CHAPITRE 6

SEMI-GROUPES ET INÉGALITÉS GAUSSIENNES

6.1. Le semi-groupe d'Ornstein-Uhlenbeck

On note γ_n la mesure gaussienne standard sur $\mathbb{R}^n$ muni de sa strucutre euclidienne usuelle. On considère aussi l'espace de Hilbert

$$L^2(\gamma_n) = L^2(\mathbb{R}^n, \gamma_n; \mathbb{R})$$

des fonctions mesurables $f : \mathbb{R}^n \to \mathbb{R}$ telles que $\displaystyle\int f^2 \, d\gamma_n < +\infty$.

Pour f une fonction de classe C^2 sur $\mathbb{R}^n$, on peut définir Lf par

$$\forall x \in \mathbb{R}^n, \qquad Lf(x) = \Delta f(x) - x \cdot \nabla f(x). \tag{6.1}$$

Ainsi L est un opérateur différentiel linéaire d'ordre 2 (elliptique). En se restreignant aux fonctions dans $L^2(\gamma_n)$, on peut voir L comme un opérateur non-borné sur $L^2(\gamma_n)$. Cet opérateur est en fait le Laplacien naturel associé au poids gaussien. En effet, si $f \in L^2(\gamma_n)$ est une fonction de classe C^2 telle que $Lf \in L^2(\gamma_n)$, alors pour toute fonction g de classe C^1 et a support compact on a, par intégration par parties,

$$\int (Lf) g \, d\gamma_n = -\int \nabla f \cdot \nabla g \, d\gamma_n,$$

ce qui permet de voir L comme un opérateur non-borné auto-adjoint négatif de $L^2(\gamma_n)$.

Il sera parfois utile de considérer L sur une algèbre de fonctions plus restreinte. On dit qu'une fonction $g : \mathbb{R}^n \to \mathbb{R}$ est à *croissance lente* si $|g|$ est dominée par une fonction polynôme, c'est-à-dire s'il existe des constantes $C, k > 0$ tel que, $\forall x \in \mathbb{R}^n, \quad |g(x)| \leq C(1 + |x|^k)$. On introduit alors

$$\mathcal{A} := \{f \in C^\infty(\mathbb{R}^n, \mathbb{R}) \ ; \quad f \text{ et toutes ses dérivées (partielles) sont à croissance lente}\}.$$

On a bien sûr que $\mathcal{A} \subset L^2(\gamma_n)$, mais aussi que l'algèbre $\mathcal{A}$ est stable par L

$$L\mathcal{A} \subset \mathcal{A}$$

On remarque que $\mathcal{A}$ contient les fonctions C^∞ à support compact, ce qui permet de voir que $\mathcal{A}$ est dense dans $L^2(\gamma_n)$. La formule d'intégration par parties est évidemment valable pour des fonctions dans $\mathcal{A}$.

Fait 6.1. — *Soit $f, g \in \mathcal{A}$. Alors*
$$\int (Lf)g\,d\gamma_n = \int fLg\,d\gamma_n = -\int \nabla f \cdot \nabla g\,d\gamma_n. \tag{6.2}$$

Le résultat suivant découle de la théorie des opérateurs différentiels elliptiques :

Théorème 6.2. — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue et à croissance lente. Alors, il existe une fonction*
$$F : [0, +\infty[\times\mathbb{R}^n \longrightarrow \mathbb{R}$$
continue sur $[0, +\infty[\times\mathbb{R}^n$ et C^∞ sur $]0, +\infty[\times\mathbb{R}^n$ telle que
$$F(0, \cdot) = f(\cdot) \tag{6.3}$$
et pour $t > 0$ et $x \in \mathbb{R}^n$,
$$\begin{aligned} \frac{\partial F}{\partial t}(t, x) &= L(F(t, x)) \\ &= \Delta_x F(t, x) - x \cdot \nabla_x F(t, x). \end{aligned} \tag{6.4}$$
De plus, si f est de classe C^2, alors l'équation précédente est vraie aussi en $t = 0$.

On notera $P_t(f) = F(t, \cdot)$ et en toute logique on écrit $P_t(f) = e^{tL}f$.

En fait, on peut se passer de ce résultat d'EDP, car on peut donner ici une expression explicite de la solution. Pour $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ continue et à croissance lente, on posera
$$\forall t \geq 0, \quad \forall x \in \mathbb{R}^n, \qquad P_t(f)(x) = \int_{\mathbb{R}^n} f\big(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\big)\,d\gamma_n(y). \tag{6.5}$$
On a bien $P_0(f) = f$. Si on pose $c_t = e^{-t}$ et $s_t = \sqrt{1 - e^{-2t}}$ on a $c_t^2 + s_t^2 = 1$ et
$$P_t(f)(x) = \int_{\mathbb{R}^n} f\big(c_t\, x + s_t\, y\big)\,d\gamma_n(y) = \mathbb{E}\big[f(c_t\, x + s_t\, Y)\big]$$
si Y est un vecteur gaussien standard de $\mathbb{R}^n$ défini sur un espace de probabilité $(\Omega, \mathbb{P})$. Pour $t > 0$ on a
$$P_t(f)(x) = \int_{\mathbb{R}^n} f(z)\, e^{-|z - c_t\, x|^2/2s_t^2} \frac{dz}{(2\pi s_t^2)^{n/2}}. \tag{6.6}$$
On note alors que

Fait 6.3. — *Si f est continue à croissance lente, alors dès que $t > 0$, on a que $P_t f$ est régulière et même que $P_t f \in \mathcal{A}$. En particulier, $\mathcal{A}$ est stable par P_t. De plus, si f est positive non-identiquement nulle, alors $P_t(f)(x) > 0$ pour tout $x \in \mathbb{R}^n$.*

Un calcul (direct mais pénible) donne que pour tout $t > 0$,
$$\frac{\partial}{\partial t} P_t(f) = L(P_t(f)) \quad \text{ sur } \mathbb{R}^n.$$

En utilisant l'une des caractérisations précédentes de P_t on montre (exercice) les propriétés suivantes. Tout d'abor P_t est un opérateur linéaire
$$P_t(\lambda f + \mu g) = \lambda P_t(f) + \mu P_t(g), \tag{6.7}$$
et il possède la propriété de semi-groupe :
$$\forall s, t \geq 0, \qquad P_t\big(P_s(f)\big) = P_{t+s}(f) \tag{6.8}$$

