolev

Rappel de l'inégalité

On dit qu'une chaîne de Markov vérifie l'inégalité de log-Sobolev si il existe une constante A tel que pour toutes fonctionnelles f :

$$A \text{Ent}(f) \leq \mathcal{E} \left(\sqrt{f}, \sqrt{f} \right).$$

La constante de log-Sobolev notée $\alpha(P)$ est alors la constante optimale.

Remarque. L'inégalité de log-Sobolev a un lien fort avec l'inégalité de Poincaré.

En effet, si on considère la classe d'inégalités suivante appelées inégalités de Sobolev :

$$\forall p \in]1, 2], \exists C_p, C_p(\|f\|_p^p - \|f\|_1^p) \leq \mathcal{E}_P(f, f^{(p-1)})$$

dans le cas $p = 2$, on retrouve exactement l'inégalité de Poincaré et, en prenant la limite lorsque p tend vers 1, on obtient l'inégalité de log-Sobolev.

Proposition 2.14. *L'inégalité de log-Sobolev implique l'inégalité de log-Sobolev modifiée.*

Démonstration. Si a et b sont deux réels positifs, on a par convexité :

$$\left(\frac{\sqrt{a} - \sqrt{b}}{a - b}\right)^2 = \left(\frac{1}{2(a - b)} \int_b^a \frac{1}{\sqrt{t}}\right)^2 \leq \frac{\log(a) - \log(b)}{4(a - b)}.$$

Aussi,

$$4(\sqrt{a} - \sqrt{b})^2 \leq (\log(a) - \log(b))(a - b).$$

On a alors par définition de la forme de Dirichlet, pour toute f fonctionnelle positive,

$$\mathcal{E}_P(f, \log(f)) \geq 2\mathcal{E}_P(\sqrt{f}, \sqrt{f}).$$

Si de plus P est réversible, cette inégalité devient :

$$\mathcal{E}_P(f, \log(f)) \geq 4\mathcal{E}_P(\sqrt{f}, \sqrt{f}).$$

D'où le résultat. □

Théorème 2.15 (Bilan des dépendances des constantes).

Si (\mathcal{X}, P, μ_0) vérifie l'inégalité de log-Sobolev et que P est irréductible, on a :

$$\frac{\gamma(P)}{2} \leq \beta(P) \leq \frac{\alpha(P)}{4}$$

Si P est réversible, on a de plus égalité des constantes.

Démonstration. C'est un corollaire direct des propriétés 2.10 et 2.14 qui montrent l'enchaînement classique des dépendances des trois inégalités. \square

On s'intéresse maintenant à la contraction induite par l'inégalité de log-Sobolev, étudiée par Laurent Miclo dans l'article [3].

Définition 2.16.

On dit qu'on observe un phénomène d'**hypercontractivité** d'une chaîne de Markov (\mathcal{X}, P, μ_0) si il existe $\alpha > 0$ tel que :

$$\forall t > 0, \forall f \in \mathcal{L}_2(\mathcal{X}), \forall q \in [2, \infty[, \quad (e^{4\alpha t} \geq q - 1) \implies (\|H_t f\|_q \leq \|f\|_2).$$

Ce phénomène induit un résultat très fort puisqu'il permet en particulier de majorer la convergence de $\bar{\mu}_t$ non seulement en norme L^2 , mais aussi en norme L^p pour p quelconque voire infini.

Proposition 2.17. *Le phénomène d'hypercontractivité est équivalent à la vérification de l'inégalité de log-Sobolev.*

Démonstration. Supposons dans un premier temps que la chaîne est hypercontractive et notons α la constante d'hypercontractivité. Posons pour tout $p \in \mathbb{N}$, f tel que $\|f\|_2 = 1$ et $p(t) = 1 + e^{4\alpha t}$:

$$\begin{aligned} \mathcal{L}_p(f) &= \sum_x |f(x)|^p \log(|f(x)|^p) \pi(x) \\ G(t) &= \|H_t f\|_{p(t)}^{p(t)} \\ F(t) &= \|H_t f\|_{p(t)}. \end{aligned}$$