Ensuite P_t préserve les constantes et les fonctions positives sur $\mathbb{R}^n$,

$$P_t(\mathbf{1}) = \mathbf{1} \quad \text{et} \quad f \geq 0 \Rightarrow P_t(f) \geq 0. \tag{6.9}$$

On peut en déduire que $|P_t(f)| \leq P_t(|f|)$, ce qui se voit aussi sur la définition (6.5) . Plus généralement, on voit que

$$\forall p \geq 1, \qquad \big|P_t(f)(x)\big|^p \leq P_t\big(|f|^p\big)(x). \tag{6.10}$$

En fait, la définition (6.5), ou plus précisément (6.6), nous dit que $P_t f$ est donné par un noyau de convolution. On peut écrire que, pour chaque $x \in \mathbb{R}^n$ et $t > 0$ donné,

$$P_t f(x) = \int f(y)\, d\mu_{t,x}(y)$$

où $\mu_{t,x}$ est une probabilité sur $\mathbb{R}^n$. On dit que P_t est un semi-groupe Markovien. Cela permet d'obtenir des inégalités ponctuelles sur $P_t f(x)$. Par exemple on a

$$P_t(fg)(x) \leq \sqrt{P_t(f^2)(x) P_t(g^2)(x)} \tag{6.11}$$

et pour toute fonction convexe $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$,

$$\phi\big(P_t f(x)\big) \leq P_t\big(\phi(f)\big)(x)$$

dès que toutes ces quantités ont bien un sens.

On a, par convergence dominée, que

$$\forall x \in \mathbb{R}^n, \qquad P_t f(x) \overset{t \to +\infty}{\longrightarrow} \int f\, d\gamma_n. \tag{6.12}$$

Examinons maintenant le comportement vis-à-vis de l'intégrale gaussienne. On a

$$\forall t \geq 0, \qquad \int P_t f\, d\gamma_n = \int f\, d\gamma_n. \tag{6.13}$$

On dit que la mesure gaussienne est invariante par P_t. On a plus généralement,

$$\forall t \geq 0, \qquad \int P_t(f)\, g\, d\gamma_n = \int f\, P_t(g)\, d\gamma_n, \tag{6.14}$$

ce qui est une manière de traduire que L (et donc P_t) est autoadjoint. Notez qu'on tire alors de (6.10) que

$$\forall p \geq 1,\ \forall t \geq 0, \qquad \|P_t(f)\|_{L_p(\gamma_n)} \leq \|f\|_{L_p(\gamma_n)}, \tag{6.15}$$

c'est-à-dire que P_t est une contraction de $L_p(\gamma_n)$.

Signalons enfin la propriété de commutation de la dérivée (gradient) et de P_t (en fait cela provient de la commutation entre le gradient et L) : pour $t > 0$,

$$\forall i \leq n, \qquad \frac{\partial}{\partial x_i}\Big|_x P_t f = e^{-t} P_t\big(\frac{\partial}{\partial x_i} f\big)(x),$$

ce que l'on résume en écrivant

$$\nabla P_t f = e^{-t} P_t(\nabla f), \tag{6.16}$$

où l'on convient que, dans le terme de gauche, le semi-goupe agit sur chaque cordonnée.

Remarque 6.4 **(Semi-groupe de la chaleur).** — *On peut aussi s'intéresser à l'évolution le long du semi-groupe de la chaleur. Pour $f : \mathbb{R}^n \to \mathbb{R}$ continue (et bornée, ou à croissance lente), il existe $F : [0, +\infty[\times \mathbb{R}^n \longrightarrow \mathbb{R}$ continue sur $[0, +\infty[\times\mathbb{R}^n$ et C^∞ sur $]0, +\infty[\times\mathbb{R}^n$ telle que*

$$F(0, \cdot) = f(\cdot) \quad \text{et} \quad \forall t > 0,\ \forall x \in \mathbb{R}^n, \quad \frac{\partial F}{\partial t}(t, x) = \Delta_x F(t, x).$$

On note $Q_t(f)(\cdot) := F(t, \cdot) = (e^{t\Delta} f)(\cdot)$. Ici encore, on peut définir Q_t explicitement par

$$Q_t(f)(x) = \int_{\mathbb{R}^n} f(x + \sqrt{2t}\, y)\, d\gamma_n(y).$$

Certaines des propriétés précédentes sont encore valables pour Q_t. En particulier, on a (6.7)-(6.8)-(6.10)-(6.9). *On a encore une écriture sous forme de noyau de convolution (de probabilité) :*

$$Q_t(f)(x) = \int_{\mathbb{R}^n} f(z) \frac{e^{-|z-x|^2/4t}}{(4\pi t)^{n/2}}\, dz,$$

et on a donc aussi (6.11). *Les TROIS grandes différences sont les suivantes. Tout d'abord, la mesure gaussienne n'est pas invariante. En fait, c'est la mesure de Lebesgue (qui n'est pas une probabilité) qui est invariante,*

$$\int g\, Q_t(f) = \int Q_t(g)\, f.$$

puisque pour des fonctions suffisamment régulières et intégrables (par exemple C^2 à support compact) on a

$$\frac{d}{dt} \int g\, Q_t(f) = \int g\, \Delta(Q_t(f)) = -\int \nabla g \cdot \nabla Q_t(f).$$

Ensuite, la comportement asymptotique est différent, puisqu'il n'y a pas de limite finie. On peut montrer que, pour x fixé, on a, lorsque $t \to +\infty$,

$$Q_t(f)(x) \sim \frac{1}{(4\pi t)^{n/2}} \int f.$$

Enfin, la commutation du gradient et du semi-groupe est différente puisque on a

$$\nabla Q_t(f)(x) = Q_t(\nabla f)(x).$$

Remarque 6.5 **(Semi-groupe d'Ornstein-Uhlenbeck généralisé)**

Soit $V : \mathbb{R}^n \to \mathbb{R}$ régulier pour lequel il existe des constantes $C, K > 0$ tel que

$$\forall |x| \geq K, \quad V(x) \geq C\, |x|.$$

De manière équivalente, on peut demander qu'il existe des constantes c_1, c_2 avec $c_2 > 0$ tel que $\forall x \in \mathbb{R}^n$ on ait $V(x) \geq c_1 + c_2\, |x|$. Cela est vérifié si V est convexe avec $V(x) \to +\infty$ lorsque $|x| \to +\infty$. Alors, quitte à ajouter à V une constante, on peut supposer que $\int e^{-V} = 1$. On définit alors la mesure de probabilité μ sur $\mathbb{R}^n$ par

$$d\mu(x) = e^{-V(x)}\, dx.$$

On on a encore que $\mathcal{A}$ est dense dans $L_p(\mu)$ pour tout $p \geq 1$. On introduit l'opérateur différentiel

$$Lf = \Delta f - \nabla V \cdot \nabla f$$

C'est l'opérateur laplacien naturel sur $L_2(\mu)$ puisque pour $f, g \in \mathcal{A}$ on vérifie que par intégration par parties

$$\int (Lf)g\, d\mu = -\int \nabla f \cdot \nabla g\, d\mu.$$

En particulier $\int Lf\, d\mu = 0$ et les fonctions constantes constituent exactement le noyau de L dans $L_2(\mu)$.

Si f est continue (et à croissance lente, de sorte que $f \in L_2(\mu)$), alors on peut définir $P_t f = e^{tL} f$ continue sur $[0, +\infty] \times \mathbb{R}^n$ et C^∞ sur $]0, +\infty[\times\mathbb{R}^n$ tel que

$$\partial_t(P_f) = L_x(P_t f).$$

À la différence du semi-groupe d'Ornstein-Uhlenbeck, on n'a pas de formule explicite pour $P_t f$ lorsque V n'est pas une forme quadratique. Mais de nombreuses propriétés essentielles restent vraies. En particulier

- *P_t est un semi-groupe (P_t est linéaire et $P_t \circ P_s = P_{t+s}$).*
- *P_t est Markovien : pour $t > 0$ et $x \in \mathbb{R}^n$, il existe une probabilité $\nu_{t,x}$ sur $\mathbb{R}^n$ tel que $P_t f(x) = \int_{\mathbb{R}^n} f(y)\, d\nu_{t,x}(y)$.*
- *μ est invariante pour P_t : $\int P_t f\, d\mu = \int f\, d\mu$ et plus généralement symétrique :*

$$\int (P_t f) g\, d\mu = \int f P_t g\, d\mu.$$

- $\|P_t\|_{L^p(\mu) \to L^p(\mu)} = 1$
- *$P_t f(x) \longrightarrow \displaystyle\int f\, d\mu$ lorsque $t \to +\infty$ pour tout $x \in \mathbb{R}^n$.*

Par contre, on a pas de formule pour calculer $\nabla P_t f$ et de fait, les relations de commutations entre le gradient et P_t sont plus compliquées (il n'y a pas d'égalités mais seulement des inégalités ponctuelles) et bien plus difficiles à obtenir.

6.2. Généralités sur l'inégalité de Sobolev logarithmique

Soit (X, μ) un espace de probabilité. Une fonction positive $f : X \longrightarrow \mathbb{R}^+$ est dite d'entropie finie si f et $f \log f$ sont dans $L_1(\mu)$, et on définit alors l'entropie de f par rapport à μ par

$$\begin{aligned}
\mathrm{Ent}_\mu(f) &:= \int f \log(f)\, d\mu - \left(\int f\, d\mu\right) \log\left(\int f\, d\mu\right) \\
&= \int f \log\left(\frac{f}{\int f\, d\mu}\right) d\mu \\
&= \int f \log(f)\, d\mu \quad \text{si } \int f\, d\mu = 1.
\end{aligned}$$

Puisque la fonction $s \longrightarrow s\log(s)$ est strictement convexe sur $\mathbb{R}^+$, on a

$$\mathrm{Ent}_\mu(f) \geq 0 \quad \text{et} \quad \mathrm{Ent}_\mu(f) = 0 \Leftrightarrow f \equiv \int f\, d\mu \ \ (\mu - pp).$$

On veut justement utiliser l'entropie pour quantifier que f est proche de sa moyenne sur un ensemble de grande mesure.

On va maintenant travailler avec une mesure de probabilité μ sur $\mathbb{R}^n$ équipé de sa structure euclidienne usuelle. Toutes les probabilités considérées seront supposées être absolument continues par rapport à la mesure de Lebesgue.

Rappelons (théorème de Rademacher) qu'une fonction localement lipschitzienne est différentiable presque partout et sa différentielle (presque partout) est une fonction localement intégrable qui coïncide avec la différentielle au sens des distributions. En particulier, si f est localement lipschitzienne avec $|\nabla f| \in L_1(\gamma_n)$, alors $P_t(\nabla f) \longrightarrow \nabla f$ lorsque $t \to 0$.

Definition 6.6 **(LSI(ρ_0))**. — *On dit qu'une probabilité μ vérifie une inégalité de Sobolev logarithmique avec constante $\rho_0 > 0$ (en abrégé « LSI(ρ_0) ») si pour toute fonction $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ suffisamment régulière (i.e. g^2 d'entropie finie et g est de classe C^1 ou localement lipschitzienne) on a :*

$$\mathrm{Ent}_\mu(g^2) \leq \frac{2}{\rho_0} \int |\nabla g|^2 \, d\mu. \tag{6.17}$$

On préfère parfois énoncer le résultat de la manière équivalente suivante : pour toute fonction positive $f : \mathbb{R}^n \longrightarrow \mathbb{R}^+$ d'entropie finie avec $\sqrt{f}$ de classe C^1 (ou localement lipschitzienne), on a

$$\mathrm{Ent}_\mu(f) \leq \frac{1}{2\rho_0} \int \frac{|\nabla f|^2}{f} \, d\mu. \tag{6.18}$$

L'intégrale de droite s'appelle *l'information de Fisher de f par rapport à μ*

Remarque 6.7. — *Notons que si $g := \sqrt{f}$ est de classe C^1, alors,*

$$2g\nabla g = \nabla f$$

et donc $\dfrac{|\nabla f|^2}{f}$ est interprété, en un point où f s'annule, comme le prolongement (par continuité) $4|\nabla g|^2$. Par ailleurs, puisque g est positive, on a $f(x) = 0 \Leftrightarrow g(x) = 0 \Rightarrow \nabla g(x) = 0$, et donc, on conclut qu'on a toujours

$$\int \frac{|\nabla f|^2}{f} \, d\mu = \int_{f>0} \frac{|\nabla f|^2}{f} \, d\mu.$$

Le résultat reste vrai lorsque $g = \sqrt{f}$ est localement lipschitzienne, puisqu'on a alors $\nabla g = 0$ presque partout sur l'ensemble $\{g = 0\}$.