On a alors :

$$G'(t) = -\frac{p'(t)}{p(t)} \mathcal{E}_P(H_t f, (H_t f)^{p(t)-1}) + \frac{p'(t)}{p(t)} \sum_x |(H_t f)|^{p(t)} \log(|(H_t f)|^{p(t)}) \pi(x).$$

On en déduit que

$$\begin{aligned} F'(t) &= \left(-\frac{p'(t)}{p(t)^2} \log(G(t)) + \frac{G'(t)}{p(t)G(t)} \right) F(t) \\ &= \left(\frac{p'(t)}{p(t)^2} \mathcal{L}_p(H_t f) - \mathcal{E}_P(H_t f, (H_t f)^{p(t)-1}) \right) F(t)^{-p(t)+1}. \end{aligned}$$

En réinjectant la valeur de $p(t)$, on obtient alors :

$$F'(t) = F(t)^{-p(t)+1} \left(\frac{4\alpha e^{4\alpha t}}{(1 + e^{4\alpha t})} \mathcal{L}_P(H_t f) - \mathcal{E}_P(H_t f, (H_t f)^{p(t)-1}) \right).$$

De plus, $p(0) = 2$, $H_0 f = f$ et par hypercontractivité ($\|H_t f\|_q \leq 1$), on obtient l'inégalité de log-Sobolev :

$$\alpha \text{Ent}(f^2) \leq \mathcal{E}_P(f, f).$$

Réciproquement, on constate dans la preuve du sens direct que supposer l'inégalité de log-Sobolev et remonter les inégalités permet d'obtenir l'hypercontractivité avec la même constante. \square

On obtient alors l'inégalité suivante sur le temps de mélange.

Théorème 2.18 (Majoration de log-Sobolev).

Lorsque P est irréductible, en temps continu on a

$$2 d_{\text{TV}}(\bar{\mu}_t, \pi)^2 \leq \log \left(\frac{1}{\min(\pi)} \right) e^{-2\alpha(P)t}$$

Si de plus P est réversible, on a même :

$$2 d_{\text{TV}}(\bar{\mu}_t, \pi)^2 \leq \log \left(\frac{1}{\min(\pi)} \right) e^{-4\alpha(P)t}$$

Pour montrer cette inégalité, on raisonne de la même manière que lors de l'étude de l'inégalité de log-Sobolev modifiée. On commence donc par montrer un résultat de contraction de l'entropie, induit par l'hypercontractivité mentionnée ci-dessus.

Proposition 2.19. *Si P est irréductible, on a :*

$$\forall \mu, \text{Ent}(\mu H_t) \leq e^{-2\alpha(P)t} \text{Ent}(\mu).$$

Si de plus P est réversible :

$$\forall \mu, \text{Ent}(\mu H_t) \leq e^{-4\alpha(P)t} \text{Ent}(\mu).$$

Démonstration.

Soit f telle que $\mu = f\pi$ et $\|f\|_2 = 1$.

On obtient alors :

$$\begin{aligned}\partial_t \text{Ent}(H_t^* f) &= -\langle (I - P^*)H_t^* f, \log(H_t^* f) \rangle \\ &\leq -\mathcal{E}_P((H_t^* f)^{\frac{1}{2}}, (H_t^* f)^{\frac{1}{2}}) \\ &\leq -2\alpha(P)\text{Ent}(H_t^* f).\end{aligned}$$

En intégrant avec le lemme de Gronwall, on trouve alors :

$$\text{Ent}(H_t^* f) \leq e^{-2\alpha(P)t}\text{Ent}(f).$$

Pour le cas réversible, on procède de même. \square

On obtient alors la majoration du temps de mélange liée à l'inégalité de log-Sobolev de la même manière exactement qu'on a appliquée pour celle de l'inégalité de log-Sobolev modifiée.

En outre, Persi Diaconis et Laurent Saloff-Coste montrent dans l'article [2] qu'on sait, de même que pour l'inégalité de Poincaré, relier les constantes de log-Sobolev de deux chaînes de Markov agissant sur un même espace.

Proposition 2.20. *Sur un même ensemble d'états \mathcal{X} , on considère deux chaînes de Markov de matrices de transitions P et Q et de mesures stationnaires respectives π et π' .*

S'il existe deux constantes strictement positives A et a telles que pour tout f , on ait :

$$\begin{cases} \mathcal{E}'_P(f, f) \leq A \mathcal{E}_P(f, f) \\ \pi \leq a\pi' \end{cases}$$

Alors, on a les deux inégalités suivantes :

$$\gamma(Q) \leq \frac{A}{a}\gamma(P) \qquad \alpha(Q) \leq \frac{A}{a}\alpha(P)$$

Démonstration. Il s'agit de remarquer que pour tous réels strictement positifs x et y , on a :

$$x \log(x) - x \log(y) - x + y > 0$$

On obtient alors le résultat en réinjectant dans les définitions de γ et α . \square

2.5 Comparaison des différents effets

Théorème 2.21 (Conclusion).

Du point de vue des temps de mélange, l'inégalité de log-Sobolev fournit :

- Un meilleur contrôle que l'inégalité de Poincaré si

$$\frac{1}{\gamma(P)} \log \left(\frac{1}{\min(\pi)} \right) \geq \frac{1}{\alpha(P)} \log \left(\log \left(\frac{1}{\min(\pi)} \right) \right)$$

- Toujours une meilleure approximation que celle de log-Sobolev.

Démonstration. On obtient ces formules en comparant les résultats des théorèmes ci-dessus. Le deuxième point s'obtient car $\alpha(P)$ est toujours supérieur à $\beta(P)$ d'après la proposition 2.15. \square

3 Implication réciproque dans le cas réversible

Nous en venons enfin à la preuve de l'équivalence entre l'inégalité de log-Sobolev et l'inégalité de log-Sobolev modifiée. La preuve présentée ici est tirée [5] de Justin Salez, Konstantin Tikhomirov et Pierre Youssef.

Théorème 3.1. Si (\mathcal{X}, P, μ_0) est une chaîne de Markov réversible qui vérifie l'inégalité de log-Sobolev, alors elle vérifie l'inégalité de log-Sobolev modifiée.

Démonstration. Dans cette preuve, nous noterons $\alpha = \alpha(P)$ et $\beta = \beta(P)$.

Notons

$$E = \{(x, y), \in \mathcal{X}^2, P(x, y) > 0\} \quad \text{et} \quad Q(x) = \sum_{y \neq x} P(x, y).$$

On pose également

$$p = \min_{(x,y) \in E} \frac{P(x, y)}{\max(P(x), P(y, x))}.$$

Pour démontrer le résultat nous allons utiliser une technique de régularisation. Notre objectif sera de nous restreindre à une famille de fonctions particulières pour pouvoir y améliorer notre inégalité, pour ensuite généraliser.

On dit que f positive est r -régulière si :

$$\forall (x, y) \in E : f(x) \leq r f(y).$$

On pose alors, pour $r > 1$:

$$\mathcal{H}(r) = \frac{\sqrt{r} + 1}{\sqrt{r} - 1} \log(r).$$

Remarquons tout d'abord qu'on a le résultat suivant.

Lemme 3.2 (Majoration avec régularisation).

Si f est positive et r -régulière, alors :

$$\mathcal{E}_P(f, \log(f)) \leq \mathcal{H}(r) \mathcal{E}_P(\sqrt{f}, \sqrt{f})$$

L'intérêt de ce résultat est directement visible puisqu'il suffit ensuite, pour montrer le résultat de contrôler \mathcal{H} .

Démonstration. On prolonge \mathcal{H} par la valeur 4 en 1. Comme $\mathcal{H}(r) = \mathcal{H}(r^{-1})$. On en déduit que pour tout $u \in [r^{-1}, r] : \mathcal{H}(u) \leq \mathcal{H}(r)$.