On cherche à vérifier $LSI(\rho_0)$ avec la plus grande constante ρ_0 possible.

Le résultat suivant indique qu'une inégalité de log-Sobolev entraîne de la concentration gaussienne.

Théorème 6.8. — *Soit μ une probabilité vérifiant* LSI(ρ_0). *Alors, pour toute fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ 1-lipschitzienne on a :*

$$\forall \lambda \geq 0, \qquad \int e^{\lambda \left[f - \int f \, d\mu \right]} \, d\mu \leq e^{\lambda^2 / 2\rho_0}.$$

On a donc, pour toute fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ 1-lipschitzienne

$$\forall r \geq 0, \qquad \mu\left(\left\{ x \in \mathbb{R}^n \, ; f(x) \geq \int f \, d\mu + r \right\} \right) \leq e^{-\rho_0 r^2 / 2},$$

et $\alpha_{(\mathbb{R}^n,|\cdot|,\mu)}(r) \leq e^{-\rho_0\, r^2/8}$.

Démonstration. — On introduit, pour $\lambda \geq 0$,

$$H(\lambda) = \int e^{\lambda\left[f - \int f\, d\mu\right]}\, d\mu.$$

En appliquant (6.17) à $g = e^{\lambda f/2}$ et en utilisant que $|\nabla f| \leq 1$ (presque partout si f est seulement lipschitzienne) on obtient,

$$\lambda H'(\lambda) - H(\lambda)\log(H(\lambda)) \leq \frac{1}{2\rho_0}\lambda^2 H(\lambda),$$

soit encore, pour $\lambda > 0$,

$$\frac{1}{\lambda}\cdot\frac{H'(\lambda)}{H(\lambda)} - \frac{1}{\lambda^2}\cdot\log(H(\lambda)) \leq \frac{1}{2\rho_0},$$

Par ailleurs, un DL en $\lambda = 0$ donne que $\frac{1}{\lambda}\log H(\lambda) \to \int f\, d\mu$ lorsque $\lambda \to 0$. Par conséquent, en intégrant l'inégalité précédente, on obtient

$$\frac{1}{\lambda}\cdot\log H(\lambda) - \int f\, d\mu \leq \frac{1}{2\rho_0}\cdot\lambda,$$

ce qui donne le résultat annoncé. $\square$

Une propriété importante de l'inégalité de Sobolev logarithmique est qu'elle se « tensorise ».

Théorème 6.9. — *Soit μ_1 et μ_2 deux probabilités sur $\mathbb{R}^{n_1}$ et $\mathbb{R}^{n_2}$, respectivement. Si μ_1 vérifie* LSI(ρ_1) *et μ_2 vérifie* LSI(ρ_2), *alors $\mu_1 \otimes \mu_2$ vérifie* LSI($\min(\rho_1, \rho_2)$).

En particulier, si μ vérifie $LSI(\rho)$, alors $\mu^{\otimes N}$ vérifie aussi $LSI(\rho)$ pour tout $N \geq 1$.

Ce qu'il y a d'important dans ce dernier énoncé, c'est que la constante est indépendante de la dimension N.

Démonstration. — Soit $G : \mathbb{R}^{n_1+n_2} \longrightarrow \mathbb{R}$ régulière. Par Fubini, on écrit (astucieusement)

$$\begin{aligned}\mathrm{Ent}_{\mu_1\otimes\mu_2}(G^2) = \int \left(\int G^2 \log\left(\frac{G^2}{\int G^2\, d\mu_1}\right) d\mu_1\right) d\mu_2 \\ + \int \left(\int G^2\, d\mu_1\right) \log\left(\frac{\int G^2\, d\mu_1}{\int G^2\, d\mu_1 d\mu_2}\right) d\mu_2.\end{aligned}$$

Dans la première intégrale, on applique LSI(ρ_1) à la fonction $x \to G(x,y)$, pour $y \in \mathbb{R}^{n_2}$ fixé. Dans la deuxième intégrale, on applique LSI(ρ_2) sous la forme (6.18) à la fonction $f(y) = \int G^2(x,y)\, d\mu_1(x)$. On obtient donc, en écrivant ∇_x pour le vecteur formé par les n_1 premières dérivées partielles (idem pour ∇_y),

$$\mathrm{Ent}_{\mu_1\otimes\mu_2}(G^2) \leq \frac{2}{\rho_1}\iint |\nabla_x G|^2\, d\mu_1 d\mu_2 + \frac{1}{2\rho_2}\int \frac{\left|\nabla_y \int G^2\, d\mu_1\right|^2}{\int G^2\, d\mu_1}\, d\mu_2.$$

On conclut en dérivant sous l'intégrale puis utilisant l'inégalité de Cauchy-Schwarz, pour $y \in \mathbb{R}^{n_2}$ fixé,

$$\left|\nabla_y \int G^2\, d\mu_1\right|^2 = 4\left|\int G\nabla_y G\, d\mu_1\right|^2 \leq \int G^2\, d\mu_1 \int |\nabla_y G|^2\, d\mu_1,$$

et enfin le fait que $|\nabla_x G|^2 + |\nabla_y G|^2 = |\nabla G|^2$. □

De manière plus systématique, on montre (voir exercice ci-dessous) une inégalité de tensorisation pour l'entropie, de laquelle la tensorisation de log-Sobolev découle immédiatement (en n'utilisant que la forme (6.17) de log-Sobolev, ce qui sera important dans le cas discret).