Par le calcul, on trouve en outre que :

$$\forall (x, y) \in E : (f(x) - f(y)) \log\left(\frac{f(x)}{f(y)}\right) \leq \mathcal{H}(r) (\sqrt{f(x)} - \sqrt{f(y)})^2.$$

On en déduit alors le résultat en multipliant par $\pi(x)P(x, y)$. □

De ce lemme, on déduit une majoration intéressante des constantes :

$$\alpha(r) \leq \mathcal{H}(r) \beta(r)$$

où $\alpha(r)$ et $\beta(r)$ sont les restrictions de α et β aux fonctions r -régulières.

On va maintenant faire le lien entre les fonctions r -régulières et les autres. Pour cela, on définit la notion de distance dans un graphe : comme la chaîne de Markov est supposée irréductible, le graphe est connexe et il existe un chemin reliant chaque couple d'états $(x, y) \in \mathcal{X}^2$. On note alors $d(x, y)$ la longueur du plus court chemin reliant x à $y : d(x, y) = \inf\{t \in \mathbb{N}, P^t(x, y) > 0\}$. Cette quantité est alors bien définie.

Notons maintenant pour toute fonction $f : \mathcal{X} \rightarrow \mathbb{R}_+^*$:

$$\forall x \in \mathcal{X} \quad f_*(x) = \max_{z \in \mathcal{X}} r^{-d(x,z)} f(z).$$

On obtient alors les très chouettes inégalités suivantes.

Proposition 3.3. *Pour $r = \frac{4}{p^2}$, on a*

$$\mathcal{E}_P(\sqrt{f_*}, \sqrt{f_*}) \leq \frac{4}{3} \mathcal{E}_P(\sqrt{f}, \sqrt{f}) \quad \text{et} \quad \mathcal{E}_P(f_*, \log f_*) \leq \frac{4}{3} \mathcal{E}_P(f, \log(f)).$$

Cette proposition découle directement de la combinaison des deux lemmes suivants :

Lemme 3.4. Pour tout $e = (x, y) \in E^2$ tel que $f_*(x) \leq f_*(y)$, il existe $e' = (x', y') \in E^2$ tel que :

$$r^{-d(e,e')} f(x') \leq f_*(x) \leq f_*(y) \leq r^{-d(e,e')} f(y').$$

Démonstration. Si $f_*(y) = f(y)$, on peut considérer $e' = e$.

Sinon, si $f_*(y) < f(y)$, on choisit y' tel que

$$f_*(y) = r^{-d(y,y')} f(y')$$

et x' l'avant dernier point sur le plus court chemin de y à y' .

On a alors :

$$d(e, e') = d(y, y') = 1 + d(y, x') \geq d(x, x').$$

On en déduit donc par définition de f_* que

$$f_*(x) \geq r^{-d(x,x')} f(x') \geq r^{-d(e,e')} f(x')$$

et on a le résultat souhaité. □

Lemme 3.5. Pour tout f , on a :

$$\mathcal{E}_P(\sqrt{f_*}, \sqrt{f_*}) \leq \kappa \mathcal{E}_P(\sqrt{f}, \sqrt{f}) \quad \text{et} \quad \mathcal{E}_P(f_*, \log(f_*)) \leq \kappa \mathcal{E}_P(f, \log(f))$$

où

$$\kappa = \max_{e' \in E} \left\{ \frac{1}{c(e')} \sum_{e \in E} c(e) r^{-d(e, e')} \right\}.$$

Où $c(x, y) = \pi(x)Q(x, y)$. La quantité κ est en fait une quantification de la distance asymptotique entre les arêtes.

Démonstration. Posons pour toute arête $e = (x, y) : \nabla f(e) = f(y) - f(x)$.

Pour tout e dans E on peut alors trouver e' dans E tel que :

$$\nabla \sqrt{f_{*+}}(e) \leq r^{-d(e, e')} \nabla \sqrt{f_+}(e')$$

On en déduit alors :

$$\begin{aligned} \mathcal{E}_P(\sqrt{f_*}, \sqrt{f_*}) &= \sum_{e \in E} c(e) \left(\nabla \sqrt{f_*} \right)_+^2(e) \\ &\leq \sum_{e, e' \in E} c(e) r^{-d(e, e')} \left(\nabla \sqrt{f} \right)_+^2(e') \\ &\leq \kappa \sum_{e' \in E} c(e') \left(\nabla \sqrt{f} \right)_+^2(e') \\ &= \kappa \mathcal{E}_P(\sqrt{f}, \sqrt{f}). \end{aligned}$$

Pour le deuxième point on effectue la même démarche avec $\mathcal{E}_P(f, \log(f))$. □

On déduit alors la proposition par définition de la constante κ .

On regarde ensuite le passage de l'autre côté de l'inégalité :

Proposition 3.6. Pour $r = \frac{4}{p^2}$ On a pour tout f :

$$\text{Ent}(f) \leq 2\text{Ent}(f_*)$$

Pour démontrer ce résultat, on pose encore trois lemmes.

Lemme 3.7. Si $r = \frac{4}{p^2}$, alors $\kappa \leq \frac{4}{3}$.

Démonstration. Soit deux arrêtes e et e' de E telles que $d(e, e') = 1$. On a alors nécessairement $e = (x, y)$ et $e' = (y, z)$.

Par réversibilité de P et définition de p on a alors :

$$\begin{aligned} c(e) &= \pi(x)P(x, y) \\ &= \pi(y)P(y, x) \\ &\leq \pi(y)P(y) \\ &\leq \frac{1}{p}\pi(y)P(y, z) \\ &\leq \frac{1}{p}c(e') \end{aligned}$$

De plus on a pour tout $e, e' \in E$, $c(e) \leq r^{-d(e, e')}c(e')$, d'où on obtient :

$$\kappa \leq \max_{e' \in E} \left\{ \sum_{e \in E} (rp)^{-d(e, e')} \right\} \leq \sum_{k \in \mathbb{N}} \deg(\mathcal{X})^k (rp)^{-k}$$

où $\deg(\mathcal{X})$ est le degré du graphe (\mathcal{X}, E) , défini comme le degré maximal des sommets du graphe, soit le nombre de voisins directs du sommet ayant le plus de voisins directs dans le graphe.

Or, par définition de p , $\deg(\mathcal{X}) \leq \frac{1}{p}$.

La choix de r impose donc bien :

$$\kappa \leq \sum_{k \in \mathbb{N}} (p^2 r)^{-k} = \frac{4}{3}.$$

□

Lemme 3.8. Pour tout f on a :

$$5 \text{Ent}(f) \leq 6 \text{Ent}(f_*) + 6 \log 6 \mathbb{E}(f_* - f)$$

Démonstration. Remarquons tout d'abord que :

$$\text{Ent}(f) = \max\{\mathbb{E}(fg), \mathbb{E}(e^g) \leq 1\}.$$

Quitte à multiplier f et f_* par une constante, on peut supposer que $\mathbb{E}(f) = 1$.

Posons maintenant $h = \log\left(\frac{1+5f}{6}\right)$.

Par hypothèse sur f , on remarque alors les trois résultats suivants :

- $\mathbb{E}(e^h) = 1$
- $h \geq \frac{5}{6} \log(f)$
- $h \geq \frac{1}{6}$.

De plus,

$$\text{Ent}(f_*) \geq \mathbb{E}(f_* h) \geq \mathbb{E}(f h) + \mathbb{E}((f_* - f)h).$$

Le résultat découle alors de la combinaison de ces propriétés. \square

Lemme 3.9. Pour $r = \frac{4}{p^2}$, on a $3 \log(6) \mathbb{E}(f_* - f) \leq \text{Ent}(f)$

Démonstration.

Commençons par poser pour tout $x \in \mathcal{X}$, $t(x) \in \mathcal{X}$ tel que $f_*(x) = r^{-d(x,t(x))} f(t(x))$.

Remarquons alors que :

- $t(x) = x$ si et seulement si $f_*(x) = f(x)$
- si $f(t(x)) < f_*(t(x))$, alors :

$$f_*(x) < r^{-d(x,t(x))} f_*(t(x)) = r^{-d(x,t(x))} r^{-d(t(x),t^2(x))} f(t^2(x)) \leq r^{-d(x,t^2(x))} f(t^2(x))$$

ce qui contredit la maximalité de $f_*(x)$.