***Remarque 6.10* (Rappel : Tensorisation de l'entropie).** — *Soit* $(\Omega_1, \mu_1), \dots, (\Omega_n, \mu_n)$ *des espaces de probabilité. On note* $\Omega = \Omega_1 \times \dots \Omega_n$ *et* $P = \mu_1 \otimes \dots \otimes \mu_n$. *Pour* $x = (x_1, \dots, x_n) \in \Omega$ *et* $i \le n$, *on notera* $\hat{x}_i = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. *Pour* $f : \Omega \longrightarrow \mathbb{R}$ *et* $i \le n$, *on introduit* $f_i : \Omega_i \longrightarrow \mathbb{R}$,

$$f_i(x_i) := f_i^{\hat{x}_i}(x_i) := f(x_1, \dots, x_n)$$

lorsque $\hat{x}_i$ *est fixé. Alors, pour* $f \ge 0$ *on peut définir, pour* $z \in \Omega$,

$$\begin{aligned} \mathrm{Ent}_{\mu_i}(f_i)(z) &:= \int f_i^{\hat{z}_i}(t) \log \frac{f_i^{\hat{z}_i}(t)}{\int f_i^{\hat{z}_i}\, d\mu_i} \, d\mu_i(t) \\ &= \int f(z_1, \dots, z_{i-1}, z_i, z_{i+1}, z_n) \log \frac{f(z_1, \dots, z_{i-1}, z_i, z_{i+1}, z_n)}{\int f(z_1, \dots, z_{i-1}, s, z_{i+1}, z_n)\, d\mu_i(s)} \, d\mu_i(z_i), \end{aligned}$$

qui ne dépend donc que de $\hat{z}_i$ *(*$\mathrm{Ent}_{\mu_i}(f_i)(z) = \mathrm{Ent}_{\mu_i}(f_i)(\hat{z}_i)$ *: on a intégré* $f(z) \log f(z)$ *par rapport à* $d\mu_i(z_i)$.

L'inégalité de tensorisation de l'entropie dit que :

$$\mathrm{Ent}_P(f) \le \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(f_i) \, dP.$$

En utilisant la tensorisation de l'entropie, on obtient directement le résultat précédent sur la tensorisation de l'inégalité de log-Sobolev, puisque du coté gradient, on aura, sur un produit d'espaces euclidiens $\Omega_i = \mathbb{R}^{n_i}$, *que* $|\nabla f|^2 = \sum |\nabla_i f|^2$.

6.3. L'inégalité de Sobolev logarithmique gaussienne

Théorème 6.11. — *La mesure gaussienne* γ_n *sur* $\mathbb{R}^n$ *vérifie* LSI(1).

Démonstration. — Soit $f : \mathbb{R}^n \to \mathbb{R}$ telle que $\sqrt{f}$ soit C^1 (localement lipschitzienne suffit) avec $\int f \, d\gamma_n = 1$. Soit $(P_t f)_{t \ge 0}$ le semi-groupe d'Ornstein-Uhlenbeck correspondant. Notez qu'on a $\int P_t f \, d\gamma_n = \int f \, d\gamma_n = 1$. Si on pose, pour $t \ge 0$,

$$S(t) = \int (P_t f) \log(P_t f) \, d\gamma_n = \mathrm{Ent}_{\gamma_n}(P_t f)$$

on a

$$S(0) = \mathrm{Ent}_{\gamma_n}(f) \quad \text{et} \quad \lim_{t \to +\infty} S(t) = 0,$$

et par conséquent

$$\mathrm{Ent}_{\gamma_n}(f) = -\int_0^{+\infty} S'(t) \, dt.$$

Puisque $\frac{d}{dt}P_t f = L(P_t f)$, on a, par intégration par parties,

$$\begin{aligned} -S'(t) &= -\int \frac{d}{dt}\Big[(P_t f)\log(P_t f)\Big]\, d\gamma_n \\ &= -\int \big(1+\log(P_t f)\big) L\big(P_t f\big)\, d\gamma_n \\ &= \int \frac{|\nabla P_t f|^2}{P_t f}\, d\gamma_n. \end{aligned}$$

On voit donc au passage que, le long du semi-groupe d'Ornstein-Uhlenbeck, l'entropie décroît et que sa dérivée est (l'opposé) de l'information de Fisher. Pour $t > 0$ et $x \in \mathbb{R}^n$ fixés on a (la notation vectorielle s'entend coordonnée par coordonnée)

$$|\nabla P_t f|^2 = |e^{-t} P_t \nabla f|^2 = e^{-2t}|P_t \nabla f|^2$$

et en utilisant Cauchy-Schwarz (6.11)

$$(P_t \partial_i f)^2 \leq P_t\left(\frac{(\partial_i f)^2}{f}\right) P_t(f)$$

soit encore

$$\frac{|P_t \nabla f|^2}{P_t f} \leq P_t\left(\frac{|\nabla f|^2}{f}\right).$$

On obtient donc,

$$-S'(t) \leq e^{-2t}\int P_t\left(\frac{|\nabla f|^2}{f}\right) d\gamma_n = e^{-2t}\int \frac{|\nabla f|^2}{f}\, d\gamma_n,$$

puisque γ_n est invariante par P_t. On a donc

$$\mathrm{Ent}_{\gamma_n}(f) \leq \left(\int_0^{+\infty} e^{-2t}\, dt\right)\int \frac{|\nabla f|^2}{f}\, d\gamma_n = \frac{1}{2}\int \frac{|\nabla f|^2}{f}\, d\gamma_n.$$

□

On verra en exercice d'autres preuves de l'inégalité de log-Sobolev gaussienne.

6.4. Isopérimétrie gaussienne

6.4.1. Généralité sur l'isopérimétrie. — Commençons par quelques généralités sur le problème isopérimétrie. Soit (X, d, μ) un espace métrique muni d'une mesure borélienne. Pour $A \subset X$ on notera

$$A_r := \{x \in X \;;\; d(x, A) \leq r\}$$

et la μ-mesure du bord ∂A (pour A borélien, bien sûr) sera par définition

$$\mu^+(A) := \liminf_{r \to 0} \frac{\mu(A_r) - \mu(A)}{r}.$$

Cette limite inf existe toujours, mais elle peut être infinie : $\mu^+(A) \in [0, +\infty]$

Le problème isopérimétrique consiste à trouver, parmi les ensembles $A \subset X$ de mesure $\mu(A)$ fixée, celui qui a le plus petit $\mu^+(A)$. En général, ce problème est beaucoup trop difficile, et on cherche au moins une fonction $F : \mathbb{R}^+ \to \mathbb{R}^+$ telle que, pour tout $A \subset X$,

$$\mu^+(A) \geq F\big(\mu(A)\big). \tag{6.19}$$

La fonction isopérimétrique I_μ est définie sur $\mathbb{R}^+$ par

$$\forall s \geq 0, \qquad I_\mu(s) := \inf \left\{ \mu^+(A) \ ; \ \ A \subset X \text{ avec } \mu(A) = s \right\}$$

La connaissance exacte de I_μ est aussi difficile que la résolution du problème isopérimétrique (même si ce dernier revient non seulement à connaître I_μ mais aussi à déterminer des A pour lequel l'infimum précédent est atteint). En général, on se contente d'une majoration $I_\mu \geq F$, ce qui équivaut à (6.19). De fait, la fonction I_μ est la meilleure fonction F pour laquelle on a (6.19).

Il est bon de noter que, pour $A \subset X$ non-vide donné, la fonction $r \longrightarrow \mu(A_r)$ est continue à droite sur $]0, +\infty[$ car $A_r = \cap_{r'>r} A_{r'}$; notez que $A_0 = \overline{A}$ et plus généralement que

$$A_r = (\overline{A})_r.$$

Le résultat suivant confirme qu'on peut se restreindre aux ensembles fermés.

Fait 6.12. — *Si $\mu^+(A) < +\infty$, alors $\mu(A) = \mu(\overline{A})$ et $\mu^+(A) = \mu^+(\overline{A})$.*

Démonstration. — Pour tout $r \geq 0$, on a $A_r \supset \overline{A}$, et donc

$$\mu^+(A) = \liminf_{r \to 0} \frac{\mu(A_r \setminus A)}{r} \geq \liminf_{r \to 0} \frac{\mu(\overline{A} \setminus A)}{r}.$$

Si $\mu^+(A) < +\infty$, on en tire que $\mu(\overline{A} \setminus A) = 0$, c'est-à-dire que $\mu(A) = \mu(\overline{A})$. □

Sur l'espace euclidien $\mathbb{R}^n$ muni de la mesure de Lebesgue (notée λ), on a déjà résolu le problème isopérimétrique : les boules euclidiennes sont extrémales, et on a $I_\lambda(t) := c_n t^{\frac{n-1}{n}}$ pour tout $t \geq 0$, où c_n est une constante dépendant de n ($c_n = |\partial B_2^n| \cdot |B_2^n|^{-\frac{n-1}{n}}$.)

Lorsque μ est une probabilité sur X, alors $I_\mu(t)$ est nulle pour $t > 1$, et donc on convient que I_μ est définie sur $[0, 1]$. On a alors $I_\mu(0) = I_\mu(1) = 0$, et (modulo un minimum de régularité) on a que $I_\mu(t) = I_\mu(1 - t)$ pour tout $t \in [0, 1]$ (symétrie par rapport à $t = 1/2$, ce qui traduit le passage de A à son complémentaire A^c).

6.4.2. L'isopérimétrie gaussienne. — Un demi-espace (fermé) de $\mathbb{R}^n$ est un ensemble de la forme

$$H = H_u(t) := \{x \in \mathbb{R}^n \ ; \ x \cdot u \leq t\}$$

où $u \in \mathbb{R}^n$ est choisi de norme 1 et $t \in \mathbb{R}$. Tous les demi-espaces s'obtiennent par une rotation d'un demi-espace de la forme $H(t) := \{x_1 \leq t\} =] - \infty, t] \times \mathbb{R}^{n-1}$. En particulier, tous les demi-espaces de même mesure gaussienne s'obtiennent par des rotations d'un même espace.

Théorème 6.13. — *À mesure gaussienne fixé, les demi-espaces minimisent la mesure gaussienne du bord. En d'autres termes, pour $A \subset \mathbb{R}^n$, si H est un demi-espace tel que $\gamma_n(A) = \gamma_n(H)$ alors*

$$\gamma_n^+(A) \geq \gamma_n^+(H).$$

Ce résultat répond donc à la question de l'isopérimétrie gaussienne. On va en voir des formulation équivalentes.

Théorème 6.14. — *Soit $A \subset \mathbb{R}^n$ et H est un demi-espace tel que $\gamma_n(A) = \gamma_n(H)$. Alors,*

$$\forall r \geq 0, \qquad \gamma_n(A_r) \geq \gamma_n(H_r).$$

Ce résultat permet de retrouver l'inégalité de concentration gaussienne. En effet, soit $A \subset \mathbb{R}^n$ et introduisons $H =]-\infty, t] \times \mathbb{R}^{n-1}$ tel que $\gamma_n(A) = \gamma_n(H) = \int_{-\infty}^t e^{-v^2/2} \frac{dv}{\sqrt{2\pi}}$. Si l'on suppose que $\gamma_n(A) \geq \frac{1}{2}$ alors $t \geq 0$ et on a par conséquent, pour tout $r \geq 0$

$$\gamma_n(A_r) \geq \gamma_n(H_r) = \int_{-\infty}^{t+r} e^{-v^2/2} \frac{dv}{\sqrt{2\pi}} = 1 - \int_{t+r}^{+\infty} e^{-v^2/2} \frac{dv}{\sqrt{2\pi}} \geq 1 - \int_{r}^{+\infty} e^{-v^2/2} \frac{dv}{\sqrt{2\pi}}$$
$$\geq 1 - e^{-r^2/2}$$

Le Théorème 6.14 implique clairement le Théorème 6.13, mais en fait il lui est complètement équivalent, comme le montre un exercice plus loin.