Aussi, on en déduit que $f(t(x)) = f_*(t(x))$, ce qui équivaut à $t^2(x) = t(x)$.

En posant maintenant

$$A = \{x \in \mathcal{X} : t(x) \neq x\} = \{x \in \mathcal{X} : f(x) \neq f_*(x)\}$$

on remarque que $t(A) = A^c$. De plus :

$$\begin{aligned} \mathbb{E}(f_* - f) &= \sum_{x \in A} \pi(x)(f_*(x) - f(x)) \\ &= \sum_{x \in A} \pi(x) f(t(x)) r^{-d(x,t(x))} - \sum_{x \in A} \pi(x) f(x) \\ &= \sum_{y \in t(A)} \pi(y) f(y) h(y) - \sum_{x \in A} \pi(x) f(x) \end{aligned}$$

où on pose pour tout $y \in t(A)$:

$$h(y) = \sum_{x \in t^{-1}(\{y\})} \frac{\pi(x)}{\pi(y)} r^{-d(x,y)}.$$

On étend alors h à \mathcal{X} tout entier en posant $h(A) = -1$, ce qui nous permet d'écrire :

$$\mathbb{E}(f_* - f) = \mathbb{E}(fh).$$

Par la caractérisation de l'entropie du lemme précédent, on remarque alors que $\mathbb{E}(6^{3h}) \leq 1$ et on en déduit par le choix de r que $h \leq \sum_{k \geq 1} (p^2 r)^{-k} = \frac{1}{3}$. Finalement, on obtient $6^{3h} \leq 6 + 15h$ sur $t(A)$.

Comme de plus $r \geq 16$, on déduit de tout cela que :

$$\mathbb{E}((6^{3h} - 1)\mathbb{1}_{t(A)}) \leq 15 \mathbb{E}(h\mathbb{1}_{t(A)}) = 15 \sum_{x \in A} \pi(x) r^{-d(x, t(x))} \leq \frac{15}{16} \pi(A).$$

De la même manière, on a $\mathbb{E}((6^{3h} - 1)\mathbb{1}_A) = (6^{-3} - 1)\pi(A) \leq -\frac{15}{16}\pi(A)$.

On obtient alors le résultat souhaité par définition de $\text{Ent}(f)$. □

On obtient alors en combinant tous ces lemmes l'inégalité entre l'entropie de f et celle de f_* .

On combine alors les inégalités obtenues sur les entropies et les formes de Dirichlet de f et f_* pour obtenir le résultat suivant.

Proposition 3.10. *Pour $r = \frac{4}{p^2}$ on a :*

$$\alpha \leq 3\alpha(r) \quad \text{et} \quad \beta \leq 3\beta(r).$$

De plus par lemme 3.2 avec le résultat pour les fonctions r -régulières, on déduit que :

$$\alpha \leq 3\alpha(r) \leq \mathcal{H}(r)\beta(r) \leq 3\mathcal{H}(r)\beta$$

car $\beta(r) \leq \beta$, ce qui achève la preuve! □

Remarque. On déduit de la démonstration que :

$$\alpha(P) \leq 20 \beta(P) \log \left(\frac{1}{p} \right)$$

4 Conclusion

Références

- [1] Justin Salez, *Mixing times of Markov chains*,
- [2] Persi Diaconis et Laurent Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, The Annals of Applied Probability 1996, Vol. 6, No. 3, 695-750
- [3] Laurent Miclo, *Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies*, Séminaire de probabilités (Strasbourg), tome 31 (1997), p. 136-167
- [4] Sergey G. Bobkov et Prasad Tetali, *Modified logarithmic Sobolev inequalities in discrete settings*
- [5] Justin Salez, Konstantin Tikhomirov et Pierre Youssef, *Upgrading MLSI to LSI for reversible Markov chains*, arXiv :2212.06028 [math.PR]
- [6] Anna Ben-Hamou, divers conseils et encadrements pendant la durée de ce mémoire