La particularité des demi-espaces, c'est que ce sont des ensembles 1-dimensionnel. Introduisons la fonction $\phi : \mathbb{R} \longrightarrow [0,1]$,

$$\phi(t) := \gamma_1(]-\infty, t]) = \int_{-\infty}^{t} e^{-v^2/2} \frac{dv}{\sqrt{2\pi}}$$

qui est un difféomorphisme croissant de $[-\infty, +\infty]$ sur $[0,1]$. On a $\phi(t) = \gamma_n(\{x_1 \leq t\})$ et plus généralement, pour tout demi-espace $H_u(t)$,

$$\forall t \in \mathbb{R}, \qquad \gamma_n(H_u(t)) = \phi(t),$$

ce que l'on peut aussi résumer en disant que, pour tout demi-espace H de $\mathbb{R}^n$

$$\gamma_n(H) = \gamma_n(H_u(t)) \Longleftrightarrow t = \phi^{-1}\big(\gamma_n(H)\big).$$

Pour un demi-espace, on a

$$\forall r \geq 0, \quad (H_u(t))_r = H_u(t+r),$$

et donc

$$\forall t \in \mathbb{R}, \qquad \gamma_n^+\big(H_u(t)\big) = \phi'(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}} =: \varphi(t).$$

Introduisons la fonction $I : [0,1] \longrightarrow \mathbb{R}^+$:

$$\forall s \in [0,1], \qquad I(s) := \varphi \circ \phi^{-1}(s) = \phi'\big(\phi^{-1}(t)\big).$$

Alors, d'après les considération précédente, que pour tout demi-espace H,

$$\gamma_n^+(H) = I\big(\gamma_n(H)\big)$$

Ainsi le Théorème 6.13 est strictement équivalent au résultat suivant :

Théorème 6.15. — *On a : $I_{\gamma_n} = I$, soit encore, pour tout $A \subset \mathbb{R}^n$,*

$$\gamma_n^+(A) \geq I(\gamma_n(A))$$

avec égalité lorsque A est un sous-espace.

De même, le Théorème 6.14 peut se formuler comme suit : pour tout $A \subset \mathbb{R}^n$ on a

$$\forall r \geq 0, \qquad \gamma_n(A_r) \geq \phi\big(\phi^{-1}(\gamma_n(A)) + r\big).$$

La fonction isopérimétrique gaussienne $I : [0,1] \to \mathbb{R}^+$ a de nombreuses propriétés utiles. Notons que $I(0) = I(1) = 0$ et que pour $I(s) = I(1-s)$. La fonction Iest régulière sur $]0,1[$ et pour tout $s \in]0,1[$ on a

$$I'(s) = -\phi^{-1}(s)$$

ainsi que la relation essentielle suivante :

$$\forall s \in]0,1[, \qquad I''(s) = -\frac{1}{I(s)}.$$

Exercice 6.1. — *Montrer que lorsque $s \to 0$, $I(s) \sim s\sqrt{2\log\frac{1}{s}}$.*

L'inégalité isopérimétrique gaussienne admet une forme fonctionnelle due à S. Bobkov (anticipée, en fait, par A.Ehrhard).

***Théorème 6.16* (Inégalité de Ehrhard-Bobkov)**. — *Pour toute fonction $f : \mathbb{R}^n \longrightarrow [0,1]$ de classe C^1 ou localement lipschitzienne, on a :*

$$I\left(\int f\, d\gamma_n\right) \leq \int \sqrt{I^2(f) + |\nabla f|^2}\, d\gamma_n. \tag{6.20}$$

Expliquons pourquoi ce résultat entraîne le Théorème 6.13. Soit $A \subset \mathbb{R}^n$ un ensemble fermé. Introduisons pour $\varepsilon \in]0,1[$,

$$\forall x \in \mathbb{R}^n, \quad f_\varepsilon(x) = h_\varepsilon\big(d(x,A)\big) \in [0,1]$$

où $h_\varepsilon : \mathbb{R}^+ \to \mathbb{R}^+$ est la fonction continue affine par morceaux valant 1 sur $[0,\varepsilon^2]$, 0 sur $[\varepsilon,+\infty[$, et affine sur $[\varepsilon^2,\varepsilon]$. On a, ponctuellement,

$$\lim_{\varepsilon\to 0} f_\varepsilon = \mathbf{1}_A.$$

Par ailleurs, la fonction f_ε est lipschitzienne sur $\mathbb{R}^n$, de constante de Lipschitz majorée par la constante de Lipschitz de h_ε, à savoir $\dfrac{1}{\varepsilon-\varepsilon^2}$. Comme f_ε est constante sur les ouverts $\{d(\cdot,A) < \varepsilon^2\} \supset A$ et $\mathbb{R}^n \setminus A_\varepsilon$, on en déduit que

$$|\nabla f_\varepsilon| \leq \frac{1}{\varepsilon-\varepsilon^2}\,\mathbf{1}_{A_\varepsilon\setminus A}.$$

En appliquant l'inégalité (6.20) à f_ε, suivi de l'inégalité $\sqrt{a^2+b^2} \leq |a|+|b|$, on obtient donc,

$$I\left(\int f_\varepsilon\, d\gamma_n\right) \leq \int \big[I(f_\varepsilon) + |\nabla f_\varepsilon|\big]\, d\gamma_n \leq \int I(f_\varepsilon)\, d\gamma_n + \frac{1}{\varepsilon-\varepsilon^2}\gamma_n(A_\varepsilon \setminus A).$$

Quand $\varepsilon \to 0$, le terme de gauche a une limite, et cette limite minore donc la limite inf du terme de droite. On a donc, en utilisant que $I(f_\varepsilon) \to I(\mathbf{1}_A) \equiv 0$, que

$$I(\gamma_n(A)) \leq \gamma_n^+(A),$$

ce qui redonne bien l'inégalité isopérimétrique.

Remarque 6.17. — *Il est plus courant de prendre la fonction*

$$f_\varepsilon(x) := \big(1 - \frac{1}{\varepsilon} d(x,A)\big)_+,$$

pour laquelle le raisonnement est identique.

On peut aussi montrer que, réciproquement, l'inégalité isopérimétrique (appliquée à un ensemble bien choisi de $(\mathbb{R}^{n+1}, \gamma_{n+1})$, pour être précis) permet de retrouver directement l'inégalité (6.20).

6.4.3. Preuve de l'inégalité fonctionnelle. — Soit $f : \mathbb{R}^n \to [0,1]$ non-nulle et non-identiquement égale à 1 (sinon il n'y a rien à montrer). On suppose que f est localement lipschitzienne avec $\int |\nabla f|^2 \, d\gamma_n < +\infty$. Pour $t > 0$, on a $P_t f(x) \in]0,1[$, ce qui permet de garantir que I est régulière aux points $P_t f(x)$. Posons, pour $t \geq 0$,

$$J(t) := \int \sqrt{I^2(P_t f) + |\nabla P_t f|^2} \, d\gamma_n.$$

On a $J(0) = \int \sqrt{I^2(f) + |\nabla f|^2} \, d\gamma_n$, et comme, lorsque $t \to +\infty$, $P_t f \to \int f \, d\gamma_n$ et $|\nabla P_t f| = e^{-t} |P_t \nabla f| \to 0 \times |\int \nabla f \, d\gamma_n| = 0$, on a

$$\lim_{t \to +\infty} J(t) = I\left(\int f \, d\gamma_n\right).$$

Ainsi, l'inégalité (6.20) sera démontrée si on montre que J est décroissante sur $\mathbb{R}^+$.

Pour simplifier les notations (!) on va faire le calcul dans le cas de la dimension 1 (voir plus loin pour le cas n-dimensionnel). On travaille donc avec le semi-groupe de générateur $Lf(x) = f''(x) - xf'$ et on notera $\gamma = \gamma_1$. La notation g' sera toujours la dérivée en espace. Pour toute fonction $g : \mathbb{R} \to [0,1]$ de classe C^1, on note

$$K(g) := I^2(g) + g'^2,$$

de sorte que $J(t) = \int \sqrt{K(P_t f)} \, d\gamma$. Pour $t > 0$ on a, par commutation de ∂_x et ∂_t (i.e. $(\partial_t g)' = \partial_t g'$),

$$J'(t) = \int \frac{1}{\sqrt{K(P_t f)}} \left[(II')(P_t f) L(P_t(f)) + (P_t f)'(L(P_t f))'\right] d\gamma.$$

On pose, t étant dorénavant fixé, $g = P_t f$, et on doit donc estimer

$$J'(t) = \int \frac{(II')(g)}{\sqrt{K(g)}} Lg \, d\gamma + \int \frac{1}{\sqrt{K(g)}} g'(Lg)' \, d\gamma.$$

La première intégrale donne, après intégration par parties,

$$\begin{aligned}
\int \frac{(II')(g)}{\sqrt{K(g)}} Lg \, d\gamma &= -\int g' \left(-\frac{II'(g)}{K(g)^{3/2}}\left[(II')(g)g' + g'g''\right] + \frac{1}{\sqrt{K(g)}}\left[I'^2(g)g' + II''(g)g'\right]\right) d\gamma \\
&= \int \frac{II'(g)}{K(g)^{3/2}}\left[(II')(g)g'^2 + g'^2 g''\right] d\gamma - \int \frac{1}{\sqrt{K(g)}}\left[I'^2(g)g'^2 - g'^2\right] d\gamma
\end{aligned}$$

où l'on a utilisé que $II'' \equiv -1$ (ce sera le seul moment où une propriété de I sera utilisée). Pour calculer la deuxième intégrale on remarque que

$$(Lg)' = L(g') - g' \tag{6.21}$$

et donc, par intégration par parties,

$$\begin{aligned}\int \frac{1}{\sqrt{K(g)}} g'(Lg)'\,d\gamma &= -\int \left(\frac{1}{\sqrt{K(g)}} g'\right)' g''\,d\gamma - \int \frac{1}{\sqrt{K(g)}} g'^2\,d\gamma \\ &= \int \frac{II'(g)}{K(g)^{3/2}} \big[(II')(g)g' + g'g''\big] g''\,d\gamma - \int \frac{1}{\sqrt{K(g)}} g''^2\,d\gamma - \int \frac{1}{\sqrt{K(g)}} g'^2\,d\gamma.\end{aligned}$$

En mettant les deux calculs précédent ensembles, on trouve (après mise au même dénominateur et simplifications) :

$$\begin{aligned}J'(t) &= -\int \frac{1}{K(g)^{3/2}} \left[I'^2(g)g'^4 - 2(II')(g)g''g'^2 + I^2(g)g''^2\right] d\gamma \\ &= -\int \frac{1}{K(g)^{3/2}} \Big(I'(g)g'^2 - I(g)g''\Big)^2 d\gamma \leq 0.\end{aligned}$$

□

Dans le cas n-dimensionnel, le calcul est formellement le même modulo les points suivants. La commutation du gradient et de L, qui remplace (6.21), est :

$$\nabla g \cdot \nabla(Lg) = \sum_{i=1}^{n} \left[\partial i L(\partial_i g) - (\partial_i g)^2\right] = \nabla g \cdot L(\nabla g) - |\nabla g|^2,$$

où l'on convient donc que l'opérateur L agit (de manière identique) sur chaque coordonnée $\partial_i g$ de ∇g. Le calcul donne alors

$$\begin{aligned}J(t) = -\int \frac{1}{K(g)^{3/2}} \Big[& I'^2(g)|\nabla g|^4 - 2(II')(g)\operatorname{Hess} g(\nabla g)\cdot \nabla g + I^2(g)\|\operatorname{Hess} g\|_2^2 \\ & + \|\operatorname{Hess} g\|_2^2 |\nabla g|^2 - |\operatorname{Hess} g(\nabla g)|^2\Big] d\gamma_n\end{aligned}$$

où $\|A\|_2$ désigne la norme de Hilbert-Schmidt d'une matrice carré A :

$$\|A\|_2^2 = \sum_{i,j\leq n} a_{i,j}^2 = \operatorname{Tr}\big({}^tAA\big).$$

Les deux derniers termes dans l'intégrale, égaux en dimension 1, se contrôlent par l'inégalité de Cauchy-Schwarz,

$$|\operatorname{Hess} g(\nabla g)|^2 = \sum_j \left(\sum_i \partial_{i,j}^2 g\, \partial_i g\right)^2 \leq \|\operatorname{Hess} g\|_2^2 \cdot |\nabla g|^2.$$

On a donc :

$$\begin{aligned}J'(t) &\leq -\int \frac{1}{K(g)^{3/2}} \Big[I'^2(g)|\nabla g|^4 - 2(II')(g)\operatorname{Hess} g(\nabla g)\cdot\nabla g + I^2(g)\|\operatorname{Hess} g\|_2^2\Big] d\gamma_n \\ &= -\int \frac{1}{K(g)^{3/2}} \Big\| I'(g)\nabla g\, {}^t\nabla g - I(g)\operatorname{Hess} g\Big\|_2^2 d\gamma_n \leq 0,\end{aligned}$$

ce qui donne bien que J est décroissante sur $\mathbb{R}^+$.